



Universidad de Valladolid

Facultad de Ciencias

TRABAJO FIN DE GRADO

Grado en Estadística

**Estudio de la abstención en los
resultados de las elecciones al
Congreso de los Diputados en España**

Autor:

Pablo Martín Gómez

Tutores:

M. Pilar Rodríguez del Tío

Miguel A. Fernández Temprano

Agradecimientos

Quiero agradecer su apoyo constante a mis tutores, María Pilar Rodríguez del Tío y Miguel Alejandro Fernández Temprano, su orientación y correcciones han sido imprescindibles durante todo el proceso.

También quiero dar las gracias a mi familia, amigos y compañeros por haber hecho este camino más llevadero.

Resumen

El objetivo de este trabajo es estudiar el efecto de la abstención en las elecciones generales al Congreso de los Diputados en España en las distintas convocatorias y su posible relación con otras variables de interés. Para ello, se han utilizado los resultados de las elecciones generales desde 1977 hasta 2019, empleando metodologías como el análisis de correspondencias y el análisis de la varianza. Estas técnicas fueron implementadas utilizando el lenguaje de programación R.

Palabras clave

Abstención, elecciones generales, España, análisis de correspondencias, ANOVA.

Abstract

The aim of this work is to study the effect of abstention in the general elections to the Congress of Deputies in Spain across different calls and their potential relationship with other variables of interest. To do this, the results of the general elections from 1977 to 2019 have been used, employing methodologies such as correspondence analysis and variance analysis. These techniques were implemented using the R programming language.

Key words

Abstention, general elections, Spain, correspondence analysis, ANOVA.

Índice general

1. Introducción	1
1.1. Asignaturas relacionadas	2
2. Metodología	3
2.1. ANOVA	3
2.1.1. Tipos de ANOVA	4
2.1.2. Hipótesis asumidas	4
2.1.3. Análisis post-hoc	5
2.1.4. Test de Kruskal-Wallis	5
2.2. Análisis de correspondencias	6
2.2.1. Conceptos básicos	6
2.2.2. Procedimiento	7
2.2.3. Elementos suplementarios	7
2.3. Modelos de regresión	7
2.3.1. Hipótesis asumidas	8
2.3.2. Construcción del modelo	9
3. Obtención y tratamiento de datos	11
3.1. Conjuntos de datos base	11
3.2. Conjunto de datos resultante	12
3.3. Clasificación de candidaturas	14
3.4. Corrección de errores	15
4. Descripción de los datos	17
5. Análisis de los datos	21
5.1. Relación entre abstención y convocatorias	21
5.1.1. Relación entre abstención y cambio de gobierno	25

5.1.2. Relación entre abstención y partido ganador	29
5.2. Relación entre abstención y tamaño de la provincia	30
5.3. Análisis de Correspondencias Simples por convocatorias	37
5.3.1. Convocatoria 2000	38
5.3.2. Convocatoria 2004	40
5.3.3. Convocatoria 2008	40
5.3.4. Convocatoria 2011	42
5.3.5. Convocatoria 2015	42
5.3.6. Conclusión del Análisis de Correspondencias	44
6. Conclusiones	45
Bibliografía	49
Apéndices	53
A. Clasificación de candidaturas	53
B. Resumen de información de las convocatorias	55
C. Código R	57
C.1. Lectura de datos	57
C.2. Preprocesamiento por convocatorias	57
C.3. Preprocesamiento global	61
C.4. Descriptivos	62
C.5. Abstención y convocatorias	63
C.6. Abstención y tamaño de la provincia	65
C.7. Análisis de Correspondencias Simples	66

Índice de figuras

4.1. Abstención a nivel nacional.	17
4.2. Abstención a nivel provincial.	18
4.3. Abstención a nivel municipal.	19
4.4. Abstención frente a tamaño de municipio.	19
4.5. Abstención frente a provincias.	20
5.1. Tabla ANOVA de abstención frente a convocatorias.	22
5.2. Test de Levene modificado.	22
5.3. Residuos frente a valores predichos.	23
5.4. Test de Duncan.	23
5.5. Tabla ANOVA en el Modelo 2	24
5.6. Residuos frente a valores predichos en el Modelo 2.	25
5.7. Test de Duncan en el Modelo 2.	25
5.8. Test Chi-cuadrado entre Cambio y Grupo.	26
5.9. Test de Fisher entre Cambio y Grupo.	26
5.10. Análisis discriminante en R	28
5.11. Análisis discriminante en Statgraphics	28
5.12. Test Chi-Cuadrado entre Gobierno y Grupo.	29
5.13. Test de Fisher entre Gobierno y Grupo.	29
5.14. Resultados ANOVA del primer modelo de regresión.	31
5.15. Gráfico del primer modelo de regresión.	31
5.16. Resultados ANOVA del segundo modelo de regresión.	32
5.17. Gráfico del segundo modelo de regresión.	32
5.18. Coeficiente R ² del tercer modelo de regresión.	33
5.19. Gráfico del tercer modelo de regresión.	33
5.20. Resultados ANOVA del tercer modelo de regresión.	34
5.21. Coeficiente R ² del cuarto modelo de regresión.	34

5.22. Gráfico de residuos del cuarto modelo de regresión.	34
5.23. Resultados ANOVA del cuarto modelo de regresión.	35
5.24. Gráfico del cuarto modelo de regresión.	35
5.25. Coeficiente R2 del quinto modelo de regresión.	36
5.26. Gráfico de residuos del quinto modelo de regresión.	36
5.27. Resultados ANOVA del quinto modelo de regresión.	36
5.28. Gráfico del quinto modelo de regresión.	37
5.29. Varianza explicada de la convocatoria del 2000.	38
5.30. Gráfico ACS de la convocatoria de 2000. Ejes 1 y 2.	39
5.31. Gráfico ACS de la convocatoria de 2000. Ejes 2 y 3.	40
5.32. Gráfico ACS de la convocatoria de 2004.	41
5.33. Gráfico ACS de la convocatoria de 2008.	42
5.34. Gráfico ACS de la convocatoria de 2011.	43
5.35. Gráfico ACS de la convocatoria de 2015.	43

Índice de tablas

5.1. Tabla Cambio contra Grupo	26
5.2. Intervalo de confianza al 95 % de la Odds Ratio Cambio y Grupo	26
5.3. Tabla Cambio contra Predicho	28
5.4. Tabla Gobierno contra Grupo	29
5.5. Intervalo de confianza al 95 % de la Odds Ratio Gobierno y Grupo	30
A.1. Clasificación de candidaturas	54
B.1. Resumen convocatorias	55

Capítulo 1

Introducción

La abstención electoral y sus implicaciones han suscitado considerables polémicas desde el establecimiento de la democracia [1].

El propósito de este trabajo es estudiar la distribución de las abstenciones en las elecciones generales al Congreso de los Diputados desde 1977 hasta 2019 en España. Los datos utilizados para llevarlo a cabo son los resultados de las elecciones asociados a las quince convocatorias que han tenido lugar en nuestro país en los últimos años. Se intentará dar respuesta a cuestiones como “¿La abstención es mayor en los municipios más pequeños que en las ciudades?”, “¿Ha habido un aumento de la abstención a lo largo de los años?” “¿Los votantes de izquierdas se abstienen más que los de derechas?” o “¿Qué relación guarda la abstención con el cambio de gobierno?”.

Para responder a estas preguntas, en primer lugar se utilizará el Análisis de la Varianza aplicado a las convocatorias, utilizando la abstención como variable respuesta, que nos permitirá agrupar las convocatorias en grupos con participación similar y extraer conclusiones al cruzar estos grupos con el cambio o permanencia de gobierno y el hecho de que gane un partido de izquierda o de derecha. También se utilizarán modelos de regresión que utilicen el censo como variable explicativa y la abstención relativa como respuesta para buscar relación entre el tamaño de las provincias y la participación electoral. Por último, se utilizará un Análisis de Correspondencias Simples para estudiar individualmente en cada convocatoria la relación entre la abstención y los bloques políticos.

En la memoria se comenzará detallando la metodología utilizada durante el estudio, se continuará explicando cómo se han obtenido y tratado los datos necesarios para el trabajo y

finalmente se realizará una descripción y un posterior análisis de estos datos.

1.1. Asignaturas relacionadas

A lo largo del Grado en Estadística, varias asignaturas han proporcionado las bases teóricas y prácticas para abordar el presente trabajo. A continuación se detallan las asignaturas más relevantes y su relación con el TFG:

- **Regresión y ANOVA:** Por un lado, en esta asignatura se sitúan los modelos lineales como herramienta básica en los procedimientos de regresión. Por otro lado, se explora en profundidad el Análisis de la Varianza, desde la validez del modelo hasta el análisis e interpretación de los resultados. Ambos puntos son una base fundamental en los análisis realizados en este trabajo.
- **Análisis Multivariante:** En esta asignatura se amplían los conocimientos relativos a los problemas multivariantes y sus técnicas. En concreto, el Análisis de Correspondencias Simples, método aprendido en esta asignatura, es el procedimiento empleado en el último análisis del TFG.
- **Computación estadística:** En esta asignatura se tiene el primer contacto con R, el software empleado en la mayor parte del trabajo, y se construyen las bases para poder utilizarlo con soltura.

Capítulo 2

Metodología

En este capítulo se describirán los distintos métodos empleados para el análisis del efecto de las abstenciones que tiene lugar en este trabajo. En la primera sección se explicará el análisis de la varianza, en la segunda sección se describirá el análisis de correspondencias simple (ACS) y en la última se hablará sobre los modelos de regresión.

2.1. ANOVA

El Análisis de la Varianza (ANOVA) es un método estadístico empleado para determinar si existen diferencias significativas de la media de la variable respuesta en los distintos grupos de una variable cualitativa o factor.[2] [3] [4].

Recibe este nombre porque para detectar las diferencias, compara la varianza dentro de cada uno de los grupos con la varianza entre los distintos grupos. Se rechazará la hipótesis del ANOVA, indicando diferencias entre las medias, cuando la varianza entre distintos niveles del factor sea significativamente mayor que dentro de los niveles.

Siendo X un factor y x_1, x_2, \dots, x_n sus niveles, se tiene que nuestra variable respuesta será $Y = \mu_i + \epsilon$, donde ϵ es una variable aleatoria normal de media 0 y varianza desconocida pero igual para todos los niveles. Bajo esta notación, el test ANOVA para un único factor comprobará si se puede aceptar la hipótesis:

- $H_0 : \mu_1 = \mu_2 = \dots = \mu_n$
- H_1 : No todas las μ_i son iguales

Este contraste se resuelve mediante el calculo de un estadístico F y el p-valor asociado a este sirve para rechazar o no la hipótesis nula anterior. [5]

2.1.1. Tipos de ANOVA

Podemos clasificar nuestro Análisis de la Varianza de varias formas. Según el número de factores:

- **Unifactorial:** solo se estudia un factor.
- **Multifactorial:** se estudia la variable respuesta en función de más de un factor.

Según la forma en la que se hayan seleccionado los niveles de los factores:

- **Efectos fijos:** todos los factores del análisis tienen sus niveles seleccionados de forma específica y como consecuencia no se pueden generalizar sus conclusiones a otros niveles no especificados previamente.
- **Efectos aleatorios:** los niveles de todos los factores son una muestra que se ha escogido de manera aleatoria sobre una población de más niveles.
- **Efectos mixtos:** algunos factores son de efectos fijos y otros de efectos mixtos. Este caso solo se puede dar en el ANOVA multifactorial.

2.1.2. Hipótesis asumidas

El ANOVA asume distintas hipótesis que hay que validar para poder interpretar los resultados de forma adecuada. Estas hipótesis son las siguientes:

- **Homocedasticidad:** La homocedasticidad (u homogeneidad de las varianzas) asume que todos los grupos tienen la misma varianza. Para comprobarlo, se puede utilizar el test de Levene o Bartlett. En caso de que no se cumpla la hipótesis se puede realizar una transformación logarítmica a los datos.

- **Independencia:** Las observaciones del análisis deben ser independientes entre sí. Si el muestreo se ha realizado de forma aleatoria, esto suele cumplirse. Para comprobarlo, basta con ver que no hay patrones o tendencias en los gráficos de residuos.
- **Normalidad:** Los residuos deben seguir una distribución normal. Cuando el tamaño de muestra es muy grande, no es una cuestión crucial. Para comprobarlo, se puede utilizar un test de normalidad como Shapiro-Wilk [6] o mediante un Q-Q plot. Una transformación logarítmica podría ayudar a que se cumpla la hipótesis.

2.1.3. Análisis post-hoc

Una vez realizado el Análisis de la Varianza, si se rechaza la hipótesis nula, podemos asegurar que hay diferencias entre los distintos niveles del factor y podemos realizar un análisis post-hoc que nos indique los grupos que son diferentes. Algunos de los análisis más utilizados son los siguientes [7]:

- **Test de Duncan:** Esta prueba se basa en comparar las medias de todos los grupos entre sí mediante comparaciones múltiples. Para empezar, calcula las diferencias entre todas las posibles combinaciones de medias de grupo y las ordena de menor a mayor. Aplicando un criterio de significancia, determina las que son estadísticamente significativas y según esta agrupa las medias en grupos que no difieran entre sí.
- **Test de Tukey:** El objetivo de esta prueba es el mismo que en el test de Duncan, pero su procedimiento varía. En primer lugar, calcula el intervalo de confianza simultáneo para todas las diferencias entre las medias de los grupos y compara cada una de las diferencias de media con dicho intervalo para determinar su significancia. Si la diferencia entre dos medias supera el intervalo de confianza, se considera significativa.

2.1.4. Test de Kruskal-Wallis

En caso de que las hipótesis del ANOVA no se cumplan, una alternativa es el test de Kruskal-Wallis. Este método no paramétrico, tiene la misma finalidad que el ANOVA, pero no asume normalidad en los datos ni homogeneidad de las varianzas.

Una desventaja frente al Análisis de la Varianza, es que no proporciona información sobre las medias de qué grupos son diferentes. Para averiguarlo, podemos comparar cada pare-

ja de categorías del factor con la prueba de Mann-Whitney y aplicar posteriormente alguna corrección de la significación como la de Bonferroni [8].

Si lo que se quiere es agrupar los niveles en grupos semejantes puede emplearse alguna técnica de Análisis Cluster como el método de las K-Medias.

2.2. Análisis de correspondencias

El análisis de correspondencias [9] es un método descriptivo que se utiliza para reducir la dimensionalidad de una tabla de contingencia y poder obtener nubes de puntos en dos o más dimensiones fácilmente interpretables. A diferencia del Análisis de Componentes Principales o del Análisis Factorial, las variables utilizadas en este análisis son variables cualitativas, donde cada una de sus categorías se representará como un punto en el gráfico que permite relacionarlas. En este gráfico, la cercanía entre categorías indica un mayor grado de asociación.

Dependiendo del número de variables que se quieren relacionar, podemos distinguir entre Análisis de Correspondencias Simples (ACS) o Análisis de Correspondencias Múltiple. En ACS, solo se relacionan dos variables mientras que en el ACM se relacionan tres o más. En este trabajo únicamente se utilizará ACS, por lo que la explicación se centrará en él.

2.2.1. Conceptos básicos

Se empezará definiendo unos términos básicos para entender mejor el método. Partiendo de una tabla de contingencia, se entiende como perfiles fila y perfiles columna las distribuciones condicionadas de las filas y las columnas. En ACS, filas y columnas son tratadas de igual forma. Como cada uno de estos perfiles va a estar formado por una distinta cantidad de individuos, resulta lógico ponderarlos para asignarle más peso a los perfiles más poblados.

La distancia empleada en medir el grado de asociación entre los perfiles es la distancia χ^2 , que cumple la propiedad de equivalencia distribucional. Esta propiedad indica que si dos perfiles fila son iguales, al agruparlas, no se modificarán las distancias entre el resto de categorías de los perfiles columna. Lo mismo ocurre con los perfiles columna.

El último elemento a definir es la inercia total. Esta se entiende como la suma de los cuadrados de las distancias ponderadas desde cada uno de los perfiles al centro de gravedad. La inercia total se utiliza como medida de calidad de representación de los ejes, cuanto mayor

sea la inercia, mejor representación obtendremos.

2.2.2. Procedimiento

En primer lugar, el ACS construye una matriz de distancias utilizando la distancia χ^2 , que calibrará la magnitud de las diferencias entre la tabla de datos analizada y una supuesta tabla en la que las variables no tienen relación. Posteriormente, se diagonalizará esta matriz para obtener los vectores y autovalores que definirán los nuevos ejes sobre los que se proyectará la nube de puntos inicial. Estos nuevos ejes formarán un sistema de coordenadas asociado a las filas y columnas de la tabla de contingencias inicial, en el que se reflejarán las relaciones entre dichas filas y columnas.[10]

El siguiente paso es elegir el número de dimensiones a extraer (suele ser dos para poder hacer gráficos bidimensionales). El número de dimensiones elegido suele depender de la cantidad de inercia total que se explica con dichos factores. Por último se proyectarán los perfiles sobre los ejes elegidos y se interpretará el gráfico.

En este trabajo se utilizan gráficos simétricos, en los que tanto las filas como las columnas se representan en sus coordenadas principales, lo que permite que la distancia entre perfiles filas y columnas sean interpretables.

2.2.3. Elementos suplementarios

Los elementos suplementarios son filas o columnas incluidos en el análisis de forma posterior o en distintas condiciones que los elementos activos. Por tanto, no se han utilizado para calcular las proyecciones del ACS.

En este trabajo se utilizarán para mejorar la interpretación de la variable de interés (las abstenciones) mediante su proyección en el gráfico.

