Universidad de Valladolid
Facultad de Ciencias

Trabajo Fin de Grado

Grado en Estadística

Análisis de resultados electorales a nivel municipal con relación a factores económicos

Autora:

María Aragón Ruiz

Tutores:

Bonifacio Salvador González

Jesús M. Rodríguez Rodríguez

Curso 2020/2021

ÍNDICE

RESUMEN

El objetivo principal de este proyecto fue analizar los resultados correspondientes a las elecciones municipales del 26 de mayo de 2019 en Castilla y León con relación a factores económicos. Habiéndose obtenido los datos de fuentes oficiales, se buscaron las relaciones entre los dos tipos de variables (electorales y económicas), mediante diferentes técnicas multivariantes. Todos los resultados se han obtenido mediante el entorno de desarrollo R-Studio, incluyendo la utilización de la herramienta Excel para el tratamiento de los datos.

ABSTRACT

Main objective of this study was to analyse the results of the municipal elections of 26 May, 2019, held in Castilla y León in relation to economic factors. The data were obtained from official sources, relation between electoral and economic variables were found using different multivariate techniques. Results were obtained using R studio development environment and Excel for data processing.

Capítulo 1 INTRODUCCIÓN

La finalidad de este trabajo consiste en buscar relaciones, si es que existen, entre los resultados de las elecciones municipales del 26 de mayo de 2019 en Castilla y León, y diferentes factores económicos que se tendrán en cuenta.

Tanto la información de la parte electoral, como la correspondiente a los datos económicos, se ha obtenido a partir de fuentes oficiales, como son: el Instituto Nacional de Estadística (INE), la Agencia Tributaria, el Servicio Público de Empleo Estatal (SEPE) y la Junta de Castilla y León. Una vez recopilados los datos anteriores, se han unido mediante la variable que identifica a cada municipio, realizando en ella los cambios necesarios para que la tabla resultante sea correcta. Por último, se ha procedido a la depuración de los datos, corrigiendo los errores producidos después de unir los archivos individuales, aunque únicamente se han tenido en cuenta los 456 municipios de la comunidad autónoma cuyo número de habitantes es superior a 500, ya que en las elecciones municipales se ha supuesto que en municipios pequeños se suele votar más en función de las personas que de los partidos.

En aquellas variables que recogen resultados totales, como es evidente que presentan valores más altos en aquellos municipios con más habitantes, ha sido necesaria una transformación de los valores absolutos a relativos. Por ello, se han tomado los datos por cada 1.000 habitantes para todas las variables exceptuando las rentas, que recogen valores medios.

Debido al desequilibrio existente en el número de habitantes, se ha tenido en cuenta la siguiente clasificación en tres tipos diferentes de municipios:

- Rural: población menor que 2.000
- Semirural: población entre 2.000 y 10.000
- Urbano: población mayor que 10.000

Se observa que, aunque el primer grupo representa el 72% de los municipios totales, el segundo el 23% y el tercero el 5%, el porcentaje de población total que recoge cada uno de ellos es 15%, 21% y 64% respectivamente.

Además, se ha observado que el número de habitantes del municipio ejerce una gran influencia en el valor del resto de variables. Por ello, todos los análisis se han realizado tanto para el conjunto inicial, como para los tres conjuntos separados por tipos.

La primera técnica utilizada ha sido el Análisis de Componentes Principales, cuya finalidad es reducir el número de variables, intentando perder la menor cantidad de información posible. El propósito de este tipo de análisis dentro del trabajo ha sido encontrar de forma gráfica alguna de las relaciones buscadas.

Por otro lado, se ha aplicado un Análisis de Correlaciones Canónicas, que resulta una técnica adecuada cuando las variables disponibles pueden dividirse en dos grupos en función de algún criterio, como es el caso de este trabajo, que se han dividido en electorales y económicas. Su principal utilidad reside en estudiar las relaciones entre ambos conjuntos.

Al aplicarse los dos análisis anteriores, en el grupo inicial de municipios no se han encontrado relaciones de interés entre los dos tipos de variables.

Además, se ha realizado un Análisis Clúster y un ANOVA para contrastar la igualdad de medias en los clústeres obtenidos.

El Análisis Clúster tiene como finalidad obtener particiones de los datos de tal forma que los inviduos dentro de cada una de ellas sean lo más similares posibles entre sí y que difieran lo máximo posible con respecto a los municipios del resto de particiones. En este caso se utilizaron el método k-means y el método jerárquico, una de cuyas diferencias es que en el primero de ellos se debe establecer a priori el número de grupos que se espera obtener. En este trabajo se consideraron 8 clústeres en ambos métodos, puesto que es el número óptimo que se obtenía para k-means, seleccionando finalmente el clustering jerárquico como el más adecuado. La conclusión obtenida del análisis clúster ha sido que, como ya se había supuesto, la población influye en el resto de variables, ya que en el clúster 1 se encuentran agrupados casi todos los municipios rurales, mientras que los municipios rurales y semirurales se encuentran repartidos de forma más equitativa entre los otros 7 clústeres.

Para comparar las respuestas medias de cada variable en los 8 grupos obtenidos mediante el método jerárquico se ha realizado un ANOVA para cada variable disponible, contrastándose en cada modelo la hipótesis:

$$H0: \mu_1 = \mu_2 = \cdots = \mu_8$$

Encontrándo diferencias significativas entre las medias de los 8 grupos para todas las variables excepto para la renta por persona, y presentándose estas diferencias entre el cúster 1 y los demás, que como ya se ha explicado, es el que representa a los municipios rurales y demostrándose que la Población ejerce gran influencia sobre el resto de variables

Por lo tanto, al no haber encontrado relaciones sólidas entre las variables económicas y los resultados electorales, se han aplicado de nuevo el análisis de componentes principales y el análisis de correlaciones canónicas en los 3 tipos de municipio. Utilizando el primero de ellos, se ha observado que, para los municipios rurales y para los semirurales**,** no se aparecen relaciones claras, mientras que, en los urbanos, se establece que Otros comparte dirección de crecimiento con Paro Total y Total Empresas por Municipio, Izquierda con Total Contratos, Número Total Licencias y Número Trabajadores y Derecha con Número Establecimientos, Número Total Declaraciones y Población. Por otra parte, mediante el Análisis de Correlaciones Canónicas, no pueden establecerse conclusiones sólidas para los municipios rurales y semirurales, aunque para los municipios urbanos se establece que la primera variable canónica económica resulta un buen predictor para Derecha e Izquierda

Como último método para establecer las relaciones buscadas, se ha aplicado la técnica de regresión logística, cuya principal utilidad es que permite estimar la probabilidad de una variable categórica binaria en función variables númericas. Al no disponer de las variables categóricas necesarias para aplicar este tipo de modelo, se han creado tres nuevas variables: *MayoríaDerecha, MayoríaIzquierda* y *MayoríaOtros,* descartando aquellos municipios en los que la mayoría de voto era a otros partidos, de tal forma que MayoríaDerecha= 1-MayoríaIzquierda. Tampoco se han tenido en cuenta los municipios rurales, puesto que el hecho de que su número de habitantes sea pequeño podría enmascarar los resultados.

Los resultados obtenidos mediante la regresión logística muestran que la probabilidad de que un municipio semirural tenga mayoría de derechas aumenta con el número de empresas y número de contratos por cada 1.000 habitantes y disminuye con el número de licencias por cada 1.000 habitantes. Mientras que, la probabilidad de que un municipio urbano tenga mayoría de derechas aumenta con el número de Trabajadores y número de parados por cada 1.000 habitantes y disminuye con el número de establecimientos por cada 1.000 habitantes.

Capítulo 2 DATOS

Este capítulo consistirá en la explicación detallada de los pasos seguidos hasta la obtención de los datos que se utilizarán para los análisis, comenzando por la recopilación de tablas proporcionadas por diferentes fuentes, y terminando por la unión de todas ellas y su posterior depuración.

Se comenzará por la definición de algunos conceptos, puesto que es importante conocer la información de la que se dispondrá para comprender las variables con las que se va a trabajar a lo largo de todo el análisis.

En primer lugar, se define la **renta**, según el Instituto Nacional de Estadística, como los ingresos percibidos durante el año anterior al de la entrevista, y el **IRPF** (Impuesto sobre la Renta de las Personas Físicas), como un impuesto que pagan las personas físicas que son residentes en España o contribuyentes de sus rentas obtenidas durante un año natural, que empieza el 1 de enero y termina el 31 de diciembre.

Por otra parte, las **Afiliaciones a la seguridad social**, son un acto administrativo mediante el que se reconoce la condición de inclusión en el Sistema de Seguridad Social a la persona física y el **IAE**, es un **impuesto** que recoge las Actividades Económicas de los establecimientos.

Además, la **Cuenta de Cotización de una empresa**, consiste en un código de 11 dígitos que se le asignan a cada empresa con el fin de identificar y controlar las responsabilidades del Sistema de Seguridad Social. El **SIE** (Sistema de Información Estadística). El **Código INE** o **Código de municipio**, sirve como identificación única de cada municipio y está compuesto por 5 dígitos, representando los dos primeros al código de la provincia y los tres restantes al del municipio dentro de ésta. Por último, el **Censo** es una lista oficial de los habitantes de una población o de un estado.

Los datos con los que se va a trabajar se agrupan en tres conjuntos diferentes, contando cada uno de ellos con diferentes variables.

Por un lado, los datos **poblacionales**, en los que se encuentran:

- Municipio: nombre del municipio.
- Código INE: código INE del municipio.
- Población: número de habitantes del municipio.

Continuando por los datos **electorales**:

- Votos en blanco: número de votos en blanco en el municipio.
- Votos nulos: número de voto nulos en el municipio.
- Derecha: número de votos a partidos de Derecha en el municipio.
- Izquierda: número de votos a partidos de Izquierda en el municipio.
- Otros: número de votos a otro tipo de partidos en el municipio.

Y por último los datos **económicos**:

- Renta neta media por persona: ingresos netos percibidos durante el año anterior al de la entrevista por persona en el municipio.
- Renta neta media por hogar: ingresos netos percibidos durante el año anterior al de la entrevista por los miembros del hogar en el municipio.
- Total Empresas por Municipio: número de empresas en el municipio.

- Número Total Declaraciones: número total de declaraciones en el municipio.
- Paro Total: número total de parados en el municipio.
- Total Contratos: número total de contratos en el municipio.
- Número Total Licencias: número total de licencias que figuran en cada municipio por el IAE.
- Numero Trabajadores: número de trabajadores en cada municipio con una Cuenta de Cotización.
- Numero Establecimientos: número de establecimientos en cada municipio por Cuentas de Cotización.

Todos estos datos se han descargado y se han sometido a distintos procedimientos hasta llegar a agrupar las variables de la forma anterior y, a lo largo de todo este capítulo, se irá detallando el proceso.

2.1 OBTENCIÓN DE DATOS

Es importante que queden reflejadas las fuentes de las que se han obtenido los datos, así como su referencia temporal, puesto que, si se produce algún fallo a la hora de trabajar con ellos, podrán revisarse de forma sencilla. Teniendo en cuenta que se utilizarán los resultados de las elecciones municipales del 26 de mayo de 2019 en Castilla y León, todos los datos se seleccionarán tan próximos a dicha fecha como sea posible.

Los datos iniciales, se han obtenido mediante consutas personalizadas a partir de las siguientes fuentes, seleccionando para cada una de las 9 provincias de Castilla y León:

1. https://www.ine.es/dynt3/inebase/index.html?padre=525

- Año 2019
- Cifras oficiales de población resultantes de la revisión del Padrón municipal a 1 de enero:
    - Municipio y código INE
    - Total población (unidad personas)

2. https://www.ine.es/experimental/atlas/exp_atlas_tab.htm

- Año 2017
- Indicadores de renta media:
    - Municipio y código INE
    - Renta neta media por persona (unidad euros)
    - Renta neta media por hogar (unidad euros)

3. https://www.agenciatributaria.es/AEAT.internet/datosabiertos/catalogo/hacienda/Estadistica_de_los_declarantes_del_IRPF_por_municipios.shtml

- Año 2018
- Declaraciones por municipios del IRPF:
    - Municipio
    - Número total declaraciones (unidad declaraciones)

4. http://www.sepe.es/HomeSepe/que-es-el-sepe/estadisticas/datos-estadisticos/municipios

- Abril 2019
- Paro y contratos registrados:
    - Municipio
    - Paro Total (unidad personas)
    - Total Contratos (unidad contratos)

5. https://www.ine.es/jaxiT3/Tabla.htm?t=4721
- Año 2019
- Empresas por municipio:
    - Municipio y código INE
    - Total empresas (unidad empresas)
6. https://estadistica.jcyl.es/web/es/estadistica.html
- Año 2019
- Número total de licencias:
    - Municipio y código INE
    - Número Total Licencias (unidad licencias)
- Número de trabajadores y número de establecimientos:
    - Municipio y código INE
    - Número de Trabajadores (unidad personas)
    - Número de Establecimientos (unidad establecimientos)
7. http://www.infoelectoral.mir.es/infoelectoral/min/areaDescarga.html?method=inicio
- 26 de mayo de 2019
- Datos electorales:
    - Nombre de la Comunidad
    - Código de Provincia
    - Nombre de Provincia
    - Código de Municipio
    - Nombre de Municipio
    - Población (unidad personas)
    - Número de mesas (unidad mesas)
    - Total censo electoral (unidad personas)
    - Total votantes (unidad personas)
    - Votos válidos (unidad votos)
    - Papeletas a candidaturas (unidad papeletas)
    - Votos en blanco (unidad votos)
    - Votos nulos (unidad votos)
    - Nombre partido (unidad votos)

Para simplificar el problema se ha decidido reducir el número de variables. Esta tarea se realizará agrupando las variables que corresponden a los votos de cada partido en tres diferentes:

- Derecha: suma por municipios de las variables: PARTIDO POPULAR, CIUDADANOS-PARTIDO DE LA CIUDADANIA y VOX. (Derecha o centro derecha)
- Izquierda: suma por municipios de las variables: PARTIDO SOCIALISTA OBRERO ESPAÑOL, UNIDAS PODEMOS IZQUIERDA UNIDA, IZQUIERDA UNIDA, PODEMOS y PODEMOS – EQUO. (Izquierda o centro izquierda)
- Otros: suma por municipios del resto.

Aclarando que dicha clasificación se ha realizado mediante información objetiva según la cual, los partidos se asignan a derecha o izquierda por su definición, y los partidos correspondientes a determinadas comunidades autónomas o provincias se han asignado al grupo “otros”.

Esta modificación se ha realizado mediante el software R, generando un nuevo archivo de datos en el que se encontrarán las 3 nuevas variables, mediante los siguientes pasos:

- Cargar las librerias “readxl”, “dplyr” y “xlsx”.

- Función "read_excel" para la lectura del archivo a modificar.
- Creación de las nuevas variables mediante las especificaciones.
- Añadir nuevas variables a los datos.
- Función write.xlsx para crear el nuevo archivo Excel.

Una vez generado el archivo con las nuevas variables, mediante Excel se añaden al archivo que contiene los datos electorales y se eliminan las correspondientes a los partidos por separado.

2.2 UNIÓN DE LOS DATOS

Una vez descargados los datos que se van a a utilizar para los análisis posteriores, el siguiente paso es juntarlos en una misma tabla, puesto que se encuentran en archivos diferentes.

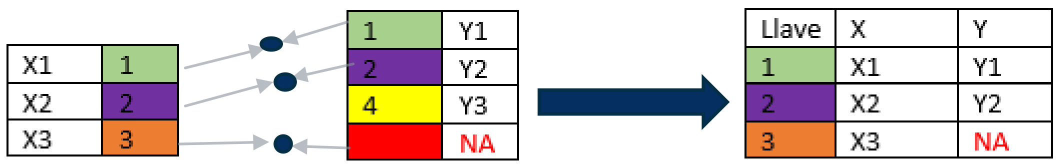
Esta tarea se realizará mediante la variable común en todos los archivos, que es **MUNICIPIO** y la dificultad que se encuentra es que no está representada de la misma forma en todos ellos, además de que en muchos casos el orden de los municipios es distinto. A continuación, se explican todas las soluciones que se han ido aplicando hasta poder realizar la unión de todas las tablas por dicha variable.

El primer lugar, se tomó la variable MUNICIPIO en el formato:

- Municipio y código INE

Al intentar juntar todos los archivos que presentan esta notación se ha encontrado el problema de que la variable no presentaba en el mismo orden en todos ellos y solo en el caso de que las tablas estuviesen ordenadas de la misma forma podrían juntarse directamente.

La solución será la unión exterior izquierda, que mantiene las observaciones que aparecen en la tabla izquierda, es decir, todas las observaciones que se encuentren en la tabla X, añadiendo una observación adicional a la observación que no se encuentre en ella, cuyo valor será NA:



.....

Ilustración 1.Unión de Datos

Lo más frecuente es utilizar este tipo de unión, puesto que se conservarán las observaciones originales incluso cuando no hay coincidencias.

Por lo tanto, se encuentran dos problemas a la hora de unir los diferentes archivos para crear una única tabla de datos, el primero de ellos, es que lo municipios se encuentran en distinto orden, y se solucionará con la unión izquierda, y el segundo es la diferencia de notaciones para la variable MUNICIPIO, según la fuente de la que provienen los datos, que habrá que solucionar antes de realizar la unión, para que la tabla final se genere de forma correcta.

Todos estos cambios en la variable MUNICIPIO se realizarán mediante funciones de **Excel**:

→Separar el código INE del municipio y extraer los municipios para que esta información se encuentre por separado:

- *=IZQUIERDA (cadenaCaracteres, numeroCaracteres)*
- *=EXTRAE (cadenaCaracteres, posicionInicial, numeroCaracteres)*

→Pasar la variable municipio a mayúscula:

- *=MAYUSC(cadenaCaracteres)*

→Quitar todos los acentos:

- *=SUSTITUIR(SUSTITUIR(SUSTITUIR(SUSTITUIR(SUSTITUIR(cadenaCaracteres; ”Á”;”A”); ”É”;”E”); ”Í”;”I”); ”Ó”;”O”); ”Ú”;”U”)*

→Por último, utilizando las opciones de Excel:

- *Buscar y seleccionar -> Reemplazar -> Reemplazar todos*

Modificar los siguientes valores para eliminar los paréntesis y comas de la variable:

- *, por (*
- *(por vacío*
- *) por vacío*

Tras todos estos cambios se encuentra la variable Municipio para todas las tablas representada sin el código INE, en mayúsculas, sin acentos, sin paréntesis y sin comas.

Además, queda también recogida la variable código INE puesto que puede ser más sencillo trabajar con ella posteriormente.

Una vez finalizado el trabajo de notación, se partirá de la tabla de datos electorales para la unión, puesto que son los más importantes en el análisis y en caso de producirse algún fallo podría traernos más problemas si tuviera lugar en esta parte de los datos.

La unión izquierda se ha realizado mediante el software R, creando un nuevo archivo de datos en el que se encontrarán las variables de todos los archivos anteriores:

- Cargar las librerias “readxl”, “dplyr” y “xlsx”, “data.table” y “tidyverse”.
- Función read_excel para cargar todos los archivos.
- Función as.data.table para convertir los archivos en data.table.
- Función left_join para unir los archivos.
- Función write.xlsx para crear el nuevo archivo Excel.

Una vez creado el archivo se encuentran dos variables de población:

- Población
- Poblacion2

Se eliminará la primera y se modificará el nombre de la segunda, puesto que corresponde a los datos electorales y como hemos explicado anteriormente serán los más importantes en el análisis, obteniéndose, una vez unidas todas las tablas, un archivo con 2.248 municipios y 17 variables: Municipio, Población, Votos en blanco, Votos nulos, Derecha, Izquierda, Otros, Renta neta media por persona, Renta neta media por hogar, Total Empresas por Municipio, Código INE, Número Total Declaraciones, Paro Total, Total Contratos, Número Total Licencias, Número Trabajadores, Número Establecimientos

2.3 DEPURACIÓN DE LOS DATOS

A partir de los datos anteriores compuestos por 2.248 municipios y 17 variables, se crearán dos nuevos archivos, puesto que se ha supuesto que, en los municipios con número pequeño de habitantes, la decisión electoral se debe más a la persona que se presenta que al partido.

Para ello, se establecerán dos restricciones, ambas en base a la variable Población.

- $Población\ > 100\ habitantes$
- $Población\ > 500\ habitantes$

La aplicación de las dos restricciones anteriores se ha realizado mediante el software R, creando dos nuevos archivos:

- Cargar las librerías “readxl”, “dplyr” y “xlsx”.
- Función read_excel para cargar el archivo.
- Aplicar restricciones a la variable Población.
- Función write.xlsx para crear los nuevos archivos Excel.

A la vista de los resultados obtenidos, se toma la decisión de seleccionar el archivo correspondiente a más de 500 habitantes, puesto que 456 municipios serán suficientes para poder realizar el análisis, por lo tanto, se continuará trabajando con el archivo de datos resultante.

Se cambia el orden de la variable Código INE, colocándola en segundo lugar, después del nombre del municipio, puesto que así todas las variables de carácter poblacional están juntas y de esta forma los datos quedan más ordenados.

En este archivo se encuentran varios datos con valor NA y puede ser por dos motivos:

- Fallos al realizar la unión, que se pueden corregir revisando los archivos iniciales.
- Valores ausentes reales en los archivos que hay que decidir cómo tratar.

Para identificar bien el tipo de error se irán analizando uno a uno todos los valores ausentes de los datos.

→En la **fila 91,** que corresponde a **VILLARCAYO DE MERINDAD DE CASTILLA LA VIEJA**, se ha producido un error en varias de las variables puesto que el nombre del municipio se ha cortado, por lo tanto, se completa la información correspondiente a este municipio:

- Código INE: 09903
- Renta neta media por persona: 10519
- Renta neta media por hogar: 23286
- Total Empresas Municipio: 473
- Número Total Declaraciones: 2089
- Paro Total: 255
- Total Contratos: 102
- Número Total Licencias: 609
- Numero Trabajadores: 309
- Numero Establecimientos: 120

Hay que destacar que, a excepción del Código INE de la fila 91 todos los valores ausentes se encuentran en las variables de tipo económico.

→Variable **Número Total Declaraciones**:

- FILA 94: ALFOZ DE QUINTANADUEÑAS: 1034
  Problema: Lectura en la letra Ñ.
- FILA 140: NOCEDA DEL BIERZO: 276
  Problema: El nombre del municipio aparece como NOCEDA.
- FILA 190: TORAL DE LOS VADOS: 839
  Problema: El municipio aparece como VILLADECANES, que es su nombre antiguo.
- FILA 320: REAL SITIO DE SAN ILDEFONSO: 2665
  Problema: El nombre del municipio aparece como ILDEFONSO.
- FILA 420: CORRALES DEL VINO: 414
  Problema: El nombre del municipio aparece como CORRALES.

→Variable **Paro Total**:

- FILA 420: CORRALES DEL VINO: 47
  Problema: El nombre del municipio aparece como CORRALES.

→Variable **Total Contratos**:

- FILA 420: CORRALES DEL VINO: 24
  Problema: El nombre del municipio aparece como CORRALES.
- El resto de NA correspondientes a esta variable son valores ausentes reales, y por lo tanto, como indica la fuente de la que proviene (www.sepe.es), se completarán con 0.
  - FILA 92: VALLE DE LAS NAVAS: 0
  - FILA 149: QUINTANA DEL CASTILLO: 0

→Variable **Número Total Licencias**: el único valor NA encontrado es un valor ausente real y como indica la fuente de la que proviene la variable (www.estadistica.jcyl.es), se completará con 0.

- FILA 334: SAN CRISTOBAL DE SEGOVIA: 0

→Variable **Numero Trabajadores**: los valores NA encontrados son valores ausentes reales y como indica la fuente de la que proviene (www.estadistica.jcyl.es), se completarán con 0.

- FILA 17: GUISANDO: 0
- FILA 92: VALLE DE LAS NAVAS: 0
- FILA 142: PALACIOS DE SIL: 0
- FILA 195: VILLAMONTAN DE LA VALDUERNA: 0
- FILA 203: VILLAZALA: 0
- FILA 221: MAGAZ DE PISUERGA: 0
- FILA 231: VILLALUENGA DE LA VEGA: 0
- FILA 234: VILLAUMBRALES: 0
- FILA 239: ALDEALEGUA: 0
- FILA 262: HINOJOSA DE DUERO: 0
- FILA 284: VILLARINO DE LOS AIRES: 0
- FILA 287: VILLORUELA: 0
- FILA 405: VILLAFRECHOS: 0
- FILA 423: FERRERAS DE ABAJO: 0
- FILA 428: MANGANESES DE LA POLVOROSA: 0

→Variable **Numero Establecimientos**: los valores NA encontrados son valores ausentes reales y como indica la fuente de la que proviene (www.estadistica.jcyl.es), se completarán con 0.

- FILA 17: GUISANDO: 0
- FILA 92: VALLE DE LAS NAVAS: 0
- FILA 142: PALACIOS DE SIL: 0
- FILA 195: VILLAMONTAN DE LA VALDUERNA: 0
- FILA 203: VILLAZALA: 0
- FILA 221: MAGAZ DE PISUERGA: 0
- FILA 231: VILLALUENGA DE LA VEGA: 0
- FILA 234: VILLAUMBRALES: 0
- FILA 239: ALDEALEGUA: 0
- FILA 262: HINOJOSA DE DUERO: 0
- FILA 284: VILLARINO DE LOS AIRES: 0
- FILA 287: VILLORUELA: 0
- FILA 405: VILLAFRECHOS: 0
- FILA 423: FERRERAS DE ABAJO: 0
- FILA 428: MANGANESES DE LA POLVOROSA: 0

Por lo tanto, el archivo final con el que se trabajará será en el que se han almacenado todos estos cambios, presentando las siguientes características:

- 456 municipios (filas)
- 17 variables (columnas)
- 0 valores ausentes

Las variables que contiene son las siguientes:

1. Municipio
2. Código INE
3. Población
4. Votos en blanco
5. Votos nulos
6. Derecha
7. Izquierda
8. Otros
9. Renta neta media por persona
10. Renta neta media por hogar
11. Total Empresas por Municipio
12. Número Total Declaraciones
13. Paro Total
14. Total Contratos
15. Número Total Licencias
16. Número Trabajadores
17. Número Establecimientos

Capítulo 3 ANÁLISIS DE LOS DATOS

En este capítulo se tratará de cumplir el objetivo del trabajo, que es establecer relaciones entre las variables económicas y electorales. Una vez obtenidos los datos con los que se trabajará, se hará uso de las ténicas multivariantes conocidas para encontrar dichas relaciones.

En primer lugar, un **análisis exploratorio** de los datos resultará de interés, en el que se presentarán las características principales de cada variable presente en la tabla.

Por otra parte, se aplicarán al conjunto de municipios inicial, técnicas multivariantes, cómo **Análisis de Componentes Principales**, para intentar establecer relaciones de forma gráfica, **Análisis de Correlaciones Canónicas**, en el que se encuentran las variables separadas en los grupos de interés, un **Análisis Clúster** sobre los municipios y un **ANOVA**, usando las características de los clústeres obtenidos.

Para terminar, se agruparán los municipios en base a una clasificación que se explicará posteriormente, y se aplicarán el **Análisis de Componentes Principales** y el de **Correlaciones Canónicas**, para comparar los resultados obtenidos en cada uno de los tipos, además de una **regresión logística** para establecer las variables que aumentan la probabilidad de voto a cada una de las opciones electorales.

3.1 EXPLORACIÓN DE LOS DATOS

Esta etapa del proceso de análisis de datos, consiste en utilizar técnicas de estadística descriptiva para poder interpretar los datos más fácilmente, creando gráficos y tablas para plantear unas conclusiones iniciales sobre las posibles relaciones existentes entre las variables económicas y electorales.

Esta etapa tiene dos objetivos principales, por un lado, presentar características de las variables individuales y por otro, descubrir patrones y relaciones entre variables.

Se separará el análisis exploratorio en base al tipo de variables con las que se trabaja: **categóricas** o **numéricas**.

Variables Categóricas

La variable **Tipo de municipio**, se ha creado por el desequilibrio existente en el número de habitantes de los municipios disponibles, clasificando los municipios según su Población de la siguiente forma:

- Rural: población menor que 2.000
- Semirural: población entre 2.000 y 10.000
- Urbano: población mayor que 10.000

Rural	Semirural	Urbano
329	104	23
0.72	0.23	0.05

Tabla 1. Tipo de municipio

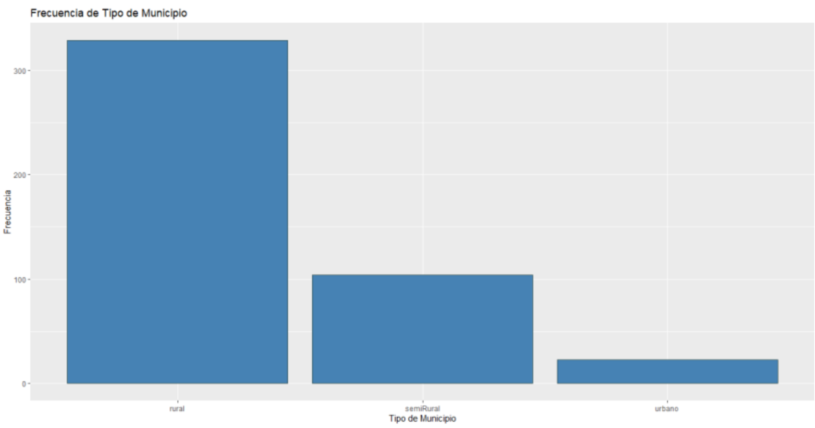


Ilustración 2. Exploración Tipo Municipio

Población Rural	Población Semirural	Población Urbana	Población Total
310.998	444.192	1.357.677	2.112.867
0.15	0.21	0.64	1

Tabla 2. Población por Tipo de Municipio

El total de la población de los 456 municipios analizados, son 2.112.867 habitantes, de los cuales 310.998 pertenecen a municipios rurales (15%), 444.193 a municipios semirurales (21%) y 1.357.677 a municipios urbanos (64%).

La idea obtenida tras el análisis de **Tipo de municipio** es que los habitantes de Castilla y León están vinculados a los 3 grupos de forma muy desigual, ya que, aunque los urbanos solo son el 5%, en ellos encontramos el 64% de la población y, por lo tanto, las conclusiones que se obtengan de este tipo de municipios serán de gran importancia.

Variables Numéricas

Min.	1º cuartil	Mediana	Media	3º cuartil	Max.
501	696	1.066	4633	2162	298.866

Tabla 3. Estadísticos Población

A pesar de que la media de **Población** es 4.633, se puede ver que el 75% de los municipios no superan los 2.162 habitantes y todos ellos se encuentran muy alejados del máximo de la variable y muy próximos al mínimo.

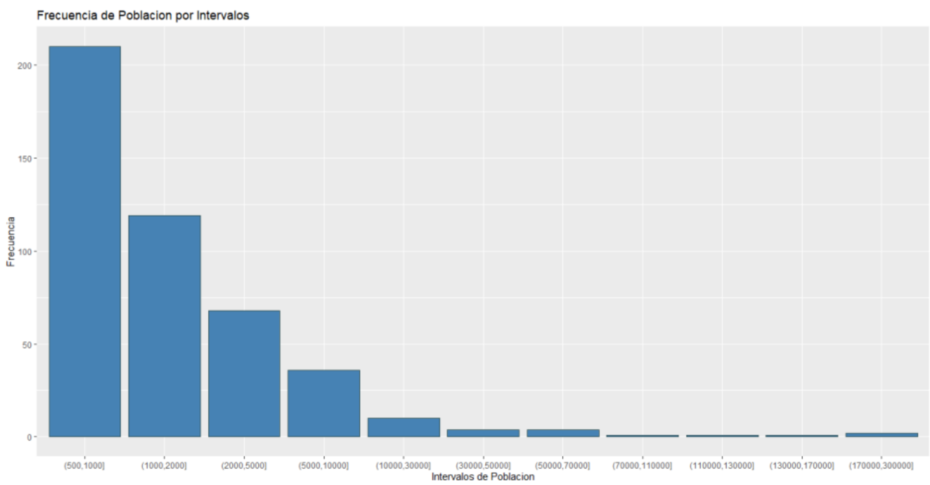


Ilustración 3. Exploración Población

Por otro lado, la moda de la variable se encuentra en el intervalo entre 500 y 1.000 habitantes, siendo más de 200 municipios los que cumplen esta característica, mientras que los intervalos en los se observa menor número de municipios corresponden a un gran número de habitantes.

Min.	1º cuartil	Mediana	Media	3º cuartil	Max.
7.715	9.921	10.656	10.829	11.681	16.544

Tabla 4. Estadísticos Renta por Persona

Los estadísticos de la **renta neta media por persona** muestran una distribución de la variable bastante centrada, puesto que la media y la mediana de los datos son muy próximas.

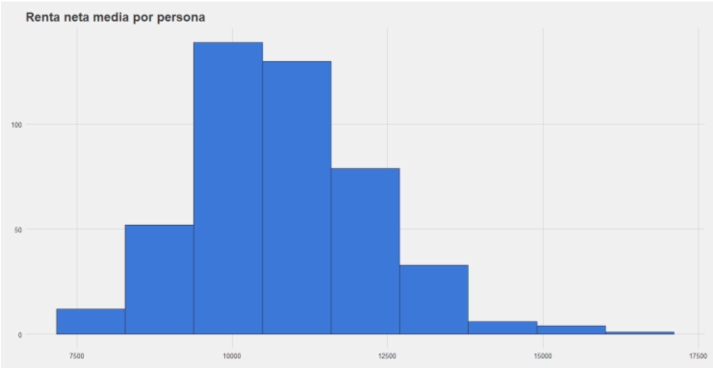


Ilustración 4. Exploración Renta por Persona

La moda de la renta neta media por persona se encuentra en 10.000, aunque los siguientes valores más comunes están entre dicho valor y 12.500, por lo tanto, se aprecia que la distribución de esta variable es ligeramente asimétrica hacia la izquierda.

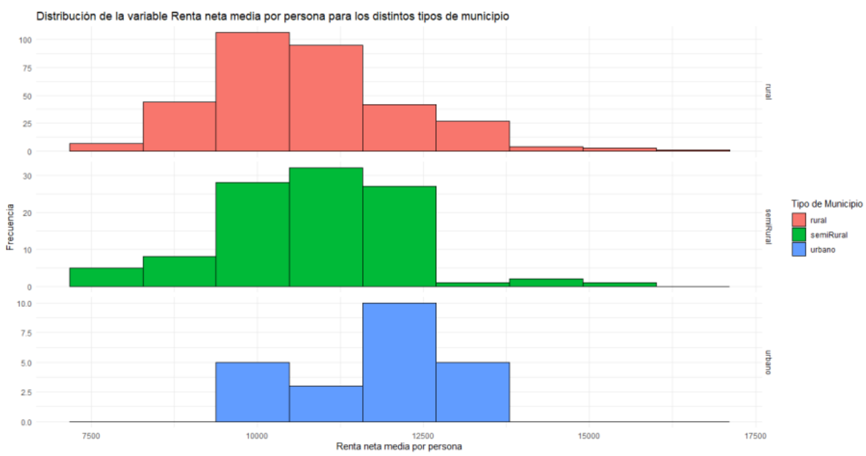


Ilustración 5. Exploración Renta por Persona por Tipo Municipio

Si se analiza esta variable según los 3 tipos de municipio se puede ver que los rurales son los que presentan menor moda, 10.000, seguidos de los semirurales, entre 10.000 y 12.500 y por último los urbanos, en 12.500.

Min.	1º cuartil	Mediana	Media	3º cuartil	Max.
16.559	22.749	25.012	25.538	27.807	47.012

Tabla 5. Estadísticos Renta por Hogar

Para la **renta renta neta media por hogar**, se encuentran características similares a las de la renta neta media por persona. Su distribución también es centrada, puesto que en este caso la media y la mediana también son bastante próximas.

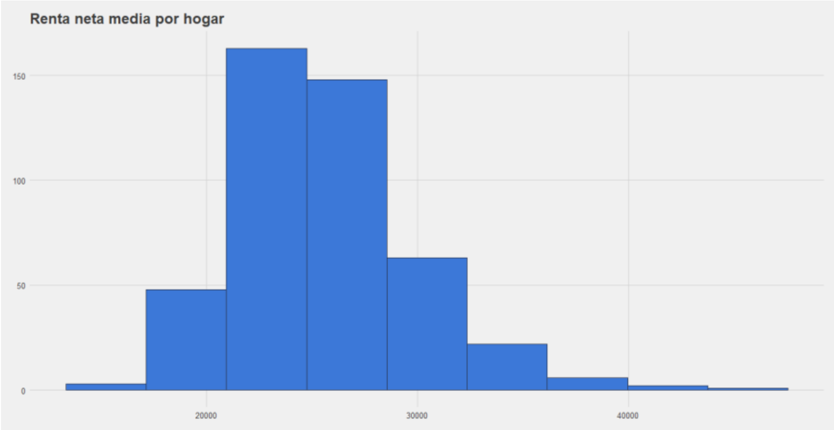


Ilustración 6. Exploración Renta por Hogar

Su distribución también es ligeramente asimétrica hacia la izquierda y la moda se encuentra muy próxima a 20.000, siendo más de 150 municipios los que presentan una renta neta media por hogar en torno a dicho valor.

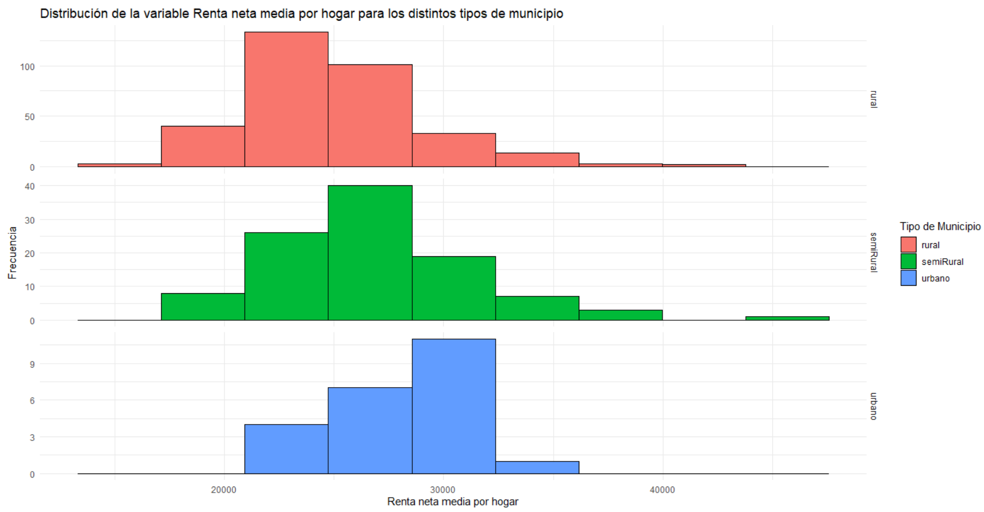


Ilustración 7. Exploración Renta por Hogar por Tipo Municipio

También se encuentran diferencias en la moda de la variable según el tipo de municipio siendo de nuevo, la más pequeña para los municipios rurales, seguido de los semirurales y la más grande para los municipios urbanos.

Para las siguientes variables, si se continúa trabajando con valores absolutos el análisis no aportaría nada puesto que, al tratarse de totales, es evidente que cualquiera de ellas presentará valores más altos en aquellos municipios con más habitantes. Por lo tanto, se continuará con la exploración de los datos con valores relativos, es decir, tomando los datos por cada 1.000 habitantes.

Min.	1º cuartil	Mediana	Media	3º cuartil	Max.
0.007	0.02	0.03	0.15	0.08	9.42

Tabla 6. Estadísticos Empresas por cada 1.000 habitantes

En cuanto a la variable **número de empresas**, se observa que el 75% de los municipios no superan las 0.08 empresas por cada 1.000 habitantes, encontrándose dicho valor muy alejado del máximo de la variable (9.42) y mucho más próximo al mínimo (0.007).

