

LLICENCIATURA EN CIÈNCIES
I TÈCNIQUES ESTADÍSTIQUES

*Influencia de los atentados del 11M en la
participación electoral del 14M*

Autor: Josep Ramon Marsal

Director: Pedro Delicado

Data d'entrega: 2 de junio de 2004

UNIVERSITAT POLITÈCNICA DE CATALUNYA
Biblioteca



1400611969



Facultat de Matemàtiques
i Estadística

UNIVERSITAT POLITÈCNICA DE CATALUNYA

DADES DEL PROJECTE

Títol del Projecte: Influencia de los atentados del 11M en la participación electoral del 14 07

Nom de l'estudiant: Josep Ramon Marsal

Titulació: LCTO

Crèdits:

Director/Ponent: Pedro Delicado


Departament: 810

QUALIFICACIO

Notable 8

MEMBRES DEL TRIBUNAL (nom i signatura)

President:

Manuel Martí Revolor 

Vocal:

Carlos Seara Giza 

Secretari:

Pedro Delicado 

Data:

27 de juny de 2005

*"Hay dos formas de ver la vida:
una es creer que no existen
milagros, la otra es creer que
todo es un milagro."*

Albert Einstein

AGRADECIMIENTOS

En primer lugar me gustaría agradecer a la Subdirección General del Centro de Sistemas de Información del Ministerio de Interior su trabajo ya que **sin** ello este proyecto no hubiera sido posible. También le doy un **agradecimiento** especial a Pedro Delicado ya que siempre ha estado dispuesto **a** trabajar y en ningún momento ha perdido la paciencia.

Por último sólo agradecer a todos los compañeros, familiares, profesores **y** en general a todas aquellas personas que durante estos años hemos **estudiado**, trabajado o simplemente convivido.

Muchas gracias a todos

ÍNDICE DE CONTENIDOS

Capitulo 1.Introducción	7
Capitulo 2.Objetivo.....	9
Capitulo 3.Bases de datos	11
3.1.Datos I. Participación por correo	11
3.2.Datos 2. Participación CERA.....	13
3.3.Datos de interés para el análisis.....	16
Capitulo 4.Participación	17
4.1.Estimación de la participación	17
4.2.Introducción a los modelos de respuesta normal	19
4.2.1. Modelo lineal general. Caso de la participación.....	21
Capitulo 5.Modelo 1	24
5.1.Estimación de la participación general	25
5.2.Estimación de los parámetros	26
5.3.Modelos con la variable CERA.....	27
5.4.Modelos con la variable AÑO	29
5.5.Modelos con la variable POSTAL.....	31
5.6.Modelos con las variables POSTAL + AÑO	33
5.7.Modelos con las variables CERA + AÑO.....	35
5.7.1. Estimación de la participación general.....	35
5.8.Modelos con las variables AÑO + CERA + CORREO.....	38
5.9.Conclusiones del modelo 1	39
Capitulo 6.Modelo 2	41
6.1.Diferencias entre modelo 1 y modelo 2	42
6.1.1. Utilización de pesos	42
6.2.Interpretación del modelo	44
6.3.Modelo 2.1.....	46
6.4.Modelo 2.2.....	50
Capitulo 7.Modelo 3	52
7.1.Otros modelos finalmente descartados	52

7.2.Años de cambio.....	54
Capitulo 8.Discusión y conclusiones.....	58
8.1.Modelo 1.....	59
8.2.Modelo 2.....	60
8.3.Modelo 3.....	61
8.4.Diferencias entre modelos.....	63
8.5.Resultados.....	64
8.6.Conclusiones.....	66
8.7.Posibles ampliaciones.....	67
Capitulo 9.Anexos.....	68
9.1.Anexo I. MODELO 1.....	68
9.2.Anexo II. MODELO 1.....	74
9.3.Anexo III. Modelo 2.....	78
9.4.Anexo IV. Modelo 3 (con cambio).....	82
9.5.Anexo V. Modelo 3 (sin cambio).....	86
9.6.Anexo VI. Errores de estimación.....	89
9.6.1. Modelo 1.....	89
9.6.2. Modelo 2.1.....	90
9.6.3. Modelo 3 con cambio.....	91
9.6.4. Modelo 3 sin cambio.....	92
Capitulo 10.Bibliografía.....	93

ÍNDICE DE TABLAS

Tabla 1.Descripción fichero 3.....	14
Tabla 2.Codificación de las variables.....	21
Tabla 3.Coeficientes modelo 1.....	40
Tabla 4.Pesos probados en el modelo 2.....	43
Tabla 5.Coeficientes estimados modelo 2.1	49
Tabla 6.Coeficientes Modelo 2.2.....	51
Tabla 7.Coeficientes del modelo 3.....	57
Tabla 8.Resumen y estimación del modelo 1	59
Tabla 9.Resumen y estimación del modelo 2	60
Tabla 10.Resumen y estimación del modelo 3 (sin cambio).....	61
Tabla 11.Resumen y estimación del modelo 3 (con cambio).....	62
Tabla 12.Estimación de los parámetros referentes al modelo 1	73
Tabla 13.Estimación de los parámetros referentes al modelo 2.1	77
Tabla 14.Estimación de los parámetros referentes al modelo 3 (sin cambio)	81
Tabla 15.Estimación de los parámetros referentes al modelo 3 (con cambio)	85
Tabla 16.Estimación de los parámetros referentes al modelo 3 (sin cambio)	88

INDICE DE FIGURAS

Figura 1. Estudios de normalidad de la participación.....	18
Figura 2. Cociente entre censo CERA y censo INE.....	26
Figura 3. Participación según participación CERA en la media española..	27
Figura 4. Participación según año de la elección en las 52 provincias	29
Figura 5. Participación según año de la elección en las 52 provincias	30
Figura 6. Participación según participación postal en la media española ..	32
Figura 7. Relación entre participación en urna y participación postal en Guipúzcoa	34
Figura 8. Estimación de la participación general a partir del modelo 1.4. ..	37
Figura 9. Participación estimada vs. error de ajuste (modelo 1)	39
Figura 10. Residuos vs. predichos en modelo 2.1	48
Figura 11. Distancia de Cook para modelo 2.1	48
Figura 12. Participación provincial según año de elección.....	54
Figura 13. Residuos vs. predichos en modelo 3	56
Figura 14. Error de estimación del modelo 1	89
Figura 15. Error de estimación del modelo 2.1	90
Figura 16. Error de estimación del modelo 3 con cambio	91
Figura 17. Error de estimación del modelo 3 sin cambio	92

Capítulo 1. Introducción

El Partido Socialista Obrero Español (PSOE), ganó las Elecciones Generales del 14 de Marzo de 2004. Aunque la mayoría de los sondeos preelectorales otorgaban la victoria al Partido Popular (PP), el PSOE ganó por una diferencia de casi un 5% (4,88%) al PP.

La inesperada victoria del PSOE hizo surgir discusiones durante los días posteriores sobre la posición adoptada por el electorado ante las urnas y sobre las circunstancias que envolvieron los comicios. Parte de la sociedad imputó a los atentados ocurridos en Madrid el 11M un cambio de voto de los electores y por lo tanto un resultado no esperado.

Por otra parte, algunos imputaron la victoria del PSOE a un cambio en la intención de voto que venía de tiempos anteriores. El cambio de tendencias, según algunas publicaciones, vino por la mala gestión del Gobierno de algunos aspectos durante el último año de poder; lo que provocó que la población fuera cambiando la intención de voto, paulatinamente. Algunos periódicos subrayaban el aumento del PSOE desde principios de año, aunque este ascenso, teóricamente, no habría llegado a darle la presidencia del Gobierno.

Algunas encuestas realizadas periódicamente, ya otorgaban al PSOE un porcentaje de voto superior al del PP, pero la mayoría de los sondeos revelaban una victoria del PP. Según "La Vanguardia" del día 18 de marzo,

donde se comentaban algunos sondeo publicados antes de los atentados, dos de cada tres encuestados creía que el PP continuaría en el Gobierno.

La participación fue superior a la media de las últimas elecciones, lo que algunos achacaron a la gestión que el Gobierno realizó después de los atentados. Éste aumento de voto y la posible redistribución, provocó que se llegara a un resultado que ningún sondeo predijo.

Que los atentados del 11M no dejaron indiferente a la población es un hecho, lo que se cuestiona en este estudio es saber si la tragedia ocurrida en Madrid fue el factor desencadenante de la inesperada victoria del PSOE. Por una parte existe la posibilidad de que hubiera un cambio de tendencias anterior o que el atentado cambiara la intención de voto y cuando el PP había de ganar, perdió. También pudo suceder que el cambio de tendencia realmente existiera y que el atentado lo acelerada, dando una victoria al PSOE.

Lo que se intentará en este estudio es saber si se hubiera obtenido la misma participación si no hubieran sucedido los trágicos acontecimientos de Madrid. Se intentará cuantificar cuánto cambió el voto el 14 de Marzo.

Por lo que el objetivo de este estudio pasará por estudiar y cuantificar qué factores son buenos predictores de la participación general en unas elecciones. Una vez estudiados estos factores se estudiarán los factores que se presentaron el 11M y por lo tanto se podrá estudiar cómo cambió la participación general en las elecciones.

Capítulo 2. Objetivo

El objetivo del proyecto pasa por cuantificar la participación en las elecciones del 2004 si no hubieran ocurrido los atentados terroristas.

Para analizar los resultados será imprescindible conocer los votos realizados por los españoles residentes ausentes (CERA), así como la participación de los electores que decidieron ejercer su derecho de voto vía correo. Los españoles que se encontraban fuera del país, así como los que lo realizaron utilizando el correo postal, tuvieron que votar antes de los atentados. De estos colectivos se tiene el censo y el voto, por lo que se podrá calcular los mismos estadísticos que para los votantes del 14 de Marzo.

La votación de los CERA no estaba *contaminada* por el efecto de los atentados, ya que votaron con anterioridad. Teniendo en cuenta que la población CERA tiene una participación inferior y normalmente con distribución diferente de votos, se podrá estimar la participación hipotética del 14 de Marzo sin el efecto de los atentados. Si los resultados a los que se llega son diferentes a los resultados obtenidos en los pasados comicios, se podrá imputar esta diferencia a los atentados de Madrid, suponiendo que el único factor significativo fuera éste.

La metodología estadística que se utilizará en el desarrollo de este estudio **pasará** por los modelos de regresión, conjuntamente con las herramientas **descriptivas** y gráficas básicas.

Capítulo 3. Bases de datos

Por un lado se obtuvo la base de datos referente a la participación por correo y por otro lado la base de datos referente a la participación de los españoles ausentes. La base de datos final fue obtenida por la unión de las dos bases de datos.

3.1. Datos I. Participación por correo

Los datos referentes a los votantes que ejercieron su derecho mediante correo se bajaron de la página Web del Instituto Nacional de Estadística (www.ine.es). Entre todo el conjunto de datos nos interesaron los datos referentes al porcentaje de votaciones gestionada por el servicio de correos.

Los datos estaban desglosados por provincias y por comicios, y hacían referencia a las elecciones del 1989, 1993, 1996 y 2000. También se adjuntaron los datos referentes al número total de peticiones recibidas, el número total de votos enviados, etc., aunque decidimos que la mejor estimación de la participación era el porcentaje de votos enviados por provincia.

Dada la oficialidad de la fuente, se decidió no realizar una depuración y considerar que esos datos estimaban la participación real de los votantes que decidieron utilizar esta vía para ejercer su derecho.

Según la normativa explicada en la página Web del Ministerio de Interior (<http://www.mir.es/pciudadada/eleccion/votocorr.htm>), para la correcta tramitación de las votaciones, este tipo de votación será aceptada antes del tercer día previo a la votación, lo que nos situaría en el día anterior al de los atentados de Madrid (10 de marzo). Por lo tanto esta votación no estaría *contaminada* por el efecto de los atentados esperando que sea un buen factor de predicción de lo que hubiera sido la participación sin tales circunstancias.

3.2. Datos 2. Participación CERA

La base de datos principal, a partir de la que se trabajó, se obtuvo de la Subdirección General del Centro de Sistemas de Información. Éste organismo depende del Ministerio del Interior. Dada la oficialidad de los datos, decidimos que era innecesario un proceso de depuración de datos. Los datos están publicados y se pueden descargar libremente desde Internet.

La base de datos que utilizaremos está constituida por los datos referentes a todas las Elecciones Generales desde el 1977 hasta la del 2004, en total 9 comicios. Para cada Elección General se han obtenido un conjunto de ficheros con la información referente a los diferentes comicios; nivel estatal, nivel comunitario, provincial y municipal. Para cada consulta se han obtenido 10 ficheros, referentes a las Elecciones al Congreso de los Diputados, con la siguiente información:

- 01** = Fichero de control.
- 02** = Fichero de identificación del proceso electoral.
- 03** = Fichero de candidaturas.
- 04** = Fichero de candidatos.
- 05** = Fichero de datos comunes de ámbito municipal.
- 06** = Fichero de datos de candidaturas de ámbito municipal.
- 07** = Fichero de datos comunes de ámbito superior al municipio.
- 08** = Fichero de datos de candidaturas de ámbito superior al municipio.
- 09** = Fichero de datos comunes de mesas.
- 10** = Fichero de datos de candidaturas de mesas.

Para cada uno de los 90 ficheros existen registros con una longitud fija (rstras de números), donde cada campo, delimitado por una longitud predeterminada, registra la información de algún parámetro. Por lo tanto cada fichero está compuesto por una tira de caracteres limitado, sin espacios en

blanco. Así por ejemplo, el fichero de candidaturas (fichero 03), registra los siguientes campos:

Posiciones		Datos		
Inicio	Fin	Tipo	Long.	Descripción
1	2	Num.	2	Tipo de elección.
3	6	Num.	4	Año del proceso electoral.
7	8	Num.	2	Mes del proceso electoral.
9	14	Num.	6	Código de la candidatura.
15	25	Alf.	11	Siglas de la candidatura.
26	75	Alf.	50	Denominación de la candidatura.
76	81	Num.	6	Código de la candidatura cabecera de acumulación a nivel provincial.
82	87	Num.	6	Código de la candidatura cabecera de acumulación a nivel autonómico.
88	93	Num.	6	Código de la candidatura cabecera de acumulación a nivel nacional.

Tabla 1. Descripción fichero 3

La Subdirección General del Centro de Sistemas de Información, envió juntamente a los ficheros de datos un fichero donde explicaba cada uno de los campos y proporcionaba una tabla como la anterior para cada uno de los 10 ficheros de cada elección.

La extracción de los datos, fue un proceso largo y laborioso que se realizó con la ayuda del paquete SPSS. En este programa existen funciones, para utilizar de forma fácil las cadenas de caracteres. Para cada uno de los ficheros se extrajo una tabla en formato SPSS. Finalmente, tras decidir los datos que utilizaríamos para el análisis del objetivo (inicialmente existían muchísimas variables) se juntaron los ficheros necesarios para quedarnos con una sola base de datos, que incorporaba inicialmente las siguientes variables:

- Año de la elección (del 77 al 04).
- Número de votos de la participación general (válidos, nulos, total, votos en primer recuento, segundo recuento)
- Censo electoral (según INE)

- Número de votos de la población CERA (válidos, nulos, total)
- Censo CERA (según INE)
- Comunidad Autónoma (según codificación del INE)
- Provincia (según codificación del INE)

Sólo a partir de las elecciones del 1986 tenemos información referente al Censo de Extranjeros Ausentes (CERA), por lo que para los análisis donde se requiere información de los CERA se utilizarán los ficheros a partir de éste año. Debemos recordar que tenemos la participación por correos desde el 1989.

El conjunto de datos inicial era muy complejo, existían registros muy heterogéneos, tanto en votos, como en número de mesas, número de votos (en blanco, válidos, nulos y en primer recuento, etc.), partidos, candidatos, votos nulos, etc. y referentes al total Estatal, al nivel de Comunidad Autónoma, provincia y hasta municipio.

Finalmente se conjuntaron los datos aportados por la Subdirección General del Centro de Sistemas de Información y los datos bajados del Instituto Nacional de Estadística en una única base de datos. Esta base fue la que se utilizó para todos los análisis, aunque posteriormente se agregaron otros datos y se definieron nuevas variables.

3.3. Datos de interés para el análisis.

Entre todas las variables que teníamos, se decidió seleccionar las que a *priori* creíamos que serían buenas candidatas para la estimación de la participación.