2.3. Modelos de regresión

El modelo de regresión [11] es un método que se utiliza cuando se quiere evaluar la relación entre una variable dependiente (denominada Y) y una serie de variables independientes,

también conocidas como variables explicativas o regresoras (con notación X_1, X_2, \dots, X_n). Dependiendo de la función que formen las variables independientes para explicar la variable dependiente, el modelo de regresión recibe distintos nombres. En este trabajo nos centraremos en el modelo de regresión lineal, que puede ser expresado como:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \varepsilon \quad (2.1)$$

En esta fórmula, epsilon representa un término aleatorio, y los betas son los parámetros del modelo que miden la influencia que los regresores tienen sobre la variable independiente.

Si el objetivo del modelo es evaluar como afecta el cambio en las variables independientes sobre la variable dependiente, se le denomina modelo con fines explicativos. Si en cambio, el objetivo es tratar de estimar el valor de Y cuando el conjunto de regresores toma un valor, se le conoce como modelo con fines predictivos.

2.3.1. Hipótesis asumidas

El modelo de regresión asume ciertas hipótesis para ser utilizado y poder interpretar sus resultados con exactitud.

- **Linealidad:** La media de la variable respuesta respecto a las variables explicativas debe guardar una relación lineal. Si esto se cumple, los residuos tendrán una media 0.
- **Homocedasticidad:** Según esta hipótesis, la varianza de la variable respuesta no cambia con los valores de las variables explicativas, lo que significa que los residuos tendrán una varianza homogénea. Para comprobarlo puede utilizarse un diagrama de dispersión de los residuos y verificar que no existen patrones.
- **Independencia:** Las correlaciones entre los residuos son nulas, es decir, no se ven influenciadas por los residuos de otras observaciones. La prueba de Durbin-Watson evalúa si hay autocorrelación entre los residuos.
- **Normalidad:** Los residuos deben seguir una distribución normal, y como se ha indicado en las hipótesis anteriores, su media debe ser 0 y su varianza constante. Para verificarlo, se puede utilizar el test de Shapiro-Wilk.

2.3.2. Construcción del modelo

Para construir el modelo, se debe elegir cuales de los regresores sirven realmente para explicar la variable dependiente. Para ello se pueden emplear distintas técnicas:

- **Pasos hacia delante:** este método introduce las variables de una en una si verifican que cumplen una condición, como puede ser una significancia de nivel menor que 0.05.
- **Pasos hacia atrás:** el modelo originalmente incluye todas las variables y se irán extrayendo de una en una, aquellas que no cumplan una condición.
- **Por pasos:** es una combinación de las dos técnicas anteriores donde las variables pueden introducirse o eliminarse. Requiere mayor coste computacional pero obtiene mejores resultados.

Para estimar el valor de los parámetros, se utilizará el método de mínimos cuadrados. El objetivo es minimizar la suma de los cuadrados de las diferencias entre los valores observados de la variable dependiente y sus valores predichos.

Capítulo 3

Obtención y tratamiento de datos

En el presente capítulo se explicará el proceso de extracción de los datos utilizados para este trabajo, así como los cambios y correcciones que haya sido necesario realizar en el preprocesamiento previo al análisis.

3.1. Conjuntos de datos base

Los datos base con los que se ha trabajado en este estudio fueron extraídos del Área de Descargas del sitio web del Ministerio del Interior [12]. Desde esta página se pueden seleccionar el tipo de elecciones en el que estamos interesados, en este caso, las elecciones al Congreso de los Diputados. Por último, se pueden descargar los archivos con información por provincias o por municipios. Se han utilizado los datos de municipios, pudiéndose agrupar estos por provincias para algunos análisis posteriores.

Cabe destacar, que al utilizar los datos a nivel municipal, quedan excluidos de este trabajo y de los posteriores análisis los votos pertenecientes a los españoles residentes en el extranjero, es decir, el voto CERA. Esto significa que solo se trabajará con los votos de los residentes en España y como consecuencia, la abstención a nivel nacional en cada una de las convocatorias puede sufrir ligeras variaciones a la encontrada en otras fuentes que consideran también el voto CERA, como puede ser Wikipedia[13], sobre todo en los últimos años en los que el número de residentes en el extranjero ha aumentado, pero al querer utilizar los datos a nivel municipal o provincial en algunos análisis, se ha optado por excluir los votos que no se asocian a ningún municipio ni provincia.

Las quince convocatorias con las que se trabajará en este estudio son: junio de 1977, marzo de 1979, octubre de 1982, junio de 1986, octubre de 1989, junio de 1993, marzo de 1996, marzo del 2000, marzo de 2004, marzo de 2008, noviembre de 2011, diciembre de 2015, junio de 2016, abril de 2019 y noviembre de 2019. A partir de ahora se referirá a ellas solamente por su año, con excepción de abril de 2019 y noviembre del mismo año que serán 2019A y 2019B, respectivamente.

Para cada una de estas convocatorias se dispone de un archivo comprimido con 7 documentos distintos. Para este estudio solo se utilizarán tres de ellos, cuya información se detalla a continuación.

- **Fichero de candidaturas:** Este archivo guarda información de todas las candidaturas que se presentaron en una convocatoria. De él, son de interés las variables que se refieren al código único de cada candidatura, la denominación de esta candidatura y las siglas que la abrevian.
- **Fichero de datos globales de ámbito municipal:** En este archivo se recoge información general de los votos para cada municipio. Identifica cada municipio con dos variables: un identificador de provincia y uno de municipio dentro de esa provincia. Estos códigos son los mismos identificadores que utiliza el Instituto Nacional de Estadística. Para cada municipio, guarda el censo registrado en esa convocatoria, el número de votos en blanco, los votos nulos y los votos totales a candidaturas.
- **Fichero de datos a candidaturas de ámbito municipal:** En este archivo se identifican los municipios de la misma manera que en el anterior, mediante un código de provincia y otro de municipio. Para cada uno de estos, se recoge información de cuantos votos ha recibido cada candidatura en él. Las candidaturas están identificadas por el código único mencionado en el primer fichero.

3.2. Conjunto de datos resultante

Ahora que ya están descritos los conjuntos de datos utilizados, se detalla el proceso a través del cual se obtiene el conjunto definitivo para este trabajo.

En primer lugar, se trata de unir la información de interés de cada uno de los tres archivos utilizando las claves primarias. La tecnología empleada es *R*, y en este software, esto se

Obtención y tratamiento de datos

puede hacer utilizando la orden *merge*. Por un lado, se unirá el fichero de candidaturas con el fichero de datos de candidaturas de ámbito municipal a través de la columna de Código de candidaturas, para saber cuantos votos se llevó cada candidatura en cada municipio. Por otro lado, se unirán el fichero de datos globales y el de datos a candidaturas a través de los códigos identificativos de cada municipio. De esta forma, se tienen 15 conjuntos de datos (uno por convocatoria) con toda la información necesaria para cada municipio.

En segundo lugar, antes de unir los 15 conjuntos de datos en uno solo, se harán otras transformaciones previas. Cada municipio está a la vez dividido en distritos, pero para este análisis solo nos interesa el nivel de municipio. Para ello, nos fijaremos en el distrito 99 de cada municipio, que contiene información de la suma de todos los distritos. Eliminando el resto de distritos, podemos olvidarnos de esta columna.

A continuación, creamos la variable de mayor interés en este estudio, las abstenciones. Esta variable no aparece en nuestros datos originales pero se puede generar fácilmente. Para cada municipio tenemos información del número de personas censadas, de cuantos votos han ido a candidaturas, de cuantos votos blancos ha habido y de cuantos nulos. Con estos datos, es fácil calcular las abstenciones que ha habido en cada municipio como se detalla a continuación.

$$Abstenciones = Censo - VotosCandidaturas - VotosBlancos - VotosNulos \quad (3.1)$$

Para poder comparar los votos entre distintos municipios, se trabajará en porcentajes. Todos los votos a candidaturas, así como las abstenciones, serán relativos al tamaño del municipio.

Como se había mencionado anteriormente, el año en el que se realiza la convocatoria es suficiente para identificarla. Por eso se eliminará el mes de convocatoria tras haber cambiado el año de las convocatorias de 2019 a 2019A y 2019B.

Por último, se agregará una columna que discretice el tamaño de los municipios para poder utilizarlo como variable categórica en algunos análisis. Para ello, se dividirá en tres categorías según una clasificación aportada por el Ministerio de Agricultura, Pesca y Alimentación para un estudio sobre los municipios rurales[14]:

- **Municipios pequeños:** menos de 5.000 censados.
- **Municipios medianos:** entre 5.000 y 30.000 censados.

- **Municipios grandes:** más de 30.000 censados.

Con estas modificaciones, ya podemos juntar los 15 archivos en uno solo, a través de la orden `rbindlist` de R [15].

3.3. Clasificación de candidaturas

Una de las partes más costosas del preprocesamiento de los datos ha sido la clasificación de candidaturas. Uno de los objetivos del análisis es relacionar las abstenciones con las distintas tendencias políticas, para ello se clasificarán los partidos en los siguientes grandes bloques: Derecha, Izquierda, Centro, Extrema Derecha, Extrema Izquierda y Nacionalistas.

Para cada una de las 15 convocatorias, existen más de 150 candidaturas, en ocasiones más de 300. Además, una misma candidatura no tiene el mismo código de una convocatoria a la siguiente. Por este motivo, la clasificación manual de todas ellas es completamente inviable.

Para este trabajo, se optó inicialmente por considerar solo las candidaturas con un porcentaje significativo de votos a nivel nacional en esa convocatoria (superiores al 5%) [13]. Puede parecer una simple solución, pero el principal inconveniente que se encontró, es que un partido a nivel nacional, puede tener incluso 20 candidaturas en cada una de las provincias. Por ejemplo, el Partido Socialista Obrero Español (PSOE) en las elecciones del 2008, se dividía a su vez en 15 candidaturas (Partido Socialista de Euskadi, PSOE de Andalucía, PSOE de Cataluña, Partido Socialista de Galicia...) cada una de ellas con unas siglas y códigos de candidaturas distintas. Si solo se incluía el PSOE general, esto hacía que la gran mayoría de municipios apenas tuviera votos en el bloque Izquierda.

Para conseguir tener unos datos sin demasiados ausentes, se utilizaron técnicas de búsquedas de patrones en las denominaciones de las candidaturas. Por ejemplo, para el caso descrito en el párrafo anterior, considerar todas las candidaturas que contengan los patrones PSOE, P.S.O.E o SOCIALISTA. A pesar de esto, muchas provincias apenas tenían representación, y fue necesario buscar manualmente aquellas con un NA, e investigar que candidatura en dicha provincia y para esa convocatoria era la que representaba a ese bloque político. Nótese que se habla de provincias y no de municipios, pues las candidaturas suelen variar por provincias, y todos los municipios de una, están representados por la misma candidatura.

Otro problema encontrado fue el caso de los bloques extremos, la extrema izquierda y la

extrema derecha. En muchas convocatorias no tuvieron una relevancia superior al 5% de los votos a nivel nacional, y en ellas, las tablas quedaban pobladas de valores ausentes. Estos NA provocan problemas en los análisis posteriores de modo que se incluyeron candidaturas que a nivel nacional pudieran ser menos significativas. Aun así, hubo casos en los que los valores ausentes se sustituyeron por una cantidad muy próxima al 0 para evitar problemas de software.

Algo similar ocurrió con el bloque nacionalista. En la mayoría de las provincias, estas candidaturas no tienen relevancia a nivel nacional, pero tienen su importancia en provincias de las Comunidades Autónomas de Cataluña, Galicia o País Vasco, por tanto se incluyeron finalmente. Esto provocaba que en el resto de provincias, el bloque nacionalista estuviera lleno de valores ausentes. Se planteó incluir las candidaturas nacionalistas de más provincias como Partido Leonés o Convergencia Democrática de Mallorca, pero su influencia era tan baja que se optó por sustituir los valores ausentes por valores próximos al 0 en todos estos municipios.

La clasificación final de qué partidos se han incluido en cada bloque, puede encontrarse en el Apéndice A. En estas tablas solo se indican los partidos, no cada una de las candidaturas para cada partido al considerarse redundante e innecesaria.

Con esta sección se ha pretendido explicar la toma de decisiones respecto a la clasificación de candidaturas y mostrar que está lejos de ser algo determinado (de forma única) debido a la cantidad masiva de datos y el trabajo que llevaría la clasificación manual de todos ellos. Aun así, las distintas técnicas utilizadas, permiten que los partidos más significativos estén bien representados y que los resultados de los análisis sean suficientemente fiables. A mayores indicar, que se ha prestado aún más atención a aquellas candidaturas que se han analizado en detalle en el Capítulo 5, para obtener unos mejores resultados.

3.4. Corrección de errores

En esta última sección, se detectarán errores en los datos para garantizar la fiabilidad de los resultados en los posteriores capítulos.

Mediante unos simples estadísticos descriptivos que muestren el máximo y el mínimo de las variables numéricas, se pueden observar las siguientes incongruencias:

- **Abstenciones relativas negativas:** En algunas observaciones, el porcentaje de abstenciones de un municipio resulta ser menor que 0. Comprobando estas, se ve que el error

proviene de que en los datos originales, el número de censados en el municipio es menor que el número de votos recibidos por las candidaturas. Como la abstención es una variable que se ha construido a través de la fórmula (3.1), esto provoca que esta sea negativa. Como solo hay 20 observaciones con este error y provenientes de municipios pequeños, se opta por eliminarlas del estudio.

- **Abstenciones relativas mayores que 1:** En el caso opuesto, el porcentaje de abstenciones de algunos municipios es mayor que 1. De forma similar, se comprueba que el error proviene de que en una observación, los votos a candidaturas son negativos. Esto provoca que al sumarse al censo, haya más abstenciones que personas censadas. Como solo ocurre en 2 observaciones, igualmente se eliminan.
- **Caracteres problemáticos:** Los nombres de algunas candidaturas como Ciudadanos (C's) o de algunos municipios (L'Ametlla), contienen caracteres que impide que el software interprete correctamente los datos. Por esto se han reemplazado todas las comillas y ? en todos los archivos de datos para garantizar una lectura adecuada.

Hagáse notar que es crucial detectar estos errores para no obtener resultados incoherentes en los siguientes análisis.

Capítulo 4

Descripción de los datos

En el siguiente capítulo se presenta una breve descripción de las variables de interés del trabajo. Estos datos constituyen el fundamento sobre el cual se basarán los análisis posteriores y en este capítulo describiremos la naturaleza de estos datos y los echaremos un primer vistazo general.

Nos centraremos en la variable de mayor interés, la abstención relativa. Se trata de una variable continua que toma valores entre 0 y 1 e indica que porcentaje de la población de un municipio no participó en cada una de las convocatorias. Si agrupamos la abstención de los municipios por convocatorias, podemos ver la evolución de la participación en las elecciones desde 1977 a 2019 (Figura 4.1).

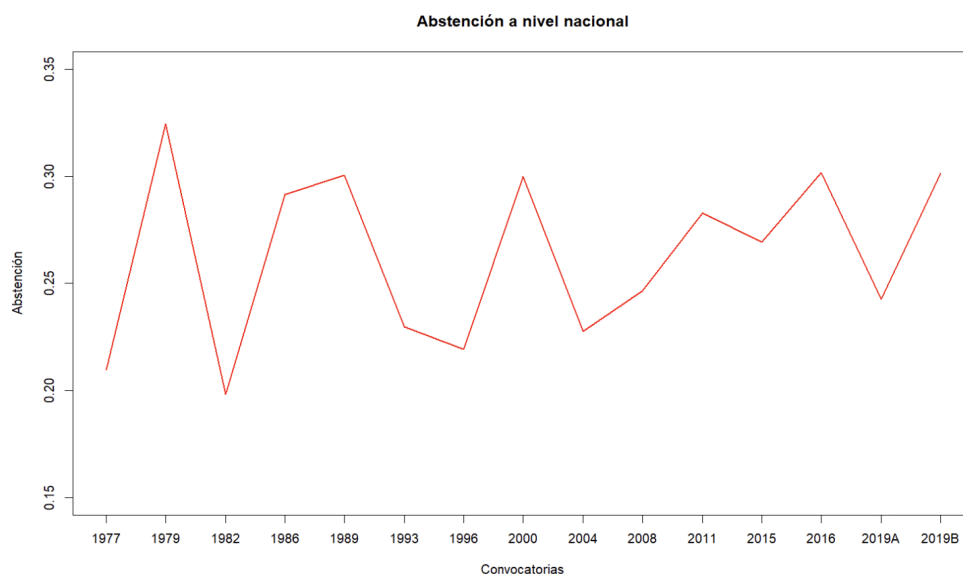


Figura 4.1: Abstención a nivel nacional.

En el gráfico anterior observamos que a nivel nacional la abstención se mantiene entre el 19% y el 33%, alcanzando su mínimo en 1982 con un 19.83% y su máximo en 1979 con un 32.44%.

Podemos ir más allá y representar estos datos en un diagrama de cajas múltiples en el que cada observación representa una provincia (Figura 4.2).

