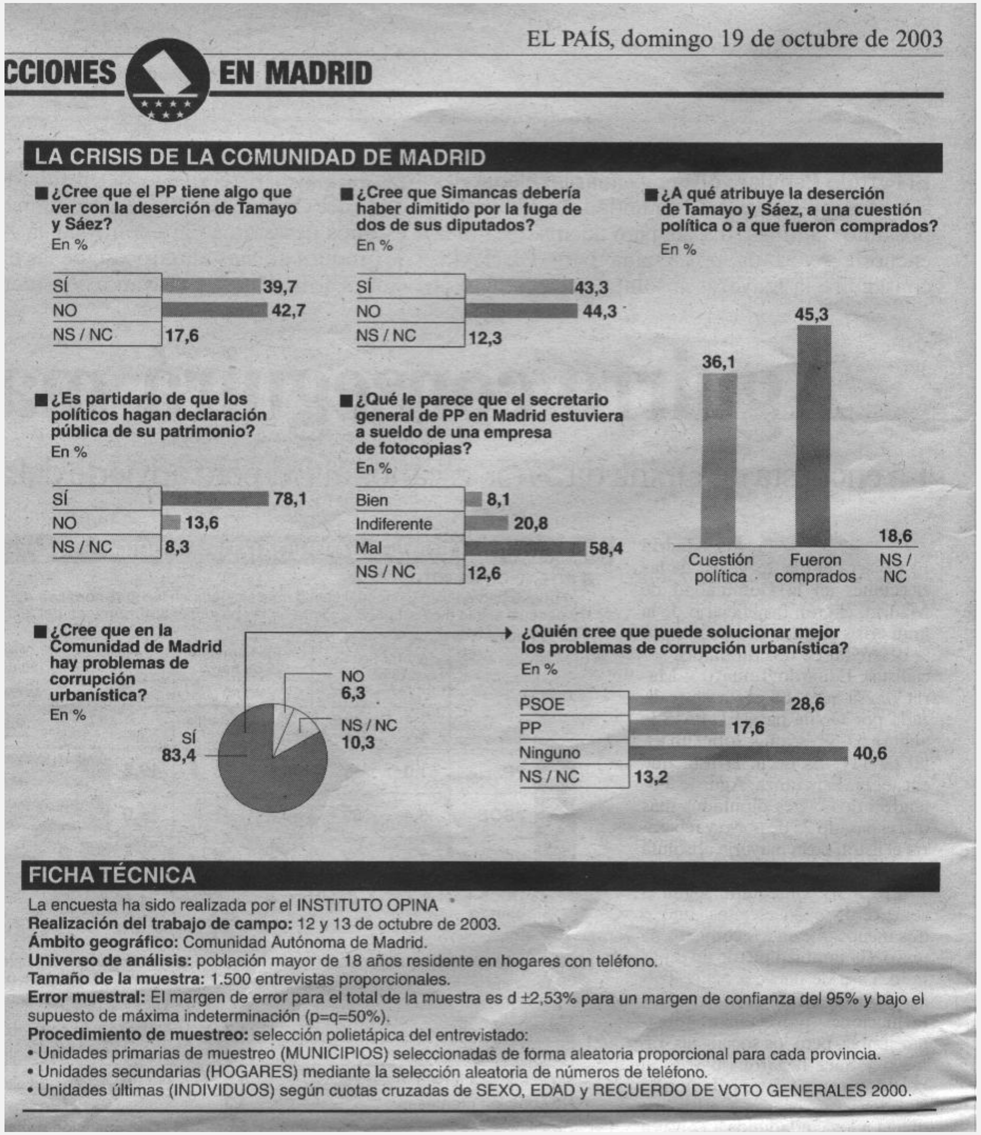
Vots i escons

Angel Corberán & Francisco Montes
Departament d'Estadística i I. O.
Universitat de València
http://www.uv.es/~montes

Abans de les eleccions

... enquestes d'opinió



EL PAÍS, domingo 19 de octubre de 2003

CCIONES EN MADRID

LA CRISIS DE LA COMUNIDAD DE MADRID

¿Cree que el PP tiene algo que ver con la deserción de Tamayo y Sáez?
En %

SÍ	39,7
NO	42,7
NS / NC	17,6

¿Cree que Simancas debería haber dimitido por la fuga de dos de sus diputados?
En %

SÍ	43,3
NO	44,3
NS / NC	12,3

¿A qué atribuye la deserción de Tamayo y Sáez, a una cuestión política o a que fueron comprados?
En %

Cuestión política	36,1
Fueron comprados	45,3
NS / NC	18,6

¿Es partidario de que los políticos hagan declaración pública de su patrimonio?
En %