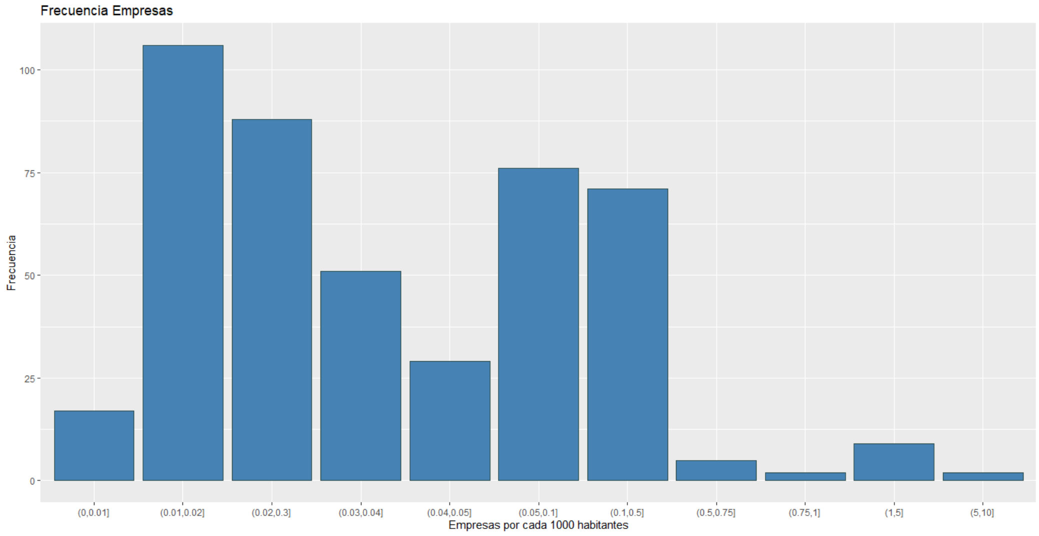


Ilustración 8. Exploración Empresas por cada 1000 habitantes

La moda de la variable se encuentra entre las 0.01 y 0.02 empresas por cada 1.000 habitantes, seguida por el intervalo comprendido entre las 0.05 y las 0.1.

Para terminar con la descripción de esta variable, se ve como el número de empresas por cada 1.000 habitantes es mayor para los municipios urbanos, si se compara con el número de empresas que presentan los municipios rurales y semirurales, puesto que las modas de cada tipo de municipio se encuentran en intervalos diferentes.

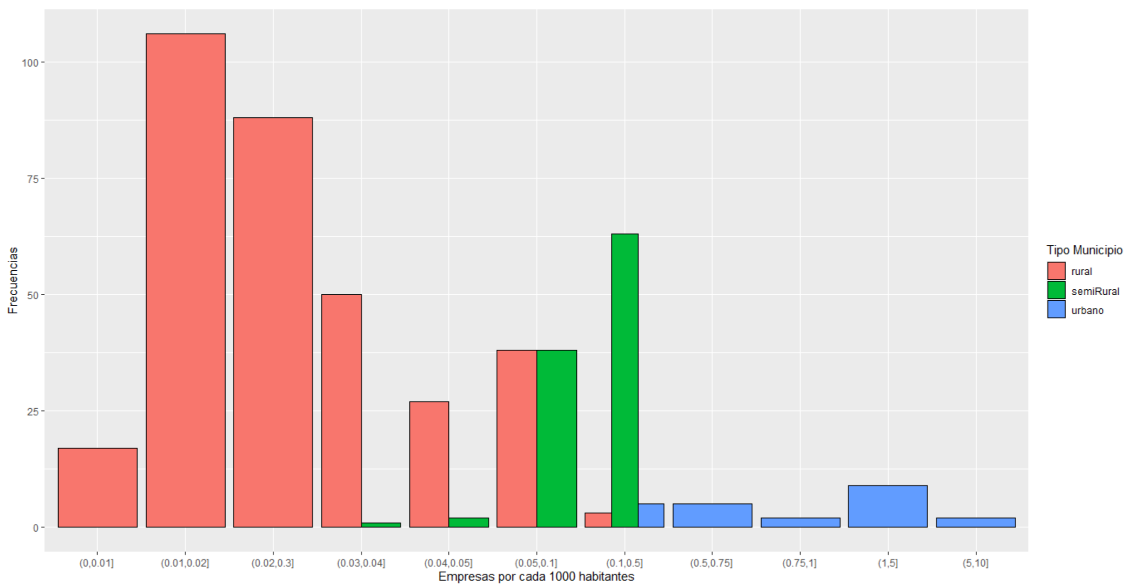


Ilustración 9. Exploración Empresas por cada 1000 habitantes Tipo Municipio

Min.	1º cuartil	Mediana	Media	3º cuartil	Max.
0.03	0.15	0.23	1.18	0.51	81.29

Tabla 7. Estadísticos Delaraciones por cada 1000 habitantes

Se observa que, para la variable **Número Declaraciones**, el 75% de los municipios no superan las 0.51 declaraciones por cada 1.000 habitantes, encontrándose todos ellos muy alejados del valor máximo de la variable, que es 81.29.

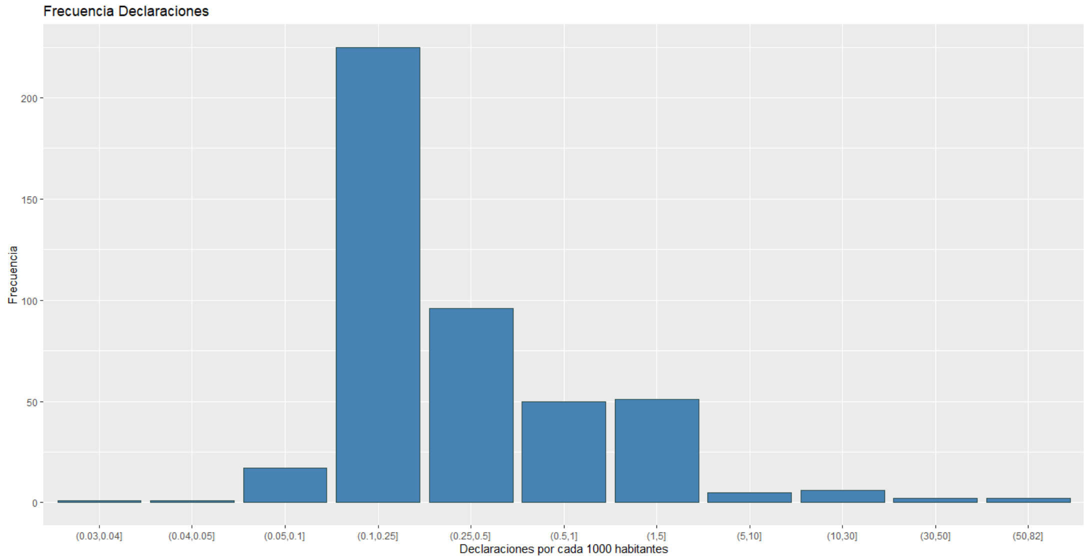


Ilustración 10. Exploración Declaraciones por cada 1000 habitantes

La moda de la variable se encuentra en el intervalo entre 0.1 y 0.25 declaraciones por cada 1.000 habitantes siendo unos 225 municipios los que presentan su número comprendido entre estos dos valores.

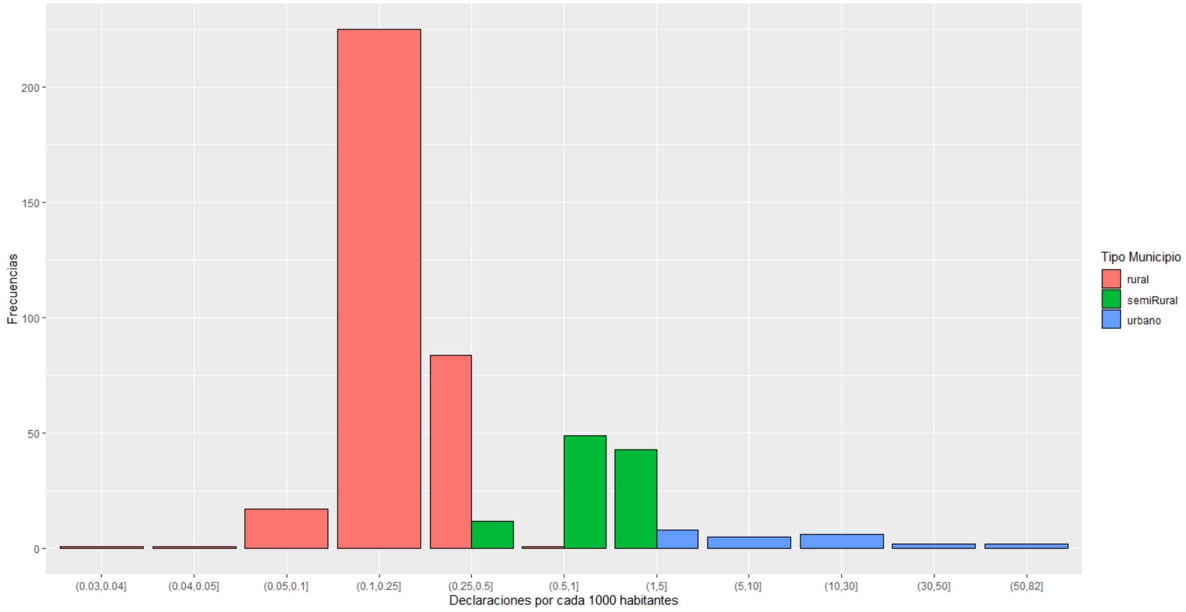


Ilustración 11. Exploración Declaraciones por cada 1000 habitantes Tipo Municipio

El número de declaraciones por cada 1.000 habitantes es mayor para los municipios urbanos, seguido de los municipios semirurales y siendo el grupo de rurales el que presenta menor valor para la variable en la mayor parte de los casos.

Min.	1º cuartil	Mediana	Media	3º cuartil	Max.
0.004	0.01	0.02	0.13	0.05	8.4

Tabla 8. Estadísticos Paro por cada 1000 habitantes

Para la variable **Paro**, el 75% de los municipios no superan los 0.05 parados por cada 1.000 habitantes, valor que se encuentra muy alejado del máximo de la variable y mucho más próximo al mínimo, que son 8.4 y 0.004, respectivamente.

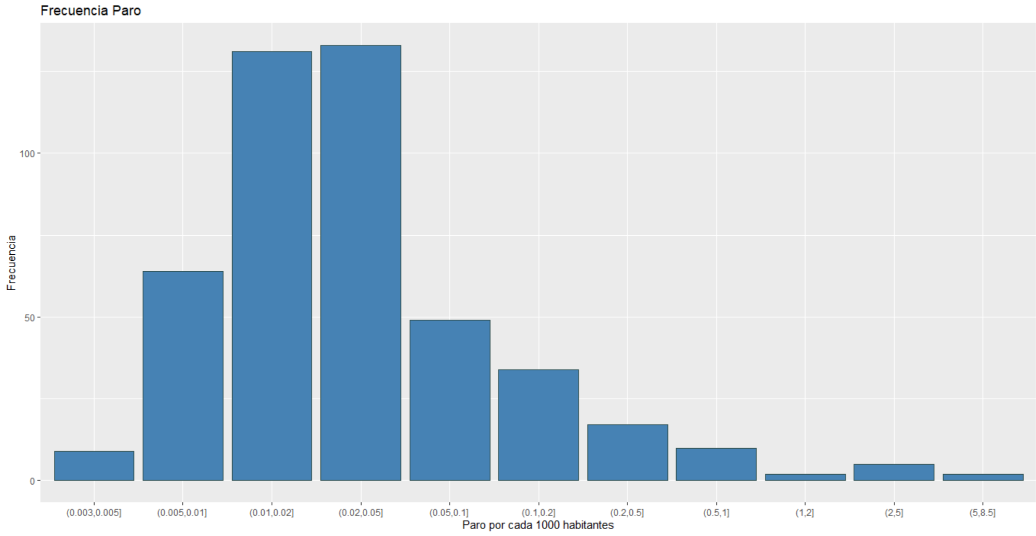


Ilustración 12. Exploración Paro por cada 1000 habitantes

La moda corresponde al intervalo que comprende entre los 0.02 y los 0.05 parados por cada 1.000 habitantes, siendo más de 125 municipios los que registran su tasa entre estos valores. Un número muy similar de municipios presentan entre 0.01 y 0.02 parados por cada 1.000 habitantes.

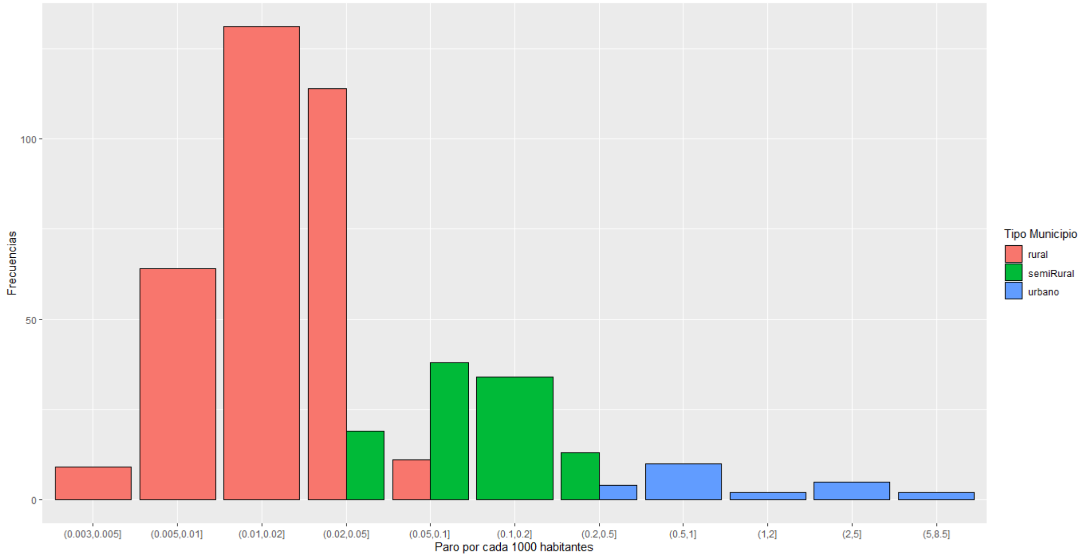


Ilustración 13. Exploración Paro por cada 1000 habitantes Tipo Municipio

El grupo los municipios que presentan mayor tasa de parados son los urbanos, y los que la presentan menor son los rurales, observándose de nuevo que la moda en cada uno de los tipos de municipio se encuentra en intervalos muy diferentes.

Min.	1º cuartil	Mediana	Media	3º cuartil	Max.
0	0.005	0.01	0.09	0.04	7.06

Tabla 9. Estadísticos Contratos por cada 1000 habitantes

En **Contratos** se osberva que el 75% de los municipios no superan los 0.04 contratos por cada 1.000 habitantes, encontrándose este valor muy alejado del máximo de la variable (7.06).

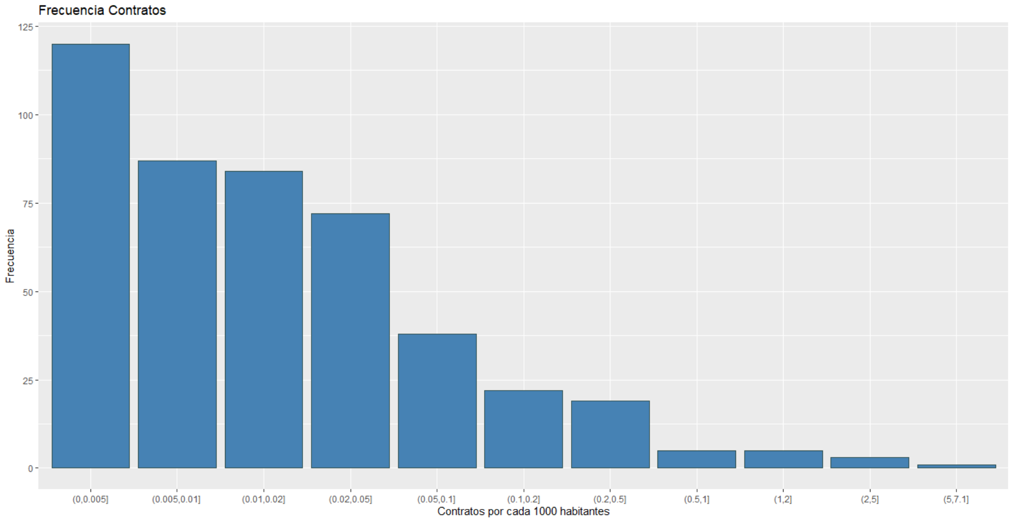


Ilustración 14. Exploración Contratos por cada 1000 habitantes

La moda de la variable se encuentra en el intervalo que comprende los valores 0 y 0.05, siendo casi de 125 municipio los que presentan el número de contratos por cada 1.000 habitantes dentro de dicho intervalo.

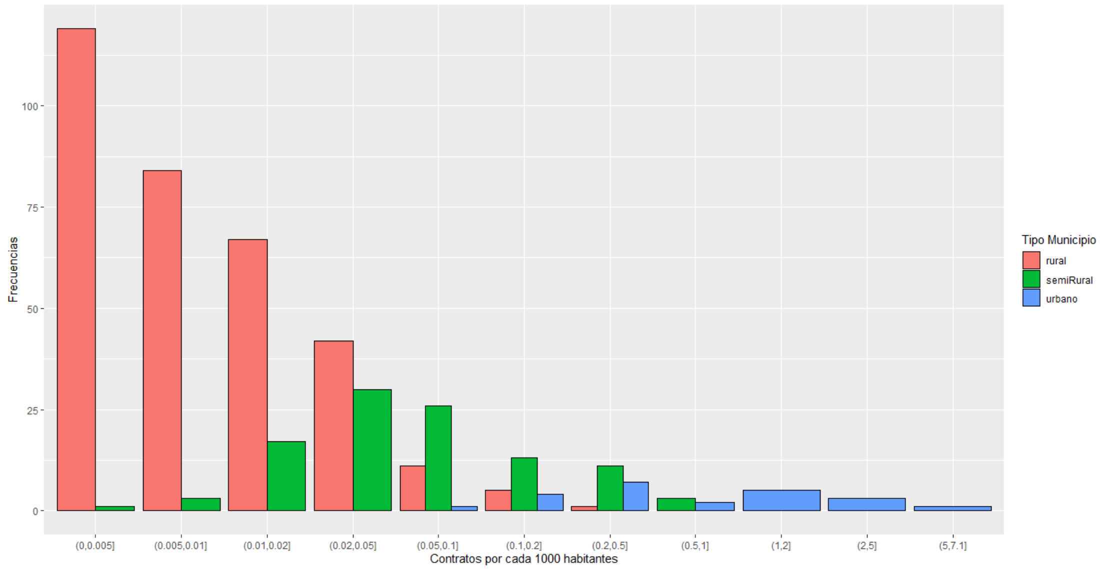


Ilustración 15. Exploración Contratos por cada 1000 habitantes Tipo Municipio

Aunque, los municipios urbanos son los que presentan un número mayor de contratos por cada 1.000 habitantes, seguidos del grupo de municipios semirurales, se aprecia que la diferencia entre el número de contratos entre los municipios rurales y los semirurales es menos evidente, incluso hay varios intervalos que presentan municipios de los 3 tipos.

Min.	1º cuartil	Mediana	Media	3º cuartil	Max.
0	0.02	0.04	0.19	0.09	13.07

Tabla 10. Estadísticos Número Licencias por cada 1000 habitantes

Para **Número Licencias** por cada 1000 habitantes, se observa que el 75% de los municipios disponibles no superan las 0.09, estando dicho valor muy alejado del máximo de la variable, que son 13.07 licencias.

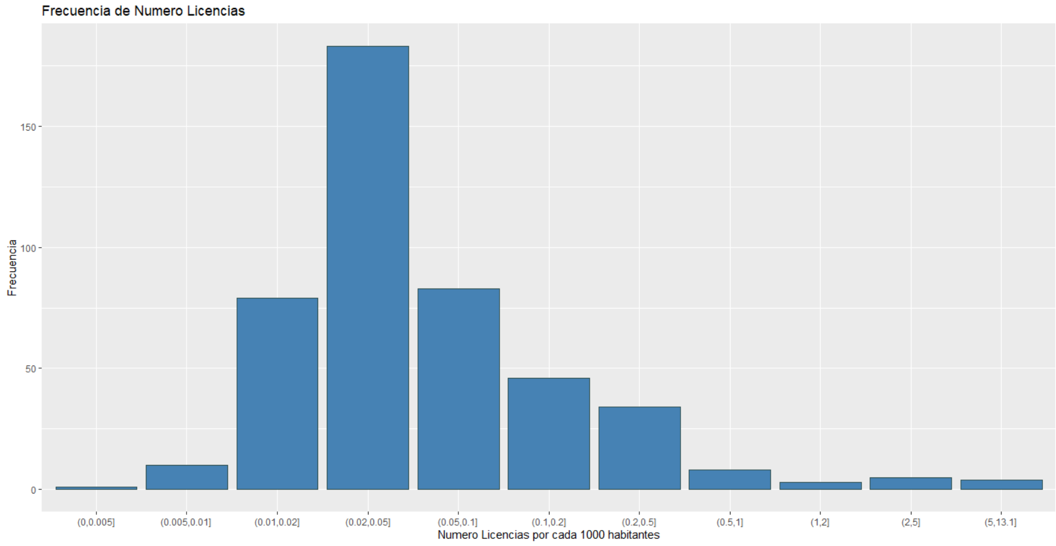


Ilustración 16. Exploración Licencias por cada 1000 habitantes

Además, se puede ver que la moda de la variable se encuentra en el intervalo [0.02,0.05], siendo más de 175 municipios los que presentan su número de licencias por cada 1.000 habitantes comprendido entre dichos valores.

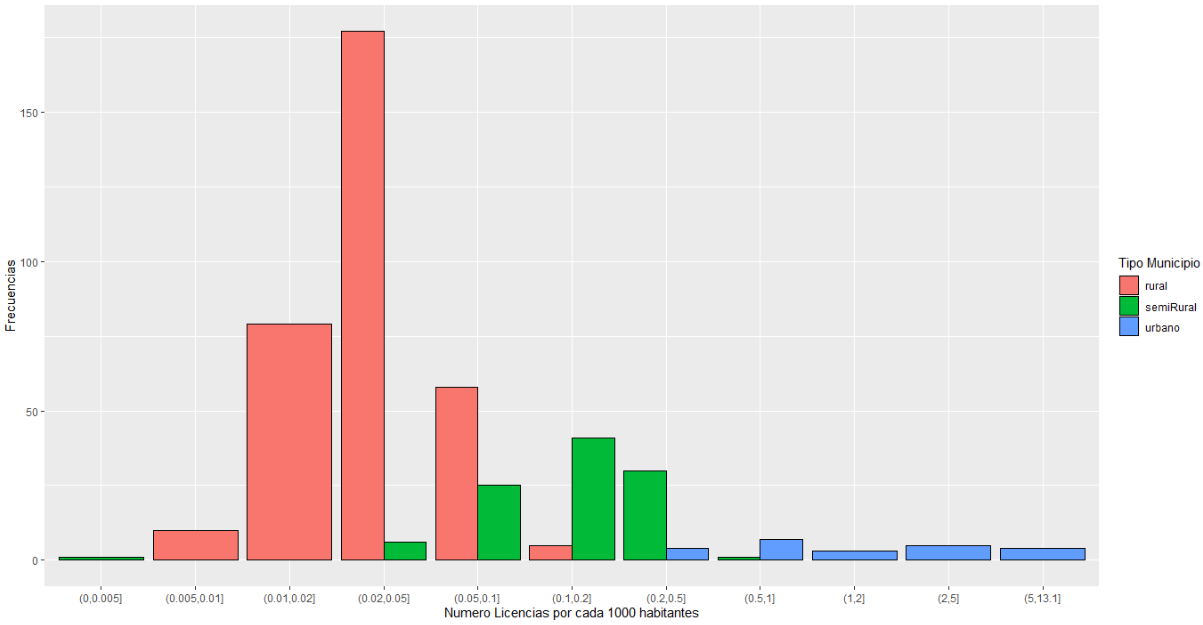


Ilustración 17. Exploración Licencias por cada 1000 habitantes Tipo Municipio

En este caso, los municipios con el número de licencias más alto por cada 1.000 habitantes son los urbanos, y los municipios semirurales se encuentran presentes en casi todos los intervalos representados.

Min.	1º cuartil	Mediana	Media	3º cuartil	Max.
0	0.007	0.016	0.14	0.06	10.58

Tabla 11. Estadísticos Trabajadores por cada 1000 habitantes

Para **Número Trabajadores** por cada 1.000 habitantes se puede ver que el 75% no supera el valor 0.06, estando este valor mucho más próximo al mínimo (0), que al máximo (10.58).

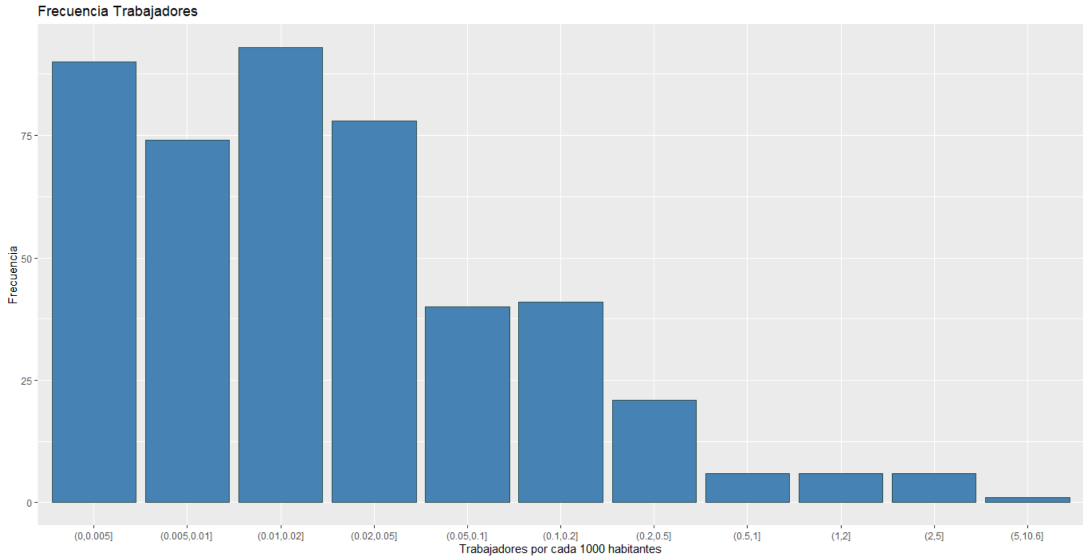


Ilustración 18. Exploración Trabajadores por cada 1000 habitantes

Se encuentra la moda de esta variable en el intervalo comprendido entre los valores valores 0.01 y 0.02, siendo más de 87 municipios los que cumplen dicha característica, además, los resultados son muy similares para el intervalo que comprende entre los 0 y los 0.005 trabajadores por cada 1.000 habitantes.

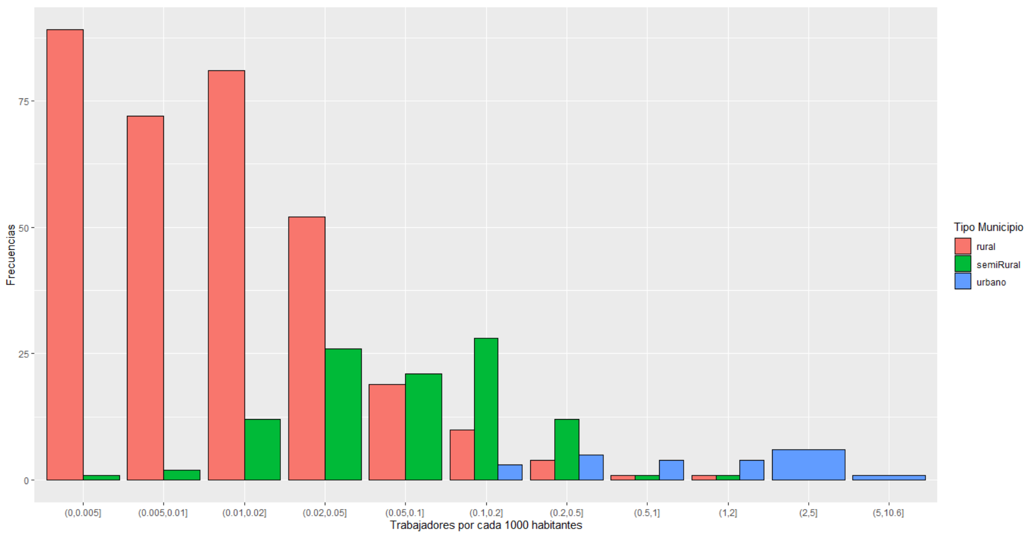


Ilustración 19. Exploración Trabajadores por cada 1000 habitantes Tipo Municipio

Para esta misma variable se observa que los municipios pertenecientes al grupo semirural se encuentran presentes en todos los intervalos de valores excepto en los dos últimos, encontrándose además municipios urbanos en los 6 últimos intervalos.

Min.	1º cuartil	Mediana	Media	3º cuartil	Max.
0	0.005	0.01	0.05	0.02	3.78

Tabla 12. Estadísticos Establecimientos por cada 1000 habitantes

Para la última variable económica, **Número Establecimientos**, se ve que el 75% de los municipios no supera los 0.02 establecimientos por cada 1.000 habitantes, encontrándose dicho valor mucho más próximo al mínimo que al máximo.

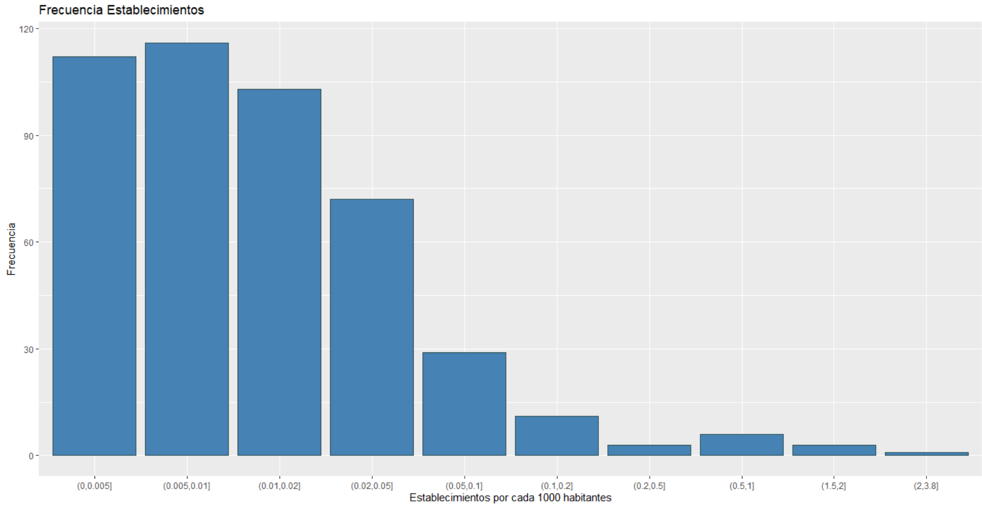


Ilustración 20. Exploración Establecimientos por cada 1000 habitantes

La moda para esta variable se encuentra en el intervalo que contiene a los municipios comprendidos entre 0.005 y 0.01 establecimientos por cada 1.000 habitantes. Los municipios que comprenden entre los 0 y 0.005, y entre los 0.01 y 0.02, presentan una frecuencia muy similar, en torno a 120.

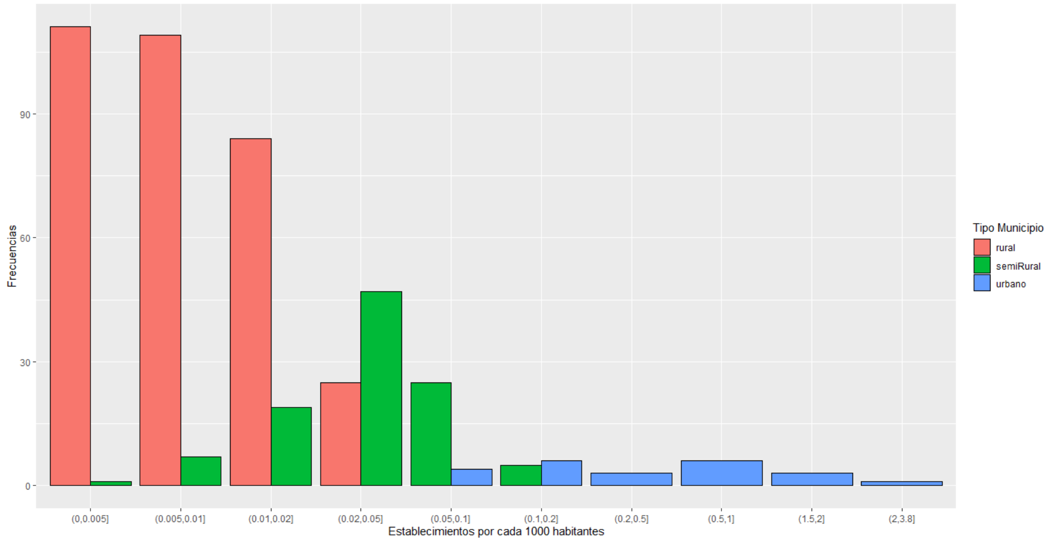


Ilustración 21. Exploración Establecimientos por cada 1000 habitantes

Como se puede observar, aunque los valores más grandes de la variable corresponden a los urbanos y los más bajos a rurales, el número de establecimientos por cada 1.000 habitantes no presenta diferencias demasiado claras entre los 3 tipos de municipio estudiados, puesto que en los intervalos intermedios se encuentran municipios de todos los grupos.

Min.	1º cuartil	Mediana	Media	3º cuartil	Max.
0	0.11	0.17	0.58	0.30	39.09

Tabla 13. Estadísticos Derecha por cada 1000 habitantes

Al analizar la primera variable electoral, **Derecha**, se ve que el 75% de los municipios disponibles no superan los 0.3 votos por cada 1.000 habitantes, valor que se encuentra muy distanciado del máximo de esta variable (39.09).

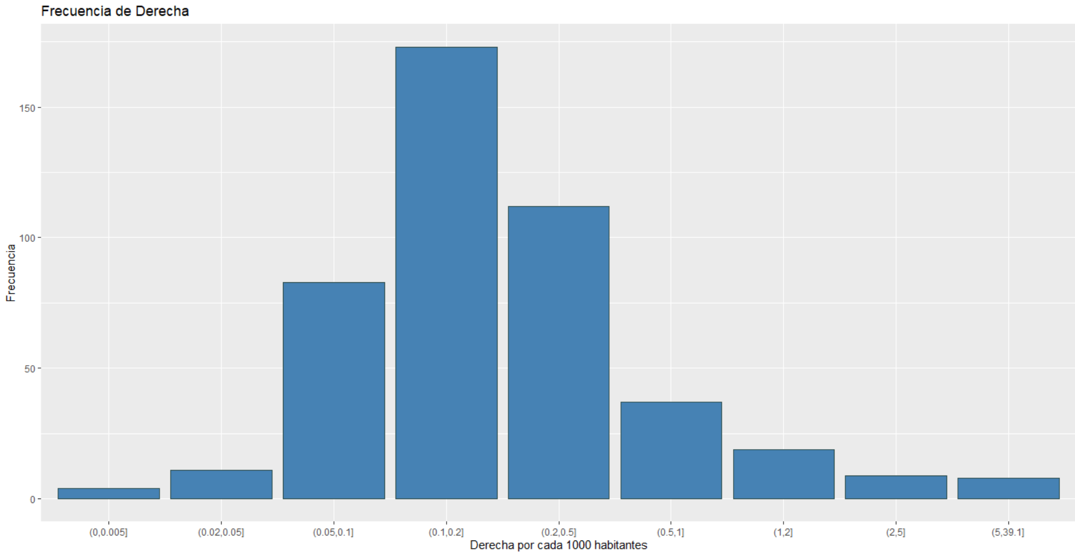


Ilustración 22. Exploración Derecha por cada 1000 habitantes

La moda de la variable Derecha se encuentra en aquellos municipios cuyo número de votos por cada 1.000 habitantes a esta tendencia está comprendido entre 0.1 y 0.2, siendo casi de 175 municipios los que se encuentran dentro de este intervalo.

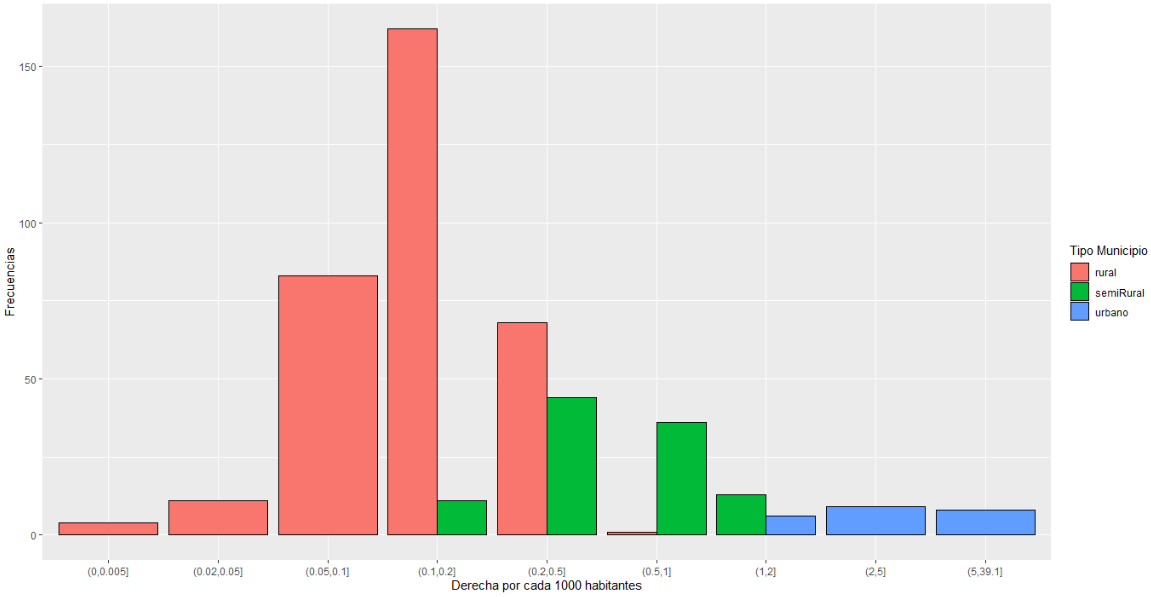


Ilustración 23. Exploración Derecha por cada 1000 habitantes Tipo Municipio

Los valores más altos en el número de votos a la derecha corresponden a los municipios urbanos y los más bajos a los rurales. Además, los intervalos que contienen valores intermedios para esta variable también corresponden al grupo semirural.

Min.	1º cuartil	Mediana	Media	3º cuartil	Max.
0	0.07	0.12	0.44	0.24	30.82

Tabla 14. Estadísticos Izquierda por cada 1000 habitantes

Para **Izquierda**, el 75% de los municipios no superan los 0.24 votos por cada 1.000 habitantes, valor que se encuentra mucho más próximo al mínimo que al máximo de la variable.

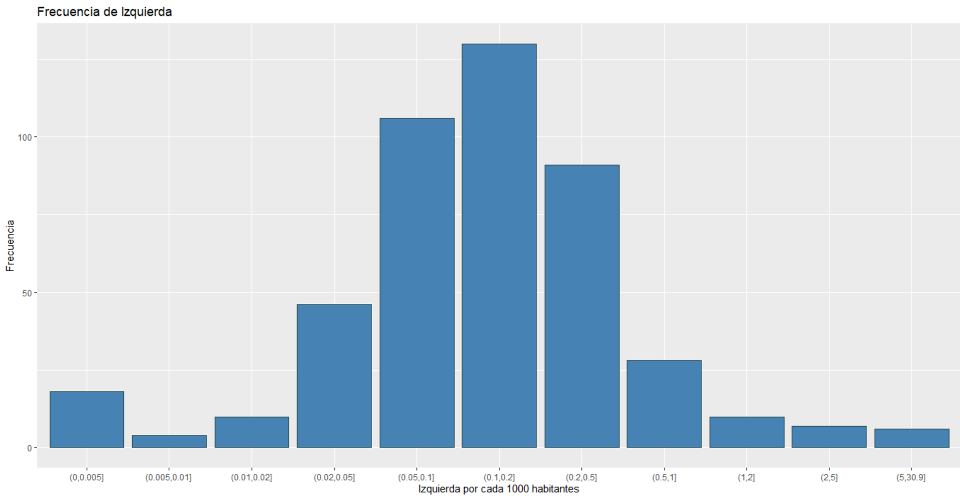


Ilustración 24. Exploración Izquierda por cada 1000 habitantes

La moda se encuentra en el intervalo que contiene a aquellos municipios cuyo número de votos a la izquierda por cada 1.000 habitantes se encuentra entre 0.1 y 0.2, siendo más de 125 municipios los que cumplen esta característica.

