Sondeos electorales: ¿Qué fue más sorprendente, la mayoría absoluta del PP en 2000, o la victoria del PSOE en 2004?

Pedro Delicado (Universitat Politècnica de Catalunya)

(en colaboración con Frederic Udina, Universitat Pompeu Fabra)

Página Web:
`http://www-eio.upc.es/~delicado/`

Esquema:

1. Introducción.
2. Representación gráfica de sondeos electorales.
3. Parlament'99 y Parlament'03.
4. Congreso'00 y Congreso'04.

   ... y la respuesta es ...

5. Otros trabajos en curso sobre la ifluencia del 11-M en el 14-M.
6. Sesgo en la estimación de un parlamento: la ley d'Hondt.
7. Una curiosidad: *Hi ha sondetjos amb cara i ulls?*
8. Conclusiones.

1. Introducción

Durante las semanas previas a unas elecciones aparecen multitud de sondeos electorales en los medios de comunicación.

Los días siguientes, aparece la discusión de por qué los sondeos se equivocaron.

Ejemplos claros:
elecciones al Parlament'03, al Parlament'99,
al Congreso'00 y al Congreso'04.

En el caso de las elecciones del 14-M de 2004 parece haber una justificación para que ningún sondeo hubiese dado al PSOE como ganador:

La influencia del 11-M sobre los resultados no pudo ser detectada por los sondeos.

En el año 2000 no ocurrió ningún 11-M y, aún así, ningún sondeo predijo la mayoría absoluta del PP.

Aquel año ETA aseseinó a Fernando Buesa (miembro de la ejecutiva del PSE-EE) y a su escolta Jorge Díez en plena campaña electoral (22 de febrero de 2000).

La repercusión de este asesinato en los resultados electorales del 2000 se percibe menos importante que la del 11-M en los del 2004.

¿Qué fue más sorprendente, la mayoría absoluta del PP en 2000, o la victoria del PSOE en 2004?

Diremos que un resultado electoral es tanto más sorprendente cuanto más se aleje de los sondeos electorales previos.

Ejemplo de sondeo electoral.

Elecciones: Elecciones al Parlament de Catalunya, 16 de noviembre de 2003.

Ejemplo de sondeo: El publicado por el **CIS** una semana antes (8-11-2003).

Tamaño muestral: 3600 entrevistas a personas con derecho a voto.

Distribución territorial de la muestra:
Barcelona 1830, Girona 600, Lleida 528, Tarragona 642.

Trabajo de campo: 13 al 26 de octubre de 2003.

Preguntas: **¿Votará? ¿Qué votará?** ¿Qué votó la última vez? Características sociales y demográficas.

Selección: Muestreo polietápico, estratificado por conglomerados, con selección aleatòria proporcional.

Tipo de entrevista: Entrevista personal en los domicilios.

Elecciones al Parlament'99. Resultados y sondeos

	CiU	PSC	PP	IC-V	ERC
Resultado	**56**	**52**	**12**	**3**	**12**
CIS	57..59	45..47	14..15	4..5	11..13
El Mundo	58..62	48..50	12..14	3	10
La Vanguardia	56..58	51..52	13	4	9..10
El País	58..60	48..50	13..14	3	10
El Periódico	57..63	40..46	14..15	5..6	12..15
ABC	53..55	45..47	15..16	7..8	12..13
Diario 16	53..56	48..50	14..16	6..7	10..12

CONCLUSIONES:

Todos los sondeos predecían mejores resultados de CiU y peores del PSC.

Elecciones al Parlament'03. Resultados y sondeos

	CiU	PSC	ERC	PPC	ICV
Resultados	**46**	**42**	**23**	**15**	**9**
ABC	49..49	48..49	18..19	13..13	6..6
CIS	48..48	47..47	18..18	14..14	8..8
El Periódico	45..47	50..51	19..20	12..13	7..8
La Vanguardia	48..49	49..50	18..19	12..12	6..7
El País	47..49	48..50	17..19	13..13	7..7
La Razón	48..49	48..51	17..18	13..15	5..6
El Mundo	46..49	46..49	18..20	15..16	5..6

CONCLUSIONES:

Todos los sondeos predecían mejores resultados del PSC y peores de CiU.

Todos predecían peores resultados de los otros tres partidos.

Tras los resultados poco satisfactorios de los sondeos, surge la siguiente pregunta:

¿Pueden ser achacadas al azar las diferencias entre las predicciones de los sondeos y los resultados reales?

Nos propusimos **comparar los parlamentos pronosticados por los sondeos** con los que se obtendrían mediante un **muestreo aleatorio ideal**.

¿Cómo comparar **sondeos reales** y **sondeos simulados por ordenador**?.

Buscamos una forma **gráfica** de comparar estos sondeos.

Si sólo hubiese dos partidos ... (I)

Ejemplo: PP y PSOE.

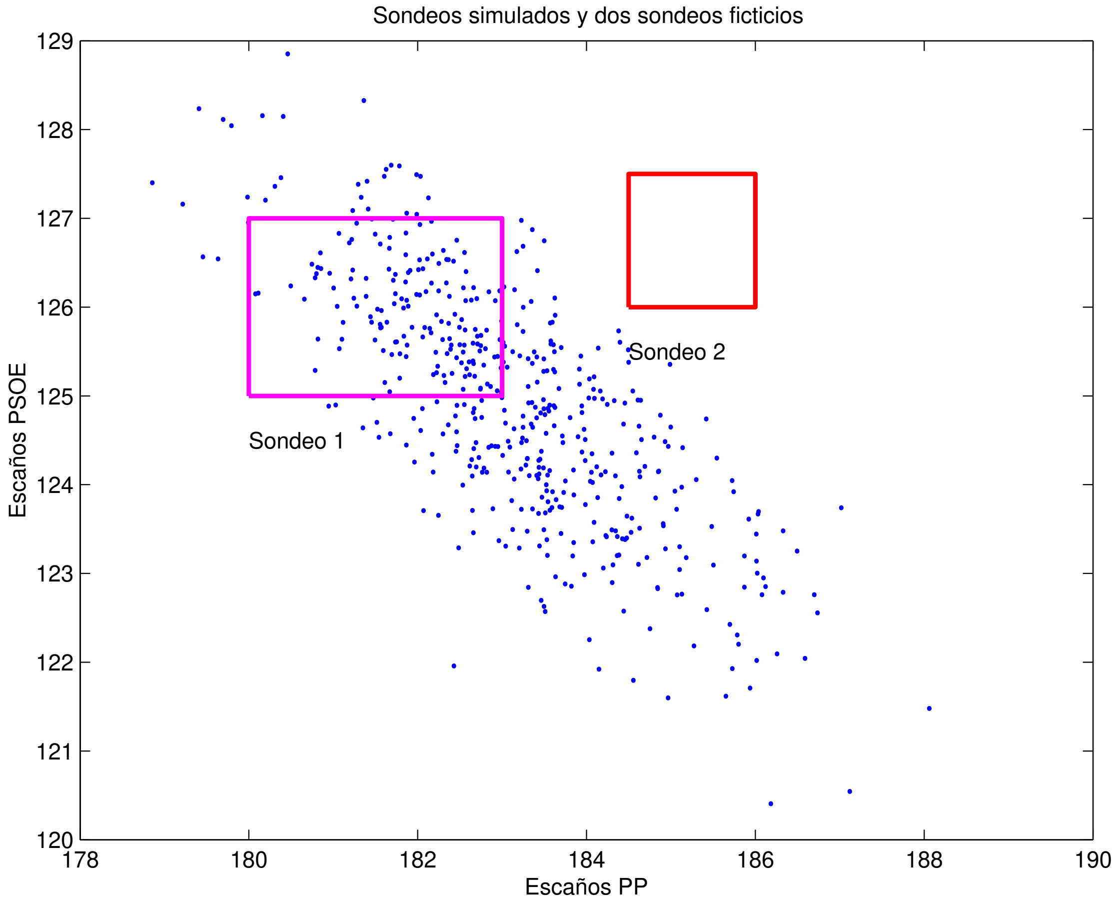
- Resultado de cada sondeo simulado:
  un par ordenado (PP,PSOE), **un punto en el plano coordenado**.
- Simulamos **500** sondeos: 500 puntos.
- Representamos en unos ejes de coordenadas los 500 puntos:

  Eje X: PP, Eje Y: PSOE

- También representamos los sondeos electorales publicados.
- **PREGUNTA:** ¿Se parecen los sondeos reales a los producidos por el ordenador?

Si sólo hubiese dos partidos ... (II)

Se obtendría una representación como ésta:



Pero hay más de dos: La mejor foto de los sondeos.

Hay muchos partidos políticos: **5** en el Parlament, más de **10** en el Congreso.

Cada sondeo simulado es un vector de **dimensión 10** (o mayor):

(PP, PSOE, CiU, IU, PNV,..., CHA).

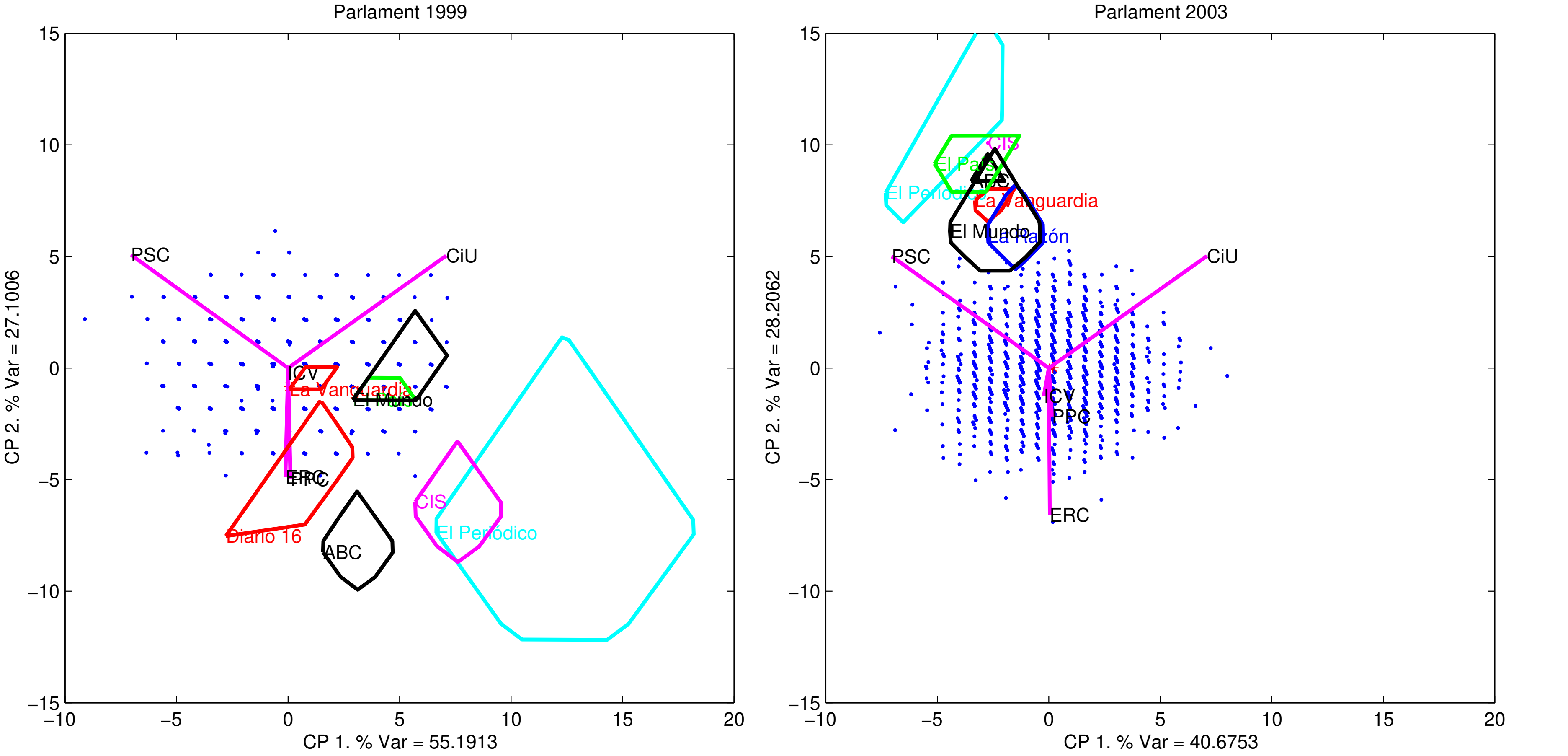
¿Cómo **representar** los sondeos en un **papel** o una **pantalla de ordenador?**

La mejor foto de los sondeos.

1. Disponemos de los **resultados reales** de las elecciones.
2. Se simula una muestra de intención de voto en cada provincia y se calcula la correspondiente asignación de escaños.
3. La suma de los escaños obtenidos por provincias da lugar a un **parlamento simulado**.
4. Se repite el paso anterior ***B* veces**.
5. Se toma **la mejor foto** del conjunto de **B** datos multidimensionales:

   COMPONENTES PRINCIPALES.
6. Se proyecta el **parlamento real** en el plano de las dos primeras componentes principales.
7. Se proyectan también las **horquillas de escaños publicadas**.