Evidentemente la variable participación era fundamental (variable a explicar), en este caso se obtuvo los datos referentes a todos los comicios desde la historia de la democracia (del 1977 al 2004) y desde el nivel de municipio hasta el agregado total de España. Inicialmente se descartó el hecho de ajustar la participación desde el nivel de los municipios, pues carecía de relevancia y podía complicar mucho el modelo (tanto estadística como computacionalmente). Se decidió estudiar la participación a nivel provincial.

Las variables explicativas que entraron en la base de datos final fueron la participación del CERA, la participación de votantes por correo (ya que estas dos variables no estaban afectadas por los acontecimientos). La provincia o comunidad autónoma, el año de cada elección y el censo de los electores para cada provincia tanto de la población general como de la población CERA, también entraron en el modelo.

Posteriormente y como se irá explicando en el desarrollo del estudio se crearán otras variables a partir de éstas.

Capítulo 4. Participación

Según el objetivo del estudio, nos planteamos predecir la participación de las elecciones sin tener en cuenta el factor de los atentados ocurridos en Madrid el 11 de marzo de 2004.

4.1. Estimación de la participación

A partir de la aplicación estricta de la teoría de los modelos lineales generalizados deberíamos aplicar la metodología de la regresión logística. Ésta dice que para ajustar modelos que relacionen la probabilidad de un suceso (votación) y las variables explicativas, es necesaria la aplicación de este tipo de modelos, pues la variable a explicar es dicotómica. Parece claro que si definimos al individuo estadístico como cada persona que tiene derecho a voto y definimos el suceso como el hecho de ejercerlo, la metodología que deberíamos utilizar es la regresión logística, aunque en este caso se tiene el agregado (*porcentaje de individuos que ejercieron el voto en cada provincia*).

La regresión logística básicamente relaciona unos datos con una variable a explicar a partir de una función (*función link*). La diferencia básica entre los modelos lineales clásicos y los modelos logísticos es que esta función de link es distinta. Mientras que en los modelos logísticos, tienen en cuenta algunas propiedades concretas, como puede ser que la variable respuesta esté dentro

de rangos (0/1), los modelos lineales no los tienen en cuenta y utilizan la función identidad como link.

En el caso que nos ocupa, optamos por estimar la participación a partir del modelo lineal clásico. A partir del estudio descriptivo de la variable respuesta nos dimos cuenta de que no era fundamental la aplicación de modelos logísticos y de esta forma se simplificaría el modelo, debido a que la variable respuesta se podía tratar como una variable continua y Normal.

La diferencia fundamental en este caso es que no se tomará el individuo estadístico a la persona votante sino la proporción de votantes en una provincia. Así podemos comprobar que la distribución de los datos no nos presenta evidencias para tratar la variable como una variable de distribución continua y Normal (requisito para la aplicación de modelos lineales).

A partir de las observaciones recogidas de las cuatro últimas elecciones se observa que aunque el histograma no se ajusta bien a la distribución normal, a raíz de la prueba de Kolmogorov-Smirnov, no se puede rechazar la hipótesis nula de normalidad.

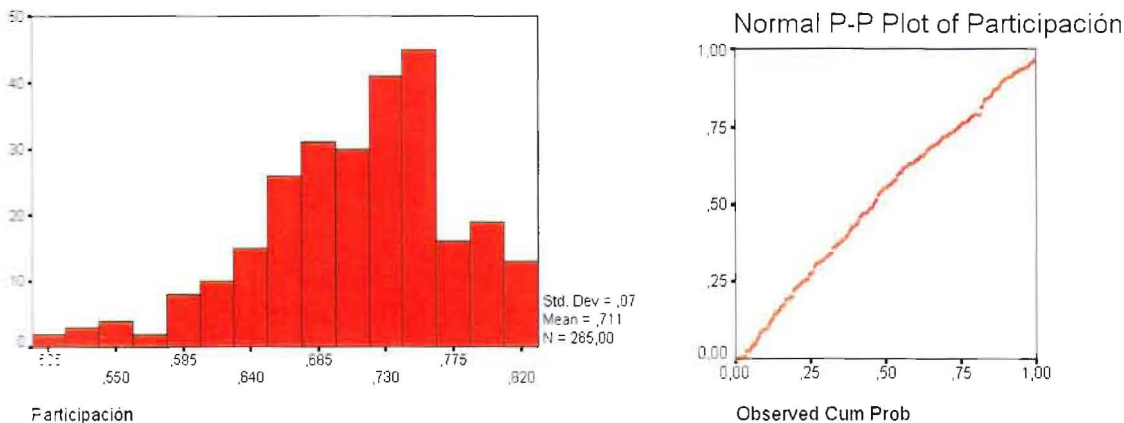


Figura 1. Estudios de normalidad de la participación

4.2. Introducción a los modelos de respuesta normal

La regresión estadística, relaciona una o más variables aleatorias de interés con una (modelo lineal simple) o varias variables respuesta (modelo lineal múltiple). A partir del estudio de las variables respuesta, se pueden deducir las causas/relaciones del suceso de interés si se cumplen unas circunstancias concretas. En este proyecto la variable de interés es la participación y las variables explicativas son los otros tipos de participación y las diferentes regiones de votación. La utilidad principal de la modelización estadística es la previsión de la variable de interés, bajo una condición concreta.

Adaptando este párrafo al análisis que nos interesa, el modelo lineal nos relacionará la participación general (*lo que queremos estimar*), con un conjunto de variables explicativas o informativas (*participación CERA, participación por correo, etc...*). A parte de ajustar un modelo que explique de forma satisfactoria los datos, esta herramienta ayudará a estimar la variable a explicar (participación general), bajo unas circunstancias concretas (atentados del 11M).

En los modelos lineales ordinarios, las componentes de la respuesta (participación), que denotaremos con una Y , tienen distribuciones normales, independientes con varianza constante.

$$Y = (y_1, y_2, \dots, y_n)' \sim N(X\beta, \sigma^2 I)$$

, donde X son las variables explicativas.

A partir de las observaciones recogidas de las cuatro últimas elecciones se observa que aunque el histograma no se ajusta bien a la distribución normal, a raíz de la prueba de Kolmogorov-Smirnov, no se puede rechazar la hipótesis nula de normalidad.

Nota: El estudio de Kolmogorov-Smirnov se realizó con las participaciones provinciales de los últimos años. En el transcurso de este capítulo entenderemos por qué utilizamos la participación provincial como individuo estadístico.

La componente sistemática del modelo consiste en la especificación de un vector η (predictor lineal), a partir de un número reducido de parámetros a estimar (β) y los regresores (X). La igualdad del modelo se ajusta mediante el error de previsión (ε), que se distribuye de forma normal, con media 0 y con desviación típica constante. Estos errores son independientes entre sí.

$$\left. \begin{array}{l} \beta' = (\beta_1, \beta_2, \dots, \beta_n) \\ X' = (X_1, X_2, \dots, X_n) \end{array} \right\} \rightarrow \eta = X\beta$$

Sea el modelo con variable de respuesta continua, $Y = \eta + \varepsilon = X\beta + \varepsilon$. La estimación de los parámetros puede realizarse de manera genérica como, $\text{Min}_{\beta} \sum_k M(\varepsilon_k) = \sum_k M(Y_k - x_k^T \beta)$, donde la función real M de los errores normalmente es el cuadrado de los mismos. El ajuste de parámetros a partir de la función de mínimos cuadrados, proporciona estimaciones no sesgadas y características deseables para la posterior inferencia de los resultados.

4.2.1. Modelo lineal general. Caso de la participación

Como se ha comentado en los dos apartados anteriores, intentaremos ajustar un modelo lineal para estimar la participación. La variable respuesta del modelo será la proporción de población que fue a votar en cada año a cada provincia. Las variables explicativas serán, inicialmente, la proporción de participación del CERA, la proporción de votantes por correo, el año de comicio y el distrito. Se prevé un último apartado del estudio donde se probarán otras variables explicativas. A continuación adjuntamos la notación de las variables que utilizaremos.

Aunque el objetivo inicial del proyecto sea el estudio inferencial de la participación general, se estimará la participación provincial y después se estimará la participación general agregándolos.

Variable respuesta	
$y_{ij}, i = 1..52, j = 1..4$	% participación en provincia i el año j .
Variables explicativas	
$c_{ij}, i = 1..52, j = 1..4$	% participación CERA en provincia i el año j .
$p_{ij}, i = 1..52, j = 1..4$	% participación correo en provincia i el año j . *
$a_j, j = 1..4$	Año del comicio
$d_i, i = 1..52$	Distrito de la elección (provincia)
$com_k, k = 1..19$	Comunidad Autónoma

*: porcentaje del censo al que se le ha tramitado el voto por correos

Tabla 2. Codificación de las variables

De forma general, cualquier modelo se puede escribir de la siguiente forma: $\pi_{ij} \equiv y_{ij} = x_{ij}^z \beta_{ij}^z + \varepsilon_{ij}$, donde los errores se distribuyen independientemente y de forma Normal, centrados en el cero. El predictor lineal ($\eta = X\beta$) es una combinación (de las z combinaciones posibles) de variables explicativas (se

han expuesto anteriormente), formado por el producto de los coeficientes y los datos. Por lo que ajustar el modelo implica encontrar la mejor combinación de variables explicativas que minimicen los errores.

El problema se abordará desde distintos puntos de vista, ajustando diferentes modelos. De esta forma podremos sacar conclusiones de ellos y podremos concluir con un único modelo. Si no se puede concluir con un único modelo, al menos se intentará estudiar los factores que influyen en la participación. Se prevé ajustar tres tipos de modelos:

Modelo 1: $\pi_{ij} \equiv y_{ij} = \beta_{0ij} + x_{ij}^z \beta_{ij}^z + \varepsilon_{ij}$, ($z = 1 \dots 6$) donde:

$$x_{ij}^1 = (c_{ij}), x_{ij}^2 = (a_j), x_{ij}^3 = (p_{ij})$$
$$x_{ij}^4 = \begin{pmatrix} c_{ij} \\ a_j \end{pmatrix}, x_{ij}^5 = \begin{pmatrix} p_{ij} \\ a_j \end{pmatrix}, x_{ij}^6 = \begin{pmatrix} c_{ij} \\ p_{ij} \\ a_j \end{pmatrix}$$

Para el modelo 1 se ajustarán seis tipos de regresiones. El modelo 1 servirá en cierto modo como descripción del problema, debido a que se ajustarán cada una de las variables una a una. En este modelo se estimará la participación de cada una de las provincias, teniendo en cuenta el año de la elección. Para estimar la participación general en el año 2004 se utilizará las variables explicativas de este año.

Modelo 2: $\pi_{ij} \equiv y_{ij} = \beta_{0ij} + \sum_{i=1}^{i=52} \beta_{1ij} d_{ij} + \beta_2 a_{ij} + \beta_3 c_{ij} + \beta_4 p_{ij} + \varepsilon_{ij}$

Se estimará un único modelo, donde se tendrá en cuenta el efecto de cada una de las provincias sobre el total de la participación. En este caso se tratará el efecto de la participación de los españoles extranjeros y los que votaron vía

postal como común a todas las provincias para predecir con el modelo estimado.

Modelo 3: Modelos con variables explicativas de nueva creación.

En este último modelo se calcularán nuevas variables explicativas a partir de los conocimientos conseguidos en los modelos 2 y 3. Lo que se pretende en este último apartado del proyecto es inyectar al mejor modelo la información que se haya adquirido en el transcurso de éste, para poder mejorar el mejor modelo encontrado hasta el momento.

Capítulo 5. Modelo 1

Modelo 1: $\pi_{ij} \equiv y_{ij} = \beta_{0ij} + x_{ij}^z \beta_{ij}^z + \varepsilon_{ij}$, donde:

$$x_{ij}^1 = (c_{ij}), x_{ij}^2 = (a_j), x_{ij}^3 = (p_{ij})$$
$$x_{ij}^4 = \begin{pmatrix} c_{ij} \\ a_j \end{pmatrix}, x_{ij}^5 = \begin{pmatrix} p_{ij} \\ a_j \end{pmatrix}, x_{ij}^6 = \begin{pmatrix} c_{ij} \\ p_{ij} \\ a_j \end{pmatrix}$$

Lo que se intentará en este capítulo es ajustar la participación en las elecciones anteriores a partir de modelos lineales de respuesta normal, para cada una de las 52 provincias. El resultado será la estimación de la participación en cada una de las provincias. Escogeremos el modelo que ajuste mejor a partir de los criterios de selección habituales de los modelos.

La participación se ajustará a partir de seis modelos que combinarán las tres variables que hemos escogido (participación CERA, año de la elección y participación por correo). Para cada provincia (y el agregado de comunidad y estado) y para cada uno de los años, tenemos los datos completos de los dos tipos de participación.

5.1. Estimación de la participación general

Una vez estimadas la participación de cada una de las provincias, para estimar la participación general del país, realizaremos la suma ponderada de cada una de las participaciones. Para ponderar cada una de las participaciones se utilizará el censo de cada una de las provincias, según el año que interese estimar. Debemos realizar la hipótesis de que la votación de una provincia es independiente a la participación de otra, por lo que consideraremos que la covarianza entre las provincias es cero.

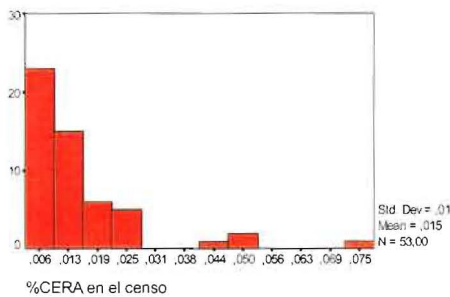
Una vez escogido el mejor conjunto de variables explicativas, se realizará la estimación de la participación ($\hat{\pi}_{ij}$), en la provincia de referencia y para el año de interés. Una vez estimadas todas las participaciones provinciales se estimará la participación general de todo el país, según la ecuación 5.1 (cuyo desarrollo se estudiará en el punto 5.7 de este capítulo).

$$\hat{\pi}_j = \frac{\sum_{i=1}^{52} w_{ij} \hat{\pi}_{ij}}{\sum_{i=1}^{52} w_{ij}}, j = 1..4 \quad (5.1)$$

Una vez realizados los cálculos, se obtendrá una estimación puntual de la variable que estamos interesados en conocer. Esta variable será la combinación lineal de un conjunto de 52 variables aleatorias (participaciones provinciales), que se ha supuesto que se distribuían de forma normal por lo que ésta también se distribuye según una Normal. Los dos momentos caracterizadores de la distribución igualmente se estimarán a partir de las 52 estimaciones provinciales.

5.2. Estimación de los parámetros

En el proceso de estimación de parámetros, se ponderarán los datos de cada año según el censo CERA. Es conocido que existen provincias que históricamente tienen población ausente emigrante diferente a otras provincias. Por ejemplo Galicia tiene una proporción no despreciable del censo electoral que depende de los gallegos que viven fuera de la comunidad. A modo de curiosidad se computó el cociente entre el censo CERA y censo electoral obteniendo los siguientes resultados:



Provincia	Cociente
Ourense	7,591%
Pontevedra	4,898%
A Coruña	4,838%
Lugo	4,299%
León	2,677%
Total España	1,4061%

Figura 2. Cociente entre censo CERA y censo INE

La estimación de los parámetros se realizará a partir del algoritmo de mínimos cuadrados, teniendo en cuenta dicha ponderación.

5.3. Modelos con la variable CERA

El objetivo de este punto es estimar la participación de cada provincia sólo con la variable explicativa participación CERA.

Se observa que la participación CERA está entre el 20% y el 35%, observándose sistemáticamente que la participación en 1993 y 1996 era elevada, acompañándose de una gran participación general. En 1989 se observó que aunque la participación CERA era elevada, la participación general no lo fue tanto. En el otro caso se observó que para participaciones del censo CERA bajas, coincidía con participaciones generales bajas. Tras este simple análisis notamos que existe una relación positiva entre ambas participaciones aunque ésta viene perturbada por las elecciones del 1989.