SÍ	78,1
NO	13,6
NS / NC	8,3

¿Qué le parece que el secretario general de PP en Madrid estuviera a sueldo de una empresa de fotocopias?
En %

Bien	8,1
Indiferente	20,8
Mal	58,4
NS / NC	12,6

¿Cree que en la Comunidad de Madrid hay problemas de corrupción urbanística?
En %

SÍ	83,4
NO	6,3
NS / NC	10,3

¿Quién cree que puede solucionar mejor los problemas de corrupción urbanística?
En %

PSOE	28,6
PP	17,6
Ninguno	40,6
NS / NC	13,2

FICHA TÉCNICA

La encuesta ha sido realizada por el INSTITUTO OPINA
Realización del trabajo de campo: 12 y 13 de octubre de 2003.
Ámbito geográfico: Comunidad Autónoma de Madrid.
Universo de análisis: población mayor de 18 años residente en hogares con teléfono.
Tamaño de la muestra: 1.500 entrevistas proporcionales.
Error muestral: El margen de error para el total de la muestra es d ±2,53% para un margen de confianza del 95% y bajo el supuesto de máxima indeterminación (p=q=50%).
Procedimiento de muestreo: selección polietápica del entrevistado:

- Unidades primarias de muestreo (MUNICIPIOS) seleccionadas de forma aleatoria proporcional para cada provincia.
- Unidades secundarias (HOGARES) mediante la selección aleatoria de números de teléfono.
- Unidades últimas (INDIVIDUOS) según cuotas cruzadas de SEXO, EDAD y RECUERDO DE VOTO GENERALES 2000.

Què vol dir tot aquest embolic de la *fitxa tècnica*?

FICHA TÉCNICA

La encuesta ha sido realizada por el INSTITUTO OPINA
Realización del trabajo de campo: 12 y 13 de octubre de 2003.
Ámbito geográfico: Comunidad Autónoma de Madrid.
Universo de análisis: población mayor de 18 años residente en hogares con teléfono.
Tamaño de la muestra: 1.500 entrevistas proporcionales.
Error muestral: El margen de error para el total de la muestra es d ±2,53% para un margen de confianza del 95% y bajo el supuesto de máxima indeterminación (p=q=50%).
Procedimiento de muestreo: selección polietápica del entrevistado:
- Unidades primarias de muestreo (MUNICIPIOS) seleccionadas de forma aleatoria proporcional para cada provincia.
- Unidades secundarias (HOGARES) mediante la selección aleatoria de números de teléfono.
- Unidades últimas (INDIVIDUOS) según cuotas cruzadas de SEXO, EDAD y RECUERDO DE VOTO GENERALES 2000.

particularment, quan parla de l'**error muestral**, què significa allò de

- marge d'error,
- marge de confianza, i
- màxima indeterminació ($p = q = 50\,\%$)?

Estimació d'una proporció

El que en una enquesta de grandària *n* ens interesa és estimar la proporció de votants a un cert partit A. L'enquesta ens proporciona el número d'enquestats que diuen van a votar-lo. Es sabut que la millor manera, i la més obvia, d'estimar p és fer-ho mitjançant la proporció en la mostra. Això és, si dels n enquestats m han dit que votaran al partit A, aleshores

$$\hat{p} = \frac{m}{n}.$$

Però, que passa amb $\hat{\mathrm{p}}$?

Que és una quantitat aleatòria que pren els valors

$$\hat{p} = \frac{m}{n}, \quad m = 0, 1, 2, \ldots, n.$$

Si podem conèixer la seua distribució de probabilitat, podrem dir alguna cosa de la seua variabilitat. Es a dir, de l'error que puga haver-hi en la seua estimació, i podrem començar a esbrinar el que diu la *fitxa tècnica*.

Distribució de probabilitat de $\hat{\mathrm{p}}$

Com que la grandària de la mostra n es fixa, el que varia en $\hat{p}$ és el numerador de la fracció, m. Si li diem X al nombre d'enquestats que han dit que votaran al partit A, és clar que

$$P\left(\hat{p} = \frac{m}{n}\right) = P(X = m), \quad m = 0, 1, 2, \ldots, n.$$

Si l'enquesta ha estat feta com cal (mostra aleatòria), $X \sim B(n, p)$ on p és la verdadera i desconeguda proporció.

