



Universidad de Valladolid

FACULTAD DE CIENCIAS

TRABAJO FIN DE GRADO

Grado en Estadística

**Caracterización de las Secciones Censales de
Valladolid en función de su comportamiento
electoral en las elecciones al Congreso**

Autora: Esther Lassalle Alonso

Tutora: M. Pilar Rodríguez del Tío

Junio de 2023

Agradecimientos

Quiero agradecer el apoyo recibido por mi tutora, María Pilar Rodríguez del Tío por su paciencia y excelente orientación, a Ángel Agúndez desde el Ayuntamiento de Valladolid por las contribuciones realizadas en este trabajo y a Miguel A. Fernández Temprano por la ayuda recibida en el desarrollo de la metodología.

También a mi familia, sobre todo a mi tía Isabel por su apoyo incondicional, y a mis amigos por los ánimos y ayuda recibida.

Índice General

Resumen	8
Abstract.....	8
1. Introducción.....	1
2. Contexto	2
2.1. Análisis de correspondencias	2
2.1.1. Simple.....	2
2.1.2. Múltiple	3
2.1.3. Representaciones gráficas	4
2.2. Análisis clúster	5
2.2.1. Métodos jerárquicos	5
2.2.2. Métodos no jerárquicos	6
3. Obtención y tratamiento de datos	7
3.1. Datos de resultados electorales.....	7
3.2 Datos cartográficos	8
4. Aplicación a los resultados electorales de Valladolid en elecciones al Congreso.....	9
4.1. Descripción de algunos resultados utilizando análisis de correspondencias.....	11
4.2. Análisis del comportamiento de las secciones censales utilizando análisis de correspondencias múltiples.....	17
4.3. Análisis del comportamiento de las secciones censales en cada convocatoria electoral.	23
4.3.1. Elecciones 24 de mayo de 2015	25
4.3.2. Elecciones 28 de junio de 1982.....	30
4.3.3. Elecciones 22 de junio de 1986.....	31
4.3.4. Elecciones 29 de octubre de 1989	32
4.3.5. Elecciones 6 de junio de 1993.....	33
4.3.6. Elecciones 3 de marzo de 1996	34
4.3.7. Elecciones 12 de marzo de 2000	34
4.3.8. Elecciones 14 de marzo de 2004	35
4.3.9. Elecciones 9 de marzo de 2008	36
4.3.10. Elecciones 20 de noviembre de 2011	37
4.3.11. Elecciones 3 de mayo de 2016	38

4.3.12. Elecciones 28 de abril de 2019.....	39
4.3.13. Elecciones 10 de noviembre de 2019.....	40
4.4. Comparación entre los resultados obtenidos en el análisis de correspondencias múltiples y los análisis de correspondencias simples por convocatoria.....	43
5. Resumen de los resultados.....	45
6. Conclusiones.....	48
Bibliografía.....	49
Anexo I.....	51
Anexo II.....	53
Anexo III	55
Anexo IV	57
Anexo V. Aplicación de análisis clúster jerárquico.	59

Índice de figuras

4.1 Mapa asimétrico con coordenadas principales los perfiles fila de las convocatorias frente a las tendencias políticas	12
4.2 Mapa simétrico de las convocatorias frente a las tendencias políticas.....	13
4.3 Mapa asimétrico con coordenadas principales los perfiles fila de las secciones frente a las tendencias políticas.....	14
4.4 Mapa simétrico de las secciones frente a las tendencias políticas.....	14
4.5 Mapa de la ciudad de Valladolid con las secciones más radicalizadas	16
4.6 Mapa simétrico del análisis de correspondencias múltiples para todas las convocatorias	18
4.7 Mapa simétrico del análisis de correspondencias múltiples para las últimas convocatorias	19
4.8 Mapa simétrico del análisis de correspondencias múltiples en frecuencias relativas para las últimas convocatorias	21
4.9 Mapa de la ciudad de Valladolid con las secciones más representativas del año 2015	29
4.10 Mapa de la ciudad de Valladolid con las secciones más representativas de la segunda convocatoria del año 2019	42
5.1 Mapa de la ciudad de Valladolid con las secciones más representativas	47

Índice de tablas

4.1 Promedio de los votos de las convocatorias en cada candidatura	15
4.2 Secciones radicalizadas de la ciudad de Valladolid	16
4.3 Secciones más representativas en el análisis de correspondencias múltiples para todas las convocatorias.....	18
4.4 Secciones más representativas en el análisis de correspondencias múltiples para las últimas cinco convocatorias.....	20
4.5 Secciones más representativas en el análisis de correspondencias múltiples en frecuencias relativas para las últimas cinco convocatorias	21
4.6 Secciones más representativas en el análisis de correspondencias múltiples en frecuencias relativas aumentando el radio de la circunferencia para las últimas cinco convocatorias	22
4.7 Valores medios del coeficiente de agregación para las cuatro metodologías estudiadas	24
4.8 Secciones representativas en la aplicación de análisis de correspondencias simple en la convocatoria de 2015	25
4.9 Secciones representativas en la aplicación de análisis de clúster no jerárquico en la convocatoria de 2015.....	26
4.10 Secciones comunes entre el análisis de correspondencias y clúster en la convocatoria de 2015	26
4.11 Secciones representativas del análisis de correspondencias en el segundo escenario para la convocatoria de 2015	27
4.12 Secciones representativas en el segundo escenario para la convocatoria de 2015 .	27
4.13 Secciones representativas del análisis de correspondencias en el tercer escenario para la convocatoria de 2015.....	28
4.14 Secciones representativas del análisis de clúster en el tercer escenario para la convocatoria de 2015.....	28
4.15 Secciones comunes entre el análisis de correspondencias y clúster para el tercer escenario en la convocatoria de 2015	28
4.16 Secciones representativas en la aplicación de análisis de correspondencias y clúster para la convocatoria de 1982	30
4.17 Secciones representativas en la aplicación de análisis de correspondencias y clúster para la convocatoria de 1986.....	31
4.18 Secciones representativas en la aplicación de análisis de correspondencias y clúster para la convocatoria de 1989	32

4.19 Secciones representativas en la aplicación de análisis de correspondencias y clúster para la convocatoria de 1993	33
4.20 Secciones representativas en la aplicación de análisis de correspondencias y clúster para la convocatoria de 1996	34
4.21 Secciones representativas en la aplicación de análisis de correspondencias y clúster para la convocatoria de 2000	34
4.22 Secciones representativas en la aplicación de análisis de correspondencias y clúster para la convocatoria de 2004	35
4.23 Secciones representativas en la aplicación de análisis de correspondencias y clúster para la convocatoria de 2008	36
4.24 Secciones representativas en la aplicación de análisis de correspondencias y clúster para la convocatoria de 2011	37
4.25 Secciones representativas en la aplicación de análisis de correspondencias y clúster para la convocatoria de 2016	38
4.26 Secciones representativas en la aplicación de análisis de correspondencias y clúster para la primera convocatoria de 2019.....	39
4.27 Secciones representativas en la aplicación de análisis de correspondencias simple en la segunda convocatoria de 2019.....	40
4.28 Secciones representativas en la aplicación de análisis de clúster no jerárquico en la segunda convocatoria de 2019.....	40
4.29 Secciones representativas en la aplicación de análisis de correspondencias y clúster para la segunda convocatoria de 2019	41
4.30 Tabla resumen con las secciones más representativas en las últimas cinco convocatorias	43
5.1 Tabla resumen con las secciones más representativas según los cuatro estudios realizados.....	46
A5.1 Tabla resumen con las secciones obtenidas en el análisis jerárquico para todas las convocatorias	59

Resumen

La finalidad del presente estudio es caracterizar las secciones censales más representativas, respecto de la distribución del voto de Valladolid en las elecciones al Congreso de los Diputados. Para llevarlo a cabo nos hemos basado en los resultados electorales de las elecciones generales celebradas desde 1982 a 2019, empleando metodologías tales como el análisis de correspondencias y el análisis clúster, implementados con el lenguaje de programación R, y comparando los resultados obtenidos, con un estudio análogo para las elecciones municipales, realizado en el año 2021.

Palabras clave

Secciones censales, elecciones generales, análisis de correspondencias, análisis clúster, clasificación, Valladolid.

Abstract

This memoir aims to obtain the most representative census tracts according to the distribution of votes in Valladolid for the elections to the Congress of Deputies. For this purpose, we have used the electoral results of the general elections occurred between 1982 and 2019. The dissertation will be carried out through methodologies such as correspondence analysis and cluster analysis implemented with the R programming language. Results are compared with an analogous study for the municipal elections held in 2021.

Keywords

Census tracts, general elections, correspondence analysis, cluster analysis, classification, Valladolid.

1. Introducción

El estudio del comportamiento de la distribución del voto en las diferentes ciudades ha resultado interesante desde el establecimiento de la democracia en España.

Teniendo en cuenta lo expuesto, creemos interesante definir como objetivo principal de este trabajo, el análisis de la distribución del voto de las secciones censales de Valladolid para determinar cuáles son las más representativas en las elecciones al Congreso de los Diputados. Los datos utilizados para llevarlo a cabo son los resultados electorales asociados a las trece convocatorias de elecciones generales que han tenido lugar en España desde 1982 a 2019.

Los métodos aplicados en el estudio para identificar las secciones serán el análisis de correspondencias aplicado a una serie de tablas de contingencia que contienen los votos de cada sección, otorgados a los diferentes partidos durante las convocatorias señaladas, y un análisis clúster basado en la distribución de los votos en estas mismas y en cada año electoral. El resultado del primer análisis será un conjunto de secciones cercanas al origen que se encuentran dentro de un radio definido, mientras que el segundo análisis permite localizar un conjunto de ellas que pertenecen al mismo grupo que la sección artificial, creada para recoger los porcentajes de votos de Valladolid.

Con el fin de reforzar estos resultados se aplicará una variante de análisis clúster, por medio de la cual se obtendrán las secciones que pertenecían al grupo en el momento de asociación con la sección artificial que contiene la distribución global del voto.

Aunque no sea el objetivo principal del estudio, también haremos notar distintas peculiaridades que se obtienen al aplicar el análisis de correspondencias.

Finalmente, se distinguirán los métodos más interesantes a aplicar en este tipo de análisis y se mostrarán las secciones más representativas del voto en las elecciones al Congreso de los Diputados.

Para acompañar los resultados obtenidos, en los análisis descritos, se muestran mapas cartográficos de la ciudad, resaltando los resultados para que nos permita visualizarlo mejor.

El contenido de la memoria se divide en tres partes. En la primera parte, se describirán los conceptos teóricos de los métodos utilizados. A continuación, se incluyen los datos con los que se va a trabajar. Por último, se aplican las metodologías descritas, obteniendo los resultados buscados.

2. Contexto

El objetivo principal de este trabajo es el análisis de la distribución del voto de las secciones censales de Valladolid para determinar cuáles son las más representativas en las elecciones al Congreso de los Diputados. Para ello, se han utilizado metodologías analíticas tales como el análisis de correspondencias y una clasificación no supervisada. En la Sección 2.1 se describirá la metodología de análisis de correspondencias con sus respectivas variantes. A continuación, en la Sección 2.2 se describirá la metodología de análisis clúster con sus respectivas ramas.

2.1. Análisis de correspondencias

El análisis de correspondencias es una técnica descriptiva multivariante cuya finalidad es estudiar relaciones existentes entre variables categóricas. El mayor interés reside en resumir la información contenida en una tabla de contingencia de dos o más dimensiones. Según el número de dimensiones que tenga la tabla se realizará un análisis de correspondencias simple (dos) o múltiples (tres o más).

Dicho análisis tiene como objetivos estudiar la relación entre categorías de columnas o filas, así como estudiar las relaciones entre categorías de filas y columnas. En otras palabras, se trata de estudiar las asociaciones entre perfiles específicos y las posibles relaciones entre ellas.

2.1.1. Simple

Para estudiar las posibles relaciones entre dos variables categóricas, es necesario definir algunos conceptos acerca de dicho análisis.

El primer concepto son los perfiles fila y columna, que representan distribuciones condicionadas por filas y por columnas. La principal característica de estos perfiles es que no todos tienen el mismo peso, ya que, cada perfil está compuesto por un número diferente de individuos. El mayor interés reside en expresar la distancia entre los perfiles fila y columna en el menor número de dimensiones posible.

Para ello, tal y como apunta [5] se va a cuantificar dicha diferencia entre los perfiles con la distancia X^2 . En este caso, dicha distancia se trata de la distancia euclídea con los perfiles ponderados y tiene la siguiente expresión.

$$d_{\chi^2}^2(i, i') = \sum_{j=1}^c \frac{1}{f_{+j}} \left(\frac{f_{i,j}}{f_{i+}} - \frac{f_{i',j}}{f_{i'+}} \right)^2$$

Dicha distancia tiene como propiedad la equivalencia distribucional, lo que garantiza que al resumir dos categorías del perfil en una no haya pérdida de información.

Además, se pretende que dicho procedimiento represente con la mayor calidad posible la diferencia entre los perfiles, para lo que definimos como medida de variabilidad, la inercia total. Esta medida evaluará la calidad de la representación de los ejes, de manera que, cuanto mayor sea la variabilidad recogida por las dos primeras componentes, mejor será la representación del análisis.

En la aplicación del análisis de correspondencias, calcularemos la proyección óptima en un espacio de bajas dimensiones, es decir, minimizaremos la distancia entre la nube de puntos correspondiente y su centro de gravedad. De esta manera la nube de puntos quedará proyectada sobre el origen de coordenadas y por ello el perfil medio de cada dimensión se situará en este mismo punto [5].

Por último, se estudiará las contribuciones absolutas y relativas a la inercia para poder identificar la aportación de cada perfil a la formación de un eje y la calidad con la que el perfil está siendo representado. La contribución absoluta mide la importancia de cada categoría de la variable analizada en la construcción de cada eje. Mientras que la contribución relativa refleja la relevancia de cada categoría en la explicación de la posición de la variable.

2.1.2. Múltiple

Constituye una generalización del ACS al caso multidimensional. Por ello, en este análisis definiremos P perfiles de los que pretendemos estudiar las posibles relaciones entre ellos.

Para su aplicación se plantearán dos enfoques esencialmente. El primero de ellos consiste en un análisis de correspondencias sobre la matriz binaria de los datos. Dicha matriz M tendrá definidos N individuos o grupos y Q variables donde cada variable tendrá asociada una serie de categorías. Así, cada celda es excluyente en el sentido de, un individuo solo puede tomar el valor 1 en una de las categorías y 0 en el resto [4].

Además, la inercia en este caso solo depende del número de variables y no de sus valores. Por todo esto, la diferencia en este caso reside en la interpretación de las relaciones entre las variables, de forma que solo se interpretan las posiciones de los puntos y sus proximidades.

Otro posible enfoque es utilizar la denominada la matriz Burt. Dicha matriz está compuesta por todos los posibles cruces dos a dos de las variables a estudiar. En este segundo método la inercia se ve aumentada, puesto que se trata de un promedio de las matrices cruzadas dos a dos.

En la comparación de las dos técnicas, los resultados que se obtienen son bastante similares, con el detalle de que en el segundo método la inercia siempre es mayor, por ser esta el promedio de inercias de las respectivas matrices con dos dimensiones.