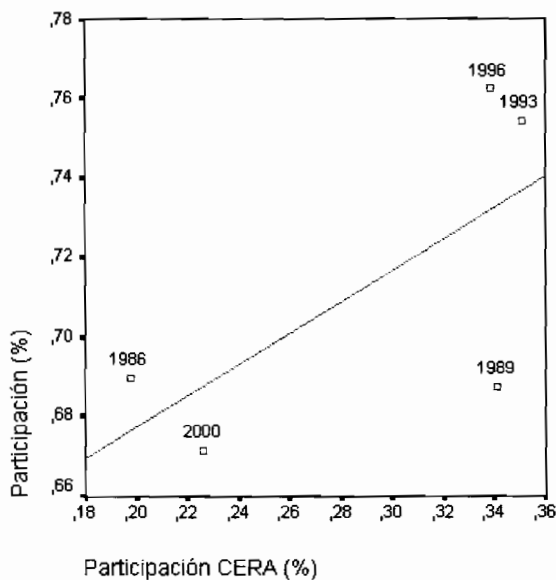


Figura 3. Participación según participación CERA en la media española

Una vez modeladas las regresiones vemos que este tipo de modelo proporciona una R^2 media del 61% de la suma de los cuadrados.

Como se comentó al principio del capítulo, los primeros modelos, servían como descripción de los datos.

5.4. Modelos con la variable AÑO

A partir de los modelos ajustados observamos que las participaciones en los distintos años no eran iguales. En los años 1986, 1989 y 2000 se observaron participaciones inferiores a los años 1993 y 1996.

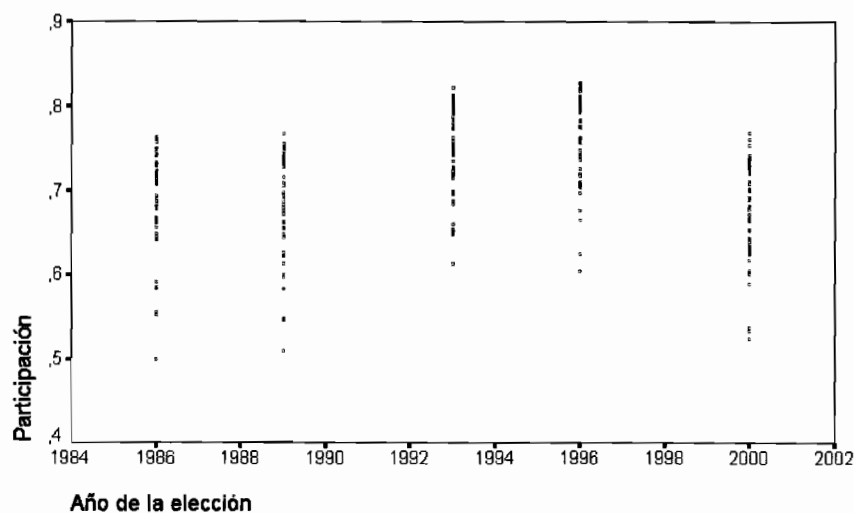


Figura 4. Participación según año de la elección en las 52 provincias

También advertimos que existen provincias que sistemáticamente tienen participaciones inferiores a la global. Por ejemplo Ceuta y Melilla tienen una participación media del 58% y del 57% respectivamente, en el caso de Sta. Cruz de Tenerife se calcula una participación media del 59%. En contra aparecen provincias que sistemáticamente tienen elevadas participaciones, así Toledo presenta una participación media del 79%. La participación media de todo el país es del 67%.

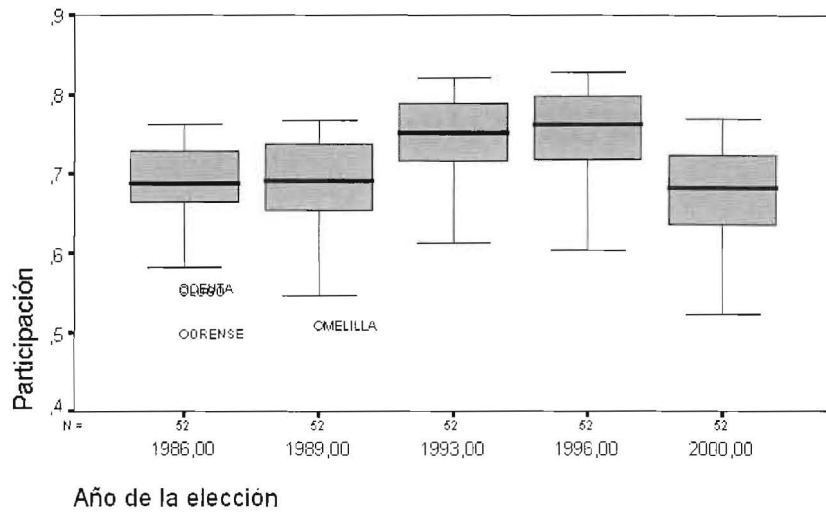


Figura 5. Participación según año de la elección en las 52 provincias

5.5. Modelos con la variable *POSTAL*

En este apartado se realizaron regresiones del mismo modo que en el apartado 5.3 pero con la participación anticipada por correos. En este caso los puntos no se ponderan por el censo CERA, puesto que no procedía dar mayor importancia según el censo de éstos.

Una vez realizadas las 52 regresiones, para cada una de las provincias, la conclusión fue que no era una buena variable explicativa por ella misma. En todos los casos menos una (Guipúzcoa) el coeficiente no era significativamente distinto de cero.

En puntos anteriores ya se han comentado comportamientos diferentes a los medios de algunas provincias en particular (Ceuta, Melilla, Sta. Cruz de Tenerife), en este punto se detecta un comportamiento particular de la provincia de Guipúzcoa. En el transcurso del documento nos daremos cuenta de que existen un conjunto de provincias que por sus características (*geográficas, políticas, ...*) presentarán continuamente resultados diferenciados del resto.

En el siguiente gráfico se observan los puntos de la participación general en función de la participación por correo en todas las provincias simultáneamente. En este caso se observa que la regresión lineal parece constante, sin pendiente, lo que se deduce del coeficiente no significativamente diferente a cero. Lo que sí parece es que haya una progresión en el tiempo, esto es que cada elección parece que la población votara más por correo y no tanto en la urna, aunque es en el siguiente punto donde se estudiará esto.

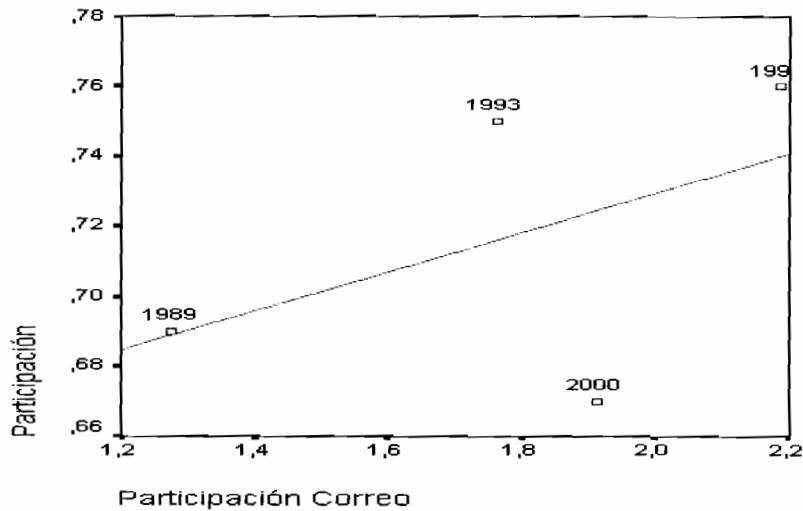


Figura 6. Participación según participación postal en la media española

Inicialmente se pensó que esta variable sería de interés, pero se observó que modelos a partir del año de comicio eran mejores que los propuestos a partir de la participación por correo. Por lo tanto, si tuviéramos que escoger entre estos dos modelos, nos quedaríamos con que incluye el año de elección, como única variable explicativa. Debemos recordar que el modelo a partir del año no nos sirve para el objetivo que persigue el proyecto ya que este dato puede ser bueno en la predicción del modelo pero no fue afectado por el efecto de los atentados.

5.6. Modelos con las variables *POSTAL + AÑO*

En este caso se estimó la participación provincial a partir de la participación gestionada por Correos y el año de la elección, tratándola como variable continua.

Una vez realizadas las 52 regresiones concluimos que este modelo no es adecuado puesto que aunque el año de la elección es una variables explicativa importante (en todas las regresiones presenta el coeficiente estadísticamente diferente de cero), no es el caso de la participación por correo.

Aunque no era significativa la participación de correo en casi todas las provincias, sí se detectó una provincia donde tomaba una importancia notable (p -valor = 0,0072). Esta provincia era Guipúzcoa, dónde ya se había detectado una relación relevante en el punto anterior. Si nos fijamos en el gráfico que se adjunta a continuación se puede observar la gran relación que guardan ambas variables, también se debe remarcar que la participación fue inferior a la media estatal (parecido fenómeno pasa en Ceuta y Melilla).

Dado el contexto político de esta provincia se puede asumir que la importancia de esta variable esta confundida con otros factores y mientras en otras provincias se vota vía correo por unos motivos determinados, en ésta podría ser por otros motivos, por lo que asumimos que esta regresión particular se aparta del objetivo del estudio y sirve como mera ilustración.

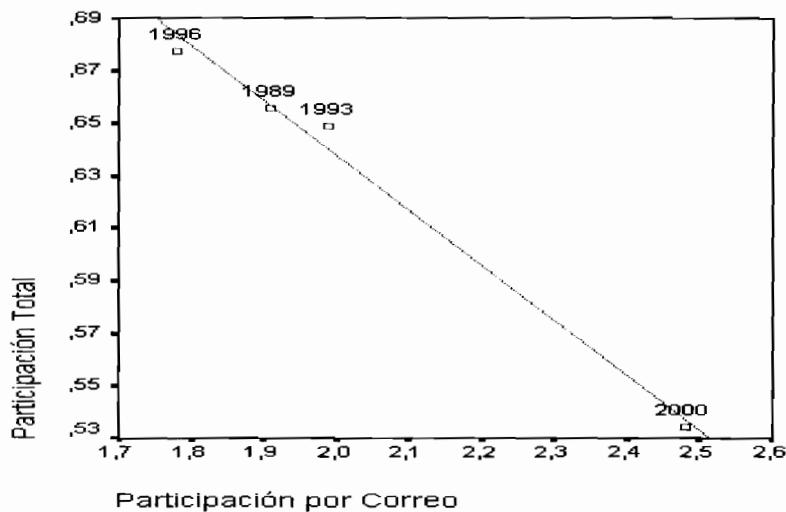


Figura 7. Relación entre participación en urna y participación postal en Guipúzcoa

La conclusión general de este punto es que la participación por correo no es un buen indicador de la participación general. Lo que se observa de forma repetida es que independientemente de la participación en las urnas, la participación por carta es parecida cada año. Esta conclusión, refuerza el caso de Guipúzcoa. En este caso se detecta una relación significativamente distinta a cero y negativa.

5.7. Modelos con las variables CERA + AÑO

En este caso se estimó la participación a partir de la participación de los españoles extranjeros y el año de la elección, tratándola como una variable continua.

Una vez realizadas todas las regresiones, se observa que en gran parte de las provincias la combinación de participación CERA y año son buenas variables predictoras, por lo que en los contrastes independientes para ver si los parámetros son significativamente distintos de cero alcanzan la significación estadística (*fijada en 0,05*).

En algunas provincias la participación CERA no es significativamente diferente a cero, pero en todas las provincias sí lo es el año (tomada como variable continua).

Una vez estimados los 52 modelos (uno por cada provincia), se debía estimar la participación general. Se ha de comentar que se decidió ajustar los modelos sin tener en cuenta un coeficiente independiente, por lo que no tiene sentido hablar de R^2 (recordar que obligamos que la recta pase por el punto (0,0)).

De momento éste es el mejor modelo que se ha encontrado, por lo que nos decidimos a estimar la participación.

5.7.1. Estimación de la participación general

Recordamos el modelo que estábamos ajustando:

Modelo 1.5: $\pi_{ij} \equiv y_{ij} = \beta_{0ij} + x_{ij}^4 \beta_{ij}^4 + \varepsilon_{ij}$, donde: $x_{ij}^4 = \begin{pmatrix} p_{ij} \\ a_j \end{pmatrix}$

Estimamos la participación en el año j , tal y como se había enunciado en la ecuación 5.1:

$$\hat{y}_j = \frac{\sum_{i=1}^{52} w_i \hat{y}_{ij}}{\sum_{i=1}^{52} w_i}, \text{ donde } w_i \text{ es el censo con derecho a ejercer el voto en la}$$

provincia i el año j (el que se quiere estimar). La participación es una variable aleatoria continua con distribución normal, por lo que debemos estudiar sus dos momentos.

$$\left\{ \begin{array}{l} E[\hat{y}_j] = E\left[\frac{\sum_{i=1}^{52} w_i \hat{y}_{ij}}{\sum_{i=1}^{52} w_i} \right] = \frac{E\left[\sum_{i=1}^{52} w_i \hat{y}_{ij} \right]}{E\left[\sum_{i=1}^{52} w_i \right]} = \frac{\sum_{i=1}^{52} w_i E[\hat{y}_{ij}]}{\sum_{i=1}^{52} w_i} \\ V[\hat{y}_j] = V\left[\frac{\sum_{i=1}^{52} w_i \hat{y}_{ij}}{\sum_{i=1}^{52} w_i} \right] = \frac{V\left[\sum_{i=1}^{52} w_i \hat{y}_{ij} \right]}{\left[\sum_{i=1}^{52} w_i \right]^2} = \frac{\sum_{i=1}^{52} w_i^2 V[\hat{y}_{ij}]}{\left[\sum_{i=1}^{52} w_i \right]^2} \end{array} \right.$$

En este caso suponemos que no existe covarianza entre la participación de dos provincias para un año concreto.

Para estimar la participación en las elecciones del año 2004, se utilizaron los datos publicados por el Instituto Nacional de Estadística. A partir de la participación CERA se pudo estimar para cada una de las provincias la participación que se esperaba en el 2004 si no hubieran afectado los factores que la hacían diferente a las otras elecciones.

Una vez estimadas las participaciones provinciales y gracias a los datos referenciados en el INE, en cuanto al censo provincial, estimamos la participación general a partir de la anterior ecuación. La estimación obtenida de la participación general fue del 70,05% y una varianza de 0,001837%. En un 95% de confianza sabemos que la participación debería haber estado entre un 69,19% y el 70,91%. Realmente la participación que hubo el 14 de marzo de 2004 fue del 75,66%, muy superior a la esperada según el modelo que acabamos de estimar.

Suponiendo que la estimación no está sesgada, podemos concluir que los hechos implicaron que un 5,61% fueran a votar, cuando no se esperaba que lo hicieran.

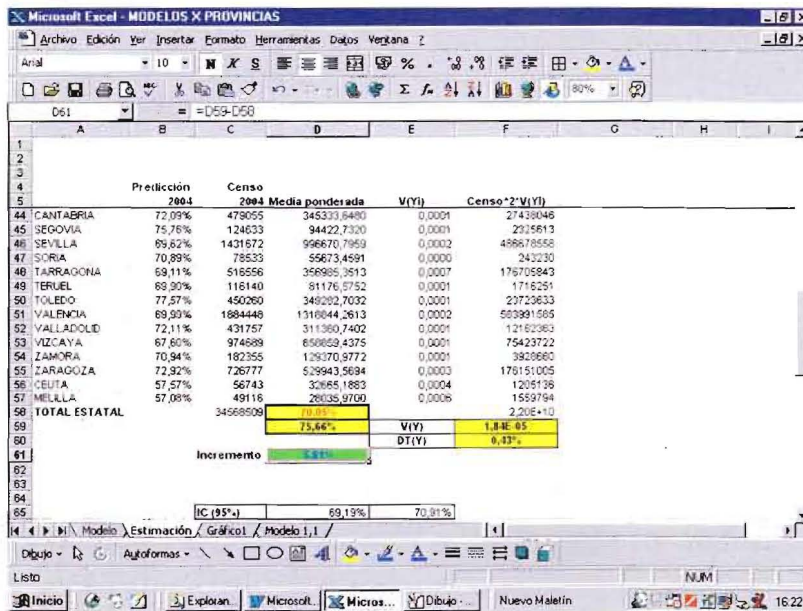


Figura 8. Estimación de la participación general a partir del modelo 1.4.

Proceso de Estimación: Para cada una de las provincias se estimó la participación en 2004 a partir de los coeficientes concretos fruto de cada una de las regresiones (predicor lineal). A partir del censo electoral de 2004 se calculó la media ponderada para cada una de las provincias. También se estimó la variabilidad para cada una de las provincias y a partir de fórmula expuesta anteriormente se calculó la estimación de la participación total, con la variabilidad estimada. De esta forma se pudo aportar una estimación puntual acompañada de un intervalo de confianza.