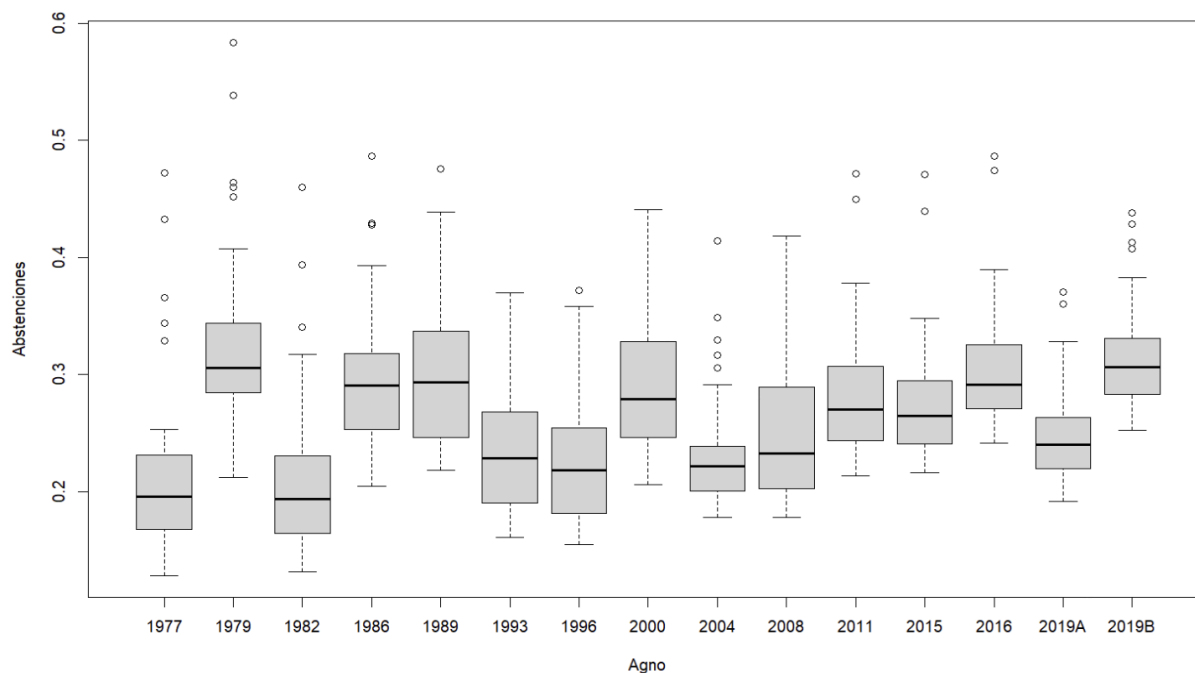


Figura 4.2: Abstención a nivel provincial.

En este gráfico observamos que a pesar de las medias obtenidas a nivel nacional, existen provincias en los que la abstención fue más que el doble de lo obtenido en todo el país para esa convocatoria, obteniendo su máximo en la provincia de Salamanca en 1979 con una abstención del 58.37%.

También se puede observar el mismo gráfico a nivel de municipios, siendo cada observación uno de ellos (Figura 4.3).

En este caso los resultados son mucho más extremos llegando a cubrir todo el rango de abstención. Se tienen municipios en los que la abstención es nula, y otros en los que es completa. Esto se explica en que existen municipios de tan solo 7 habitantes, donde es más fácil conseguir una participación total o inexistente. Cabe destacar que existe una fuerte asimetría en la cola derecha en todas las convocatorias, existiendo una gran cantidad de municipios que superan la abstención media.

Descripción de los datos

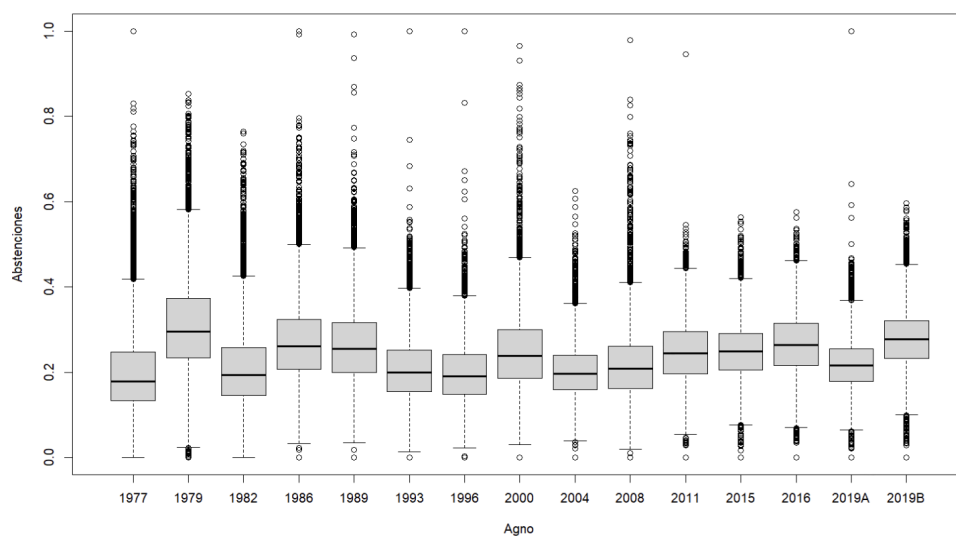


Figura 4.3: Abstención a nivel municipal.

También podemos estudiar la abstención en un gráfico de cajas múltiples que distinga el tamaño del municipio según la clasificación expuesta en la Sección 3.2 (Figura 4.4).

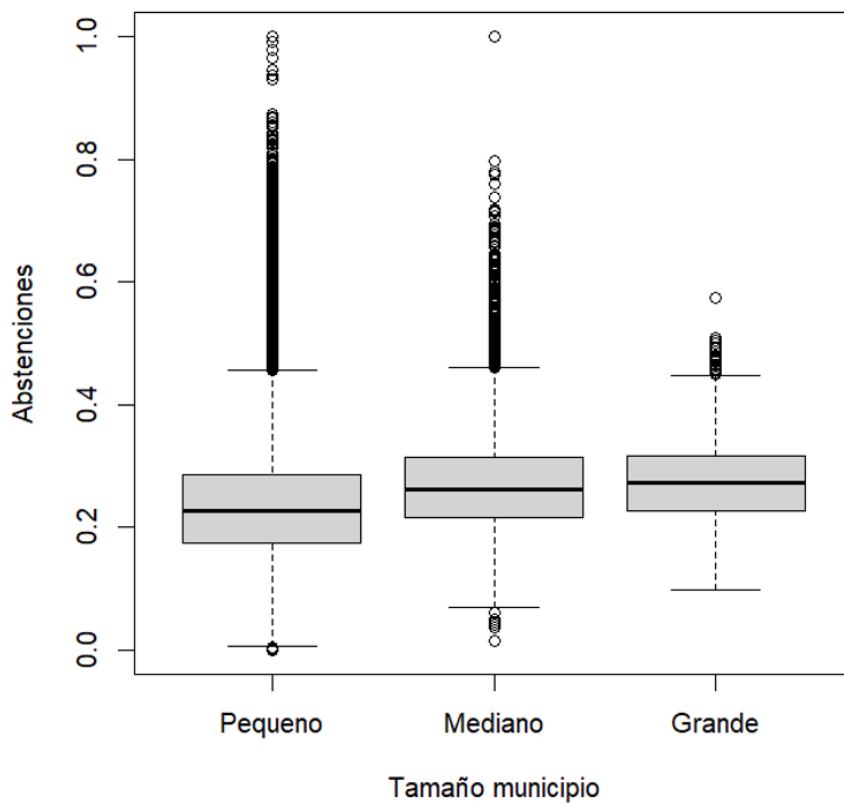


Figura 4.4: Abstención frente a tamaño de municipio.

Observando la Figura 4.4, vemos que existe una fuerte asimetría hacia la derecha en los tres niveles de tamaño de municipio. A su vez, se aprecia un nivel de abstención más bajo en los municipios pequeños que en los grandes y medianos, pero para confirmarlo se hará un análisis más profundo en el siguiente capítulo.

Por último, podemos tomar las convocatorias como observaciones y realizar un gráfico de cajas múltiples con cada provincia como un nivel distinto (Figura 4.5). Este gráfico no parece muy representativo, pero se puede intuir un nivel mayor de abstención en las provincias 32, 51 y 52, correspondientes a Lugo, Ceuta y Melilla [16]. En los análisis del siguiente capítulo, veremos como tratar con ellas.

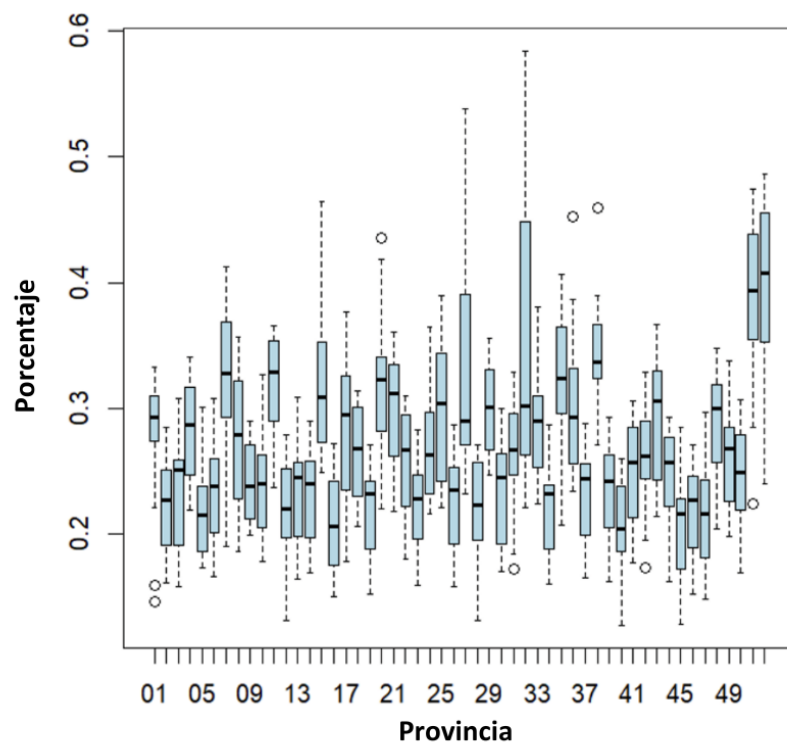


Figura 4.5: Abstención frente a provincias.

Capítulo 5

Análisis de los datos

En este capítulo se pretende responder a las preguntas planteadas en la introducción del trabajo. Para ello se intentará establecer una relación entre la abstención y las distintas variables descritas anteriormente. En primer lugar, se estudiará si la abstención cambia significativamente con las convocatorias y a continuación se estudiará la posible relación entre los niveles de abstención en las convocatorias y el cambio de gobierno en unas elecciones o el partido que gobierna tras las mismas. Posteriormente, se utilizará la población de las provincias para explicar la abstención de las mismas mediante un modelo de regresión. También se compararán los modelos de regresión que genera cada convocatoria para buscar relación entre ellos. Por último, exploraremos la relación entre la abstención y los distintos bloques políticos en algunas convocatorias.

5.1. Relación entre abstención y convocatorias

Para estudiar si la abstención varía significativamente entre las distintas convocatorias a las elecciones del Congreso de los Diputados, comenzaremos realizando un Análisis de la Varianza para ver si la convocatoria es un factor significativo. Este procedimiento comparará los valores medios de la variable Abstención para los 15 niveles de la variable Año (Convocatoria) para detectar posibles diferencias significativas entre ellos. Aunque inicialmente el objetivo era utilizar los municipios como unidades del análisis, la gran variedad y cantidad de estos hizo que finalmente se usaran las provincias, obteniendo así mejores resultados. La Figura 5.1 muestra un resumen de este análisis.

5.1. Relación entre abstención y convocatorias

```
          Df Sum Sq Mean Sq F value Pr(>F)
Agno      14  1.123  0.08018   26.04 <2e-16 ***
Residuals 765  2.356  0.00308
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figura 5.1: Tabla ANOVA de abstención frente a convocatorias.

Como el p-valor del estadístico F es menor que 0.001, existen diferencias significativas entre las medias de las abstenciones en las distintas convocatorias. Antes de realizar un análisis post-hoc para ver cuáles de las convocatorias son distintas, se validarán las hipótesis del ANOVA.

En primer lugar, se utilizará el test de Levene [17] modificado para asegurar la homogeneidad de las varianzas (Figura 5.2).

```
Levene's Test for Homogeneity of Variance (center = median)
  group    Df F value    Pr(>F)
  ---
  14      14  1.44559    0.1261
 765
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figura 5.2: Test de Levene modificado.

Como el p-valor del test es mayor que 0.1, no se puede asegurar que haya diferencias significativas entre las varianzas.

En segundo lugar, comprobaremos que no hay patrones o tendencias en el gráfico de los residuos frente a valores predichos por el modelo y que se distribuyen en torno al cero, tal y como muestra la Figura 5.3. Nótese en esta figura que ciertas observaciones tienen residuos muy elevados. Más adelante se tratará este asunto.

Validadas las hipótesis del modelo, realizaremos un análisis post-hoc. El objetivo es agrupar las convocatorias en grupos, de forma que las que pertenecen a un mismo grupo no tengan diferencias significativas en su abstención media. Una vez agrupadas, podremos cruzarlas con distintas variables como el cambio de gobierno en unas elecciones o el partido que gobierna tras esta, para intentar extraer conclusiones.

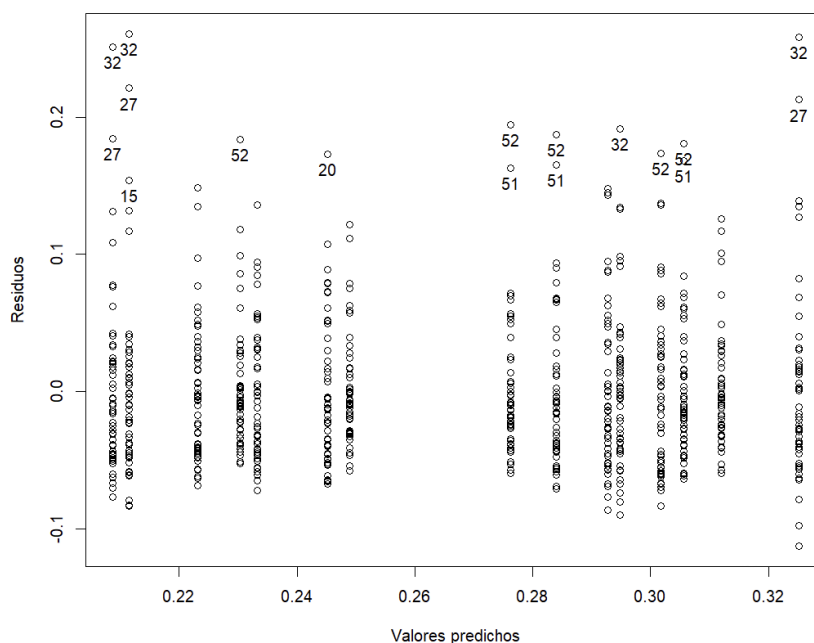


Figura 5.3: Residuos frente a valores predichos.

Para agrupar las convocatorias podemos utilizar distintos métodos. Por un lado, el test de Duncan [18] permite realizar comparaciones múltiples tras haber rechazado la hipótesis de un ANOVA. Con ayuda del software Statgraphics, obtenemos la Figura 5.4 que indica los grupos homogéneos para cada convocatoria.

Agno	Casos	Media	Grupos Homogéneos
1982	52	0.1983	A
1977	52	0.2097	AB
1996	52	0.2193	ABC
2004	52	0.2276	ABCD
1993	52	0.2297	BCD
2008	52	0.2465	CD
2019A	52	0.2425	D
2015	52	0.2692	E
2011	52	0.2840	EF
2000	52	0.2999	EFG
1986	52	0.2915	EFG
1989	52	0.3003	FG
2016	52	0.3016	FGH
2019B	52	0.3013	GH
1979	52	0.3244	H

Figura 5.4: Test de Duncan.

Según esta clasificación, podríamos dividir las convocatorias en dos grandes grupos. El primero incluiría los grupos homogéneos A, B, C y D, y el segundo los grupos E, F, G y H. Si nos fijamos en los valores límites que separan los dos grandes grupos (0.2425 y 0.2692)

5.1. Relación entre abstención y convocatorias

podemos describir el primer grupo como el conjunto de convocatorias en las que la abstención fue menor del 25 %, y el segundo grupo, las convocatorias en las que la abstención supera esta cifra.

Por otro lado, se puede utilizar un análisis cluster para agrupar las convocatorias en dos grupos. Para esto se utiliza el método *KMeans()* [19] de R, y los resultados son los mismos que los obtenidos por el test de Duncan.