2.1.3. Representaciones gráficas

Las representaciones gráficas de las proyecciones de los perfiles pueden hacerse diferentes formas. Las más comunes son mapas simétricos, asimétricos y biplots.

En los gráficos asimétricos se representará un perfil en las coordenadas principales y otro en coordenadas estandarizadas. De esta forma, el perfil que esté en coordenadas principales representará la distancia entre sus puntos como distancias aproximadas X^2 , mientras que, el otro perfil tendrá proyectado sus puntos en el espacio del primero. El motivo de actuar, así como apuntaba [5] es buscar relaciones entre las categorías de fila y columna. De forma que, se permite interpretar intuitivamente las distancias entre los puntos de fila y los puntos de columna, sobre todo cuando hay una gran cantidad de inercia recogida por los ejes.

En los gráficos simétricos se representarán ambos perfiles en coordenadas principales . Así, las distancias entre los respectivos puntos de cada perfil son distancias X^2 aproximadas. Es por esto que como apunta [5] en este tipo de representaciones podrán interpretarse tanto las distancias entre los perfiles fila como las distancias entre los perfiles columna. Sin embargo, no debe interpretarse la asociación entre las filas y las columnas. También, en este tipo de representaciones se debe tener en cuenta que aquellos puntos que están más lejos del origen representan las categorías más influyentes.

En los biplots se representarán uno de los perfiles en coordenadas principales. Aunque es cierto que, no se trata de una representación realizada en este estudio vamos a comentar puntos comunes e interesantes. En este caso, se describirán los biplots estándares que muestra uno de los perfiles en coordenadas principales, al mismo tiempo que acercan los vértices del otro perfil, multiplicando por la raíz cuadrado de sus masas, para mejorar la representación conjunta [4]. No debe interpretarse en este tipo de representaciones las distancias de los puntos en coordenadas estándares, pero sí las proyecciones de los puntos representados en coordenadas principales sobre estos ejes que son indicativas de la distancia entre esos puntos y la clase correspondiente del vértice en coordenadas estándares [5].

2.2. Análisis clúster

El análisis clúster es un conjunto de técnicas multivariantes utilizadas para agrupar unos individuos en una serie de grupos homogéneos. El objetivo principal consiste en encontrar particiones óptimas de conjuntos, las cuales garanticen que los individuos pertenecientes a un grupo sean tan similares entre sí como sea posible, y los distintos grupos tan diferentes [14].

Para ello es importante una buena cuantificación de los criterios de similitud y divergencia entre las observaciones. La medida encargada de definir dichos criterios se denomina índice de disimilaridad y puede ser calculada de numerosas maneras. En nuestro caso hemos utilizado la distancia de Kullback-Leibler, puesto que, se trata de una medida robusta y apropiada para nuestro estudio por no ser sensible a valores atípicos.

También, es necesario definir un criterio que mida la distancia entre grupos o individuos a lo largo del análisis que se denominará índice de agregación. Dicho índice, puede ser calculado de las siguientes cuatro maneras principalmente:

1. Single linkage: Considera la distancia entre clústeres como la distancia entre los individuos más próximos.
2. Complete linkage: Considerando la distancia entre clústeres como la distancia entre los individuos más alejados.
3. Average linkage: Considera la distancia entre un clúster y otro como el promedio desde cualquier miembro del clúster a cualquier otro.
4. Método de Ward: Este método minimiza la varianza total dentro del grupo. En cada paso, se unirán los dos clústeres / individuos con la menor suma de cuadrados de error.

Según la manera de clasificación de este análisis podemos definir dos métodos esencialmente, jerárquicos y no jerárquicos.

2.2.1. Métodos jerárquicos

Los individuos no se dividen en clústeres de una sola vez, sino que se realizan particiones sucesivas a distintos niveles de agrupamiento. De manera que, inicialmente, definimos una solución inicial con un grupo por cada observación, para continuar, seleccionamos los dos grupos con menor índice de agrupamiento y finalmente el algoritmo termina cuando todas las observaciones se localizan en un grupo. El número de grupos es una decisión subjetiva y dependerá de la finalidad del análisis. Una representación clásica de este tipo de análisis es el dendrograma, que constituye una representación esquemática de la aplicación de este algoritmo.

2.2.2. Métodos no jerárquicos

Los individuos en este caso se dividen inicialmente en un número de grupos prefijado, de manera que, aunque estos individuos puedan cambiar de grupo, el número de grupos no se modificaría. Para su aplicación definimos como número de grupos aquel que nos interese según el objetivo del análisis.

Los pasos que siguen estos tipos de algoritmos son los siguientes:

1. Definimos una partición inicial de las observaciones en los k grupos.
2. Calculamos el índice de agrupación al mover un individuo de su grupo a otro.
3. Realizamos el cambio óptimo.

En este campo se utilizará como algoritmo el de *k-medianas*, que tiene como objetivo encontrar la ubicación del centroide que optimice la distancia definida entre el centroide de un grupo y los puntos asignados a ese grupo.

3. Obtención y tratamiento de datos

La elaboración de este trabajo ha requerido de dos bases distintas de datos, la formada por los resultados electorales de Valladolid en las elecciones generales desde 1982 hasta 2019, y la constituida por los datos cartográficos más recientes empleados en la elaboración de mapas.

3.1. Datos de resultados electorales

Estos datos se han tomado de la página web del Ministerio del interior, dentro del subapartado de información electoral, en el que se ubica el apartado “área de descargas” y en el que se visualiza un menú “convocatorias” con un desplegable que nos permite seleccionar la pestaña Congreso y posteriormente ver las convocatorias con los respectivos años a los que corresponden. Seleccionamos los años desde 1982 a 2019 [1] y descargamos únicamente el archivo “Mesa” de cada una de ellas. Tras la descarga de estas carpetas, se observa que contienen una serie de datos de los que solo necesitamos una parte. Es necesario, por tanto, seleccionar (por ser este el más completo), el archivo 10 de cada año electoral, nombrado como “Fichero de DATOS DE CANDIDATURAS DE MESAS y del C.E.R.A.” donde se presentan las votaciones en las distintas mesas de los distritos y clasificadas por candidaturas.

Posteriormente, realizamos un filtrado a través del software R que nos permitió trabajar con las secciones censales únicamente de la ciudad de Valladolid. A continuación, realizamos la decodificación de los datos que contaban con doce campos por convocatoria, creando así los archivos que contengan el conjunto de votos asociados a cada candidatura, en las secciones correspondientes para poder trabajar con más comodidad.

Puesto que la pretensión es analizar las convocatorias de forma independiente y también conjunta, los partidos se agruparon según su tendencia política, tal y como se muestra en el Anexo I. Con este resumen, se obtuvo una base de datos constituida por los años electorales, las tendencias políticas, las secciones censales y los votos asociados a cada una de estas combinaciones de variables. A partir de dicha base de datos resultará más sencillo la construcción de tablas de contingencia para los análisis a desarrollar, teniendo en cuenta que en segundo análisis será necesario definir tablas de contingencia en frecuencias relativas, donde cada fila, representa la distribución del voto de cada sección en tantos por 1.

3.2 Datos cartográficos

En este caso, los datos espaciales han sido obtenidos mediante la página web del INE en el apartado de “Productos y servicios “[2]. Dichos archivos cuentan con la información de todas las provincias y de todas las comunidades pertenecientes a España, por lo cual es necesario una selección de las filas únicamente de este municipio, a través del software R. Estos datos se encuentran en formato *(.shp)* y contienen los contornos georreferenciados de las secciones censales de toda España para todos los años desde el 2001.

Dado que las secciones censales de Valladolid han ido fluctuando a lo largo de los años, teniendo en algunas convocatorias hasta 270 secciones, utilizamos el año 2019 como base para la representación de los resultados genéricos, por ser la convocatoria más reciente.

4. Aplicación a los resultados electorales de Valladolid en elecciones al Congreso

Aplicamos las técnicas de análisis de correspondencias, tanto simple como múltiple y análisis clúster jerárquico y no jerárquico explicadas anteriormente para las trece convocatorias de elecciones generales celebradas desde 1982 a 2019 en España, centrándonos en los resultados electorales en la ciudad de Valladolid.

Con el objetivo de encontrar cuáles serían las secciones más representativas de la ciudad, se va a realizar un estudio sobre los resultados electorales de cada convocatoria, inicialmente de manera conjunta y posteriormente de forma independiente. De esta forma, se descubrirá que secciones son las que cumplen nuestro objetivo y cuál es la metodología que mejores resultados obtiene.

Para lograr este propósito, la primera tarea consiste en establecer las variables de las que se quiere estudiar el tipo de relación. Para ello vamos a analizar los denominados perfiles fila y columna, que no son más que distribuciones condicionadas de la variable fila (en este caso, las secciones) y la variable columna (en este caso, los partidos). En el análisis conjunto, esto resulta más complejo, definiendo una tabla tridimensional en la que N secciones votan a los Q partidos en cada una de las convocatorias. El análisis conjunto requiere, por tanto, definir que N y Q deben ser iguales para todas las convocatorias y, por ello, es necesario quedarnos solo con las secciones comunes (doscientas, en este caso) y agrupar los partidos políticos en tendencias (Anexo I).

En las otras técnicas empleadas, para conseguir nuestro objetivo de encontrar las secciones más representativas, realizamos un estudio por año electoral de manera independiente. Para lo cual vamos a crear una sección artificial que contenga la distribución global de voto, de forma que, tras aplicar un análisis clúster jerárquico, se identificarán las secciones que pertenecen al clúster a la hora de anexionarse con la sección artificial, mientras que, para el análisis clúster no jerárquico se identificarán aquellas secciones que pertenezcan al mismo grupo que la sección con la distribución del voto global.

Por ello inicialmente vamos a definir un índice de disimilaridad encargado de cuantificar las diferencias entre las distribuciones del voto de las distintas secciones censales. En este caso hemos elegido la distancia de Kullback-Leibler, ya que es una medida apropiada y robusta, cuya expresión asociada es la siguiente.

$$D_{KL}(g, h) = \sum_{i=1}^n P(i) \cdot \ln \left(\frac{P(i)}{Q(i)} \right)$$

Donde P y Q son distribuciones de probabilidad de una variable aleatoria que toma N valores, en este caso el número de candidaturas que se presentaron en la convocatoria.

El paquete *seewave* en R contiene la función *kl.dist()* [15] que nos servirá para cuantificar las diferencias entre cada par de individuos. A través de esta función, se construirá la matriz de disimilaridades.

Para el análisis clúster no jerárquico es necesario, además, elegir el número de clúster adecuado que cumpla el objetivo de localizar un número moderado de secciones en el mismo clúster que la sección objetivo.

Tras aplicar ambos análisis a cada convocatoria, se comparan los resultados obtenidos y se extraerán las secciones comunes en todos los casos. De este modo, aquellas secciones que fueron representativas en ambos análisis serían candidatas para cumplir nuestro objetivo. Puesto que, el análisis clúster jerárquico es una herramienta de analítica que obtiene resultados similares, su estudio se realizará para ratificar los resultados obtenidos.

Todo esto irá acompañado por mapas cartográficos de la ciudad de Valladolid, destacando las secciones representativas en cada momento.

4.1. Descripción de algunos resultados utilizando análisis de correspondencias.

El propósito de este apartado es analizar de manera conjunta las trece convocatorias. De forma que, aplicando análisis de correspondencias, realizaremos posibles estudios sobre los resultados electorales que nos aportarán diferentes informaciones acerca de ellos.

Es importante señalar que respecto al perfil secciones, las categorías se definen como combinaciones de cuatro dígitos, de los cuales los dos primeros representan el distrito y los dos siguientes la numeración de las secciones en cada distrito. También hay que destacar que para este primer análisis los cruces entre perfiles serán dos a dos en un análisis de correspondencias simple.

Con este fin, inicialmente los perfiles a cruzar fueron las convocatorias con las tendencias políticas. Para representar este análisis, primero se obtendrá un mapa asimétrico cuyas coordenadas principales serán las convocatorias, que servirá para interpretar las diferencias entre las categorías del perfil fila y el grado de asociación entre ambos perfiles, y posteriormente, se representará un mapa simétrico, que nos permitirá estudiar las distancias entre los perfiles fila, así como las distancias entre los perfiles columna, sin embargo, la distancia entre filas y columnas no debe interpretarse directamente en esta representación.

El segundo análisis que se elaborará tratará de un cruce entre las secciones y las tendencias políticas. En este tipo de análisis, se realizarán las mismas representaciones que anteriormente para estudiar tanto el grado de asociación entre las secciones y las tendencias políticas, como las posibles diferencias entre cada uno de los perfiles.

Además, es necesario que en este tipo de análisis se tenga en cuenta la inercia que recogen los ejes principales, ya que, si esta es baja se subestimarán la calidad de la representación obtenida.

Con todos los aspectos indicados se aplicará dicho análisis utilizando el paquete *ca* de R que nos proporciona la función *ca()* [16] para el análisis de correspondencias simple .

Dicha función recibe tablas de contingencia con frecuencias absolutas, teniendo en cuenta que el peso de cada individuo es distinto. En el análisis por convocatorias se planteará otro escenario en el cual las tablas de contingencia contendrán frecuencias relativas en porcentaje, de esta manera, todos los perfiles fila tendrán la misma masa. Esto nos servirá para descubrir si aplicando el mismo análisis con todas las secciones en igualdad de condiciones, es decir, sin tener en cuenta el número de votantes, se obtienen resultados similares.

Aplicación

En este primer análisis se cruzan las variables convocatorias y tendencias políticas. Aplicando un análisis de correspondencias simple en la figura 4.1 se muestra un mapa asimétrico donde las coordenadas principales son las convocatorias y las coordenadas estándares las tendencias políticas.

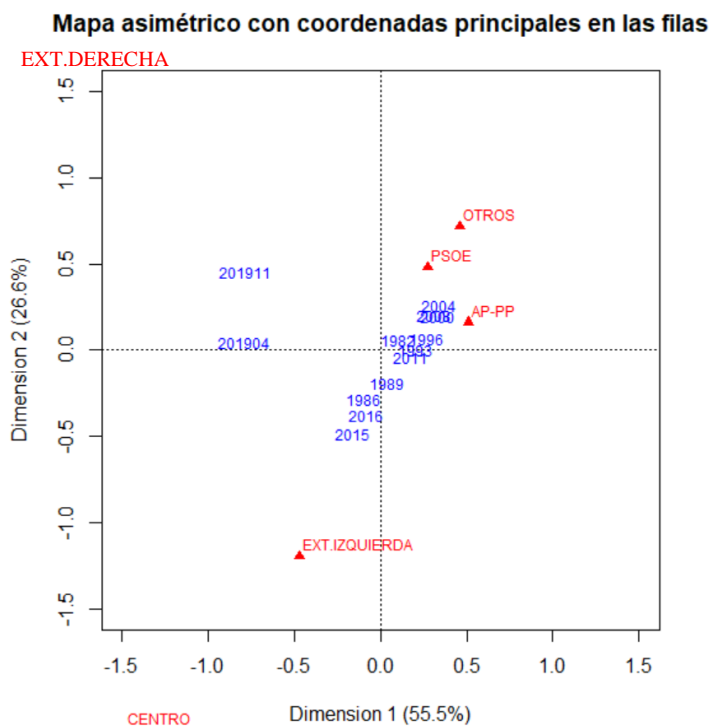


Figura 4.1 Mapa asimétrico con coordenadas principales los perfiles fila de las convocatorias frente a las tendencias políticas

En este gráfico se observa que el grado de asociación entre las convocatorias y las tendencias políticas no es muy alto. Destacando que las últimas convocatorias han tendido a la radicalización, puesto que en 2015 la candidatura de EXT.IZQUIERDA obtuvo un 22% frente a un 11% como promedio de todas las convocatorias, de igual forma que lo sucedido en la convocatoria de abril de 2019 donde la candidatura de EXT.DERECHA obtendría un 13% frente al 2% obtenido como promedio de todas las convocatorias. Además, los vértices situados más hacia la derecha son las categorías de mayor peso, un hecho que resulta verificado, puesto que se tratan de las convocatorias con mayor número de votos.