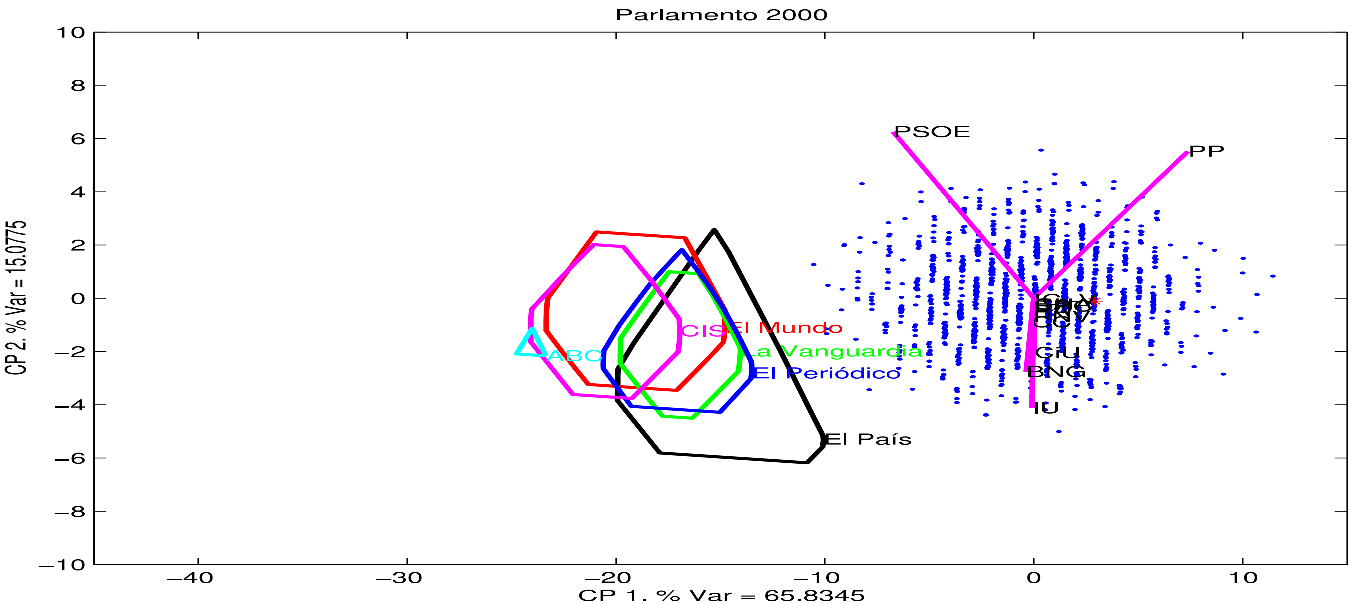
3. Parlament'99 y Parlament'03



Congreso, 2000

Part.	Res.	El País	El Mundo	La Vang.	El Perd.	CIS	ABC
PP	**183**	165..171	164..170	166..170	165..170	163..168	158..164
PSOE	**125**	131..139	137..143	136..140	135..140	138..143	138..144
CiU	**15**	14..15	16	14..15	13..14	14..15	9..12
IU	**8**	9..13	9..11	12..14	9..12	9..11	15..16
PNV	**7**	6..7	6..7	6	7..8	7	5..7
CC	**4**	4	4	4	4..5	3..4	4
BNG	**3**	4..5	3..5	3	4..5	4..5	4
PA	**1**	0	0..1	0..1	2	0	0
ERC	**1**	1..2	1	1..2	1	1	0
IC-V	**1**	1	0	0..1	0	1..2	0
EA	**1**	1	1	1	0	1	0
CHA	**1**	1	0	0	0	1	0

CONCLUSIONES: Ningún sondeo predijo la mayoría absoluta del PP.

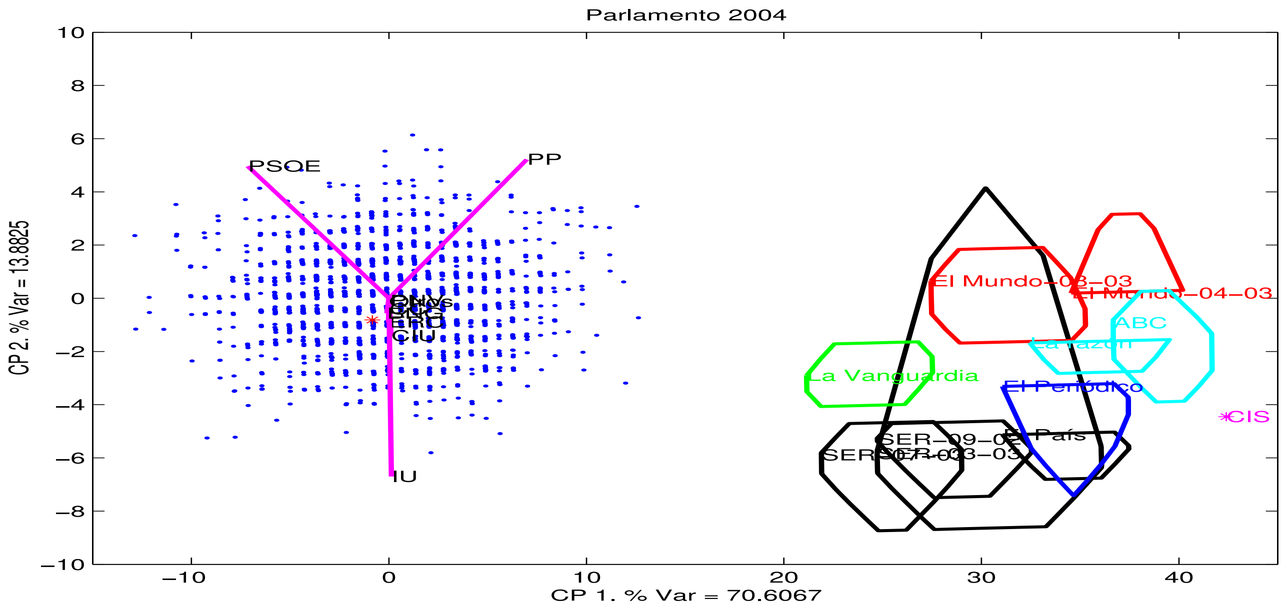


Parlamento 2000
CP 2. % Var = 15.0775
CP 1. % Var = 65.8345
PSOE
PP
CiU
BNG
IU
CIS
El Mundo
La Vanguardia
El Periódico
El País
ABC
10
8
6
4
2
0
−2
−4
−6
−8
−10
−40
−30
−20
−10
0
10