Antes de intentar comprender que características comunes tienen estas convocatorias para estar agrupadas juntas, afinaremos algo más el modelo ANOVA. Notese que en la Figura 5.3, las observaciones con un error mayor suelen ser las mismas provincias en distintas convocatorias. Esto nos hace plantear un nuevo modelo ANOVA excluyendo estas provincias para comprobar si el modelo varía al eliminar los residuos más altos. Este nuevo modelo será igual que el anterior pero excluyendo las provincias con ID 51, 52, 32 y 27, correspondientes a Ceuta, Melilla, Ourense y Lugo, respectivamente.

```
          Df Sum Sq Mean Sq F value Pr(>F)
Agno      14  1.123  0.08018   38.10 <2e-16 ***
Residuals 765  2.356  0.00308
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figura 5.5: Tabla ANOVA en el Modelo 2

En la Figura 5.5 se comprueba que el test ANOVA sigue detectando diferencias significativas para el valor medio de la abstención en los distintos niveles de la variable *Convocatoria*. El gráfico de valores predichos frente a residuos, muestra ahora unos errores menores que en el anterior modelo (Figura 5.6).

Tras un análisis post-hoc utilizando el test de Duncan (Figura 5.7), se ve que los grupos que se forman son los mismos que se tenían en el modelo con todas las provincias, a pesar de que dentro de cada grupo hay más diferencias significativas que en el anterior análisis.

Esto nos acaba de confirmar que los grupos a utilizar en lo que queda de análisis son los siguientes:

- **Grupo A:** Convocatorias con una abstención inferior al 25 %, formado por las convocatorias: 1977, 1982, 1993, 1996, 2004, 2008 y 2019A.
- **Grupo B:** Convocatorias con una abstención superior al 25 %, formado por las convoca-

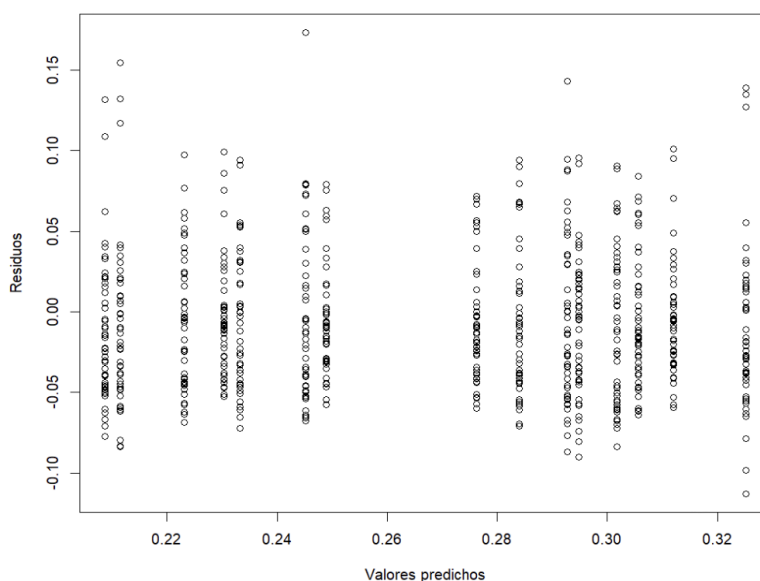


Figura 5.6: Residuos frente a valores predichos en el Modelo 2.

Agno	Casos	Media	Grupos Homogéneos
1982	48	0.1954	A
1977	48	0.2006	AB
1996	48	0.2157	BC
2004	48	0.2241	CD
1993	48	0.2258	CDE
2008	48	0.2417	DE
2019A	48	0.2432	E
2015	48	0.2685	F
2011	48	0.2774	FG
1986	48	0.2832	FGH
2000	48	0.2872	FGHI
1989	48	0.2906	GHI
2016	48	0.2982	HIJ
2019B	48	0.3055	IJ
1979	48	0.3121	J

Figura 5.7: Test de Duncan en el Modelo 2.

torias: 1979, 1986, 1989, 2000, 2011, 2015, 2016 y 2019B.

5.1.1. Relación entre abstención y cambio de gobierno

Una de las preguntas que nos planteábamos en la introducción de este trabajo, es si la abstención afecta al cambio de gobierno. Para responderla, se creará una tabla de contingencia donde las convocatorias son observaciones y las variables cruzadas serán la pertenencia al grupo A o B y si hay cambio o no de gobierno tras esas elecciones. La tabla con toda la información resumida para la siguiente parte del análisis junto con el partido que gobernó tras cada convocatoria se encuentra en el Apéndice B. La tabla cruzada obtenida es la Tabla 5.1

5.1. Relación entre abstención y convocatorias

	Grupo A	Grupo B
Cambio	5	1
No cambio	2	7

Tabla 5.1: Tabla Cambio contra Grupo

Lo primero que debemos preguntarnos es si existe independencia entre ambas variables. Para ello, utilizamos el test Chi-cuadrado[20], como muestra la Figura 5.8.

```
Pearson's Chi-squared test with Yates' continuity correction
data: Cambio and Grupo
X-squared = 5.4022, df = 1, p-value = 0.0201
```

Figura 5.8: Test Chi-cuadrado entre Cambio y Grupo.

Como el p-valor del test es menor que 0.05, rechazamos la independencia de las variables. También se puede comprobar a través del test de Fisher[21], cuya hipótesis nula es que la Odds Ratio[22] es igual a 1, es decir, que las variables son independientes (Figura 5.9).

```
Fisher's Exact Test for Count Data
data: Tabla
p-value = 0.04056
alternative hypothesis: true odds ratio is not equal to 1
```

Figura 5.9: Test de Fisher entre Cambio y Grupo.

De nuevo el p-valor menor que 0.05, rechaza la independencia. Podemos ir más allá y generar un intervalo de confianza para la OR de estas dos variables, con *Cambio* en el numerador y *No cambio* en el denominador para el grupo A, obteniendo los resultados de la Tabla 5.2.

OR	Ext. inferior	Ext. superior
17.5	1.22	250.36

Tabla 5.2: Intervalo de confianza al 95 % de la Odds Ratio Cambio y Grupo

Por un lado, el hecho de que el 1 esté fuera del intervalo de confianza, confirma lo que ya habíamos visto, existe relación entre las dos variables. La OR para el grupo A, el que tenía las convocatorias con menos abstención, es 17.5, lo que indica que es 17 veces mayor la probabilidad de que haya un cambio en el gobierno si hay baja abstención que de que no haya cambio en el gobierno. Análogamente, esto significa que la probabilidad de que el gobierno

Análisis de los datos

se mantenga cuando la abstención es mayor, es 17 veces más grande que la probabilidad de que el gobierno cambie. Esto puede significar que cuando hay disconformidad con el gobierno actual, la gente tiende a participar más en las elecciones, lo que propicia que haya cambios de gobierno cuando la abstención es más baja.

Podemos utilizar los resultados de la convocatoria de julio de 2023 (excluida del análisis al no encontrarse disponibles sus datos al comienzo de este trabajo) para ver si en ella se cumple nuestra hipótesis. En esta convocatoria, la abstención nacional de los residentes en España ha sido del 29.59%. Según nuestra clasificación, pertenecería al grupo B, por ser superior al 25% y por tanto no haber cambio de gobierno. Aunque el partido más votado en esta convocatoria haya sido uno perteneciente al bloque de la derecha, finalmente la izquierda ha continuado gobernando, por lo que no ha habido cambio de gobierno y se cumple la conclusión extraída en el análisis.

Se puede realizar un análisis discriminante utilizando el valor medio de la abstención en cada convocatoria en vez de la pertenencia al grupo A o B, para asegurarnos de que este resultado no está sesgado por la agrupación previa. Este procedimiento construye un modelo en base a la variable predictora *Abstención* para discriminar en los dos niveles de la variable *Cambio*.

Para realizar dicho análisis utilizamos tanto R como Statgraphics, ya que la función *lda* de R[23] no aporta los p-valores de los coeficientes del modelo. En los resultados de R (Figura 5.10), vemos que el coeficiente de la variable *Abstención* es 33.02, lo que sugiere una fuerte relación positiva con la variable *Cambio*.

En la figura 5.11 comprobamos en primer lugar, que el p-valor menor que 0.01 para la variable *Abstención* indica que esta es significativa para el análisis. Además, la correlación de 0.68 nos indica una dependencia moderadamente fuerte entre esta variable y el cambio de gobierno. La evaluación de los errores de clasificación obtenidos se resume en la Tabla 5.3, con un acierto en el 80% de los casos.

5.1. Relación entre abstención y convocatorias

```

call:
lda(Cambio ~ abstencion_nacional)

Prior probabilities of groups:
  NO  SI
0.6  0.4

Group means:
      abstencion_nacional
NO          0.2849333
SI          0.2305333

Coefficients of linear discriminants:
              LD1
abstencion_nacional 33.02453
  
```

Figura 5.10: Análisis discriminante en R

Análisis Discriminante

Variable de clasificación: cambio

Variables independientes:

Porc abstención

Número de casos completos: 15

Número de grupos: 2

Función Discriminante	Eigenvalor	Porcentaje Relativo	Correlación Canónica
1	0,878393	100,00	0,68383

Funciones Derivadas	Lambda de Wilks	Chi-Cuadrada	GL	Valor-P
1	0,53237	7,8802	1	0,0050

Figura 5.11: Análisis discriminante en Statgraphics

	Predice Cambio	Predice No Cambio
Cambio	5	1
No cambio	2	7

Tabla 5.3: Tabla Cambio contra Predicho

5.1.2. Relación entre abstención y partido ganador

Aquí, exploraremos la posibilidad de que la abstención esté más relacionada con un gobierno de izquierda o con uno de derecha. Para ello, comenzaremos elaborando una tabla de contingencia en la que las columnas serán los dos grupos previamente formados (uno con las convocatorias con más abstención y otro con las que menos) y en las filas las convocatorias en los que ha gobernado la derecha o la izquierda después de las elecciones.

Los resultados se encuentran en la Tabla 5.4:

	Grupo A	Grupo B
Gobierna Izquierda	5	3
Gobierna Derecha	2	5

Tabla 5.4: Tabla Gobierno contra Grupo

En primer lugar, para comprobar si existe relación o no entre las dos variables, aplicaremos un test Chi-Cuadrado (Figura 5.12).

```
Pearson's Chi-squared test with Yates' continuity correction
data: AbstGobierna
X-squared = 0.63257, df = 1, p-value = 0.4264
```

Figura 5.12: Test Chi-Cuadrado entre Gobierno y Grupo.

El p-valor del test, 0.42 es superior a 0.05 por lo que no podemos asegurar descartar la independencia entre las variables. Podemos utilizar el test de Fisher como en la subsección anterior, para comprobar si la Odds Ratio es distinta de 1. En la Figura 5.13 se recogen los resultados.

```
Fisher's Exact Test for Count Data
data: AbstGobierna
p-value = 0.3147
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 0.3253661 65.7047428
sample estimates:
odds ratio
 3.759962
```

Figura 5.13: Test de Fisher entre Gobierno y Grupo.

Por un lado, el p-valor del test es mayor que 0.05, por tanto no podemos asegurar que la Odds Ratio sea significativamente distinta de 1, como indica su hipótesis nula. Podemos esti-

5.2. Relación entre abstención y tamaño de la provincia

mar la Odds Ratio y construir un intervalo de confianza para esta a nivel 95 %. Los resultados se recogen en la Tabla 5.5. El intervalo de confianza contiene el 1, y este resultado apoya la hipótesis anterior: no existe una relación significativa entre el hecho de que haya más abstención y el gobierno de un partido u otro.

OR	Ext. inferior	Ext. superior
3.75	0.32	65.7

Tabla 5.5: Intervalo de confianza al 95 % de la Odds Ratio Gobierno y Grupo

Como conclusión a toda esta sección, podemos decir que la abstención no está más relacionada con un gobierno de izquierdas o de derechas, como se podía suponer, si no con el hecho de que siga el mismo gobierno o haya un cambio en este. En las convocatorias con menor abstención, es más probable que haya un cambio de gobierno, mientras que, en las convocatorias con mayor abstención, es más probable que el gobierno se mantenga. Además, cabe destacar que el valor que determina si en una convocatoria hay abstención alta o baja es el 25 %, punto de corte para clasificar la participación en las elecciones.

5.2. Relación entre abstención y tamaño de la provincia

El objetivo inicial de esta sección era realizar un análisis de la abstención considerando como variable explicativa el censo con los municipios como unidades, pero los resultados obtenidos no eran nada buenos por la cantidad y variedad de municipios a lo largo de quince convocatorias. En base a esto, se decidió simplificar el modelo utilizando la provincia como unidad de estudio.

Para ello, se construirán diversos modelos de regresión que utilicen el censo de las provincias como variable explicativa y la abstención relativa como variable respuesta. Cabe destacar que se utilizará el censo como variable continua y no la versión discreta que se describió en el Capítulo 3.

En primer lugar, partiremos del modelo más simple, uno que utiliza como observaciones la abstención en cada provincia y en cada convocatoria. En todos los modelos utilizaremos el logaritmo del censo como variable explicativa para paliar el efecto de la asimetría que segrega las provincias más grandes por un lado y las pequeñas por otro.

En este primer modelo (Figura 5.14) vemos que el efecto del censo no es significativo para

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.297710	0.033769	8.816	<2e-16 ***
log(Censo)	-0.002443	0.002615	-0.934	0.35

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.06681 on 778 degrees of freedom
Multiple R-squared: 0.001121, Adjusted R-squared: -0.0001634
F-statistic: 0.8728 on 1 and 778 DF, p-value: 0.3505

Figura 5.14: Resultados ANOVA del primer modelo de regresión.

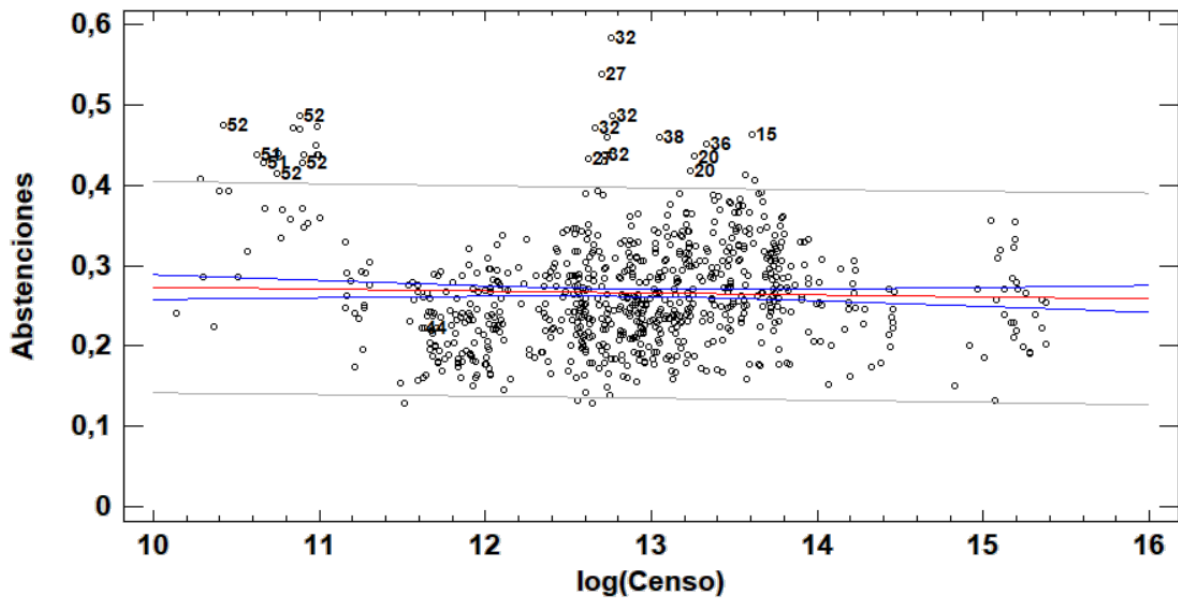


Figura 5.15: Gráfico del primer modelo de regresión.

5.2. Relación entre abstención y tamaño de la provincia

explicar la abstención, su p-valor es mayor de 0,3. Además, si vemos el gráfico junto a las bandas de predicción (Figura 5.15), comprobamos que las provincias que se encuentran fuera de dichas bandas son siempre las mismas. En consecuencia, el siguiente paso será crear un modelo igual pero esta vez excluyendo las provincias que peor están explicadas por el modelo. Estas provincias son las que tienen ID 51, 52, 27 y 32 (correspondientes a Ceuta, Melilla, Ourense y Lugo, respectivamente).

Comprobamos ahora que los resultados (Figura 5.16 y Figura 5.17) son mucho más significativos. Por un lado, el efecto del logaritmo del censo sobre la abstención relativa es relevante, con un pvalor mucho menor que 0.01. Por otro lado, el coeficiente R2 es bajo, pero eso no significa que el modelo no sirva a nuestro objetivo.