También, se refleja que, con el avance de las convocatorias, las categorías de los perfiles tendieron a separarse, principalmente por el hecho de que el voto tendió a dispersarse, perdiendo significación la tendencia bipartidista.

Finalmente, hay que comentar que la calidad de la representación es buena, con una inercia total recogida por los ejes de un 82.6%.

En la figura 4.2, se muestra el gráfico simétrico para poder estudiar de manera más visual las diferencias entre las modalidades de los perfiles.

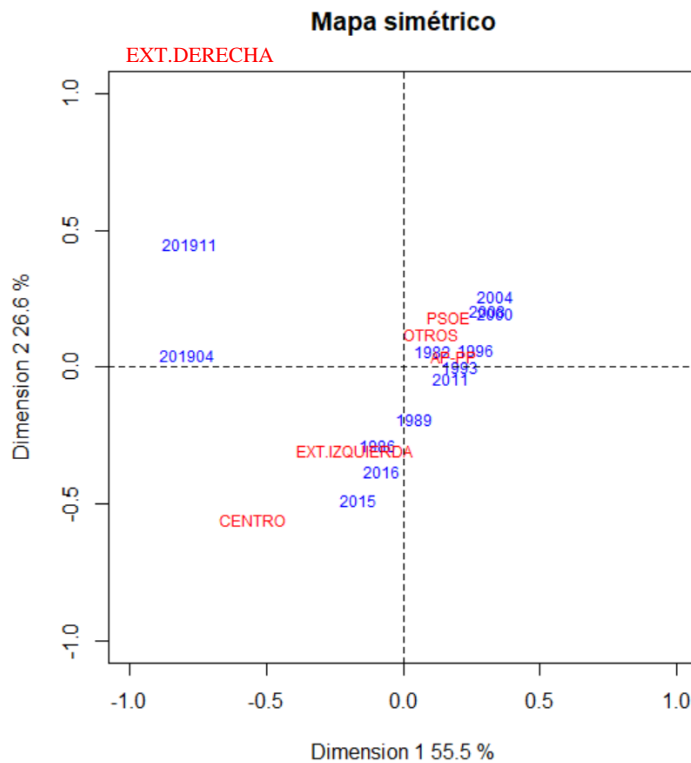


Figura 4.2 Mapa simétrico de las convocatorias frente a las tendencias políticas

De igual forma que en el anterior análisis, la convocatoria de noviembre de 2019 es la más alejada, reflejándose como punto influyente del análisis. Este hecho se probará posteriormente realizando una analítica mucho más detallada, ya que se trata de una convocatoria especial (Sección 4.3.13).

También se observa que los partidos más situados a la derecha resultan ser los más votados, teniendo en cuenta que estos serían el PP y el PSOE, que son los partidos con mayor cantidad de votos en España. Los más alejados son el CENTRO y la EXT.DERECHA, que en el perfil medio son los que contienen el menor número de votos.

En otra perspectiva del análisis se comentará como es la tendencia política de las secciones de los diferentes distritos comunes en todos los años. Hay que recordar que en este caso el perfil de las secciones tiene identificadas sus categorías por cuatro dígitos, los dos primeros corresponden al distrito y los dos siguientes a la numeración de las secciones de cada distrito. Para la aplicación de este análisis, primeramente, se obtendrá un mapa asimétrico de los perfiles fila, es decir, las secciones. En la figura 4.3, se podrá estudiar cuál será el grado de asociación entre los dos perfiles.

Mapa asimétrico con coordenadas principales las filas

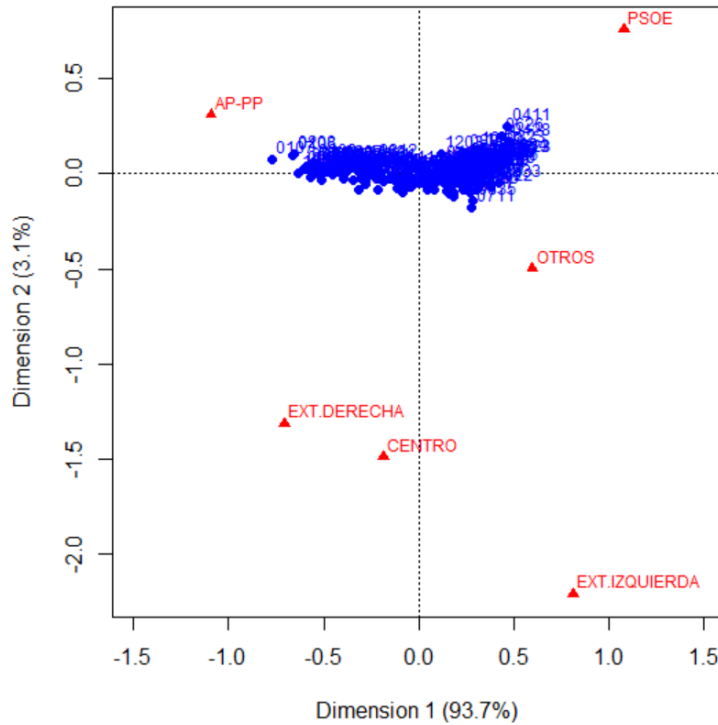


Figura 4.3 Mapa asimétrico con coordenadas principales los perfiles fila de las secciones frente a las tendencias políticas

En este gráfico se ve reflejado, que el grado de asociación entre las secciones y las tendencias políticas es muy bajo. Además, la calidad de la representación es buena, con una inercia total recogida por los ejes de un 96,8%. En la figura 4.4 se representa el mapa simétrico asociado a dicho estudio en el cual se comparan las diferencias entre las categorías de los perfiles.

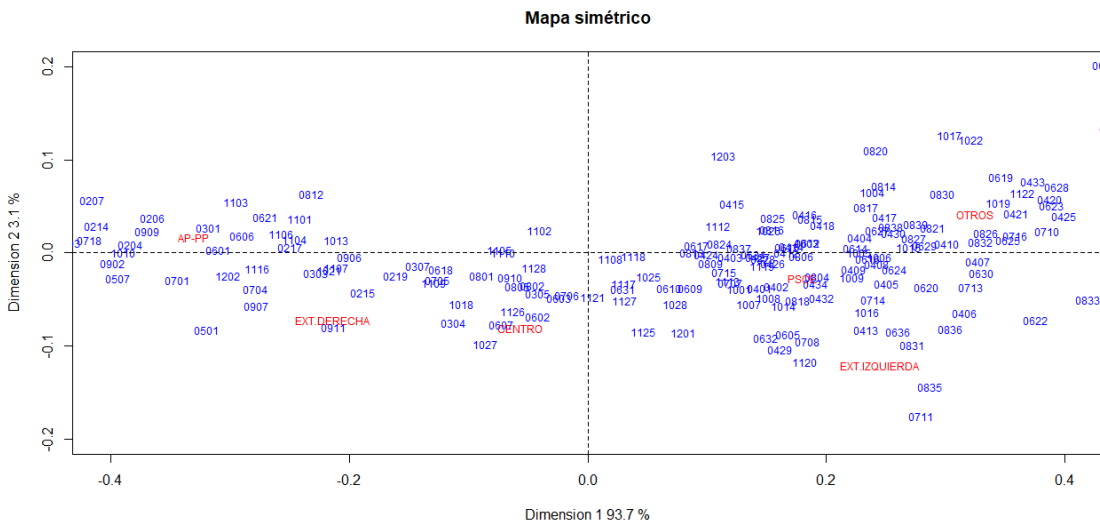


Figura 4.4 Mapa simétrico de las secciones frente a las tendencias políticas

Para el análisis se han seleccionado solamente las secciones comunes a lo largo de los años, ya que algunas secciones han ido apareciendo o desapareciendo, por lo que no existe ningún punto de comparación en la línea temporal.

En este análisis, se observa que existen puntos atípicos, como pueden ser la sección 7 del distrito 1 (no representada en el gráfico por excederse de los límites), con un promedio de votos al PP del 78%, frente a un 41% que se obtendría en promedio de todas las secciones. De igual forma, detectamos que la sección 11 del distrito 4 (no representada en el gráfico por excederse de los límites) obtiene un promedio de votos al PSOE del 62% frente a un 35% que se obtendría en promedio de todas las secciones. Es claro destacar que, a pesar de que en el mapa asimétrico no se detectará asociación entre los perfiles, existirán secciones que serán representativas de algún partido político. A mayores, las diferencias en las tendencias políticas son claras, situándose el CENTRO lo más próximo al origen de coordenadas y las ideologías políticas opuestas a diferentes lados del eje principal.

En la tabla 4.1, se obtiene el promedio de los votos de las respectivas agrupaciones de candidaturas en las convocatorias:

	AP-PP	PSOE	EXT.IZQUIERDA	CENTRO	EXT.DERECHA	OTROS
% de votos	41.25	35.88	10.76	8.44	2.19	1.46

Tabla 4.1 Promedio de los votos de las convocatorias en cada candidatura

Con el resultado de dicho análisis se representará en la figura 4.5 un mapa de la ciudad con las secciones comunes, en el cual destacamos aquellas secciones cuyo porcentaje de votos a una determinada agrupación de candidaturas aumente en un 70% el porcentaje promedio de voto a esa agrupación en Valladolid.

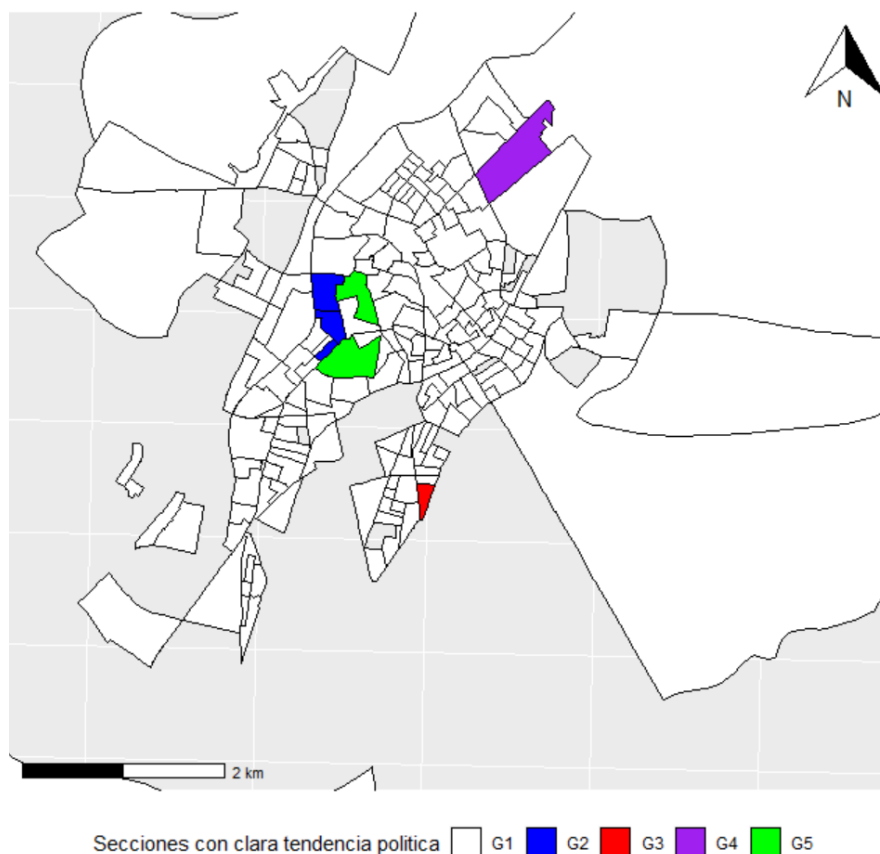


Figura 4.5 Mapa de la ciudad de Valladolid con las secciones más radicalizadas

De manera que, las secciones pertenecientes al grupo 2, en azul, son aquellas que obtuvieron al menos un 70% de los votos a la agrupación de candidaturas AP-PP, consideramos a dichas secciones como las más representativas de esa opción política. De igual modo, el grupo 3 está formado por aquellas secciones que obtuvieron más del 61% de los votos para la opción PSOE, el grupo 4 lo forman aquellas secciones que obtuvieron más del 18% de los votos para la opción de EXT.IZQUIERDA y el grupo 5 aquellas secciones que obtuvieron más del 4% de los votos para la opción de EXT.DERECHA.

En la tabla 4.2, se obtiene las secciones pertenecientes a los grupos G2, G3, G4 y G5:

Distrito	Sección	Lugar	Grupo	Tendencia
01	01	Centro-Plaza Mayor	G5	EXT.DERECHA
01	06	Centro-Doctrinos	G2	AP-PP
01	07	Centro-Doctrinos	G2	AP-PP
02	02	Centro-Acera Recoletos	G5	EXT.DERECHA
02	08	Centro- Plaza Tenerías	G2	AP-PP
04	11	Delicias – Las Viudas	G3	PSOE
08	35	San Pedro Regalado	G4	EXT.IZQUIERDA

Tabla 4.2 Secciones radicalizadas de la ciudad de Valladolid

4.2. Análisis del comportamiento de las secciones censales utilizando análisis de correspondencias múltiples

El objetivo de este apartado es estudiar el comportamiento de las secciones censales localizando aquellas que resultan más representativas del voto global de la ciudad. Para su aplicación, se definirá una circunferencia centrada en el origen y de radio, el adecuado para conseguir un número moderado de secciones. Por todo ello, dicho radio ha sido seleccionado para cada caso concreto.

De manera que, en dicho estudio, se ha realizado un análisis de correspondencias múltiples cruzando las tres variables. Para la representación de dicho análisis se obtendrá un mapa simétrico donde sólo los perfiles fila, en este caso las secciones, aparecerán representados. Así aquellas que se localicen más próximas al centro de gravedad serán las más representativas del análisis.

Además, es necesario tener en cuenta, igual que en el anterior análisis, que la inercia que recogen los ejes principales sea la mayor posible en cada caso.

Con todos los aspectos indicados se aplicará dicho análisis utilizando el paquete *ca* de R que nos proporciona la función *mjca()* [17] para el análisis de correspondencias múltiples.