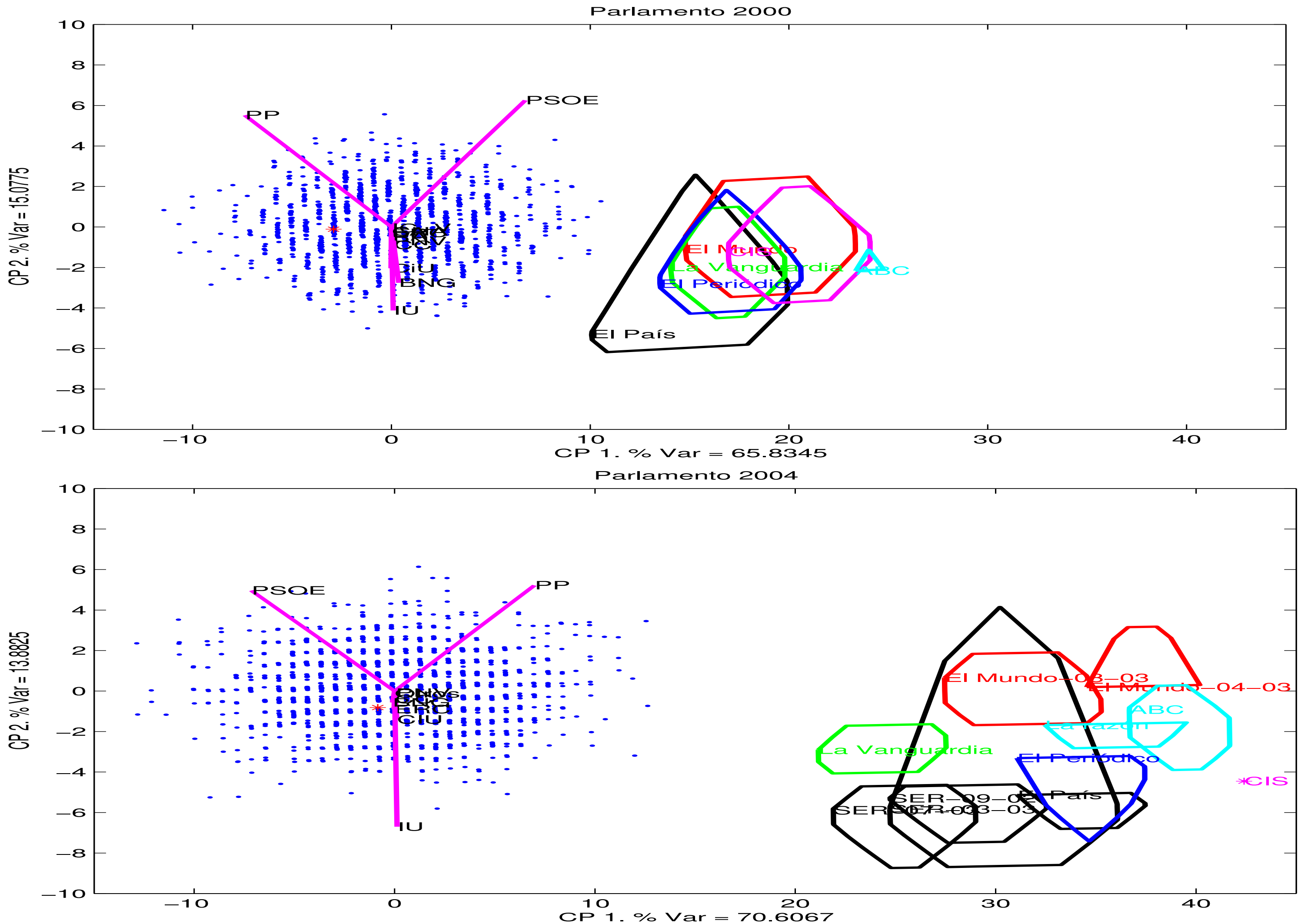
Congreso, 2004

Medio.Comunic.	Muestra	PSOE	PP	CiU	ERC	PNV	IU/ICV
SER 29 02	3000	138..144	164..169	10..11	6	7..8	12
SER 03 03	3000	135..143	163..171	10..11	6..8	7..8	11..12
SER 07 03	3000	140..145	161..166	10..11	7..8	7..8	11..12
El Mundo 04 03	12500	134..139	172..177	11	5	7	5..7
El Mundo 08 03	1000	138..144	168..173	11	5..6	7	6..8
La razón	2404	135..141	171..176	11..12	5	7..8	9..9
El País	4000	134..141	168..172	10..11	7..7	7..8	11
CIS	24109	131	176	12	6	7	10
ABC	12760	133..137	174..177	10..11	6..7	7	8..10
La Vanguardia	2200	143..147	162..167	10	5..6	7	9..10
El Periódico	2071	135..140	169..173	10..11	6..7	7	10..12
Res 2004		164	148	10	8	7	5

CONCLUSIONES: Ningún sondeo predijo la victoria del PSOE.

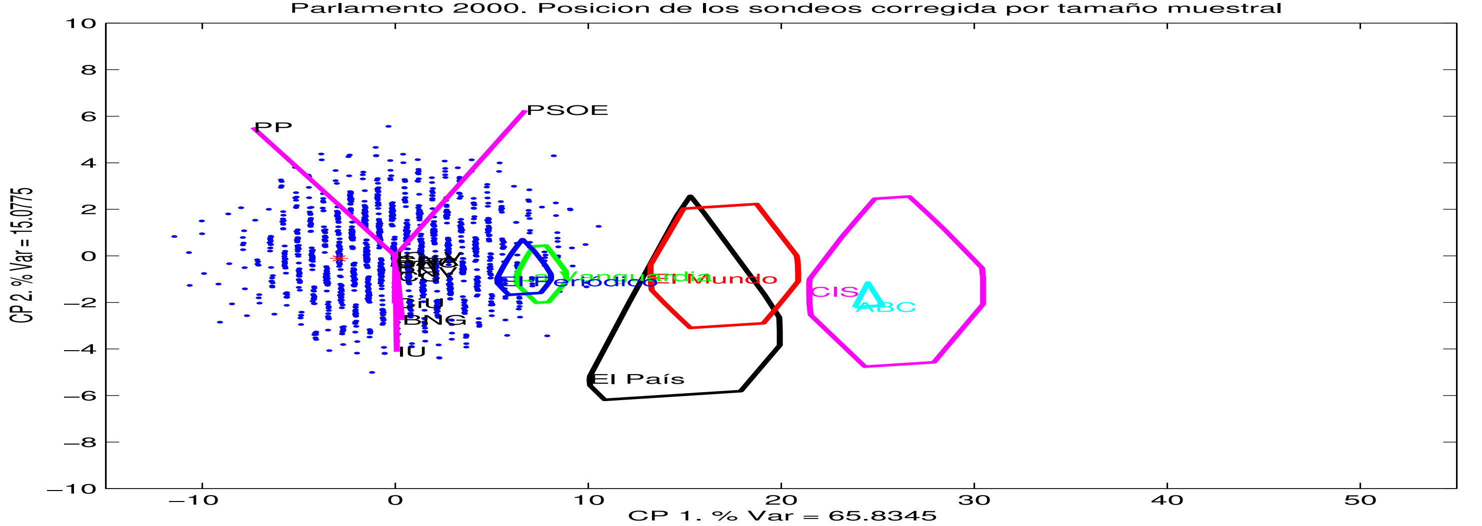


Parlamento 2004
CP 2. % Var = 13.8825
CP 1. % Var = 70.6067
PSOE
PP
IU
CIU
La Vanguardia
El Mundo-04-03
ABC
El Periódico
CIS
El País