```
Coefficients:
      Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.071140   0.032756   2.172  0.0302 *
log(Censo)   0.014401   0.002518   5.719 1.57e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.05631 on 718 degrees of freedom
Multiple R-squared:  0.04357, Adjusted R-squared:  0.04223
F-statistic: 32.71 on 1 and 718 DF, p-value: 1.573e-08
```

Figura 5.16: Resultados ANOVA del segundo modelo de regresión.

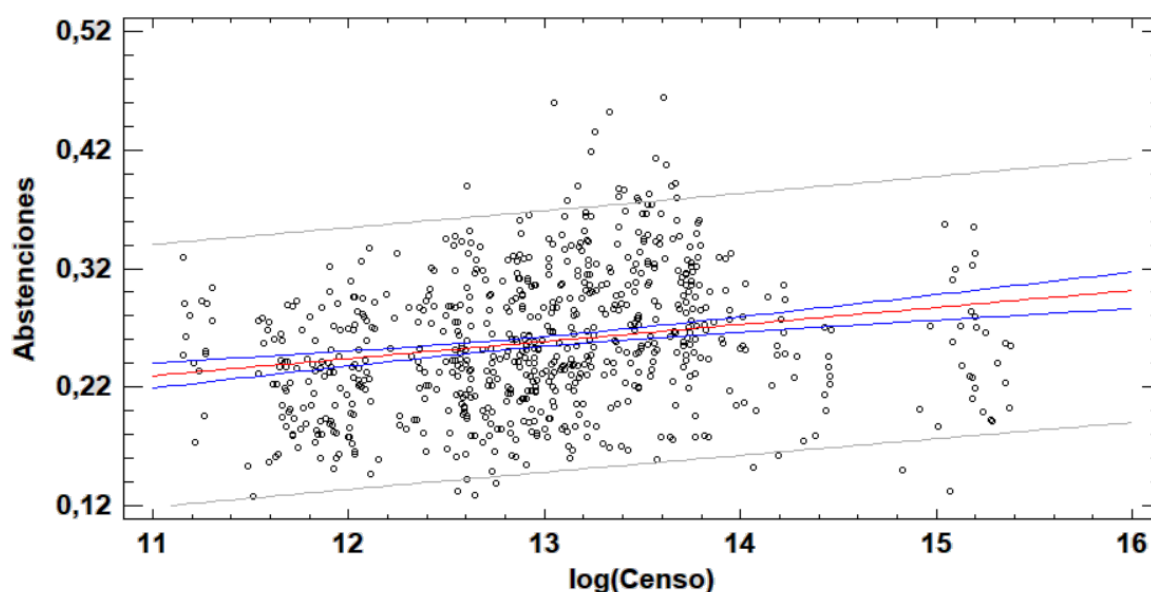


Figura 5.17: Gráfico del segundo modelo de regresión.

Análisis de los datos

Podemos considerar ahora una nueva variable explicativa en este modelo, la convocatoria a la que pertenecen los resultados de la abstención. Como ya hemos visto, la abstención varía en función de la convocatoria, y puede que esta nueva variable ayude a mejorar el modelo. Se empieza construyendo un modelo que incluye todas las provincias.

Residual standard error: 0.05573 on 750 degrees of freedom
Multiple R-squared: 0.3299, Adjusted R-squared: 0.304
F-statistic: 12.73 on 29 and 750 DF, p-value: < 2.2e-16

Figura 5.18: Coeficiente R2 del tercer modelo de regresión.

En un primer vistazo, observamos que el R2 ha mejorado mucho con el efecto de la convocatoria (Figura 5.18), esto significa que algo de relevancia tiene que tener. En la Figura 5.19, vemos que el modelo construye una recta de regresión para cada una de las convocatorias. En la Figura 5.20, comprobamos que hay diferencias significativas entre los intercepts de estas rectas (la variable Año), pero no entre sus pendientes (la interacción entre el Censo y la convocatoria, Año).

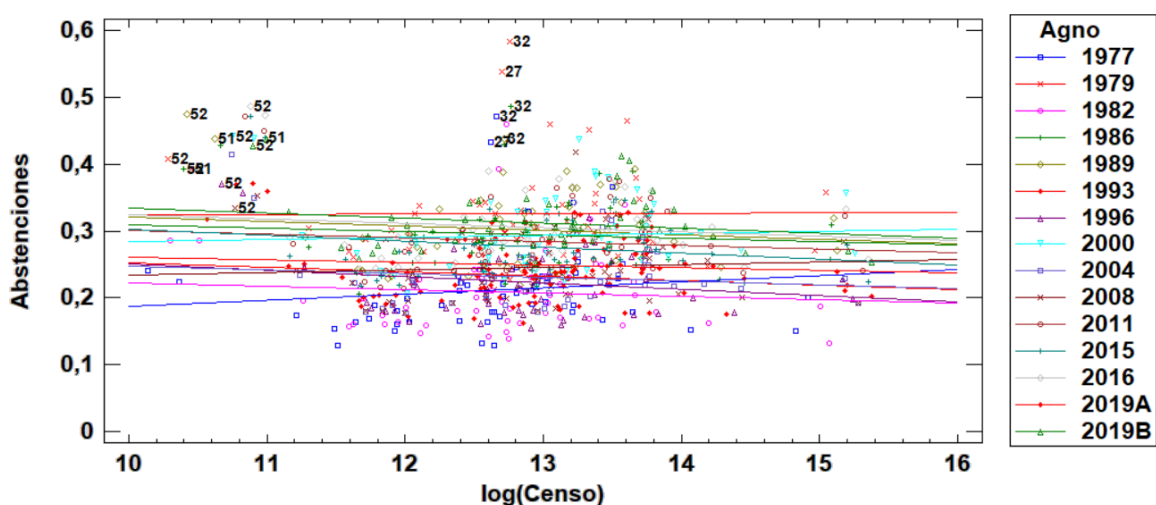


Figura 5.19: Gráfico del tercer modelo de regresión.

5.2. Relación entre abstención y tamaño de la provincia

Analysis of Variance Table

Response: porcentaje

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
log(Censo)	1	0.00390	0.003895	1.2541	0.2631
Agno	14	1.12621	0.080444	25.9002	<2e-16 ***
log(Censo):Agno	14	0.01662	0.001187	0.3823	0.9799
Residuals	750	2.32943	0.003106		

 Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Figura 5.20: Resultados ANOVA del tercer modelo de regresión.

Antes de entrar en profundidad en este aspecto, planteemos un último modelo igual que este pero excluyendo las provincias mencionadas anteriormente, ya que en la Figura 5.19 vemos que sus valores ajustados siguen estando lejos del valor observado.

Residual standard error: 0.04281 on 690 degrees of freedom
 Multiple R-squared: 0.4688, Adjusted R-squared: 0.4464
 F-statistic: 21 on 29 and 690 DF, p-value: < 2.2e-16

Figura 5.21: Coeficiente R2 del cuarto modelo de regresión.

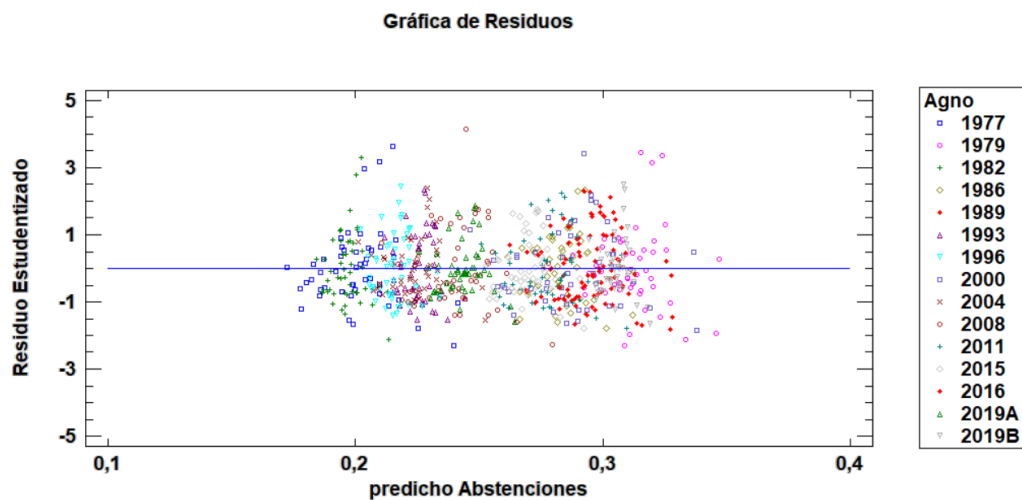


Figura 5.22: Gráfico de residuos del cuarto modelo de regresión.

Análisis de los datos

```
Analysis of Variance Table
Response: porcentaje
          Df Sum Sq Mean Sq F value    Pr(>F)
log(Censo)  1  0.10372  0.103718  56.5869  1.68e-13 ***
Agno       14  1.00181  0.071558  39.0407 < 2.2e-16 ***
log(Censo):Agno 14  0.01045  0.000747  0.4074  0.9728
Residuals  690  1.26471  0.001833
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figura 5.23: Resultados ANOVA del cuarto modelo de regresión.

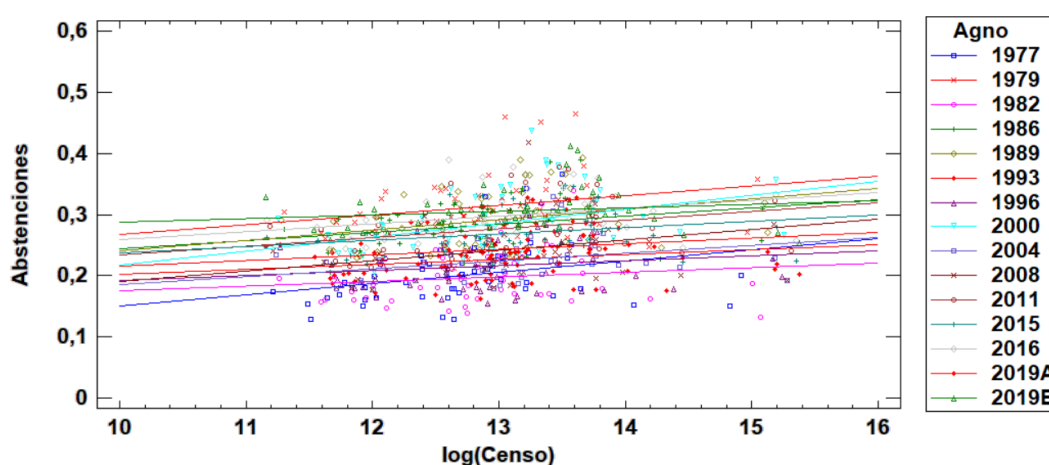


Figura 5.24: Gráfico del cuarto modelo de regresión.

Este modelo mejora significativamente su R^2 (Figura 5.21). También vemos un análisis residual bastante bueno (Figura 5.22) donde los residuos no tienen patrones ni tendencias y además, se mantienen los resultados que ya habíamos visto: las rectas de regresión de las distintas convocatorias tienen intercepts significativamente distintos, pero no hay evidencias de que no sean paralelas (Figura 5.23). Como la pendiente de las rectas es positiva, podemos concluir que hay un mayor nivel de abstención en las provincias con más censo, como ya se podía intuir tras los gráficos descriptivos del Capítulo 4. A mayores concluimos que el efecto de la convocatoria es significativo, pero no lo es su interacción con el censo, pues no hay diferencias entre las pendientes de las distintas rectas.

En vista a estos resultados, se añade un último modelo en el que no se tiene en cuenta la interacción de la convocatoria con el censo, es decir, asumiendo pendientes iguales. Las figuras 5.25, 5.26, 5.27 y 5.28 muestran los resultados, muy similares al modelo anterior, pero ahora podemos ver que las rectas son totalmente paralelas.

5.2. Relación entre abstención y tamaño de la provincia

Residual standard error: 0.04256 on 14 degrees of freedom
 Multiple R-squared: 0.4645, Adjusted R-squared: 0.4531
 F-statistic: 21 on 29 and 690 DF, p-value: <2.2e-16

Figura 5.25: Coeficiente R2 del quinto modelo de regresión.

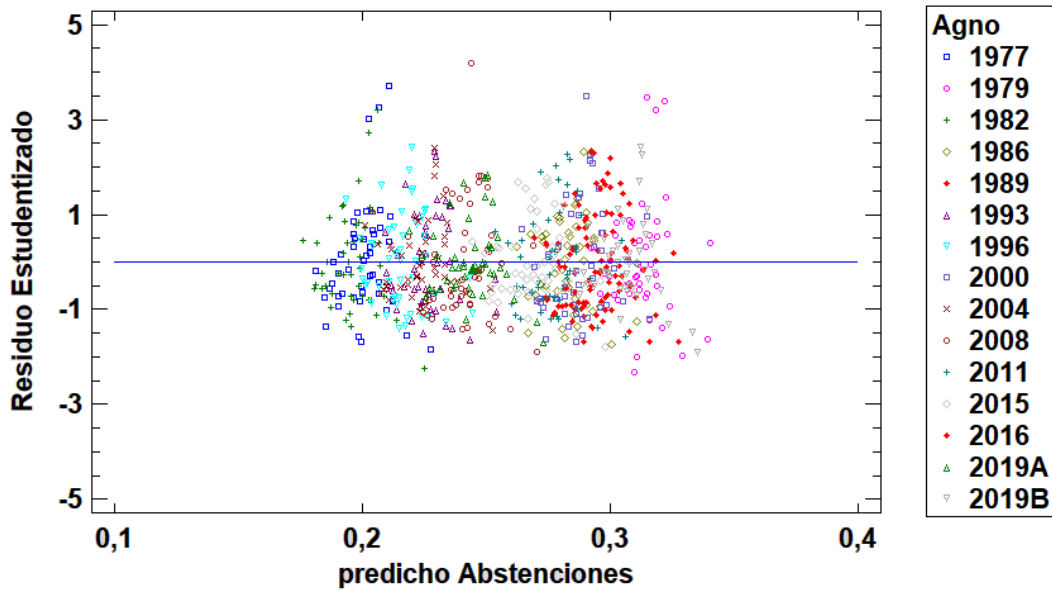


Figura 5.26: Gráfico de residuos del quinto modelo de regresión.

Analysis of Variance Table

Response: porcentaje

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
log(Censo)	1	0.10377	0.103771	57.2809	1.65e-13 ***
Agno	14	1.00306	0.071647	39.5415	< 2.2e-16 ***
Residuals	704	2.38237	0.001811		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Figura 5.27: Resultados ANOVA del quinto modelo de regresión.

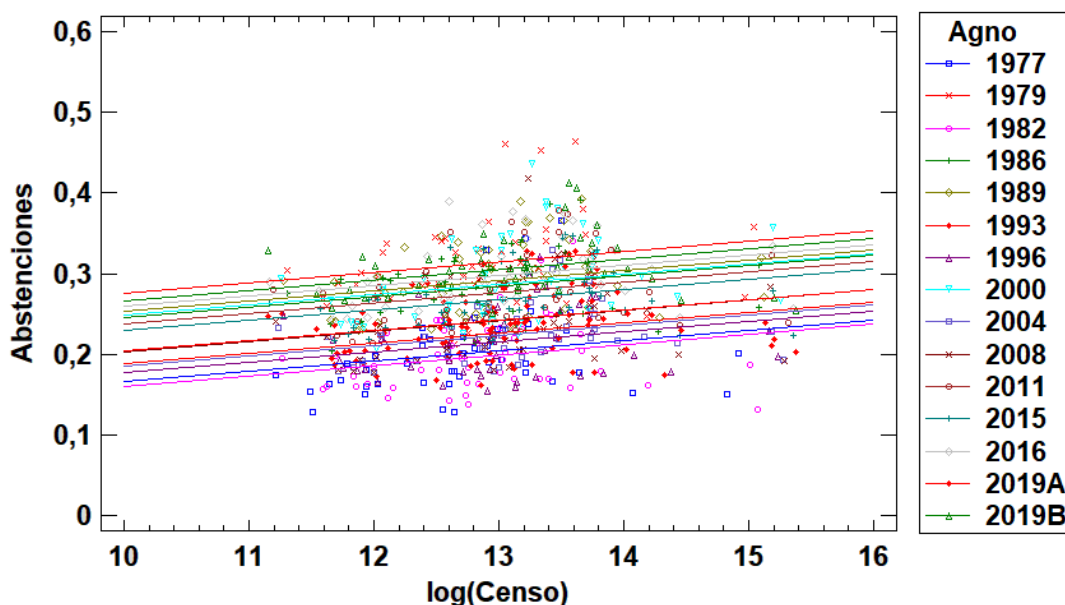


Figura 5.28: Gráfico del quinto modelo de regresión.

Como conclusión a esta sección podemos asegurar que existe un efecto del censo de las provincias sobre la abstención: a mayor población mayor abstención. Podríamos preguntarnos si este efecto aumenta cuando hay un mayor nivel de abstención a nivel nacional, como ocurre en distintas convocatorias, y la respuesta es negativa: independientemente del nivel de abstención nacional, el efecto del censo será el mismo, como se refleja en la igualdad de las pendientes de las distintas rectas de regresión.

5.3. Análisis de Correspondencias Simples por convocatorias

Como ya hemos visto en la sección 5.1, no existe una relación directa entre una mayor o menor abstención con el hecho de que gobierne un partido u otro. Sin embargo, podemos analizar para una convocatoria concreta cuales son los partidos con los que está más relacionada la abstención con el fin de extraer alguna conclusión.

Antes de comenzar con el análisis, cabe recordar que, como ya se dijo en la sección 3.3, la clasificación de las candidaturas ha sido una parte muy costosa del preprocesamiento y los resultados no son exactos al no poderse clasificar manualmente las más de 150 candidaturas distintas para cada convocatoria. Por eso, en esta sección, solo analizaremos en detalle unas pocas convocatorias en las que, además de aplicar las técnicas descritas en la sección 3.3, se

5.3. Análisis de Correspondencias Simples por convocatorias

han clasificado manualmente las candidaturas más significativas restantes para poder garantizar unos resultados precisos en los que en ninguna de las provincias haya valores ausentes para grandes bloques políticos, como ya se ha explicado que podía pasar.

Para la elección de que convocatorias utilizar en este análisis, inicialmente se eligieron el 2008 y el 2011, pues en la primera de ellas gana la izquierda y en la segunda la derecha. Cuando se obtuvo una posible conclusión a raíz de estos análisis, se eligieron algunas convocatorias consecutivas por delante y por detrás a estas fechas. De esta forma, las convocatorias analizadas en esta sección van del 2000 al 2015.

En esta sección usaremos Análisis de Correspondencias Simples que muestren gráficamente la relación entre los niveles de las variables *Provincia* y *Bloque Político*. Inicialmente se intentó realizar el análisis con la variable municipio, pero los más de 8000 niveles de esta variable, imposibilitaban que el software pudiera completar el trabajo. Además, como se indica en la sección 2.2.3, utilizaremos los elementos suplementarios para proyectar el nivel *Abstenciones* de forma posterior al resto de bloques políticos y mejorar su interpretación. A continuación se exponen los resultados para cada convocatoria y una conclusión final.

5.3.1. Convocatoria 2000

Los resultados más relevantes del ACS para la convocatoria del 2000 son los siguientes. Por un lado la Figura 5.29 representa la variabilidad explicada en cada uno de los ejes. Como se ve, el primer eje explica la gran mayoría de la varianza, y con los dos primeros ejes se alcanza casi un 94 %.

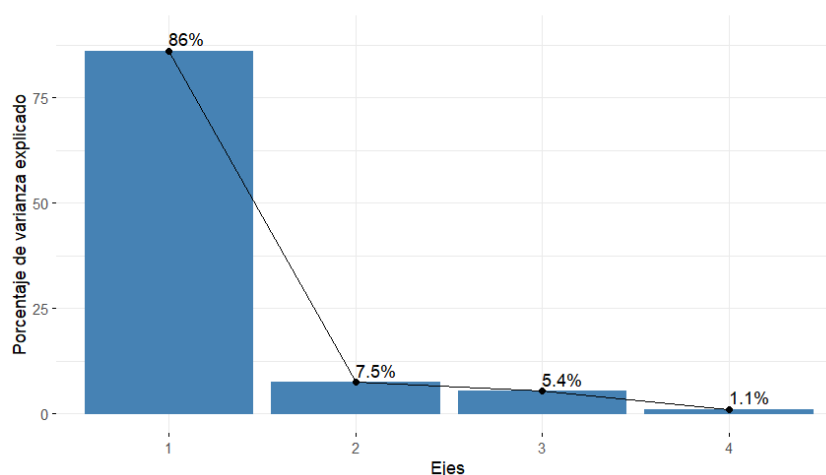


Figura 5.29: Varianza explicada de la convocatoria del 2000.