5.8. Modelos con las variables AÑO + CERA + CORREO

En este caso se intentó estimar la participación provincial a partir del modelo maximal.

Por falta de efectivos no se pudo estimar los parámetros de este modelo. Debemos recordar que sólo se tenían los datos completos de las tres variables, para las cuatro últimas elecciones.

El modelo que se plantea en este punto consume más grados de libertad que observaciones tiene por grupo. Se deben estimar los tres coeficientes referentes a las variables explicativas, más el coeficiente constante.

Al intentar estimar la participación a partir de las explicativas y sin tener en cuenta el coeficiente independiente, se constató que en cada una de las provincias los parámetros estimados no alcanzaban la significación estadística.

5.9. Conclusiones del modelo 1

A partir de los seis modelos que comprendían el modelo 1, se decidió estimar el modelo a partir de las variables CERA y AÑO. La participación general esperada en las pasadas elecciones era inferior (5,61%) a la participación que realmente hubo.

A continuación se adjunta el gráfico con los errores de ajuste en función de las participaciones esperadas (*diferencia entre participación real y participación esperada*). Se puede observar que casi todas las provincias tuvieron una participación superior a la que se les estimó.

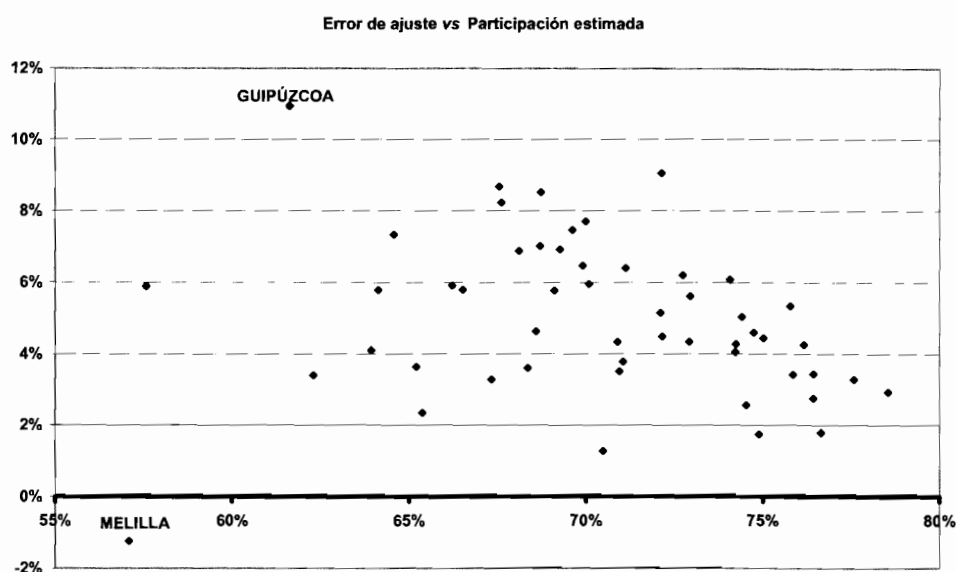


Figura 9. Participación estimada vs. error de ajuste (modelo 1)

Particularmente se debe comentar la situación de Melilla y de Guipúzcoa, la primera presentó una participación incluso inferior a la que se le otorgaba. En el caso de Guipúzcoa, se debe comentar que aunque se le había previsto una participación baja (*como normalmente se da en todas la elecciones*), fue en este caso muy superior a la prevista (+ 10,94%).

En los anexos (Anexo I) se adjuntan las estimaciones realizadas.

Los coeficientes ajustados se adjuntan a continuación.

	CERA	AÑO		CERA	AÑO		CERA	AÑO		CERA	AÑO
ALAVA	0,00%	3,4	CORDOBA	33,07%	3,3	LUGO	0,00%	3,3	SEGOVIA	0,00%	3,8
ALBACETE	45,73%	3,1	A CORUÑA	58,54%	2,5	MADRID	43,90%	3,1	SEVILLA	48,81%	2,9
ALICANTE	62,04%	2,9	CUENCA	0,00%	3,9	MALAGA	45,29%	2,7	SORIA	34,67%	2,9
ALMERIA	41,15%	2,9	GIRONA	0,00%	3,5	MURCIA	53,54%	3,1	TARRAGONA	0,00%	3,4
AVILA	35,21%	3,3	GRANADA	58,13%	2,7	NAVARRA	60,02%	2,6	TERUEL	37,08%	3,0
BADAJOS	38,35%	3,3	GUADALAJARA	40,86%	3,2	OURENSE	85,65%	2,1	TOLEDO	29,66%	3,5
BALEARIS	0,00%	3,3	GUIPUZCOA	0,00%	3,1	ASTURIAS	43,58%	2,9	VALENCIA	73,54%	2,8
BARCELONA	0,00%	3,5	HUELVA	0,00%	3,4	PALENCIA	41,20%	3,2	VALLADOLID	45,86%	3,1
BURGOS	46,08%	3,0	HUESCA	0,00%	3,6	LAS PALMAS	0,00%	3,3	VIZCAYA	38,94%	2,8
CACERES	0,00%	3,8	JAEN	27,61%	3,5	PONTEVEDRA	0,00%	3,3	ZAMORA	42,02%	2,9
CADIZ	41,03%	2,8	LEON	0,00%	3,5	SALAMANCA	51,46%	3,0	ZARAGOZA	0,00%	3,6
CASTELLON	0,00%	3,8	LLEIDA	0,00%	3,4	TENERIFE	0,00%	3,1	CEUTA	0,00%	2,9
C. REAL	36,75%	3,3	LA RIOJA	0,00%	3,7	CANTABRIA	48,62%	3,0	MELILLA	0,00%	2,8

Nota: Todos los coeficientes de año están multiplicados por 10⁴.

Tabla 3. Coeficientes modelo 1

Capítulo 6. Modelo 2

Modelo 2: $\pi_{ij} \equiv y_{ij} = \beta_{0+} + \sum_{i=1}^{i=52} \beta_{1i} d_i + \beta_2 a_j + \beta_3 c_{ij} + \beta_4 p_{ij} + \varepsilon_{ij}$

El modelo que se plantea en este capítulo ajusta la participación en las elecciones anteriores a partir de modelos lineales de respuesta normal, para todas las provincias conjuntamente.

Se predecirá la participación de la provincia i en el año j a partir de los siguientes parámetros:

- a) Coeficiente independiente
- b) Coeficiente de la provincia (*distrito*) i ($i = 1 \dots 52$)
- c) Coeficiente del año j ($j = 1 \dots 4$)
- d) Coeficiente de participación CERA en la provincia i el año j
- e) Coeficiente de participación postal en la provincia i el año j

6.1. Diferencias entre modelo 1 y modelo 2

En este modelo, a diferencia del presentado en el capítulo anterior, pretende estimar la participación general de cada una de las provincias, pero teniendo en cuenta todos los datos. En el capítulo anterior se ajustaba una regresión (una ecuación) para cada una de las provincias; en este caso se ajustará una única regresión, que servirá para todas las provincias.

Así, en el primer caso se ajustaba la selección de participaciones referentes a una provincia para todos los años (4 datos → 4 individuos); en cambio en este caso se ajustarán todas las participaciones a la vez (4*52 datos → 4*52 individuos) y tras obtener un modelo general, se estimará cada una de las participaciones en cada provincia. Una vez estimada cada una de las participaciones se deberá estimar la participación general de la misma forma que en el capítulo anterior.

Sin apartarnos del objetivo principal del trabajo, se introducirán en el modelo las variables de participación de los españoles extranjeros (CERA) y los que decidieron votar vía postal (CORREO). Gracias a estas dos variables se estudiará la participación que hubiera habido si no se hubieran producido las condiciones que alteraron la participación (como ya se ha comentado estos datos no fueron afectados por los atentados). Debemos comentar que en este caso y como diferencia importante con el modelo precedente, hacemos la hipótesis que los coeficientes de las variables CERA y CORREO afectan por igual en todas las provincias (*la interpretación del modelo se estudiará en el siguiente punto*).

6.1.1. Utilización de pesos

La temporalidad de la participación se corregirá utilizando pesos en función del año de la elección. Se probaron los diferentes modelos utilizando cuatro tipos de pesos:

	1989	1993	1996	2000
Pesos 1	1/4	1/4	1/4	1/4
Pesos 2	1/8	1/8	1/4	1/2
Pesos 3	1/6	1/6	1/6	1/2
Pesos 4	1/15	2/15	4/15	8/15

Tabla 4. Pesos probados en el modelo 2

El primer peso pondera cada año con la misma importancia y supone que no existen diferencias entre ellos. En el caso del segundo peso, presupone que el 1989 y el 1993 tuvieron la misma importancia en la votación y da mayor importancia a la participación de los años 1996 y 2000, pesando el año 2000 el doble que 1996 y cuatro veces más que los dos primeros. En el tercero se presupone que los tres primeros años son iguales entre sí (en cuanto a importancia) y le asigna la mitad del peso al último año. Finalmente el último peso supone que la importancia de cada año se diluye de forma que al 1996 se le da la mitad del peso que la última elección, al 1993 cuatro veces menos, etc.

Los cuatro pesos fueron probados en todos los modelos de este capítulo, comprobando que el segundo y el cuarto eran los mejores. Se decidió expresar los resultados con el segundo de los pesos. El peso 1 y el peso 3 proporcionaban peores predictores que los otros.

No obstante en cada uno de los modelos que se plantean se introducirá el año de la elección y también se ajustará sin esta variable.

6.2. Interpretación del modelo

Según el modelo que se presenta se intentará estimar la participación de la provincia i ($i=1...52$) en el año j ($j=1989, 1993, 1996$ i 2000). Tal participación se estimará a partir de una constante, un parámetro para cada una de las provincias (menos una de referencia), un parámetro para la participación del CERA, otra para CORREO y según el modelo uno para el año de la elección (se estimarán con el año y sin el año). En total, estamos estimando $1 + 51 + 1 + 1 + (1) = 54$ (55) a partir de $4 \cdot 52$ datos, por lo que nos aseguramos de largo los grados de libertad. El modelo es éste:

$$\pi_{ij} \equiv y_{ij} = \beta_{0ij} + \sum_{i=1}^{i=52} \beta_{1i} d_i + \beta_2 a_j + \beta_3 c_{ij} + \beta_4 p_{ij} + \varepsilon_{ij}$$

La participación de la provincia i en el año j , será el resultado de sumar la estimación (objetivo de este estudio) más un error (error de ajuste).

Para introducir cada una de las provincias se construyó una variable dicotómica que distinguía cada una de las opciones, menos una. En este caso no se introdujo la primera de las provincias (Álava), esta decisión fue arbitraria pero irrelevante. Inicialmente se introdujeron los 51 parámetros que controlaban las 52 provincias y tras ajustar el modelo nos quedamos con aquellos que fueron significativamente distintos de cero. Para los coeficientes que son estadísticamente iguales a cero se supone una participación igual a Álava (provincia de referencia). Los coeficientes que entraron en el modelo estimaban la diferencia media entre la provincia de referencia (Álava) y la provincia que representaba.

El parámetro independiente (constante), no depende de ninguna variable explicativa y resulta la media de las participaciones de la provincia referencia (ya que todas las otras variables se consideran cero), asumiendo una participación media tanto del CERA como de los electores postales.

Si no se utilizara el término independiente se forzaría que la regresión pasara por el origen y por lo tanto la variabilidad explicada por los β_1 sería más elevada. En este caso todos los coeficientes provinciales serían significativos, pues seguro que se explicaría mejor la respuesta con un coeficiente de provincia (participación media de la provincia i) que con ninguno (absorbería la información los otros parámetros). En este caso también se reducen los grados de libertad, puesto que se detectarían provincias cuya participación sea equiparable a la de referencia.

Los coeficientes de la participación CERA y por correo, suponen que estas variables afectan de forma similar en cada una de las provincias. Sabemos que parte de la variabilidad que pueda existir entre este coeficiente para cada una de las provincias, lo podrá asumir el coeficiente estimado de cada una de las provincias. Así debemos considerar que el coeficiente estimado para la participación de la población CERA y la población postal, es la 'afectación' media de estas variables en la participación general.

6.3. Modelo 2.1

En una primera fase se probaron los modelos explicados con anterioridad, teniendo en cuenta los siguientes puntos:

- a) Los cuatro pesos (*peso1*, *peso2*, *peso3* y *peso4*).
- b) Modelos, ajustados según el algoritmo Stepwise implantado en el paquete estadístico S-PLUS.
- c) Modelos con coeficiente independiente y sin variable año.

Una vez estimados los parámetros de los cuatro primeros modelos, observamos que no existían muchas diferencias entre ellos. En todos los modelos entraba en juego las variables (CERA y postal). En los cuatro modelos había muchas provincias que adoptaban un coeficiente significativo.

En la mayoría de los casos se repetían las provincias que tenían una participación significativamente diferente a la de referencia. En todo caso se observaba que los modelos con peso 2 y peso 3 asignaban significación a provincias que el primer peso no detectaba y el peso 3 a provincias que el peso 2 no tenía en cuenta. Por lo que se detectó una asociación entre el hecho de ponderar los años próximos con mayor peso con el hecho de detectarse mayor número de provincias diferentes. Se debería estudiar en mayor profundidad qué cambió en los últimos años entre estas provincias para considerarse diferentes a la de referencia (Álava). Una primera hipótesis podría ser que estas provincias estaban al borde de la significación y al darse mayor importancia en algunos datos, sobrepasaran el límite. En todo caso nos apartaríamos de nuestro objetivo, por lo que continuaremos con el desarrollo de este punto.

Por otra parte se observó que existían provincias que sistemáticamente sobresalían de la normalidad del modelo. Así por ejemplo se detectó otra vez provincias de Euskadi y Galicia con comportamientos diferentes. Lo mismo sucedió con las ciudades autónomas de Ceuta y Melilla. En este caso se

observó que Vizcaya, Guipúzcoa y Navarra tenían una participación inferior, al igual que Ourense y ambas ciudades autónomas (*coeficientes provinciales estadísticamente diferentes de cero y negativos*). También se detectaron participaciones inferiores en las provincias de Soria y Santa Cruz de Tenerife. Para cada una de estas provincias seguramente existe algún factor no recogido como puede ser la localización geográfica o alguna variable que recogiera la percepción del voto particular en relación al estado.

En cuanto a los coeficientes de la participación CERA y CORREO, indicar que en todos los modelos registraba significación estadística. También observar que el coeficiente de la población CERA era superior al de la postal, que a sabiendas que este tipo de participación también es más elevada, se concluye que parte de la participación general explicada será imputada a este colectivo. Esto es que la magnitud de la participación CERA es superior a la participación CORREO y su coeficiente de regresión también, por lo que la magnitud de participación general también será superior a la del CORREO.

La elección del modelo se realizó a partir del R^2 y de la prueba F del modelo. En este caso nos quedamos con el segundo modelo (*peso 2*), ya que recogía mejor la variabilidad ($R^2=0,8485$ - $F=26,6$).

Una vez estimado el modelo se graficaron los residuos, observándose que algunas provincias (Guipúzcoa, Melilla, Ourense, Lleida y Girona) para algunos años tenían un comportamiento atípico. En general el gráfico que se adjunta a continuación (*residuos vs. valores predichos*), concuerda con lo esperable. Debemos fijarnos que a la izquierda de él aparecen las provincias con poca participación (Ceuta, Melilla y provincias vascas).

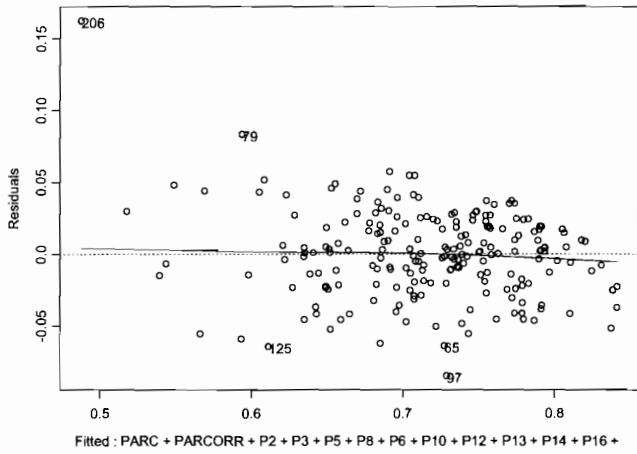


Figura 10. Residuos vs. predichos en modelo 2.1

La importancia que ejerce cada una de las provincias en la estimación de los parámetros se adjunta a continuación (*distancia de Cook*). En ella, se observa como algunas provincias ejercen mucha influencia en la estimación y todas las observaciones de alguna de las provincias en bloque tienen una importancia atípica. Así por ejemplo las observaciones 79 y 80 (Guipúzcoa en 1996 y 2000 respectivamente) son importantes en la estimación de los parámetros. En el mismo caso está la observación de Melilla en 1993 (observación 206).