Aplicación

A continuación, vamos a realizar un análisis de correspondencias múltiples a tres dimensiones: año, partido y sección. En un primer análisis se incluyen todas las convocatorias y se seleccionan únicamente las secciones comunes para todos los años. Para cubrir nuestro objetivo utilizamos una circunferencia centrada en el origen y con radio 0.06. En el mapa simétrico de la figura 4.6, la circunferencia representa los límites para definir qué secciones son las más similares a la distribución global del voto. Debido a la heterogeneidad entre las convocatorias que se ve reflejada en el primer gráfico (Figura 4.1), los resultados serán poco esclarecedores.

Además, debido a la condición de quedarnos solo con las secciones comunes, doscientas en este caso, se produce una clara pérdida de información, ya que en estudios posteriores veremos que existen secciones no comunes a todos los años que cubren nuestro objetivo.

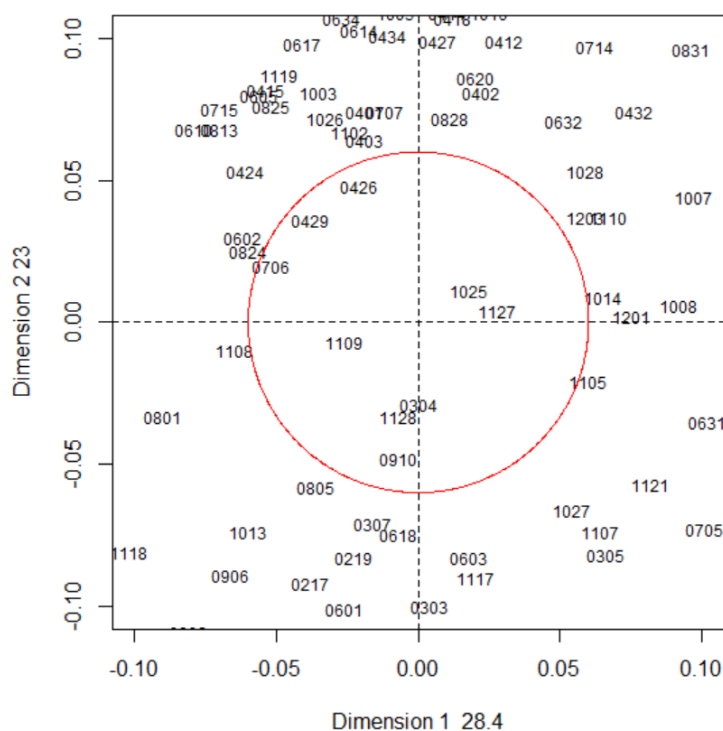


Figura 4.6 Mapa simétrico del análisis de correspondencias múltiples para todas las convocatorias

En este análisis el resultado obtuvo una calidad de la representación bastante baja, con una inercia total recogida por los ejes del 51.4%, esto puede deberse a la heterogeneidad de las convocatorias y a su carácter bidimensional.

En la tabla 4.3 mostramos las nueve secciones localizadas en el interior de dicho círculo:

Distrito	Secciones
3	4
4	26,29
7	6
9	10
10	25
11	9,27,28

Tabla 4.3 Secciones más representativas en el análisis de correspondencias múltiples para todas las convocatorias

En este estudio se observa que la sección más próxima al origen de coordenadas y por ello la más similar al perfil medio es la sección 25 del distrito 10, otra sección también interesante por su proximidad sería la sección 27 del distrito 11. Ambas secciones serán relevantes en los posteriores análisis realizados.

Una vez observado el resultado obtenido se decide realizar un análisis de correspondencias múltiples, con convocatorias más homogéneas, para ello se seleccionan las convocatorias acontecidas desde 2008 hasta la primera convocatoria de 2019. Descartamos la segunda convocatoria de 2019, ya que como se había concluido en el primer análisis, se trataba de un punto influyente. Esta selección se debe a dos factores, el primero es que al realizar la comparación con todas las convocatorias se obtenían doscientas secciones comunes, mientras que ahora se ha aumentado a doscientas cuarenta y nueve, sin eliminar secciones que podrían ser representativas y el segundo factor está fundamentado en la similitud de los resultados electorales obtenidos durante esos años, con una clara segregación del voto en las diferentes tendencias. Aplicando dicho análisis vamos a representar en el mapa simétrico de la figura 4.7, donde se incluye una circunferencia centrada en el origen y con radio 0.03 para definir cuáles de las secciones son las más similares a la distribución global del voto.

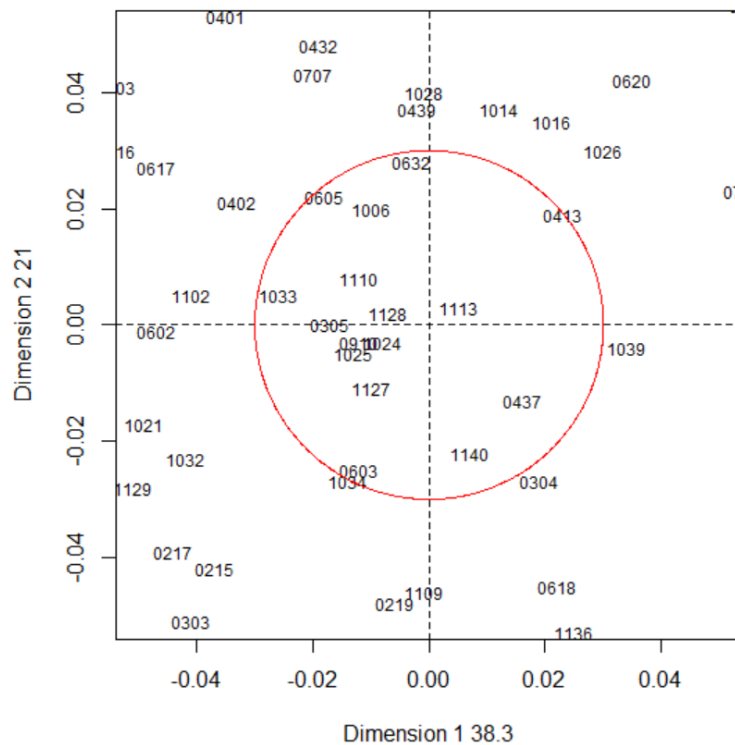


Figura 4.7 Mapa simétrico del análisis de correspondencias múltiples para las últimas convocatorias

En la tabla 4.4 mostramos las dieciséis secciones localizadas en el interior de dicho círculo:

Distrito	Secciones
3	5
4	13,37
6	3,5,32
9	10
10	6,25,28,33
11	10,13,27,28,40

Tabla 4.4 Secciones más representativas en el análisis de correspondencias múltiples para las últimas cinco convocatorias

En este segundo análisis podemos observar que la calidad de la representación ha mejorado teniendo un 60% de inercia total recogida por los ejes principales. Además de lo anterior se observa un resultado mucho más claro con un número de secciones próximas al origen, lo que produce que el objetivo se cumpla con un mayor rigor. La sección más próxima al origen ahora sería la 13 del distrito 11, aunque también están próximas al origen las secciones 25 del distrito 10 y 27 del distrito 11, que destacaban en el análisis anterior.

Finalmente, aplicamos un análisis de correspondencias múltiples en frecuencias relativas para verificar si los resultados obtenidos por el análisis en frecuencias absolutas para estas últimas cinco convocatorias son similares o, por el contrario, algunas de las secciones no resultaron representativas porque contenían un número de votantes significativamente menor. En dicho análisis se elaborará un mapa simétrico donde se incluye una circunferencia centrada en el origen y con radio 0.03. Dicha circunferencia representa la frontera para definir qué secciones son las más similares a la distribución global del voto.

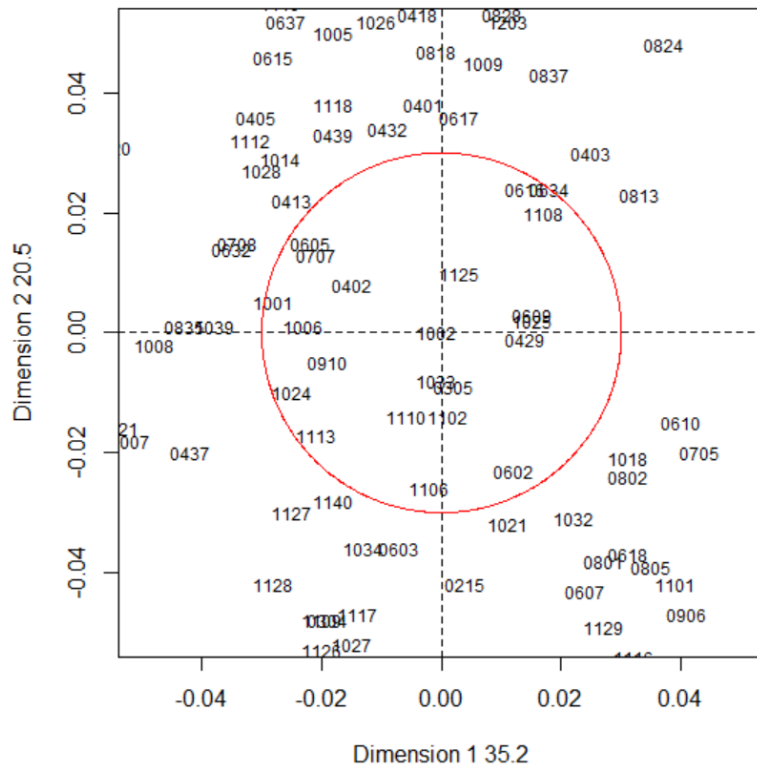


Figura 4.8 Mapa simétrico del análisis de correspondencias múltiples en frecuencias relativas para las últimas convocatorias

En la tabla 4.5 mostramos las veintidós secciones localizadas en el interior de dicha circunferencia:

Distrito	Secciones
3	5
4	2,29
6	2,5,9,16,34
7	7
9	10
10	1,2,6,24,25,33
11	2,6,8,10,13,25

Tabla 4.5 Secciones más representativas en el análisis de correspondencias múltiples en frecuencias relativas para las últimas cinco convocatorias

En este último análisis, como se puede comprobar, la calidad de la representación se ha visto reducida en un 3%, con una inercia total recogida por los ejes principales del 56%. Observando además que la sección 33 del distrito 10 es una de las más próximas al origen junto con la sección 25 del distrito 10, como ocurría en el estudio en frecuencias absolutas.

También, destacamos que en este análisis la sección más cercana al origen es la 2 del distrito 10, la cual no resultaba representativa en el análisis en frecuencias absolutas por estar formado por un número de votantes menor.

Si aumentamos 0.01 el radio de la circunferencia añadimos dieciocho secciones más, es decir, obtendríamos cuarenta secciones en su interior:

Distrito	Secciones
3	5
4	1,2,3,13,29,32,39
6	2,3,5,9,16,17,32,34
7	7,8
8	2
9	10
10	1,2,6,14,18,21,24,25,32,33,34,39
11	2,6,8,10,13,25,27,40

Tabla 4.6 Secciones más representativas en el análisis de correspondencias múltiples en frecuencias relativas aumentando el radio de la circunferencia para las últimas cinco convocatorias

Observamos que la sección 37 del distrito 4, la 28 del distrito 10 y la 25 del distrito 11, no resultaron representativas en el análisis de correspondencias múltiples en frecuencias relativas, pero sí, el resto de las secciones.

Por último, añadir que al realizar el aumento del radio obtuvimos la sección 27 del distrito 11 como representativa, al igual que en el estudio en frecuencias absolutas.

4.3. Análisis del comportamiento de las secciones censales en cada convocatoria electoral.

Para cada convocatoria se van a realizar dos tipos de análisis, uno de correspondencias simple y otro de clasificación no jerarquizada.

Análisis de correspondencias simple:

Para cubrir el propósito de identificar las secciones que fueron más representativas en cada año electoral, vamos a representar los resultados del análisis de correspondencias simple en un mapa simétrico donde un círculo centrado en el origen y de radio definido en cada convocatoria determinará cuáles serán las secciones que cumplan el objetivo, por ser las más cercanas al perfil medio de la nube de puntos.

Para realizar el procedimiento se va a utilizar la función en R *ca()* [16] del paquete *ca*. Esta función recibe una tabla de contingencia con los votos de cada sección a cada partido en frecuencias absolutas, de manera que cada perfil fila posee un peso distinto según el número de individuos que vote en la sección. En relación con lo anterior, planteamos otro escenario en el cual todos los perfiles fila, es decir, las secciones, posean los mismos pesos, de manera que, expresaremos las frecuencias en porcentaje y aplicaremos el mismo procedimiento.

Análisis no jerárquico:

Realizamos un análisis no jerarquizado con la particularidad de que el número de clústeres será seleccionado de manera que pueda cubrir mejor el objetivo de este trabajo. Nuestro interés es agrupar las secciones en un número de divisiones, de manera que se establece un conjunto reducido de secciones censales representativas en cada candidatura. De esta forma, el grupo al que pertenezca la sección artificial será del que se obtengan las secciones más representativas.

Para realizar este procedimiento vamos a aplicar la función en R *pam()* [18] incluida en el paquete *cluster*. Esta función tiene la particularidad de que puede utilizar para aplicar el algoritmo de *k-medoids*. En la práctica recibe como entrada una matriz de disimilitud calculada con el método de Kullback-Leiber.

La aplicación de ambos análisis nos permitirá comparar los resultados, es decir, los dos conjuntos de secciones representativas. Si tomamos la intersección de ambos conjuntos, los resultados obtenidos estarán avalados por ambos análisis.

Nos hemos planteado algunas variantes en la implementación de los métodos anteriores:

- Utilizar frecuencias en porcentajes en el análisis de correspondencias para que todas las secciones tengan el mismo peso.
- Utilizar una clasificación más fina de las candidaturas que la que se propone en (Anexo I), dejando en la categoría de otros aquellas con un porcentaje de voto menor del 5%.

Vamos a tomar como modelo, para estudiar la influencia de estas variantes y decidir los pasos a llevar a cabo con cada convocatoria, la convocatoria del 24 de mayo de 2015 (Sección 4.3.1), puesto que se trata de una convocatoria compartida con las elecciones municipales. De este modo, podremos comparar los resultados con los obtenidos en [3] para el análisis de las elecciones municipales, trabajo en el que se utilizaron distintos métodos de clasificación, pero no se utilizó análisis de correspondencias.

Finalmente, comentamos la posibilidad de utilizar un análisis clúster jerárquico. En este tipo de análisis no es necesario definir previamente el número de grupos, pero sí el método que se utiliza para definir el índice de agrupamiento. Tal y como se explicó en la Sección 2.2 existen cuatro métodos de los cuales se probará la eficacia para este conjunto de datos.

Para realizar este procedimiento vamos a aplicar la función en R *agnes()* [19] incluida en el paquete *cluster*. Esta función tiene la particularidad de recibir una matriz de disimilaridades, con la que aplicando el método óptimo para la selección del índice de agregación, se obtendrá las secciones más representativas. En la selección de dicho método se busca el que genere la partición más robusta. Un criterio para cuantificar esta propiedad es el coeficiente de agregación, con el que obteniendo una media de dicha medida para las trece convocatorias con los métodos estudiados, se obtendrían los siguientes resultados en la tabla 4.7.