Análisis de los datos

Si representamos en un gráfico de dos dimensiones nuestras filas, las provincias, y nuestras columnas, los bloques políticos, añadiendo la abstención como variable suplementaria se obtiene la Figura 5.30.

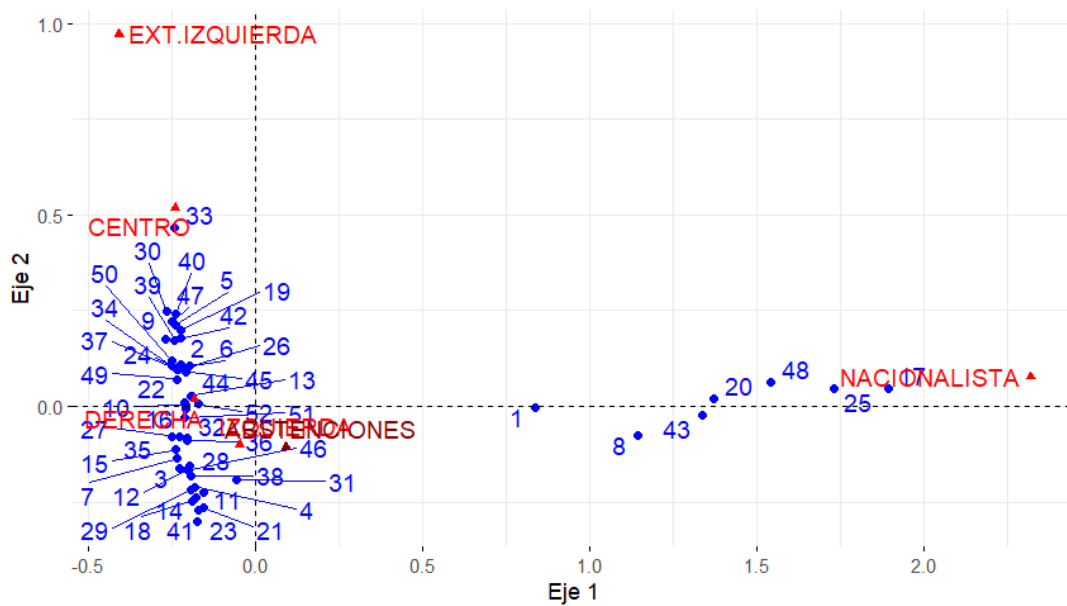


Figura 5.30: Gráfico ACS de la convocatoria de 2000. Ejes 1 y 2.

El eje 1 separa las provincias más relacionadas con el bloque nacionalista del resto. No es sorprendente que estas provincias sean Girona, Lleida, Bizkaia, Gipuzcoa o Tarragona, como podemos comprobar con los códigos del Instituto Nacional de Estadística [16]. Pero a nosotros lo que nos interesa en este análisis es apreciar el nivel de cercanía de la abstención al bloque de derechas o al de izquierdas. Para esto, será mejor representar en el gráfico los ejes 2 y 3 y no perder información con el bloque nacionalista (Figura 5.31).

Los resultados están más claros ahora y vemos que las abstenciones se sitúan más cerca del bloque de derechas. Nótese que en esta convocatoria fue la derecha quien gobernó tras las elecciones.

5.3. Análisis de Correspondencias Simples por convocatorias

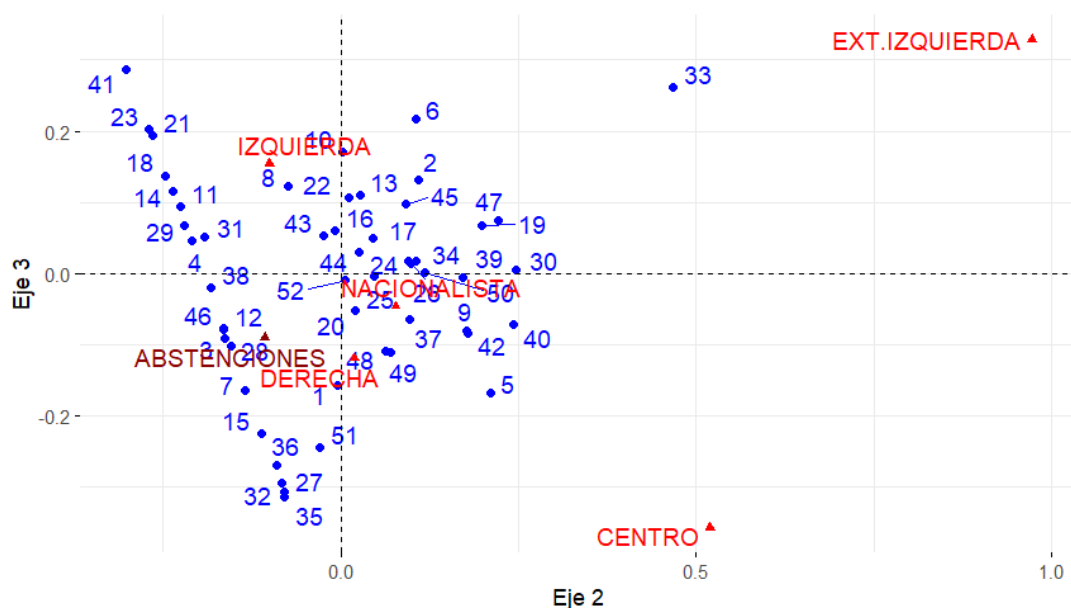


Figura 5.31: Gráfico ACS de la convocatoria de 2000. Ejes 2 y 3.

5.3.2. Convocatoria 2004

De esta convocatoria en adelante, solo se incluirá el gráfico final que nos permitirá analizar los resultados, pues la varianza explicada en cada eje es muy similar en todas las convocatorias (el primer eje supera el 85 % de varianza, y junto con el segundo eje se llega a más del 90 %) y el gráfico en el que se representan los dos primeros ejes separa siempre las provincias más nacionalistas del resto. Para no sobrecargar con esta información que no nos interesa en la presente sección, solo se mostrará el gráfico que representan el eje 2 y el eje 3 (Figura 5.32).

Como se puede ver en el gráfico, la abstención se encuentra más cerca de la izquierda que de la derecha. En estas elecciones, el PSOE consiguió la mayoría y gobernó tras ellas.

5.3.3. Convocatoria 2008

El gráfico obtenido tras el ACS en la convocatoria del 2008 con los ejes 2 y 3 (el primer eje solo separaba a los nacionalistas) se encuentra en la Figura 5.33.

En esta convocatoria, de nuevo el Partido Socialista fue el que obtuvo la mayoría y observamos claramente que la abstención se encuentra más cerca de la izquierda que de la derecha.

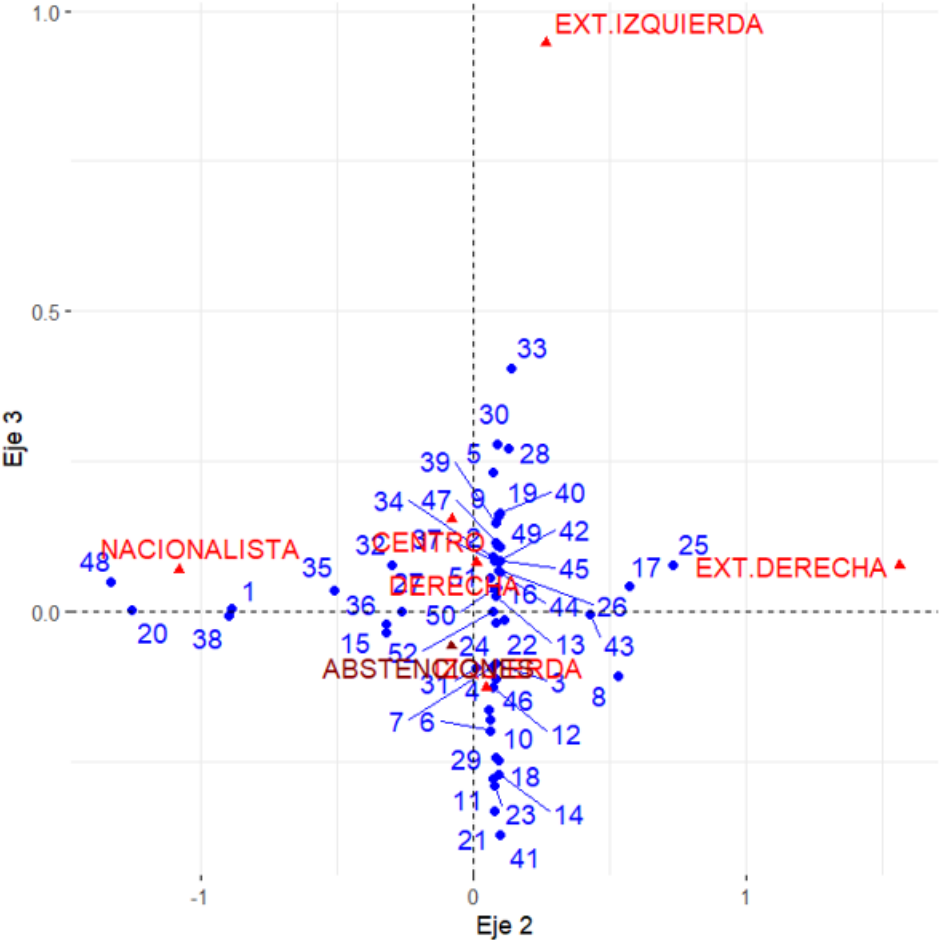


Figura 5.32: Gráfico ACS de la convocatoria de 2004.

5.3. Análisis de Correspondencias Simples por convocatorias

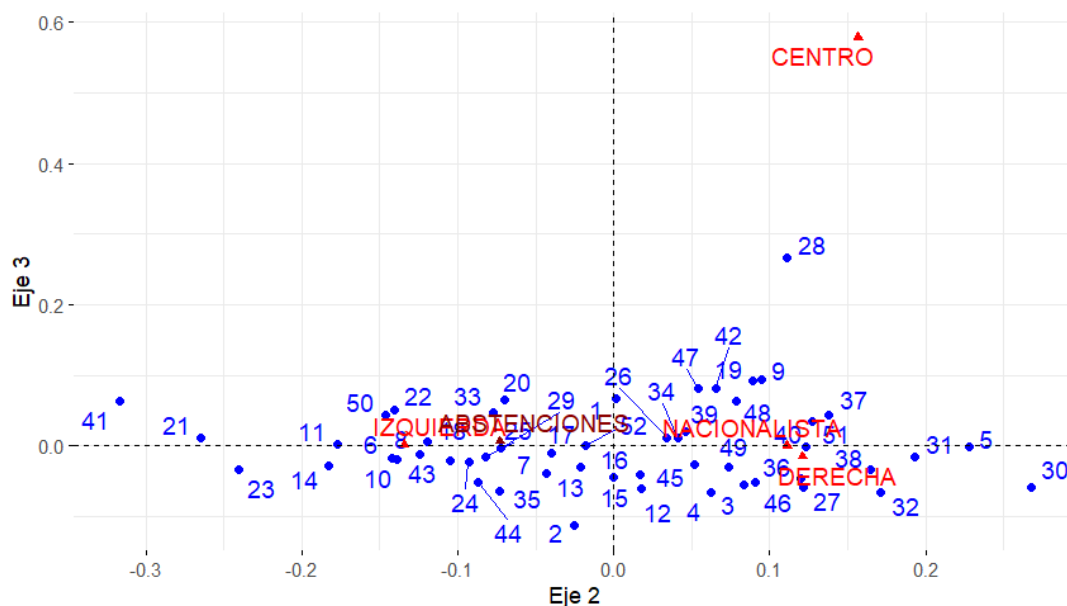


Figura 5.33: Gráfico ACS de la convocatoria de 2008.

5.3.4. Convocatoria 2011

Representando el segundo y tercer eje tras el análisis de correspondencias simples de la convocatoria del 2011 obtenemos el siguiente resultado (Figura 5.34). En esta convocatoria fue el Partido Popular quien obtuvo la mayoría, y en el gráfico observamos que las abstenciones se encuentran más cerca de la derecha que de la izquierda.

5.3.5. Convocatoria 2015

Por último, analicemos los resultados en la convocatoria del 2015. La figura 5.35 es el gráfico obtenido tras realizar el ACS.

Parece que la cercanía de la variable abstenciones a uno de los grandes bloques políticos no es tan directa como en las anteriores convocatorias, en todo caso, se encuentra más cerca del centro o de los nacionalistas que de la izquierda o la derecha. En esta convocatoria ganó el Partido Popular, pero estas elecciones supusieron el fin del bipartidismo, y el PP ganó con un 28% de los votos, frente al más del 40% de los votos que habían obtenido los partidos ganadores en el resto de las convocatorias. Además la diferencia entre el partido ganador y el siguiente en las convocatorias anteriores era mayor que la que hubo en estas elecciones. Puede que por estos motivos, la abstención no se relacione tan directamente con un partido.

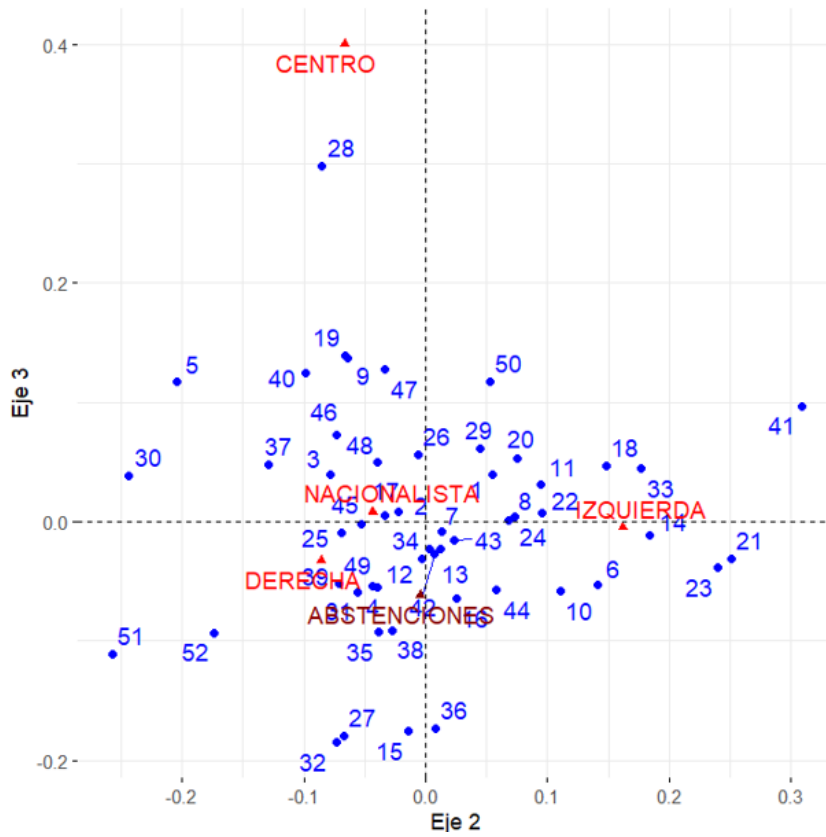


Figura 5.34: Gráfico ACS de la convocatoria de 2011.

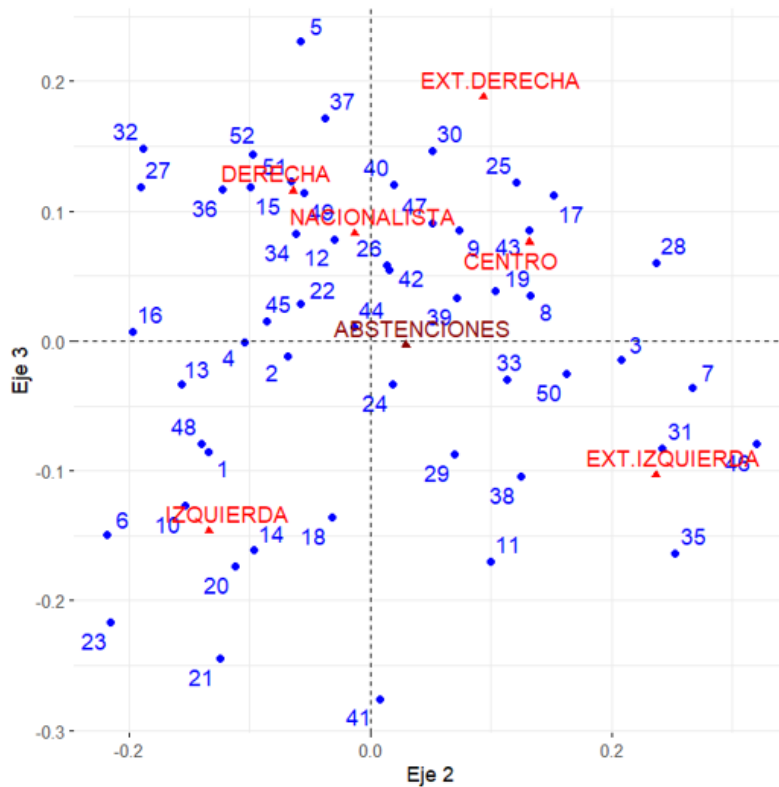


Figura 5.35: Gráfico ACS de la convocatoria de 2015.

5.3.6. Conclusión del Análisis de Correspondencias

Como hemos visto en las convocatorias anteriores, en casi todos los casos, la abstención se sitúa más cercana al bloque ganador de dicha convocatoria. Como ya se explicó, esta cercanía significa una correlación positiva alta, es decir, que cuando una tiende a aumentar, la otra también lo hará.

Podemos decir, de manera general, que en las provincias en las que obtiene más votos el partido que gana las elecciones, la abstención es mayor. Esto podría interpretarse como una menor motivación entre los votantes del partido que no gana dichas elecciones, es decir, que las provincias en las que obtiene más votos (relativamente al total global de esas elecciones) el partido que no gana las elecciones son aquellas con menor abstención.