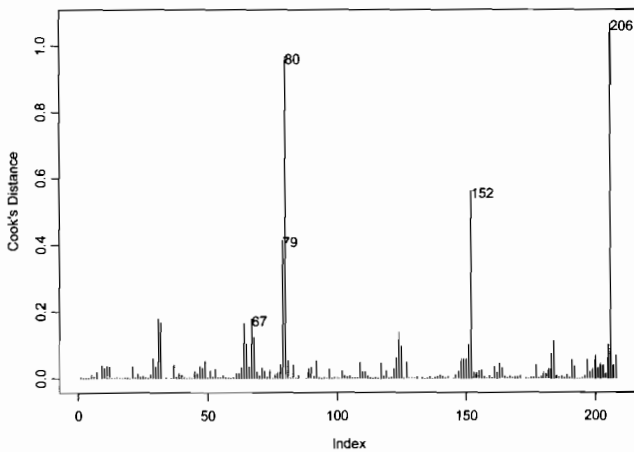


Figura 11. Distancia de Cook para modelo 2.1

Se estimó la participación a partir de los parámetros referentes al año 2004, calculando una estimación del 70,01% (recordar que en el capítulo anterior se estimó una participación del 70,05%). Comparando los dos modelos no se observaron diferencias notables. La verdadera participación a partir de este modelo, con una confianza del 95%, se estimó entre un 67,15% y 72,88%. El intervalo de confianza para este modelo fue notablemente mayor que el calculado para el primer modelo. Por último comentar que la participación en las pasadas elecciones se situó en el 75,66% de la población con derecho, por lo que se calcula un aumento de 5,61 puntos.

A la vista de estos resultados se deduce que la participación en las últimas elecciones fue diferente a la participación esperada.

Constante	51,04%	CASTELLON	7,45%	LA RIOJA	0,00%	SEGOVIA	6,08%
CERA	38,29%	CIUDAD REAL	6,52%	LUGO	0,00%	SEVILLA	6,16%
POSTAL	3,71%	CORDOBA	6,99%	MADRID	6,40%	SORIA	-6,55%
ALAVA	0,00%	A CORUÑA	0,00%	MALAGA	0,00%	TARRAGONA	0,00%
ALBACETE	6,74%	CUENCA	6,39%	MURCIA	9,22%	TERUEL	0,00%
ALICANTE	9,29%	GIRONA	2,54%	NAVARRA	-2,39%	TOLEDO	10,48%
ALMERIA	0,00%	GRANADA	2,45%	OURENSE	-2,55%	VALENCIA	10,00%
AVILA	4,28%	GUADALAJARA	4,41%	ASTURIAS	0,00%	VALLADOLID	6,05%
BADAJOS	8,23%	GUIPUZCOA	-10,82%	PALENCIA	3,18%	VIZCAYA	-3,04%
BALEARS	0,00%	HUELVA	3,44%	LAS PALMAS	0,00%	ZAMORA	0,00%
BARCELONA	3,13%	HUESCA	0,00%	PONTEVEDRA	0,00%	ZARAGOZA	2,55%
BURGOS	0,00%	JAEN	6,06%	SALAMANCA	2,81%	CEUTA	-22,21%
CACERES	5,33%	LEON	0,00%	TENERIFE	-2,49%	MELILLA	-20,30%
CADIZ	0,00%	LLEIDA	0,00%	CANTABRIA	3,77%		

Tabla 5. Coeficientes estimados modelo 2.1

6.4. Modelo 2.2

Los modelos que en este punto se proponen son idénticos a los modelos propuestos en el punto anterior excepto que se introduce la variable año, como variable continua.

Las provincias que tenían una participación significativamente diferente a la de referencia en los modelos 2.1, en este caso también alcanzaban la significación estadística. Los parámetros referentes a la participación CERA y la participación postal eran igualmente significativos.

Sabemos que el año no se puede introducir en el modelo como una variable continua, ya que realmente no mide ninguna magnitud sino que referencia momentos exactos de tiempo, debería considerarse de otras formas. Aunque podría haber tendencia lineal, parece razonable no introducirla en el modelo.

Aunque sabemos que existen modelos adecuados para tratar la variable año, decidimos pasar al próximo capítulo ya que en él se realiza un tratamiento especial de esta variable.

En todo caso se estimaron los parámetros observando resultados parecidos (el año aunque no era estadísticamente significativo se estimó).

Var	Coef	Var	Coef	Var	Coef	Var	Coef
Constante	-276,69%	CADIZ	0,00%	LLEIDA	0,00%	CANTABRIA	6,95%
AÑO	0,16%	CASTELLON	7,66%	LA RIOJA	2,23%	SEGOVIA	6,50%
CERA	45,20%	CIUDAD REAL	7,28%	LUGO	0,00%	SEVILLA	-5,20%
POSTAL	3,18%	CORDOBA	7,43%	MADRID	7,14%	SORIA	11,13%
ALAVA	0,00%	A CORUÑA	0,00%	MALAGA	0,00%	TARRAGONA	0,00%
ALBACETE	7,35%	CUENCA	7,59%	MURCIA	9,84%	TERUEL	0,00%
ALICANTE	9,64%	GIRONA	2,67%	NAVARRA	4,31%	TOLEDO	10,46%
ALMERIA	2,16%	GRANADA	3,05%	OURENSE	0,00%	VALENCIA	6,77%
AVILA	5,26%	GUADALAJARA	5,34%	ASTURIAS	0,00%	VALLADOLID	0,00%
BADAJOS	9,25%	GUIPUZCOA	-9,99%	PALENCIA	4,14%	VIZCAYA	0,00%
BILBAO	0,00%	HUELVA	4,29%	LAS PALMAS	0,00%	ZAMORA	0,00%

Var	Coef	Var	Coef	Var	Coef	Var	Coef
BALEARNS							
BARCELONA	3,33%	HUESCA	0,00%	PONTEVEDRA	0,00%	ZARAGOZA	3,17%
BURGOS	0,00%	JAEN	7,19%	SALAMANCA	4,70%	CEUTA	-19,70%
CACERES	6,38%	LEON	0,00%	SANTA CRUZ DE TENERIFE	0,00%	MELILLA	-17,94%

Tabla 6. Coeficientes Modelo 2.2

La estimación de la participación general obtenida fue del 63,81% [60,95% - 66,68%], por lo que este modelo le otorgaría a los atentados un aumento de la participación del 11,85%.

En el siguiente capítulo se intentarán ajustar modelos incorporando la información recogida en los capítulos anteriores, corrigiendo el hecho de tomar el año como variable continua.

Capítulo 7. Modelo 3

En este capítulo se proponen modelos con variables explicativas de nueva creación, esto es, definiendo variables a partir de las peculiaridades que se han visto en los capítulos anteriores.

7.1. Otros modelos finalmente descartados

La primera variable que probamos fue la agregación de las provincias por **comunidad autónoma**. Al ajustar estos modelos observamos que eran peores que los conseguidos hasta el momento, pues suponíamos que un conjunto de provincias juntas tenía la misma participación y se comprobó que esto no era así. Por lo que la conclusión que debíamos sacar es que las diferentes provincias de una misma comunidad tenían participaciones independientes. Aunque se ganaban grados de libertad el modelo dejaba de explicar mucha variabilidad. Los modelos probados tenían la estructura definida en el modelo 2.

Aún se realizó otra agregación y unimos las comunidades de **centro** y las **autonómicas**, empeorando otra vez el modelo. En este punto se quería ver si existía diferencias según "*ideologías políticas*", diferenciando las provincias "*autonómicas*" de las "*centralistas*". La justificación para rechazar estos modelos era parecida a las anteriores.

La conclusión que debemos sacar de los modelos probados es que el supuesto aumento de participación, no venía introducido por el hecho de votar en una comunidad autónoma en concreto o por la existencia de partidos autonómicos o centralistas.

A partir de los modelos ajustados hasta el momento, se entreveía que existía otro factor que no quedaba reflejado en las variables valoradas hasta el momento y que no dependía de la participación CERA ni de la participación postal.

7.2. Años de cambio

El año de los comicios tomado como variable continua en las regresiones lineales no era correcto, pues realmente no actuaba como tal sino como indicador de la elección. Ante este problema cabían varias posibilidades, la primera crear variables *dummy*, como se realizó con las provincias en el capítulo anterior, aunque esta posibilidad era factible a nivel exploratorio de los datos, no lo era para predecir ya que no entra la *dummy* de 2004. También se podía realizar regresión utilizando modelos mixtos y seguramente se contemplarían otras vías.

Nos dimos cuenta de que los años tomados como variables *dummy*, tenían coeficientes muy diferentes. Observamos que los años donde podía haber un posible cambio de gobierno, la población acudía a las urnas en mayor proporción que en elecciones donde los sondeos daban una amplia victoria al partido que ya estaba gobernando. Como se ha comentado, este tipo de modelo introducirá los coeficientes de los años de las últimas elecciones pero evidentemente no introducirá el coeficiente del año 2004, por lo que no servirá para predecir la participación.

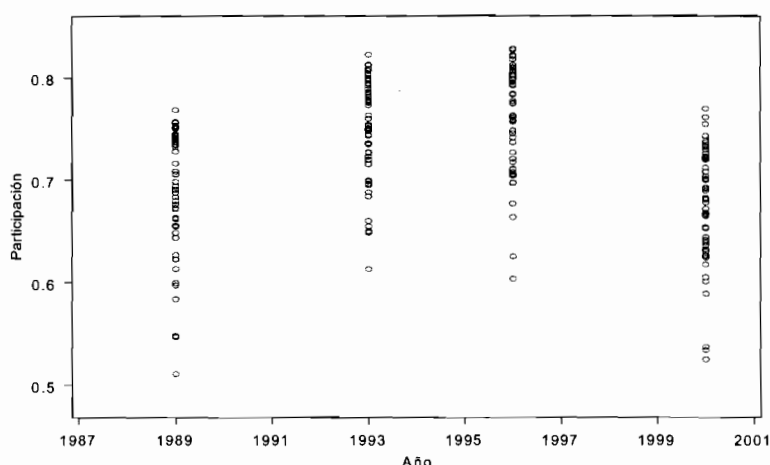


Figura 12. Participación provincial según año de elección

Decidimos crear una nueva variable que tomaba el valor 0 si el año de la elección era el 1989 (reelección del PSOE) y 2000 (reelección del PP), indicando **no cambio** político. El 1996 tomó valor 1 puesto que fue el año donde el PP accedió a la Moncloa y el 1993 tomó el mismo valor, puesto que fue el año en el que el PSOE ganó por muy poco, cuando los sondeos preelectorales pronosticaban un cambio de poder.

La nueva variable daba mucho juego en nuestro proyecto pues podíamos estimar la participación de 2004 teniendo en cuenta ambas situaciones (como si la población hubiera participado asumiendo un cambio político o sin asumirlo).

Entrando en la modelización estadística, se decidió utilizar el modelo del capítulo anterior (puesto que era más general que el primero) y añadir la nueva

variable $(\pi_{ij} \equiv y_{ij} = \beta_0 + \sum_{i=1}^{i=52} \beta_{1i} d_i + \beta_2 cam_j + \beta_3 c_{ij} + \beta_4 p_{ij} + \varepsilon_{ij})$. Una vez

estimado el nuevo modelo se observó que el R^2 aumenta hasta un 95,69%, cuando el modelo sin la variable *cambio* tenía una $R^2 = 84,85\%$.

Se observó que al introducir la nueva variable, que era estadísticamente muy diferente de cero (p-valor $\ll 0,001$; $t = 20,83$). El coeficiente estimado fue de 7.55 puntos más en las elecciones donde existía posibilidad de cambio, por lo que se deduce que en una provincia determinada existía un aumento medio de 7,55 puntos en unas elecciones con posible cambio político, que en unas elecciones sin tal suceso.

Casi todas las variables indicadoras de provincia entraban en el modelo (excepto las provincias catalanas, Huelva, Málaga, Lugo, Navarra y Vizcaya). Este fenómeno se podría explicar por el hecho que la nueva variable explica gran cantidad de la información que antes explicaban las provincias y el resto la explican las peculiaridades de cada una de las provincias.

Gráficamente, se observa que existe un patrón parecido en cada año, por lo que explicado el efecto de la nueva variable se podría observar que para

cada elección existe un comportamiento parecido. Un ejemplo muy claro se da en Huesca, los años donde no se preveía un cambio de presidente hubo una participación del 69% (tanto en 1989 como 2000), pero cuando se preveía un cambio de presidente la votación aumentó hasta el 76% (misma participación en ambos años, dos a dos).

Según el algoritmo *Stepwise* incorporado en el paquete S-plus, la variable participación por correo no entraba en el modelo a priori, aunque se decidió introducirla en el modelo ya que es una de las variables principales.

A continuación se adjunta el gráfico de residuos, observándose que en el 1989 Guipúzcoa (observación 77), Ourense (observación 125) y Lugo (observación 105) quedaban mal representadas (error de predicción elevado) por el modelo. A parte de estas observaciones, el conjunto del gráfico no nos induce a sospechar de la distribución aleatoria de los residuos.

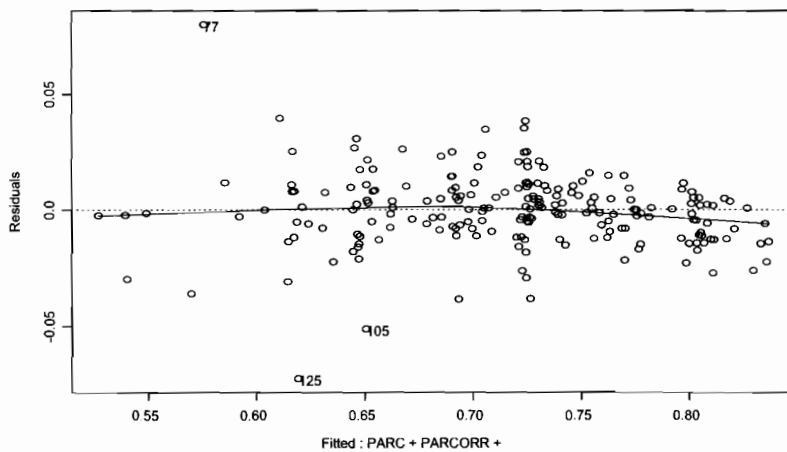


Figura 13. Residuos vs. predichos en modelo 3

La participación estimada a partir de este modelo, suponiendo el **no cambio** político, fue del 66,06% (64,49% - 67,63%); 9,6 puntos inferior a la que **realmente** hubo.

Pero la participación estimada, asumiendo un cambio de poder, fue del **73,60%**, sólo 2,06% inferior a la que realmente hubo. Si en este caso se tienen

en cuenta los intervalos de confianza, se observa que a partir del modelo se estima una participación entre el 72,03% y el 75,17%, por lo que la participación real (75,66%) quedaba fuera de la estimación.