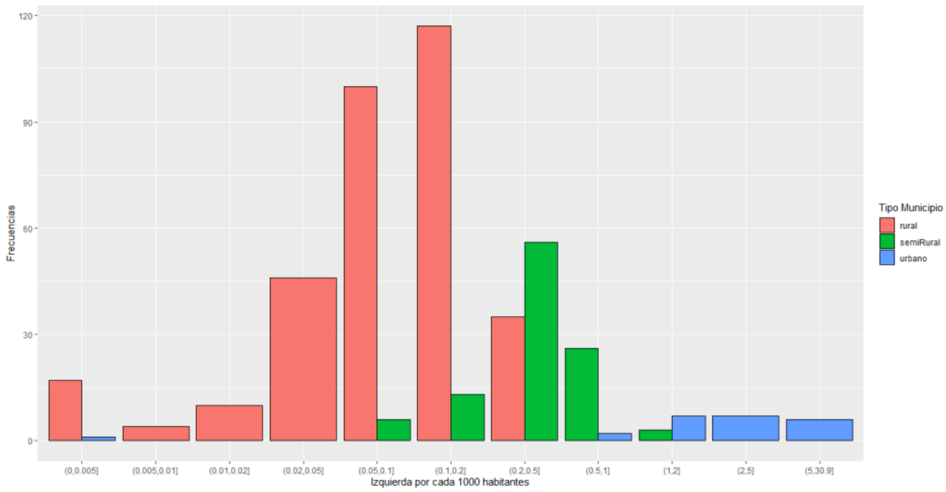


Ilustración 25. Exploración Izquierda por cada 1000 habitantes Tipo Municipio

Los intervalos correspondientes a un número pequeño de votos a la izquierda por cada 1.000 habitantes, recogen a municipios rurales y los correspondientes a un gran número de votos a la izquierda a los urbanos, a excepción de un número mínimo de municipios de este mismo grupo que se encuentran en el intervalo *(0,0.005].*

Min.	1º cuartil	Mediana	Media	3º cuartil	Max.
0	0	0.02	0.19	0.11	9.29

Tabla 15. Estadísticos Otros por cada 1000 habitantes

Para **Otros**, el 75% de los municipios no supera los 0.11 votos a otros partidos por cada 1.000 habitantes, siendo un valor muy extremo en comparación con el máximo (9.29). Además, el 25% de los municipios no presentan ningún voto hacia esta tendencia electoral y el 50% no superan los 0.02 votos por cada 1.000 habitantes.

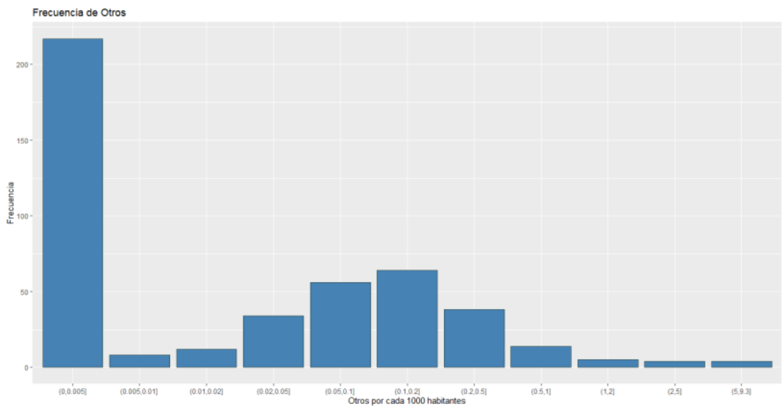


Ilustración 26. Exploración Otros por cada 1000 habitantes

La moda de la variable, como cabía esperar después de las apreciaciones de los estadísticos, se encuentra en el intervalo que representa a los municipios cuya tasa de voto a otros partidos está entre 0 y 0.005.

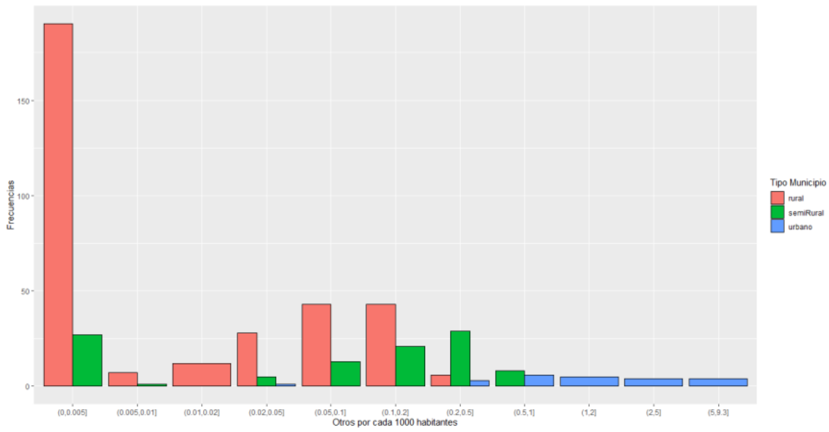


Ilustración 27. Exploración Otros por cada 1000 habitantes Tipo Municipio

Por último, se observa que los intervalos que representan un número de votos a otros partidos más elevado corresponden mayoritariamente a los municipios urbanos, a excepción de algunos municipios de este tipo que se encuentran en otros intervalos que representan menor tasa de votos.

	RURAL	SEMIRURAL	URBANO	TOTAL
Población	0.15	0.21	0.64	1
Derecha	0.19	0.21	0.60	1
Izquierda	0.17	0.21	0.62	1
Otros	0.14	0.22	0.64	1
Rent.Persona	0.72	0.23	0.05	1
Rent.Hogar	0.70	0.24	0.06	1
Empresas	0.14	0.22	0.64	1
Declaracion.	0.12	0.20	0.68	1
Paro	0.11	0.20	0.69	1
Contratos	0.12	0.24	0.64	1
Licencias	0.13	0.19	0.68	1
Trabajadores	0.14	0.19	0.67	1
Establecimie.	0.13	0.19	0.68	1

Tabla 16. Variables por Tipo de Municipio

Como idea principal, tras obtener los resultados de la exploración de los datos, se observa en la tabla superior que la Población parece tener gran influencia sobre el resto de variables, puesto que, el porcentaje de población que se encuentra dentro de cada uno de los 3 tipos de municipio es similar al porcentaje que se encuentra del resto de variables, a excepción de las rentas.

3.2 RESULTADOS GENERALES

En este punto, se presentan los resultados obtenidos en los diferentes análisis aplicados al conjunto inicial de municipios y las conclusiones que pueden extraerse de cada uno de ellos, en cuanto a establecer relaciones entre variables económicas y electorales.

Análisis de Componentes Principales

El primer paso a realizar será un **análisis de componentes principales**, para tratar de encontrar de forma gráfica alguna de las relaciones que se buscan.

Esta técnica, se encuentra dentro de las correspondientes al aprendizaje no supervisado, puesto que se quiere estudiar la posible existencia de subgrupos entre las variables u observaciones, no se pretende obtener predicciones.

Su finalidad es la reducción del número de variables o dimensiones intentando que se pierda la menor cantidad de información posible, midiendo dicha información como el porcentaje de varianza explicada.

Cada componente principal representará una combinación lineal de las variables iniciales, siendo todas ellas independientes entre sí, y representándose de la siguiente forma:

$$Componente_p = Coeficiente_{1p}Variable_1 + Coeficiente_{2p}Variable_2 + \cdots + Coeficiente_{pp}Variable_p$$

$$\{p = número\ de\ la\ variable\}$$

Por lo tanto, en este caso concreto se obtendrán 13 componentes principales, cada una de ellas formada por una combinación líneal de las 13 variables con las que se va a trabajar:

1. Población
2. Derecha
3. Izquierda
4. Otros
5. Renta neta media por persona
6. Renta neta media por hogar
7. Total empresas por municipio
8. Número Total Declaraciones
9. Paro total
10. Total contratos
11. Número total licencias
12. Número trabajadores
13. Número establecimientos

Por otra parte, cada componente principal está representada por un autovector y su correspondiente autovalor, indicando, el primero la dirección de la dimensión y el segundo la cantidad de varianza explicada, siendo siempre mayor por orden de componentes, lo que hace que puedan utilizarse para determinar el número de componentes necesarias para explicar los datos.

En este caso, los autovalores del análisis serán 13, puesto que se trabaja con 13 variables numéricas, y por lo tanto con 13 componentes principales. Además, al tratarse de un análisis normado, debido a las diferencias en las respectivas escalas, la suma de todos ellos será 13:

	Autovalor	%Varianza	%Varianza Acumulada
1	6.17	47.43	47.43
2	1.60	12.33	59.76
3	1.02	7.88	67.64
4	0.86	6.61	74.25
5	0.66	5.08	79.33
6	0.59	4.51	83.84
7	0.42	3.24	87.08
8	0.37	2.85	89.93
9	0.34	2.64	92.57
10	0.30	2.28	94.85
11	0.28	2.14	96.99
12	0.25	1.92	98.93
13	0.14	1.07	100

Tabla 17. Varianza Explicada por Componentes

Observando la tabla superior, se puede ver que la primera componente es la que presenta mayor autovalor y mayor porcentaje de varianza explicada (47.43%). Podrían seleccionarse hasta 4 componentes para superar el 70%, alcanzando así el 74.25%. Sin embargo, a partir de la segunda componente se ve que el porcentaje de varianza explicada que aportarían las nuevas componentes es mínimo y por lo tanto habría que seleccionar demasiadas dimensiones para superar el 80%.

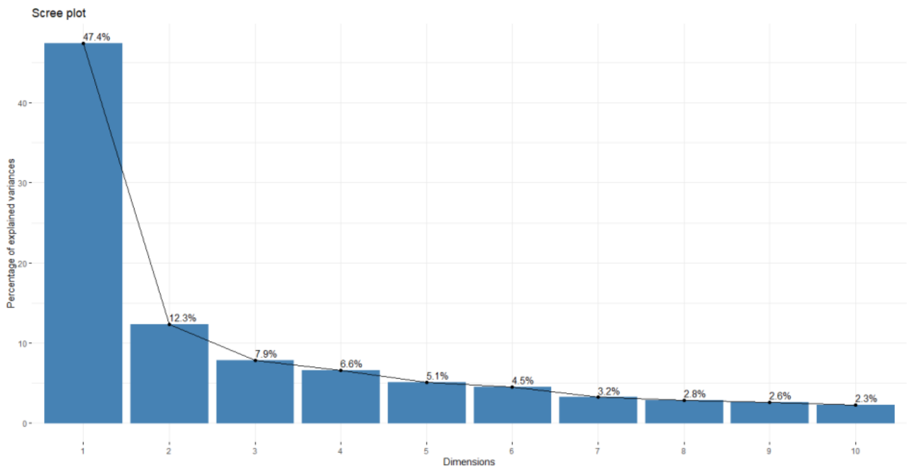


Ilustración 28. Gráfico de Autovalores

Analizando el gráfico de autovalores se llega a la misma conclusión. Se puede pensar que 2 componentes son suficientes, puesto que es el número en el que la representación empieza a descender más lentamente, aunque, con esta elección, únicamente se obtendría un 59.7% de varianza explicada. Además, se seleccionarán 2 dimensiones para poder interpretar los resultados gráficamente.

Mediante el autovector correspondiente a la primera componente, que determina los coeficientes, se observa que todas las variables presentan un valor negativo, siendo la variable renta neta media por persona la que tiene el coeficiente más pequeño, seguido de la renta neta media por hogar.

$$PC1 = -0.36Población - 0.31Derecha - 0.26Izquierda - 0.18Otros - 0.04RentaNetaMediaPersona - 0.10RentaNetaMediaHogar - 0.35TotalEmpresasMunicipio - 0.35NumeroDeclaraciones - 0.34ParoTotal - 0.21TotalContratos - 0.32TotalLicencias - 0.25NumeroTrabajadores - 0.32NumeroEstablecimientos$$

Para la segunda componente, se encuentran signos tanto positivos como negativos en los coeficientes, aunque son mucho mayores los de las variables de renta que los pertenecientes al resto de variables, teniendo ambas signo negativo:

$$PC2 = 0.03Población + 0.02Derecha + 0.04Izquierda - 0.08Otros - 0.71RentaNetaMediaPersona - 0.68RentaNetaMediaHogar + 0.06TotalEmpresasMunicipio + 0.003NumeroDeclaraciones + 0.08ParoTotal - 0.007TotalContratos + 0.03TotalLicencias + 0.04NumeroTrabajadores + 0.06NumeroEstablecimientos$$

	Coordenadas1	Coordenadas2	Coordenadas3	Correlacion1	Correlacion2	Correlacion3
Poblacion	0.89	-0.04	-0.23	0.89	-0.04	-0.23
Derecha	0.78	-0.02	-0.16	0.78	-0.02	-0.16
Izquierda	0.65	-0.05	-0.15	0.65	-0.05	-0.15
Otros	0.45	0.11	0.00	0.45	0.11	0.00
Renta neta media por persona	0.09	0.90	-0.07	0.09	0.90	-0.07
Renta neta media por hogar	0.25	0.86	0.07	0.25	0.86	0.07
Total Empresas por Municipio	0.86	-0.07	-0.14	0.86	-0.07	-0.14
Numero Total Declaraciones	0.86	0.00	-0.23	0.86	0.00	-0.23
Paro Total	0.84	-0.10	-0.22	0.84	-0.10	-0.22
Total Contratos	0.51	0.01	0.64	0.51	0.01	0.64
Numero Total Licencias	0.81	-0.04	0.04	0.81	-0.04	0.04
Numero Trabajadores	0.61	-0.05	0.54	0.61	-0.05	0.54
Numero Establecimientos	0.79	-0.07	0.29	0.79	-0.07	0.29

	Contribucion1	Contribucion2	Contribucion3	Cos2.1	Cos2.2	Cos2.3
Poblacion	12.77	0.09	4.95	0.79	0.00	0.05
Derecha	9.85	0.03	2.58	0.61	0.00	0.03
Izquierda	6.84	0.17	2.32	0.42	0.00	0.02
Otros	3.28	0.77	0.00	0.20	0.01	0.00
Renta neta media por persona	0.14	50.81	0.49	0.01	0.81	0.01
Renta neta media por hogar	1.01	46.61	0.43	0.06	0.75	0.00
Total Empresas por Municipio	11.91	0.32	1.99	0.73	0.01	0.02
Numero Total Declaraciones	12.00	0.00	5.22	0.74	0.00	0.05
Paro Total	11.33	0.58	4.70	0.70	0.01	0.05
Total Contratos	4.23	0.00	40.55	0.26	0.00	0.42
Numero Total Licencias	10.55	0.11	0.17	0.65	0.00	0.00
Numero Trabajadores	6.02	0.16	28.19	0.37	0.00	0.29
Numero Establecimientos	10.06	0.35	8.43	0.62	0.01	0.09

Ilustración 29. Contribución por Variables

Todas las variables están correladas positivamente y en grado alto con la primera componente, excepto las rentas, como puede observarse en la columna correspondiente a Coordenadas1. Se observa de nuevo, que, en el segundo eje, las variables que presentan mayor correlación con dicha dimensión son los dos tipos de renta (Coordenadas2).

Si se analizan las contribuciones de las variables a las 2 primeras componentes, que indican lo que ha contribuido cada variable a la definición de cada uno de los ejes:

- Para la primera componente las contribuciones más bajas corresponden a los dos tipos de renta, como se ve en la columna Contribucion1.
- Para la segunda, las rentas son las que aportan una mayor contribución como puede verse en la columna Contribucion2.

Por otro lado, los cosenos cuadrados informan de la calidad de la representación de la variable en la dimensión correspondiente:

- Para la primera componente todas las variables presentan cierta calidad de representación excepto las rentas (Cos2.1).
- En la segunda, las únicas variables bien representadas son las rentas (Cos2.2).

Tras este análisis no se obtienen conclusiones muy claras en cuanto a establecer relaciones entre las variables electorales y las económicas, pero si se han observado ciertos comportamientos:

- La primera dimensión representa el comportamiento de todas las variables a excepción de las rentas, mientras que la segunda componente está representada por las rentas.
- Con respecto a la primera dimensión todas las variables parecen crecer de forma simultánea, aunque las rentas en menor medida.

Puesto que mediante el **análisis de componentes principales** centrado en las dos primeras dimensiones no se obtienen las relaciones buscadas, antes de concluir que no hay relación entre las variables económicas y las electorales, se representarán los gráficos con las coordenadas de las variables por pares de componentes.

La primera dimensión, con la que se obtiene un 47.43% de variabilidad explicada, parece representar gran parte de las características económicas de los municipios, encontrándose entre las variables que más contribuyen dentro de dicha componente: Población, Número Total Declaraciones, Total Empresas por Municipio, Paro Total, Número Total Licencias y Número Establecimientos.

$$
\begin{aligned}
PC1 = -0.36Población - 0.31Derecha - 0.26Izquierda - 0.18Otros \\
- 0.04RentaNetaMediaPersona - 0.10RentaNetaMediaHogar \\
- 0.35TotalEmpresasMunicipio - 0.35NumeroDeclaraciones \\
- 0.34ParoTotal - 0.21TotalContratos - 0.32TotalLicencias \\
- 0.25NumeroTrabajadores - 0.32NumeroEstablecimientos
\end{aligned}
$$

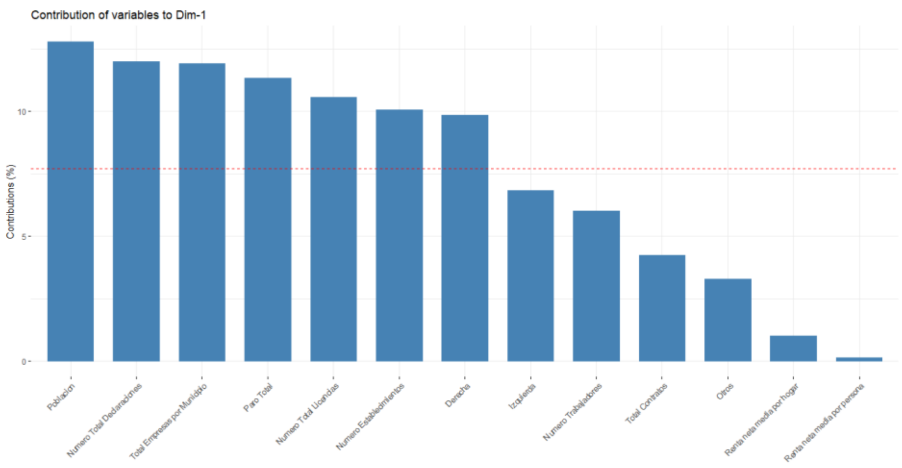


Ilustración 30. Contribuciones a la Componente 1

La segunda componente, representa los dos tipos de rentas, puesto que estas dos variables son las que presentan mayor importancia en la dimensión 2.

$$
\begin{aligned}
PC2 = 0.03Población + 0.02Derecha + 0.04Izquierda - 0.08Otros \\
- 0.71RentaNetaMediaPersona - 0.68RentaNetaMediaHogar \\
+ 0.06TotalEmpresasMunicipio + 0.003NumeroDeclaraciones \\
+ 0.08ParoTotal - 0.007TotalContratos + 0.03TotalLicencias \\
+ 0.04NumeroTrabajadores + 0.06NumeroEstablecimientos
\end{aligned}
$$

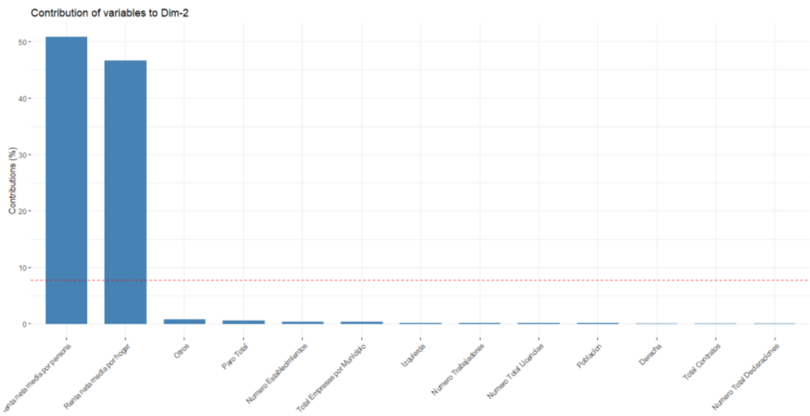


Ilustración 31. Contribuciones a la Componente 2

La cuarta dimensión, se utilizará para la representación de la elección de voto a otros partidos, puesto que, es la variable Otros la que presenta una mayor contribución en la componente 4.

$$\begin{aligned} PC4 = {} & 0.05Población + 0.22Derecha + 0.03Izquierda - 0.92Otros \\ & + 0.05RentaNetaMediaPersona + 0.10RentaNetaMediaHogar \\ & - 0.03TotalEmpresasMunicipio + 0.04NumeroDeclaraciones \\ & + 0.004ParoTotal - 0.16TotalContratos + 0.006TotalLicencias \\ & + 0.24NumeroTrabajadores + 0.09NumeroEstablecimientos \end{aligned}$$

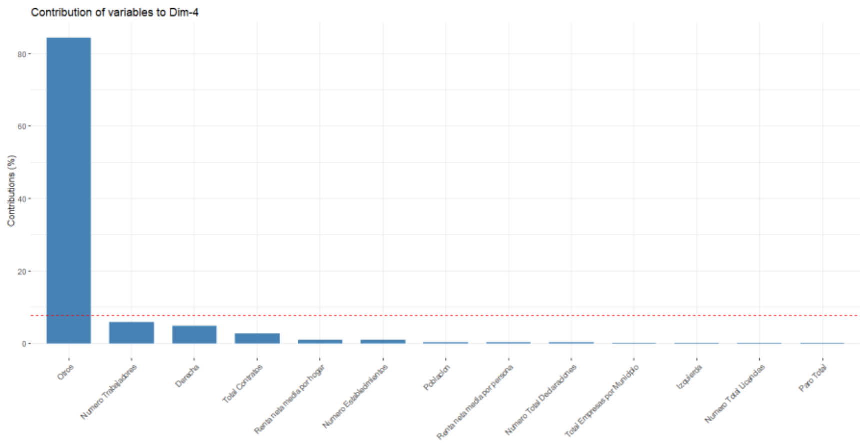


Ilustración 32. Contribuiones a la Componente 4

La componente 5, representará a la tendencia de voto a partidos de izquierdas, aunque también a las dos variables económica que faltan por considerar, puesto que las variables que aportan una mayor contribución en esta dimensión son: Izquierda, Total Contratos y Número Trabajadores.

$$\begin{aligned} PC5 = {} & -0.095Población - 0.108Derecha + 0.668Izquierda - 0.226Otros \\ & + 0.0117RentaNetaMediaPersona + 0.00371RentaNetaMediaHogar \\ & - 0.0507TotalEmpresasMunicipio - 0.059NumeroDeclaraciones \\ & - 0.000026ParoTotal + 0.5TotalContratos + 0.126TotalLicencias \\ & - 0.4NumeroTrabajadores - 0.236NumeroEstablecimiento \end{aligned}$$

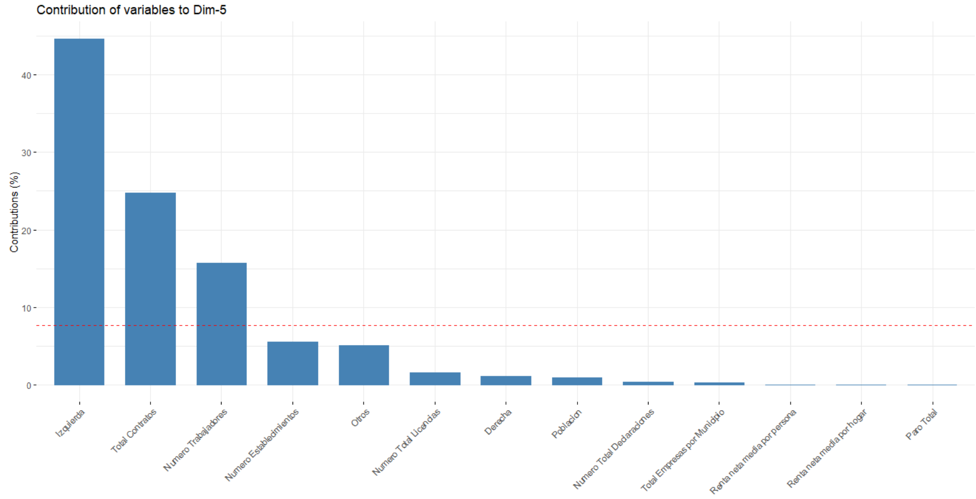


Ilustración 33. Contribuciones a la Componente 5

Por último, la octava dimensión corresponde a los votos a partidos considerados de derechas, puesto que la variable con mayor contribución en la componente 8 es Derecha.

$$
\begin{aligned}
PC8 = {} & -0.29Población + 0.76Derecha + 0.15Izquierda + 0.14Otros \\
& + 0.22RentaNetaMediaPersona - 0.25RentaNetaMediaHogar \\
& - 0.01TotalEmpresasMunicipio - 0.27NumeroDeclaraciones \\
& - 0.13ParoTotal + 0.02TotalContratos - 0.25TotalLicencias \\
& - 0.06NumeroTrabajadores + 0.16NumeroEstablecimiento
\end{aligned}
$$

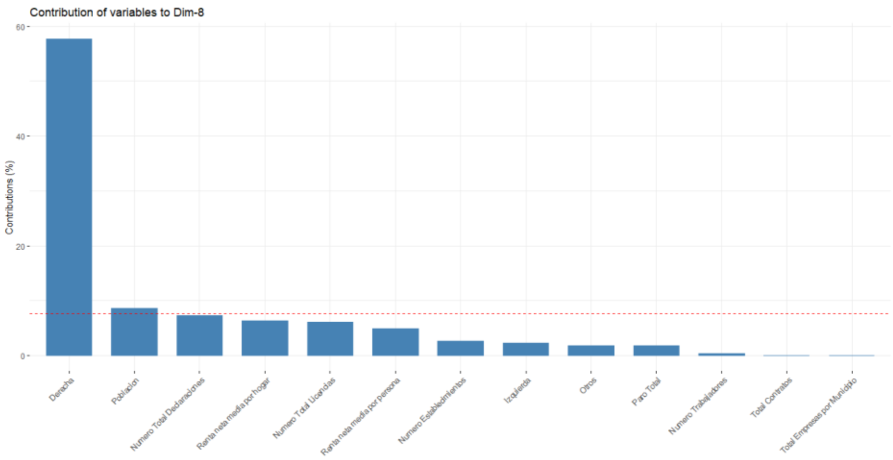


Ilustración 34. Contribuciones a la componente 8

Puesto que todas las variables disponibles se encuentran representadas por alguna de las 5 componentes analizadas anteriormente (1,2,4,5 y 8), serán las utilizadas para crear los planos por pares de componentes, obteniéndose 10 gráficos diferentes.

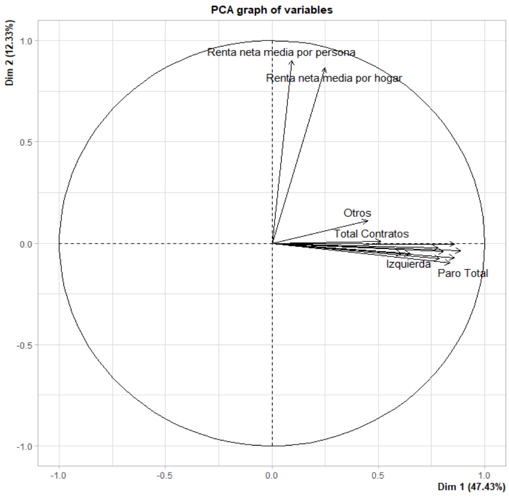


Ilustración 35. Plano por componentes 1-2

En el gráfico superior se observan dos grupos de variables, por un lado, las rentas, que presentan coordenadas próximas a 1 con la segunda dimensión y próximas a 0 con la primera, y por otro, el resto de variables, en cuyo caso ocurre contrario. Puesto que las rentas muestran coordenadas próximas a 0 con la primera componente, no presentan relación lineal con los valores que toman los municipios en dicho eje. Ninguna de las variables está bien representada sobre el plano formado por las 2 componentes, aunque lo estén en cada una de ellas por separado. Por lo tanto, se puede decir que los municipios con rentas altas por persona también presentarán rentas altas por hogar.

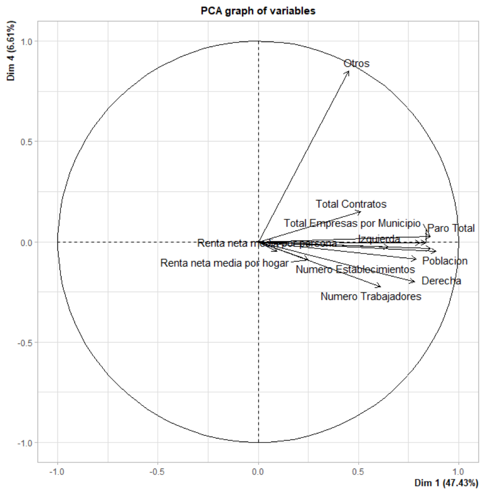


Ilustración 36. Plano por componentes 1-4

En el plano anterior, se observan 2 grupos de variables diferenciados. Por un lado, la variable Otros, que presenta un valor de coordenada próxima a 1 con la componente 4, en contraposición con el resto de variables, cuya coordenada con la cuarta dimensión no supera 0.25 (en valor absoluto). Además, la variable Otros es la mejor representada en el plano formado por las componentes 1 y 4, puesto que es la única variable que se encuentra en el borde del círculo y sus coordenadas con ambos ejes son relativamente altas. Los municipios con un número de votos alto para la derecha, presentan un número alto de trabajadores y un gran número de habitantes. En el caso de los votos a la izquierda, los municipios con valores altos presentan gran número de parados y un valor alto de población.

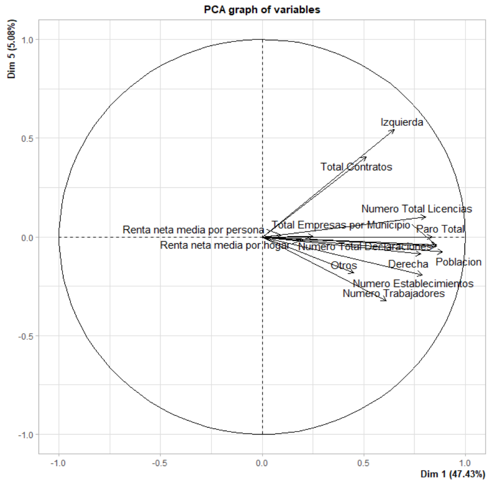


Ilustración 37. Plano pr componentes 1-5

En el gráfico anterior, se aprecia como la variable Izquierda se encuentra junto con Total Contratos, diferenciada del resto de variables, puesto que presentan coordenadas muy similares en los dos ejes. Además, como las coordenadas que presentan en ambas dimensiones son altas, son las variables mejor representadas en el plano. Los municipios con número un número alto de votos a la izquierda presentan también un número alto de contratos. Por otra parte, los municipios con gran número de votos a la derecha presentarán un gran número de habitantes, un número alto de trabajadores y un gran número de establecimientos. Para la variable Otros, se aprecia una dirección de crecimiento similar a la que presentan Número Trabajadores y Número de Establecimientos.

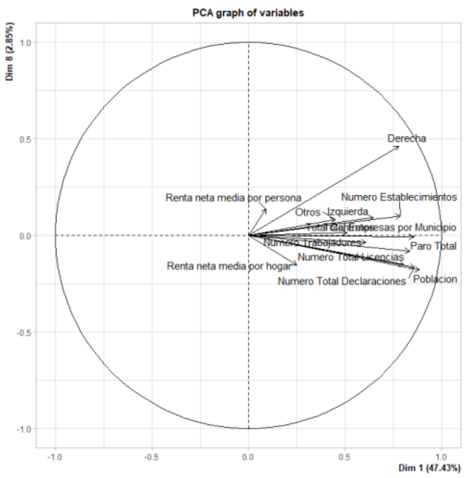


Ilustración 38. Plano por componentes 1-8

En el último plano cuyo primer eje es la componente 1, no se observan grandes diferencias entre grupos de variables. La variable Derecha es la mejor representada por el plano formado por la primera y octava dimensión, puesto que es la que presenta mayores coordenadas en ambos ejes a la vez. Sin embargo, este gráfico no resulta útil para establecer relaciones interesantes entre variables electorales y económicas. Simplemente, se establece que aquellos municipios con valores altos en el número de votos para cualquiera de las 3 opciones, presentarán gran número de establecimientos y una renta neta media por persona alta, puesto que todas estas variables presentan direcciones de crecimiento similares.

Si se analizan los 4 gráficos anteriores, se puede ver que la primera componente resulta ser factor de tamaño, puesto que todas las coordenadas de las variables en el primer eje tienen signo positivo. Al desplazarse en la dirección de crecimiento de esta dimensión, todas las variables crecen a la vez, o decrecen al moverse en la dirección contraria.

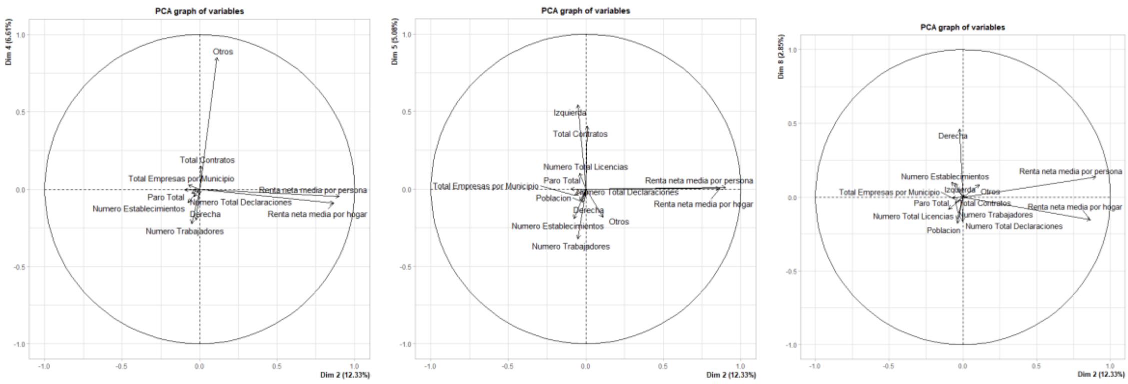


Ilustración 39. Plano por componentes 2-4, 2-5 y 2-8

En el gráfico correspondiente a las dimensiones 2 y 4, se observan 4 grupos de variables diferentes. Por un lado, con respecto a la componente 2, se puede ver que las rentas presentan coordenadas positivas y altas, mientras que Paro Total, Número Establecimientos y Total Empresas por Municipio presentan coordenadas negativas y pequeñas, de tal forma que, si aumenta el valor de uno de los dos grupos anteriores, el otro disminuye. Con respecto a la componente 4 se observan otros dos grupos de variables, el primero de ellos formado por Otros y Total Contratos, cuyas coordenadas en el eje son positivas, y el segundo formado por Número Total Declaraciones, Derecha y Número Trabajadores, cuyas coordenadas son negativas. En cuanto a las relaciones entre variables electorales y económicas en este gráfico se observa que los grupos de municipios con un número alto de votos a la derecha presentan un gran número de declaraciones y de trabajadores. Los municipios con un gran número de votos a otros partidos presentan un número alto de contratos.

En la representación de las dimensiones 2 y 5, se observan diferencias entre grupos de variables. En primer lugar, para la dimensión 2, los dos tipos de renta presentan coordenadas positivas, mientras que Paro Total, Número Establecimientos y Total Empresas por Municipio presentan coordenadas negativas para esta misma componente. Para la dimensión 5, se encuentra la misma diferencia para dos grupos de variables, el primero formado por Izquierda, Total Contratos y Número Total Licencias, y el segundo por Derecha, Número Establecimientos y Número Trabajadores. Por lo tanto, los municipios con mayor número de votos a la izquierda presentarán mayor número de contratos y de licencias, mientras que los que tengan mayor número de votos a la derecha presentarán mayores números de establecimientos y trabajadores.

En el último plano, correspondiente a las dimensiones 2 y 8, no se observan diferencias tan marcadas entre las variables como en los casos anteriores, aunque se puede ver que, de nuevo, los dos tipos de renta presentan coordenadas positivas y altas con la dimensión 2. Tampoco resulta útil para establecer relaciones entre variables económicas y electorales, puesto que las 3 variables correspondientes al segundo grupo se encuentran bastante concentradas en el gráfico, aunque podría decirse que los municipios con mayor valor para Derecha presentan valores altos para Número Establecimientos y que Otros y Renta neta media por persona presentan una dirección de crecimiento similar.

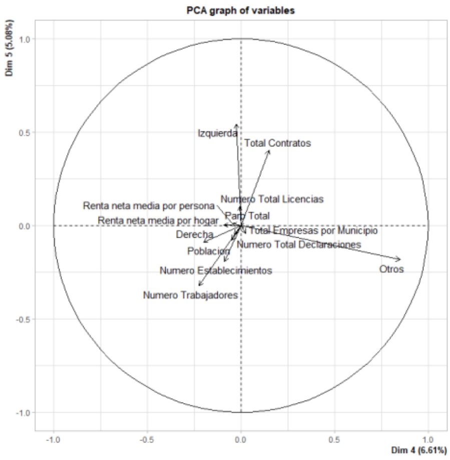


Ilustración 40. Plano por componentes 4-5

Aunque en el caso del gráfico superior, correspondiente a las componentes 4 y 5, no se encuentran agrupaciones claras de variables por valores extremos de sus coordenadas en cada uno de los ejes, esta representación resulta muy interesante para establecer las relaciones buscadas, puesto que ninguna de las variables electorales presenta la misma dirección de crecimiento. Los municipios con gran número de votos a la derecha, presentan también altos valores para Población, Número de Establecimientos y Número de Trabajadores. En cuanto a los votos para los partidos de izquierda, serán más altos en aquellos municipios con mayor número de contratos, y también crecerán con el paro y el número total de licencias. Por último, los votos a otros partidos serán mayores en aquellos municipios con mayor número de declaraciones y de empresas.

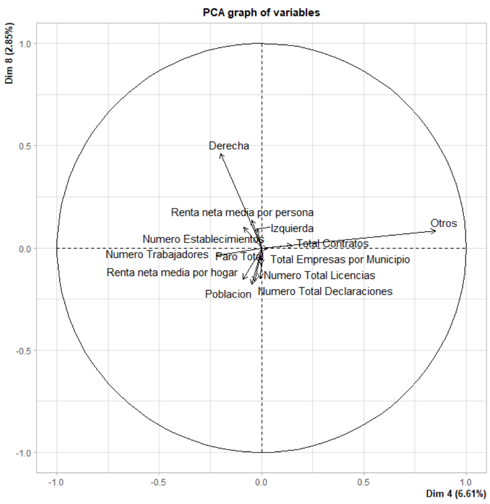


Ilustración 41. Plano por componentes 4-8

En el último plano correspondiente a las dimensiones 4 y 8, no se establecen relaciones claras entre las variables económicas y las electorales, puesto que Derecha e Izquierda presentan la misma dirección de crecimiento. Los municipios con mayor número de votos a derecha e izquierda presentarán una renta neta media por persona más alta y mayor número de establecimientos y aquellos con mayor número de votos a otros partidos presentarán mayor número de contratos.

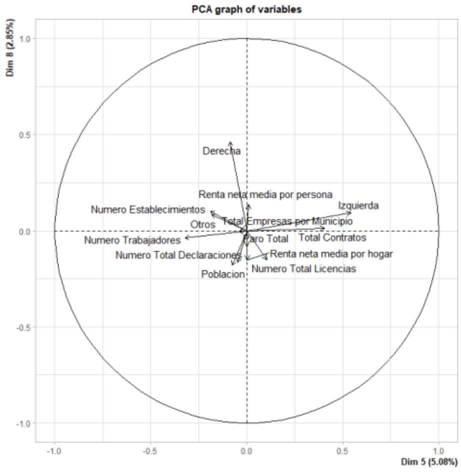


Ilustración 42. Plano por componentes 5-8

En el último plano obtenido, formado por la componente 5 en el primer eje y la 8 en el segundo, se observan relaciones interesantes entre las variables electorales y las económicas. Los grupos de municipios con mayor número de votos a la derecha presentan una gran renta neta media por persona y un alto número de empresas. Además, este tipo de voto disminuirá a medida que aumente el paro. Los municipios con un número alto de votos a la izquierda, presentarán un mayor número de contratos con respecto al resto y la variable Izquierda aumentará a medida que disminuya el número de trabajadores. Por último, los municipios con mayor número de votos a otros partidos, presentarán mayor número de trabajadores, declaraciones, población y establecimientos. Además, se observa que izquierda y otros presentan direcciones opuestas de crecimiento, lo que quiere decir que al aumentar una de las variables disminuirá la otra.