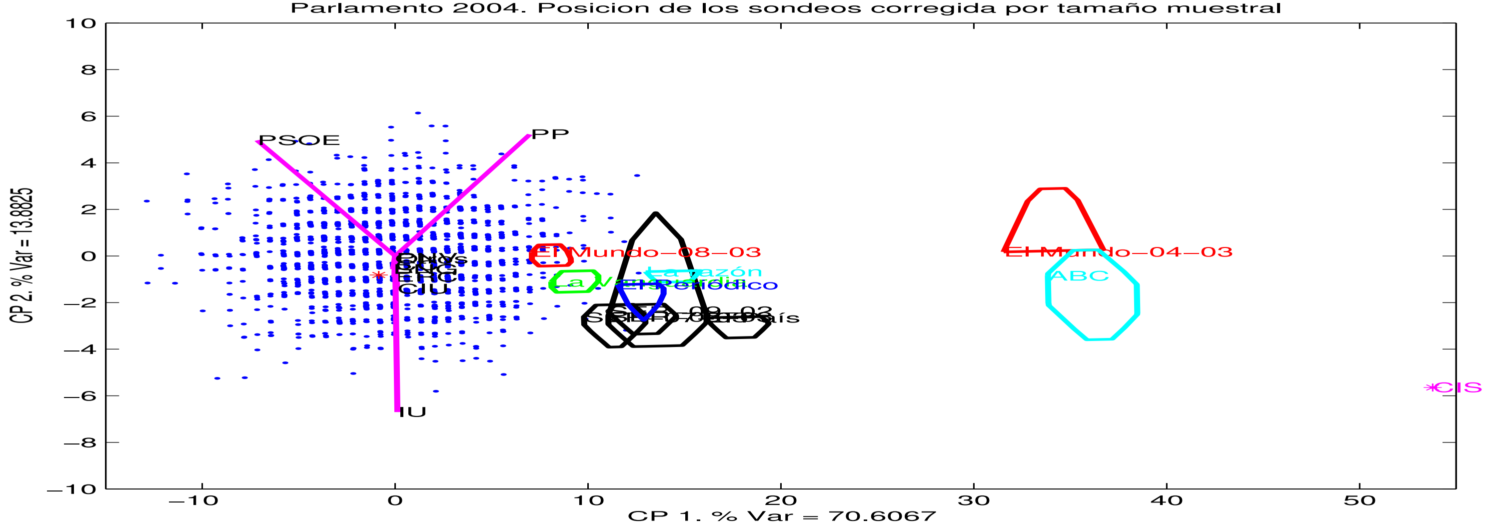


Parlamento 2000
CP 2. % Var = 15.0775
CP 1. % Var = 65.8345
PP
PSOE
CiU
BNG
IU
El Mundo
La Vanguardia
El Periodico
ABC
El País
Parlamento 2004
CP 2. % Var = 13.8825
CP 1. % Var = 70.6067
PSOE
PP
CiU
IU
El Mundo−03−03
ABC
La Vanguardia
El Periodico
SER−09−03
*CIS

Parlamento 2000. Posicion de los sondeos corregida por tamaño muestral



Parlamento 2004. Posicion de los sondeos corregida por tamaño muestral



Distancias de Mahalanobis de los sondeos al sondeo medio simulado.

2000		**2004**	
MedCom	**d.Mah.**	**MedCom**	**d.Mah.**
El País	5,11	SER-09-02	3,80
El Mundo	5,77	SER-03-03	3,71
La Vanguardia	3,77	SER-07-03	3,34
EL Periódico	3,66	El Mundo-04-03	9,51
CIS	9,11	El Mundo-08-03	2,19
ABC	14,12	La Razón	4,33
		El País	5,02
		CIS	15,60
		ABC	9,30
		La Vanguardia	2,79
		El Periódico	3,39

Variable	Año	N	Mean	Median	TrMean	StDev
dMah	2000	6	6,92	5,44	6,92	4,04
	2004	11	5,73	3,80	5,02	4,09

Variable	Año	SE Mean	Minimum	Maximum	Q1	Q3
dMah	2000	1,65	3,66	14,12	3,74	10,36
	2004	1,23	2,19	15,60	3,34	9,30

H_0: La distribución de las distancias de los sondeos al centro de la nube es la misma en 2000 y en 2004.

H_1: Las distribuciones son diferentes (o la de 2004 domina a la otra).

Contraste no paramétrico de Kolmogorov-Smirnov para dos muestras: **p-valor=0.583**.

Logaritmos de las distancias: normalidad SÍ, igualdad de varianzas SÍ, contraste de igualdad de medias con **p-valor=0.576**.

Conclusión.

¿Qué fue más sorprendente, la mayoría absoluta del PP en 2000, o la victoria del PSOE en 2004?

No hay evidencia estadísticamente significativa para afirmar que la victoria del PSOE en 2004 fuese más sorprendente que la mayoría absoluta de PP en 2000.

5. Otros trabajos en curso sobre la ifluencia del 11-M en el 14-M.

Los únicos resultados que no se vieron afectados por el 11-M son los correspondientes al **Censo de Residentes Ausentes**.

A partir de los resultados del **Censo de Residentes Ausentes**, ¿es posible predecir los resultados generales que se hubiesen producido si no hubiese habido el 11-M?

- ¿Participación?
- ¿Porcentaje de votos de un partido?
- ¿Incremento de ese porcentaje respecto a las últimas elecciones?
- ¿A nivel provincial, autonómico, nacional?

En estos momentos un estudiante de la LCTE de la UPC (Josep Ramon Marsal) está realizando su Proyecto Fin de Carrera sobre este tema.

Reparto de M escaños entre K partidos:

δ : umbral mínimo para optar a un escaño. *(Entre un 3 % y un 5 %)*

$\{d_1, d_2, \ldots, d_M\}$ lista creciente de *denominadores*

$(f_1, f_2, \ldots, f_K)$ proporciones de votos obtenidos.

$$0 \leq f_i \leq 1, (i = 1..K), \qquad \sum_{i=1}^{K} f_i = 1.$$

Sean $q_{i,j}$, $(i = 1..K,\ j = 1..M)$ los *cocientes*:

$$\text{Si } f_i \geq \delta, \quad q_{i,j} = \tfrac{f_i}{d_j}, \qquad \text{para } j = 1..M \tag{1}$$

$$\text{Si } f_i < \delta, \quad q_{i,j} = 0 \qquad \text{para } j = 1..M \tag{2}$$

Se asigna un escaño a cada uno de los M cocientes mayores.

Listas de denominadores d_j:

Ley d'Hondt: $1, 2, 3, \ldots, d_j = j, \ldots$

Regla de Saint-Lagüe: $1, 3, 5, \ldots, d_j = 2j - 1, \ldots$

Regla modificada de Saint-Lagüe: 1,4, $3, 5, 7 \ldots$

Fichero `gauss.upf.es/papers/epolls/hondt.xls`

Elecciones al parlamento vasco 13-05-01. Resultados en Vizcaya.

5 partidos y hasta 41 escaños

INFO	Partido	Votos s/1000	Escaños
1	PNV	441	*12*
2	PP	237	*6*
3	PSE	183	*4*
4	EH	82	*2*
5	IU	57	*1*
	5	1000	*25*

umbral	3%
num. escaños	25
num. Cocientes	125

último: *36,75*

Cocientes

	1	2	3	4	5	6	7	8	9	10	11	12
1	441,0	220,5	147,0	110,3	88,2	73,5	63,0	55,1	49,0	44,1	40,1	36,8
2	237,0	118,5	79,0	59,2	47,4	39,5	33,9	29,6	26,3	23,7	21,5	19,7
3	183,0	91,5	61,0	45,7	36,6	30,5	26,1	22,9	20,3	18,3	16,6	15,2
4	82,0	41,0	27,3	20,5	16,4	13,7	11,7	10,2	9,1	8,2	7,5	6,8
5	57,0	28,5	19,0	14,2	11,4	9,5	8,1	7,1	6,3	5,7	5,2	4,7

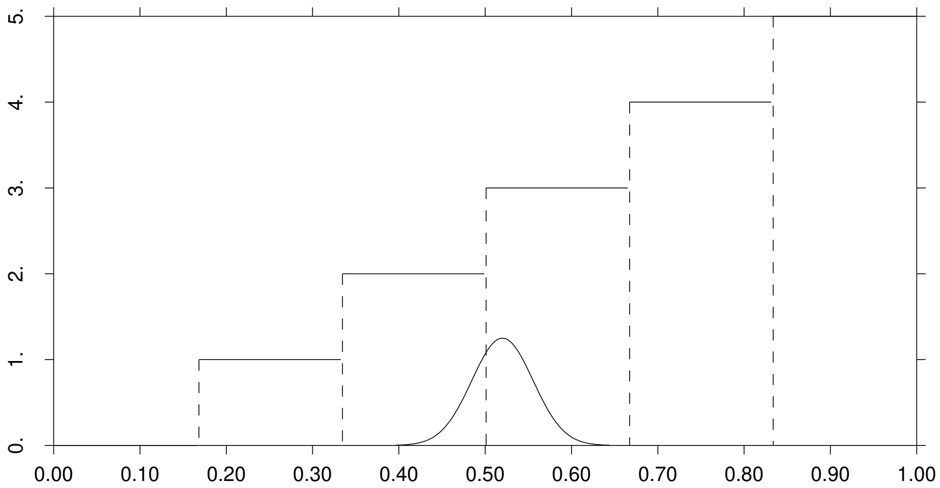
Rango de los cocientes

0	2	4	6	8	11	12	16	17	20	22	24
1	5	10	14	18	23	27	30	35	39	43	48
3	7	13	19	25	29	36	41	46	52	56	59
9	21	34	45	57	65	71	78	85	88	92	95
15	32	50	61	73	82	89	94	96	99	101	104

Cáceres, elecciones generales 2000. Dos partidos, cinco escaños.