CONSTANTE	63,89%	CADIZ	-2,12%	LLEIDA	7,31%	CANTABRIA	5,44%
CAMBIO	7,55%	CASTELLON	7,65%	LA RIOJA	0,00%	SEGOVIA	7,66%
CERE	5,05%	CIUDAD REAL	8,72%	LUGO	5,96%	SEVILLA	4,51%
CORREO	-1,45%	CORDOBA	8,67%	MADRID	0,00%	SORIA	2,51%
ALAVA	0,00%	A CORUÑA	-3,22%	MALAGA	8,12%	TARRAGONA	5,09%
ALBACETE	9,70%	CUENCA	11,44%	MURCIA	0,00%	TERUEL	0,00%
ALICANTE	7,72%	GIRONA	4,03%	NAVARRA	-2,92%	TOLEDO	11,18%
ALMERIA	2,80%	GRANADA	0,00%	OURENSE	0,00%	VALENCIA	7,28%
AVILA	8,25%	GUADALAJARA	8,55%	ASTURIAS	1,69%	VALLADOLID	8,13%
BADAJOS	9,57%	GUIPUZCOA	-7,38%	PALENCIA	8,71%	VIZCAYA	4,01%
BALEARS	-3,12%	HUELVA	4,16%	LAS PALMAS	-3,05%	ZAMORA	0,00%
BARCELONA	5,95%	HUESCA	0,00%	PONTEVEDRA	-1,74%	ZARAGOZA	4,99%
BURGOS	0,00%	JAEN	10,84%	SALAMANCA	5,84%	CEUTA	-9,19%
CACERES	9,07%	LEON	3,57%	TENERIFE	-5,69%	MELILLA	-10,72%

Tabla 7. Coeficientes del modelo 3

Capítulo 8. Discusión y conclusiones

En el desarrollo del documento se han discutido tres posibles modelos. Los tres modelos se suponen correctos en cada uno de los capítulos por separado. En este capítulo se discutirá la selección de un único modelo a partir del estudio de los tres.

8.1. Modelo 1

En cada uno de los modelos se pretendía estimar la participación de cada una de las provincias, para una posterior estimación de la participación general.

Modelo 1: $\pi_{ij} \equiv y_{ij} = \beta_{0ij} + x_{ij}^z \beta_{ij}^z + \varepsilon_{ij}$, donde:

$$x_{ij}^1 = (c_{ij}), x_{ij}^2 = (a_j), x_{ij}^3 = (p_{ij})$$

$$x_{ij}^4 = \begin{pmatrix} c_{ij} \\ a_j \end{pmatrix}, x_{ij}^5 = \begin{pmatrix} p_{ij} \\ a_j \end{pmatrix}, x_{ij}^6 = \begin{pmatrix} c_{ij} \\ p_{ij} \\ a_j \end{pmatrix}$$

El primer modelo buscaba una recta de regresión para cada una de las provincias, por lo que a partir de los datos teníamos una estimación diferente para cada una de ellas. Las variables explicativas que entraron en el modelo fueron el año de elección y la participación de la población CERA (x_{ij}^4).

Posteriormente, estimando la participación general ($\hat{y}_j = \frac{\sum_{i=1}^{52} w_i \hat{y}_{ij}}{\sum_{i=1}^{52} w_i}$), se calculó

una participación del 70,05% (69,19% - 70,91%), cuando realmente fue del 75,66%, lo que implica un aumento del 5,61% (ver Anexo, para estimación y coeficientes).

MODELO 1			
Participación estimada	70,05%	V(Y)	1,84E-05
Participación real	75,66%	DT(Y)	0,43%
Incremento	5,61%		
IC (95%)	69,19%	70,91%	

Tabla 8. Resumen y estimación del modelo 1

8.2. Modelo 2

Modelo 2: $\pi_{ij} \equiv y_{ij} = \beta_{0+} + \sum_{i=1}^{i=52} \beta_{1i} d_i + \beta_2 a_j + \beta_3 c_{ij} + \beta_4 p_{ij} + \varepsilon_{ij}$

En el segundo modelo se buscaba un único modelo de regresión para la participación de todas las provincias. Las variables explicativas que entraron en el modelo fueron las indicadoras de provincia, la participación de la población CERA y la participación postal. Posteriormente, calculando la participación

general ($\hat{y}_j = \frac{\sum_{i=1}^{52} w_i \hat{y}_{ij}}{\sum_{i=1}^{52} w_i}$, donde $\hat{y}_{ij} (\forall i, j)$) a partir del producto de cada una de

los coeficientes de provincia y las participaciones del CERA y POSTAL. A partir del error estándar residual se calculó una participación del 70,01% (67,15% - 72,88%), cuando realmente fue del 75,66%, lo que implica un aumento del 5,65% (ver Anexo III).

MODELO 2			
Participación estimada	70,01%	V(Y)	2,05E-04
Participación real	75,66%	DT(Y)	1,43%
Incremento	5,65%		
IC (95%)	67,15%	72,88%	

Tabla 9. Resumen y estimación del modelo 2

8.3. Modelo 3

Modelo 3: $\pi_{ij} \equiv y_{ij} = \beta_{0ij} + \sum_{i=1}^{i=52} \beta_{1ij} d_{ij} + \beta_2 cam_j + \beta_3 c_{ij} + \beta_4 p_{ij} + \varepsilon_{ij}$

En el tercer modelo se buscaba una única recta de regresión, igual que en el segundo, introduciendo además la variable “*posibilidad de cambio*”. Las variables explicativas que entraron en el modelo fueron las indicadoras de provincia, la participación de la población CERA y la participación postal. En este caso no se tomó el año como variable continua sino que se creó una nueva variable indicadora de la “*posibilidad de cambio*”. Finalmente la estimación se realizó como en el segundo modelo (algoritmo Stepwise).

En este caso se podía realizar la estimación, suponiendo que los electores habían ido a ejercer esperando un cambio de gobierno o por lo contrario sin creer en esta posibilidad.

Si la estimación se realiza suponiendo **NO** cambio político, se estima una participación del 66,06% (64,49% - 67,63%), realizando la estimación a partir del cambio político la estimación fue superior: 73,60% (72,03% - 75,17%).

MODELO 3 No CAMBIO			
Participación estimada	66,06%	V(Y)	6,16E-05
Participación real	75,66%	DT(Y)	1,43%
Incremento	9,60%		
IC (95%)	64,49%	67,63%	

Tabla 10. Resumen y estimación del modelo 3 (sin cambio)

MODELO 3 Sí CAMBIO			
Participación estimada	73,60%	V(Y)	6,16E-05
Participación real	75,66%	DT(Y)	0,78%
Incremento	2,06%		
IC (95%)	72,03%	75,17%	

Tabla 11. Resumen y estimación del modelo 3 (con cambio)

Nota: En los Anexos IV y V se adjuntan los modelos completos.

8.4. Diferencias entre modelos

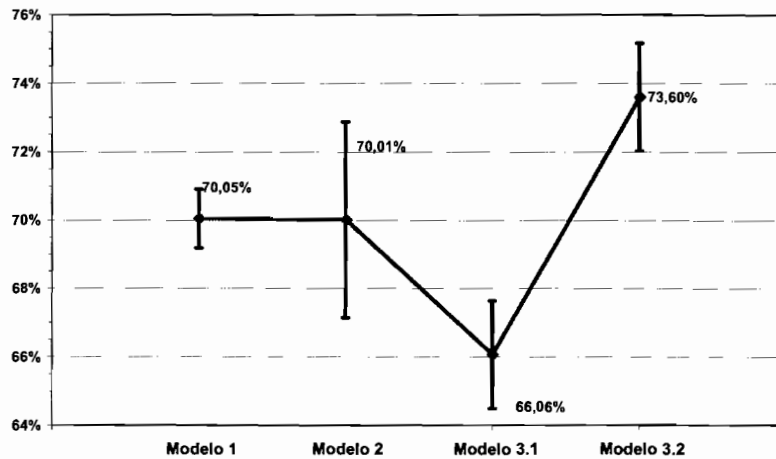
La diferencia fundamental en el tercer modelo, en referencia con el segundo, era la variable indicadora de cambio. Si nos fijamos, esta variable discriminaba entre dos situaciones que el primero no detectaba y que modelizaba estimando el efecto intermedio. El primero de ellos suponía que no existían diferencias de participación si todos los otros factores estaban fijados. Debemos observar que el tercer modelo, según el valor de la variable explicativa *cambio*, estimaba o bien una participación inferior a la del modelo 2 o bien superior.

A partir de los parámetros estimados, se calculó que la participación en un año de posible cambio era en medida de 7,55 puntos superior a la de un año sin cambio, aunque debemos pensar que existe una relación no nula entre los coeficientes de los otros parámetros estimados. Lo que se intenta transmitir en este punto es que no podemos afirmar con rotundidad que la participación aumenta justamente un 7,55% en años con previsión de cambio y los que no, sino que parte de esta información puede estar representada en la estimación de otros coeficientes (no estamos en un diseño perpendicular).

La diferencia fundamental estaba entre el primer modelo y el segundo/tercero. Como se comentó en el capítulo referente al segundo modelo, en los dos últimos se asumían que el efecto de la participación CERA en la participación general era igual en todas las provincias, sólo estimaba un único parámetro para todas las provincias. En esta ocasión seguramente la parte de variabilidad (suma de cuadrados) explicada por la participación CERA se asumía por otros parámetros.

8.5. Resultados

A continuación se adjunta gráficamente los resultados alcanzados tras la estimación de los cuatro modelos.



Los primeros modelos estiman, prácticamente la misma participación. La diferencia entre los dos modelos es esencialmente la variabilidad del intervalo de confianza, el primer modelo estima la participación de forma más precisa que el segundo modelo.

Este hecho es normal, ya que en el primer modelo se estima cada una de las provincias por separado, tomando en cada caso las 4 observaciones que se tenían que explicar a partir de dos parámetros. En el caso del segundo modelo, a partir de un número mayor de observaciones se pretendía estimar un número reducido de parámetro. Resumiendo, para cada estimación provincial a partir del primer modelo la variabilidad era menor que en la misma estimación realizada por el segundo modelo y por tanto la suma de todas ellas también lo era.

Aparte de los intervalos de confianza, se puede observar que los dos modelos apuntan hacia el mismo valor y se comprueba que los intervalos de confianza se solapan. Si tuviéramos que escoger entre uno de los dos

modelos, nos quedaríamos con el primero, puesto que apunta al mismo valor pero con mayor precisión, aunque debe estimar más coeficientes y es más complejo (estima una ecuación diferente para cada provincia).

En el tercer modelo entra en juego la variable *cambio*, que explica con mayor exactitud el modelo. Los dos modelos entre ellos son distintos, se observa claramente a partir del gráfico y por la estimación de parámetros. Para comparar los modelos con y sin cambio con el modelo 2, nos creamos la diferencia de participación para cada una de las provincias. A partir de la estimación nos calculamos el intervalo de confianza y comprobamos que no contiene el cero, por lo que deducimos que los tres modelos son diferentes entre sí. En el caso en que se compara el modelo dos con el modelo sin cambio se calcula que el primero estima una participación entre 3,44% y 5,83% superior. En el caso contrario éste estima una participación entre el 1,72% y el 4,11% inferior al modelo que incorpora la variable cambio.

8.6. Conclusiones

A la vista de los resultados, podemos concluir lo siguiente:

No existen evidencias suficientes para pensar que la participación real fue la esperada. Si asumimos que los ciudadanos hubiesen ido a votar sin esperanzas de cambio político, podemos imputar a los atentados un aumento en la votación de 9,6 puntos. En cambio si asumimos que fueron a votar asumiendo dicho cambio, esta diferencia se reduce hasta 2,06 puntos, aunque ambos estadísticamente diferentes a cero.

La diferencia entre lo que se podía predecir y lo que realmente sucedió (9,6% o 2,06%) podía imputarse a los atentados en sí o a otros motivos (guerra de Irak, problema Prestige, etc.). El segundo implicaba directamente que el votante entendiera un cambio político, pasando del modelo 3 sin cambio al modelo 3 con cambio, ya que podía provocar que el ciudadano contemplara la posibilidad de cambio. También los atentados en sí podían influir en esta percepción pero incluso con esto más gente de lo que se esperaba (2,06%) participó. Por lo que se puede concluir que los atentados tuvieron una lectura política que hizo que la participación aumentara más de lo esperado.

Por lo tanto el atentado del 11M consiguió:

- Que los votantes asumieran la posibilidad de un cambio político y por lo tanto aumentara la participación (aunque no tanto como la que realmente hubo).
- Que el 2,06% que aumentó la participación fuera un 'castigo' por la naturaleza del atentado.

8.7. Posibles ampliaciones

Si en un futuro este proyecto continuara, sería interesante seguir por una de estas dos líneas:

- Si se desea continuar con el tema de la participación, sería interesante estimar en qué situación se fue a votar. Esto pasaría básicamente por el hecho de hacer mayor hincapié en el efecto *cambio político*.
- Lo realmente interesante por hacer sería ver a partir del aumento de participación, qué partido político hubiera alcanzado la presidencia si no hubieran habido los atentados. Por esta línea ha investigado ya José García Montalvo de la Universitat Pompeu Fabra (*Voting after bombing: the electoral effect of Madrid's 11M terrorist attack*). Parece claro que aumentó la participación más de lo esperado, pero sería interesante ver qué proporción de votantes que fue a votar sin ser esperado votaba a un grupo político o a otro y por lo tanto estimar si hubiera ganado el PSOE o no sin las circunstancias ocurridas.

Capítulo 9. Anexos

En este apartado se adjuntan los resúmenes de los modelos estimados.

9.1. Anexo I. MODELO 1

Participación CERA				
Año	Total		T4,0,025	2,776
2004		26,57%	T3,0,025	3,182

Provincia	Coefficientes	Varianzas - Covarianzas		X observada	Xbi	E(Yi)	Var(Yi)	LI	LS
ÁLAVA	3,43E-04	1,64E-11	0	2004	68,71%	68,71%	0,0001	66,46%	70,96%
	0	0	0	29,43%	0,00%				
ALBACETE	3,12E-04	4,51E-10		2004	62,51%	75,85%	0,0002	71,85%	79,85%
	0,457321	-2,65E-06	0,01655	29,16%	13,34%				
ALICANTE	2,91E-04	1,39E-09		2004	58,25%	71,11%	0,0003	66,04%	76,17%
	0,620407	-9,08E-06	0,06166	20,72%	12,85%				
ALMERIA	2,92E-04	3,24E-10		2004	58,58%	68,34%	0,0001	64,63%	72,06%
	0,411507	-2,01E-06	0,01389	23,73%	9,77%				
AVILA	3,28E-04	2,69E-10		2004	65,65%	76,43%	0,0001	73,50%	79,37%

Participación CERA			
Año	Total	T4,0,025	2,776
2004	26,57%	T3,0,025	3,182

Provincia	Coefficientes	Varianzas - Covarianzas		X observada	Xbi	E(Yi)	Var(Yi)	LI	LS
	0,352053	-1,68E-06	0,01128	30,64%	10,79%				
BADAJOZ	3,34E-04	3,47E-10		2004	66,97%	74,73%	0,0001	71,09%	78,37%
	0,383508	-2,42E-06	0,01855	20,23%	7,76%				
BALEARES	3,25E-04	1,62E-10		2004	65,20%	65,20%	0,0007	58,11%	72,29%
	0	0	0	25,49%	0,00%				
BARCELONA	3,46E-04	2,20E-10		2004	69,25%	69,25%	0,0009	61,01%	77,50%
	0	0	0	25,06%	0,00%				
BURGOS	2,97E-04	5,81E-10		2004	59,56%	74,22%	0,0001	71,21%	77,23%
	0,460826	-3,82E-06	0,0258	31,82%	14,66%				
CACERAS	3,81E-04	5,80E-11		2004	76,43%	76,43%	0,0002	72,19%	80,66%
	0	0	0	17,67%	0,00%				
CADIZ	2,79E-04	5,38E-10		2004	55,82%	64,13%	0,0002	59,29%	68,96%
	0,410307	-3,73E-06	0,02892	20,25%	8,31%				
CASTELLON	3,82E-04	8,94E-11		2004	76,63%	76,63%	0,0004	71,37%	81,89%
	0	0	0	27,58%	0,00%				
CIUDAD REAL	3,27E-04	5,06E-10		2004	65,52%	72,94%	0,0002	68,90%	76,99%
	0,367451	-3,25E-06	0,02249	20,21%	7,43%				
CORDOBA	3,31E-04	3,23E-10		2004	66,39%	74,24%	0,0001	70,64%	77,84%
	0,330655	-1,90E-06	0,01217	23,73%	7,85%				
CORUÑA	2,45E-04	1,29E-10		2004	49,16%	67,32%	0,0000	65,49%	69,16%
	0,585441	-9,09E-07	0,00686	31,03%	18,17%				
CUENCA	3,92E-04	5,12E-11		2004	78,53%	78,53%	0,0002	74,55%	82,51%
	0	0	0	24,92%	0,00%				
GIRONA	3,50E-04	1,83E-10		2004	70,08%	70,08%	0,0007	62,54%	77,61%