Análisis de Correlaciones Canónicas

Puesto que los resultados obtenidos mediante el análisis de componentes principales no han resultado de gran utilidad para el propósito del trabajo, se realizará un **análisis de correlaciones canónicas**. Esta técnica multivariante es adecuada cuando las variables disponibles pueden dividirse en dos grupos en función de algún criterio, y su principal utilidad es estudiar las relaciones entre ambos conjuntos. En este caso las variables se dividirán en 3 electorales:

- X1=Derecha
- X2=Izquierda
- X3=Otros

Y 10 variables económicas:

- Y1=Población
- Y2=Renta neta media por persona
- Y3=Renta neta media por hogar
- Y4=Total Empresas por Municipio
- Y5=Número Total Declaraciones
- Y6=Paro Total
- Y7=Total Contratos
- Y8=Número Total Licencias
- Y9=Numero Trabajadores
- Y10=Numero Establecimientos

Este tipo de análisis presenta cierta similitud con la técnica de componentes principales, en cuanto al estudio de autovalores y autovectores. Sin embargo, las dimensiones se estructurarán mediante correlaciones canónicas, que permiten estudiar la relación entre dos vectores aletatorios, de dimensiones p y q, como se muestra a continuación:

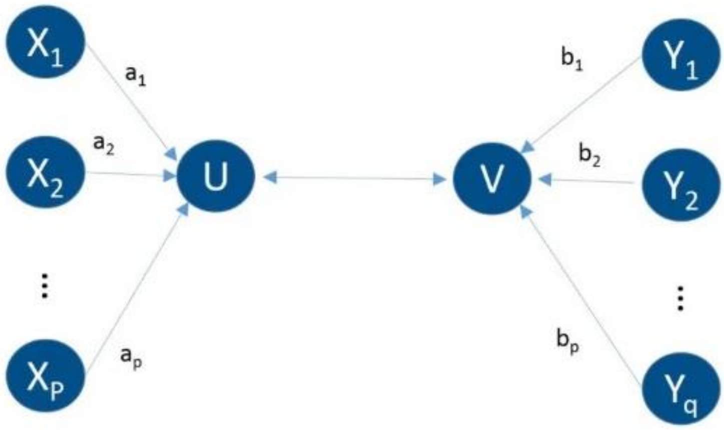


Ilustración 43. Representación Variables Canónicas

En este caso, se tienen dos vectores $X = (X1, X2, X3)$ e $Y = (Y1, Y2, \dots, Y10)$, y el análisis de correlaciones canónicas trata de encontrar dos variables U y V tal que:

$$U = a1X1 + a2X2 + a3X3$$

$$V = b1Y1 + \cdots + b10Y10$$

de forma que la correlación entre U y V sea máxima.

Una vez finalizado este proceso se habrán obtenido las primeras variabes canónicas (U y V), los primeros vectores canónicos (a y b) y la primera correlación canónica ($corr(U, V)$), de tal forma que las primeras variables canónicas presentan la máxima correlación entre una combinación lineal de X y de Y. Aunque existen otras relaciones posibles entre los dos conjuntos de variables, siempre serán "menores".

En concreto, el número de parejas de variables canónicas queda definido por:

$$\min(3,10) = 3$$

Y todas las variables canónicas serán independientes entre sí.

En primer lugar, interesa analizar las correlaciones simples entre variables electorales y económicas, puesto que son los dos grupos que se pretende relacionar:

	Y1	Y2	Y3	Y4	Y5	Y6	Y7	Y8	Y9	Y10
X1	0.68	0.06	0.17	0.66	0.66	0.65	0.31	0.58	0.39	0.56
X2	0.53	0.04	0.10	0.51	0.54	0.54	0.29	0.48	0.30	0.44

X3	0.35	0.095	0.13	0.37	0.35	0.35	0.25	0.31	0.18	0.31

Tabla 18. Correlaciones entre variables económicas y electorales

Puede verse que todas estas correlaciones son bastante suaves, aunque mayores en todos los casos para Derecha (X1), seguidas de Izquierda (X2) y por último Otros (X3), si se compara con una misma variable en el grupo de las económicas. La variable que mayor correlación presenta con las variables electorales es la Población (Y1), con un valor de 0.68 para con Derecha (X1), 0.63 con Izquierda (X2) y 0.35 con Otros (X3). Por otro lado, las variables Total Empresas por Municipio (Y4), Número Total Declaraciones (Y5) y Paro Total (Y6), también presentan correlaciones destacables con las 3 variables electorales.

El siguiente paso será analizar las variables canónicas y su significancia:

```
Canonical correlation analysis of:
         3   Electorales  variables:
             Derecha, Izquierda, Otros
  with   10   Economicas  variables:
             Poblacion, Renta neta media por persona,
             Renta neta media por hogar, Total Empresas por Municipio,
             Numero Total Declaraciones, Paro Total, Total Contratos,
             Numero Total Licencias, Numero Trabajadores, Numero Establecimientos
```

```
    CanR  CanRSQ   Eigen percent    cum                         scree
1 0.8291 0.68746 2.19961  98.635  98.63 ******************************
2 0.1378 0.01899 0.01936   0.868  99.50
3 0.1047 0.01096 0.01108   0.497 100.00

Test of H0: The canonical correlations in the
current row and all that follow are zero

     CanR LR test stat approx F numDF denDF Pr(> F)
1 0.82913      0.30324  21.7516    30  1301  <2e-16 ***
2 0.13780      0.97026   0.7505    18   888  0.7594
3 0.10470      0.98904   0.6166     8   445  0.7642
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Ilustración 44. Variables Canónicas y su significancia

La primera correlación canónica es 0.8291, considerablemente mayor que cualquiera de las correlaciones obtenidas entre los dos conjuntos de variables.

El p-valor para contrastar la hipótesis nula:

$$H0: Las\ correlaciones\ canónicas\ en\ la\ fila\ actual\ y\ las\ siguientes\ son\ 0$$

es próximo a 0 para la primera correlación canónica, 0.7594 para la segunda y 0.7642 para la tercera, por lo tanto, se rechazará en el primer caso y en los otros dos no se podrá, lo que supone que únicamente se considere la primera correlación canónica para el análisis, obteniéndose:

	Xcan1	Xcan2	Xcan3
X1	-0.68	0.67	-0.59
X2	-0.37	-0.217	1.04
X3	-0.26	-0.90	-0.45

Tabla 19. Coeficientes Estandarizados Variables Electorales

$$Xcan1 = -0.68Derecha - 0.37Izquierda - 0.26Otros$$

	Ycan1	Ycan2	Ycan3
Y1	-0.12	0.73	-0.71
Y2	-0.04	-0.46	-0.0009
Y3	-0.02	0.11	-0.59
Y4	-0.23	-0.29	-0.99
Y5	-0.25	-0.35	1.08
Y6	-0.29	-0.09	0.45
Y7	-0.07	-0.87	0.16
Y8	-0.10	0.33	0.42
Y9	0.02	0.62	-0.008
Y10	-0.14	-0.24	-0.14

Tabla 20. Coeficientes Estandarizados Variables Económicas

$$\begin{aligned} Ycan1 = {} & -0.12Población - 0.04Renta\ neta\ media\ por\ persona \\ & - 0.02Renta\ neta\ media\ por\ hogar \\ & - 0.23Total\ Empresas\ por\ Municipio \\ & - 0.25Número\ Total\ Declaraciones - 0.29Paro\ Total \\ & - 0.07Total\ Contratos - 0.10Número\ Total\ Licencias \\ & + 0.02Número\ Trabajadores - 0.14Número\ Establecimientos \end{aligned}$$

	Xcan1	Xcan2	Xcan3
X1	-0.90	0.38	-0.24
X2	-0.72	-0.14	0.68
X3	-0.50	-0.80	-0.34

Tabla 21. Correlaciones entre variables electorales y variables canónicas del mismo grupo

	Xcan1	Xcan2	Xcan3
Y1	-0.75	0.03	-0.009
Y2	-0.08	-0.05	-0.04
Y3	-0.19	-0.03	-0.05
Y4	-0.73	-0.003	-0.03
Y5	-0.74	0.006	0.02
Y6	-0.73	0.01	0.02
Y7	-0.38	-0.08	0.009
Y8	-0.65	0.01	0.01
Y9	-0.43	0.04	-0.002
Y10	-0.62	0.0002	-0.009

Tabla 22. Correlaciones entre variables económicas y variables canónicas del grupo opuesto

Las correlaciones entre las 3 variables electorales y la primera variable canónica presentan signo negativo en todos los casos, igual que los coeficientes analizados anteriormente.

	Ycan1	Ycan2	Ycan3
X1	-0.74	0.05	-0.03
X2	-0.60	-0.02	0.07
X3	-0.41	-0.11	-0.04

Tabla 23. Correlaciones entre variables electorales y variables canónicas del grupo opuesto

	Ycan1	Ycan2	Ycan3
Y1	-0.90	0.23	-0.08
Y2	-0.10	-0.40	-0.34
Y3	-0.23	-0.19	-0.52
Y4	-0.88	-0.02	-0.28
Y5	-0.89	0.05	0.21
Y6	-0.88	0.08	0.15
Y7	-0.45	-0.57	0.09
Y8	-0.79	0.07	0.10
Y9	-0.52	0.28	-0.02
Y10	-0.75	0.001	-0.08

Tabla 24. Correlaciones entre variables económicas y variables canónicas del mismo grupo

Sin embargo, en el caso de las variables económicas, se observa que el coeficiente correspondiente a la variable Número Trabajadores era positivo, mientras que su correlación con la primera variable canónica no lo es, lo que quiere decir que se trata de una **variable supresora**, ya que presenta signos opuestos en su coeficiente y en su correlación.

El siguiente paso, ha sido un análisis de redundancia canónico, cuyos resutados se muestran a continuación:

```
Redundancies for the Electorales variables & total X canonical redundancy

    Xcan1     Xcan2     Xcan3 total X|Y
 0.358627  0.005051  0.002328  0.366006

Redundancies for the Economicas variables & total Y canonical redundancy

    Ycan1     Ycan2     Ycan3 total Y|X
0.3353161 0.0012540 0.0006142 0.3371844
```

Ilustración 45. Análisis de redundancia canónico

Puede observarse que ninguna de las dos primeras variables canónicas (Xcan1 e Ycan1) es un buen predictor global del conjunto opuesto de variables, puesto que las proporciones de varianza explicadas son 0.36 y 0.34, respectivamente. Mediante el coeficiente de redundancia total, podemos ver que ni las variables electorales ni las económicas presentan buena capacidad global para predecir las del grupo opuesto, ya que los valores obtenidos son 0.366 para el primer grupo y 0.337 para el segundo.

	Ycan1	Ycan2	Ycan3
X1	0.5509	0.5535	0.5542
X2	0.3568	0.3571	0.3622
X3	0.1682	0.1803	0.1816

Ilustración 46. Correlaciones múltiples cuadradas entre v.electorales y v.canónicas opuestas

Puede verse que la primera variable canónica de las variables económicas (Ycan1) tiene buena capacidad predictiva para Derecha (X1), aunque más pobre para Izquierda (X2) y aún menor para Otros (X3), puesto que las correlaciones múltiples al cuadrado que presenta dicha variable canónica con las 3 variables electorales son 0.55, 0.36 y 0.17, respectivamente.

	Xcan1	Xcan2	Xcan3
Y1	0.5615	0.5624	0.5624
Y2	0.0065	0.0095	0.0108
Y3	0.0365	0.0372	0.0402
Y4	0.5341	0.5341	0.5349
Y5	0.5452	0.5453	0.5457
Y6	0.5342	0.5343	0.5346
Y7	0.1420	0.1482	0.1483
Y8	0.4250	0.4251	0.4252
Y9	0.1838	0.1853	0.1853
Y10	0.3843	0.3843	0.3844

Ilustración 47. Correlaciones múltiples cuadradas entre v.económicas y v.canónicas opuestas

Por otro lado, la primera variable canónica de las variables electorales (Xcan1) resulta un buen predictor para Población (Y1), Total Empresas por Municipio (Y4), Número Total Declaraciones (Y5) y Paro Total (Y6), puesto que los valores correspondientes de la correlación múltiple al cuadrado son 0.5615, 0.5341, 0.5452, 0.5342. Además, Xcan1 resulta un predictor más pobre para Número Total Licencias (Y8) y Número Establecimientos (Y10) y casi inservible para Renta neta media por persona (Y2) y Renta neta media por hogar (Y3).

Análisis Clúster

Teniendo en cuenta que los resultados de los dos análisis anteriores no han resultado concluyentes, se realizará un **análisis clúster** sobre los municipios para ver si con las características de los grupos que se obtengan se puede deducir alguna relación entre las variables económicas y las electorales. Además, se utilizará para intentar esclarecer la sospecha de que la variable Población enmascara los resultados reales.

Esta técnica realiza un aprendizaje no supervisado, puesto que se desconocen los grupos iniciales a los que pertenece cada individuo, buscando crear subconjuntos similares entre sí en relación a una determinada medida de similitud entre los municipios. Por lo tanto, su finalidad es obtener particiones de los datos de tal forma que los municipios dentro de cada una de ellas sean lo más similares posibles entre sí y que difieran lo máximo posible con respecto a los municipios del resto de particiones.

Se utilizarán dos métodos clúster diferentes, para posteriormente comparar sus resultados: método k-means y método jerárquico, estandarizando de nuevo las variables para que todas tengan una importancia equivalente en el método utilizado y no dependan de la escala en la que están representadas.

Método k-means

Este método establece K particiones de los datos distintas y que no se solapan, aunque hay que seleccionar el número de grupos requerido antes de aplicar el algoritmo. Si se denomina n al número de individuos disponibles y K al número de grupos que se quieren crear, el algoritmo de k-means proporcionará un óptimo local de las K^n formas posibles que se tienen de agrupar todas las observaciones:

- En primer lugar, de los K grupos buscados, se asigna uno aleatoriamente a cada individuo.
- Se repiten los 2 siguientes pasos hasta que no cambie el resultado que se obtiene, sabiendo que entonces habrá alcanzado el óptimo buscado:
    - Para cada grupo se calcula el centroide, como el vector de medias de las variables para las observaciones que se encuentran en ese clúster.
    - Se asigna cada individuo al grupo cuyo centroide esté más cerca.

El primer paso a realizar es estimar el número óptimo de clústeres. Para esta tarea se utilizará el método del codo, que puede automatizarse mediante la función de R "fviz_nbclust" que además de obtener el número de grupos buscado, genera una representación gráfica de los resultados:

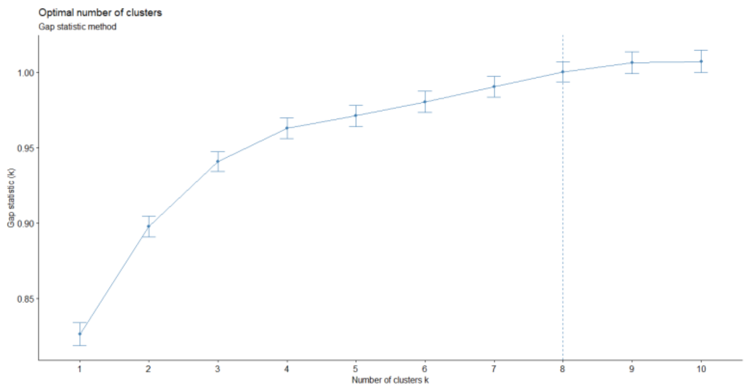


Ilustración 48. Número óptimo de clústeres

Como se puede observar en el gráfico superior, el número óptimo de clústeres para los datos es 8 y tras aplicar el algoritmo k-means para el número óptimo de grupos se obtiene:

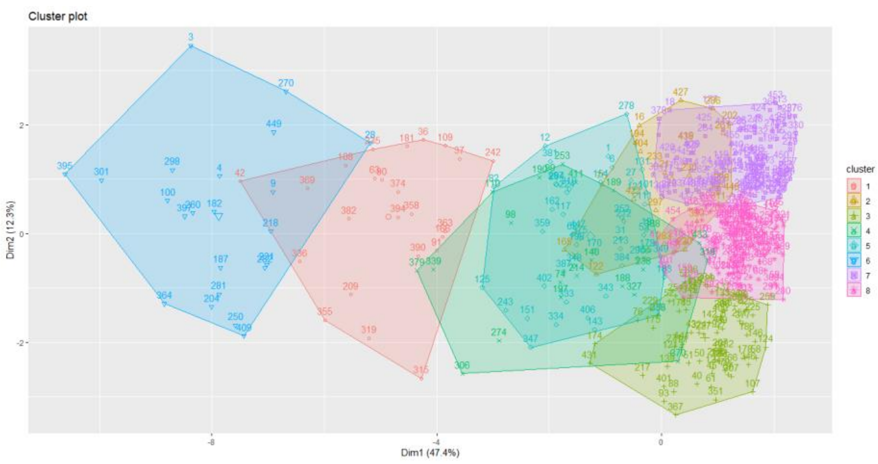


Ilustración 49. Resultado Algoritmo K-means

Método jerárquico

Este método presenta una ventaja con respecto a k-means, y es que no necesita que se establezca a priori el número de grupos que se quieren obtener. Los métodos jerárquicos obtienen representaciones con forma de árbol que se denominan dendogramas y por lo tanto se mostrará una figura jerárquica entre los municipios. A medida que se asciende en el dendograma se van uniendo las diferentes ramas, de tal forma que aquellos municipios que se encuentren dentro de la misma unión serán similares entre sí. Las uniones que estén en la parte baja del dendograma contendrán municipios más similares que las que se encuentren en la parte alta. Por lo tanto, las conclusiones que puedan extraerse de este tipo de gráficos residirán en la altura a la que se unen los diferentes municipios, ya que, en base a esta recta horizontal podrá verse que municipios se encuentran dentro de cada grupo.

Por otro lado, hay que elegir la medida de similitud, que en este caso será la distancia euclídea:

- Para dos individuos i y j, su distancia euclídea denominada D_{ij} será:

$$D_{ij} = \sqrt{\sum_{k=1}^{n} (x_{ki} - x_{kj})^2}$$

  Siendo X_{ki} el valor de la variable x_k para el individuo i y X_{kj} el valor de la variable x_k para el individuo j.

Finalmente, hay que seleccionar el algoritmo que medirá la distancia entre clústeres:

- Complete Linkage: calcula por pares la distancia entre los elementos de 2 grupos distintos y selecciona la máxima que se encuentre, siendo ésta la distancia entre ambos.
- Average Linkage: funciona de la misma forma que el método anterior, pero, en este caso, la distancia entre los grupos será la media de todas las obtenidas entre pares de individuos.
- Single Linkage: utiliza el mismo procedimiento, pero seleccionando el mínimo de todas las distancias.

Una vez seleccionados los métodos para medir la distancia entre individuos y entre clústeres, se formará el dendograma de manera iterativa, de tal forma que, primero se tratará cada elemento como un clúster individual y se irán uniendo los individuos más similares en cada paso hasta que todos ellos pertenezcan a un único grupo y se complete el gráfico.

Al aplicar el método, el primer paso ha sido calcular la matriz de distancia euclídea para los datos después de estandarizarlos y, posteriormente, aplicar los 3 métodos Linkage explicados para seleccionar aquel que obtenga mejores resultados.

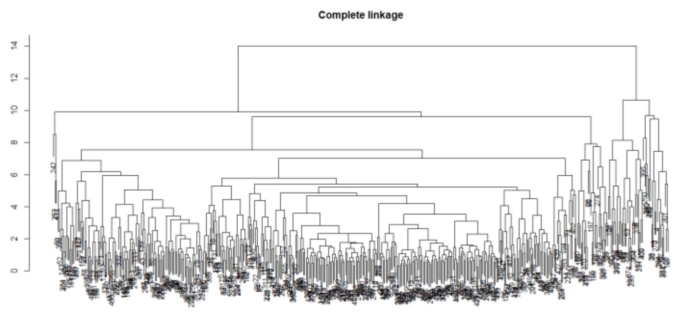


Ilustración 50. Dendograma Complete Linkage

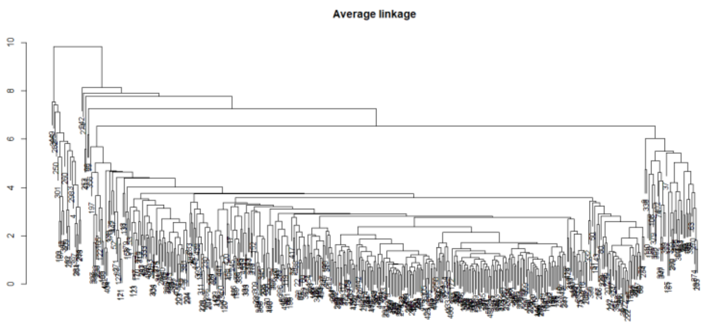


Ilustración 51. Dendograma Average Linkage

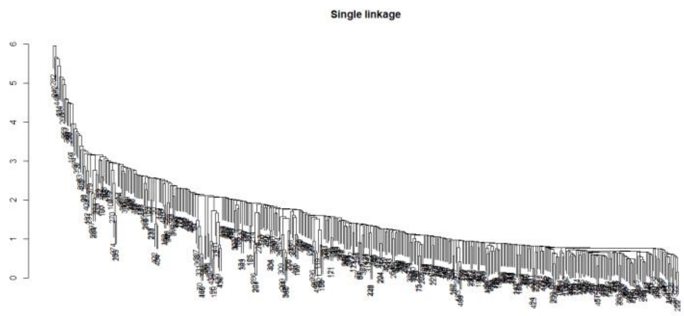


Ilustración 52. Dendograma Single Linkage

Como puede verse, el método Complete Linkage presenta un dendograma más equilibrado. Por lo tanto, se seleccionará este método para aplicar el clustering jerárquico a los datos y la distancia euclídea para medir la distancia entre los municipios, estableciendo el corte del dendograma en aquella altura que proporcione 8 grupos, que es el número óptimo de clúster que se obtenía para k-means y así se podrán comparar los resultados.

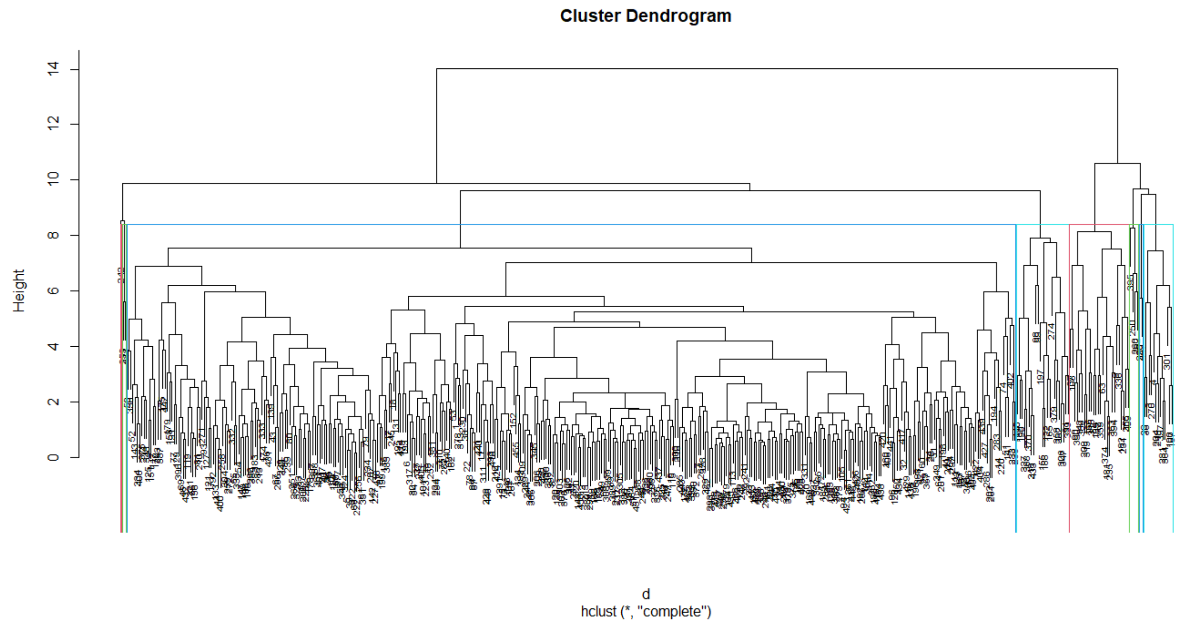


Ilustración 53. Dendograma Complete Linkage y distancia euclídea

Los resultados obtenidos utilizando el método jerárquico difieren con respecto al método k-means, por lo que el siguiente paso será seleccionar aquel que obtenga mejores resultados. Para la selección del mejor algoritmo se utilizarán 3 medidas diferentes. Por un lado, la conectividad de cada método, que será mejor cuanto menor sea su valor y, por otra parte, el índice de Dunn y el ancho de la silueta, que, en contraposición con la primera medida deben ser lo más grandes posible.

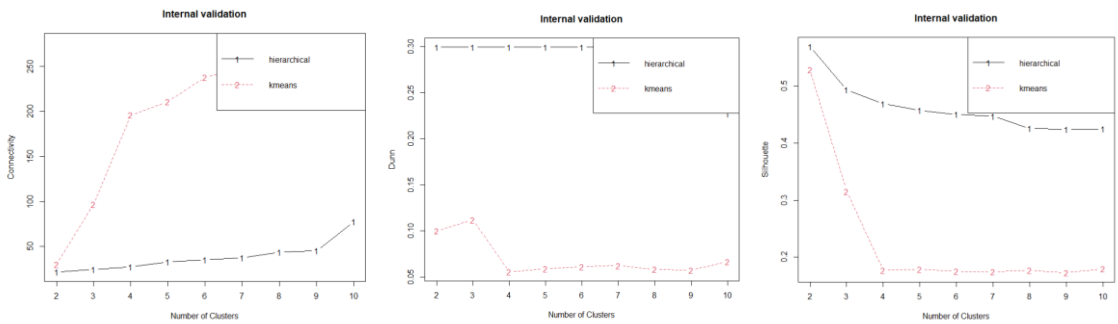


Ilustración 54. Conectividad. Índice de Dunn. Silueta

Como se puede observar en los gráficos superiores, la conectividad es menor para el método jerárquico sea cual sea el número de clústeres que se seleccione, además el índice de Dunn y el ancho de la silueta también son siempre mayores para este mismo método. Por lo tanto, se utilizará el clustering jerárquico con distancia euclídea y complete Linkage para la creación de los 8 grupos de municipios.

Como ya se ha comentdo en puntos anteriores**,** se tiene la sospecha de que la variable Población, puede condicionar los datos tanto económicos como electorales, excepto para el caso de las rentas, puesto que al tratarse de medias por municipio y no de totales, el número de habitantes no influye en el valor de dichas variables. Para contrastar esta teoría, se plantea una clasificación de los municipios según su población:

Grupo HC	Rural	Semirural	Urbano
1	321	50	14
2	0	13	0
3	0	26	0
4	8	9	6
5	0	0	1
6	0	4	0
7	0	0	2
8	0	2	0

Tabla 25. Clasifiación Clúster por tipo de municipio

De los 456 municipios municipios totales, 329 son rurales, 104 son semirurales y 23 urbanos. En cuanto a la agrupación de los individuos en los 8 clústeres, según el tipo de municipio al que pertenecen, se puede ver que de los 329 rurales, 321 se encuentran en el primer clúster, mientras que los otros dos tipos, semirural y urbano se encuentran más dispersos en los 8 grupos obtenidos. Por lo tanto, parece que la idea que se planteaba de que la Población ejerce una gran influencia en el resultado de las variables podría ser cierta.

ANOVA para clúster

Para comparar las respuestas medias a las 13 variables de los 8 grupos obtenidos con el clustering jerárquico se realizará un ANOVA, trabajando en cada análisis con dos variables. Por un lado, la variable clúster, que establece los 8 grupos de interés, y por otro, la variable respuesta, que será cada una de las 13 variables numéricas por separado. Siendo la hipótesis a contrastar en cada caso:

$$H0: \mu_1 = \mu_2 = \cdots = \mu_8$$

En caso de que se estableciese que hay diferencias significativas entre las medias de los 8 grupos, se utilizará el test de Tukey para saber entre que grupos se encuentran dichas diferencias.

Además, tras la realización del ANOVA, se debe contrastar que se cumplan tres condiciones para validar el modelo:

- Independencia de los residuos.
- Homocedasticidad, o equivalencia de las varianzas.
- Normalidad de los residuos.

Para todas las variables númericas, excepto las rentas, ha sido necesaria una transformación logarítmica para que se cumpliesen las 3 condiciones anteriores. Para las variables a las que se ha aplicado la transformación logarítmica se rechaza la igualdad de medias, puesto que los p-valores obtenidos para el contraste planteado son muy próximos a 0, y por lo tanto menores que 0.05. Después de rechazar la igualdad de medias para todas las variables anteriores, se ha aplicado el test de Tukey para averiguar entre qué grupos se producen:

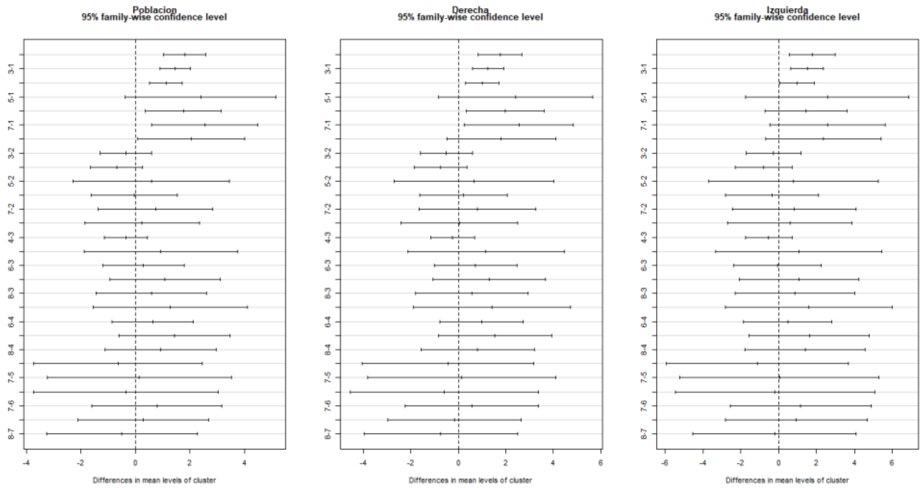


Ilustración 55. Test de Tukey para Población, Derecha e Izquierda

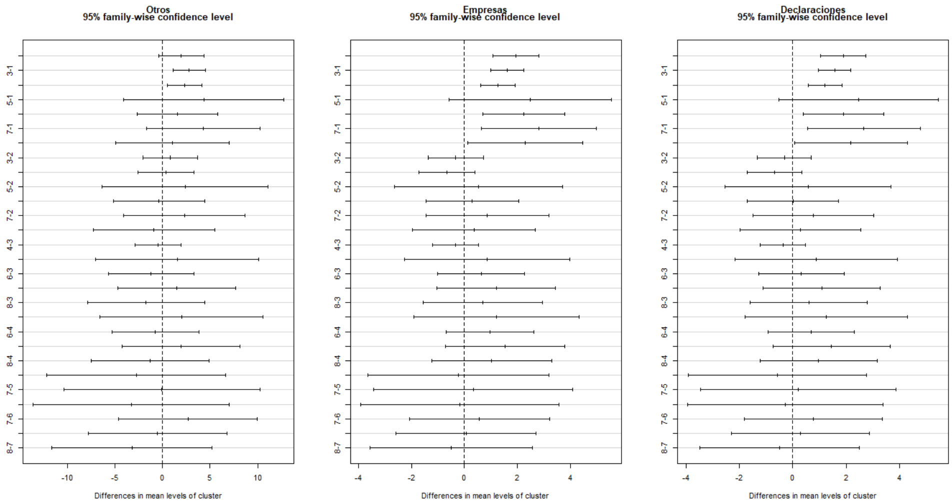


Ilustración 56. Test de Tukey para Otros, Empresas y Declaraciones

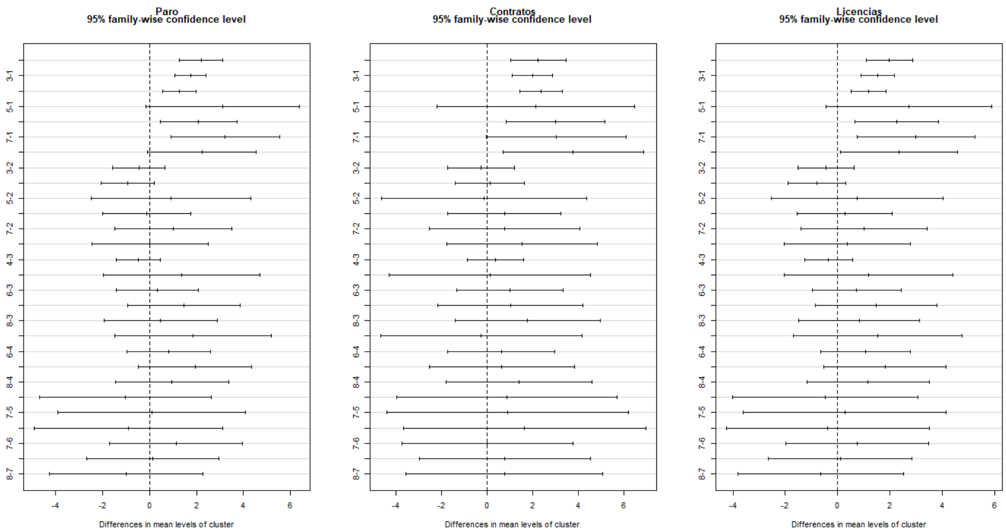


Ilustración 57. Test de Tukey para Paro, Contratos y Licencias

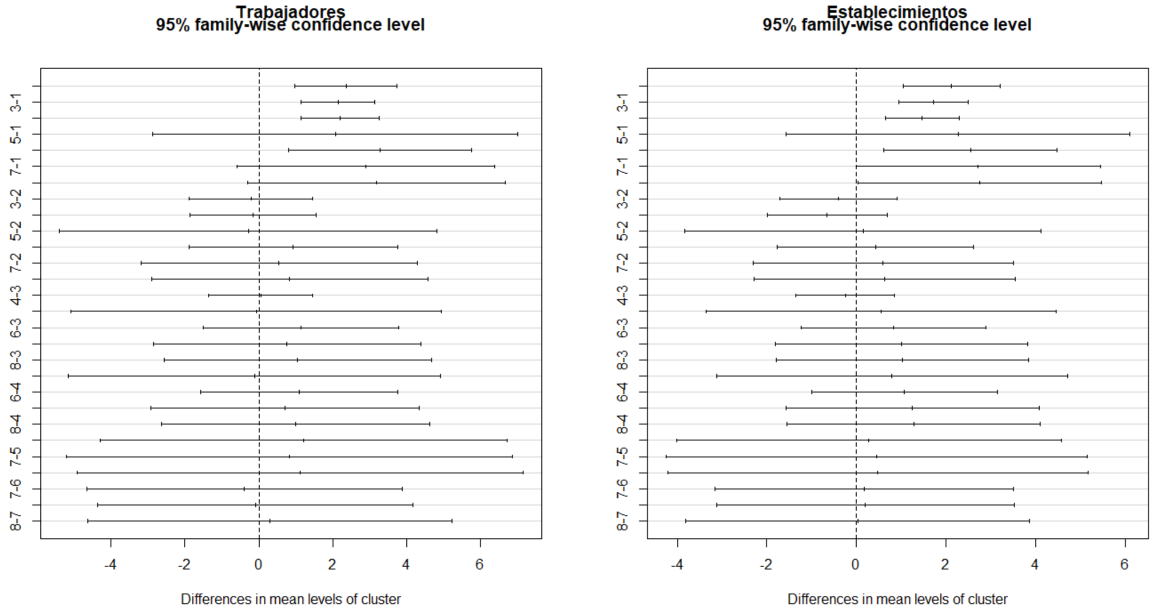


Ilustración 58. Test de Tukey para Trabajadores y Establecimientos

Como se puede observar, la mayor parte de las diferencias entre grupos residen entre el clúster 1, que está representado principalmente por los municipios rurales y los otros 7.

Por último, para las 11 variables se muestran los gráficos para comprobar que, efectivamente, se cumplen las 3 condiciones para que el modelo ANOVA sea válido:

- Independencia y homocedasticidad:

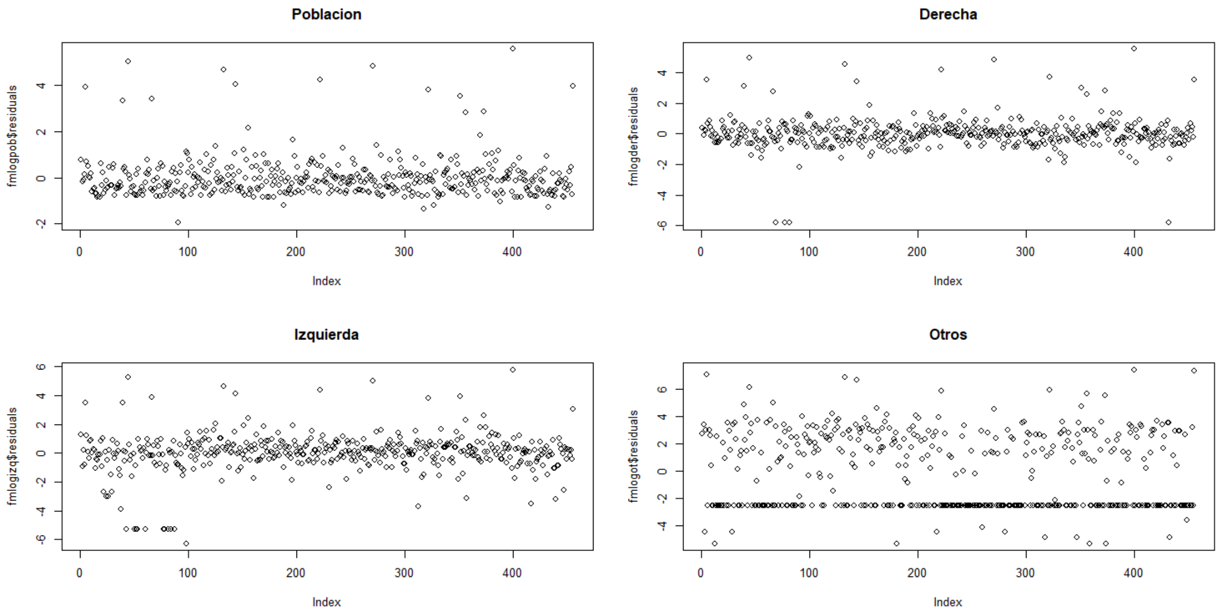


Ilustración 59. Independencia y homocedasticidad para Población, Derecha, Izquierda y Otros

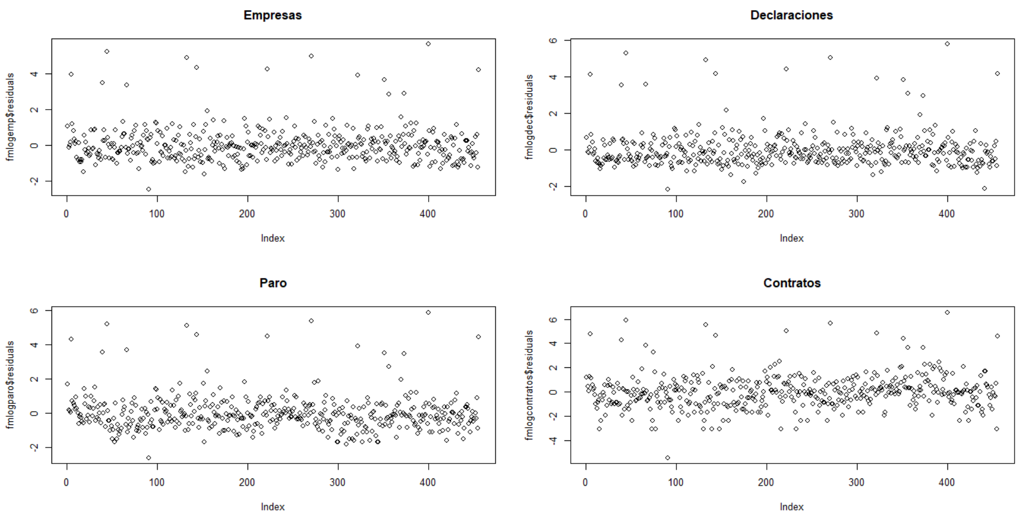


Ilustración 60. Independencia y homocedasticidad para Empresas, Declaraciones, Paro y Contratos

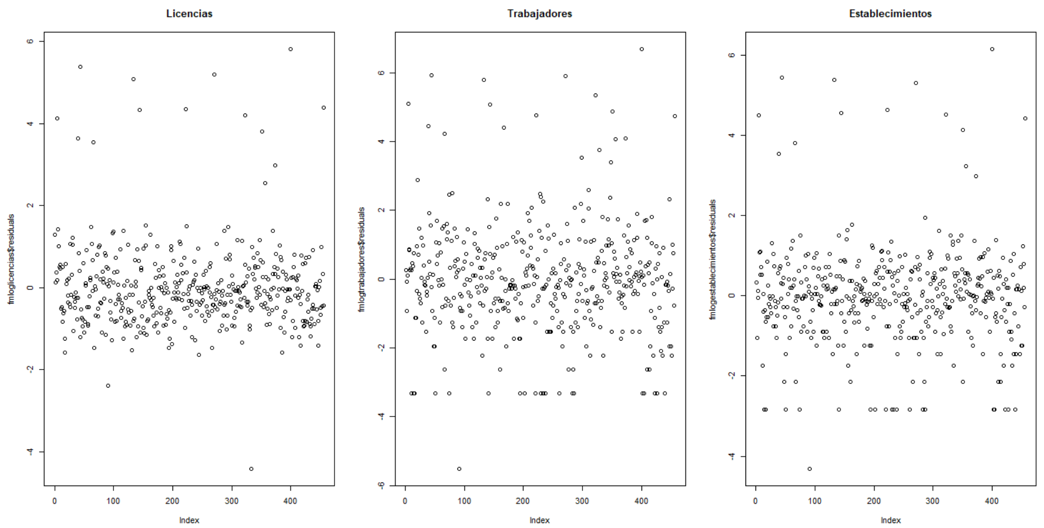


Ilustración 61. Independencia y homocedasticidad para Licencias, Trabajadores y Establecimientos

- Normalidad de los residuos:

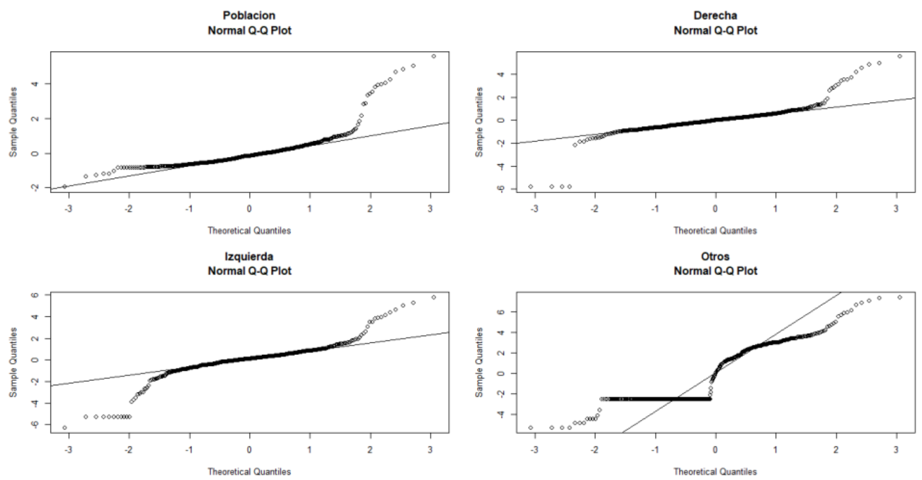


Ilustración 62. Normalidad residuos para Población, Derecha, Izquierda y Otros

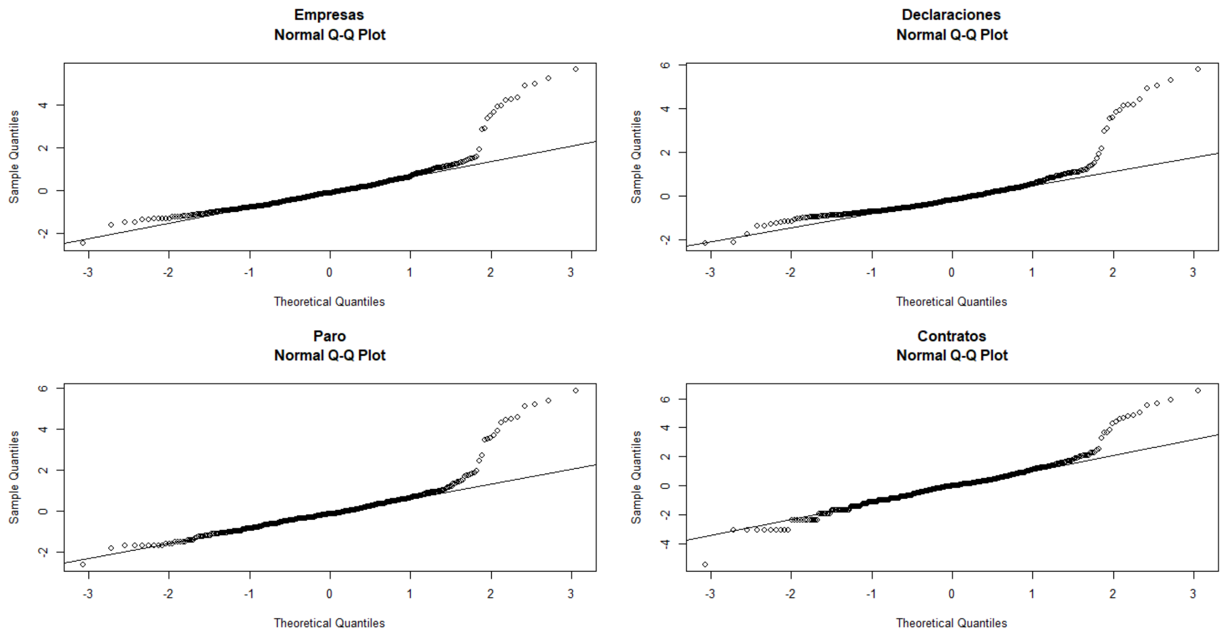


Ilustración 63. Normalidad residuos para Empresas, Declaraciones, Paro y Contratos

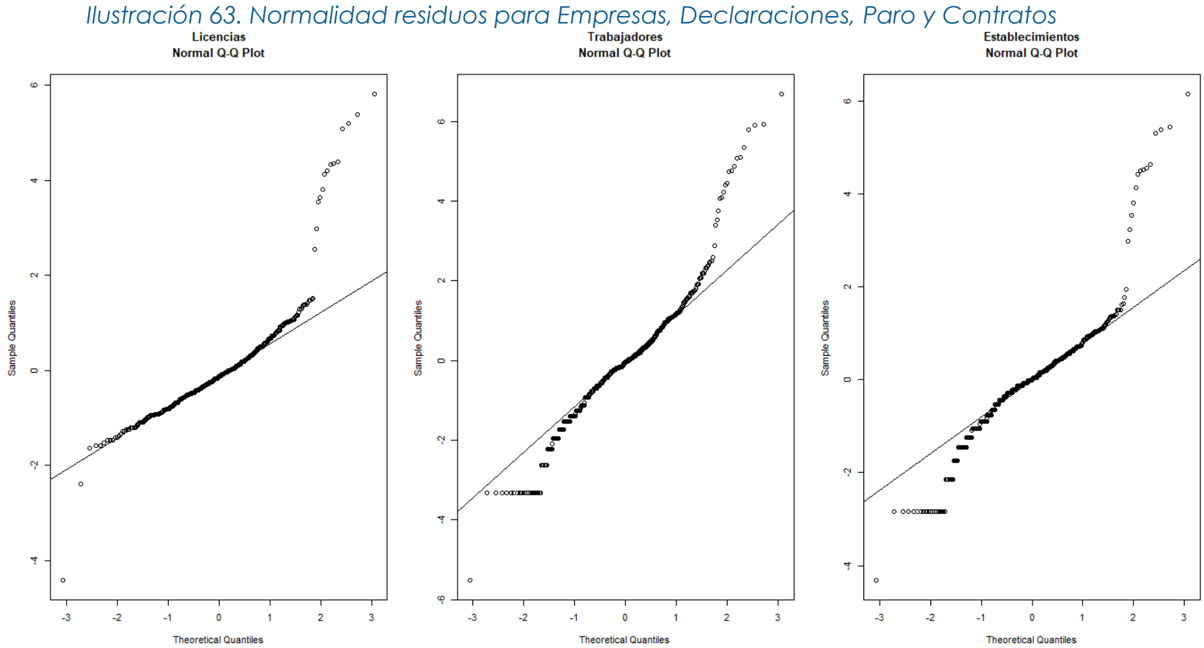


Ilustración 64. Normalidad residuos para Licencias, Trabajadores y Establecimientos

Para terminar con el ANOVA, se muestran los resultados obtenidos para la Renta neta media por persona y la renta neta media por hogar.

En el caso de la primera variable, no se puede rechazar la igualdad de medias entre los 8 grupos, puesto que el p-valor obtenido es 0.655>0.05:

```
             Df    Sum Sq Mean Sq F value Pr(>F)
cluster       7  10129198 1447028    0.72  0.655
Residuals   448 900311061 2009623
```

Ilustración 65. ANOVA renta neta media por persona

Además, se comprueba que el ANOVA es válido, puesto que se cumplen la independencia de los residuos, la homocedasticidad y la normalidad:

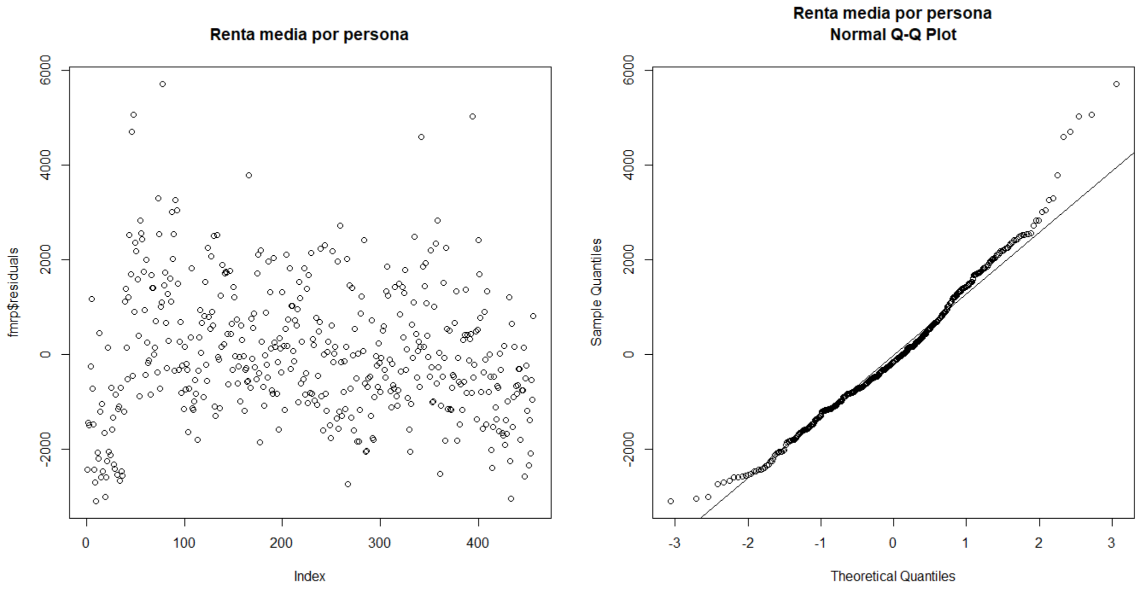


Ilustración 66. Condiciones ANOVA renta neta media por persona

En el caso de la renta neta media por hogar, se rechaza la igualdad de medias entre los 8 clústeres, puesto que el p-valor es 0.0146, menor que 0.05:

```
             Df    Sum Sq  Mean Sq F value Pr(>F)  
cluster       7 3.175e+08 45354213   2.531 0.0146 *
Residuals   448 8.027e+09 17918525                 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Ilustración 67. ANOVA renta neta media por hogar

Aunque no es posible establecer entre que grupos se producen dichas diferencias mediante el test de Tukey, parece que se encuentran las mayores desigualdades entre el grupo 1 con el 3 y el 4:

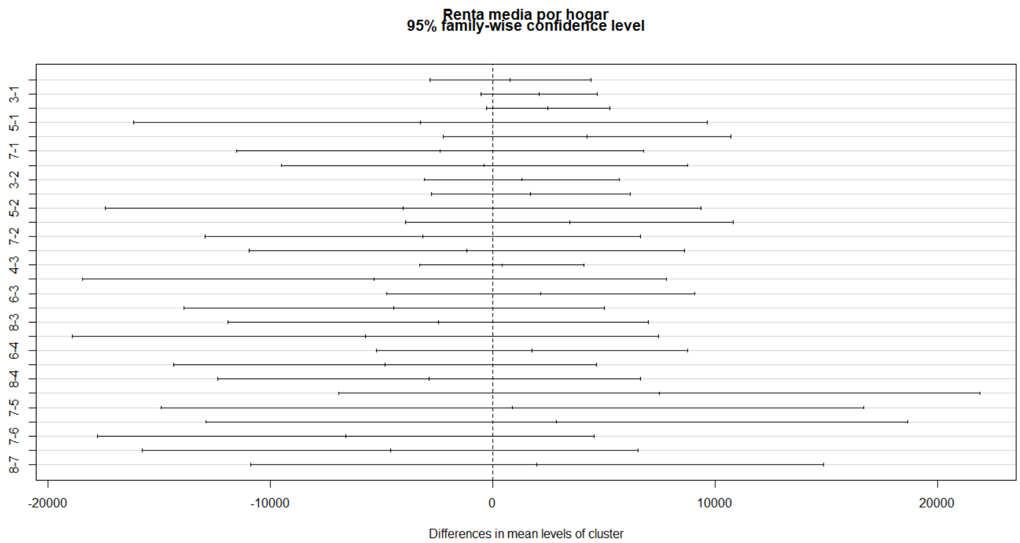


Ilustración 68. Tukey renta neta media por hogar

El ANOVA realizado para esta variable también cumple las condiciones de independencia de los residuos, homocedasticidad y normalidad:

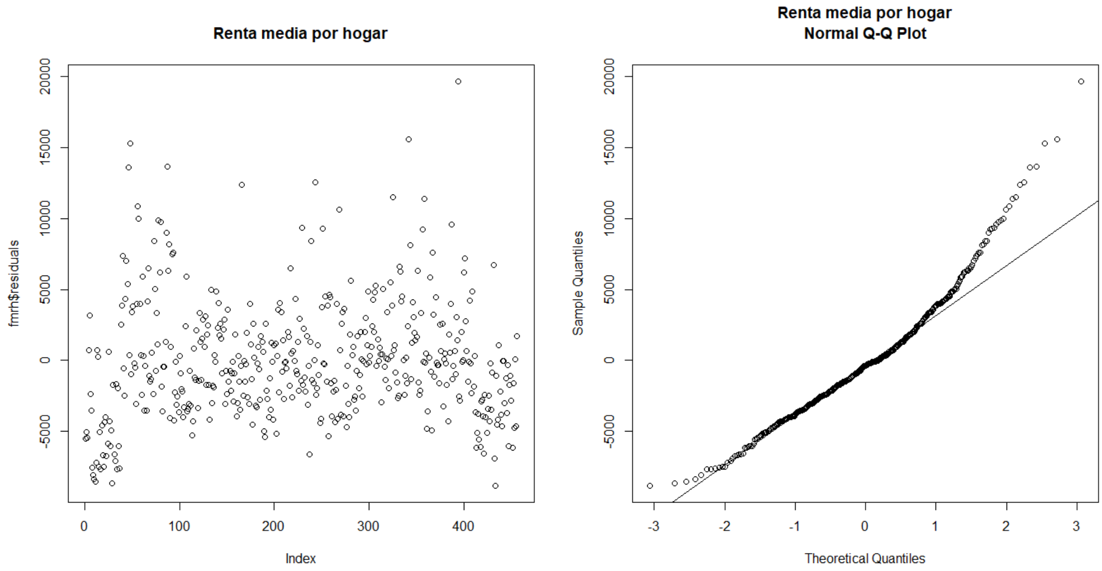


Ilustración 69. Condiciones ANOVA renta neta media por hogar

3.3 RESULTADOS POR TIPO DE MUNICIPIO

Debido a la diferencia de Población entre los municipios no es posible establecer relaciones claras entre las variables económicas y las electorales, por lo tanto, el siguiente paso será repetir los análisis anteriores separando los individuos en los 3 grupos establecidos según su tipo de población:

- Rural si su población es menor de 2000 habitantes.
- Semirural si su población es mayor de 2000 habitantes y hasta 10000.
- Urbano si su población es mayor de 10000 habitantes.

Además, se utilizará la ténica de regresión logística, para establecer qué variables aumentan la probabilidad de que la mayoría de voto se produzca en cada una de las opciones electorales.

Análisis de componentes principales

Mediante el **análisis de componentes principales**, se puede ver cómo se obtienen diferentes resultados si se separan los datos por tipos de municipio, ya que la influencia de la población será menor en grupos de municipios cuyo valor para esta variable sea lo más homogéneo posible:

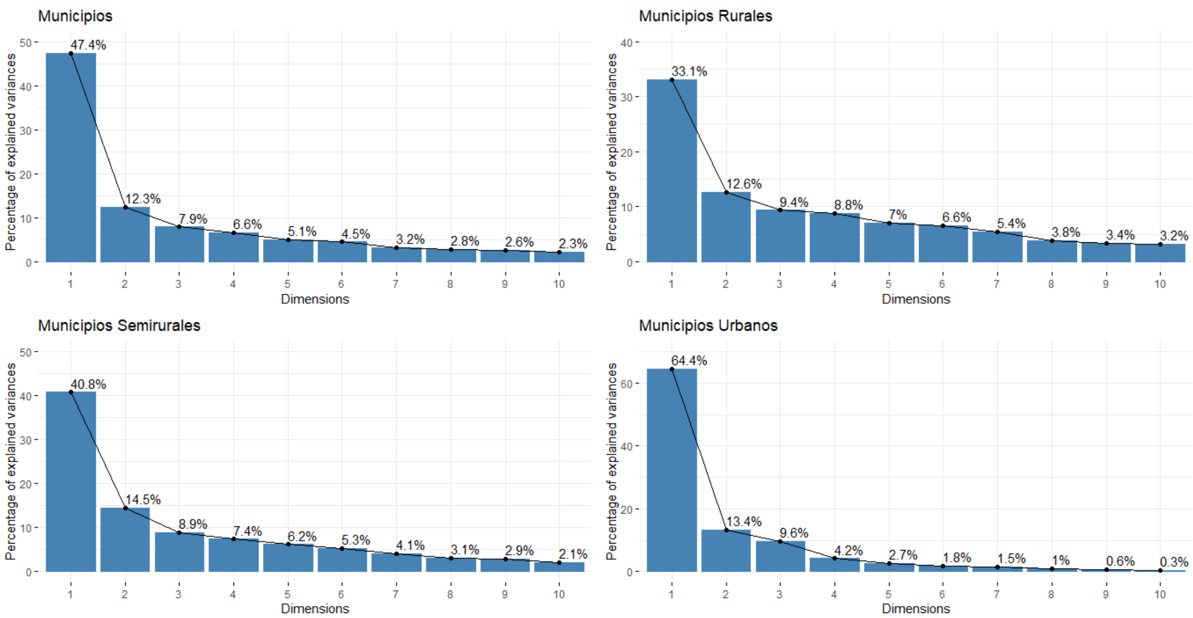


Ilustración 70. Gráfico de autovalores por tipo de municipio

En cualquiera de los cuatro casos representados, lo mejor es extraer dos dimensiones, puesto que, tanto para los municipios en general como para las tres clases, es el número de componentes en el que se encuentra el "codo" del gráfico. Sin embargo, mientras para los municipios sin división se obtiene un 59.7% de varianza explicada, para los rurales se obtiene un 45.7%, para los semirurales un 55.3% y para los urbanos un 77.8%, lo que parece indicar que el último grupo permitirá encontrar relaciones más claras entre las variables económicas y electorales, ya que, aunque es el que presenta menor número de municipios, son los que tienen mayor población.

A continuación, se muestran las ecuaciones de las dos primeras componentes para los 3 tipos de municipio:

$$
\begin{aligned}
PC1_{rurales} = {} & -0.41Población - 0.26Derecha - 0.25Izquierda - 0.03Otros \\
& - 0.05RentaNetaMediaPersona - 0.09RentaNetaMediaHogar \\
& - 0.39TotalEmpresasMunicipio - 0.40NumeroDeclaraciones \\
& - 0.29ParoTotal - 0.23TotalContratos - 0.33TotalLicencias \\
& - 0.21NumeroTrabajadores - 0.28NumeroEstablecimientos
\end{aligned}
$$

$$
\begin{aligned}
PC2_{rurales} = {} & 0.08Población + 0.07Derecha + 0.04Izquierda - 0.11Otros \\
& - 0.68RentaNetaMediaPersona - 0.67RentaNetaMediaHogar \\
& - 0.03TotalEmpresasMunicipio + 0.01NumeroDeclaraciones \\
& + 0.23ParoTotal - 0.02TotalContratos + 0.01TotalLicencias \\
& - 0.07NumeroTrabajadores - 0.05NumeroEstablecimientos
\end{aligned}
$$

Para los municipios **rurales**, los coeficientes de la primera dimensión siguen siendo negativos para todas las variables, aunque el valor del correspondiente a la variable Otros varía y es aún menor que el de las rentas. En la segunda componente se observa que los coeficientes de Otros y de Paro Total crecen de forma considerable, aunque el de la segunda variable con signo opuesto al de las rentas.

$$
\begin{aligned}
PC1_{semirurales} = {} & 0.37Población + 0.33Derecha + 0.19Izquierda + 0.13Otros \\
& + 0.03RentaNetaMediaPersona + 0.08RentaNetaMediaHogar \\
& + 0.36TotalEmpresasMunicipio + 0.37NumeroDeclaraciones \\
& + 0.32ParoTotal + 0.18TotalContratos + 0.34TotalLicencias \\
& + 0.25NumeroTrabajadores + 0.34NumeroEstablecimientos
\end{aligned}
$$

$$
\begin{aligned}
PC2_{semirurales} = {} & -0.02Población - 0.03Derecha - 0.15Izquierda + 0.28Otros \\
& + 0.67RentaNetaMediaPersona + 0.64RentaNetaMediaHogar \\
& - 0.003TotalEmpresasMunicipio + 0.07NumeroDeclaraciones \\
& - 0.14ParoTotal - 0.06TotalContratos - 0.12TotalLicencias \\
& + 0.006NumeroTrabajadores + 0.01NumeroEstablecimientos
\end{aligned}
$$

En cuanto a los municipios **semirurales**, para la primera componente se encuentran signos positivos de los coeficientes para todas las variables, aunque de nuevo, los correspondientes a las rentas resultan insignificantes. Por otro lado, la segunda dimensión presenta coeficientes considerablemente altos y de signo positivo para las rentas y Otros, destacando también los de Izquierda, Paro y Licencias, pero en este caso con signo negativo.

$$
\begin{aligned}
PC1_{urbanos} = {} & -0.33Población - 0.32Derecha - 0.30Izquierda - 0.14Otros \\
& - 0.07RentaNetaMediaPersona - 0.05RentaNetaMediaHogar \\
& - 0.33TotalEmpresasMunicipio - 0.33NumeroDeclaraciones \\
& - 0.19ParoTotal - 0.32TotalContratos - 0.32TotalLicencias \\
& - 0.31NumeroTrabajadores - 0.33NumeroEstablecimientos
\end{aligned}
$$

$$
\begin{aligned}
PC2_{urbanos} = {} & -0.11Población - 0.09Derecha + 0.08Izquierda - 0.17Otros \\
& + 0.64RentaNetaMediaPersona + 0.64RentaNetaMediaHogar \\
& - 0.09TotalEmpresasMunicipio - 0.09NumeroDeclaraciones \\
& - 0.28ParoTotal + 0.09TotalContratos + 0.10TotalLicencias \\
& + 0.11NumeroTrabajadores - 0.008NumeroEstablecimientos
\end{aligned}
$$

Por último, en el caso de la primera componente para los municipios **urbanos**, se observan las mismas ideas que para el conjunto inicial de municipios, lo que quiere decir que las rentas no tienen coeficientes representativos en esta dimensión. Sin embargo, en la segunda componente las rentas presentan coeficientes altos y positivos, destacando también el de Numero Trabajadores con este mismo signo, y, con signo opuesto, Población, Paro Total y Número Trabajadores.

Seguidamente, se analizan las contribuciones de las variables a las dos primeras componentes en los tres tipos de municipio:

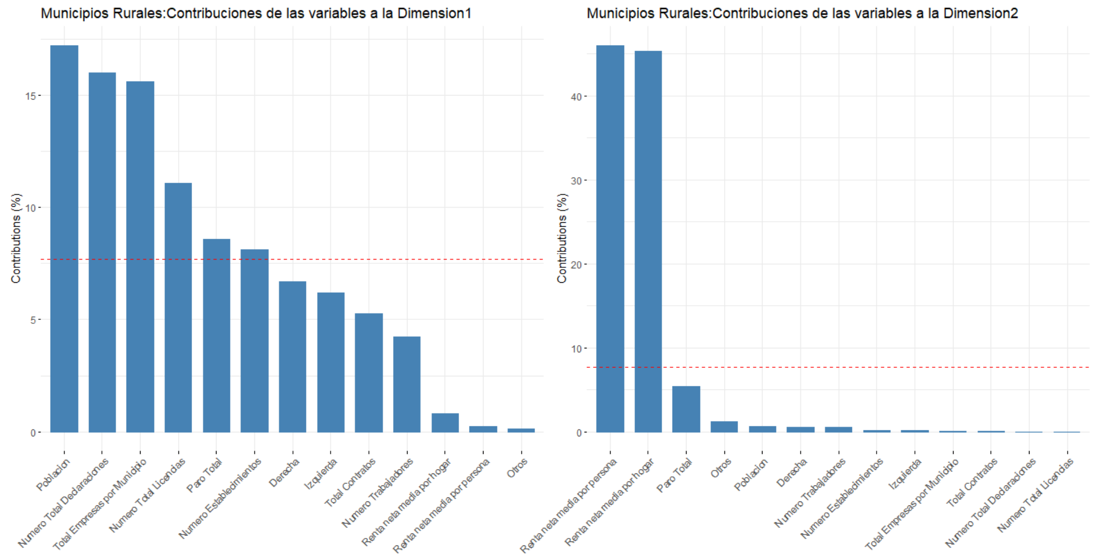


Ilustración 71. Contribuciones a las componentes 1 y 2 para rurales

Para los municipios **rurales**, se aprecia que las variables cuya contribución es significativa son Población, Número Total Declaraciones, Total Empresas por Municipio, Número Total Licencias, Paro Total y Número de Establecimientos, en el caso de la primera dimensión. Mientras que para la segunda vuelven a ser las rentas. Al no encontrarse ninguna de las variables electorales entre las que contribuyen de forma significativa, no parece adecuado utilizar estas dos dimensiones para buscar relaciones entre los dos tipos de variables.

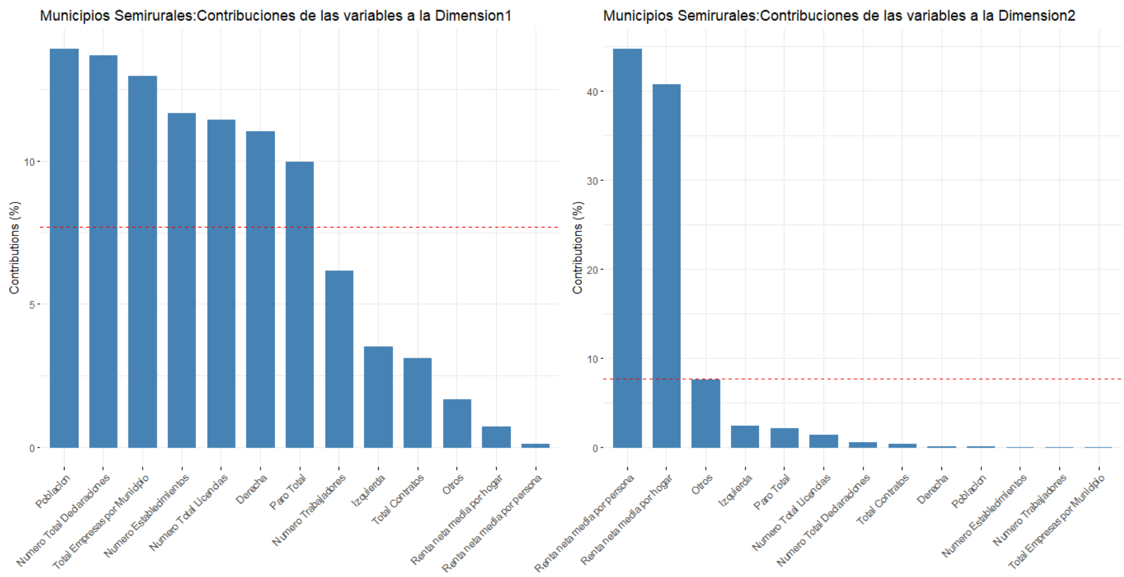


Ilustración 72. Contribuciones a las componentes 1 y 2 para semirurales

En el caso de los municipios **semirurales**, se observa que, mediante la primera componente, podrían encontrarse relaciones entre la variable Derecha y las variables Población, Número Total Declaraciones, Total Empresas por Municipio, Número Establecimientos, Número Total Licencias y Paro, puesto que son las que influyen de forma significativa en la primera dimensión. De la misma forma, podría existir alguna relación entre las rentas y los votos a otros partidos, como puede observarse en el gráfico de la segunda componente.

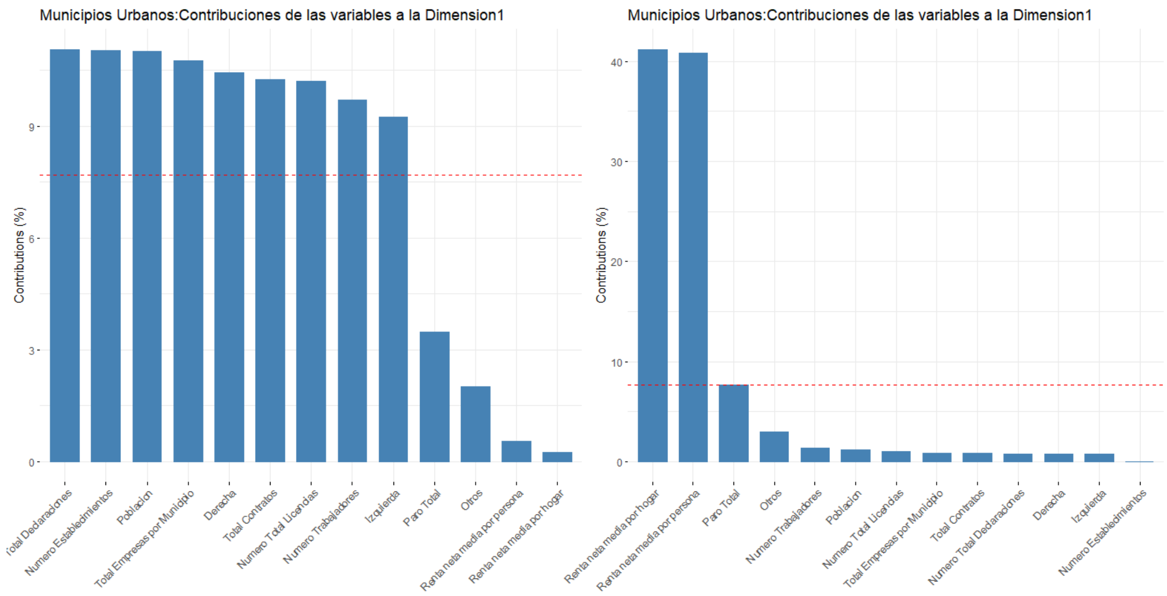


Ilustración 73. Contribuciones a las componentes 1 y 2 para urbanos

Observando las contribuciones de las variables a la primera dimensión para los municipios **urbanos**, vemos que las más significativas son Número Total Declaraciones, Número Establecimientos, Población, Total Empresas por Municipio, Derecha, Total Contratos, Número Total Licencias, Número Trabajadores e Izquierda. Por lo tanto, se intentará utilizar todas estas variables económicas para explicar los votos a Derecha e Izquierda. Para este tipo de municipios no se utilizará la segunda componente puesto que no interviene ninguna de las variables electorales y por lo tanto no sirve para encontrar las relaciones que se buscan.

A continuación, se muestran los 3 gráficos con las coordenadas de las variables en las 2 primeras componentes, cada uno correspondiente a uno de los tipos de municipio:

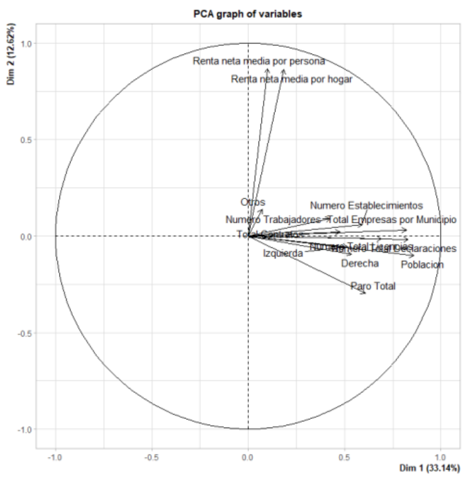


Ilustración 74. Plano por componentes 1-2 rurales

Para los municipios **rurales**, se encuentran algunas relaciones entre variables, considerando que tienen la misma dirección de crecimiento. Por un lado, Otros y las rentas. Por otro, se relaciona el voto a la Izquierda con el Paro Total y el voto a la Derecha con Número Total Licencias.

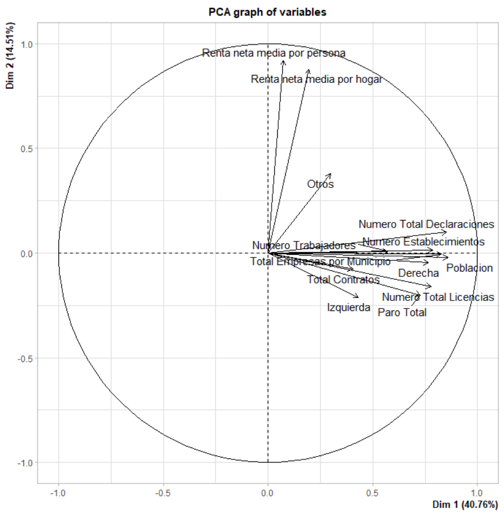


Ilustración 75. Plano por componentes 1-2 semirurales

En el caso de los municipios **semirurales**, teniendo en cuenta la dirección de crecimiento de las variables, es de interés destacar que se observa una relación entre los votos a otros partidos y las rentas, así como Izquierda con Paro Total y Derecha con Número Total Licencias.

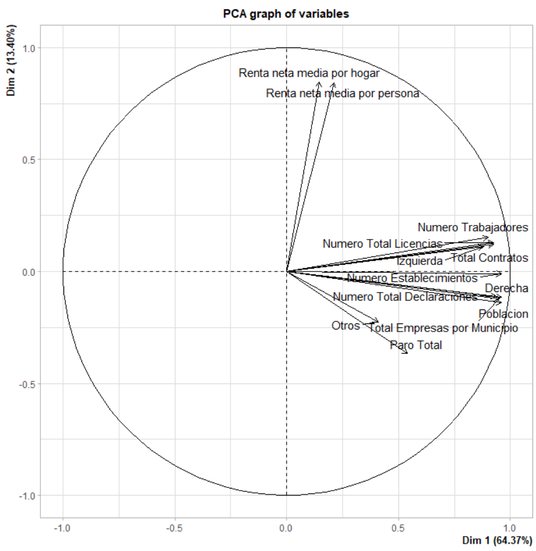


Ilustración 76. Plano por componentes 1-2 urbanos

El último gráfico, correspondiente a los municipios **urbanos**, se encuentra una clara relación entre Otros con Total Empresas por Municipio y Paro Total. Los votos a la izquierda parecen relacionarse con Total Contratos, Número Total Licencias y Número Trabajadores, mientras que los votos a la derecha se relacionan con Número Establecimientos, Número Total Declaraciones y Población. Por lo tanto, las únicas variables que parecen no relacionarse con la elección de voto son los dos tipos de renta.

Análisis de Correlaciones Canónicas

Para el **análisis de correlaciones canónicas** también se obtienen diferentes resultados si se separan los datos disponibles por tipo de municipio, recordando que las variables se encuentran agrupadas en tres electorales y diez económicas:

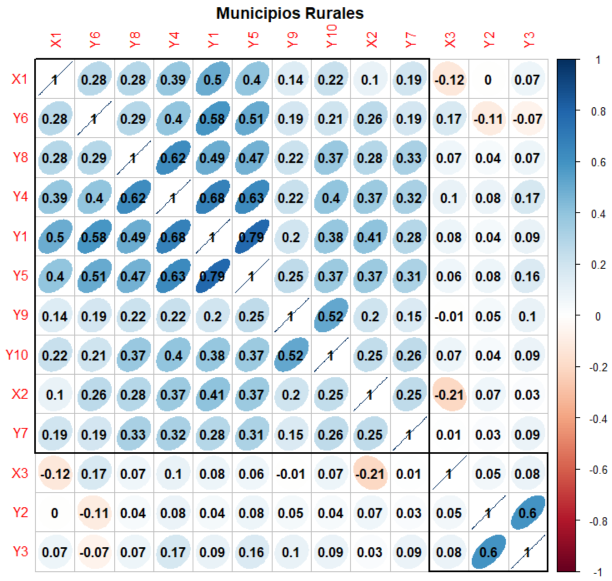


Ilustración 77. Correlaciones entre variables Rurales

Observando las correlaciones entre variables para los municipios **rurales**, se puede ver que la variable que muestra una correlación mayor con el grupo electoral es Población (Y1), tanto para Derecha como para Izquierda. Además, puede verse como los votos a otros partidos presentan correlaciones muy pequeñas con el resto de variables. Por lo tanto, se concluye que Otros apenas tiene importancia dentro de los municipios rurales.

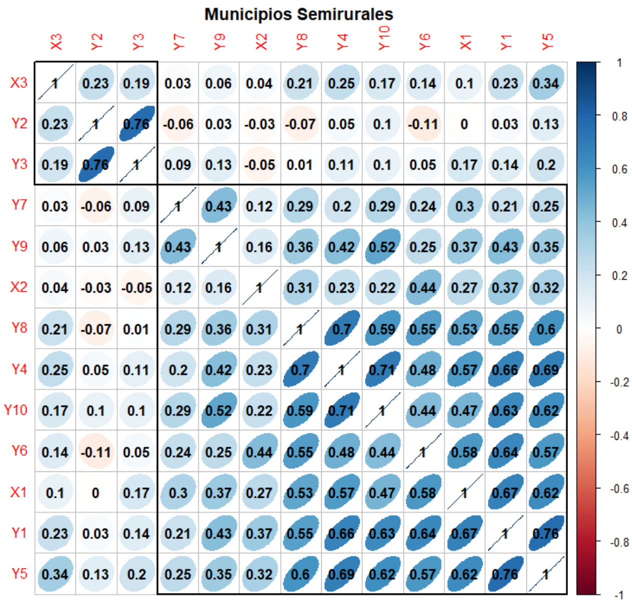


Ilustración 78. Correlaciones entre variables Semirurales

Al observar la misma tabla para los municipios **semirurales**, se ve que Derecha presenta las correlaciones más altas con Población, Total Empresas por Municipio, Número Total

Declaraciones, Paro Total y Número Total Licencias, mientras que la variable Izquierda presenta la mayor correlación con el Paro. También, se observa como las correlaciones son mucho mayores para el grupo Otros, si lo comparamos con las correspondientes al grupo de municipios rurales, lo que quiere decir que esta variable toma mayor importancia en los municipios semirurales, siendo la correlación más destacable la que presenta con Número Total Declaraciones.

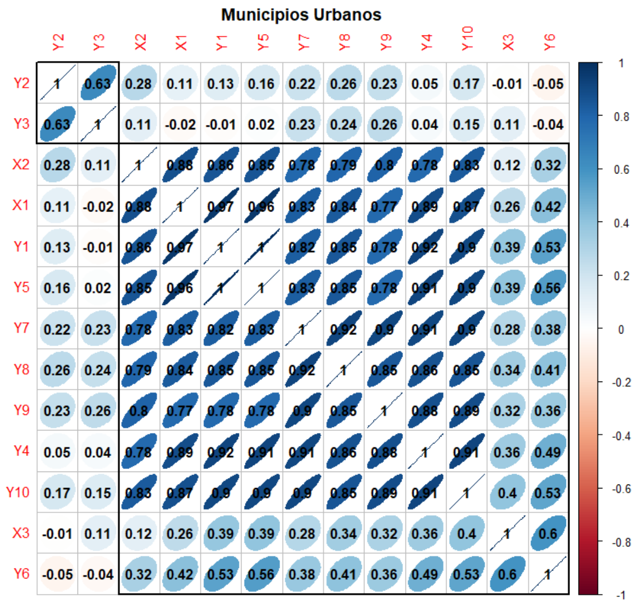


Ilustración 79. Correlaciones entre variables Urbanos

Por último, se muestran las correlaciones entre variables económicas y electorales para el grupo de municipios **urbanos** y se observa como las correlaciones de los tres grupos electorales son considerablemente mayores en este caso, sobre todo para los grupos Derecha e Izquierda, para los que ya se encuentran correlaciones muy significativas. Las variables económicas que presentan mayor correlación con las electorales son Población y Número Total Declaraciones.