Capítulo 6

Conclusiones

En este trabajo se ha realizado un estudio de la abstención electoral al Congreso de los Diputados en España desde 1977 hasta 2019 mediante los softwares de R y Statgraphics. Para llevar a cabo este análisis, se han utilizado metodologías como el Análisis de la Varianza, modelos de regresión y Análisis de Correspondencias Simples.

En primer lugar, el ANOVA realizado entre la abstención y las convocatorias muestra una fuerte relación entre las convocatorias en las que la abstención es inferior al 25% y el cambio de gobierno, y aquellas con una abstención superior al 25% y la permanencia del mismo gobierno. Podemos concluir que una mayor participación en las elecciones tiene como consecuencia la entrada de un nuevo partido al gobierno. Además, queda descartada la posible relación entre la mayor o menor abstención y un gobierno de izquierdas o derechas.

En segundo lugar, tras estudiar la relación entre la abstención y el tamaño de las provincias mediante modelos de regresión, llegamos a la conclusión de que en aquellas provincias con mayor población, hay una mayor abstención. De este análisis también se extrae que este efecto no aumenta cuando la abstención es mayor a nivel nacional, es decir, independientemente de la abstención nacional, el efecto del censo en la abstención será el mismo.

Por último, se estudió la relación entre la abstención y las distintas candidaturas en convocatorias individuales mediante Análisis de Correspondencias Simples. En la gran mayoría de convocatorias, la abstención tiene una correlación positiva alta con el partido que ha ganado dicha convocatoria. Podemos interpretar este comportamiento como poca motivación entre los votantes del partido que no gana las elecciones, ya que las provincias en las que este partido tiene más votos, son aquellas con más participación.

Aunque estos resultados se basan en análisis estadísticos sólidos, es importante considerarlos con cautela ya que las elecciones están influenciadas por el comportamiento humano, el cual a menudo es impredecible y difícil de predecir con precisión mediante modelos.

Bibliografía

- [1] J. Font, *La abstención electoral en España: certezas e interrogantes*. Universitat Autònoma de Barcelona, 1996. [Online]. Available: https://digital.csic.es/bitstream/10261/21074/1/REIS_071_072_04.pdf
- [2] Analysis of variance, statgraphics. Visitado: 2024-04-01. [Online]. Available: <https://www.statgraphics.com/analysis-of-variance>
- [3] Analysis of variance, wikipedia. Visitado: 2024-04-01. [Online]. Available: https://es.wikipedia.org/wiki/An%C3%A1lisis_de_la_varianza#:~:text=En%20estad%C3%ADstica%2C%20el%20an%C3%A1lisis%20de,debidos%20a%20diferentes%20variables%20explicativas.
- [4] I. Fernández, *Apuntes Análisis de la Varianza*. Universidad de Valladolid, 2021.
- [5] Fórmulas del modelo anova. Visitado: 13-04-2024. [Online]. Available: <https://www3.uji.es/~gregori/docencia/ei1012-1516/ei1012-1314-labo-s5-regres-anova-doc.html#:~:text=El%20modelo%20ANOVA%20supone%20que,ERROR>
- [6] Shapiro wilk r documentation. Visitado: 2024-04-08. [Online]. Available: <https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/shapiro.test>
- [7] Análisis post-hoc anova. Visitado: 13-04-2024. [Online]. Available: <https://www.ibm.com/docs/es/spss-statistics/beta?topic=anova-one-way-post-hoc-tests>
- [8] Prueba de kruskal willis. Visitado: 2024-05-20. [Online]. Available: https://gauss.inf.um.es/feir/50/#1_contrastes_param%C3%A9tricos_frente_a_no_param%C3%A9tricos
- [9] M. A. Fernández, *Apuntes Análisis de Correspondencias*. Universidad de Valladolid, 2021.

- [10] Procedimiento de análisis de correspondencias. Visitado: 2024-04-15. [Online]. Available: <https://www.fuenterrebollo.com/Economicas/ECONOMETRIA/REDUCIR-DIMENSION/CORRESPONDENCIAS/correspondencias.pdf>
- [11] Modelos de regresión. Visitado: 2024-03-27. [Online]. Available: <https://gauss.inf.um.es/feir/40/>
- [12] Área de descarga de datos. Visitado: 2023-11-15. [Online]. Available: <https://infoelectoral.interior.gob.es/es/elecciones-celebradas/area-de-descargas/>
- [13] Resultados elecciones generales (wikipedia). Visitado: 2023-12-11. [Online]. Available: https://es.wikipedia.org/wiki/Elecciones_generales_de_Espa%C3%B1a
- [14] Clasificación de los municipios. Visitado: 2023-11-29. [Online]. Available: https://www.mapa.gob.es/es/ministerio/servicios/analisis-y-prospectiva/ayp_demografiaenlapoblacionrural2020_tcm30-583987.pdf
- [15] Documentation r: rbindlist. Visitado: 2023-12-10. [Online]. Available: <https://www.rdocumentation.org/packages/data.table/versions/1.15.0/topics/rbindlist>
- [16] Códigos provincias. Visitado: 2024-4-1. Instituto Nacional de Estadística. [Online]. Available: https://www.ine.es/daco/daco42/codmun/cod_provincia.htm
- [17] Levene.test r documentation. Visitado: 2024-04-09. [Online]. Available: <https://www.rdocumentation.org/packages/LaplacesDemon/versions/16.1.6/topics/Levene.Test>
- [18] Duncan.test r documentation. Visitado: 2024-04-08. [Online]. Available: <https://www.rdocumentation.org/packages/agricolae/versions/1.3-7/topics/duncan.test>
- [19] Documentation r: kmeans. Visitado: 2023-12-10. [Online]. Available: <https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/kmeans>
- [20] Chisq.test r documentation. Visitado: 2024-04-08. [Online]. Available: <https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/chisq.test>
- [21] Fisher.test r documentation. Visitado: 2024-04-08. [Online]. Available: <https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/fisher.test>
- [22] Odds ratio. wikipedia. Visitado: 2024-03-12. [Online]. Available: https://es.wikipedia.org/wiki/Odds_ratio
- [23] Discriminante lineal. r documentation. Visitado: 2024-04-15. [Online]. Available: <https://www.rdocumentation.org/packages/MASS/versions/7.3-60.0.1/topics/lda>

Apéndices

Apéndice A

Clasificación de candidaturas

En el siguiente apéndice se expone la clasificación de los distintos partidos políticos en los bloques mencionados en la sección 3.3. Son muchas las coaliciones políticas que se han hecho entre partidos desde 1977, por eso la gran mayoría se ha omitido refiriéndose solo a los nombres de los partidos. De igual manera, las distintas candidaturas para un mismo partido como puede ser PSOE-Galicia o PSOE-Aragón se han resumido únicamente en el nombre del partido. La asignación a un bloque político u otro ha sido asignada a través de la clasificación de Wikipedia [13].

Candidatura	Siglas	Bloque
Unión Centro Democrático	UCD	DERECHA
Partido Socialista Obrero Español	PSOE	IZQUIERDA
Partido Comunista Español	PCE	EXT.IZQUIERDA
Alianza Popular	AP	DERECHA
Partido Demócrata Gallego	PDG	NACIONALISTA
Acción Nacionalista Vasca	ANV	NACIONALISTA
Frente Navarro Independiente	FNI	NACIONALISTA
Liga Comunista	LC	EXT.IZQUIERDA
Partido Nacionalista Vasco	PNV	NACIONALISTA
Falange Española Auténtica	FEA	EXT.DERECHA
Unión Navarra de Izquierdas	UNAI	NACIONALISTA
Partido Socialista de los Trabajadores	PST	IZQUIERDA
Partido Obrero Socialista Internacionalista	POSI	IZQUIERDA
Partido Socialista Gallego	PSG	NACIONALISTA
Convergencia Democrática de Navarra	CDN	NACIONALISTA
Los Verdes-Grupo Verde	LV-GV	EXT.IZQUIERDA
Partido Aragonés	PAR	NACIONALISTA
Liga Comunista	LC	EXT.IZQUIERDA
Falange Española Tradicionalista	FJONS	EXT.DERECHA
Coalición Democrática	CD	DERECHA
Convergencia y Unión	CiU	NACIONALISTA
Coalición Popular	CP	DERECHA
Centro Democrático y Social	CDS	CENTRO
Izquierda Unida	IU	EXT.IZQUIERDA
Partido Popular	PP	DERECHA
Iniciativa per Catalunya Verds	ICV	NACIONALISTA
Unión Progreso y Democracia	UPyD	CENTRO
Ciudadanos	C's	CENTRO
Podemos	Podemos	EXT.IZQUIERDA
Unidas Podemos	UP	EXT.IZQUIERDA
VOX	VOX	EXT.DERECHA
Esquerra Republicana de Catalunya	ERC	NACIONALISTA
Junts per Catalunya	JxCAT	NACIONALISTA
Euskal Herria Bildu	Bildu	NACIONALISTA
Bloque Nacionalista Gallego	BNG	NACIONALISTA
Coalición Canaria	CC	NACIONALISTA
Unión del Pueblo Navarro	UPN	NACIONALISTA
Amaiur	AMAIUR	NACIONALISTA
Nafarroa Bai	NaBai	NACIONALISTA
Libertad Navarra	Ln	NACIONALISTA
Esquerra Unida i Alternativa	EUiA	NACIONALISTA

Tabla A.1: Clasificación de candidaturas

Apéndice B

Resumen de información de las convocatorias

Se expone a continuación un resumen de la información utilizada en este trabajo para cada una de las convocatorias. La primera columna indica la convocatoria, la segunda cual fue el bloque político más votado, la tercera qué bloque político gobernó tras las elecciones, la cuarta si hubo cambio o no de gobierno, la quinta el porcentaje de abstención a nivel nacional (voto CERA no incluido) y la sexta a qué grupo (A o B) del análisis de la sección 5.1 pertenece.

Convocatoria	Más votado	Gobierna	Cambio	Abstención	Grupo
1977	DERECHA	DERECHA	SI	0.2115	A
1979	DERECHA	DERECHA	NO	0.3244	B
1982	IZQUIERDA	IZQUIERDA	SI	0.1983	A
1986	IZQUIERDA	IZQUIERDA	NO	0.2915	B
1989	IZQUIERDA	IZQUIERDA	NO	0.3003	B
1993	IZQUIERDA	IZQUIERDA	NO	0.2297	A
1996	DERECHA	DERECHA	SI	0.2193	A
2000	DERECHA	DERECHA	NO	0.2999	B
2004	IZQUIERDA	IZQUIERDA	SI	0.2276	A
2008	IZQUIERDA	IZQUIERDA	NO	0.2465	A
2011	DERECHA	DERECHA	SI	0.2840	B
2015	DERECHA	DERECHA	NO	0.2692	B
2016	DERECHA	DERECHA	NO	0.3016	B
2019A	IZQUIERDA	IZQUIERDA	SI	0.2425	A
2019B	IZQUIERDA	IZQUIERDA	NO	0.3013	B
2023	DERECHA	IZQUIERDA	NO	0.2959	B

Tabla B.1: Resumen convocatorias

Nota: En la convocatoria de 1977 se considera que hubo cambio de gobierno, pues el anterior gobierno no había sido elegido democráticamente.

Apéndice C

Código R

En esta sección se incluye el código R utilizado para el desarrollo del proyecto.

C.1. Lectura de datos

Se muestra el código de la primera convocatoria. Las mismas líneas son idénticas para las 14 convocatorias restantes.

```
1 #datos por candidaturas por municipios
2 datos1977<-read.table("06027706.DAT", header=FALSE, sep="\t", colClasses="character")
3
4 #datos generales por municipios
5 datosB1977<-read.table("05027706.DAT", header=FALSE, sep="\t")
6
7 #relacion de candidaturas con partidos
8 datosC1977<-read.table("03027706.DAT", header=FALSE, sep="\t", colClasses="character")
```

Listing C.1: Lectura de datos

C.2. Preprocesamiento por convocatorias

Como en la sección anterior, se muestra únicamente el código de una convocatoria. En el resto el procedimiento es similar, pero cambiando manualmente los partidos que más influencia tienen en dicha convocatoria.

C.2. Preprocesamiento por convocatorias

```
1 #separadores para decodificar cada uno de los ficheros
2 sep<-c(2,4,2,1,2,3,2,6,8,3)
3 sepB<-c(2,4,2,1,2,2,3,2,100,1,3,3,3,8,5,8,8,8,8,8,8,8,8,3,8,8,1)
4 sepC<-c(2,4,2,6,50,150,6,6,6)
5
6 #decodifico los datos del primer dataset
7 c1977<-data.frame(split_by_size(datos1977[1:dim(datos1977)[1],],sep))
8 colnames(c1977)<-c("Eleccion", "Agno", "Mes", "Vuelta", "Prov", "Municipio", "Distrito", "Candidatura",
9 "Votos", "Candidatos_obtenidos")
10 c1977$Votos<-as.numeric(c1977$Votos)
11
12 #decodifico los datos del segundo dataset
13 cB1977<-data.frame(split_by_size(datosB1977[1:dim(datosB1977)[1],],sepB))
14 colnames(cB1977)<-c("Eleccion", "Agno", "Mes", "Vuelta", "CCAA", "Prov", "Municipio", "Distrito",
15 "Nombre", "DE", "PJ", "DP", "Comarca", "PD", "Mesas", "Censo", "CE", "CCERE",
16 "TVCERE", "V1", "V2", "Blanco", "Nulo", "Candidaturas", "Escagnos", "VAR", "VNR", "DO")
17 cB1977$Censo<-as.numeric(cB1977$Censo)
18 cB1977$Nulo<-as.numeric(cB1977$Nulo)
19 cB1977$Blanco<-as.numeric(cB1977$Blanco)
20 cB1977$Candidaturas<-as.numeric(cB1977$Candidaturas)
21
22 #decodifico los datos del tercer dataset
23 cC1977<-data.frame(split_by_size(datosC1977[1:dim(datosC1977)[1],],sepC))
24 colnames(cC1977)<-c("Eleccion", "Agno", "Mes", "Candidatura", "Siglas", "Denominacion", "Provincial",
25 "Autonomico", "Nacional")
26
27 # aado la columna con el nombre de los partidos al primer dataset
28 c1977 <- merge(x = c1977, y = cC1977)
29 c1977 <- subset(c1977, select = -Denominacion)
30 c1977 <- subset(c1977, select = -Provincial)
31 c1977 <- subset(c1977, select = -Autonomico)
32 c1977 <- subset(c1977, select = -Nacional)
33
34 #creo el data.table con los datos definitivos
35 aux <- data.table(cB1977[, c("Prov", "Municipio", "Censo", "Blanco", "Nulo", "Candidaturas",
36 "Distrito")]) #convierto a data.table
37
38 #en aux, hay municipios con varios distritos, nos quedamos unicamente con el total (99)
39 aux<-aux[Distrito %in% "99"]
40 aux1<-aux1[Distrito %in% "99"]
41
42 resultado1977 <- aux1[aux, on = .(Prov, Municipio)] #combino ambas tablas usando como clave del
```

Código R

```
    join Prov y Municipio
43 resultado1977[, Abstencion := Censo-Candidaturas-Nulo-Blanco]
44 resultado1977[, Porcentaje := Votos / Censo]
45 resultado1977[, Eleccion := NULL] #elimino columnas que no necesito
46 resultado1977[, Vuelta := NULL]
47 resultado1977[, Distrito := NULL]
48 resultado1977[, i.Distrito := NULL]
49 resultado1977[, Candidatos_obtenidos := NULL]
50 resultado1977[, Blanco := NULL]
51 resultado1977[, Nulo := NULL]
52 resultado1977[, Candidaturas := NULL]
53
54 #filtrado de los partidos relevantes para dicha convocatoria
55 resultado1977<-resultado1977[Siglas %in% c("ANV",
56     "PCC-PCE",
57     "PCE-EPK",
58     "PCE",
59     "FNI",
60     "FJONSA",
61     "FJONS",
62     "FEI",
63     "PNV",
64     "PDG",
65     "PDPC",
66     "AP",
67     "PSCAN",
68     "PSDE",
69     "PSE-PSOE",
70     "PSG",
71     "PSOE",
72     "PSOE-H",
73     "PSP-US",
74     "PSPV",
75     "PSUC-PCE",
76     "UCD"
    ")] #me
    quedo solo con estas candidaturas
77 resultado1977[resultado1977[, Siglas %in% "UCD"
    "]]$Candidatura <- "CENTRO"
78 resultado1977[resultado1977[, Siglas %in% "PSUC-PCE"
    "]]$Candidatura <- "IZQUIERDA"
79 resultado1977[resultado1977[, Siglas %in% "PSPV"
    "]]$Candidatura <- "IZQUIERDA"
80 resultado1977[resultado1977[, Siglas %in% "PSP-US"
```