	<i>Simple linkage</i>	<i>Complete linkage</i>	<i>Average linkage</i>	<i>Método Ward</i>
<i>Ca</i>	0.988983	0.997506	0.9946157	0.9987272

Tabla 4.7 Valores medios del coeficiente de agregación para las cuatro metodologías estudiadas

Como podemos observar, el mayor coeficiente de agregación se consigue con el método de Ward. Tras aplicarlo se utilizará como representación el dendrograma (procedimiento donde se observa los pasos que siguió dicho algoritmo).

La aplicación de este análisis a nuestros datos se detalla en el Anexo V, posteriormente compararemos en la Sección 5 los resultados obtenidos por clasificación jerárquica con los del método que acabamos de describir, que utiliza análisis de correspondencias y análisis clúster no jerárquico.

Por otra parte, con el propósito de conseguir una visualización más clara de los resultados del análisis realizado, se representarán mapas de la ciudad divididos por secciones censales en los cuales el color representará si dicha sección resultaría representativa o no. La implementación de los mapas se ha llevado a cabo con los paquetes *sf* [20] y *ggplot2* [21].

4.3.1. Elecciones 24 de mayo de 2015

En dichas elecciones se presentaron 14 candidaturas, las cuales se han resumido en 6 tendencias políticas (Anexo I). La división censal de dicha convocatoria consistía en 266 secciones agrupadas en 12 distritos.

Comenzamos realizando un análisis de correspondencias simple para analizar cuáles son las secciones más próximas al origen, por ser estas las más representativas del resultado global para Valladolid en dicha convocatoria.

El otro método que vamos a utilizar para localizar las secciones más representativas es un análisis no jerárquico, pero en este caso generando una sección artificial que representa la distribución del voto en Valladolid. Así, el grupo en el que se encuentre dicha sección será el que contenga las secciones más parecidas a la distribución del voto real.

A partir de estos dos métodos vamos a poder obtener las secciones más representativas para el año electoral 2015.

En el análisis de correspondencias, se elaborará un mapa simétrico donde se incluye una circunferencia centrada en el origen y con radio 0.05. Dicha circunferencia obtenida de manera experimental representa los límites para definir qué secciones son las más similares a la distribución global del voto. En la tabla 4.8 mostramos las ocho secciones localizadas en el interior del círculo descrito:

Distrito	Secciones
3	4
6	3
10	27,33,34
11	10,25,40

Tabla 4.8 Secciones representativas en la aplicación de análisis de correspondencias simple en la convocatoria de 2015

A continuación, consideramos los resultados obtenidos aplicando análisis clúster no jerárquico con un número de grupos de 25. Hemos elegido este número de grupos después de probar con distintos escenarios, estimando que se trata de un número adecuado a nuestro objetivo para esta convocatoria. El número de secciones obtenidas dentro del grupo donde se localiza la sección artificial que contiene la distribución global del voto fue un total de 20 secciones en este caso.

Este número de secciones es ligeramente elevado, pero también es cierto que en los otros escenarios planteados no se conseguía reducir sustancialmente ese número de secciones.

En la tabla 4.9 se muestran las veinte secciones pertenecientes al mismo grupo que la sección objetivo.

Distrito	Secciones
3	4
4	37
6	3,9
9	10
10	18,27,32,33,34,35,39
11	09,10,17,25,29,36,40
12	3

Tabla 4.9 Secciones representativas en la aplicación de análisis de clúster no jerárquico en la convocatoria de 2015

Realizando una intersección de ambos conjuntos de resultados, las secciones que resultaron como las más próximas a la distribución global del voto con los dos métodos aplicados fueron las incluidas en la tabla 4.10.

Distrito	Secciones
3	4
6	3
10	27,33,34
11	10,25,40

Tabla 4.10 Secciones comunes entre el análisis de correspondencias y clúster en la convocatoria de 2015

En este caso obtenemos un total de ocho secciones que parecen satisfacer nuestro objetivo por no ser demasiadas y resultar representativa en las elecciones municipales del 24 de mayo de 2015 [3] pág. 50, a excepción de la 34 del distrito 10.

Otro escenario planteado en la aplicación de análisis de correspondencias consiste en establecer los datos en porcentajes con el fin de que todas las secciones tengan el mismo peso siendo las que estén más próximas al origen, las más representativas sin que influya el número de votantes del distrito.

Con este análisis, en el círculo de radio 0.05 obtenemos las cinco secciones siguientes:

Distrito	Secciones
6	9
8	1
10	33
11	10,25

Si aumentamos 0.01 el radio del círculo añadimos seis secciones más, es decir, obtendríamos once secciones en su interior:

Distrito	Secciones
3	4
6	2,3,9
8	1,2
10	27,33
11	10,25,40

Tabla 4.11 Secciones representativas del análisis de correspondencias en el segundo escenario para la convocatoria de 2015

Observamos que las secciones 2 y 9 del distrito 6 y las secciones 1 y 2 del distrito 8 no resultaban representativas en el análisis de correspondencias en frecuencias absolutas, pero sí, el resto de las secciones.

En la intersección con la metodología de clasificación no supervisada no jerarquizada, obtenemos como secciones resultantes de la intersección de ambas:

Distrito	Secciones
3	4
6	3,9
10	27,33
11	10,25,40

Tabla 4.12 Secciones representativas en el segundo escenario para la convocatoria de 2015

En este caso obtenemos un total de ocho secciones, de las cuales la sección 9 del distrito 6 no resultaba representativa en el análisis en frecuencias absolutas y la sección 34 del distrito 10 sí era representativa, pero no en el análisis en frecuencias relativas. Este resultado puede deberse a que el número de votantes de la sección 9 del distrito 6 era significativamente menor y por ello, se justifica que en este análisis resulte representativa al igualar los pesos de todas las secciones.

Con todo esto, podemos concluir que es una metodología similar a la primera, aportando una mejora en el número de secciones representativas en comparación con el estudio realizado en las elecciones municipales. Se observa que, si no se aumentara el radio para encontrar un número de secciones mayor, obtendríamos en la intersección de los análisis un total de cuatro secciones comunes, lo que puede ser insuficiente para extraer conclusiones de la comparación con todas las convocatorias.

El último escenario que nos planteamos es no resumir las candidaturas en tendencias políticas, ya que puede ocurrir que estemos ignorando alguna modificación debido a este resumen realizado. Por lo cual, vamos a realizar el análisis, agrupando aquellas candidaturas que no hayan obtenido el 5% de los votos.

Actuando de este modo en el análisis de correspondencias simple, la inercia total se ha visto reducido en un 6%, lo cual quiere decir que la calidad de la representación se ha visto reducida. A mayores ha aumentado el número de secciones, teniendo ahora un total de once. Estas secciones son:

Distrito	Secciones
3	4
6	3
8	1
10	27,33,34
11	9,10,25,27,40

Tabla 4.13 Secciones representativas del análisis de correspondencias en el tercer escenario para la convocatoria de 2015

Respecto al análisis clúster, aumentamos el número de grupos a 25, vamos a obtener las secciones que se encuentran dentro del grupo donde se localiza la sección artificial con la distribución del voto global. Para este caso el número de secciones son las once siguientes:

Distrito	Secciones
4	34
6	3
10	18,27,35
11	9,10,17,29,36,40

Tabla 4.14 Secciones representativas del análisis de clúster en el tercer escenario para la convocatoria de 2015

De manera que, comparando con el análisis de correspondencias nos quedarían como comunes las siguientes secciones:

Distrito	Secciones
6	3
10	27
11	9,10,40

Tabla 4.15 Secciones comunes entre el análisis de correspondencias y clúster para el tercer escenario en la convocatoria de 2015

En este último escenario nos encontramos que comparando con las elecciones municipales de 2015[3] Pág. 50, todas estas secciones obtenidas fueron representativas. Esto supone que este último escenario, igual que el anterior, no parece aportar ninguna mejora, por lo cual no se plantea como posible opción de modificación para aplicar a todas las convocatorias.

En la figura 4.9 se representa el mapa de la ciudad de Valladolid dividido por secciones, destacando las ocho obtenidas como representativas en la intersección del análisis de correspondencias simple y clúster no jerárquico.

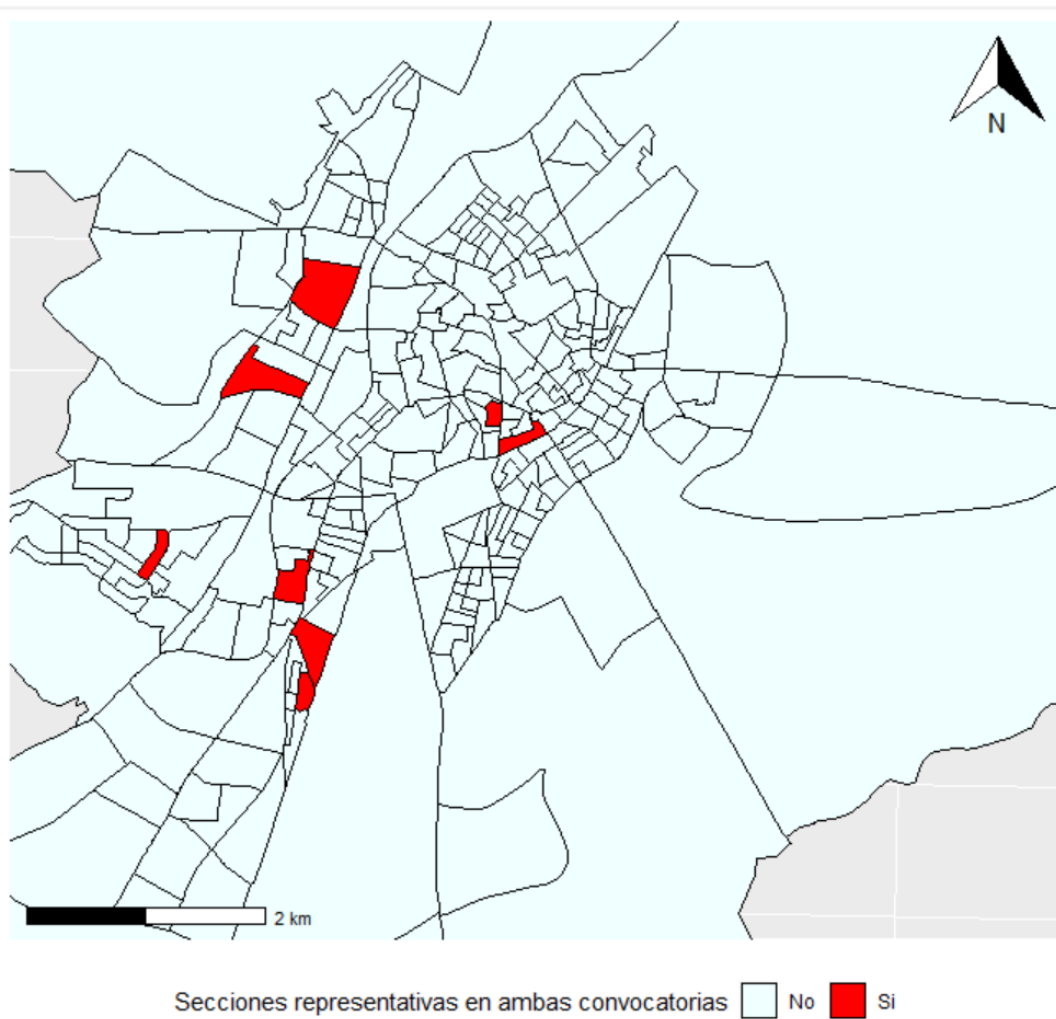


Figura 4.9 Mapa de la ciudad de Valladolid con las secciones más representativas del año 2015

4.3.2. Elecciones 28 de junio de 1982

En dichas elecciones se presentaron 16 candidaturas, las cuales se han resumido en 5 tendencias políticas. La división censal de dicha convocatoria consistía en 245 secciones agrupadas en 12 distritos.

En el análisis de correspondencias, para cubrir nuestro objetivo elaboramos un mapa simétrico donde se incluye una circunferencia centrada en el origen y con radio 0.075. Dicha circunferencia representa la frontera para definir qué secciones son las más similares a la distribución global del voto. En la aplicación del análisis clúster no jerárquico utilizamos como número de grupos 20. Dicho número de grupos es adecuado para que en este escenario se cumpla nuestro objetivo. En la tabla 4.16, se obtiene la intersección de los resultados obtenidos en ambos análisis.

Distrito	Secciones
6	9,10,17
7	15
8	4
10	28
11	3,27,29

Tabla 4.16 Secciones representativas en la aplicación de análisis de correspondencias y clúster para la convocatoria de 1982

En este caso obtenemos un total de nueve secciones que parecen satisfacer nuestro objetivo, ya que no son demasiadas cumpliendo con el criterio establecido. Por todo esto podemos concluir que es una metodología eficaz para alcanzar nuestro objetivo.

Los otros escenarios planteados no aportan soluciones mejores y únicamente ratifican lo obtenido por esta metodología, por lo cual no se incluyen para esta convocatoria.

4.3.3. Elecciones 22 de junio de 1986

En dichas elecciones se presentaron 13 candidaturas, las cuales se han resumido en 6 candidaturas más genéricas. La división censal de dicha convocatoria consistía en 248 secciones agrupadas en 12 distritos.

En el análisis de correspondencias, para cubrir nuestro objetivo utilizamos un círculo centrado en el origen y con radio 0.09. En la aplicación del análisis clúster no jerárquico utilizamos como número de grupos 20. En la tabla 4.17, se obtiene la intersección de los resultados obtenidos en ambos análisis.

Distrito	Secciones
6	2
7	6
8	9,41
11	2,11,28

Tabla 4.17 Secciones representativas en la aplicación de análisis de correspondencias y clúster para la convocatoria de 1986

El resultado de aplicar esta metodología está formado por siete secciones que, en comparación con las elecciones municipales del 10 de junio de 1987 [3] Pág. 46, se obtendrían como secciones comunes la sección 6 el distrito 7 y las secciones 2, 11 y 28 del distrito 11. Esta comparación no supone un criterio, pero verifica que los resultados obtenidos contienen patrones comunes a la convocatoria de las elecciones municipales.

4.3.4. Elecciones 29 de octubre de 1989

En dichas elecciones se presentaron 14 candidaturas, las cuales se han resumido en 6 tendencias políticas. La división censal de dicha convocatoria en 242 secciones agrupadas 12 distritos.

En el análisis de correspondencias, para cubrir nuestro objetivo elaboramos un mapa simétrico donde se incluye un círculo centrado en el origen y con radio 0.075. En la aplicación del análisis clúster no jerárquico utilizamos como número de grupos 20. En la tabla 4.18, se obtiene la intersección de los resultados obtenidos en ambos análisis.

Distrito	Secciones
3	6
6	8,17
7	6
8	2
10	25
11	2,11,26,28

Tabla 4.18 Secciones representativas en la aplicación de análisis de correspondencias y clúster para la convocatoria de 1989

El resultado de aplicar esta metodología obtiene diez secciones que, en comparación con las elecciones municipales del 10 de junio de 1987 [3] Pág. 46, resultarían como secciones comunes la sección 6 del distrito 3, la sección 17 del distrito 6, la sección 6 del distrito 7 y las secciones 2 ,11 ,26 y 28 del distrito 11. Esta comparación no supone un criterio, aunque es cierto que permite certificar los resultados obtenidos en dicho estudio.