PP obtuvo el 52 % de los votos: 3 escaños.

El muestreo con $N = 199$ da lugar a proporciones muestrales con distribuciones aproximadamente normales tal como se indica en el gráfico: el 28 % de las muestras asignan 2 escaños al PP.



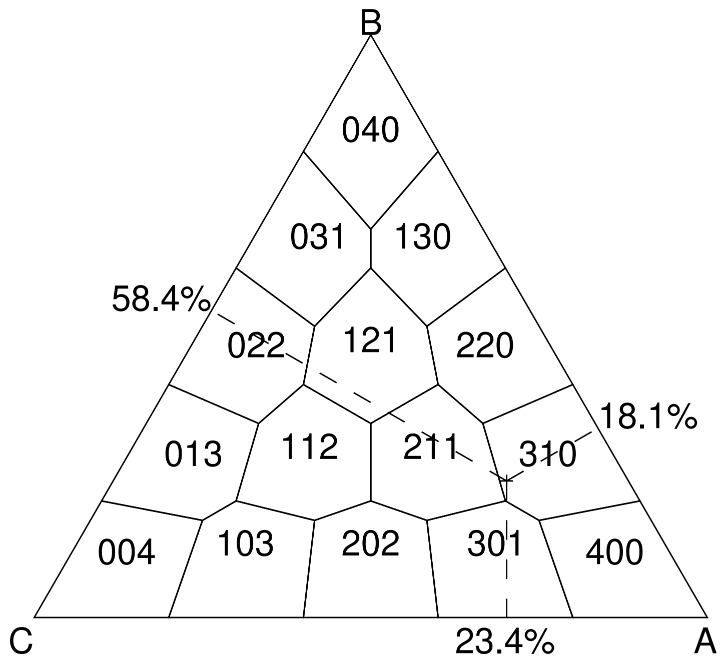
Sesgo$= (0{,}28 \times 2 + 0{,}72 \times 3) - 3 = 2{,}72 - 3 =$**-0.28**

La ley d'Hondt puede escribirse como

$$S_d(f_1,\ldots,f_K) = (m_1,\ldots,m_K) \Longleftrightarrow$$
$$\forall i,j \in \{1,\ldots,K\}, i \neq j, \quad m_i = 0 \text{ o } \frac{f_i}{d_{m_i}} > \frac{f_j}{d_{m_j+1}} \tag{3}$$

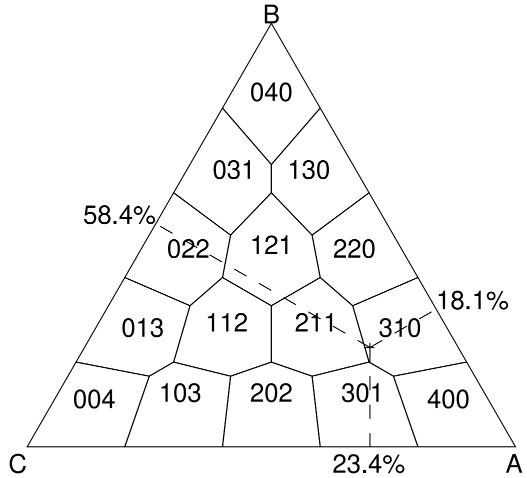
que dice que el último cociente del partido i al que le corresponde un escaño debe ser mayor que el primer cociente del partido j al que no le corresponde escaño. Según esta expresión, las regiones del espacio de proporciones de votos con **asignación de escaños constante** (constant seat allocation, CSA) están limitadas por hiperplanos y, por tanto, son poliedros convexos.
Llamamos **celdas CSA** a esas regiones.