C.2. Preprocesamiento por convocatorias

```
    ]}$Candidatura <- "IZQUIERDA"
81 resultado1977[resultado1977[, Siglas %in% "PSOE-H
    ]}$Candidatura <- "IZQUIERDA"
82 resultado1977[resultado1977[, Siglas %in% "PSOE
    ]}$Candidatura <- "IZQUIERDA"
83 resultado1977[resultado1977[, Siglas %in% "PSG
    ]}$Candidatura <- "IZQUIERDA"
84 resultado1977[resultado1977[, Siglas %in% "PSE-PSOE
    ]}$Candidatura <- "IZQUIERDA"
85 resultado1977[resultado1977[, Siglas %in% "PSDE
    ]}$Candidatura <- "IZQUIERDA"
86 resultado1977[resultado1977[, Siglas %in% "PSCAN
    ]}$Candidatura <- "IZQUIERDA"
87 resultado1977[resultado1977[, Siglas %in% "PCE
    ]}$Candidatura <- "EXT.IZQUIERDA"
88 resultado1977[resultado1977[, Siglas %in% "PCE-EPK
    ]}$Candidatura <- "EXT.IZQUIERDA"
89 resultado1977[resultado1977[, Siglas %in% "PCC-PCE
    ]}$Candidatura <- "EXT.IZQUIERDA"
90 resultado1977[resultado1977[, Siglas %in% "AP
    ]}$Candidatura <- "DERECHA"
91 resultado1977[resultado1977[, Siglas %in% "PDPC
    ]}$Candidatura <- "NACIONALISTA"
92 resultado1977[resultado1977[, Siglas %in% "PDG
    ]}$Candidatura <- "NACIONALISTA"
93 resultado1977[resultado1977[, Siglas %in% "PNV
    ]}$Candidatura <- "NACIONALISTA"
94 resultado1977[resultado1977[, Siglas %in% "FEI
    ]}$Candidatura <- "EXT.DERECHA"
95 resultado1977[resultado1977[, Siglas %in% "FJONS
    ]}$Candidatura <- "EXT.DERECHA"
96 resultado1977[resultado1977[, Siglas %in% "FJONSA
    ]}$Candidatura <- "EXT.DERECHA"
97 resultado1977[resultado1977[, Siglas %in% "FNI
    ]}$Candidatura <- "NACIONALISTA"
98 resultado1977[resultado1977[, Siglas %in% "ANV
    ]}$Candidatura <- "NACIONALISTA"
99 resultado1977[, Siglas := NULL]
100
101
102 #pasar la columna abstencion, a una fila. despues eliminar la columna abst
103 provincias<-unique(resultado1977$Prov)
104 for (i in provincias){
```

Código R

```
105 municipios<-unique(resultado1977[Prov==i]$Municipio)
106 for(j in municipios){
107   censo <- unique(resultado1977[Municipio == j & Prov == i, Censo])
108   abst <- unique(resultado1977[Municipio == j & Prov == i, Abstencion])
109   resultado1977 <- rbindlist(list(resultado1977, data.table(Agno="1977",
110     Mes="03",Candidatura="Abstenciones",Prov=i, Municipio=j, Votos=abst, Censo=censo,
111     Abstencion=1, Porcentaje=abst/censo)))
112 }
113 }
114 resultado1977[, Abstencion := NULL]
115
116 #etiquetado de los municipios por tamaño
117 cortes<-c(0, 5000, 30000, Inf)
118 etiquetas<-c("Pequeno", "Mediano", "Grande")
119 resultado1977<-resultado1977[, Tamano := cut(Censo, breaks = cortes, labels = etiquetas, right
120   = FALSE)]
```

Listing C.2: Preprocesamiento por convocatorias

C.3. Preprocesamiento global

A continuación se expone el código para juntar los datasets resultantes de cada convocatoria y un preprocesamiento final.

```
1  datos<-rbindlist(list(resultado1977, resultado1979,
2     resultado1982,resultado1986,resultado1989,resultado1993,resultado1996,resultado2000,
3     resultado2004,resultado2008,resultado2011,resultado2015,resultado2016,resultado2019A,
4     resultado2019B), use.names=TRUE)
5  datos[, Mes := NULL]
6  datos[, Agno:= factor(Agno)]
7  datos[, Prov:= factor(Prov)]
8  datos[, Candidatura:= factor(Candidatura)]
9  datos[, Municipio:= factor(paste(Prov, Municipio))] #para q la id municipio sea unica
10 #errores en algunas observaciones (pocas, unas 20): tienen mas votos en candidaturas que
11   censo->abstenciones negativas
12 #las elimino
13 datos <- datos[Votos >= 0]
14 #hay un municipio con censo 9 habitantes y 10 votos. tambien lo elimino
15 datos <- datos[Porcentaje<=1]
```

```

16
17 #agrupar las filas con mismo ano, municipio y candidatura (varias derechas por ejemplo, que
    vienen de distintos partidos)
18 datos <- datos[, .(Votos = sum(Votos), Porcentaje =sum(Porcentaje)), by = .(Año, Prov,
    Municipio, Candidatura, Censo, Tamaño)]

```

Listing C.3: Preprocesamiento global

C.4. Descriptivos

Código para realizar los gráficos descriptivos.

```

1 #abstenciones medias por ao
2 media_agno<- datos[, .(media_abstenciones = mean(Porcentaje)), by = Año]; media_agno
3 ggplot(data = media_agno, aes(x = Año, y = media_abstenciones)) +
4   geom_line(aes(group = 1), color = "skyblue", linewidth = 1) +
5   geom_point(color = "blue", size = 3) + # Puntos de dispersin
6   labs(x = "Convocatoria", y = "Media de abstenciones", title = "Abstenciones medias por ao")
7
8 #segun el tamaño de municipios (poco significativo)
9 media_tamaño<- datos[, .(media_abstenciones = mean(Porcentaje)), by = Tamaño]; media_tamaño
10 ggplot(data = media_tamaño, aes(x = Tamaño, y = media_abstenciones)) +
11   geom_bar(stat = "identity", fill = "skyblue") +
12   labs(x = "Tamaño", y = "Media de abstenciones", title = "Abstenciones medias por tamaño de
    municipio")
13
14 #segun las provincias
15 media_provincia<- datos[, .(media_abstenciones = mean(Porcentaje)), by = Prov]; media_provincia
16 ggplot(data = media_provincia, aes(x = Prov, y = media_abstenciones)) +
17   geom_bar(stat = "identity", fill = "skyblue") +
18   labs(x = "Provincia", y = "Media de abstenciones", title = "Abstenciones medias por
    provincia")
19
20 #boxplot para el tamaño de municipios y la media de abstenciones
21 datos_numericos<-datos$Porcentaje[datos$Candidatura=="Abstenciones"]
22 datos_categ<-datos$Tamaño[datos$Candidatura=="Abstenciones"]
23 boxplot(datos_numericos~datos_categ, ylab="Abstenciones", xlab="Tamaño municipio")

```

Listing C.4: Descriptivos

C.5. Abstención y convocatorias

Código para realizar los análisis relativos a la relación entre la abstención y las convocatorias.

```

1 #preparar el dataset con una fila por provincia y convocatoria
2 datoscopia<-datos
3 datoscopia[,Votos:=NULL]
4 datoscopia[,Tamano:=NULL]
5 datoscopia<-dcast(datoscopia, Prov + Censo + Municipio + Agno ~ Candidatura, value.var =
   "Porcentaje", fun.aggregate = sum)
6 #recuperar los votos en vez de porcentajes
7 datoscopia[, Abstenciones:= Abstenciones*Censo]
8 datoscopia[, DERECHA:= DERECHA*Censo]
9 datoscopia[, NACIONALISTA:= NACIONALISTA*Censo]
10 datoscopia[, IZQUIERDA:= IZQUIERDA*Censo]
11 datoscopia[, EXT.IZQUIERDA:= EXT.IZQUIERDA*Censo]
12 datoscopia[, EXT.DERECHA:= EXT.DERECHA*Censo]
13 datoscopia[, CENTRO:= CENTRO*Censo]
14 datoscopia[,Municipio:=NULL]
15 datoscopia <- datoscopia[, .(Censo = sum(Censo), Abstenciones=sum(Abstenciones),
   NACIONALISTA=sum(NACIONALISTA), IZQUIERDA=sum(IZQUIERDA), DERECHA=sum(DERECHA),
   CENTRO=sum(CENTRO), EXT.DERECHA=sum(EXT.DERECHA), EXT.IZQUIERDA=sum(EXT.IZQUIERDA)), by =
   .(Prov, Agno)]
16
17
18 #recuperar los porcentajes
19 datoscopia[, Abstenciones:= Abstenciones/Censo]
20 datoscopia[, DERECHA:= DERECHA/Censo]
21 datoscopia[, NACIONALISTA:= NACIONALISTA/Censo]
22 datoscopia[, IZQUIERDA:= IZQUIERDA/Censo]
23 datoscopia[, EXT.IZQUIERDA:= EXT.IZQUIERDA/Censo]
24 datoscopia[, EXT.DERECHA:= EXT.DERECHA/Censo]
25 datoscopia[, CENTRO:= CENTRO/Censo]
26
27 fwrite(datoscopia, "TOTAL_ABSTENCIONES.csv")
28 datoscopia <- datoscopia[, .(Censo = sum(Censo), Abstenciones=sum(Abstenciones)), by = .(Agno)]
29
30 #hacer un anova con abstencion como respuesta y convocatoria como factor
31 fm=aov(lm(Abstenciones~Agno,data=datoscopia))
32 summary(fm) #factor debe salir significativo, si q lo es
33 intervals=TukeyHSD(fm)
34 plot(intervals) #ver grupos. explica diferencias dos a dos, si esta el 0 incluido, son iguales

```

```

    esas dos convocatorias
35 # Identificar grupos con medias iguales con Tukey
36 grupos_iguales <- subset(intervals$Agno, abs(intervals$Agno[,1]) < 0.02)
37 print(grupos_iguales)
38
39 #otra forma
40 library(PMCMRplus)
41 kwAllPairsNemenyiTest(datoscopia$Abstenciones, datoscopia$Agno, dist = "Chisquare")
42
43 #validacion del modelo anova
44 #independencia
45 plot(fm$residuals[puntos_a_mostrar] ~ predict(fm)[puntos_a_mostrar], xlab = "Valores
    predichos", ylab = "Residuos")
46 puntos_a_mostrar <- datoscopia$Prov != "51" & datoscopia$Prov != "52" & datoscopia$Prov != "32"
    & datoscopia$Prov != "27"
47 text(predict(fm)[puntos_a_mostrar], fm$residuals[puntos_a_mostrar], labels =
    datoscopia$Prov[puntos_a_mostrar], pos=1)
48
49 #normalidad
50 library(car)
51 qqPlot(fm$residuals ~ predict(fm))
52
53 title(main="Sin transformacion logaritmica")
54 shapiro.test(fm$residuals) #debe ser mayor q 0.05
55
56 fm2=aov(lm(log(Abstenciones)~Agno,data=datoscopia))
57 summary(fm2)
58 qqPlot(fm2$residuals)
59 title(main="Con transformacion logaritmica")
60 shapiro.test(fm2$residuals)
61
62 bartlett.test(fm2$residuals ~ datoscopia$Agno) #debe ser mayor q 0.05 para iguales varianzas
63 leveneTest(fm2$residuals ~ datoscopia$Agno)
64
65 #si no se cumplen las condiciones: kruskal wallis
66 kruskal.test(datoscopia$Abstenciones, datoscopia$Agno)
67 kwAllPairsNemenyiTest(datoscopia$Abstenciones, datoscopia$Agno, method="Chisq")
68
69
70 #boxplot multiple por provincias
71 boxplot(Abstenciones ~ Agno, data = datoscopia, col = "lightblue", main = "Boxplot
    Mltiple", xlab = "Ao", ylab = "Porcentaje")
72 boxplot(Abstenciones ~ Prov, data = datoscopia, col = "lightblue", main = "Boxplot

```

Código R

```
Multiple", xlab = "Año", ylab = "Porcentaje")
73
74
75
76 #juntar por grupos convocatorias parecidas (clustering)
77 m1<-lm(Abstenciones~Año,data=datoscopia)
78 ajustados<-fitted(m1)
79 datos_cluster <- data.frame(año=datoscopia$Año, abstencion=ajustados) #para la misma
    candidatura predice la misma abstencion, no depende d la provincia
80 datos_cluster <- distinct(datos_cluster, datos_cluster$año, .keep_all=TRUE)
81 resultado_kmeans <- kmeans(datos_cluster$abstencion, centers = 2)
82 datos_cluster$grupo <- factor(resultado_kmeans$cluster)
```

Listing C.5: Abstención y convocatorias

C.6. Abstención y tamaño de la provincia

Código para realizar los análisis relativos a la relación entre la abstención y el tamaño de las provincias.

```
1 #me quedo con las abstenciones por provincias
2 abstenciones <- datos[Candidatura == "Abstenciones"]
3
4 regresion<-lm(Porcentaje~Censo,data=abstenciones)
5 summary(regresion)
6 predichos<-predict(regresion)
7 plot(abstenciones$Porcentaje~predichos)
8 abline(a = 0, b = 1, col = "red")
9 residuos<-resid(regresion)
10
11 #validar hipotesis
12
13 #homocedasticidad
14 plot(regresion$fitted.values, residuos, xlab = "Valores ajustados", ylab = "Residuos",
15       main = "Grfico de residuos vs valores ajustados")
16 abline(h = 0, col = "red") # Aade una linea en y=0 para referencia
17 library(lmtest)
18 bptest(regresion) #H0: homocedasticidad
19
20 #independencia
21 acf(residuos) #que las barras no sobresalgan las bandas
```

```
22
23 #media cero
24 media_residuos <- mean(residuos)
25 media_residuos #proximo a 0
26
27 #normalidad
28 shapiro.test(residuos) #fallo, max 5000 obs
29 qqnorm(residuos)
30 qqline(residuos, col = "red")
```

Listing C.6: Abstención y tamaño de provincia

C.7. Análisis de Correspondencias Simples

Código para realizar los ACS. Se incluye únicamente el de una convocatoria.

```
1 d2008<-resultado2008
2 d2008[, Agno := NULL]
3 d2008[, Mes := NULL]
4 d2008[, Tamano := NULL]
5 d2008[, Porcentaje := NULL]
6
7 censos<-data.table(Prov=d2008$Prov, Municipio=d2008$Municipio, Censo=d2008$Censo)
8 censos<-unique(censos)
9 censos[, total_censo:=sum(Censo), by=.(Prov)]
10 censos[, Municipio:=NULL]
11 censos[, Censo:=NULL]
12 censos<-unique(censos)
13
14 d2008[, Municipio := NULL]
15 d2008<-d2008[, total_votos:=sum(Votos), by=.(Candidatura,Prov)]
16 d2008[, Censo := NULL]
17 d2008[, Votos := NULL]
18 d2008<-unique(d2008)
19 d2008 <- merge(d2008, censos, by = "Prov", all.x = TRUE)
20
21 d2008[Candidatura == "Abstenciones", ABST := total_votos]
22 d2008[, ABST := ifelse(is.na(ABST), ABST[!is.na(ABST)][1], ABST), by = Prov]
23 d2008[,Participacion := total_censo-ABST]
24 d2008[,Porc_votos := total_votos/Participacion] #porcentaje de votos sobre los q han votado
25 d2008 <- d2008[Candidatura != "Abstenciones"]
```

Código R

```
26 d2008[,Porc_abst := ABST/total_censo] #porcentaje de abstencion sobre el total
27 d2008[, total_censo := NULL]
28 d2008[, total_votos := NULL]
29 d2008[, ABST := NULL]
30 d2008[, Participacion := NULL]
31 dat2008 <- dcast(d2008, Prov ~ Candidatura, value.var = "Porc_votos")
32
33 #mantener el resultado como un data.table
34 dat2008 <- data.table(dat2008) #ver cuantos NA hay y aadir mas partidos para que disminuyan #44
    NA
35 #quito la columna EXT izquierda porque aporta muy poco
36 dat2008[, EXT.IZQUIERDA := NULL] #elimino la variable provincia
37 dat2008$NACIONALISTA[is.na(dat2008$NACIONALISTA)] <- 0.00005183+runif(40,min=0, max=0.01)
38 #dat2008[,OTROS := 1-CENTRO-DERECHA-IZQUIERDA]
39 dat2008[, Prov := NULL] #elimino la variable provincia
40
41
42 #ya esta el data.table listo para el ACS
43
44 #Anlisis de correspondencia simple
45 ACS <- CA(dat2008,graph=FALSE)
46
47 #% de varianza explicado
48 valores_propios=ACS$eig; valores_propios
49 #Grfica del porcentaje de varianza explicado
50 fviz_screplot(ACS, addlabels = TRUE, ylim = c(0, 90))+ggtitle("")+
51   ylab("Porcentaje de varianza explicado") + xlab("Ejes")
52
53 #Representacin simultnea
54 fviz_ca_biplot(ACS, repel = TRUE)+ggtitle("")+ylab("Eje 2")+xlab("Eje 1")
55
56 #Aadimos ahora la variable suplementaria ABSTENCIONES
57 abs<-unique(d2008$Porc_abst)
58 dat2008[, ABSTENCIONES := abs]
59
60 #Hacemos el ACS indicando que ABSTENCIONES es suplementaria
61 ACS <- CA(dat2008,graph=FALSE,col.sup = 5)
62
63 #% de varianza explicado
64 valores_propios=ACS$eig; valores_propios
65 #Grfica del porcentaje de varianza explicado
66 fviz_screplot(ACS, addlabels = TRUE, ylim = c(0, 90))+ggtitle("")+
67   ylab("Porcentaje de varianza explicado") + xlab("Ejes")
```

C.7. Análisis de Correspondencias Simples

```
68
69 #Obtencin del estadstico de prueba a partir de los valores propios
70 estadistico=sum(valores_propios[1]+valores_propios[2])*1000; estadistico
71 #Estadstico de contraste
72 qchisq(0.95,4)
73
74 #Representacin simultnea
75 fviz_ca_biplot(ACS, axes=c(2,3),repel = TRUE)+ggtitle("")+ylab("Eje 3")+xlab("Eje 2") ##AQUI
    CAMBIO EJES
76
77 #Contribuciones por fila
78 contribuciones_fila=variables_fila$contrib;contribuciones_fila
79 #Cosenos por fila
80 variables_fila=get_ca_row(ACS)
81 cosenos_fila=variables_fila$cos2;cosenos_fila
82
83 #contribuciones por columna
84 contribuciones_columna=variables_columna$contrib; contribuciones_columna
85 #Cosenos por columna
86 variables_columna=get_ca_col(ACS)
87 cosenos_columna=variables_columna$cos2;cosenos_columna
```

Listing C.7: ACS