4.3.5. Elecciones 6 de junio de 1993

En dichas elecciones se presentaron 18 candidaturas, las cuales se han resumido en 6 candidaturas más genéricas. La división censal de dicha convocatoria consistía en 252 secciones agrupadas en 12 distritos.

En el análisis de correspondencias, elaboramos un mapa simétrico donde se incluye un círculo centrado en el origen y con radio 0.06. En la aplicación del análisis clúster no jerárquico utilizamos como número de grupos 25. En la tabla 4.19, se obtiene la intersección de los resultados obtenidos en ambos análisis.

Distrito	Secciones
3	6
6	3,8
7	6
10	2,25
11	11,26,29

Tabla 4.19 Secciones representativas en la aplicación de análisis de correspondencias y clúster para la convocatoria de 1993

El resultado de aplicar esta metodología obtiene nueve secciones que, en comparación con las elecciones municipales del 26 de mayo de 1991 [3] Pág. 46, resultarían como secciones comunes la sección 3 del distrito 6, la sección 6 del distrito 7 y la sección 25 del distrito 10. Esta comparación no supone un criterio, aunque es cierto que permite certificar los resultados obtenidos en dicho estudio.

4.3.6. Elecciones 3 de marzo de 1996

En dichas elecciones se presentaron 18 candidaturas, las cuales se han resumido en 6 tendencias políticas. La división censal de dicha convocatoria consistía en 252 secciones agrupadas en 12 distritos.

En el análisis de correspondencias, para cubrir nuestro objetivo elaboramos un mapa simétrico donde se incluye un círculo centrado en el origen y con radio 0.07. En la aplicación del análisis clúster no jerárquico utilizamos como número de grupos 25. En la tabla 4.20, se obtiene la intersección de los resultados obtenidos en ambos análisis.

Distrito	Secciones
6	17
9	10
10	2,25,43
11	13,21,26

Tabla 4.20 Secciones representativas en la aplicación de análisis de correspondencias y clúster para la convocatoria de 1996

El resultado de aplicar esta metodología obtiene ocho secciones que, en comparación con las elecciones municipales del 28 de mayo de 1995 [3] Pág. 47, resultarían como secciones comunes la sección 17 del distrito 6 y la sección 26 del distrito 11. Esta comparación no supone un criterio, aunque es cierto que permite certificar los resultados obtenidos en dicho estudio.

4.3.7. Elecciones 12 de marzo de 2000

En dichas elecciones se presentaron 15 candidaturas, las cuales se han resumido en 6 candidaturas más genéricas. La división censal de dicha convocatoria consistía en 258 secciones agrupadas en 12 distritos.

En el análisis de correspondencias, elaboramos un mapa simétrico donde se incluye un círculo centrado en el origen y con radio 0.05. En la aplicación del análisis clúster no jerárquico utilizamos como número de grupos 25. En la tabla 4.21, se obtiene la intersección de los resultados obtenidos en ambos análisis.

Distrito	Secciones
6	17
10	25,28,33
11	12

Tabla 4.21 Secciones representativas en la aplicación de análisis de correspondencias y clúster para la convocatoria de 2000

El resultado de aplicar esta metodología obtiene cinco secciones que, en comparación con las elecciones municipales del 13 de junio 1999 [3] Pág. 48, obtendríamos como secciones comunes las secciones 28 y 33 del distrito 10 y la sección 12 del distrito 11. Esta comparación no supone un criterio, aunque es cierto que permite certificar los resultados obtenidos en dicho estudio.

4.3.8. Elecciones 14 de marzo de 2004

En dichas elecciones se presentaron 19 candidaturas, las cuales se han resumido en 6 candidaturas más genéricas. La división censal de dicha convocatoria consistía en 270 secciones agrupadas en 12 distritos.

En el análisis de correspondencias, para cubrir nuestro objetivo elaboramos un mapa simétrico donde se incluye un círculo centrado en el origen y con radio 0.05. En la aplicación del análisis clúster no jerárquico utilizamos como número de grupos 20. En la tabla 4.22, se obtiene la intersección de los resultados obtenidos en ambos análisis.

Distrito	Secciones
3	5
4	37
8	5
10	27,33,35,42,43
11	27,40

Tabla 4.22 Secciones representativas en la aplicación de análisis de correspondencias y clúster para la convocatoria de 2004

El resultado de aplicar esta metodología está formado por diez secciones que, en comparación con las elecciones municipales del 25 de mayo de 2003 [3] Pág. 49, obtendríamos como secciones comunes la sección 37 del distrito 4, la sección 5 del distrito 8 y la sección 43 del distrito 10. Esta comparación no supone un criterio, aunque es cierto que permite certificar los resultados obtenidos en dicho estudio.

4.3.9. Elecciones 9 de marzo de 2008

En dichas elecciones se presentaron 21 candidaturas, las cuales se han resumido en 6 candidaturas más genéricas. Además, la división censal en esta convocatoria consistía en 266 secciones agrupadas en 12 distritos.

En el análisis de correspondencias, elaboramos un mapa simétrico donde se incluye un círculo centrado en el origen y con radio 0.05. En la aplicación del análisis clúster no jerárquico utilizamos como número de grupos 30. En la tabla 4.23, se obtiene la intersección de los resultados obtenidos en ambos análisis.

Distrito	Secciones
3	5
4	37
6	10
10	2,33,39
11	27

Tabla 4.23 Secciones representativas en la aplicación de análisis de correspondencias y clúster para la convocatoria de 2008

El conjunto resultado de aplicar dicha metodología está formado por siete secciones que, en comparación con las elecciones municipales del 27 de mayo de 2007 [3] Pág. 49, obtendríamos como secciones comunes la sección 37 del distrito 4 y la sección 10 del distrito 6. Esta comparación no supone un criterio, aunque es cierto que permite certificar los resultados obtenidos en dicho estudio.

4.3.10. Elecciones 20 de noviembre de 2011

En dichas elecciones se presentaron 14 candidaturas, las cuales se han resumido en 6 candidaturas más genéricas. Además, la división censal de dicha convocatoria consistía en 270 secciones agrupadas en 12 distritos.

En el análisis de correspondencias, elaboramos un mapa simétrico donde se incluye un círculo centrado en el origen y con radio 0.05. En la aplicación del análisis clúster no jerárquico utilizamos como número de grupos 25. En la tabla 4.24, se obtiene la intersección de los resultados obtenidos en ambos análisis.

Distrito	Secciones
6	3,17
8	2
9	10
10	33
11	27,40

Tabla 4.24 Secciones representativas en la aplicación de análisis de correspondencias y clúster para la convocatoria de 2011

El resultado de aplicar esta metodología está formado por siete secciones que, en comparación con las elecciones municipales del 22 de mayo de 2011[3] Pág. 50, obtendríamos como secciones comunes la sección 17 del distrito 6, la sección 2 del distrito 8, la sección 33 del distrito 10 y la sección 27 del distrito 11. Esta comparación no supone un criterio, aunque es cierto que permite certificar los resultados obtenidos en dicho estudio.

4.3.11. Elecciones 3 de mayo de 2016

En dichas elecciones se presentaron 14 candidaturas, las cuales se han resumido en 6 candidaturas más genéricas. Además, la división censal de dicha convocatoria consistía en 263 secciones agrupadas en 12 distritos.

En el análisis de correspondencias, para cubrir nuestro objetivo utilizamos un círculo centrado en el origen y con radio 0.05. En la aplicación del análisis clúster no jerárquico utilizamos como número de grupos 20. En la tabla 4.25, se obtiene la intersección de los resultados obtenidos en ambos análisis.

Distrito	Secciones
3	4
4	37
6	3
10	33,34
11	9,10

Tabla 4.25 Secciones representativas en la aplicación de análisis de correspondencias y clúster para la convocatoria de 2016

El resultado de aplicar esta metodología está formado por siete secciones que, en comparación con las elecciones municipales del 26 de mayo de 2019[3] Pág. 51, obtendríamos como secciones comunes la sección 37 del distrito 4, la sección 33 del distrito 10 y las secciones 9 y 10 del distrito 11. Esta comparación no supone un criterio, aunque es cierto que permite certificar los resultados obtenidos en dicho estudio.

Hay que destacar que para esta convocatoria los resultados, por lo que respecta al aumento de secciones comunes, mejoran aplicando el segundo escenario. Resultando representativas en este caso, la sección 2 del distrito 6 y las secciones 12 y 25 del distrito 11. Este resultado puede deberse a que secciones como la 12 del distrito 11 tiene un número de votantes significativamente menor a los de las secciones obtenidas en la primera metodología.

4.3.12. Elecciones 28 de abril de 2019

En dichas elecciones se presentaron 12 candidaturas, las cuales se han resumido en 6 candidaturas más genéricas. Además, la división censal de esta convocatoria consistía en 258 secciones agrupadas en 12 distritos.

En el análisis de correspondencias, para cubrir nuestro objetivo utilizamos un círculo centrado en el origen y con radio 0.05. En la aplicación del análisis clúster no jerárquico utilizamos como número de grupos 25. En la tabla 4.26, se obtiene la intersección de los resultados obtenidos en ambos análisis.

Distrito	Secciones
6	3
10	2,28
11	9,10,27
12	3

Tabla 4.26 Secciones representativas en la aplicación de análisis de correspondencias y clúster para la primera convocatoria de 2019

El resultado de aplicar esta metodología está formado por siete secciones que, en comparación con las elecciones municipales del 26 de mayo de 2019[3] Pág. 51, obtendríamos como secciones comunes la sección 28 del distrito 10 y las secciones 9 y 10 del distrito 11. Esta comparación no supone un criterio, aunque es cierto que permite certificar los resultados obtenidos en dicho estudio.

4.3.13. Elecciones 10 de noviembre de 2019

En dichas elecciones se presentaron 12 candidaturas, las cuales se han resumido en 6 tendencias políticas. La división censal de esta convocatoria consistía en 258 secciones agrupadas en 12 distritos. Cabe señalar que esta convocatoria, muestra un comportamiento anómalo con respecto a todas las convocatorias estudiadas hasta el momento, puesto que, tras lo obtenido en el análisis global (Sección 4.1), se trataba de una convocatoria atípica.

Comenzamos realizando un análisis de correspondencias simple para analizar cuáles son las secciones que están más próximas al origen, por ser estas las más representativas del resultado global para Valladolid. En esta convocatoria, la definición del círculo centrado de origen es compleja, puesto que, al definir el radio adecuado nos encontramos con un número escaso de secciones y por ello en este caso su medida es de 0.06.

El otro método que vamos a utilizar para localizar las secciones más representativas es un análisis no jerárquico, pero en este caso generando una sección artificial que representa la distribución global del voto en Valladolid. Así, el grupo en el que se encuentre dicha sección será el que contenga las secciones más parecidas a la distribución del voto real.

En este primer análisis se recogió un 77% de la inercia total, lo cual quiere decir que está infravalorada la calidad de la representación y que, comparándolo con otras convocatorias, esta variabilidad que recogen los ejes se ha visto reducida en un 12%. Además, se observa que a pesar de tener dos secciones muy próximas al origen (secciones 5 y 33 del distrito 8), las otras se localizan demasiado alejadas, provocando el aumento del radio. Con todos estos planteamientos se obtienen las once secciones siguientes:

Distrito	Secciones
6	32
8	5,33
10	1,27,35
11	5,13,17,31,36

Tabla 4.27 Secciones representativas en la aplicación de análisis de correspondencias simple en la segunda convocatoria de 2019

A continuación, consideramos los resultados obtenidos aplicando análisis clúster con un número de grupos de 30. El resultado obtenido por este análisis se observa en la tabla 4.28 y fueron las siguientes diecisiete secciones:

Distrito	Secciones
2	15
3	4
10	1,5,27,29,34,35,42
11	5,9,17,27,28,29,31,36

Tabla 4.28 Secciones representativas en la aplicación de análisis de clúster no jerárquico en la segunda convocatoria de 2019

Realizando una intersección de ambos conjuntos de resultados, las secciones que resultaron como las más próximas a la distribución global del voto con los dos métodos aplicados fueron las incluidas en la tabla 4.29.

Distrito	Secciones
10	1,27,35
11	5,17,31,36

Tabla 4.29 Secciones representativas en la aplicación de análisis de correspondencias y clúster para la segunda convocatoria de 2019

Para esta convocatoria dicha metodología resulta eficaz con un total de siete secciones comunes entre ambos métodos, observándose que, en la aplicación de los otros dos escenarios planteados, los resultados no mejoran en cuanto al número de secciones. En comparación con las elecciones municipales del 26 de mayo de 2019[3] Pág. 51, resultaron como secciones comunes las secciones 1 y 27 del distrito 10. Esta comparación no supone un criterio, aunque es cierto que permite certificar los resultados obtenidos en dicho estudio.

En la figura 4.10, se obtiene una representación de la ciudad de Valladolid en dicha convocatoria, destacando las secciones que resultaron representativas en las elecciones generales y las que lo fueron en las elecciones municipales. Se definen, para ello, tres grupos en la representación.

G1: Secciones no obtenidas como representativas en el análisis.

G2: Secciones representativas en las elecciones generales, pero no resultaban representativas en las municipales

G3: Secciones representativas en las elecciones generales y municipales

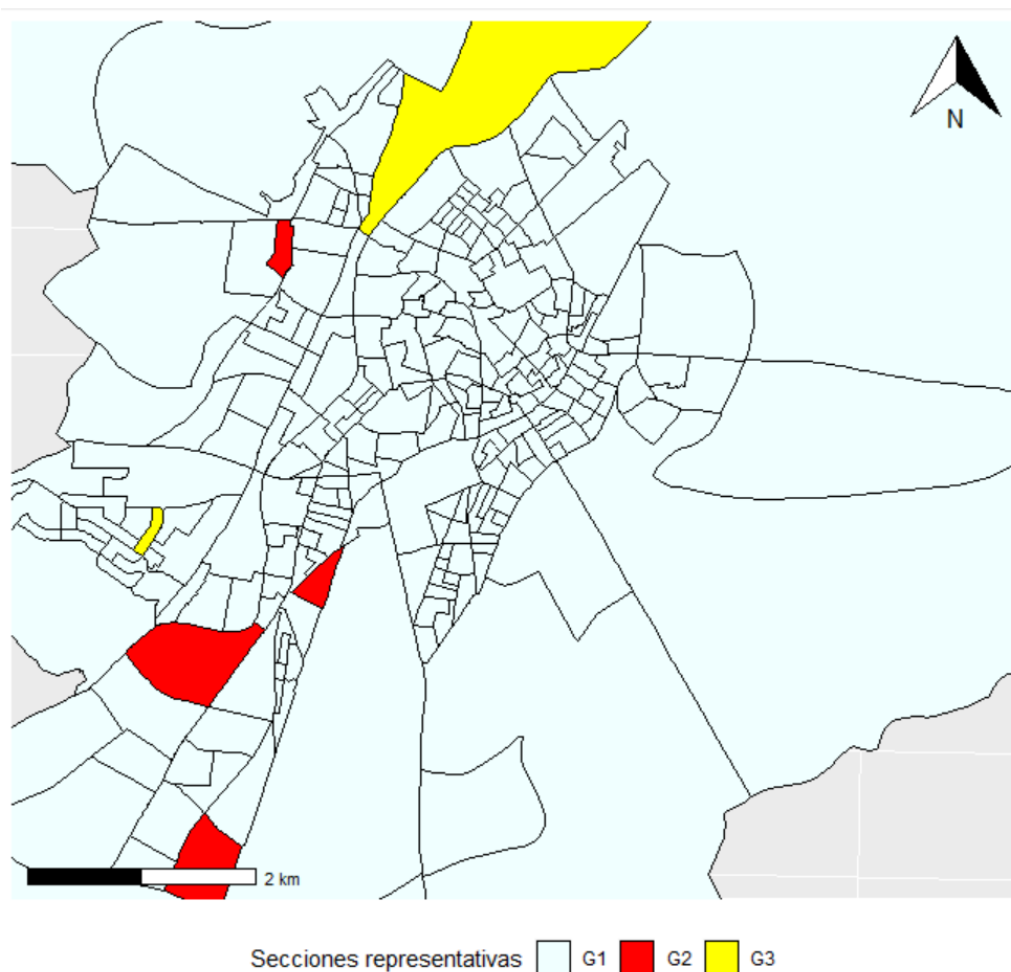


Figura 4.10 Mapa de la ciudad de Valladolid con las secciones más representativas de la segunda convocatoria del año 2019

4.4. Comparación entre los resultados obtenidos en el análisis de correspondencias múltiples y los análisis de correspondencias simples por convocatoria

Una vez realizado el análisis independiente de las trece convocatorias, así como el global de todas ellas, se ha obtenido un conjunto de secciones censales que poseen un comportamiento similar a la distribución global del voto en cada convocatoria.