Ahora, se compararán los resultados obtenidos sobre las variables canónicas y su significancia para los tres tipos de municipios:

```
    CanR  CanRSQ   Eigen percent    cum                          scree
1 0.6824 0.46569 0.87157  92.824  92.82 ******************************
2 0.2049 0.04196 0.04380   4.665  97.49 **
3 0.1518 0.02304 0.02358   2.511 100.00 *

Test of H0: The canonical correlations in the
current row and all that follow are zero

     CanR LR test stat approx F numDF denDF Pr(> F)
1 0.68241      0.50010   8.2387    30 928.2  <2e-16 ***
2 0.20485      0.93597   1.1849    18 634.0  0.2674
3 0.15178      0.97696   0.9373     8 318.0  0.4857
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Ilustración 80. Variables canónicas y su significancia Rurales

```
    CanR  CanRSQ   Eigen percent    cum                          scree
1 0.7954 0.63259 1.72179  89.089  89.09 ******************************
2 0.3269 0.10685 0.11963   6.190  95.28 **
3 0.2892 0.08362 0.09125   4.721 100.00 **

Test of H0: The canonical correlations in the
current row and all that follow are zero

     CanR LR test stat approx F numDF  denDF   Pr(> F)
1 0.79536      0.30071   4.5155    30 267.78 6.664e-12 ***
2 0.32688      0.81847   1.0769    18 184.00    0.3786
3 0.28917      0.91638   1.0608     8  93.00    0.3974
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Ilustración 81. Variables canónicas y su significancia Semirurales

```
    CanR CanRSQ   Eigen percent    cum                          scree
1 0.9867 0.9736 36.8624  96.657  96.66 ******************************
2 0.6768 0.4581  0.8452   2.216  98.87 *
3 0.5482 0.3006  0.4297   1.127 100.00

Test of H0: The canonical correlations in the
current row and all that follow are zero

     CanR LR test stat approx F numDF  denDF   Pr(> F)
1 0.98671      0.01001   3.8033    30 30.028 0.0002262 ***
2 0.67680      0.37906   0.7629    18 22.000 0.7176464
3 0.54823      0.69945   0.6446     8 12.000 0.7283922
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Ilustración 82. Variables canónicas y su significancia Urbanos

Como se puede observar, en los 3 casos la primera correlación canónica es mayor que las otras dos, siendo 0.68 para los municipios rurales, 0.80 para los semirurales y 0.99 para los urbanos. Esto parece indicar que, como ya se había observado, el conjunto de los municipios rurales parece ser el menos útil para establecer relaciones entre los 2 grupos de variables y el más útil el de los municipios urbanos, puesto que es el que contiene a la mayor parte de la población de Castilla y León.

Además, para contrastar la hipótesis nula:

$$H0\text{: } Las\ correlaciones\ canónicas\ en\ la\ fila\ actual\ y\ las\ siguientes\ son\ 0$$

los p-valores obtenidos son próximos a 0 en los 3 tipos de municipio, para la primera correlación canónica y mucho mayores que 0.05 para la segunda y la tercera, por lo tanto, se rechazará en el primer caso y en los otros dos no se podrá, lo que supone que únicamente se considere la primera.

El siguiente paso será comparar los coeficientes estandarizados de las variables canónicas para cada uno de los casos:

$$Xcan1_{rurales} = -0.71Derecha - 0.66Izquierda - 0.37Otros$$

$$Xcan1_{semirurales} = -0.82Derecha - 0.28Izquierda - 0.27Otros$$

$$Xcan1_{urbanos} = -0.75Derecha - 0.23Izquierda - 0.16Otros$$

La primera variable canónica para el grupo de variables electorales presenta en los tres casos una diferencia ponderada con mayor peso en Derecha. Sin embargo, al comparar el caso correspondiente al conjunto de municipios inicial con el resto, se observa como para los municipios **rurales** los coeficientes para Derecha e Izquierda se igualan y el correspondiente a Otros crece, lo que quiere decir que la aportación de los datos electorales para este grupo de municipios se equipara. Por otro lado, al comparar con los **semirurales** se observa como la diferencia entre Derecha e Izquierda se incrementa, mientras que Izquierda y Otros quedan igualados. Por último, en los municipios **urbanos** se aprecia una gran diferencia entre el coeficiente correspondiente a Derecha e Izquierda y a su vez una diferencia también bastante marcada entre Izquierda y Otros. Por lo tanto, se supone que la variable que más aporta en los cuatro casos es Derecha.

$$\begin{aligned} Ycan1_{rurales} = &-0.64 Población - 0.04 Renta\ neta\ media\ por\ persona \\ &- 0.008 Renta\ neta\ media\ por\ hogar \\ &- 0.23 Total\ Empresas\ por\ Municipio \\ &+ 0.006 Número\ Total\ Declaraciones - 0.12 Paro\ Total \\ &- 0.13 Total\ Contratos - 0.03 Número\ Total\ Licencias \\ &- 0.08 Número\ Trabajadores - 0.06 Número\ Establecimientos \end{aligned}$$

Para los municipios **rurales**, se encuentra un único coeficiente positivo, aunque resultará insignificante puesto que es prácticamente nulo. A diferencia del primer caso analizado, la variable que presenta mayor coeficiente es Población, presentando una gran diferencia con la siguiente variable que aporta mayor peso, que es Total Empresas por Municipio.

$$\begin{aligned} &Ycan1_{semirurales} \\ &= -0.38 Población - 0.01 Renta\ neta\ media\ por\ persona \\ &- 0.08 Renta\ neta\ media\ por\ hogar \\ &- 0.13 Total\ Empresas\ por\ Municipio \\ &- 0.29 Número\ Total\ Declaraciones - 0.28 Paro\ Total \\ &- 0.10 Total\ Contratos - 0.14 Número\ Total\ Licencias \\ &- 0.04 Número\ Trabajadores + 0.16 Número\ Establecimientos \end{aligned}$$

En cuanto a los **semirurales**, de nuevo se encuentra un único coeficiente positivo, correspondiente al Número de Establecimientos, pero en este caso presenta una importancia considerable. Por otra parte, se observa que la variable con mayor peso vuelve a ser Población, pero en este caso con menor diferencia respecto a las siguientes que son Número Total Declaraciones y Paro Total.

$$\begin{aligned} Ycan1_{urbanos} = &-1.21 Población + 0.08 Renta\ neta\ media\ por\ persona \\ &- 0.03 Renta\ neta\ media\ por\ hogar \\ &+ 0.36 Total\ Empresas\ por\ Municipio \\ &+ 0.08 Número\ Total\ Declaraciones + 0.03 Paro\ Total \\ &- 0.11 Total\ Contratos + 0.004 Número\ Total\ Licencias \\ &- 0.25 Número\ Trabajadores + 0.07 Número\ Establecimientos \end{aligned}$$

Para terminar, en el grupo de municipios **urbanos** se observan varios coeficientes positivos, aunque el único de importancia es Total Empresas por Municipio. Además, se encuentra la mayor diferencia entre la variable Población y el resto.

A continuación, se muestran las correlaciones entre los 2 conjuntos iniciales de variables y sus variables canónicas, recordando que las variables supresoras son aquellas que presentan signos opuestos en su coeficiente y en la correlación correspondiente.

	Xcan1R	Xcan1S	Xcan1U
X1	-0.72	-0.92	-0.99
X2	-0.65	-0.52	-0.90
X3	-0.15	-0.36	-0.38

Tabla 26. Correlaciones v.electorales y v.canónicas del mismo grupo

	Xcan1R	Xcan1S	Xcan1U
Y1	-0.65	-0.72	-0.98
Y2	-0.06	-0.05	-0.14
Y3	-0.10	-0.18	-0.03
Y4	-0.55	-0.60	-0.89
Y5	-0.55	-0.70	-0.97
Y6	-0.43	-0.64	-0.48
Y7	-0.30	-0.29	-0.84
Y8	-0.41	-0.58	-0.86
Y9	-0.23	-0.36	-0.81
Y10	-0.35	-0.49	-0.90

Tabla 27. Correlaciones v.económicas y v.canónicas del grupo opuesto

En el caso de las variables electorales, no se encuentra ninguna variable supresora, puesto que las correlaciones entre estas 3 variables y la primera variable canónica correspondiente a su grupo son negativas en los cuatro casos, igual que ocurría con sus coeficientes.

	Ycan1R	Ycan1S	Ycan1U
X1	-0.49	-0.73	-0.97
X2	-0.45	-0.41	-0.89
X3	-0.10	-0.29	-0.37

Tabla 28. Correlaciones v.electorales y v.canónicas del grupo opuesto

	Ycan1R	Ycan1S	Ycan1U
Y1	-0.95	-0.90	-0.99
Y2	-0.09	-0.07	-0.14
Y3	-0.15	-0.23	-0.03
Y4	-0.81	-0.75	-0.91
Y5	-0.80	-0.88	-0.99
Y6	-0.64	-0.80	-0.49
Y7	-0.45	-0.37	-0.85
Y8	-0.60	-0.72	-0.87
Y9	-0.34	-0.45	-0.82
Y10	-0.51	-0.62	-0.91

Tabla 29. Correlaciones v.económicas y v.canónicas del mismo grupo

Sin embargo, para las variables económicas se encuentran variables supresoras en los tres casos. Para el conjunto de datos inicial se encontraba como variable supresora Número Trabajadores, mientras que, para los municipios **rurales** la variable supresora es Número Total Declaraciones, para los **semirurales** Número Establecimientos y para los **urbanos** Renta neta media por persona, Total Empresas por Municipio, Número Total Declaraciones, Paro Total, Número Total Licencias y Número Establecimientos.

Para continuar, se muestra un análisis de redundancia canónico para los diferentes tipos de municipio según su población:

```
Redundancies for the Electorales variables & total X canonical redundancy

    Xcan1     Xcan2     Xcan3 total X|Y
 0.151220  0.015391  0.007107  0.173718

Redundancies for the Economicas variables & total Y canonical redundancy

    Ycan1     Ycan2     Ycan3 total Y|X
 0.166108  0.002181  0.001388  0.169676
```

Ilustración 83. Análisis Redundancia Canónico Rurales

Observando los resultados para el conjunto de municipios **rurales**, se ve que la capacidad predictiva de las dos primeras variables canónicas para el conjunto opuesto de variables es prácticamente inútil, puesto que se obtienen proporciones de varianza explicada de 0.15 y 0.17.

```
Redundancies for the Electorales variables & total X canonical redundancy

   Xcan1    Xcan2    Xcan3 total X|Y
 0.26242  0.03468  0.02179   0.31889

Redundancies for the Economicas variables & total Y canonical redundancy

   Ycan1    Ycan2    Ycan3 total Y|X
 0.25897  0.01177  0.00334   0.27408
```

Ilustración 84. Análisis Redundancia Canónico Semirurales

Lo mismo ocurre para los municipios **semirurales**, puesto que, aunque los resultados son algo mejores, las proporciones de varianza explicadas por las dos primeras variables canónicas son aproximadamente 0.26. Ocurre lo mismo observando los coeficientes de redundancia totales, que en este caso son 0.32 y 0.27.

```
Redundancies for the Electorales variables & total X canonical redundancy

   Xcan1    Xcan2    Xcan3 total X|Y
 0.62452  0.04712  0.07684   0.74848

Redundancies for the Economicas variables & total Y canonical redundancy

   Ycan1    Ycan2    Ycan3 total Y|X
 0.58691  0.02955  0.02413   0.64058
```

Ilustración 85. Análisis Redundancia Canónico Urbanos

Sin embargo, en el último grupo, correspondiente a los municipios **urbanos**, se observa que las dos primeras variables canónicas resultan ser predictores mejores para el conjunto opuesto de variables, puesto que las proporciones de varianza explicadas por cada una de ellas son 0.63 y 0.59. Los coeficientes de redundancia totales indican buena capacidad de las variables electorales para predecir la económicas (0.75) y buena capacidad de las variables económicas para predecir las electorales (0.64).

Por último, se muestran las correlaciones múltiples cuadradas entre los dos tipos de variables y las canónicas del grupo opuesto:

	Ycan1R	**Ycan1S**	**Ycan1U**
X1	0.2444	0.5385	0.9465
X2	0.1992	0.1675	0.7890
X3	0.0100	0.0812	0.1381

Tabla 30. Correlaciones múltiples cuadradas v.electorales y v.canónicas opuestas

En el caso del conjunto global de municipios, la primera variable canónica de las variables económicas presentaba cierta capacidad predictiva para Derecha (0.55), aunque más pobre para Izquierda (0.36) y aún menor para Otros (0.17). Para los municipios **rurales** se observa una disminución en dicha capacidad predictiva puesto que, en este caso, las correlaciones múltiples cuadradas son 0.24 para Derecha, 0.20 para Izquierda y 0.01 para Otros. En el grupo de municipios **semirurales** ocurre algo similar, aunque con una mejora considerable en la capacidad predictiva para Derecha. Las ideas más destacables se encuentran en el grupo de municipios **urbanos**, puesto que la primera variable canónica correspondiente al grupo económico presenta muy buenas capacidades predictivas para Derecha (0.95) e Izquierda (0.79).

	Xcan1R	Xcan1S	Xcan1U
Y1	0.4198	0.5135	0.9570
Y2	0.0038	0.0028	0.0197
Y3	0.0104	0.0324	0.0007
Y4	0.3066	0.3594	0.8008
Y5	0.2989	0.4848	0.9482
Y6	0.1876	0.4050	0.2316
Y7	0.0922	0.0854	0.7047
Y8	0.1685	0.3315	0.7368
Y9	0.0540	0.1307	0.6573
Y10	0.1193	0.2443	0.8123

Tabla 31. Correlaciones múltiples cuadradas v.económicas y v.canónicas opuestas

Por otro lado, para el conjunto inicial de municipios, la primera variable canónica de las variables electorales resultaba un buen predictor para Población, Total Empresas por Municipio, Número Total Declaraciones y Paro Total, puesto que los valores correspondientes de la correlación múltiple al cuadrado eran 0.56, 0.53, 0.55, 0.53. Para los municipios **rurales** la única variable para la que la primera variable canónica correspondiente al grupo electoral presenta una capacidad predictiva medianamente alta es para Población, puesto que el valor de la correlación múltiple al cuadrado es 0.42, y resulta mucho mayor que los valores para el resto de variables. Algo parecido ocurre para los municipios **semirurales**, ya que la primera variable canónica en este caso solo obtiene capacidades predictivas destacables para Población y para Número Total Declaraciones. Sin embargo, como ocurría en el caso de la primera variable canónica económica, en el caso electoral, los mejores resultados en cuanto a la capacidad de predecir las variables económicas se obtienen en el grupo de municipios **urbanos**, puesto que en la mayoría de las variables se presenta dicho valor por encima de 0.65, a excepción de las rentas y el Paro Total.

Regresión Logística

Debido a que, con los dos análisis anteriores, tampoco se han obtenido relaciones claras entre los dos tipos de variables, se planteará para cada uno de los tres tipos de municipio, una regresión logística con las variables electorales como respuesta y como explicativas las variables económicas. Se utilizarán para analizar las posibles diferencias entre modelos, en base a que variables aumentan la probabilidad de que la mayoría de voto se encuentre en cada tipo de opción electoral.

La principal utilidad de la regresión logística, es que permite estimar la probabilidad de una variable categórica en función de una o varias variables numéricas, siempre que la variable categórica sea binaria. Todos los modelos que se plantearán en este punto, corresponderán a regresión logística múltiple, puesto que presentarán más de un predictor, y presentarán la siguiente forma, suponiendo *Y* como variable respuesta y *X1, X2, …, Xn* como predictores:

$$\ln\left(\frac{P(Y=1|X)}{(1-P(Y=1|X))}\right)=\beta_0+\beta_1X1+\beta_2X2+\cdots+\beta_nXn$$

Por lo tanto, el valor de la probabilidad de que la variable respuesta sea 1 dados los valores de las covariables, se obtendrá:

$$\mathrm{P(Y=1|X)}=\frac{e^{\beta_0+\beta_1X1+\beta_2X2+\cdots+\beta_nXn}}{1+e^{\beta_0+\beta_1X1+\beta_2X2+\cdots+\beta_nXn}}$$

Una vez obtenidas las estimaciones de cada coeficiente del modelo, podrá conocerse la probabilidad de que un individuo pertenezca a un grupo u otro de la variable respuesta, dado un valor concreto de cada variable predictora, de la siguiente forma:

$$\hat{P}(\mathrm{Y=1|X})=\frac{e^{\widehat{\beta 0}+\widehat{\beta 1}X1+\widehat{\beta 2}X2+\cdots+\widehat{\beta n}Xn}}{1+e^{\widehat{\beta 0}+\widehat{\beta 1}X1+\widehat{\beta 2}X2+\cdots+\widehat{\beta n}Xn}}$$

Por lo tanto, cada coeficiente indica el cambio en el logaritmo debido al incremento en una unidad del valor de la variable explicativa asociada.

Para evaluar la calidad de cada modelo, se utilizará la matriz de confusión, que tiene como función principal mostrar el número de individuos cuya asignación ha sido errónea. Se trata de una tabla comparativa entre observaciones y predicciones:

Predicciones	0	1
Observaciones		
0	VN	FP
1	FN	VP

Tabla 32. Matriz de confusión

Los valores que se encuentran en la diagonal principal, corresponden con los valores estimados correctamente por el modelo, puesto que son los “Verdaderos Negativos” y los “Verdaderos Positivos”. Por el contrario, la diagonal opuesta muestra los individuos que se han asignado de forma errónea.

Existen 3 medidas que pueden extraerse a partir de la matriz. En primer lugar, la exactitud, que muestra la proporción de predicciones correctas frente al total. Por otro lado, la sensibilidad, que muestra la tasa de verdaderos positivos. Si un modelo es “poco sensible”, podría decirse que clasifica muchos individuos como negativos, cuando realmente son positivos. Por último, la especificidad, que representa la tasa de verdaderos negativos. Cómo el conjunto de datos disponible es poco equilibrado, la excatitud no resultará útil para evaluar la calidad de los modelos, por lo tanto, se utilizarán únicamente la sensibilidad y la especificidad, calculadas de la siguiente forma:

$$sensibilidad=\frac{VP}{VP+FN}$$

$$especificidad=\frac{VN}{VN+FP}$$

En primer lugar, se plantean 3 nuevas variables que se utilizarán para representar las variables electorales como respuestas a los los distintos modelos:

- $MayoríaDerecha \begin{Bmatrix} 1 \; si \; D > I \; y \; O \\ 0 \; en \; otro \; caso \end{Bmatrix}$
- $MayoríaIzquierda \begin{Bmatrix} 1 \; si \; I > D \; y \; O \\ 0 \; en \; otro \; caso \end{Bmatrix}$
- $MayoríaOtros \begin{Bmatrix} 1 \; si \; O > I \; y \; D \\ 0 \; en \; otro \; caso \end{Bmatrix}$

Antes de aplicar la regresión logística, resulta interesante observar la distribución de las 3 variables electorales, para ver cómo se reparten los votos en cada tipo de municipio:

	MayoríaDerecha	MayoríaIzquierda	MayoríaOtros
General	274	128	51
	0.60	0.28	0.12
Rurales	197	90	39
	0.60	0.28	0.12
Semirurales	63	33	8
	0.60	0.32	0.08
Urbanos	14	5	4
	0.61	0.22	0.17

Tabla 33. Distribución de mayorías electorales por tipo de municipio

Como se puede observar en la tabla superior, en cualquiera de los 3 tipos, se da la mayoría de voto a partidos de derechas en aproximadamente el 60% de los municipios. Además, puesto que los municipios **rurales** son el grupo mayoritario (329 de los 456), la distribución de la mayoría de voto es la misma que en el conjunto inicial. Las diferencias se producen en los municipios **semirurales** y en los **urbanos**. En el primer grupo, disminuye la proporción de municipios que presentan su mayoría en Otros, y aumentan los que la presentan en Izquierda, con porcentajes del 8% y 32%, respectivamente. En los municipios **urbanos**, ocurre lo contrario, aumenta la proporción de municipios con mayoría de voto a Otros partidos y disminuye para la Izquierda, obteniéndose el 17% y 22%.

Puesto que los municipios en los que la mayoría de voto se da en el grupo Otros, son un porcentaje muy pequeño en los 3 tipos de municipio, se eliminarán aquellos que cumplan dicha condición, de tal forma que se cumplirá la siguiente relación para las dos variables restantes:

$$MayoríaDerecha = 1 - MayoríaIzquierda$$

Y se realizarán los modelos logísticos con los municipios que queden, de esta forma podrán encontrarse conclusiones interesantes que indiquen con qué variables aumenta $P(MayoríaDerecha = 1)$ y, por lo tanto disminuye $P(MayoríaIzquierda = 1)$.

Por otra parte, se seleccionará el mejor modelo en base al criterio del menor AIC, puesto que se debe elegir un modelo final para cada tipo de municipio. Esta tarea se realizará mediante el método Backward de selección de variables, que comienza por incluir todas las variables posibles en el modelo y las va eliminando una a una, expluyendo en cada paso la variable menos influyente.

En cuanto a las posibles variables a incluir en los modelos, para el número de empresas, declaraciones, paro, contratos, licencias, trabajadores y establecimientos, se considerar valores relativos por cada 1000 habitantes. De esta forma no será necesario considerar la población.

Para los municipios rurales, no se ha encontrado ningún modelo que obtenga resultados de interés. Esto se debe a que en este tipo de municipios los resultados electorales dependen más de las personas que se presentan que de las ideologías y este hecho podría enmascarar las relaciones entre variables. Por lo tanto, únicamente se tendrán en cuenta los municipios semirurales y los urbanos, es decir, aquellos cuya población sea superior a los 2000 habitantes.

En primer lugar, se muestran los resultados obtenidos para los municipios **semirurales.** El primer paso, ha sido generar el modelo completo, para aplicar posteriormente el algoritmo Backward y seleccionar de esta forma las variables más últiles:

```
Coefficients:
                                    Estimate Std. Error z value Pr(>|z|)
(Intercept)                       -3.483e-01  2.859e+00  -0.122   0.9030
dats$`Renta neta media por persona` -3.325e-04  3.896e-04  -0.854   0.3934
dats$`Renta neta media por hogar`   1.312e-04  1.289e-04   1.018   0.3088
dats$emp100                        1.213e+01  7.512e+00   1.615   0.1064
dats$decl1000                      3.330e-01  8.975e-01   0.371   0.7106
dats$paro1000                     -1.431e+00  5.764e+00  -0.248   0.8039
dats$contratos1000                 1.059e+01  4.931e+00   2.147   0.0318 *
dats$licencias1000                -4.915e+00  5.241e+00  -0.938   0.3484
dats$trab1000                     -2.491e+00  3.948e+00  -0.631   0.5280
dats$estab1000                    -1.301e+01  2.054e+01  -0.633   0.5264
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 123.55  on 95  degrees of freedom
Residual deviance: 111.32  on 86  degrees of freedom
AIC: 131.32
```

Ilustración 86. Regresión Logística Semirural 1

Como se puede observar, el AIC es 131.32, y la única variable que resulta significativa a la hora de ser incluida en el modelo es el número de contratos por cada 1000 habitantes. Por lo tanto, tras aplicar el algoritmo Backward se obtienen los siguientes resultados:

```
Call:  glm(formula = dats$mayoriaDerechaSemi ~ dats$contratos1000, family = binomial, 
    data = dats)

Coefficients:
       (Intercept)  dats$contratos1000  
          -0.04618             9.84254  

Degrees of Freedom: 95 Total (i.e. Null);  94 Residual
Null Deviance:      123.6 
Residual Deviance: 117  AIC: 121
```

Ilustración 87. Regresión logística Semirural 2

El modelo óptimo presenta un AIC de 121, sin embargo, parece demasiado simple, puesto que únicamente tiene en cuenta la variable contratos, y por lo tanto se intentará encontrar un modelo más complejo de tal forma que su AIC no crezca demasiado. Si se examina la traza que ha generado el algoritmo:

```
Step:  AIC=122.17
dats$mayoriaDerechaSemi ~ dats$emp100 + dats$contratos1000 +
    dats$licencias1000

                     Df Deviance    AIC
- dats$licencias1000  1   115.94 121.94
<none>                    114.17 122.17
- dats$emp100         1   117.01 123.01
- dats$contratos1000  1   120.63 126.63
```

Ilustración 88. Regresión logística Semirural 3

Se seleccionará el modelo anterior, puesto que su AIC (122.17) es bastante similar al mínimo obtenido e incluye las variables que presentaban los p-valores más pequeños, en cuanto a contrastar su significancia en el modelo. Por lo tanto, se obtiene:

$$\ln\left(\frac{P(mayoriaDerecha = 1)}{(1 - P(mayoriaDerecha = 1))}\right) = -0.4 + 9.77Empresas + 10.14Contratos - 6.08Licen$$

Predicciones	0	1
Observaciones		
0	9	24
1	4	59

Tabla 34. Matriz Confusión Semirurales

$$sensibilidad = \frac{(59)}{(59 + 4)} = 0.94$$

$$especificidad = \frac{(9)}{(24 + 9)} = 0.27$$

El modelo estimado en los municipios semirurales, indica que, el voto mayoritario a la derecha está asociado a valores grandes de las variables Empresas y Contratos y a valores pequeños de la variable licencias.

El modelo presenta valores altos para la sensibilidad, del 94%, lo que quiere decir que apenas se le escapan casos en los que la mayoría de voto sea para la derecha, al contrario de lo que ocurre con la especificidad, que es del 27%, y por lo tanto no resulta útil para discriminar los casos en los que la mayoría de voto es a la izquierda.

Para terminar, se muestran los resultados para los municipios **urbanos**, donde se encuentran los resultados más relevantes, puesto que, aunque se trata de un pequeño número de municipios, es donde se encuentra recogida la mayor parte la población.

Se comenzará por generar el modelo completo, para aplicar posteriormente el algoritmo Backward, obteniendo:

```
Coefficients:
                                   Estimate Std. Error z value Pr(>|z|)
(Intercept)                      -0.5293226 12.3501488  -0.043    0.966
datu$`Renta neta media por persona`  0.0004063  0.0011428   0.356    0.722
datu$`Renta neta media por hogar`  -0.0001653  0.0003742  -0.442    0.659
datu$emp100                       1.2759734  3.5674882   0.358    0.721
datu$decl1000                    -0.0852060  0.3306166  -0.258    0.797
datu$paro1000                     0.5631957  1.0703003   0.526    0.599
datu$contratos1000                0.9758724  2.5285520   0.386    0.700
datu$licencias1000               -0.3194424  0.9298542  -0.344    0.731
datu$trab1000                     2.3509496  1.9727443   1.192    0.233
datu$estab1000                   -7.5840725  8.7229032  -0.869    0.385

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 21.901  on 18  degrees of freedom
Residual deviance: 18.744  on  9  degrees of freedom
AIC: 38.744
```

Ilustración 89. Regresión Logística Urbanos 1

El AIC es 38.744, y ninguna de las variables resulta significativa al ser incluida en el modelo, por lo tanto, tras aplicar el algoritmo Backward se obtienen los siguientes resultados:

```
Call:  glm(formula = datu$mayoriaDerechaU ~ 1, family = binomial, data = datu)

Coefficients:
(Intercept)
       1.03

Degrees of Freedom: 18 Total (i.e. Null);  18 Residual
Null Deviance:      21.9
Residual Deviance: 21.9        AIC: 23.9
```

Ilustración 90. Regresión Logística Urbanos 2

El modelo obtenido no es viable, puesto que no incluye ninguna variable, por lo tanto, se buscará un modelo adecuado intentando que su AIC no sea muy superior a 23.9, que es el que presenta el modelo vacío. Examinando la traza que ha generado el algoritmo:

```
Step:  AIC=29.2
datu$mayoriaDerechaU ~ datu$`Renta neta media por hogar` + datu$paro1000 +
    datu$trab1000 + datu$estab1000

                                    Df Deviance    AIC
- datu$paro1000                      1   19.384 27.384
- datu$`Renta neta media por hogar`  1   19.679 27.679
- datu$estab1000                     1   20.180 28.180
- datu$trab1000                      1   21.139 29.139
<none>                                   19.201 29.201
```

Ilustración 91. Regresión Logística Urbanos 3

El modelo anterior presenta un AIC próximo al del modelo óptimo (29.2) y de nuevo, incluye las variables con los coeficientes más pequeños en cuanto a su significancia en el modelo, destacando que ninguna de ellas es realmente significativa, obteniéndose:

$$\ln\left(\frac{P(mayoriaDerecha = 1)}{(1 - P(mayoriaDerecha = 1))}\right) = 4.6 + 2.2Trab - 4.6Estab + 0.3Paro - 0.0002RH$$

Predicciones	0	1
Observaciones		
0	0	5
1	1	13

Tabla 35. Matriz Confusión Urbanos

$$sensibilidad = \frac{(13)}{(13+1)} = 0.93$$

$$especificidad = \frac{(0)}{(0+5)} = 0$$

Observando los resultados para los municipios **urbanos**, se puede ver que el modelo presenta una gran sensibilidad, del 93%, lo que parece indicar que es un buen modelo para detectar los casos en los que la mayoría de voto es a la derecha, aunque sigue sin serlo para los contrarios.

El modelo estimado en los municipios urbanos indica que, el voto mayoritario a la derecha está asociado a valores grandes de las variables Trabajo y Paro y a valores pequeños de las variables Establecimientos y Renta por Hogar.

Por último, se estimarán las odds para los dos modelos anteriores, también conocidas como la razón de probabilidad de verdadero, que se definen como la ratio entre la probabilidad de que la variable respuesta sea 1 y la probabilidad de que sea 0:

$$odds = \frac{P(Y=1|X=x)}{1-P(Y=1|X=x)} = e^{\beta_0+\beta_1 X1+\beta_2 X2+\cdots+\beta_n Xn}$$

A modo de ejemplo, se aplicará en los valores medios de las variables incluidas en los modelos, para los 274 municipios cuya mayoría de voto es a la derecha y en los 128 cuya mayoría es a la izquierda.

En primer lugar, se muestran dos tablas. La primera de ellas corresponde a los valores medios de las 3 variables del modelo 1, en los municipios semirurales cuya mayoría de voto es la derecha y aquellos en los que es a la izquierda. La segunda tabla recoge los valores medios de las 4 variables del modelo 2, en los municipios con mayoría a derecha entre los urbanos y en los municipios con mayoría a izquierda dentro de este mismo grupo.

Variable	Valor medio
Empresas por cada 1000 habitantes D	0.066
Empresas por cada 1000 habitantes I	0.065
Contratos por cada 1000 habitantes D	0.038
Contratos por cada 1000 habitantes I	0.021
Licencias por cada 1000 habitantes D	0.092
Licencias por cada 1000 habitantes I	0.075

Tabla 36. Valores medios Modelo 1

Observando la primera tabla, se puede ver que los municipios semirurales cuya mayoría de voto es a la derecha, presentan valores medios más altos para las 3 variables incluidas en el modelo aplicado.

Variable	Valor medio
Trabajadores por cada 1000 habitantes D	0.082
Trabajadores por cada 1000 habitantes I	0.069
Establecimientos por cada 1000 habit D	0.023
Establecimientos por cada 1000 habit I	0.038
Paro por cada 1000 habitantes D	0.065
Paro por cada 1000 habitantes I	0.100
Renta neta media por hogar D	22886.96
Renta neta media por hogar I	25776.69

Tabla 37. Valores medios Modelo 2

Sin embargo, en la segunda tabla puede verse que para las variables que afectan al modelo 2, los municipios urbanos presentan valores medios más altos para el número de trabajadores si su mayoría es a la derecha, y valores medios más altos para el número de establecimientos, el número de parados y la renta neta media por hogar, si su mayoría es a la izquierda.

- En los municipios semirurales:
  - Utilizando los valores medios de las variables en los municipios con mayoría en la derecha entre los semirurales:

$$odds = \frac{P(MayDere = 1|X = x)}{1 - P(MayDer = 1|X = x)} = \frac{P(MayDere = 1|X = x)}{P(MayIzq = 1|X = x)} = 1.08$$

  - Utilizando los valores medios de las variables en los municipios con mayoría en la izquierda entre los semirurales:

$$odds = \frac{P(MayDere = 1|X = x)}{1 - P(MayDer = 1|X = x)} = \frac{P(MayDere = 1|X = x)}{P(MayIzq = 1|X = x)} = 0.98$$

En los municipios semirurales, los valores de las odds, calculados en los dos conjuntos de valores de las variables consideradas en el modelo, son muy parecidos, 1.08 y 0.98 respectivamente, indicando que la probabilidad de que un municipio semirural tenga mayoría de derechas es poco dependiente de las variables consideradas.

- En los municipios urbanos:
  - Utilizando los valores medios de las variables en los municipios con mayoría en la derecha entre los urbanos:

$$odds = \frac{P(MayDere = 1|X = x)}{1 - P(MayDer = 1|X = x)} = \frac{P(MayDere = 1|X = x)}{P(MayIzq = 1|X = x)} = 1.13$$

  - Utilizando los valores medios de las variables en los municipios con mayoría en la izquierda entre los urbanos:

$$odds = \frac{P(MayDere = 1|X = x)}{1 - P(MayDer = 1|X = x)} = \frac{P(MayDere = 1|X = x)}{P(MayIzq = 1|X = x)} = 0.59$$

En los municipios urbanos, los valores de las odds, calculados en los dos conjuntos de valores de las variables considerados son muy diferentes, 1.13 y 0.59 respectivamente, indicando que la probabilidad de que un municipio urbano tenga mayoría de derecha o de izquierda es muy dependiente de las 4 variables consideradas en el modelo.

Capítulo 4 CONCLUSIONES

Tras aplicar las diferentes técnicas multivariantes a la tabla obtenida después de la descarga y tratamiento de los datos, no se ha encontrado una relación clara entre las variables económicas analizadas y el sentido del voto. Sin embargo, si se han obtenido algunas conclusiones individuales en cada uno de los análisis.

Para el conjunto inicial de municipios, se ha concluido que los resultados proporcionados por el análisis de componentes principales y el de correlaciones canónicas no son de gran utilidad, puesto que el análisis clúster y el ANOVA muestran que la Población ejerce gran influencia sobre el resto de variables. Esto se debe a que las diferencias de medias significativas se producen entre el grupo 1 y los demás, y dicho clúster representa a los municipios con menor número de habitantes. La expliación de esta idea puede residir en que en municipios pequeños se suele votar más en función de las personas que de los partidos. Por ello, se tomó la decisión de separar los municipios según su número de habitantes y repetir los análisis para después comparar los resultados, obteniéndose las siguientes conclusiones, que continúan para los tres **tipos de municipios** por separado.

Utilizando el Análisis de Componentes Principales, se ha observado que, en el caso de los municipios **rurales** y de los **semirurales,** no se han encontrado relaciones claras, mientras que, en los **urbanos**, se establecen las siguientes:

- Otros, presenta relación con Paro Total y Total Empresas por Municipio.
- Izquierda con Total Contratos, Número Total Licencias y Número Trabajadores.
- Derecha con Número Establecimientos, Número Total Declaraciones y Población.

Observando el **Análisis de Correlaciones Canónicas**, para el conjunto de municipios **rurales** no se encuentran conclusiones que establezcan relaciones claras entre las variables económicas y las electorales, puesto que la primera variable canónica económica, resulta un mal predictor para las 3 variables electorales. Algo parecido ocurre con los municipios **semirurales**, cuya primera variable canónica económica solo resulta buen predictor para la variable Derecha. Sin embargo, en el grupo de municipios urbanos la primera variable canónica económica, resulta un buen predictor para Derecha e Izquierda.

La **Regresión Logística**, se ha aplicado sin tener en cuenta aquellos municipios en los que la mayoría de voto era a otros partidos, puesto que se trataba de un grupo minoritario, descartándose también aquellos en los que el número de habitantes no llegaba a 2000, que son los municipios **rurales**, ya que esta situación se producía en la mayor parte de los municipios disponibles y el hecho de que presentasen un valor de población tan pequeño enmascaraba los resultados. En el grupo de municipios **semirurales**, la probabilidad de que un municipio tenga mayoría de derechas aumenta con el número de empresas y número de contratos por cada 1000 habitantes y disminuye con el número de licencias por cada 1000 habitantes. En los municipios **urbanos**, la probabilidad de que un municipio tenga mayoría de derechas aumenta con el número de Trabajadores y número de parados por cada 1000 habitantes y disminuye con el número de establecimientos por cada 1000 habitantes.

REFERENCIAS

[1] INE (2019). “Cifras oficiales de población resultantes de la revisión del Padrón municipal a 1 de enero”. https://www.ine.es/dynt3/inebase/index.html?padre=525 Consultado el 5 de octubre de 2020.

[2] INE (2017).” Altas de distribución de renta de los hogares”. https://www.ine.es/experimental/atlas/exp_atlas_tab.htm Consultado el 5 de octubre de 2020.

[3] Agencia Tributaria (2018). “Estadística de los declarantes del IRPF por municipios”. https://www.agenciatributaria.es/AEAT.internet/datosabiertos/catalogo/hacienda/Estadistica_de_los_declarantes_del_IRPF_por_municipios.shtml Consultado el 5 de octubre de 2020.

[4] SEPE (abril 2019). “Paro registrado y contratos por municipios”. http://www.sepe.es/HomeSepe/que-es-el-sepe/estadisticas/datos-estadisticos/municipios Consultado el 5 de octubre de 2020.

[5] Seguridad Social: Estadísticas (2019). “Datos por municipios. Trabajadores en alta. Último día del mes. DATOS RECIENTES”. http://www.seg-social.es/wps/portal/wss/internet/EstadisticasPresupuestosEstudios/Estadisticas/EST8/EST10/EST305/1836 Consultado el 5 de octubre de 2020.

[6] INE (2019). “Resultados municipales. Empresas por municipio y actividad principal”. https://www.ine.es/jaxiT3/Tabla.htm?t=4721 Consultado el 5 de octubre de 2020.

[7] Junta de Castilla y León (2019). “Estadística de Castilla y León. SIE.IAE”. https://estadistica.jcyl.es/web/es/estadistica.html Consultado el 5 de octubre de 2020.

[8] Dirección general de política interior (26 de mayo de 2019). “Consulta de resultados electorales”. http://www.infoelectoral.mir.es/infoelectoral/min/areaDescarga.html?method=inicio Consultado el 5 de octubre de 2020.

[9] Gónzalez Argüelles, Fco Javier (20 de agosto de 2018). “Paquetes avanzados con R”. https://rstudio-pubs-static.s3.amazonaws.com/414477_1dbd5b3ef6854d35bef3402e987695a4.html Consultado el 12 de noviembre de 2020.

[10] R for Data Science, de Hadley Wickham y Garrett Grolemund (enero de 2017). “Datos relacionales”. https://es.r4ds.hadley.nz/datos-relacionales.html Consultado el 12 de noviembre de 2020.

[11] RYTE WIKI (2012). “Análisis Multivariante”. https://es.ryte.com/wiki/An%C3%A1lisis_Multivariante Consultado el 3 de diciembre de 2020.

[12] Gil Martínez, Cristina (junio de 2018). “Análisis de Componentes Principales (PCA)”. https://rpubs.com/Cristina_Gil/PCA Consultado el 11 de enero de 2021.

[13] Amat Rodrigo, Joaquín (junio 2017). "Análisis de Componentes Principales (Principal Component Analysis, PCA) y t-SNE". https://rpubs.com/Joaquin_AR/287787 Consultado el 11 de enero de 2021.

[14] Sanchez Pantigoso, Christian Francisco (18 de noviembre de 2019) "Análisis de componentes". https://rpubs.com/Csanchez15/551258 Consultado el 11 de enero de 2021.

[15] Ayala, Jairo (17 de febrero de 2020). "Minería de Datos. Análisis de Correlación Canónica". https://rpubs.com/JairoAyala/575486#:~:text=El%20An%C3%A1lisis%20de%20Correlaci%C3%B3n%20Can%C3%B3nica,econ%C3%B3mico%2C%20demogr%C3%A1fico%2C%20social%20etc. Consultado el 1 de marzo de 2021.

[16] Hidalgo, Kevin (21 de agosto de 2017). "Cluster". https://rpubs.com/kfhidalgoh/300948 Consultado el 9 de enero de 2021.

[17] Universidad de Valencia (2010). "Introducción al Análisis Cluster" https://www.uv.es/ceaces/multivari/cluster/CLUSTER2.htm Consultado el 9 de enero de 2021.

[18] De la Fuente Fernández, Santiago (2011). "Análisis Conglomerados. Fac. Ciencias Económicas y Empresariales. Universidad Autónoma de Madrid". https://www.estadistica.net/Master-Econometria/Analisis_Cluster.pdf Consultado el 10 de enero de 2021.

[19] García García, Francisco (1 de enero de 2014). "Análisis de la Varianza (ANOVA) con R". https://biocosas.github.io/R/050_anova.html Consultado el 3 de marzo de 2021.

[20] Telefónica Tech (9 de septiembre de 2020). "Cómo interpretar la matriz de confusión: ejemplo práctico". https://empresas.blogthinkbig.com/como-interpretar-la-matriz-de-confusion-ejemplo-practico/ Consultado el 13 de mayo de 2021.

[21] Ollé, Jordi. Conceptos Claros (2019). "Qué es y cómo interpretar una regresión logística". https://conceptosclaros.com/que-es-regresion-logistica/ Consultado el 13 de mayo de 2021.

[22] Gil Martínez, Cristina (mayo de 2018). "Regresión Logística (Simple y Múltiple)". https://rpubs.com/Cristina_Gil/Regresion_Logistica Consultado el 13 de mayo de 2021.

[23] Amat Rodrigo, Joaquín (agosto de 2016). "Regresión logística simple y múltiple". https://rpubs.com/Joaquin_AR/229736 Consultado el 13 de mayo de 2021.

LISTA DE FIGURAS

LISTA DE TABLAS

ANEXO DE CÓDIGOS

OBTENCIÓN DE DATOS

Creación de las nuevas variables electorales