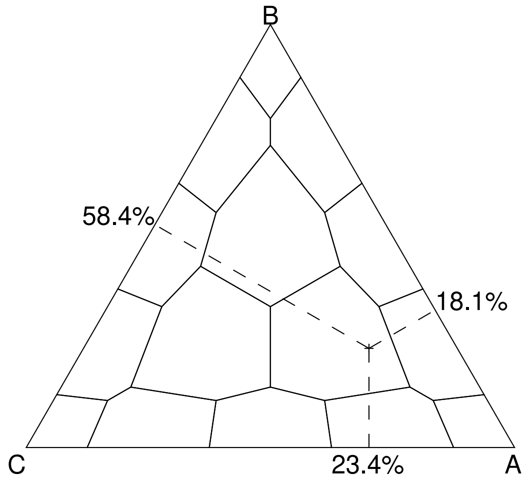
Ourense: 3 partidos, 4 escaños

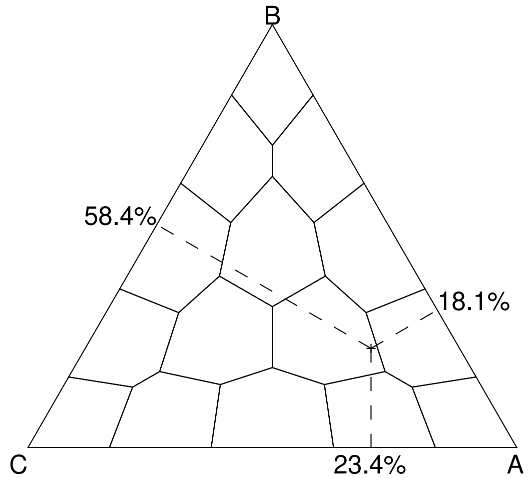


Gráficos de las reglas *proporcionales*

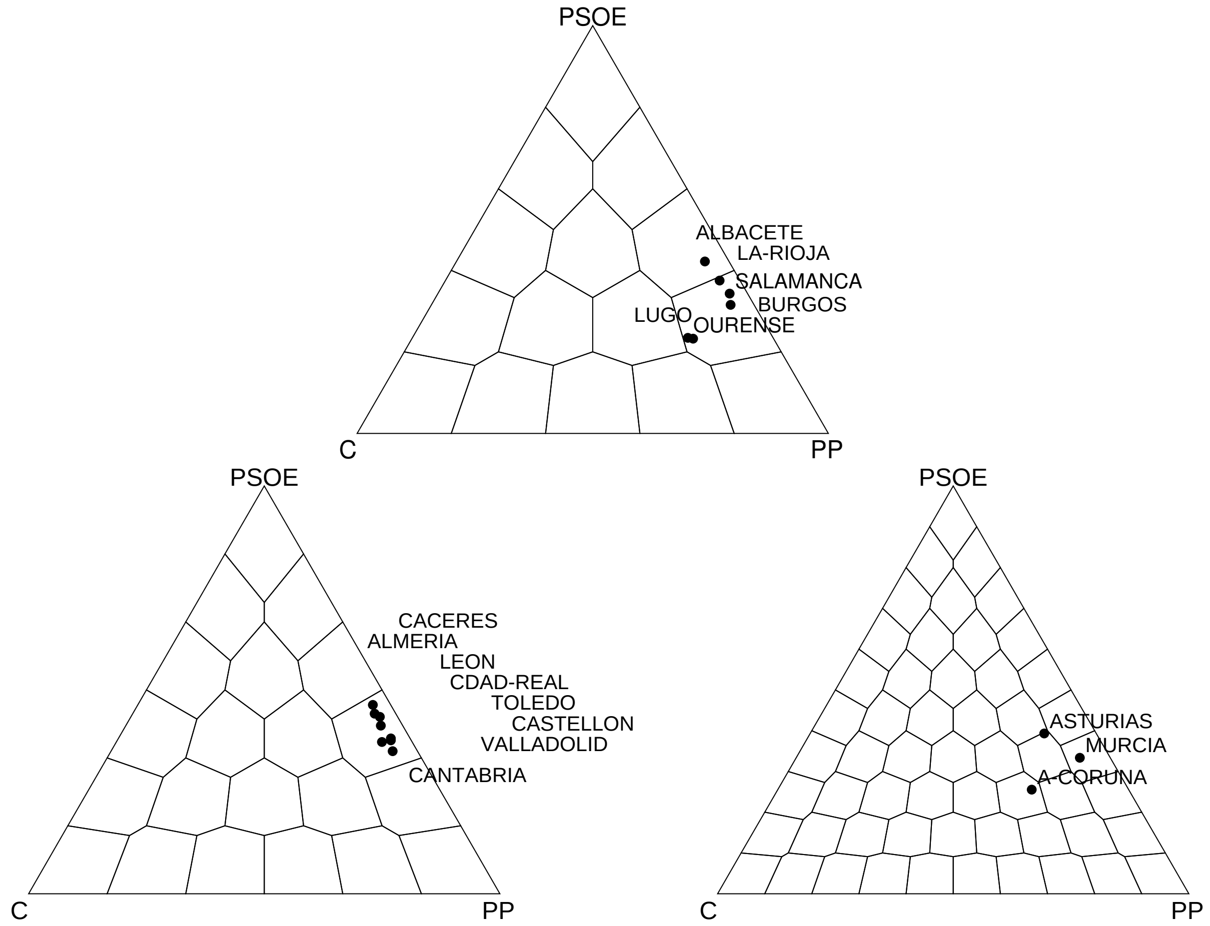
d'Hondt: $d = \{1, 2, 3, \ldots\}$
St. Lagüe: $d = \{1, 3, 5, 7, \ldots\}$
St. Lagüe Modif.: $d = \{1{,}4, 3, 5, 7, \ldots\}$



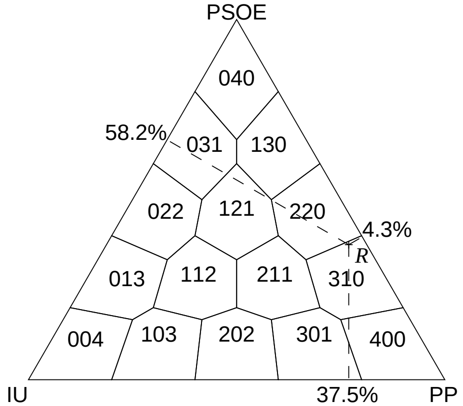




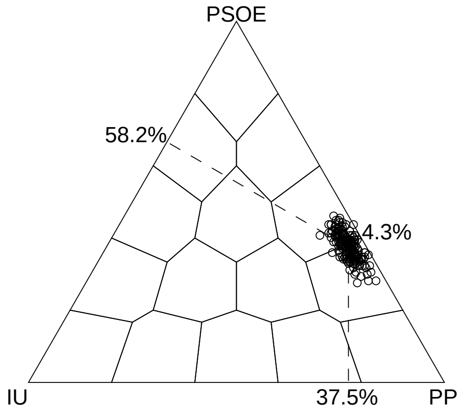
Varias provincias en un gráfico



La Rioja 2000 (4 escaños)



$B = 200$ muestras de tamaño $n = 167$

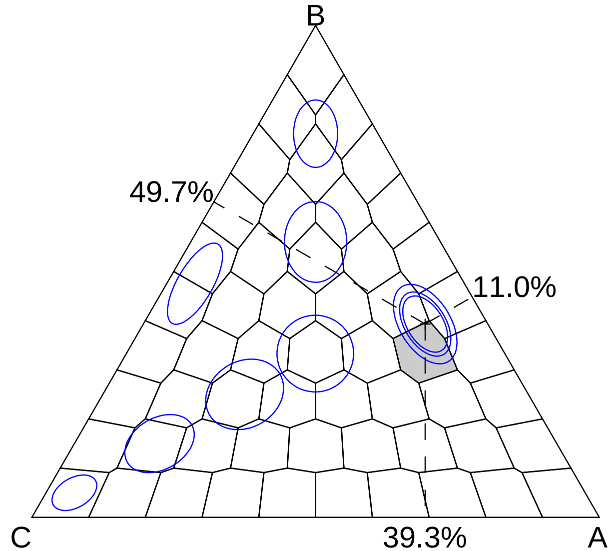


Asturias 2000 (9 escaños)

Tamaño muestral: $n = 337$

Probabilidades: 0,90, 0,95, 0,99 alrededor del punto marcado,

0,99 para las demás.



Usando la aproximación de la multinomial por la normal multivariante se pueden dibujar las elipses en las que las proporciones muestrales ponen probabilidad $(1 - \alpha)$.

Reducción del sesgo mediante bootstrap paramétrico (I)

1. Sea S_0 un sondeo electoral con tamaños muestrales n_j en cada provincia.
   Sean $\widehat{f}_{ij}, i = 1 \ldots K, j = 1 \ldots C$ las estimaciones de las proporciones de votos obtenidas por el partido i en la provincia j.
   Sea $\widehat{m}_{ij}$ el número de escaños asignados al partido i en la provincia j.

2. Repetir B_1 veces:

   a) Para cada provincia $j = 1 \ldots C$, obtener una muestra de tamaño n_j de una multinomial de parámetros $\widehat{f_{ij}}$, $i = 1 \ldots K$.

   b) Estimar los escaños: $\widehat{m}^*_{ij}$.

3. A partir de los escaños bootstrap $\widehat{m}^*_{ij}$ obtenidos en el paso 2, calcular la media $\bar{m}^*_{ij}$ de las B_1 muestral bootstrap.
   Estimar el sesgo $E(\widehat{m}_{ij}) - m_{ij}$ mediante $\bar{m}^*_{ij} - \widehat{m}_{ij}$.
   El estimador bootstrap de m_{ij} será

$$\widehat{m}^B_{ij} = \widehat{m}_{ij} - (\bar{m}^*_{ij} - \widehat{m}_{ij}).$$

Reducción del sesgo mediante bootstrap paramétrico (II)

Elecciones generales 2000.

Usamos $B_0 = B_1 = 1000$, $N = 15000$, $K = 12$ partidos en $C = 52$ provincias.