Participación CERA			
Año	Total	T4,0,025	2,776
2004	26,57%	T3,0,025	3,182

Provincia	Coefficientes	Varianzas - Covarianzas		X observada	Xbi	E(Yi)	Var(Yi)	LI	LS
	0	0	0	25,26%	0,00%				
GRANADA	2,71E-04	8,53E-10		2004	54,39%	68,70%	0,0002	64,32%	73,09%
	0,581306	-5,44E-06	0,03623	24,62%	14,31%				
GUADALAJARA	3,18E-04	5,41E-10		2004	63,81%	74,06%	0,0001	70,59%	77,54%
	0,408624	-3,41E-06	0,02226	25,09%	10,25%				
GUIPÚZCOA	3,07E-04	2,45E-10		2004	61,61%	61,61%	0,0010	52,90%	70,33%
	0	0	0	28,47%	0,00%				
HUELVA	3,42E-04	9,38E-11		2004	68,58%	68,58%	0,0004	63,19%	73,97%
	0	0	0	16,92%	0,00%				
HUESCA	3,60E-04	8,29E-11		2004	72,15%	72,15%	0,0003	67,08%	77,21%
	0	0	0	29,62%	0,00%				
JAEN	3,50E-04	2,92E-10		2004	70,13%	76,15%	0,0001	72,67%	79,64%
	0,276121	-1,85E-06	0,01297	21,81%	6,02%				
LEON	3,54E-04	8,28E-11		2004	71,03%	71,03%	0,0003	65,97%	76,10%
	0	0	0	28,08%	0,00%				
LLEIDA	3,40E-04	1,46E-10		2004	68,11%	68,11%	0,0006	61,39%	74,84%
	0	0	0	26,99%	0,00%				
RIOJA	3,74E-04	7,41E-11		2004	75,02%	75,02%	0,0003	70,23%	79,81%
	0	0	0	34,02%	0,00%				
LUGO	3,32E-04	1,40E-10		2004	66,52%	66,52%	0,0006	59,94%	73,11%
	0	0	0	34,90%	0,00%				
MADRID	3,11E-04	5,62E-10		2004	62,27%	72,73%	0,0001	68,93%	76,53%
	0,439037	-3,99E-06	0,03024	23,83%	10,46%				

Participación CERA			
Año	Total	T4,0,025	2,776
2004	26,57%	T3,0,025	3,182

Provincia	Coefficientes	Varianzas - Covarianzas		X observada	Xbi	E(YI)	Var(YI)	LI	LS
MALAGA	2,71E-04	5,61E-10		2004	54,26%	64,56%	0,0002	60,03%	69,09%
	0,452923	-3,41E-06	0,02245	22,74%	10,30%				
MURCIA	3,08E-04	3,46E-10		2004	61,72%	74,52%	0,0001	71,52%	77,52%
	0,535375	-2,37E-06	0,01734	23,91%	12,80%				
NAVARRA	2,56E-04	7,41E-10		2004	51,38%	67,55%	0,0001	63,76%	71,33%
	0,600221	-5,20E-06	0,03831	26,93%	16,16%				
OURENSE	2,07E-04	2,31E-09		2004	41,48%	63,93%	0,0004	57,61%	70,24%
	0,856495	-1,69E-05	0,12948	26,21%	22,45%				
ASTURIAS	2,85E-04	5,40E-10		2004	57,21%	70,48%	0,0002	66,18%	74,78%
	0,435806	-3,53E-06	0,02506	30,45%	13,27%				
PALENCIA	3,18E-04	2,22E-10		2004	63,75%	74,40%	0,0001	72,11%	76,69%
	0,412022	-1,40E-06	0,00923	25,85%	10,65%				
LAS PALMAS	3,26E-04	1,40E-10		2004	65,37%	65,37%	0,0006	58,79%	71,95%
	0	0	0	34,80%	0,00%				
PONTEVEDRA	3,30E-04	1,05E-10		2004	66,22%	66,22%	0,0004	60,52%	71,91%
	0	0	0	32,88%	0,00%				
SALAMANCA	3,03E-04	5,95E-10		2004	60,66%	74,89%	0,0002	70,91%	78,87%
	0,514551	-4,42E-06	0,03503	27,66%	14,23%				
TENERIFE	3,11E-04	7,92E-11		2004	62,29%	62,29%	0,0003	57,34%	67,24%
	0	0	0	44,27%	0,00%				
CANTABRIA	3,00E-04	5,11E-10		2004	60,05%	72,09%	0,0001	68,61%	75,57%
	0,486237	-3,53E-06	0,02583	24,75%	12,03%				
SEGOVIA	3,78E-04	3,73E-11		2004	75,76%	75,76%	0,0001	72,36%	79,16%
	0	0	0	26,75%	0,00%				

Participación CERA			
Año	Total	T4,0,025	2,776
2004	26,57%	T3,0,025	3,182

Provincia	Coefficientes	Varianzas - Covarianzas		X observada	Xbi	E(YI)	Var(YI)	LI	LS
SEVILLA	2,94E-04	7,40E-10		2004	58,88%	69,62%	0,0002	64,71%	74,52%
	0,488055	-4,73E-06	0,03258	21,99%	10,73%				
SORIA	2,92E-04	1,20E-10		2004	58,59%	70,89%	0,0000	68,89%	72,89%
	0,346661	-7,14E-07	0,0045	35,50%	12,31%				
TARRAGONA	3,45E-04	1,65E-10		2004	69,11%	69,11%	0,0007	61,96%	76,25%
	0	0	0	26,09%	0,00%				
TERUEL	3,02E-04	4,43E-10		2004	60,51%	69,90%	0,0001	66,31%	73,48%
	0,370789	-2,63E-06	0,01631	25,31%	9,38%				
TOLEDO	3,51E-04	3,07E-10		2004	70,33%	77,57%	0,0001	74,13%	81,02%
	0,296551	-1,88E-06	0,01246	24,42%	7,24%				
VALENCIA	2,80E-04	1,04E-09		2004	56,08%	69,99%	0,0002	65,91%	74,07%
	0,735419	-7,30E-06	0,05324	18,91%	13,91%				
VALLADOLID	3,10E-04	2,73E-10		2004	62,21%	72,11%	0,0001	69,54%	74,69%
	0,458613	-1,76E-06	0,01186	21,60%	9,91%				
VIZCAYA	2,83E-04	4,10E-10		2004	56,73%	67,60%	0,0001	64,76%	70,43%
	0,389414	-2,83E-06	0,0205	27,90%	10,86%				
ZAMORA	2,93E-04	3,92E-10		2004	58,64%	70,94%	0,0001	67,49%	74,40%
	0,420175	-2,46E-06	0,01645	29,28%	12,30%				
ZARAGOZA	3,64E-04	8,30E-11		2004	72,92%	72,92%	0,0003	67,85%	77,99%
	0	0	0	28,36%	0,00%				
CEUTA	2,87E-04	9,32E-11		2004	57,57%	57,57%	0,0004	52,20%	62,94%
	0	0	0	19,65%	0,00%				
MELILLA	2,85E-04	1,61E-10		2004	57,08%	57,08%	0,0006	50,02%	64,14%

Participación CERA			
Año	Total	T4,0,025	2,776
2004	26,57%	T3,0,025	3,182

Provincia	Coeficientes	Varianzas - Covarianzas		X observada	Xbi	E(Yi)	Var(Yi)	Li	LS
	0	0	0	11,92%	0,00%				

Tabla 12. Estimación de los parámetros referentes al modelo 1

9.2. Anexo II. MODELO 1.

Estimada	70,05%	V(Y)	1,84E-05
Real	75,66%	DT(Y)	0,43%
Incremento	5,61%		
IC (95%)	69,19%	70,91%	

	Predicción	Censo					
	2004	2004	$\hat{Y}_i w_i$	$V(\hat{Y}_i)$	Censo ^{2*} $V(\hat{Y}_i)$	Real 04	ΔY_i
ÁLAVA	68,71%	247350	169957,5203	0,0001	4029611	77,23%	8,52%
ALBACETE	75,85%	298751	226597,4244	0,0002	14096833	79,27%	3,42%
ALICANTE	71,11%	1154642	821046,3832	0,0003	337866582	77,52%	6,41%
ALMERIA	68,34%	427300	292034,7047	0,0001	24917924	71,94%	3,59%
AVILA	76,43%	143633	109781,4768	0,0001	1757097	79,86%	3,43%
BADAJOS	74,73%	538519	402433,3210	0,0001	38009320	79,32%	4,59%
BALEARES	65,20%	687769	448455,2949	0,0007	308508158	68,84%	3,63%
BARCELONA	69,25%	4007119	2775083,4282	0,0009	14160945457	76,17%	6,91%
BURGOS	74,22%	303158	225007,3870	0,0001	8235771	78,27%	4,05%
CACERAS	76,43%	346699	264977,7203	0,0002	27983621	79,18%	2,75%

Estimada	70,05%	V(Y)	1,84E-05
Real	75,66%	DT(Y)	0,43%
Incremento	5,61%		
IC (95%)	69,19%	70,91%	

	Predicción	Censo					
	2004	2004	$\hat{Y}_i w_i$	$V(\hat{Y}_i)$	$\text{Censo}^{2+} V(\hat{Y}_i)$	Real 04	ΔY_i
CADIZ	64,13%	927135	594558,0795	0,0002	198579717	69,91%	5,78%
CASTELLON	76,63%	401529	307697,1206	0,0004	57859201	78,41%	1,78%
CIUDAD REAL	72,94%	393118	286751,1048	0,0002	24974587	78,57%	5,63%
CORDOBA	74,24%	633108	470003,9336	0,0001	51269983	78,51%	4,27%
CORUÑA	67,32%	1053549	709286,6155	0,0000	36870761	70,60%	3,28%
CUENCA	78,53%	165560	130020,8149	0,0002	5637169	81,45%	2,92%
GIRONA	70,08%	472424	331058,0778	0,0007	164294038	76,03%	5,95%
GRANADA	68,70%	694244	476947,8060	0,0002	91534469	75,72%	7,02%
GUADALAJARA	74,06%	151073	111892,1367	0,0001	2722235	80,15%	6,08%
GUIPÚZCOA	61,61%	581772	358458,7247	0,0010	333289423	72,55%	10,94%
HUELVA	68,58%	378388	259498,8173	0,0004	53923786	73,20%	4,62%
HUESCA	72,15%	176976	127679,3387	0,0003	10428722	76,62%	4,47%
JAEN	76,15%	525882	400479,0330	0,0001	33166600	80,40%	4,24%
LEON	71,03%	450239	319827,2124	0,0003	67424368	74,82%	3,78%
LLEIDA	68,11%	312352	212751,5481	0,0006	57283743	74,99%	6,88%

Estimada	70,05%	V(Y)	1,84E-05
Real	75,66%	DT(Y)	0,43%
Incremento	5,61%		
IC (95%)	69,19%	70,91%	

	Predicción	Censo					
	2004	2004	$\hat{Y}_{ij}w_i$	$V(\hat{Y}_i)$	$\text{Censo}^{2*} V(\hat{Y}_i)$	Real 04	ΔY_i
RIOJA	75,02%	234758	176108,7253	0,0003	16395974	79,46%	4,44%
LUGO	66,52%	347409	231104,7324	0,0006	67907181	72,31%	5,79%
MADRID	72,73%	4458159	3242389,4969	0,0001	2837891839	78,93%	6,20%
MALAGA	64,56%	1033244	667066,1930	0,0002	216746374	71,90%	7,34%
MURCIA	74,52%	941101	701321,7446	0,0001	78599833	77,06%	2,54%
NAVARRA	67,55%	466143	314856,7962	0,0001	30685138	76,22%	8,68%
OURENSE	63,93%	354892	226877,9426	0,0004	49607839	68,03%	4,10%
ASTURIAS	70,48%	986307	695128,2590	0,0002	177759176	71,73%	1,26%
PALENCIA	74,40%	152663	113581,2914	0,0001	1211995	79,44%	5,04%
LAS PALMAS	65,37%	736100	481220,1310	0,0006	304429535	67,72%	2,34%
PONTEVEDRA	66,22%	848580	561895,4579	0,0004	303358736	72,13%	5,91%
SALAMANCA	74,89%	312366	233930,7520	0,0002	15293534	76,62%	1,73%
S.CRUZ							
TENERIFE	62,29%	721764	449579,0670	0,0003	165632864	65,67%	3,38%
CANTABRIA	72,09%	479055	345333,6480	0,0001	27438046	77,23%	5,14%

Estimada	70,05%	V(Y)	1,84E-05
Real	75,66%	DT(Y)	0,43%
Incremento	5,61%		
IC (95%)	69,19%	70,91%	

	Predicción	Censo					
	2004	2004	$\hat{Y}_{ij} w_i$	$V(\hat{Y}_i)$	Censo ^{2*} $V(\hat{Y}_i)$	Real 04	ΔY_i
SEGOVIA	75,76%	124633	94422,7320	0,0001	2325613	81,11%	5,35%
SEVILLA	69,62%	1431672	996670,7959	0,0002	486678558	77,08%	7,46%
SORIA	70,89%	78533	55673,4591	0,0000	243230	75,22%	4,33%
TARRAGONA	69,11%	516556	356985,3513	0,0007	176705843	74,88%	5,78%
TERUEL	69,90%	116140	81176,5752	0,0001	1716251	76,36%	6,47%
TOLEDO	77,57%	450260	349282,7032	0,0001	23723633	80,84%	3,27%
VALENCIA	69,99%	1884448	1318844,2613	0,0002	583991585	77,68%	7,70%
VALLADOLID	72,11%	431757	311360,7402	0,0001	12162363	81,16%	9,05%
VIZCAYA	67,60%	974689	658859,4375	0,0001	75423722	75,83%	8,23%
ZAMORA	70,94%	182355	129370,9772	0,0001	3928660	74,44%	3,50%
ZARAGOZA	72,92%	726777	529943,5694	0,0003	176151005	77,25%	4,33%
CEUTA	57,57%	56743	32665,1883	0,0004	1205136	63,45%	5,89%
MELILLA	57,08%	49116	28035,9700	0,0006	1559794	55,84%	-1,24%
TOTAL		34568509			2,20E+10		

Tabla 13. Estimación de los parámetros referentes al modelo 2.1

9.3. Anexo III. Modelo 2

Estimada	70,01%	V(Y)	2,05E-04
Real	75,66%	DT(Y)	1,43%
Incremento	5,65%		
IC (95%)	67,15%	72,88%	

	Predicción	Censo	$\hat{Y}_{ij} w_i$
	2004	2004	
ÁLAVA	70,82%	247350	175165,8292
ALBACETE	75,90%	298751	226759,7382
ALICANTE	72,57%	1154642	837881,4311
ALMERIA	66,22%	427300	282972,5759
AVILA	76,47%	143633	109832,5790
BADAJOS	74,48%	538519	401067,8621
BALEARES	68,90%	687769	473846,4252
BARCELONA	67,04%	4007119	2686468,9308
BURGOS	74,10%	303158	224643,1604
CACERAS	72,25%	346699	250495,8960
CADIZ	63,34%	927135	587203,3101

Estimada	70,01%	V(Y)	2,05E-04
Real	75,66%	DT(Y)	1,43%
Incremento	5,65%		
IC (95%)	67,15%	72,88%	

	Predicción	Censo	$\hat{Y}_{ij} w_i$
	2004	2004	
CASTELLON	74,11%	401529	297556,0647
CIUDAD REAL	72,15%	393118	283628,8699
CORDOBA	73,05%	633108	462494,0665
CORUÑA	66,95%	1053549	705374,4566
CUENCA	78,23%	165560	129510,4042
GIRONA	67,40%	472424	318431,9946
GRANADA	69,69%	694244	483798,4988
GUADALAJARA	73,86%	151073	111590,0173
GUIPÚZCOA	58,71%	581772	341563,5621
HUELVA	66,66%	378388	252215,5088
HUESCA	71,75%	176976	126977,6017
JAEN	74,74%	525882	393039,4043
LEON	70,97%	450239	319535,0401
LLEIDA	67,39%	312352	210504,7371
RIOJA	73,86%	234758	173392,6298

Estimada	70,01%	V(Y)	2,05E-04
Real	75,66%	DT(Y)	1,43%
Incremento	5,65%		
IC (95%)	67,15%	72,88%	