Ahora vamos a comparar si con la aplicación del análisis de correspondencias múltiples se obtienen los mismos resultados que con la aplicación de los análisis de correspondencias simples. Con esta comparación se podrá concluir si es suficiente un análisis conjunto o, por el contrario, es necesario la realización de un análisis por convocatoria para obtener resultados concluyentes.

Para comenzar, estudiamos el primer análisis obtenido, es decir, el análisis conjunto de todas las convocatorias. En este, los resultados no lograban alcanzar nuestro objetivo puesto que, las convocatorias son bastante dispares entre sí, además de tener una baja inercia recogida por los ejes principales. Por ello realizamos un segundo análisis conjunto de las siguientes cinco convocatorias (2008,2011,2015,2016,2019-04), ya que estas presentaban una mayor homogeneidad entre sí, garantizando que alcanzarían mejores resultados. En la tabla 4.30, obtenemos las secciones más representativas en los respectivos análisis independientes, quedándonos con aquellas que al menos fueran representativas en tres años electorales.

Distrito	Sección	Convocatorias
6	3	2011,2015,2016,2019-04
10	33	2008,2011,2015,2016
11	10	2015,2016,2019-04
11	27	2008,2011,2016,2019-04

Tabla 4.30 Tabla resumen con las secciones más representativas en las últimas cinco convocatorias

De manera que, comparando con los resultados del análisis de correspondencias múltiples mostrados en la tabla 4.4 las cuatro secciones obtenidas en la tabla anterior son representativas en ambos estudios, destacando que la sección 27 del distrito 11 era una de las más próximas al origen, en el análisis conjunto de todas las convocatorias. También, se observó que la sección 13 del distrito 11, era la más próxima a la distribución del perfil global. Sin embargo, esta misma sección no resulta representativa en el análisis por convocatorias.

Así mismo, podría dudarse de los resultados obtenidos debido a la baja inercia recogida por los ejes principales, sin embargo, teniendo en cuenta que en este tipo de análisis conjunto se estudia tanto las convocatorias por separado como posibles interacciones entre ellas, se podría afirmar que se trata de un análisis mucho más riguroso.

Finalmente, se observa que la aplicación del análisis conjunto vendrá condicionado por la homogeneidad de las convocatorias seleccionadas. De manera que, si estas cumplen estos requisitos, con un análisis de correspondencias múltiples lograríamos nuestro objetivo. Y si fuese necesario entrar en detalle, aplicaríamos un análisis por convocatoria.

5. Resumen de los resultados

En este último apartado, se exponen todos los resultados obtenidos por las distintas metodologías respecto de las secciones más representativas en la distribución real del voto en las trece convocatorias electorales al Congreso de los Diputados en la ciudad de Valladolid.

Para ello, debemos recordar los valores de los distintos parámetros definidos en cada análisis:

- **El radio de la circunferencia** centrada en el origen, que representa los límites para definir qué secciones son las más similares a la distribución global del voto, osciló entre 0.05 y 0.1. Esta oscilación se debe principalmente, a que en algunos casos las secciones estaban muy próximas al centro de gravedad y por ello bastaba con una medida próxima a 0.05, mientras que, en otros casos, se debía aumentar el radio para poder encontrar un número razonable.
- **El número de clúster** seleccionado alcanzó valores de 20 a 30. En este caso, dichos valores están condicionados a localizar el menor número de secciones posibles con distribución similar a la real en cada convocatoria.

A continuación, se presentan los resultados obtenidos:

A través de la tabla resumen con todas las secciones censales representativas por convocatoria (Anexo II) se extraen las obtenidas como representativas en al menos cuatro convocatorias. Destacando la **sección 2 del distrito 10** situada en (**La victoria**), la **sección 25 del distrito 10** situada en (**Huerta del Rey-Media**), la **sección 3 del distrito 6** situada en (**Plaza Circular**), la **sección 27 del distrito 11** situada en (**La rubia**) y la **sección 33 del distrito 10** situada en (**Huerta del Rey-Feria de muestras**) representativa esta última en seis convocatorias.

Fijándonos en la tabla 4.4 en el análisis de correspondencias múltiples, estas mismas secciones resultaron representativas también en las convocatorias (2008,2011,2015,2016 y 2019/04).

Destacable sería también, que **la sección 25 del distrito 10** y **la sección 27 del distrito 11** resultaron significativas en el análisis de correspondencias múltiples para todas las convocatorias.

Con respecto a los resultados obtenidos en el análisis de correspondencias múltiples en frecuencias relativas (Tabla 4.6) la **sección 2 del distrito 10** situada en (**La victoria**) resultó la más representativa, aunque también las demás secciones obtenidas en el Anexo II se obtuvieron como representativas y por ello quizá sea la metodología más adecuada, para lograr nuestro objetivo.

En los anexos III y IV se obtuvieron las secciones representativas de los dos escenarios alternativos planteados (frecuencias en tantos por ciento y candidaturas sin resumir), que tal y como pudimos comprobar, únicamente obtuvieron mejores resultados en la convocatoria de 2016, en cuanto al número de secciones resultantes.

Respecto a la aplicación del análisis clúster jerárquico (Anexo V) se obtuvieron como secciones más representativas **la sección 37 del distrito 4**, **la sección 33 del distrito 10** y **la sección 27 del distrito 11**. De manera que, comparando con los resultados obtenidos en la primera metodología nos quedarían como secciones comunes la **sección 33 del distrito 10** situada en (**Huerta del Rey-Feria de muestras**) y la **sección 27 del distrito 11** situada en (**La rubia**).

En la tabla 5.1 se realiza una tabla resumen con todas las secciones obtenidas en estos análisis que resultaron representativas.

			Tablas			
Distrito	Sección	Ubicación	Anexo 2	Anexo 3	Anexo 4	Anexo 5
04	37	Campo grande				✓
06	03	Plaza Circular	✓		✓	
10	02	La victoria	✓	✓	✓	
10	25	Huerta del Rey- Media	✓	✓	✓	
10	33	Huerta del Rey-Feria de muestras	✓	✓	✓	✓
11	27	La Rubia	✓	✓	✓	✓

Tabla 5.1 Tabla resumen con las secciones más representativas según los cuatro estudios realizados.

Por último, comparando con los resultados obtenidos en las elecciones municipales [3] Pág. 58, tendríamos que **la sección 33 del distrito 10** situada en **Huerta del Rey-Feria de muestras**, resulta ser representativa para ambos tipos de elecciones y por ello se trata de una sección realmente significativa. Con todo este análisis se concluye que sería la sección a la cual nos interesaría recoger los resultados en una jornada electoral.

En la figura 5.1 representamos un mapa de la ciudad con las secciones más representativas obtenidas en este análisis. Se toma como referencia de representación el mapa del año 2019 por ser esta la convocatoria más reciente.

En dicha representación se definen cinco agrupaciones de secciones en la figura, según la característica que la identifique:

G1: Secciones que no resultaron representativas en ningún análisis.

G2: Secciones que resultaron representativas en el análisis clúster jerárquico.

G3: Secciones que resultaron representativas en la aplicación por convocatorias de los métodos de análisis clúster no jerárquico y análisis de correspondencias simple.

G4: Secciones que resultaron representativas en el análisis clúster y en el análisis de correspondencias por convocatorias.

G5: Secciones que resultaron representativas en los tres métodos aplicados y en el trabajo de caracterización de secciones censales en las elecciones municipales [3], Pág. 57.

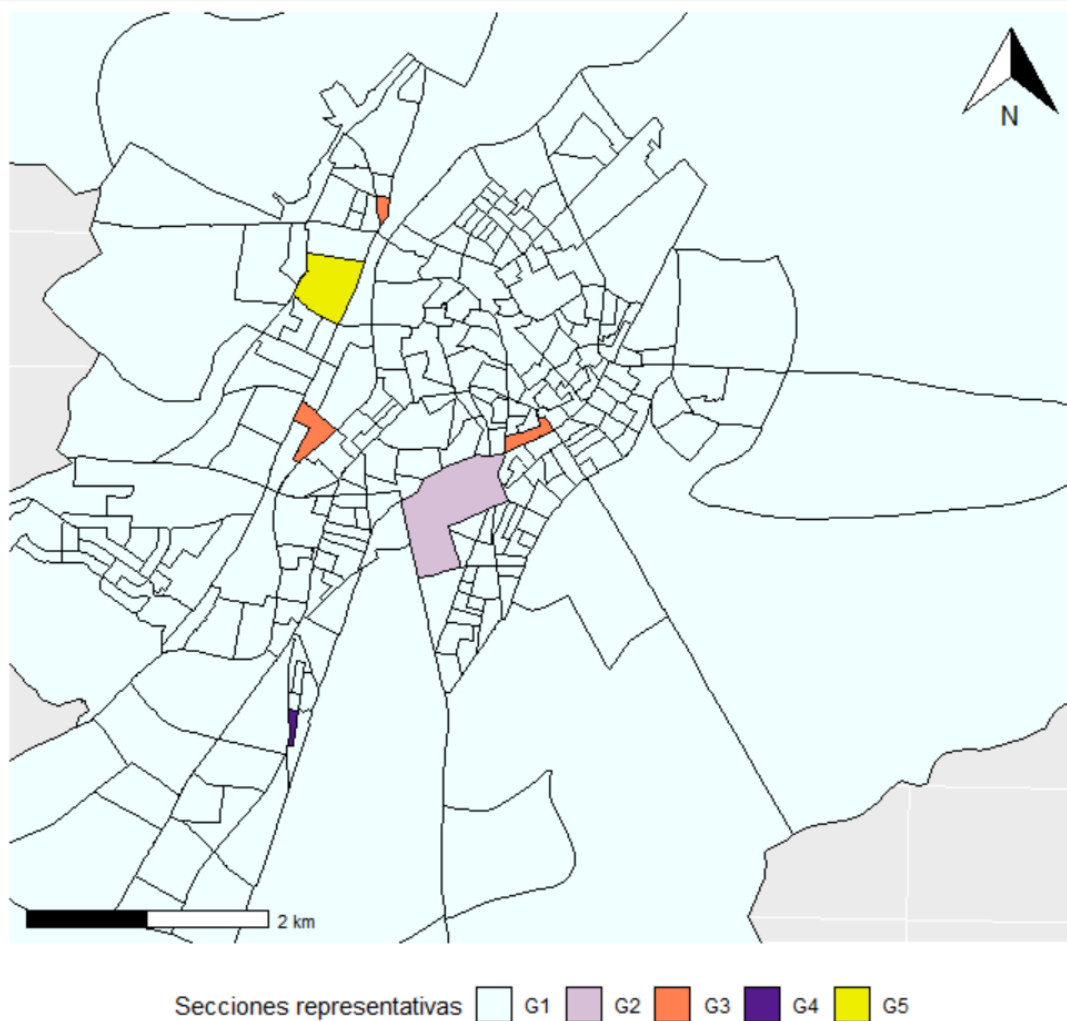


Figura 5.1 Mapa de la ciudad de Valladolid con las secciones más representativas

6. Conclusiones

Se ha realizado un estudio mediante el software R en relación con las elecciones generales celebradas en España entre 1982 y 2019, en la ciudad de Valladolid, con el objetivo de localizar las secciones más representativas, es decir, aquellas que mejor representaron la distribución del voto a lo largo de las convocatorias.

Para llevarlo a cabo se han utilizado metodologías analíticas tales como el análisis de correspondencias y el análisis clúster con diferentes escenarios planteados.

Junto al análisis de correspondencias, y aplicando un estudio general, se obtuvieron un listado de las secciones con tendencias políticas más marcadas y un resumen de la evolución política de las convocatorias.

Con todo ello se ha podido probar, que el mejor planteamiento para este tipo de investigación es la combinación del análisis de correspondencias en frecuencias absolutas contrastado con un análisis clúster. También se comprobó que con un estudio por convocatorias se obtienen resultados similares al estudio realizado de manera conjunta.

El resumen final de dicho estudio muestra cinco secciones como las más representativas, de las cuales, la sección 33 del distrito 10 situada en Huerta del Rey- Feria de Muestras, fue una de las secciones obtenidas en el estudio de las elecciones municipales [3] Pág. 58, convirtiéndose en la más representativa e interesante en cuanto a la búsqueda del voto. Las otras cuatro secciones, localizadas en la Plaza Circular, el barrio de la Victoria y la Rubia, son zonas importantes, ya que destacaron también en el estudio de las elecciones municipales.

Adicionalmente, se aplicaron otros métodos de clasificación no supervisada como el algoritmo de partición óptimo y la aplicación de distintas distancias para cuantificar el índice de disimilaridad sin obtener resultados significativamente distintos.

Después de aplicar las diferentes metodologías, podemos concluir que si las secciones son relativamente estables en el tiempo, quizás bastaría con un análisis de correspondencias múltiples en frecuencias relativas para lograr nuestro objetivo de un modo más directo.

Estos resultados, a pesar de estar sustentados por análisis estadísticos adecuados, es preferible valorarlos en su justa medida, puesto que, en todo caso, las elecciones son actividades condicionadas por el comportamiento del ser humano que en muchas ocasiones resulta impredecible y, por tanto, difícil de modelizar.