```
#CARGAR LAS LIBRERIAS
library(readxl)
library(dplyr)
library (xlsx)
#LECTURA DEL ARCHIVO
paraDepurar<-read_excel("C:/…/auxiliar.xlsx")
#DERECHA:
#PARTIDO POPULAR, CIUDADANOS-PARTIDO DE LA CIUDADANIA,VOX
Derecha<-paraDepurar$`PP`+paraDepurar$`Cs`+paraDepurar$VOX
#IZQUIERDA:
#PARTIDO SOCIALISTA OBRERO ESPAÑOL,UNIDAS PODEMOS IZQUIERDA UNIDA,IZQUIERDA
UNIDA, PODEMOS, PODEMOS-EQUO
Izquierda<-
paraDepurar$`PSOE`+paraDepurar$`PODEMOS.IU`+paraDepurar$`IU`+paraDepurar$PODEM
OS+paraDepurar$`PODEMOS EQUO`
#Se eliminan los partidos de DERECHA Y DE IZQUIERDA
paraDep<-select(paraDepurar,-PP,-Cs,-VOX,-PSOE,-PODEMOS.IU,-IU,-PODEMOS,-
`PODEMOS EQUO`)
#OTROS:RESTO DE PARTIDOS
Otros<-rowSums(paraDep)
#Se añaden las 3 nuevas variables a los datos
datosDIO<-cbind(Derecha,Izquierda,Otros)
#Creacion del nuevo archivo
auxiliarDIO<-
write.xlsx(datosDIO,"C:/Users/maria/Desktop/TFGNoviembre/escribir2/a8DIO.xlsx"
,col.names = TRUE, row.names = TRUE)
```

UNIÓN DE LOS DATOS

Unión Izquierda

```
#Cargar las librerías
library(readxl)
library(dplyr)
library (xlsx)
library(data.table)
library(tidyverse)
#Cargar los archivos
archivo7<-read_excel("C:/…/archivo7j.xlsx")
archivo12<-read_excel("C:/…/archivo12j.xlsx")
archivo3<-read_excel("C:/…/archivo3j.xlsx")
archivo4<-read_excel("C:/…/archivo4j.xlsx")
archivo5<-read_excel("C:/…/archivo5j.xlsx")
archivo6<-read_excel("C:/…/archivo6j.xlsx")
#Convertir los archivos en data.table
a7DT<-as.data.table(archivo7)
a12DT<-as.data.table(archivo12)
a3DT<-as.data.table(archivo3)
a4DT<-as.data.table(archivo4)
a5DT<-as.data.table(archivo5)
a6DT<-as.data.table(archivo6)
#Unir los archivos
a712DT<-a7DT %>%
  left_join(a12DT, by = "Municipio")

a7123DT<-a712DT %>%
  left_join(a3DT, by = "Municipio")

a71234DT<-a7123DT %>%
  left_join(a4DT, by = "Municipio")

a712345DT<-a71234DT %>%
```

```
  left_join(a5DT, by = "Municipio")

a7123456DT<-a712345DT %>%
  left_join(a6DT, by = "Municipio")
#Crear nuevo archivo
a7123456<-write.xlsx(a7123456DT,"C:/…/archivoJ.xlsx",col.names = TRUE,
row.names = TRUE)
```

DEPURACIÓN DE LOS DATOS

Restricciones para la variable Población

```
#Cargar las librerías
library(readxl)
library(dplyr)
library (xlsx)
#Cargar el archivo
datosLimpios<-read_excel("C:/…/archivoJ.xlsx")
#Aplicar restricciones a la variable Población
d500<-datosLimpios[as.numeric(datosLimpios$Poblacion)>500,]
d100<-datosLimpios[as.numeric(datosLimpios$Poblacion)>100,]
#Crear los nuevos archivos
datos1.500<-write.xlsx(d500,"C:/…/archivo500J.xlsx",col.names = TRUE, row.names
= TRUE)
datos1.100<-write.xlsx(d100,"C:/…/archivo100J.xlsx",col.names = TRUE, row.names
= TRUE)
```

EXPLORACIÓN DE LOS DATOS

```
#Cargar las librerías
library(readxl)
library (xlsx)
library(ggplot2)
library(ggthemes)
#Cargar el archivo
datos<-read_excel("C:/…/archivo500J0.xlsx")
names(datos)
#Crear la variable tipoMunicipio
tipoMunicipio<-rep(0,length(datos$Poblacion))
pob<-datos$Poblacion
for (i in 1:length(tipoMunicipio)) {
  if(pob[i]<2000){
    tipoMunicipio[i]<-"rural"
  }
  if(pob[i]>2000 & pob[i]<=10000){
    tipoMunicipio[i]<-"semiRural"
  }
  if(pob[i]>10000){
    tipoMunicipio[i]<-"urbano"
  }
}
#Añadir la variable a los datos
datos<-cbind(datos,tipoMunicipio)
#Modificar los datos
datos$`Renta neta media por persona`<-as.numeric(datos$`Renta neta media por
persona`)
datos$`Renta neta media por hogar`<-as.numeric(datos$`Renta neta media por
hogar`)
datos$`Total Empresas por Municipio`<-as.numeric(datos$`Total Empresas por
Municipio`)
datos$`Numero Total Declaraciones`<-as.numeric(datos$`Numero Total
Declaraciones`)
datos$tipoMunicipio<-as.factor(datos$tipoMunicipio)
#Frecuencia de la variable tipoMunicipio
table(datos$tipoMunicipio)
prop.table(table(datos$tipoMunicipio))
ggplot(data = datos, aes(x = datos$tipoMunicipio)) +
  geom_bar(color = 'darkslategray', fill = 'steelblue') +
```

```
  xlab("Tipo de Municipio") +
  ylab("Frecuencia") +
  ggtitle("Frecuencia de Tipo de Municipio")
poblacionTotal<-sum(datos$Poblacion)
rural<-datos[which(datos$tipoMunicipio=="rural"),]
poblacionRural<-sum(rural$Poblacion)
propPobRural<-poblacionRural/poblacionTotal
semiRural<-datos[which(datos$tipoMunicipio=="semiRural"),]
poblacionSemi<-sum(semiRural$Poblacion)
propPobSemi<-poblacionSemi/poblacionTotal
urbano<-datos[which(datos$tipoMunicipio=="urbano"),]
poblacionUrbana<-sum(urbano$Poblacion)
propPobUrbana<-poblacionUrbana/poblacionTotal
#Poblacion
summary(datos$Poblacion)
L<-c(500,1000,2000,5000,10000,30000,50000,70000,110000,130000,170000,300000)
intervaloPoblacion<-cut(datos$Poblacion, breaks = L, right = FALSE,
include.lowest = TRUE,
                        labels = c("(500,1000]",
"(1000,2000]","(2000,5000]","(5000,10000]","(10000,30000]","(30000,50000]"

,"(50000,70000]","(70000,110000]","(110000,130000]","(130000,170000]",
                                   "(170000,300000]"))
datos<-cbind(datos,intervaloPoblacion)
ggplot(data = datos, aes(x = datos$intervaloPoblacion)) +
  geom_bar(color = 'darkslategray', fill = 'steelblue') +
  xlab("Intervalos de Poblacion") +
  ylab("Frecuencia") +
  ggtitle("Frecuencia de Poblacion por Intervalos")
#Renta neta media por persona
summary(datos$`Renta neta media por persona`)
ggplot(datos, aes(x=datos$`Renta neta media por persona`)) +
  ggtitle("Renta neta media por persona") +
  theme_fivethirtyeight() +
  geom_histogram(bins = 9,color="#28324a", fill="#3c78d8")
ggplot(datos) +
  geom_histogram(bins = 9, aes(x = datos$`Renta neta media por persona`, fill
= datos$tipoMunicipio), color = 'black') +
  facet_grid(tipoMunicipio~., scales = 'free') +
  xlab("Renta neta media por persona") +
  ylab("Frecuencia") +
  ggtitle("Distribución de la variable Renta neta media por persona para los
distintos tipos de municipio") +
  theme_minimal()+
  labs(fill = 'Tipo de Municipio')
#Renta neta media por hogar
summary(datos$`Renta neta media por hogar`)
ggplot(datos, aes(x=datos$`Renta neta media por hogar`)) +
  ggtitle("Renta neta media por hogar") +
  theme_fivethirtyeight() +
  geom_histogram(bins = 9,color="#28324a", fill="#3c78d8")
ggplot(datos) +
  geom_histogram(bins = 9, aes(x = datos$`Renta neta media por hogar`, fill =
datos$tipoMunicipio), color = 'black') +
  facet_grid(tipoMunicipio~., scales = 'free') +
  xlab("Renta neta media por hogar") +
  ylab("Frecuencia") +
  ggtitle("Distribución de la variable Renta neta media por hogar para los
distintos tipos de municipio") +
  theme_minimal()+
  labs(fill = 'Tipo de Municipio')
#Total Empresas
pobTotal<-sum(datos$Poblacion)
emp1000<-rep(0,length(datos$`Total Empresas por Municipio`))
emp<-datos$`Total Empresas por Municipio`
for (i in 1:length(emp1000)) {
  emp1000[i]<-(emp[i]/pobTotal)*1000
```

```
}
summary(emp1000)
datos<-cbind(datos,emp1000)
L2<-c(0,0.01,0.02,0.03,0.04,0.05,0.75,0.1,0.5,1,5,10)
intervaloEmpresas<-cut(datos$emp1000, breaks = L2, right = FALSE,
include.lowest = TRUE,
                            labels = c("(0,0.01]",
"(0.01,0.02]","(0.02,0.3]","(0.03,0.04]","(0.04,0.05]","(0.05,0.1]"

,"(0.1,0.5]","(0.5,0.75]","(0.75,1]","(1,5]",
                                       "(5,10]"))
datos<-cbind(datos,intervaloEmpresas)
ggplot(data = datos, aes(x = datos$intervaloEmpresas)) +
  geom_bar(color = 'darkslategray', fill = 'steelblue') +
  xlab("Empresas por cada 1000 habitantes") +
  ylab("Frecuencia") +
  ggtitle("Frecuencia Empresas ")
ggplot(datos, aes(datos$intervaloEmpresas,  fill=datos$tipoMunicipio)) +
  geom_bar(position="dodge",colour="black")+
  labs(x= "Empresas por cada 1000 habitantes",  y="Frecuencias", fill="Tipo
Municipio")
#Numero Declaraciones
decl1000<-rep(0,length(datos$`Numero Total Declaraciones`))
decl<-datos$`Numero Total Declaraciones`
for (i in 1:length(decl1000)) {
  decl1000[i]<-(decl[i]/pobTotal)*1000

}
summary(decl1000)
datos<-cbind(datos,decl1000)
L3<-c(0.03,0.04,0.05,0.1,0.25,0.5,1,5,10,30,50,82)
intervaloDeclaraciones<-cut(datos$decl1000, breaks = L3, right = FALSE,
include.lowest = TRUE,
                            labels = c("(0.03,0.04]",
"(0.04,0.05]","(0.05,0.1]","(0.1,0.25]","(0.25,0.5]","(0.5,1]"
                                       ,"(1,5]","(5,10]","(10,30]","(30,50]",
                                       "(50,82]"))
datos<-cbind(datos,intervaloDeclaraciones)
ggplot(data = datos, aes(x = datos$intervaloDeclaraciones)) +
  geom_bar(color = 'darkslategray', fill = 'steelblue') +
  xlab("Declaraciones por cada 1000 habitantes") +
  ylab("Frecuencia") +
  ggtitle("Frecuencia Declaraciones")
ggplot(datos, aes(datos$intervaloDeclaraciones,  fill=datos$tipoMunicipio)) +
  geom_bar(position="dodge",colour="black")+
  labs(x= "Declaraciones por cada 1000 habitantes",  y="Frecuencias",
fill="Tipo Municipio")
#Paro Total
paro1000<-rep(0,length(datos$`Paro Total`))
paro<-datos$`Paro Total`
for (i in 1:length(paro1000)) {
  paro1000[i]<-(paro[i]/pobTotal)*1000

}
summary(paro1000)
datos<-cbind(datos,paro1000)
L4<-c(0.003,0.005,0.01,0.02,0.05,0.1,0.2,0.5,1,2,5,8.5)
intervaloParo<-cut(datos$paro1000, breaks = L4, right = FALSE, include.lowest
= TRUE,
                   labels = c("(0.003,0.005]",
"(0.005,0.01]","(0.01,0.02]","(0.02,0.05]","(0.05,0.1]","(0.1,0.2]"
                              ,"(0.2,0.5]","(0.5,1]","(1,2]","(2,5]",
                              "(5,8.5]"))
datos<-cbind(datos,intervaloParo)
ggplot(data = datos, aes(x = datos$intervaloParo)) +
  geom_bar(color = 'darkslategray', fill = 'steelblue') +
  xlab("Paro por cada 1000 habitantes") +
  ylab("Frecuencia") +
```

```
  ggtitle("Frecuencia Paro")
ggplot(datos, aes(datos$intervaloParo,  fill=datos$tipoMunicipio)) +
  geom_bar(position="dodge",colour="black")+
  labs(x= "Paro por cada 1000 habitantes",  y="Frecuencias", fill="Tipo
Municipio")
#Total Contratos
contratos1000<-rep(0,length(datos$`Total Contratos`))
contratos<-datos$`Total Contratos`
for (i in 1:length(contratos1000)) {
  contratos1000[i]<-(contratos[i]/pobTotal)*1000

}
summary(contratos1000)
datos<-cbind(datos,contratos1000)
L5<-c(0,0.005,0.01,0.02,0.05,0.1,0.2,0.5,1,2,5,7.1)
intervaloContratos<-cut(datos$contratos1000, breaks = L5, right = FALSE,
include.lowest = TRUE,
                        labels = c("(0,0.005]",
"(0.005,0.01]","(0.01,0.02]","(0.02,0.05]","(0.05,0.1]","(0.1,0.2]"
                                   ,"(0.2,0.5]","(0.5,1]","(1,2]","(2,5]",
                                   "(5,7.1]"))
datos<-cbind(datos,intervaloContratos)
ggplot(data = datos, aes(x = datos$intervaloContratos)) +
  geom_bar(color = 'darkslategray', fill = 'steelblue') +
  xlab("Contratos por cada 1000 habitantes") +
  ylab("Frecuencia") +
  ggtitle("Frecuencia Contratos")
ggplot(datos, aes(datos$intervaloContratos,  fill=datos$tipoMunicipio)) +
  geom_bar(position="dodge",colour="black")+
  labs(x= "Contratos por cada 1000 habitantes",  y="Frecuencias", fill="Tipo
Municipio")
#Numero Licencias
licencias1000<-rep(0,length(datos$`Numero Total Licencias`))
licencias<-datos$`Numero Total Licencias`
for (i in 1:length(licencias1000)) {
  licencias1000[i]<-(licencias[i]/pobTotal)*1000

}
summary(licencias1000)
datos<-cbind(datos,licencias1000)
L6<-c(0,0.005,0.01,0.02,0.05,0.1,0.2,0.5,1,2,5,13.1)
intervaloLicencias<-cut(datos$licencias1000, breaks = L6, right = FALSE,
include.lowest = TRUE,
                        labels = c("(0,0.005]",
"(0.005,0.01]","(0.01,0.02]","(0.02,0.05]","(0.05,0.1]","(0.1,0.2]"
                                   ,"(0.2,0.5]","(0.5,1]","(1,2]","(2,5]",
                                   "(5,13.1]"))
datos<-cbind(datos,intervaloLicencias)
ggplot(data = datos, aes(x = datos$intervaloLicencias)) +
  geom_bar(color = 'darkslategray', fill = 'steelblue') +
  xlab("Numero Licencias por cada 1000 habitantes") +
  ylab("Frecuencia") +
  ggtitle("Frecuencia de Numero Licencias")
ggplot(datos, aes(datos$intervaloLicencias, fill=datos$tipoMunicipio)) +
  geom_bar(position="dodge",colour="black")+
  labs(x= "Numero Licencias por cada 1000 habitantes",  y="Frecuencias",
fill="Tipo Municipio")
#Numero Trabajadores
trabajadores1000<-rep(0,length(datos$`Numero Trabajadores`))
trabajadores<-datos$`Numero Trabajadores`
for (i in 1:length(trabajadores1000)) {
  trabajadores1000[i]<-(trabajadores[i]/pobTotal)*1000

}
summary(trabajadores1000)
datos<-cbind(datos,trabajadores1000)
L7<-c(0,0.005,0.01,0.02,0.05,0.1,0.2,0.5,1,2,5,10.6)
```

```
intervaloTrabajadores<-cut(datos$trabajadores1000, breaks = L7, right = FALSE,
include.lowest = TRUE,
                           labels = c("(0,0.005]",
"(0.005,0.01]","(0.01,0.02]","(0.02,0.05]","(0.05,0.1]","(0.1,0.2]"
                                      ,"(0.2,0.5]","(0.5,1]","(1,2]","(2,5]",
                                      "(5,10.6]"))
datos<-cbind(datos,intervaloTrabajadores)
ggplot(data = datos, aes(x = datos$intervaloTrabajadores)) +
  geom_bar(color = 'darkslategray', fill = 'steelblue') +
  xlab("Trabajadores por cada 1000 habitantes") +
  ylab("Frecuencia") +
  ggtitle("Frecuencia Trabajadores")
ggplot(datos, aes(datos$intervaloTrabajadores, fill=datos$tipoMunicipio)) +
  geom_bar(position="dodge",colour="black")+
  labs(x= "Trabajadores por cada 1000 habitantes",  y="Frecuencias",
fill="Tipo Municipio")
#Numero Establecimientos
estab1000<-rep(0,length(datos$`Numero Establecimientos`))
estab<-datos$`Numero Establecimientos`
for (i in 1:length(estab1000)) {
  estab1000[i]<-(estab[i]/pobTotal)*1000

}
summary(estab1000)
datos<-cbind(datos,estab1000)
L8<-c(0,0.005,0.01,0.02,0.05,0.1,0.2,0.5,1,1.5,2,3.8)
intervaloEstablecimientos<-cut(datos$estab1000, breaks = L8, right = FALSE,
include.lowest = TRUE,
                               labels = c("(0,0.005]",
"(0.005,0.01]","(0.01,0.02]","(0.02,0.05]","(0.05,0.1]","(0.1,0.2]"

,"(0.2,0.5]","(0.5,1]","(1,1.5]","(1.5,2]",
                                          "(2,3.8]"))
datos<-cbind(datos,intervaloEstablecimientos)
ggplot(data = datos, aes(x = datos$intervaloEstablecimientos)) +
  geom_bar(color = 'darkslategray', fill = 'steelblue') +
  xlab("Establecimientos por cada 1000 habitantes") +
  ylab("Frecuencia") +
  ggtitle("Frecuencia Establecimientos")
ggplot(datos, aes(datos$intervaloEstablecimientos, fill=datos$tipoMunicipio))
+
  geom_bar(position="dodge",colour="black")+
  labs(x= "Establecimientos por cada 1000 habitantes",  y="Frecuencias",
fill="Tipo Municipio")
#Derecha
derecha1000<-rep(0,length(datos$Derecha))
derecha<-datos$Derecha
for (i in 1:length(derecha1000)) {
  derecha1000[i]<-(derecha[i]/pobTotal)*1000

}
summary(derecha1000)
datos<-cbind(datos,derecha1000)
L9<-c(0,0.005,0.01,0.02,0.05,0.1,0.2,0.5,1,2,5,39.1)
intervaloDerecha<-cut(datos$derecha1000, breaks = L9, right = FALSE,
include.lowest = TRUE,
                      labels = c("(0,0.005]",
"(0.005,0.01]","(0.01,0.02]","(0.02,0.05]","(0.05,0.1]","(0.1,0.2]"
                                 ,"(0.2,0.5]","(0.5,1]","(1,2]","(2,5]",
                                 "(5,39.1]"))
datos<-cbind(datos,intervaloDerecha)
ggplot(data = datos, aes(x = datos$intervaloDerecha)) +
  geom_bar(color = 'darkslategray', fill = 'steelblue') +
  xlab("Derecha por cada 1000 habitantes") +
  ylab("Frecuencia") +
  ggtitle("Frecuencia de Derecha")
ggplot(datos, aes(datos$intervaloDerecha, fill=datos$tipoMunicipio)) +
  geom_bar(position="dodge",colour="black")+
```

```
  labs(x= "Derecha por cada 1000 habitantes",  y="Frecuencias", fill="Tipo
Municipio")
#Izquierda
izquierda1000<-rep(0,length(datos$Izquierda))
izquierda<-datos$Izquierda
for (i in 1:length(izquierda1000)) {
  izquierda1000[i]<-(izquierda[i]/pobTotal)*1000

}
summary(izquierda1000)
datos<-cbind(datos,izquierda1000)
L10<-c(0,0.005,0.01,0.02,0.05,0.1,0.2,0.5,1,2,5,30.9)
intervaloIzqu<-cut(datos$izquierda1000, breaks = L10, right = FALSE,
include.lowest = TRUE,
                        labels = c("(0,0.005]",
"(0.005,0.01]","(0.01,0.02]","(0.02,0.05]","(0.05,0.1]","(0.1,0.2]"
                                   ,"(0.2,0.5]","(0.5,1]","(1,2]","(2,5]",
                                   "(5,30.9]"))
datos<-cbind(datos,intervaloIzqu)
ggplot(data = datos, aes(x = datos$intervaloIzqu)) +
  geom_bar(color = 'darkslategray', fill = 'steelblue') +
  xlab("Izquierda por cada 1000 habitantes") +
  ylab("Frecuencia") +
  ggtitle("Frecuencia de Izquierda")
ggplot(datos, aes(datos$intervaloIzqu, fill=datos$tipoMunicipio)) +
  geom_bar(position="dodge",colour="black")+
  labs(x= "Izquierda por cada 1000 habitantes",  y="Frecuencias", fill="Tipo
Municipio")


#Otros
otros1000<-rep(0,length(datos$Otros))
otros<-datos$Otros
for (i in 1:length(otros1000)) {
  otros1000[i]<-(otros[i]/pobTotal)*1000

}
summary(otros1000)
datos<-cbind(datos,otros1000)
L11<-c(0,0.005,0.01,0.02,0.05,0.1,0.2,0.5,1,2,5,9.3)
intervaloOtros<-cut(datos$otros1000, breaks = L11, right = FALSE,
include.lowest = TRUE,
                    labels = c("(0,0.005]",
"(0.005,0.01]","(0.01,0.02]","(0.02,0.05]","(0.05,0.1]","(0.1,0.2]"
                               ,"(0.2,0.5]","(0.5,1]","(1,2]","(2,5]",
                               "(5,9.3]"))
datos<-cbind(datos,intervaloOtros)
ggplot(data = datos, aes(x = datos$intervaloOtros)) +
  geom_bar(color = 'darkslategray', fill = 'steelblue') +
  xlab("Otros por cada 1000 habitantes") +
  ylab("Frecuencia") +
  ggtitle("Frecuencia de Otros")
ggplot(datos, aes(datos$intervaloOtros, fill=datos$tipoMunicipio)) +
  geom_bar(position="dodge",colour="black")+
  labs(x= "Otros por cada 1000 habitantes",  y="Frecuencias", fill="Tipo
Municipio")
#Proporciones de variables por tipo de municipio
rural<-datos[which(datos$tipoMunicipio=="rural"),]
semiRural<-datos[which(datos$tipoMunicipio=="semiRural"),]
urbano<-datos[which(datos$tipoMunicipio=="urbano"),]
#Población
poblacionTotal<-sum(datos$Poblacion)
poblacionRural<-sum(rural$Poblacion)
propPobRural<-poblacionRural/poblacionTotal
poblacionSemi<-sum(semiRural$Poblacion)
propPobSemi<-poblacionSemi/poblacionTotal
poblacionUrbana<-sum(urbano$Poblacion)
propPobUrbana<-poblacionUrbana/poblacionTotal
```

```
#Derecha
derechaTotal<-sum(datos$derecha1000)
derechaRural<-sum(rural$derecha1000)
propDerechaRural<-derechaRural/derechaTotal
derechaSemi<-sum(semiRural$derecha1000)
propDerechaSemi<-derechaSemi/derechaTotal
derechaUrbana<-sum(urbano$derecha1000)
propDerechaUrbana<-derechaUrbana/derechaTotal
#Izquierda
izquierdaTotal<-sum(datos$izquierda1000)
izquierdaRural<-sum(rural$izquierda1000)
propIzquierdaRural<-izquierdaRural/izquierdaTotal
izquierdaSemi<-sum(semiRural$izquierda1000)
propIzquierdaSemi<-izquierdaSemi/izquierdaTotal
izquierdaUrbana<-sum(urbano$izquierda1000)
propIzquierdaUrbana<-izquierdaUrbana/izquierdaTotal
#Otros
otrosTotal<-sum(datos$otros1000)
otrosRural<-sum(rural$otros1000)
propOtrosRural<-otrosRural/otrosTotal
otrosSemi<-sum(semiRural$otros1000)
propOtrosSemi<-otrosSemi/otrosTotal
otrosUrbana<-sum(urbano$otros1000)
propOtrosUrbana<-otrosUrbana/otrosTotal
#Renta neta media por persona
rpTotal<-sum(datos$`Renta neta media por persona`)
rpRural<-sum(rural$`Renta neta media por persona`)
proprpRural<-rpRural/rpTotal
rpSemi<-sum(semiRural$`Renta neta media por persona`)
proprpsSemi<-rpSemi/rpTotal
rpUrbana<-sum(urbano$`Renta neta media por persona`)
proprpUrbana<-rpUrbana/rpTotal
#Renta neta media por hogar
rhTotal<-sum(datos$`Renta neta media por hogar`)
rhRural<-sum(rural$`Renta neta media por hogar`)
proprhRural<-rhRural/rhTotal
rhSemi<-sum(semiRural$`Renta neta media por hogar`)
proprhSemi<-rhSemi/rhTotal
rhUrbana<-sum(urbano$`Renta neta media por hogar`)
proprhUrbana<-rhUrbana/rhTotal
#Total empresas por municipio
empresasTotal<-sum(datos$emp1000)
empresasRural<-sum(rural$emp1000)
propempresasRural<-empresasRural/empresasTotal
empresasSemi<-sum(semiRural$emp1000)
propempresasSemi<-empresasSemi/empresasTotal
empresasUrbana<-sum(urbano$emp1000)
propempresasUrbana<-empresasUrbana/empresasTotal
#Número total declaraciones
decTotal<-sum(datos$decl1000)
decRural<-sum(rural$decl1000)
propdecRural<-decRural/decTotal
decSemi<-sum(semiRural$decl1000)
propdecSemi<-decSemi/decTotal
decUrbana<-sum(urbano$decl1000)
propdecUrbana<-decUrbana/decTotal
#Paro total
paroTotal<-sum(datos$paro1000)
paroRural<-sum(rural$paro1000)
propparoRural<-paroRural/paroTotal
paroSemi<-sum(semiRural$paro1000)
propparoSemi<-paroSemi/paroTotal
paroUrbana<-sum(urbano$paro1000)
propparoUrbana<-paroUrbana/paroTotal
#Total contratos
contTotal<-sum(datos$contratos1000)
contRural<-sum(rural$contratos1000)
propcontRural<-contRural/contTotal
```

```
contSemi<-sum(semiRural$contratos1000)
propcontSemi<-contSemi/contTotal
contUrbana<-sum(urbano$contratos1000)
propcontUrbana<-contUrbana/contTotal
#Número total licencias
licTotal<-sum(datos$licencias1000)
licRural<-sum(rural$licencias1000)
proplicRural<-licRural/licTotal
licSemi<-sum(semiRural$licencias1000)
proplicSemi<-licSemi/licTotal
licUrbana<-sum(urbano$licencias1000)
proplicUrbana<-licUrbana/licTotal
#Número trabajadores
trabTotal<-sum(datos$trabajadores1000)
trabRural<-sum(rural$trabajadores1000)
proptrabRural<-trabRural/trabTotal
trabSemi<-sum(semiRural$trabajadores1000)
proptrabSemi<-trabSemi/trabTotal
trabUrbana<-sum(urbano$trabajadores1000)
proptrabUrbana<-trabUrbana/trabTotal
#Número establecimientos
estabTotal<-sum(datos$estab1000)
estabRural<-sum(rural$estab1000)
propestabRural<-estabRural/estabTotal
estabSemi<-sum(semiRural$estab1000)
propestabSemi<-estabSemi/estabTotal
estabUrbana<-sum(urbano$estab1000)
propestabUrbana<-estabUrbana/estabTotal
```

RESULTADOS GENERALES

```
#Cargar las librerias
library(readxl)
library (xlsx)
library(psych)
library(GGally)
library(ggplot2)
library(FactoMineR)
library(factoextra)
library(fields)
library(CCA)
library(vegan)
library(candisc)
library(cluster)
library("clValid")
library(fmsb)
#Cargar los datos
datos<-read_excel("C:/…/archivo500J0.xlsx")
#Crear la variable tipoMunicipio
tipoMunicipio<-rep(0,length(datos$Poblacion))
pob<-datos$Poblacion
for (i in 1:length(tipoMunicipio)) {
  if(pob[i]<2000){
    tipoMunicipio[i]<-"rural"
  }
  if(pob[i]>2000 & pob[i]<=10000){
    tipoMunicipio[i]<-"semiRural"
  }
  if(pob[i]>10000){
    tipoMunicipio[i]<-"urbano"
  }
}
tipoMunicipio<-tipoMunicipio
datos<-cbind(datos,tipoMunicipio)
datos$tipoMunicipio<-as.factor(datos$tipoMunicipio)
datos$`Renta neta media por persona`<-as.numeric(datos$`Renta neta media por
persona`)
```

```
datos$`Renta neta media por hogar`<-as.numeric(datos$`Renta neta media por
hogar`)
datos$`Total Empresas por Municipio`<-as.numeric(datos$`Total Empresas por
Municipio`)
datos$`Numero Total Declaraciones`<-as.numeric(datos$`Numero Total
Declaraciones`)
#Outliers
impute_outliers <- function(x, removeNA = TRUE){
  quantiles <- quantile(x, c(0.05, 0.95), na.rm = removeNA)
  x[x<quantiles[1]] <- mean(x, na.rm = removeNA)
  x[x>quantiles[2]] <- median(x, na.rm = removeNA)
  x
}
datosPob<-impute_outliers(datos$Poblacion)
datosDer<-impute_outliers(datos$Derecha)
datosIz<-impute_outliers(datos$Izquierda)
datosOtros<-impute_outliers(datos$Otros)
datosRentPers<-impute_outliers(datos$`Renta neta media por persona`)
datosRentHog<-impute_outliers(datos$`Renta neta media por hogar`)
datosEmp<-impute_outliers(datos$`Total Empresas por Municipio`)
datosDec<-impute_outliers(datos$`Numero Total Declaraciones`)
datosParo<-impute_outliers(datos$`Paro Total`)
datosContratos<-impute_outliers(datos$`Total Contratos`)
datosLic<-impute_outliers(datos$`Numero Total Licencias`)
datosTrabajadores<-impute_outliers(datos$`Numero Trabajadores`)
datosEstab<-impute_outliers(datos$`Numero Establecimientos`)
dat<-
cbind(datosPob,datosDer,datosIz,datosOtros,datosRentPers,datosRentHog,datosEmp
,datosDec,datosParo,datosContratos,datosLic,datosTrabajadores,datosEstab)
dat<-as.data.frame(dat)
names(dat)<-c("Poblacion","Derecha","Izquierda",
              "Otros","Renta neta media por persona","Renta neta media por
hogar",
              "Total Empresas por Municipio","Numero Total
Declaraciones","Paro Total","Total Contratos",
              "Numero Total Licencias","Numero Trabajadores","Numero
Establecimientos")

table(datos$tipoMunicipio)
#ANÁLISIS DE COMPONENTES PRINCIPALES
datpca<-dat
pca1<-prcomp(datpca,scale.=TRUE)
#Contribucion de las variables a cada dimensión
fviz_contrib(pca1, choice = "var", axes = 1, top = 13)
fviz_contrib(pca1, choice = "var", axes = 2, top = 13)
fviz_contrib(pca1, choice = "var", axes = 3, top = 13)
fviz_contrib(pca1, choice = "var", axes = 4, top = 13)
fviz_contrib(pca1, choice = "var", axes = 5, top = 13)
fviz_contrib(pca1, choice = "var", axes = 6, top = 13)
fviz_contrib(pca1, choice = "var", axes = 7, top = 13)
fviz_contrib(pca1, choice = "var", axes = 8, top = 13)
fviz_contrib(pca1, choice = "var", axes = 9, top = 13)
fviz_contrib(pca1, choice = "var", axes = 10, top = 13)
fviz_contrib(pca1, choice = "var", axes = 11, top = 13)
fviz_contrib(pca1, choice = "var", axes = 12, top = 13)
fviz_contrib(pca1, choice = "var", axes = 13, top = 13)
#Componenete 1
PCA(dat,ncp=13,axes = c(1,2))
PCA(dat,ncp=13,axes = c(1,4))
PCA(dat,ncp=13,axes = c(1,5))
PCA(dat,ncp=13,axes = c(1,8))
#Componente 2 PCA(dat,ncp=13,axes = c(2,4))
PCA(dat,ncp=13,axes = c(2,5))
PCA(dat,ncp=13,axes = c(2,8))
#Componente 4
PCA(dat,ncp=13,axes = c(4,5))
PCA(dat,ncp=13,axes = c(4,8))
#Componente 5
```

```
PCA(dat,ncp=13,axes = c(5,8))
#ANALISIS DE CORRELACIONES CANONICAS
#Datos de variables electorales
X<-cbind(dat$Derecha,dat$Izquierda,dat$Otros)
colnames(X)<-c("Derecha","Izquierda","Otros")
#Datos de variables economicas
Y<-cbind(dat$Poblacion,dat$`Renta neta media por persona`,dat$`Renta neta
media por hogar`,dat$`Total Empresas por Municipio`,dat$`Numero Total
Declaraciones`,dat$`Paro Total`,dat$`Total Contratos`,dat$`Numero Total
Licencias`,dat$`Numero Trabajadores`,dat$`Numero Establecimientos`)
colnames(Y)<-c("Poblacion","Renta neta media por persona","Renta neta media
por hogar",
            "Total Empresas por Municipio","Numero Total Declaraciones","Paro
Total","Total Contratos",
            "Numero Total Licencias","Numero Trabajadores","Numero
Establecimientos")
velectorales<-as.data.frame(X)
veconomicas<-as.data.frame(Y)
cor(veconomicas,velectorales)
colnames(X)
XX<-X
colnames(XX)<-c("X1","X2","X3")
colnames(Y)
YY<-Y
colnames(YY)<-c("Y1","Y2","Y3","Y4","Y5","Y6","Y7","Y8","Y9","Y10")
cor(XX)
cor(YY)
cor(XX,YY)
cc<-cancor(X,Y,set.names = c("Electorales","Economicas"))
cc
summary(cc)
cc$structure
redundancy(cc)
coef(cc,type = "both",standardize = TRUE)
#ANALISIS CLUSTER
my_data<-scale(dat)
d<-dist(my_data, method = "euclidean")
#Representación de los dendogramas con los 3 metodos
plot(hclust(d, method = "complete"),
     main = "Complete linkage",
     xlab = "",
     ylab = "",
     cex = 0.8,
     sub = "")

plot(hclust(d, method = "average"),
     main = "Average linkage",
     xlab = "",
     ylab = "",
     cex = 0.8,
     sub = "")

plot(hclust(d, method = "single"),
     main = "Single linkage",
     xlab = "",
     ylab = "",
     cex = 0.8,
     sub = "")
#Clustering jerarquico con complete linkage
clustComp<-hclust(d)
clustComp
#Corte del dendograma resultante en 8 clusters.
clusterHC<-cutree(tree = clustComp,k=8)
clusterHC
#Representación gráfica del corte en el dendograma
plot(clustComp, cex = 0.6)
rect.hclust(clustComp, k = 8, border = 2:5)
table(clusterHC, datos$tipoMunicipio)
```

```
table(datos$tipoMunicipio, clusterHC)
#K-MEANS clustering
set.seed(123)
fviz_nbclust(my_data, kmeans, nstart = 25, method = "gap_stat", nboot = 50)+
  labs(subtitle = "Gap statistic method")
#K=8 y 20 asignaciones aleatorias de clusters iniciales
kMeans<-kmeans(x=my_data,centers = 8,nstart = 20)
fviz_cluster(kMeans, data = my_data, frame.type = "convex")
#Asignacion de clusters de cada observacion
clustersKM<-kMeans$cluster
table(clustersKM, datos$tipoMunicipio)
table(datos$tipoMunicipio, clustersKM)
#Elegir algoritmo de agrupacion adecuado
intern <- clValid(my_data, nClust = 2:10,
                  clMethods = c("hierarchical","kmeans"),
                  validation = "internal")
summary(intern)
plot(intern)
#ANOVA
#Seleccionamos el método jerárquico
datos<-cbind(datos,clusterHC)
datos$clusterHC<-as.factor(datos$clusterHC)
n1<-datos[which(datos$clusterHC==1),]
nombres1<-n1$Municipio
n2<-datos[which(datos$clusterHC==2),]
nombres2<-n2$Municipio
n3<-datos[which(datos$clusterHC==3),]
nombres3<-n3$Municipio
n4<-datos[which(datos$clusterHC==4),]
nombres4<-n4$Municipio
n5<-datos[which(datos$clusterHC==5),]
nombres5<-n5$Municipio
n6<-datos[which(datos$clusterHC==6),]
nombres6<-n6$Municipio
n7<-datos[which(datos$clusterHC==7),]
nombres7<-n7$Municipio
n8<-datos[which(datos$clusterHC==8),]
nombres8<-n8$Municipio
#Poblacion
poblacion<-datos$Poblacion
cluster<-datos$clusterHC
boxplot(poblacion ~ cluster, col = c(1,2,3,4,5,6,7,8), ylab = "Poblacion")
tapply(poblacion, cluster, mean)
fm<-aov(lm(poblacion~cluster))
summary(fm)
intervals<-TukeyHSD(fm)
intervals
plot(intervals)
plot(fm$residuals)
qqnorm(fm$residuals)
qqline(fm$residuals)
boxplot(fm$residuals~cluster, col = c(1,2,3,4,5,6,7,8))
logpoblacion<-log(poblacion)
boxplot(logpoblacion ~ cluster, col = c(1,2,3,4,5,6,7,8), ylab = "Poblacion")
tapply(logpoblacion, cluster, mean)
fmlogpob<-aov(lm(logpoblacion~cluster))
summary(fmlogpob)
intervalslogpob<-TukeyHSD(fmlogpob)
intervalslogpob
plot(intervalslogpob)
plot(fmlogpob$residuals)
qqnorm(fmlogpob$residuals)
qqline(fmlogpob$residuals)
boxplot(fmlogpob$residuals~cluster, col = c(1,2,3,4,5,6,7,8))
#DERECHA
derecha<-datos$Derecha
boxplot(derecha ~ cluster, col = c(1,2,3,4,5,6,7,8), ylab = "Votos Derecha")
fmder<-aov(lm(derecha~cluster))
```

```r
summary(fmder)
intervalsder<-TukeyHSD(fmder)
intervalsder
plot(intervalsder)
plot(fmder$residuals)
qqnorm(fmder$residuals)
qqline(fmder$residuals)
boxplot(fmder$residuals~cluster, col = c(1,2,3,4,5,6,7,8))
for (i in 1:length(derecha)){
  if(derecha[i]==0){
    derecha[i]<-0.000001
  }
}
logder<-log(derecha)
for (i in 1:length(logder)){
  if(logder[i]==log(0.000001)){
    logder[i]<-0
  }
}
boxplot(logder ~ cluster, col = c(1,2,3,4,5,6,7,8), ylab = "Votos Derecha")
fmlogder<-aov(lm(logder~cluster))
summary(fmlogder)
intervalslogder<-TukeyHSD(fmlogder)
intervalslogder
plot(intervalslogder)
plot(fmlogder$residuals)
qqnorm(fmlogder$residuals)
qqline(fmlogder$residuals)
boxplot(fmlogder$residuals~cluster, col = c(1,2,3,4,5,6,7,8))
#IZQUIERDA
izquierda<-datos$Izquierda
boxplot(izquierda ~ cluster, col = c(1,2,3,4,5,6,7,8), ylab = "Votos
Izquierda")
fmizq<-aov(lm(izquierda~cluster))
summary(fmizq)
intervalsizq<-TukeyHSD(fmizq)
intervalsizq
plot(intervalsizq)
plot(fmizq$residuals)
qqnorm(fmizq$residuals)
qqline(fmizq$residuals)
boxplot(fmizq$residuals~cluster, col = c(1,2,3,4,5,6,7,8))
for (i in 1:length(izquierda)){
  if(izquierda[i]==0){
    izquierda[i]<-0.000001
  }
}
logizq<-log(izquierda)
for (i in 1:length(logizq)){
  if(logizq[i]==log(0.000001)){
    logizq[i]<-0
  }
}
boxplot(logizq ~ cluster, col = c(1,2,3,4,5,6,7,8), ylab = "Votos Izquierda")
fmlogizq<-aov(lm(logizq~cluster))
summary(fmlogizq)
intervalslogizq<-TukeyHSD(fmlogizq)
intervalslogizq
plot(intervalslogizq)
plot(fmlogizq$residuals)
qqnorm(fmlogizq$residuals)
qqline(fmlogizq$residuals)
boxplot(fmlogizq$residuals~cluster, col = c(1,2,3,4,5,6,7,8))
#OTROS
otros<-datos$Otros
boxplot(otros ~ cluster, col = c(1,2,3,4,5,6,7,8), ylab = "Votos Otros")
fmot<-aov(lm(otros~cluster))
summary(fmot)
```