	Predicción	Censo	$\hat{Y}_y w_i$
	2004	2004	
LUGO	70,23%	347409	243970,1330
MADRID	73,09%	4458159	3258594,6479
MALAGA	63,83%	1033244	659507,8612
MURCIA	73,61%	941101	692736,0677
NAVARRA	66,50%	466143	309990,9114
OURENSE	64,94%	354892	230450,0955
ASTURIAS	69,04%	986307	680913,3610
PALENCIA	75,27%	152663	114913,6093
LAS PALMAS	69,77%	736100	513551,0417
PONTEVEDRA	67,63%	848580	573853,7265
SALAMANCA	73,90%	312366	230850,4600
S.CRUIZ			
TENERIFE	70,60%	721764	509599,4289
CANTABRIA	71,93%	479055	344603,6758
SEGOVIA	75,15%	124633	93666,5176

Estimada	70,01%	V(Y)	2,05E-04
Real	75,66%	DT(Y)	1,43%
Incremento	5,65%		
IC (95%)	67,15%	72,88%	

	Predicción	Censo	$\hat{Y}_j w_i$
	2004	2004	
SEVILLA	69,95%	1431672	1001514,6806
SORIA	71,30%	78533	55990,2161
TARRAGONA	65,33%	516556	337486,6006
TERUEL	72,28%	116140	83946,8277
TOLEDO	76,83%	450260	345947,1424
VALENCIA	72,81%	1884448	1372130,6625
VALLADOLID	73,12%	431757	315715,3770
VIZCAYA	68,13%	974689	664059,2551
ZAMORA	72,12%	182355	131515,0844
ZARAGOZA	70,73%	726777	514049,2922
CEUTA	54,24%	56743	30776,0400
MELILLA	54,12%	49116	26582,8594
TOTAL		34568509	

Tabla 14. Estimación de los parámetros referentes al modelo 3 (sin cambio)

9.4. Anexo IV. Modelo 3 (con cambio)

Estimada	73,60%	V(Y)	6,16E-05
Real	75,66%	DT(Y)	0,78%
Incremento	2,06%		
IC (95%)	72,03%	75,17%	

	Predicción	Censo	$\hat{Y}_{ij} w_i$
	2004	2004	
ÁLAVA	69,60%	247350	172157,7292
ALBACETE	79,89%	298751	238673,1463
ALICANTE	78,52%	1154642	906597,4712
ALMERIA	73,05%	427300	312150,1456
AVILA	77,56%	143633	111405,3626
BADAJOS	79,12%	538519	426071,1739
BALEARES	66,45%	687769	456996,0076
BARCELONA	77,37%	4007119	3100248,3587
BURGOS	68,80%	303158	208566,4909
CACERAS	77,84%	346699	269869,1918
CADIZ	68,56%	927135	635639,8753

Estimada	73,60%	V(Y)	6,16E-05
Real	75,66%	DT(Y)	0,78%
Incremento	2,06%		
IC (95%)	72,03%	75,17%	

	Predicción	Censo	$\hat{Y}_{ij} w_i$
	2004	2004	
CASTELLON	78,50%	401529	315211,1546
CIUDAD REAL	78,50%	393118	308583,9886
CORDOBA	78,98%	633108	500051,5457
CORUÑA	68,21%	1053549	718594,8463
CUENCA	79,74%	165560	132015,0521
GIRONA	75,12%	472424	354892,3198
GRANADA	70,03%	694244	486204,2568
GUADALAJARA	77,81%	151073	117551,7331
GUIPÚZCOA	62,53%	581772	363773,0366
HUELVA	74,22%	378388	280841,8159
HUESCA	69,28%	176976	122601,2038
JAEN	79,75%	525882	419405,8516
LEON	72,84%	450239	327973,6993
LLEIDA	77,76%	312352	242884,2575
RIOJA	69,33%	234758	162761,3464

Estimada	73,60%	V(Y)	6,16E-05
Real	75,66%	DT(Y)	0,78%
Incremento	2,06%		
IC (95%)	72,03%	75,17%	

	Predicción	Censo	$\hat{Y}_{ij} w_i$
	2004	2004	
LUGO	76,88%	347409	267094,6806
MADRID	70,09%	4458159	3124747,5287
MALAGA	79,11%	1033244	817421,1699
MURCIA	71,01%	941101	668235,5333
NAVARRA	66,93%	466143	312008,8121
OURENSE	70,26%	354892	249336,2813
ASTURIAS	72,19%	986307	711982,6383
PALENCIA	77,10%	152663	117699,4024
LAS PALMAS	68,03%	736100	500790,5613
PONTEVEDRA	69,80%	848580	592295,2936
SALAMANCA	74,98%	312366	234196,7969
S.CRUZ			
TENERIFE	65,99%	721764	476306,1177
CANTABRIA	75,14%	479055	359985,4113
SEGOVIA	77,40%	124633	96466,2433

Estimada	73,60%	V(Y)	6,16E-05
Real	75,66%	DT(Y)	0,78%
Incremento	2,06%		
IC (95%)	72,03%	75,17%	

	Predicción	Censo	$\hat{Y}_{ij} w_i$
	2004	2004	
SEVILLA	75,37%	1431672	1079000,9586
SORIA	70,58%	78533	55432,2340
TARRAGONA	76,16%	516556	393421,7477
TERUEL	68,21%	116140	79215,2752
TOLEDO	81,52%	450260	367060,1116
VALENCIA	77,90%	1884448	1467906,3904
VALLADOLID	77,63%	431757	335157,3162
VIZCAYA	73,16%	974689	713117,8999
ZAMORA	69,06%	182355	125939,2020
ZARAGOZA	75,40%	726777	547994,7670
CEUTA	56,26%	56743	31921,0917
MELILLA	53,98%	49116	26512,5533
TOTAL		34568509	

Tabla 15. Estimación de los parámetros referentes al modelo 3 (con cambio)

9.5. Anexo V. Modelo 3 (sin cambio)

Estimada	66,06%	V(Y)	6,16E-05
Real	75,66%	DT(Y)	0,78%
Incremento	9,60%		
IC (95%)	64,49%	67,63%	

	Predicción	Censo	$\hat{Y}_{ij} w_i$
	2004	2004	
ALAVA	62,05%	247350	153493
ALBACETE	72,34%	298751	216129
ALICANTE	70,97%	1154642	819468
ALMERIA	65,51%	427300	279906
AVILA	70,02%	143633	100567
BADAJOS	71,57%	538519	385435
BALEARES	58,90%	687769	405097
BARCELONA	69,82%	4007119	2797871
BURGOS	61,25%	303158	185690
CACERES	70,29%	346699	243707
CADIZ	61,01%	927135	565678
CASTELLON	70,96%	401529	284912
CIUDAD REAL	70,95%	393118	278919
CORDOBA	71,44%	633108	452277
CORUÑA	60,66%	1053549	639094
CUENCA	72,19%	165560	119522
GIRONA	67,58%	472424	319243
GRANADA	62,49%	694244	433817

Estimada	66,06%	V(Y)	6,16E-05
Real	75,66%	DT(Y)	0,78%
Incremento	9,60%		
IC (95%)	64,49%	67,63%	

	Predicción	Censo	$\hat{Y}_{ij} w_i$
	2004	2004	
GUADALAJARA	70,27%	151073	106152
GUIPUZCOA	54,98%	581772	319873
HUELVA	66,67%	378388	252289
HUESCA	61,73%	176976	109247
JAEN	72,21%	525882	379723
LEON	65,30%	450239	293999
LLEIDA	70,21%	312352	219314
RIOJA	61,79%	234758	145047
LUGO	69,34%	347409	240879
MADRID	62,54%	4458159	2788335
MALAGA	71,57%	1033244	739453
MURCIA	63,46%	941101	597220
NAVARRA	59,39%	466143	276834
OURENSE	62,71%	354892	222556
ASTURIAS	64,64%	986307	637556
PALENCIA	69,55%	152663	106179
LAS PALMAS	60,49%	736100	445244
PONTEVEDRA	62,25%	848580	528261
SALAMANCA	67,43%	312366	210626
S.CRUZ TENERIFE	58,45%	721764	421842
CANTABRIA	67,60%	479055	323836
SEGOVIA	69,85%	124633	87061
SEVILLA	67,82%	1431672	970967

Estimada	66,06%	V(Y)	6,16E-05
Real	75,66%	DT(Y)	0,78%
Incremento	9,60%		
IC (95%)	64,49%	67,63%	

	Predicción		Censo	$\hat{Y}_j w_i$
	2004	2004		
SORIA	63,04%	78533	49506	49506
TARRAGONA	68,62%	516556	354442	354442
TERUEL	60,66%	116140	70451	70451
TOLEDO	73,98%	450260	333083	333083
VALENCIA	70,35%	1884448	1325706	1325706
VALLADOLID	70,08%	431757	302577	302577
VIZCAYA	65,62%	974689	639568	639568
ZAMORA	61,52%	182355	112179	112179
ZARAGOZA	67,85%	726777	493152	493152
CEUTA	48,71%	56743	27639	27639
MELILLA	46,43%	49116	22806	22806
TOTAL ESTATAL		34568509		

Tabla 16. *Estimación de los parámetros referentes al modelo 3 (sin cambio)*

9.6. Anexo VI. Errores de estimación

9.6.1. Modelo 1

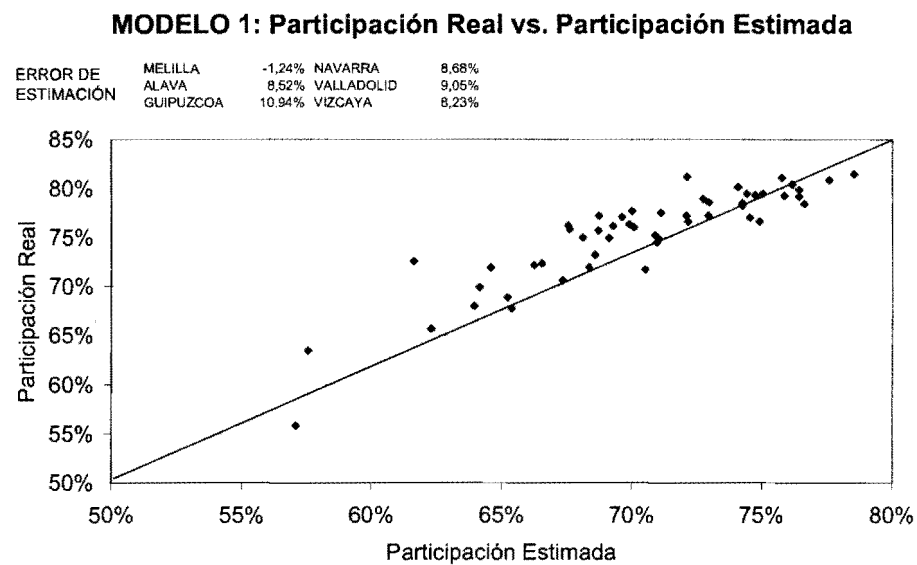


Figura 14. Error de estimación del modelo 1

9.6.2. Modelo 2.1

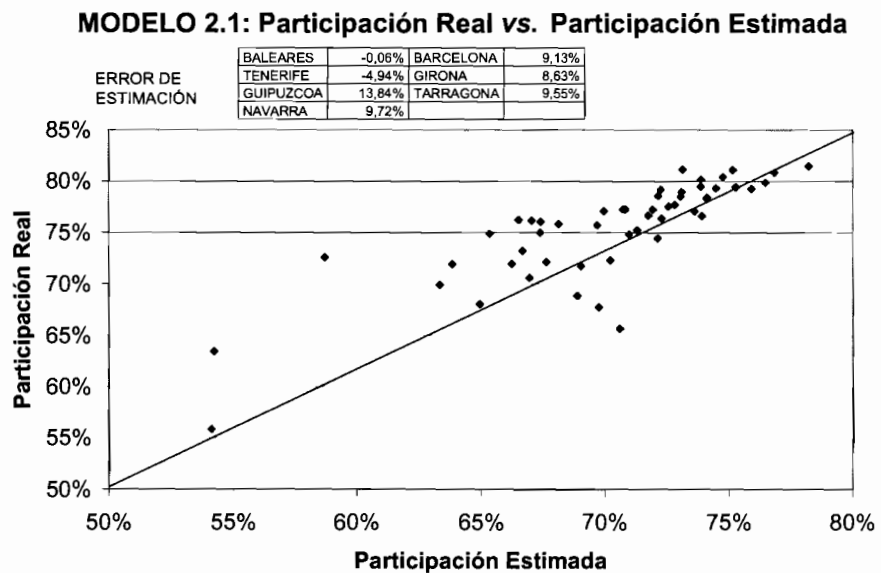


Figura 15. Error de estimación del modelo 2.1

9.6.3. Modelo 3 con cambio

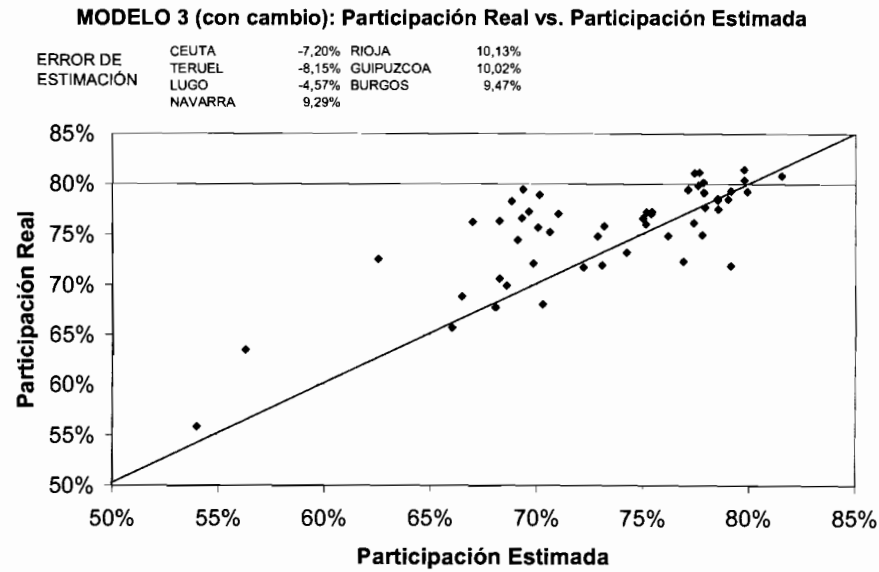


Figura 16. Error de estimación del modelo 3 con cambio

9.6.4. Modelo 3 sin cambio

MODELO 3v (sin cambio): Participación Real vs. Participación Estimada

ERROR DE ESTIMACIÓN	CEUTA	14,74%	RIOJA	17,67%
	TERUEL	15,70%	GUIPUZCOA	17,57%
	LUGO	2,98%	BURGOS	17,02%
	NAVARRA	16,84%		

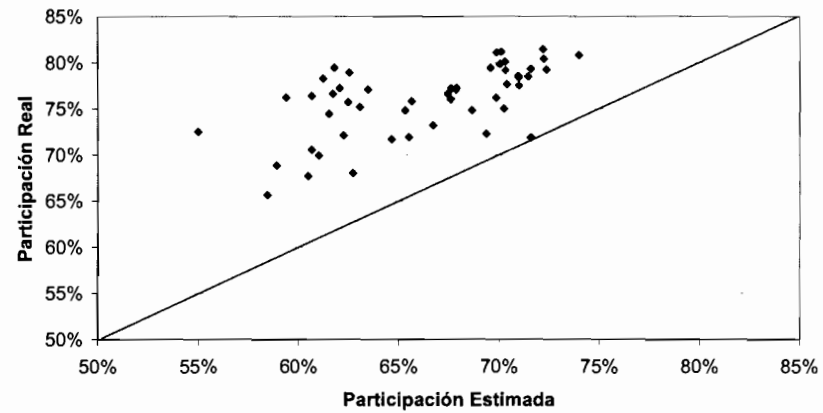


Figura 17. Error de estimación del modelo 3 sin cambio

Capítulo 10. Bibliografía

Documentación consultada:

- Josep M. Colomer. (2004). The general election in Spain, March 2004. An Introduction to *R.W.N. Benables, D.M.Smith* and the R Development Core Team.
- Lúdia Montero. (2003). Models lineals generalitzats. Departament d'Estadística i Investigació Operativa. Facultat de Matemàtiques i Estadística.
- José García Montalvo. (2004). Voting after the bombing: the electoral effect of Madrid's 11M terrorist attack. IVIE and Universitat Pompeu Fabra.

Páginas Web consultadas:

- Ministerio de Interior de España. www.mir.es
- Instituto Nacional de Estadística. www.ine.es