Característiques de $\hat{\mathrm{p}}$

Com que $\hat{p} = X/n$,

$$E(\hat{p}) = E(X/n) = p, \qquad var(\hat{p}) = \frac{1}{n^2} var(X) = \frac{p(1-p)}{n}. \tag{1}$$

La variança ens pot ajudar a calcular l'error de l'estimació (puntual) de p, però tenim un entrebanc insalvable, p és desconeguda. El que podem fer és estimar la variança substituint en (1) p per la seua estimació, el que ens donarà

$$\widehat{var}(\hat{p}) = \frac{1}{n^2} \widehat{var}(X) = \frac{\hat{p}(1-\hat{p})}{n}. \tag{2}$$

Interval de confiança per a p

La manera clàssica d'introduir l'error en la estimació puntual és construir un interval de confiança. Es tracta d'un interval aleatori que ens informa del conjunt de possibles valors que pot prendre p, sempre a partir de l'estimació que hem fet mitjançant $\hat{p}$. En concret, un Interval de Confiança al $q\,\%$, $0 \leq q \leq 100$, IC_q verifica

$$P(p \in IC_q) = \frac{q}{100}. \tag{3}$$

Com obtenir IC_q?

$\longrightarrow$

Obtenció d' $\mathbf{IC_q}$

La forma d'IC_q depèn de $\hat{p}$ i de la seua distribució. Farem ús del fet que

$$B(n,p) \underset{n\to\infty}{\longrightarrow} N\left(np, np(1-p)\right).$$

La definició de l'interval és aleshores,

$$IC_q = \left[\hat{p} - t_{n-1,q}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} - t_{n-1,q}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right], \tag{4}$$

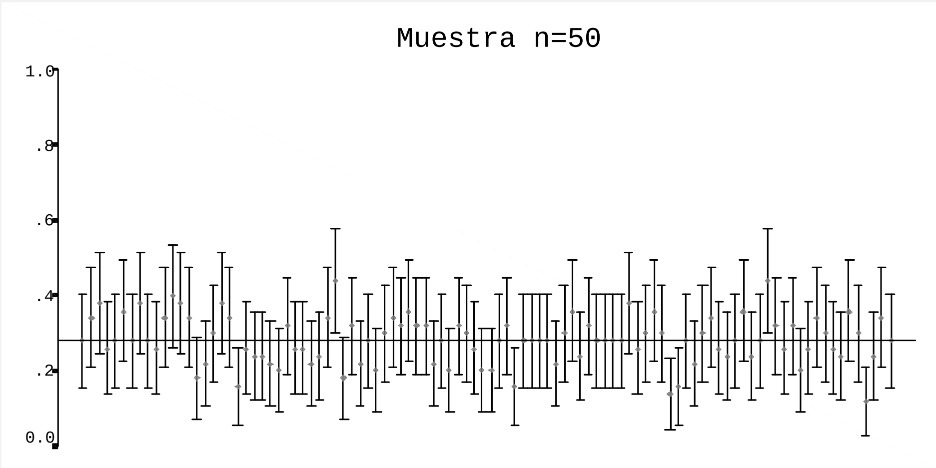
on $t_{n-1,q}$ es el percentil q d'una t de Student amb $n-1$ graus de llibertat.

Propietats d' $\mathbf{IC_q}$

1. És simètric respecte de $\hat{p}$.
2. La longitud de l'interval depèn de dos factors que actúen oposadament, com el sentit comú exigeix.
   - El primer, $t_{n-1,q}$, és un valor que depèn sobretot de q i que augmenta conforme ho fa q.
   - El segon, n, actua inversament. Més gran és n més menut és IC_q.
3. La interpretació de la igualtat (3) significa que si tenim moltes mostres de la mateixa grandària, aproximadament el $q\,\%$ dels IC_q cobriran la vertadera proporció, p.

Un exemple d'IC$_\mathbf{q}$

En la gràfica hem representat els 100 IC_{95} obtinguts a partir de 100 mostres de grandària 50 tretes d'una població en la que hi ha una proporció de lectors de la revista `Diez minutos a la semana` coneguda, $p = 0{,}28$. Conèixer p ens permet contar aquells IC_{95} que no tallen la recta $p = 0{,}28$.



Tornem a la fitxa tècnica

Quan la fitxa diu

> El margen de error para el total de la muestra es $\pm 2{,}53\,\%$ para un margen de confianza del 95 % y ...

el que diu, en llenguatge probabilístic és, senzillament

$$P(|p - \hat{p}| \leq 0{,}0253) = 0{,}95. \tag{5}$$

Però

$$|p - \hat{p}| \leq 0{,}0253$$

equival a

$$\hat{p} - 0{,}0253 \leq p \leq \hat{p} + 0{,}0253$$

Marge d'error i interval de confiança

És a dir

$$P(|p - \hat{p}| \leq 0{,}0253) = 0{,}95$$
$$\Updownarrow$$
$$P(\hat{p} - 0{,}0253 \leq p \leq \hat{p} + 0{,}0253) = 0{,}95 \tag{6}$$

i si recordem les característiques d'IC_{95} el que diu (6) és que la semilongitud d'IC_{95} i el marge d'error són el mateix,

$$\delta = t_{n-1,q}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}. \tag{7}$$