```
intervalsot<-TukeyHSD(fmot)
intervalsot
plot(intervalsot)
plot(fmot$residuals)
qqnorm(fmot$residuals)
qqline(fmot$residuals)
boxplot(fmot$residuals~cluster, col = c(1,2,3,4,5,6,7,8))
for (i in 1:length(otros)){
  if(otros[i]==0){
    otros[i]<-0.000001
  }
}
logot<-log(otros)
for (i in 1:length(logot)){
  if(logot[i]==log(0.000001)){
    logot[i]<-0
  }
}
boxplot(logot~ cluster, col = c(1,2,3,4,5,6,7,8), ylab = "Votos Otros")
fmlogot<-aov(lm(logot~cluster))
summary(fmlogot)
intervalslogot<-TukeyHSD(fmlogot)
intervalslogot
plot(intervalslogot)
plot(fmlogot$residuals)
qqnorm(fmlogot$residuals)
qqline(fmlogot$residuals)
boxplot(fmlogot$residuals~cluster, col = c(1,2,3,4,5,6,7,8))
#RENTA NETA MEDIA POR PERSONA
rentaPersona<-datos$`Renta neta media por persona`
boxplot(rentaPersona ~ cluster, col = c(1,2,3,4,5,6,7,8), ylab = "Renta por
persona")
fmrp<-aov(lm(rentaPersona~cluster))
summary(fmrp)
intervalsrp<-TukeyHSD(fmrp)
intervalsrp
plot(intervalsrp)
plot(fmrp$residuals)
qqnorm(fmrp$residuals)
qqline(fmrp$residuals)
boxplot(fmrp$residuals~cluster, col = c(1,2,3,4,5,6,7,8))
#RENTA NETA MEDIA POR HOGAR
rentaHogar<-datos$`Renta neta media por hogar`
boxplot(rentaHogar ~ cluster, col = c(1,2,3,4,5,6,7,8), ylab = "Renta por
hogar")
fmrh<-aov(lm(rentaHogar~cluster))
summary(fmrh)
intervalsrh<-TukeyHSD(fmrh)
intervalsrh
plot(intervalsrh)
plot(fmrh$residuals)
qqnorm(fmrh$residuals)
qqline(fmrh$residuals)
boxplot(fmrh$residuals~cluster, col = c(1,2,3,4,5,6,7,8))
#TOTAL EMPRESAS POR MUNICIPIO
empresas<-datos$`Total Empresas por Municipio`
boxplot(empresas ~ cluster, col = c(1,2,3,4,5,6,7,8), ylab = "Numero
Empresas")
fmemp<-aov(lm(empresas~cluster))
summary(fmemp)
intervalsemp<-TukeyHSD(fmemp)
intervalsemp
plot(intervalsemp)
plot(fmemp$residuals)
qqnorm(fmemp$residuals)
qqline(fmemp$residuals)
boxplot(fmemp$residuals~cluster, col = c(1,2,3,4,5,6,7,8))
logemp<-log(empresas)
```

```
boxplot(logemp~ cluster, col = c(1,2,3,4,5,6,7,8), ylab = "Numero Empresas")
fmlogemp<-aov(lm(logemp~cluster))
summary(fmlogemp)
intervalslogemp<-TukeyHSD(fmlogemp)
intervalslogemp
plot(intervalslogemp)
plot(fmlogemp$residuals)
qqnorm(fmlogemp$residuals)
qqline(fmlogemp$residuals)
boxplot(fmlogemp$residuals~cluster, col = c(1,2,3,4,5,6,7,8))
#NUMERO TOTAL DECLARACIONES
declaraciones<-datos$`Numero Total Declaraciones`
boxplot(declaraciones ~ cluster, col = c(1,2,3,4,5,6,7,8), ylab = "Numero
Declaraciones")
fmdec<-aov(lm(declaraciones~cluster))
summary(fmdec)
intervalsdec<-TukeyHSD(fmdec)
intervalsdec
plot(intervalsdec)
plot(fmdec$residuals)
qqnorm(fmdec$residuals)
qqline(fmdec$residuals)
boxplot(fmdec$residuals~cluster, col = c(1,2,3,4,5,6,7,8))
logdec<-log(declaraciones)
boxplot(logdec~ cluster, col = c(1,2,3,4,5,6,7,8), ylab = "Numero
Declaraciones")
fmlogdec<-aov(lm(logdec~cluster))
summary(fmlogdec)
intervalslogdec<-TukeyHSD(fmlogdec)
intervalslogdec
plot(intervalslogdec)
plot(fmlogdec$residuals)
qqnorm(fmlogdec$residuals)
qqline(fmlogdec$residuals)
boxplot(fmlogdec$residuals~cluster, col = c(1,2,3,4,5,6,7,8))
#PARO TOTAL
paro<-datos$`Paro Total`
boxplot(paro ~ cluster, col = c(1,2,3,4,5,6,7,8), ylab = "Numero Parados")
fmparo<-aov(lm(paro~cluster))
summary(fmparo)
intervalsparo<-TukeyHSD(fmparo)
intervalsparo
plot(intervalsparo)
plot(fmparo$residuals)
qqnorm(fmparo$residuals)
qqline(fmparo$residuals)
boxplot(fmparo$residuals~cluster, col = c(1,2,3,4,5,6,7,8))
logparo<-log(paro)
boxplot(logparo~ cluster, col = c(1,2,3,4,5,6,7,8), ylab = "Numero Parados")
fmlogparo<-aov(lm(logparo~cluster))
summary(fmlogparo)
intervalslogparo<-TukeyHSD(fmlogparo)
intervalslogparo
plot(intervalslogparo)
plot(fmlogparo$residuals)
qqnorm(fmlogparo$residuals)
qqline(fmlogparo$residuals)
boxplot(fmlogparo$residuals~cluster, col = c(1,2,3,4,5,6,7,8))
#TOTAL COTRATOS
contratos<-datos$`Total Contratos`
boxplot(contratos ~ cluster, col = c(1,2,3,4,5,6,7,8), ylab = "Numero
Contratos")
fmcontratos<-aov(lm(contratos~cluster))
summary(fmcontratos)
intervalscontratos<-TukeyHSD(fmcontratos)
intervalscontratos
plot(intervalscontratos)
plot(fmcontratos$residuals)
```

```
qqnorm(fmcontratos$residuals)
qqline(fmcontratos$residuals)
boxplot(fmcontratos$residuals~cluster, col = c(1,2,3,4,5,6,7,8))
for (i in 1:length(contratos)){
  if(contratos[i]==0){
    contratos[i]<-0.000001
  }
}
logcontratos<-log(contratos)
for (i in 1:length(logcontratos)){
  if(logcontratos[i]==log(0.000001)){
    logcontratos[i]<-0
  }
}
boxplot(logcontratos~ cluster, col = c(1,2,3,4,5,6,7,8), ylab = "Numero
Contratos")
fmlogcontratos<-aov(lm(logcontratos~cluster))
summary(fmlogcontratos)
intervalslogcontratos<-TukeyHSD(fmlogcontratos)
intervalslogcontratos
plot(intervalslogcontratos)
plot(fmlogcontratos$residuals)
qqnorm(fmlogcontratos$residuals)
qqline(fmlogcontratos$residuals)
boxplot(fmlogcontratos$residuals~cluster, col = c(1,2,3,4,5,6,7,8))
#NUMERO TOTAL LICENCIAS
licencias<-datos$`Numero Total Licencias`
boxplot(licencias ~ cluster, col = c(1,2,3,4,5,6,7,8), ylab = "Numero
Licencias")
fmlicencias<-aov(lm(licencias~cluster))
summary(fmlicencias)
intervalslicencias<-TukeyHSD(fmlicencias)
intervalslicencias
plot(intervalslicencias)
plot(fmlicencias$residuals)
qqnorm(fmlicencias$residuals)
qqline(fmlicencias$residuals)
boxplot(fmlicencias$residuals~cluster, col = c(1,2,3,4,5,6,7,8))
for (i in 1:length(licencias)){
  if(licencias[i]==0){
    licencias[i]<-0.000001
  }
}
loglicencias<-log(licencias)
for (i in 1:length(loglicencias)){
  if(loglicencias[i]==log(0.000001)){
    loglicencias[i]<-0
  }
}
boxplot(loglicencias~ cluster, col = c(1,2,3,4,5,6,7,8), ylab = "Numero
Licencias")
fmloglicencias<-aov(lm(loglicencias~cluster))
summary(fmloglicencias)
intervalsloglicencias<-TukeyHSD(fmloglicencias)
intervalsloglicencias
plot(intervalsloglicencias)
plot(fmloglicencias$residuals)
qqnorm(fmloglicencias$residuals)
qqline(fmloglicencias$residuals)
boxplot(fmloglicencias$residuals~cluster, col = c(1,2,3,4,5,6,7,8))
#NUMERO TRABAJADORES
trabajadores<-datos$`Numero Trabajadores`
boxplot(trabajadores ~ cluster, col = c(1,2,3,4,5,6,7,8), ylab = "Numero
Trabajadores")
fmtrabajadores<-aov(lm(trabajadores~cluster))
summary(fmtrabajadores)
intervalstrabajadores<-TukeyHSD(fmtrabajadores)
intervalstrabajadores
```

```
plot(intervalstrabajadores)
plot(fmtrabajadores$residuals)
qqnorm(fmtrabajadores$residuals)
qqline(fmtrabajadores$residuals)
boxplot(fmtrabajadores$residuals~cluster, col = c(1,2,3,4,5,6,7,8))
for (i in 1:length(trabajadores)){
  if(trabajadores[i]==0){
    trabajadores[i]<-0.000001
  }
}
logtrabajadores<-log(trabajadores)
for (i in 1:length(logtrabajadores)){
  if(logtrabajadores[i]==log(0.000001)){
    logtrabajadores[i]<-0
  }
}
boxplot(logtrabajadores~ cluster, col = c(1,2,3,4,5,6,7,8), ylab = "Numero
Trabajadores")
fmlogtrabajadores<-aov(lm(logtrabajadores~cluster))
summary(fmlogtrabajadores)
intervalslogtrabajadores<-TukeyHSD(fmlogtrabajadores)
intervalslogtrabajadores
plot(intervalslogtrabajadores)
plot(fmlogtrabajadores$residuals)
qqnorm(fmlogtrabajadores$residuals)
qqline(fmlogtrabajadores$residuals)
boxplot(fmlogtrabajadores$residuals~cluster, col = c(1,2,3,4,5,6,7,8))
#NUMERO ESTABLECIMIENTOS
establecimientos<-datos$`Numero Establecimientos`
boxplot(establecimientos ~ cluster, col = c(1,2,3,4,5,6,7,8), ylab = "Numero
Establecimientos")
fmestablecimientos<-aov(lm(establecimientos~cluster))
summary(fmestablecimientos)
intervalsestablecimientos<-TukeyHSD(fmestablecimientos)
intervalsestablecimientos
plot(intervalsestablecimientos)
plot(fmestablecimientos$residuals)
qqnorm(fmestablecimientos$residuals)
qqline(fmestablecimientos$residuals)
boxplot(fmestablecimientos$residuals~cluster, col = c(1,2,3,4,5,6,7,8))
for (i in 1:length(establecimientos)){
  if(establecimientos[i]==0){
    establecimientos[i]<-0.000001
  }
}
logestablecimientos<-log(establecimientos)
for (i in 1:length(logestablecimientos)){
  if(logestablecimientos[i]==log(0.000001)){
    logestablecimientos[i]<-0
  }
}
boxplot(logestablecimientos~ cluster, col = c(1,2,3,4,5,6,7,8), ylab = "Numero
Establecimientos")
fmlogestablecimientos<-aov(lm(logestablecimientos~cluster))
summary(fmlogestablecimientos)
intervalslogestablecimientos<-TukeyHSD(fmlogestablecimientos)
intervalslogestablecimientos
plot(intervalslogestablecimientos)
plot(fmlogestablecimientos$residuals)
qqnorm(fmlogestablecimientos$residuals)
qqline(fmlogestablecimientos$residuals)
boxplot(fmlogestablecimientos$residuals~cluster, col = c(1,2,3,4,5,6,7,8))
```

RESULTADOS POR TIPOS DE MUNICIPIO

```
#Cargar librerías
library(readxl)
library (xlsx)
library(psych)
library(GGally)
library(ggplot2)
library(FactoMineR)
library(factoextra)
library(fields)
library(CCA)
library(vegan)
library(candisc)
library(ggpubr)
library(gridExtra)
library(corrplot)
#Cargar los datos
datos<-read_excel("C:/…/archivo500J0.xlsx")
names(datos)
#Crear variable tipoMunicipio
tipoMunicipio<-rep(0,length(datos$Poblacion))
pob<-datos$Poblacion
for (i in 1:length(tipoMunicipio)) {
  if(pob[i]<2000){
    tipoMunicipio[i]<-"rural"
  }
  if(pob[i]>2000 & pob[i]<=10000){
    tipoMunicipio[i]<-"semiRural"
  }
  if(pob[i]>10000){
    tipoMunicipio[i]<-"urbano"
  }
}
#Crear variables mayoriaElección
mayoriaDerecha<-rep(0,length(datos$Derecha))
derecha<-datos$Derecha
izquierda<-datos$Izquierda
otros<-datos$Otros
for (i in 1:length(mayoriaDerecha)) {
  if(derecha[i]>izquierda[i] && derecha[i]>otros[i]){
    mayoriaDerecha[i]<-1
  }
  else{
    mayoriaDerecha[i]<-0
  }
}

mayoriaIzquierda<-rep(0,length(datos$Izquierda))
for (i in 1:length(mayoriaIzquierda)) {
  if(izquierda[i]>derecha[i] && izquierda[i]>otros[i]){
    mayoriaIzquierda[i]<-1
  }
  else{
    mayoriaIzquierda[i]<-0
  }
}

mayoriaOtros<-rep(0,length(datos$Otros))
for (i in 1:length(mayoriaOtros)) {
  if(otros[i]>derecha[i] && otros[i]>izquierda[i]){
    mayoriaOtros[i]<-1
  }
  else{
    mayoriaOtros[i]<-0
  }
}
```

```
datos<-cbind(datos,tipoMunicipio,mayoriaDerecha,mayoriaIzquierda,mayoriaOtros)

datos$`Renta neta media por persona`<-as.numeric(datos$`Renta neta media por
persona`)
datos$`Renta neta media por hogar`<-as.numeric(datos$`Renta neta media por
hogar`)
datos$`Total Empresas por Municipio`<-as.numeric(datos$`Total Empresas por
Municipio`)
datos$`Numero Total Declaraciones`<-as.numeric(datos$`Numero Total
Declaraciones`)
datos$tipoMunicipio<-as.factor(datos$tipoMunicipio)
datos$mayoriaDerecha<-as.factor(datos$mayoriaDerecha)
datos$mayoriaIzquierda<-as.factor(datos$mayoriaIzquierda)
datos$mayoriaOtros<-as.factor(datos$mayoriaOtros)
#Crear variables por cada 1000 habitantes
pobTotal<-sum(datos$Poblacion)
emp1000<-rep(0,length(datos$`Total Empresas por Municipio`))
emp<-datos$`Total Empresas por Municipio`
for (i in 1:length(emp1000)) {
  emp1000[i]<-(emp[i]/pobTotal)*1000
}
decl1000<-rep(0,length(datos$`Numero Total Declaraciones`))
decl<-datos$`Numero Total Declaraciones`
for (i in 1:length(decl1000)) {
  decl1000[i]<-(decl[i]/pobTotal)*1000
}
paro1000<-rep(0,length(datos$`Paro Total`))
paro<-datos$`Paro Total`
for (i in 1:length(paro1000)) {
  paro1000[i]<-(paro[i]/pobTotal)*1000
}
contratos1000<-rep(0,length(datos$`Total Contratos`))
contratos<-datos$`Total Contratos`
for (i in 1:length(contratos1000)) {
  contratos1000[i]<-(contratos[i]/pobTotal)*1000
}
licencias1000<-rep(0,length(datos$`Numero Total Licencias`))
licencias<-datos$`Numero Total Licencias`
for (i in 1:length(licencias1000)) {
  licencias1000[i]<-(licencias[i]/pobTotal)*1000
}
trab1000<-rep(0,length(datos$`Numero Trabajadores`))
trab<-datos$`Numero Trabajadores`
for (i in 1:length(trab1000)) {
  trab1000[i]<-(trab[i]/pobTotal)*1000
}
estab1000<-rep(0,length(datos$`Numero Establecimientos`))
estab<-datos$`Numero Establecimientos`
for (i in 1:length(estab1000)) {
  estab1000[i]<-(estab[i]/pobTotal)*1000
}
datos<-
cbind(datos,emp1000,decl1000,paro1000,contratos1000,licencias1000,trab1000,est
ab1000)

table(datos$mayoriaDerecha)
table(datos$mayoriaIzquierda)
table(datos$mayoriaOtros)

rurales<-datos[which(datos$tipoMunicipio=="rural"),]
nombres1<-rurales$Municipio
semi<-datos[which(datos$tipoMunicipio=="semiRural"),]
nombres2<-semi$Municipio
urbano<-datos[which(datos$tipoMunicipio=="urbano"),]
nombres3<-urbano$Municipio
#Tratar outliers
impute_outliers <- function(x, removeNA = TRUE){
  quantiles <- quantile(x, c(0.05, 0.95), na.rm = removeNA)
```

```
  x[x<quantiles[1]] <- mean(x, na.rm = removeNA)
  x[x>quantiles[2]] <- median(x, na.rm = removeNA)
  x
}

datosPob<-impute_outliers(datos$Poblacion)
datosDer<-impute_outliers(datos$Derecha)
datosIz<-impute_outliers(datos$Izquierda)
datosOtros<-impute_outliers(datos$Otros)
datosRentPers<-impute_outliers(datos$`Renta neta media por persona`)
datosRentHog<-impute_outliers(datos$`Renta neta media por hogar`)
datosEmp<-impute_outliers(datos$`Total Empresas por Municipio`)
datosDec<-impute_outliers(datos$`Numero Total Declaraciones`)
datosParo<-impute_outliers(datos$`Paro Total`)
datosContratos<-impute_outliers(datos$`Total Contratos`)
datosLic<-impute_outliers(datos$`Numero Total Licencias`)
datosTrabajadores<-impute_outliers(datos$`Numero Trabajadores`)
datosEstab<-impute_outliers(datos$`Numero Establecimientos`)
em<-impute_outliers(datos$emp1000)
dc<-impute_outliers(datos$decl1000)
pa<-impute_outliers(datos$paro1000)
co<-impute_outliers(datos$contratos1000)
li<-impute_outliers(datos$licencias1000)
tr<-impute_outliers(datos$trab1000)
es<-impute_outliers(datos$estab1000)

dat<-
cbind(datosPob,datosDer,datosIz,datosOtros,datosRentPers,datosRentHog,datosEmp
,datosDec,datosParo,datosContratos,datosLic,datosTrabajadores,datosEstab,
           em,dc,pa,co,li,tr,es)
dat<-as.data.frame(dat)
names(dat)<-c("Poblacion","Derecha","Izquierda",
              "Otros","Renta neta media por persona","Renta neta media por
hogar",
              "Total Empresas por Municipio","Numero Total
Declaraciones","Paro Total","Total Contratos",
              "Numero Total Licencias","Numero Trabajadores","Numero
Establecimientos",

"emp100","decl1000","paro1000","contratos1000","licencias1000","trab1000","est
ab1000")

datosPobr<-impute_outliers(rurales$Poblacion)
datosDerr<-impute_outliers(rurales$Derecha)
datosIzr<-impute_outliers(rurales$Izquierda)
datosOtrosr<-impute_outliers(rurales$Otros)
datosRentPersr<-impute_outliers(rurales$`Renta neta media por persona`)
datosRentHogr<-impute_outliers(rurales$`Renta neta media por hogar`)
datosEmpr<-impute_outliers(rurales$`Total Empresas por Municipio`)
datosDecr<-impute_outliers(rurales$`Numero Total Declaraciones`)
datosParor<-impute_outliers(rurales$`Paro Total`)
datosContratosr<-impute_outliers(rurales$`Total Contratos`)
datosLicr<-impute_outliers(rurales$`Numero Total Licencias`)
datosTrabajadoresr<-impute_outliers(rurales$`Numero Trabajadores`)
datosEstabr<-impute_outliers(rurales$`Numero Establecimientos`)
emr<-impute_outliers(rurales$emp1000)
dcr<-impute_outliers(rurales$decl1000)
par<-impute_outliers(rurales$paro1000)
cor<-impute_outliers(rurales$contratos1000)
lir<-impute_outliers(rurales$licencias1000)
trr<-impute_outliers(rurales$trab1000)
esr<-impute_outliers(rurales$estab1000)

datr<-
cbind(datosPobr,datosDerr,datosIzr,datosOtrosr,datosRentPersr,datosRentHogr,da
tosEmpr,datosDecr,datosParor,datosContratosr,datosLicr,datosTrabajadoresr,dato
sEstabr,
            emr,dcr,par,cor,lir,trr,esr)
```

```
datr<-as.data.frame(datr)
names(datr)<-c("Poblacion","Derecha","Izquierda",
               "Otros","Renta neta media por persona","Renta neta media por
hogar",
               "Total Empresas por Municipio","Numero Total
Declaraciones","Paro Total","Total Contratos",
               "Numero Total Licencias","Numero Trabajadores","Numero
Establecimientos",

"emp100","decl1000","paro1000","contratos1000","licencias1000","trab1000","est
ab1000")

datosPobs<-impute_outliers(semi$Poblacion)
datosDers<-impute_outliers(semi$Derecha)
datosIzs<-impute_outliers(semi$Izquierda)
datosOtross<-impute_outliers(semi$Otros)
datosRentPerss<-impute_outliers(semi$`Renta neta media por persona`)
datosRentHogs<-impute_outliers(semi$`Renta neta media por hogar`)
datosEmps<-impute_outliers(semi$`Total Empresas por Municipio`)
datosDecs<-impute_outliers(semi$`Numero Total Declaraciones`)
datosParos<-impute_outliers(semi$`Paro Total`)
datosContratoss<-impute_outliers(semi$`Total Contratos`)
datosLics<-impute_outliers(semi$`Numero Total Licencias`)
datosTrabajadoress<-impute_outliers(semi$`Numero Trabajadores`)
datosEstabs<-impute_outliers(semi$`Numero Establecimientos`)
ems<-impute_outliers(semi$emp1000)
dcs<-impute_outliers(semi$decl1000)
pas<-impute_outliers(semi$paro1000)
cos<-impute_outliers(semi$contratos1000)
lis<-impute_outliers(semi$licencias1000)
trs<-impute_outliers(semi$trab1000)
ess<-impute_outliers(semi$estab1000)

dats<-
cbind(datosPobs,datosDers,datosIzs,datosOtross,datosRentPerss,datosRentHogs,da
tosEmps,datosDecs,datosParos,datosContratoss,datosLics,datosTrabajadoress,dato
sEstabs,
            ems,dcs,pas,cos,lis,trs,ess)
dats<-as.data.frame(dats)
names(dats)<-c("Poblacion","Derecha","Izquierda",
               "Otros","Renta neta media por persona","Renta neta media por
hogar",
               "Total Empresas por Municipio","Numero Total
Declaraciones","Paro Total","Total Contratos",
               "Numero Total Licencias","Numero Trabajadores","Numero
Establecimientos",

"emp100","decl1000","paro1000","contratos1000","licencias1000","trab1000","est
ab1000")

datosPobu<-impute_outliers(urbano$Poblacion)
datosDeru<-impute_outliers(urbano$Derecha)
datosIzu<-impute_outliers(urbano$Izquierda)
datosOtrosu<-impute_outliers(urbano$Otros)
datosRentPersu<-impute_outliers(urbano$`Renta neta media por persona`)
datosRentHogu<-impute_outliers(urbano$`Renta neta media por hogar`)
datosEmpu<-impute_outliers(urbano$`Total Empresas por Municipio`)
datosDecu<-impute_outliers(urbano$`Numero Total Declaraciones`)
datosParou<-impute_outliers(urbano$`Paro Total`)
datosContratosu<-impute_outliers(urbano$`Total Contratos`)
datosLicu<-impute_outliers(urbano$`Numero Total Licencias`)
datosTrabajadoresu<-impute_outliers(urbano$`Numero Trabajadores`)
datosEstabu<-impute_outliers(urbano$`Numero Establecimientos`)
emu<-impute_outliers(urbano$emp1000)
dcu<-impute_outliers(urbano$decl1000)
pau<-impute_outliers(urbano$paro1000)
cou<-impute_outliers(urbano$contratos1000)
liu<-impute_outliers(urbano$licencias1000)
```

```
tru<-impute_outliers(urbano$trab1000)
esu<-impute_outliers(urbano$estab1000)

datu<-
cbind(datosPobu,datosDeru,datosIzu,datosOtrosu,datosRentPersu,datosRentHogu,da
tosEmpu,datosDecu,datosParou,datosContratosu,datosLicu,datosTrabajadoresu,dato
sEstabu,
            emu,dcu,pau,cou,liu,tru,esu)
datu<-as.data.frame(datu)
names(datu)<-c("Poblacion","Derecha","Izquierda",
               "Otros","Renta neta media por persona","Renta neta media por
hogar",
               "Total Empresas por Municipio","Numero Total
Declaraciones","Paro Total","Total Contratos",
               "Numero Total Licencias","Numero Trabajadores","Numero
Establecimientos",

"emp100","decl1000","paro1000","contratos1000","licencias1000","trab1000","est
ab1000")

#Análisis Componentes Principales
pca<-prcomp(dat,scale.=TRUE)
aux<-fviz_screeplot(pca, addlabels = TRUE,main="Municipios",ylim = c(0, 50))
pcar<-prcomp(datr,scale.=TRUE)
auxr<-fviz_screeplot(pcar, addlabels = TRUE,main="Municipios Rurales",ylim =
c(0, 40))
pcas<-prcomp(dats,scale.=TRUE)
auxs<-fviz_screeplot(pcas, addlabels = TRUE,main="Municipios Semirurales",ylim
= c(0, 50))
pcau<-prcomp(datu,scale.=TRUE)
auxu<-fviz_screeplot(pcau, addlabels = TRUE,main="Municipios Urbanos",ylim =
c(0, 70))

ggpubr::ggarrange(aux,auxr,auxs,auxu)
gridExtra::grid.arrange(aux,auxr,auxs,auxu)

pca$rotation
pcar$rotation
pcas$rotation
pcau$rotation

au1<-fviz_contrib(pca, choice = "var", axes = 1, top = 13,title
="Municipios:Contribuciones de las variables a la Dimension1")
au2<-fviz_contrib(pca, choice = "var", axes = 2, top = 13,title
="Municipios:Contribuciones de las variables a la Dimension2")
ggpubr::ggarrange(au1,au2)

au3<-fviz_contrib(pcar, choice = "var", axes = 1, top = 13,title ="Municipios
Rurales:Contribuciones de las variables a la Dimension1")
au4<-fviz_contrib(pcar, choice = "var", axes = 2, top = 13,title ="Municipios
Rurales:Contribuciones de las variables a la Dimension2")
ggpubr::ggarrange(au3,au4)

au5<-fviz_contrib(pcas, choice = "var", axes = 1, top = 13,title ="Municipios
Semirurales:Contribuciones de las variables a la Dimension1")
au6<-fviz_contrib(pcas, choice = "var", axes = 2, top = 13,title ="Municipios
Semirurales:Contribuciones de las variables a la Dimension2")
ggpubr::ggarrange(au5,au6)

au7<-fviz_contrib(pcau, choice = "var", axes = 1, top = 13,title ="Municipios
Urbanos:Contribuciones de las variables a la Dimension1")
au8<-fviz_contrib(pcau, choice = "var", axes = 2, top = 13,title ="Municipios
Urbanos:Contribuciones de las variables a la Dimension1")
ggpubr::ggarrange(au7,au8)

PCA(dat)
PCA(datr)
PCA(dats)
```

```
PCA(datu)
#Correlaciones Canónicas
X<-cbind(dat$Derecha,dat$Izquierda,dat$Otros)
colnames(X)<-c("Derecha","Izquierda","Otros")
Y<-cbind(dat$Poblacion,dat$`Renta neta media por persona`,dat$`Renta neta
media por hogar`,dat$`Total Empresas por Municipio`,dat$`Numero Total
Declaraciones`,dat$`Paro Total`,dat$`Total Contratos`,dat$`Numero Total
Licencias`,dat$`Numero Trabajadores`,dat$`Numero Establecimientos`)
colnames(Y)<-c("Poblacion","Renta neta media por persona","Renta neta media
por hogar",
               "Total Empresas por Municipio","Numero Total
Declaraciones","Paro Total","Total Contratos",
               "Numero Total Licencias","Numero Trabajadores","Numero
Establecimientos")
velectorales<-as.data.frame(X)
veconomicas<-as.data.frame(Y)

Xr<-cbind(datr$Derecha,datr$Izquierda,datr$Otros)
colnames(Xr)<-c("Derecha","Izquierda","Otros")
Yr<-cbind(datr$Poblacion,datr$`Renta neta media por persona`,datr$`Renta neta
media por hogar`,datr$`Total Empresas por Municipio`,datr$`Numero Total
Declaraciones`,datr$`Paro Total`,datr$`Total Contratos`,datr$`Numero Total
Licencias`,datr$`Numero Trabajadores`,datr$`Numero Establecimientos`)
colnames(Yr)<-c("Poblacion","Renta neta media por persona","Renta neta media
por hogar",
               "Total Empresas por Municipio","Numero Total
Declaraciones","Paro Total","Total Contratos",
               "Numero Total Licencias","Numero Trabajadores","Numero
Establecimientos")
velectoralesr<-as.data.frame(Xr)
veconomicasr<-as.data.frame(Yr)

Xs<-cbind(dats$Derecha,dats$Izquierda,dats$Otros)
colnames(Xs)<-c("Derecha","Izquierda","Otros")
Ys<-cbind(dats$Poblacion,dats$`Renta neta media por persona`,dats$`Renta neta
media por hogar`,dats$`Total Empresas por Municipio`,dats$`Numero Total
Declaraciones`,dats$`Paro Total`,dats$`Total Contratos`,dats$`Numero Total
Licencias`,dats$`Numero Trabajadores`,dats$`Numero Establecimientos`)
colnames(Ys)<-c("Poblacion","Renta neta media por persona","Renta neta media
por hogar",
                "Total Empresas por Municipio","Numero Total
Declaraciones","Paro Total","Total Contratos",
                "Numero Total Licencias","Numero Trabajadores","Numero
Establecimientos")
velectoraless<-as.data.frame(Xs)
veconomicass<-as.data.frame(Ys)

Xu<-cbind(datu$Derecha,datu$Izquierda,datu$Otros)
colnames(Xu)<-c("Derecha","Izquierda","Otros")
Yu<-cbind(datu$Poblacion,datu$`Renta neta media por persona`,datu$`Renta neta
media por hogar`,datu$`Total Empresas por Municipio`,datu$`Numero Total
Declaraciones`,datu$`Paro Total`,datu$`Total Contratos`,datu$`Numero Total
Licencias`,datu$`Numero Trabajadores`,datu$`Numero Establecimientos`)
colnames(Yu)<-c("Poblacion","Renta neta media por persona","Renta neta media
por hogar",
                "Total Empresas por Municipio","Numero Total
Declaraciones","Paro Total","Total Contratos",
                "Numero Total Licencias","Numero Trabajadores","Numero
Establecimientos")
velectoralesu<-as.data.frame(Xu)
veconomicasu<-as.data.frame(Yu)

colnames(X)
XX<-X
colnames(XX)<-c("X1","X2","X3")
colnames(Y)
YY<-Y
colnames(YY)<-c("Y1","Y2","Y3","Y4","Y5","Y6","Y7","Y8","Y9","Y10")
```

```r
cor(XX)
cor(YY)
cor(XX,YY)

if (require(corrplot)) {
  M <- cor(cbind(XX,YY))
  corrplot(M, method="ellipse", order="hclust", addrect=2,
addCoef.col="black")
  title("Municipios",line=3)
}

colnames(Xr)
XXr<-Xr
colnames(XXr)<-c("X1","X2","X3")
colnames(Yr)
YYr<-Yr
colnames(YYr)<-c("Y1","Y2","Y3","Y4","Y5","Y6","Y7","Y8","Y9","Y10")
cor(XXr)
cor(YYr)
cor(XXr,YYr)

if (require(corrplot)) {
  M <- cor(cbind(XXr,YYr))
  corrplot(M, method="ellipse", order="hclust", addrect=2,
addCoef.col="black")
  title("Municipios Rurales",line=3)
}

colnames(Xs)
XXs<-Xs
colnames(XXs)<-c("X1","X2","X3")
colnames(Ys)
YYs<-Ys
colnames(YYs)<-c("Y1","Y2","Y3","Y4","Y5","Y6","Y7","Y8","Y9","Y10")
cor(XXs)
cor(YYs)
cor(XXs,YYs)

if (require(corrplot)) {
  M <- cor(cbind(XXs,YYs))
  corrplot(M, method="ellipse", order="hclust", addrect=2,
addCoef.col="black")
  title("Municipios Semirurales",line=3)
}

colnames(Xu)
XXu<-Xu
colnames(XXu)<-c("X1","X2","X3")
colnames(Yu)
YYu<-Yu
colnames(YYu)<-c("Y1","Y2","Y3","Y4","Y5","Y6","Y7","Y8","Y9","Y10")
cor(XXu)
cor(YYu)
cor(XXu,YYu)

if (require(corrplot)) {
  M <- cor(cbind(XXu,YYu))
  corrplot(M, method="ellipse", order="hclust", addrect=2,
addCoef.col="black")
  title("Municipios Urbanos",line=3)
}

cc<-cancor(X,Y,set.names = c("Electorales","Economicas"))
cc
summary(cc)

ccr<-cancor(Xr,Yr,set.names = c("Electorales","Economicas"))
ccr
```

```
summary(ccr)

ccs<-cancor(Xs,Ys,set.names = c("Electorales","Economicas"))
ccs
summary(ccs)

ccu<-cancor(Xu,Yu,set.names = c("Electorales","Economicas"))
ccu
summary(ccu)

coef(cc,type = "both",standardize = TRUE)
coef(ccr,type = "both",standardize = TRUE)
coef(ccs,type = "both",standardize = TRUE)
coef(ccu,type = "both",standardize = TRUE)

cc$structure
ccr$structure
ccs$structure
ccu$structure

redundancy(cc)
redundancy(ccr)
redundancy(ccs)
redundancy(ccu)
#REGRESION LOGISTICA
mayoriaDerechaRural<-mayoriaDerecha[which(datos$tipoMunicipio=="rural")]
mayoriaIzquierdaRural<-mayoriaIzquierda[which(datos$tipoMunicipio=="rural")]
mayoriaOtrosRural<-mayoriaOtros[which(datos$tipoMunicipio=="rural")]
mayoriaDerechaSemi<-mayoriaDerecha[which(datos$tipoMunicipio=="semiRural")]
mayoriaIzquierdaSemi<-
mayoriaIzquierda[which(datos$tipoMunicipio=="semiRural")]
mayoriaOtrosSemi<-mayoriaOtros[which(datos$tipoMunicipio=="semiRural")]
mayoriaDerechaU<-mayoriaDerecha[which(datos$tipoMunicipio=="urbano")]
mayoriaIzquierdaU<-mayoriaIzquierda[which(datos$tipoMunicipio=="urbano")]
mayoriaOtrosU<-mayoriaOtros[which(datos$tipoMunicipio=="urbano")]

dat<-cbind(dat,mayoriaDerecha,mayoriaOtros,mayoriaIzquierda)
dat<-dat[dat$mayoriaOtros!=1,]
datr<-cbind(datr,mayoriaDerechaRural,mayoriaOtrosRural,mayoriaIzquierdaRural)
datr<-datr[datr$mayoriaOtros!=1,]
dats<-cbind(dats,mayoriaDerechaSemi,mayoriaOtrosSemi,mayoriaIzquierdaSemi)
dats<-dats[dats$mayoriaOtros!=1,]
datu<-cbind(datu,mayoriaDerechaU,mayoriaOtrosU,mayoriaIzquierdaU)
datu<-datu[datu$mayoriaOtros!=1,]


modeloCompletoSemi<-glm(dats$mayoriaDerechaSemi ~ dats$`Renta neta media por
persona`+dats$`Renta neta media por
hogar`+dats$emp100+dats$decl1000+dats$paro1000+dats$contratos1000+dats$licenci
as1000+dats$trab1000+dats$estab1000, family = binomial, data = dats)
semiBack<-stepAIC(modeloCompletoSemi, trace = T, direction = "backward")
modelo1Semi<-glm(dats$mayoriaDerechaSemi ~
dats$emp100+dats$contratos1000+dats$licencias1000, family = binomial, data =
dats)
predicciones1<- ifelse(test = modelo1Semi$fitted.values > 0.5, yes = 1, no =
0)
matriz1<- table(modelo1Semi$model$`dats$mayoriaDerechaSemi`, predicciones1,
                   dnn = c("observaciones", "predicciones"))
matriz1

modeloCompletoUrbanos<-glm(datu$mayoriaDerechaU~ datu$`Renta neta media por
persona`+datu$`Renta neta media por
hogar`+datu$emp100+datu$decl1000+datu$paro1000+datu$contratos1000+datu$licenci
as1000+datu$trab1000+datu$estab1000, family = binomial, data = datu)
urbanoBack<-stepAIC(modeloCompletoUrbanos, trace = T, direction = "backward")
modelo1U<-glm(datu$mayoriaDerechaU~
datu$trab1000+datu$estab1000+datu$paro1000+datu$`Renta neta media por
hogar`,family = binomial, data = datu)
```

```
predicciones2 <- ifelse(test = modelo1U$fitted.values > 0.5, yes = 1, no = 0)
matriz2<- table(modelo1U$model$`datu$mayoriaDerechaU`, predicciones2,
                   dnn = c("observaciones", "predicciones"))
matriz2
#Calular odds
datMayDS<-dat[which(dats$mayoriaDerechaSemi==1),]
datMayIS<-dat[which(dats$mayoriaIzquierdaSemi==1),]
empresasDS<-mean(datMayDS$emp100)
contratosDS<-mean(datMayDS$contratos1000)
licenciasDS<-mean(datMayDS$licencias1000)
empresasIS<-mean(datMayIS$emp100)
contratosIS<-mean(datMayIS$contratos1000)
licenciasIS<-mean(datMayIS$licencias1000)
a1D<-(-0.4)+(9.77*empresasDS)+(10.14*contratosDS)+(-6.08*licenciasDS)
a1I<-(-0.4)+(9.77*empresasIS)+(10.14*contratosIS)+(-6.08*licenciasIS)
a2D<-exp(a1D)
a2I<-exp(a1I)
odds1<-((a2D)/(1+a2D))/(1-((a2D)/(1+a2D)))
odds2<-((a2I)/(1+a2I))/(1-((a2I)/(1+a2I)))

datMayDU<-dat[which(datu$mayoriaDerechaU==1),]
datMayIU<-dat[which(datu$mayoriaIzquierdaU==1),]
trabajadoresDU<-mean(datMayDU$trab1000)
estabecimientosDU<-mean(datMayDU$estab1000)
paroDU<-mean(datMayDU$paro1000)
RHDU<-mean(datMayDU$`Renta neta media por hogar`)
trabajadoresIU<-mean(datMayIU$trab1000)
estabecimientosIU<-mean(datMayIU$estab1000)
paroIU<-mean(datMayIU$paro1000)
RHIU<-mean(datMayIU$`Renta neta media por hogar`)
a3D<-(4.6)+(2.2*trabajadoresDU)+(-4.6*estabecimientosDU)+(0.3*paroDU)+(-
0.0002*RHDU)
a3I<-(4.6)+(2.2*trabajadoresIU)+(-4.6*estabecimientosIU)+(0.3*paroIU)+(-
0.0002*RHIU)
a4D<-exp(a3D)
a4I<-exp(a3I)
odds3<-((a4D)/(1+a4D))/(1-((a4D)/(1+a4D)))
odds4<-((a4I)/(1+a4I))/(1-((a4I)/(1+a4I)))
```