Bibliografía

- [1] Ministerio del Interior información electoral – Resultados electorales – Área de Descargas. Recuperado en Mayo de 2023. URL: <https://infoelectoral.interior.gob.es/opencms/es/elecciones-celebradas/area-de-descargas/>
- [2] INE – Productos y servicios – Información estadística – Cartografía secciones censales y callejero de Censo Electoral. Recuperado en Junio de 2023. URL: https://www.ine.es/ss/Satellite?L=es_ES&c=Page&cid=1259952026632&p=1259952026632&pagename=ProductosYServicios%2FPYSLayout
- [3] González Duque, C. *Caracterización de las Secciones Censales de Valladolid en función de su comportamiento electoral*. Recuperado en Junio de 2023. URL: <https://uvadoc.uva.es/handle/10324/50493> Trabajo fin de grado, Grado en Estadística, Universidad de Valladolid, 2021.
- [4] Greenacre, M.J. *La práctica del análisis de correspondencias*. BBVA Foundation, 2008
- [5] Fernández Temprano, Miguel A. *Apuntes análisis de correspondencias*. 2021
- [6] Greenacre, M. *Correspondence Analysis in Practice, Third Edition* Chapman and Hall/CRC, 2016
- [7] Merola, G. M., Chen, G. *Projection sparse principal component analysis: An efficient least squares method*. Journal of Multivariate Analysis, 2019
- [8] Cuadras, Carles M. *Nuevos métodos de análisis multivariante*. CMC Editions, 2008.
- [9] Kassambara, A. *Practical guide to cluster analysis in R: Unsupervised machine learning, First Edition*. Multivariate Analysis I, 2019.
- [10] Le Roux, B., & Rouanet, H. *Geometric Data Analysis, From Correspondence Analysis to Structured Data Analysis*. Springer Science & Business Media, 2004.
- [11] Leish, F. *A toolbox for K-centroids cluster analysis, Computational Statistics & Data Analysis*, Vol.51, 2006.
- [12] Mora, R., & Ruiz-Castillo, J. *A Kullback-Leibler measure of conditional segregation*. Departamento de Economía, Universidad Carlos III de Madrid, Junio 2010.
- [13] Wickham, H. *ggplot2: Elegant Graphics for Data Analysis, Second Edition*. Springer-Verlag New York, 2016
- [14] Universitat de Valencia. Recuperado en Junio de 2023. URL: <https://www.uv.es/ceaces/multivari/cluster/CLUSTER2>

- [15] Sueur J. *RDocumentation kl.dist : Kullback-Leibler distance*. Recuperado en mayo de 2023. URL: <https://rdocumentation.org/packages/seewave/versions/2.2.0/topics/kl.dist>
- [16] Nenadic O. *RDocumentation ca : Simple correspondence analysis*. Recuperado en Mayo de 2023. URL: <https://rdocumentation.org/packages/ca/versions/0.71.1/topics/ca>
- [17] Nenadic O. *RDocumentation mjca: Multiple and joint correspondence analysis*. Recuperado en Mayo de 2023. URL: <https://rdocumentation.org/packages/ca/versions/0.71.1/topics/mjca>
- [18] Maechler M. *RDocumentation pam : Partitioning Around Medoids*. Recuperado en Mayo de 2023. URL: <https://www.rdocumentation.org/packages/cluster/versions/2.1.4/topics/pam>
- [19] Maechler M. *RDocumentation agnes : Agglomerative Nesting (Hierarchical Clustering)*. Recuperado en Junio de 2023. URL: <https://www.rdocumentation.org/packages/cluster/versions/2.1.4/topics/agnes>
- [20] Pebesma E. *RDocumentation sf library: Simple Features for R sf version 1.0-13*. Recuperado en Junio 2023. URL: <https://rdocumentation.org/packages/sf/versions/1.0-13>
- [21] Lind Pedersen, T. *RDocumentation ggplot2 library: Create Elegant Data Visualisations Using the Grammar of Graphics*. ggplot2 version 3.4.2. Recuperado en Junio de 2023. URL: <https://rdocumentation.org/packages/ggplot2/versions/3.4.2>
- [22] López Martínez, D. *Creación de un paquete de análisis de series temporales para RStudio*. Recuperado en Junio de 2023. URL: <https://uvadoc.uva.es/handle/10324/57958>. Trabajo fin de grado, Grado en Estadística, Universidad de Valladolid, 2022.
- [23] Castiñeiras Rella, J. *Asociación entre variables cualitativas; ejemplo práctico de análisis de correspondencias*. Recuperado en Junio de 2023. URL: <http://hdl.handle.net/10347/26306> Trabajo fin de grado, Grado en Matemáticas, Universidad de Santiago de Compostela, 2019.

Anexo I

Se han resumido las candidaturas en 6 tendencias políticas, este apartado se va a encargar de determinar que partidos formaron cada tendencia política con sus abreviaturas correspondientes:

Partido político	Tendencia política
ALIANZA POPULAR - PARTIDO DEMOCRATA POPULAR	AP-PP
CENTRO DEMOCRATICO Y SOCIAL	CENTRO
FALANGE ESPAÑOLA DE LAS JONS	EXT.DERECHA
ASOCIACION POLITICA FUERZA NUEVA	EXT.DERECHA
LIGA COMUNISTA (C. OBREROS SOCIAL.INTERNACIONALIS.)	EXT.IZQUIERDA
LIGA COMUNISTA REVOLUCIONARIA	EXT.IZQUIERDA
MOVIMIENTO COMUNISTA DE CASTILLA-LEON	EXT.IZQUIERDA
PARTIDO COMUNISTA DE ESPAÑA	EXT.IZQUIERDA
PARTIDO COMUNISTA DE ESPAÑA (MARXISTA-LENINISTA)	EXT.IZQUIERDA
PARTIDO SOCIALISTA	PSOE
PARTIDO SOCIALISTA OBRERO ESPAÑOL	PSOE
PARTIDO SOCIALISTA DE LOS TRABAJADORES	EXT.IZQUIERDA
UNION DE CENTRO DEMOCRATICO	CENTRO
UNIFICACION COMUNISTA DE ESPAÑA	EXT.IZQUIERDA
COALICION POPULAR	AP-PP
COALICION IZQUIERDA UNIDA	EXT.IZQUIERDA
MESA PARA LA UNIDAD DE LOS COMUNISTAS	EXT.IZQUIERDA
PARTIDO NACIONALISTA DE CASTILLA Y LEON	OTROS
PARTIDO OBRERO SOCIALISTA INTERNACIONALISTA	EXT.IZQUIERDA
UNIDAD POPULAR REPUBLICANA	EXT.IZQUIERDA
PARTIDO POPULAR	AP-PP
PARTIDO REFORMISTA DEMOCRATICO	CENTRO
PARTIDO TRABAJADORES DE ESPAÑA-UNIDAD COMUNISTA	EXT.IZQUIERDA
PARTIDO COMUNISTA DE LOS PUEBLOS DE ESPAÑA	EXT.IZQUIERDA
AGRUPACION RUIZ-MATEOS	OTROS
PARTIDO HUMANISTA	OTROS
LOS VERDES ECOLOGISTAS	OTROS
UNIDAD CENTRISTA-P.E.D.	CENTRO
LOS ECOLOGISTAS	OTROS
PARTIDO DE LA LEY NATURAL	OTROS
UNION CASTELLANISTA	OTROS
MONARQUIA CRISTIANA ESPAÑOLA SOCIALDEMOCRATA	OTROS
UNIDAD REGIONALISTA DE CASTILLA Y LEON	OTROS
TIERRA COMUNERA-PARTIDO NACIONALISTA CASTELLANO	OTROS
PARTIDO DE EL BIERZO	OTROS
UNION DEL PUEBLO LEONES	OTROS
COALICION POR UN NUEVO PARTIDO SOCIALISTA	EXT.IZQUIERDA

Partido político	Tendencia política
FALANGE ESPAÑOLA AUTENTICA	EXT.DERECHA
PARTIDO REVOLUCIONARIO DE LOS TRABAJADORES	OTROS
TIERRA COMUNERA-PARTIDO NACIONALISTA CASTELLANO	OTROS
UNION CENTRISTA	CENTRO
ESPAÑA 2000	OTROS
FALANGE ESPAÑOLA INDEPENDIENTE-FALANGE 2000	EXT.DERECHA
PARTIDO DEMOCRATA ESPAÑOL	OTROS
CANDIDATURA INDEP. PARTIDO DE CASTILLA Y LEON	OTROS
PARTIDO CARLISTA	OTROS
CIUDADANOS EN BLANCO	OTROS
DEMOCRACIA NACIONAL	EXT.DERECHA
IZQUIERDA REPUBLICANA	EXT.IZQUIERDA
MOVIMIENTO SOCIAL REPUBLICANO	EXT.IZQUIERDA
PART. CANNABIS POR LA LEGALIZACION Y NORMALIZACION	OTROS
PARTIDO DEMOCRATA ESPAÑOL	EXT.DERECHA
PARTIDO FAMILIA Y VIDA	OTROS
ALIANZA NACIONAL	EXT.DERECHA
ALTERNATIVA ESPAÑOLA	EXT.DERECHA
CIUDADANOS-PARTIDO DE LA CIUDADANIA	CENTRO
PARTIDO ANIMALISTA CONTRA EL MALTRATO ANIMAL	OTROS
POR UN MUNDO MAS JUSTO	OTROS
SOLIDARIDAD Y AUTOGESTION INTERNACIONALISTA	OTROS
TIERRA COMUNERA	OTROS
UNION PROGRESO Y DEMOCRACIA	CENTRO
EQUO	OTROS
RECORTES CERO-GRUPO VERDE	OTROS
UNIDAD POPULAR: IZQUIERDA UNIDA	EXT.IZQUIERDA
PODEMOS	EXT.IZQUIERDA
VOX	EXT.DERECHA
CIUDADANOS DE CENTRO DEMOCRÁTICO-	CENTRO
CANDIDATURA INDEP	
PARTIDO LIBERTARIO	OTROS
RECORTES CERO-GRUPO VERDE	OTROS
PARTIDO COMUNISTA DE LOS TRABAJADORES DE ESPAÑA	EXT.IZQUIERDA

Anexo II: Tabla resumen con las secciones representativas de cada una de las trece convocatorias en la comparación análisis clúster y análisis de correspondencias simple

	1982	1986	1989	1993	1996	2000	2004	2008	2011	2015	2016	2019-04	2019-11	Resumen
0304														2
0305														2
0306														2
0437														3
0602														1
0603														5
0608														2
0609														1
0610														2
0617														5
0706														2
0715														1
0802														2
0804														1
0805														1
0809														1
0841														1
0910														2
1001														1
1002														4
1025														4
1027														3
1028														3
1033														6
1034														2

	1982	1986	1989	1993	1996	2000	2004	2008	2011	2015	2016	2019-04	2019-11	Resumen
1035														2
1039														1
1042														1
1043														2
1102														2
1103														1
1105														1
1109														2
1110														3
1111														3
1112														1
1113														1
1117														1
1121														1
1125														1
1126														3
1127														5
1128														2
1129														2
1131														1
1136														1
1140														3
1203														1

Anexo III: Tabla resumen con las secciones representativas de cada una de las trece convocatorias en la comparación análisis clúster y análisis de correspondencias en tantos por 100. Se incluyen el promedio del número de votantes en los años electorales donde resultaron representativas.

	Votantes	1982	1986	1989	1993	1996	2000	2004	2008	2011	2015	2016	2019-04	2019-11	Resumen
0216	420	■													1
0304	554										■	■			2
0305	912							■	■	■					3
0306	862				■										1
0437	1156							■	■			■			3
0602	792	■	■									■			3
0603	970			■	■						■				3
0608	700			■	■										2
0609	424										■				1
0610	842								■						1
0617	736	■		■	■	■	■			■					5
0706	690		■	■											2
0707	780			■											1
0801	892	■													1
0802	1176			■						■					2
0804	1110	■													1
0805	922							■							1
0809	685		■	■											2
0813	997						■								1
0841	665		■												1
0910	612									■					1
1001	388										■		■		2
1002	584				■	■			■				■	■	4
1005	1077													■	1
1025	768			■	■	■	■								4

	Votantes	1982	1986	1989	1993	1996	2000	2004	2008	2011	2015	2016	2019-04	2019-11	Resumen
1027	726														4
1028	671														2
1033	1068														6
1034	857														3
1035	1102														2
1039	952														1
1042	1200														1
1043	809														2
1102	1291														2
1103	816														1
1109	449														1
1110	870														3
1111	635														3
1112	531														2
1113	643														1
1117	768														2
1121	1215														1
1125	1147														2
1126	1029														3
1127	510														4
1128	694														1
1129	813														2
1140	745														2

Anexo IV: Tabla resumen con las secciones representativas de cada una de las trece convocatorias en la comparación análisis clúster y análisis de correspondencias sin resumir candidaturas.

	1982	1986	1989	1993	1996	2000	2004	2008	2011	2015	2016	2019-04	2019-11	Resumen
0305														2
0306														1
0437														3
0602														1
0603														5
0608														2
0609														1
0610														2
0617														4
0631														1
0706														3
0707														1
0715														1
0802														2
0804														1
0805														1
0809														2
0813														1
0910														1
1001														1
1002														4
1013														1
1025														4
1027														2
1028														4

	1982	1986	1989	1993	1996	2000	2004	2008	2011	2015	2016	2019-04	2019-11	Resumen
1033														5
1034														2
1035														2
1039														1
1042														1
1043														2
1102														2
1103														1
1105														1
1109														2
1110														3
1111														3
1112														1
1113														1
1117														2
1121														1
1126														3
1127														4
1129														2
1131														1
1136														1
1140														2
1203														1

Anexo V. Aplicación de análisis clúster jerárquico.

En este apartado, se va a resumir la aplicación de dicho análisis para las trece convocatorias, tal y como se explicó en la (Sección 2.2.1). El estudio se realizará con los resultados obtenidos por el dendrograma.

Según lo expuesto en la sección 4.3 el método que obtiene mayor coeficiente de agregación es el método Ward por lo que será el que utilizemos para la resolución.

El procedimiento que se seguirá para la selección de secciones significativas consistirá en la extracción de aquellas secciones que formaban parte del clúster en el momento en el que se anexiona con la sección que contiene la distribución global de la convocatoria.

Vamos a presentar una tabla resumen que contenga aquellas secciones que se obtuvieron como representativas para las trece convocatorias. En la tabla A5.1 se muestran las secciones que fueron representativas en al menos, dos convocatorias.

Distrito	Sección	Convocatoria
3	6	1982,1996
4	37	2004,2016,2019-04
6	17	2000,2004
10	28	1982,1996
10	33	1993,2011,2015
11	27	1982,2011,2019-04
11	28	1982,1993

Tabla A5.1 Tabla resumen con las secciones obtenidas en el análisis jerárquico para todas las convocatorias

Hacemos notar que, en la aplicación del análisis, en algunas convocatorias los resultados de secciones principales fueron una única sección, puesto que, en el momento de la anexión el clúster se encontraba vacío.

Únicamente existen tres secciones que son representativas en tres convocatorias. La primera fue **la sección 37 del distrito 4**, la segunda **la sección 33 del distrito 10** y la tercera **la sección 27 del distrito 11**. Todas ellas resultaron representativas en el análisis de correspondencias múltiples realizado para las últimas cinco convocatorias (Sección 4.2). Además, las dos últimas secciones resultaron representativas en el análisis clúster no jerárquico. Hay que añadir que **la sección 33 del distrito 10** también resultó representativa en el análisis clúster jerárquico de las elecciones municipales [3] pág. 55.