Primera fila: promedio de los sesgos observados en los 1000 sondeos simulados.
Segunda fila: promedio de los 1000 sesgos estimados mediante bootstrap.

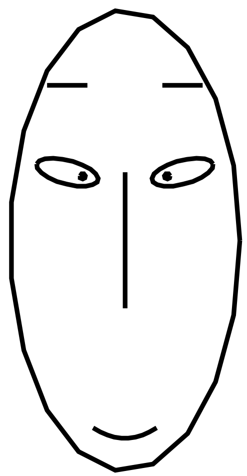
PP	PSOE	CIU	IU	PNV	CC	BNG	PA	ERC	ICV	EA	CHA
-2.51	2.04	0.45	-0.85	-0.26	0.39	0.86	-0.08	0.19	-0.02	-0.11	-0.09
-0.45	0.48	0.01	0.03	-0.12	0.00	0.08	0.01	0.12	-0.04	-0.03	-0.09

La magnitud del sesgo se ha infraestimado.
La dirección de los sesgos estimados es la correcta:
las dos filas tienen una correlación de 0,91.

Gráfico de las caras de Chernoff. (I)

Se asocia cada partido con una característica de la cara.

Partido	Característica de la cara
PP	Área de la cara
PSOE	Forma de la cara
CiU	Longitud de la nariz
IU	Posición de la boca
PNV	Curvatura de la sonrisa
CC	Anchura de la boca
BNG	Posición de los ojos
PA	Separación de los ojos
ERC	Ángulo de los ojos
IC-V	Forma de los ojos
EA	Anchura de los ojos
CHA	Posición de las pupilas



Escanyos.12M

Gráfico de las caras de Chernoff. (II)

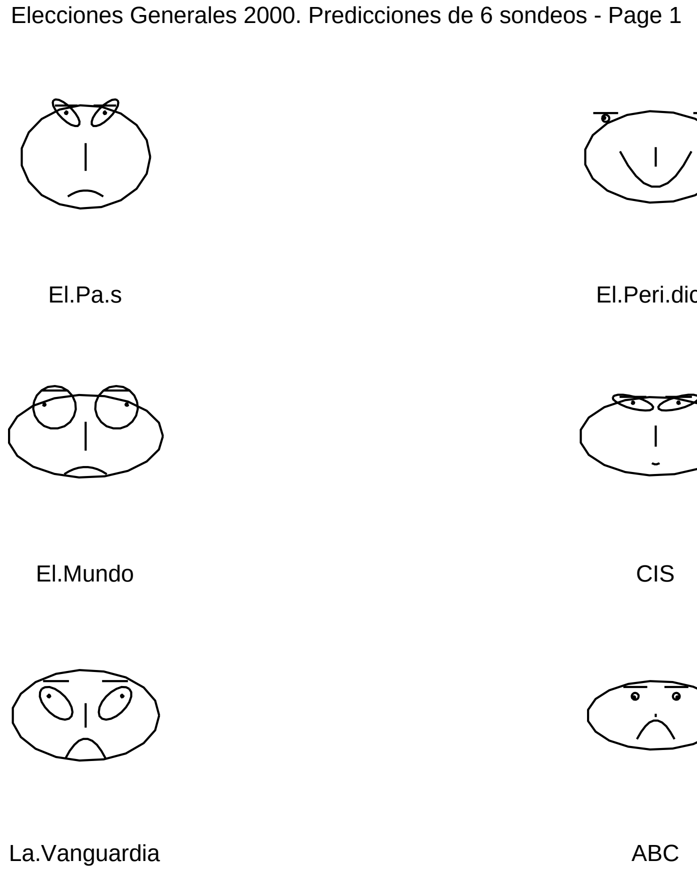
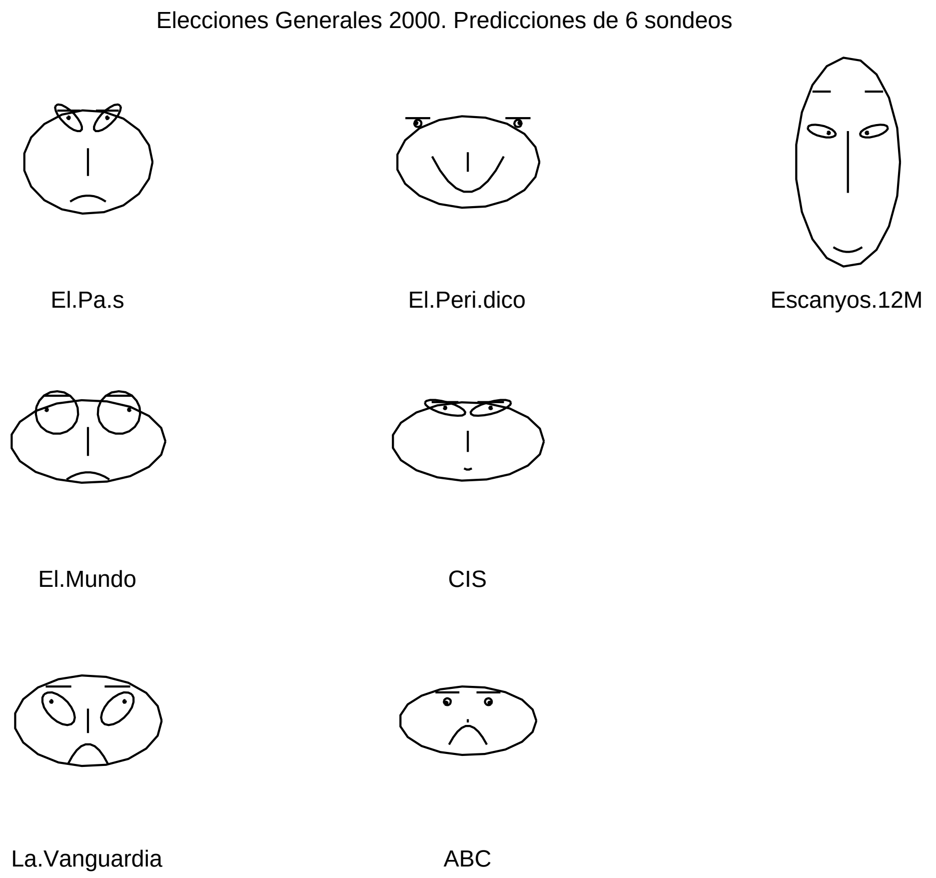


Gráfico de las caras de Chernoff. (III)



8. Conclusiones.

- Los sondeos electorales publicados en la prensa no parecen responder al mismo modelo que los sondeos simulados.
- Posibles razones:
  - Se toman muestras del **censo electoral**, no de los **votantes reales**.
  - **Elección sesgada de los entrevistados:** Sólo se entrevista a aquellas personas que pueden y quieren ser entrevistadas.
  - Muchos entrevistados **no saben qué votarán** con seguridad.
  - Hay gente que **no manifiesta su intención real de voto**.
  - Es difícil realizar la **imputación de voto.**

- Las encuestas bien hechas dan buenos resultados. **Ejemplo:**
  - La noche electoral del 16-Nov-2003, a las 22:45 se había recontado el 66 % de los votos (3.400.000).
  - El PSC era entonces la fuerza con más representación.
  - TV3 anunció el resultado de un muestreo de las primeras 100 papeletas de 200 mesas electorales representativas (estratégicamente seleccionadas).
  - Ese muestreo (20.000 votos) acertó de pleno.