Càlcul del marge d'error

Si volem comprovar (7), la fitxa tècnica diu que $n = 1500$ i sabem que $t_{1499,95} = 1{,}96$, però ens falta conèixer $\hat{p}$. Necessitem fer l'enquesta per a saber-ho. Ara s'entén allò que diu la darrera frase *"... bajo el supuesto de máxima indeterminación ($p = 1 - p = 50\,\%$)"*. Significa que en no conèixer $\hat{p}$ se li assigna el valor més desfavorable, $\hat{p} = 0{,}5$, que fa màxim el producte $\hat{p}(1 - \hat{p})$. Ara sí,

$$\delta = 1{,}96\sqrt{\frac{1}{4 \times 1500}} = 0{,}0253$$

Grandària de la mostra i marge d'error

Però (7) pot també emprar-se en sentit contrari: fixat el marge d'error, δ, hom pot determinar la grandària de mostra necessària per assolir-lo. Assumida la situació més desfavorable per a $\hat{p}$, $\hat{p}(1-\hat{p}) = 1/4$, i com que per a n gran la t de Student depèn tan sols de q, tindrem

$$\delta = \frac{t_q}{2\sqrt{n}} \quad \Rightarrow \quad n = \frac{t_q^2}{4\delta^2}, \tag{8}$$

	valor de p			
error	**0,2**	**0,3**	**0,4**	**0,5**
0,020	1961	2246	2401	2451
0,025	1255	1438	1537	1569
0,030	872	998	1067	1089
0,035	641	734	784	801
0,040	491	562	601	613
0,045	388	444	475	484
0,050	314	360	385	393

Nivell de confiança 95 %

Després de les eleccions

... l'escrutini

Cal parlar de l'escrutini?

A mesura que els recomptes de vots de cada taula arriben a l'organisme encarregat de l'escrutini oficial, es van fent públics els resultats parcials: distribució de vots i d'escons.

Cal pensar que si l'arribada dels resultats de cada col.legi és aleatòria, escrutat un cert percentatge les distribucions de vots i escons no variaran gaire. Amb un percentatge del 50 % ja comença a albirar-se com pot concloure la cosa. Però ...

Alguns escrutinis fan parlar

De l'escrutini de les darreres eleccions al parlament de la Comunitat de Madrid se'n ha parlat i se'n ha escrit molt. La raó estaria en l'evolució dels resultats al llarg de la nit. Des d'una derrota clara del PP a l'inici, les dades varen evolucionar de manera constantment favorable als populars, que assoliren a la fi de l'escrutini una majoria absoluta.

Vegem el curios escrutini.

L'escrutini del 26-O

$\longrightarrow$

	escons			
cens escrutat	**PP**	**PSOE**	**IU**	**majoria**
25 %	50	51	10	PSOE + IU
50 %	52	49	10	PSOE + IU
75 %	54	47	10	PSOE + IU
90 %	55	46	10	PSOE + IU
92 %	56	45	10	PP
100 %	57	45	9	PP

L'escrutini del 26-O

Als diaris s'han publicat articles comentant aquesta evolució del vot. Com exemple dos publicats a EL PAÍS:

- *El escrutinio del 26-O, magia potagia* (29-10-2003). L'autor, Antonio Kindelán, manté la tesi d'un escrutini manipulat, i
- *Poca magia y mucha demagogia* (31-10-2003). Els autors, Josu Mezo y Juan C. Rodríguez, no troben res d'anòmal en el recompte i l'evolució dels escons.

Paga la pena pegar-los una ullada.

Com és l'escrutini parcial?

Si acceptem la tesi d'En Kindelán, la informació de les taules arriba aleatòriament, el escrutini parcial és aleatori i ho será també qualsevol característica numèrica que d'ell se'n derive. Per això diu que *"La probabilidad de que todas esas circunstancias se produzcan al tiempo por azar es prácticamente nula."*

Si coneguérem la seua distribució de probabilitat baix l'hipòtesi d'arribada aleatòria, podríem fer un contrast d'aquesta hipòtesi, però no hi ha forma senzilla de conèixer-la. Alguna cosa hauria de poder fer-se.

El Mètode de Montecarlo i l'escrutini parcial

El que podem fer és simular n escrutinis amb l'hipòtesi d'arriba aleatòria dels resultats i obtenir per a cadascun d'ells una taula d'evolució del repartiment d'escons com la d'aquella nit. Després compararem aquella taula amb les n simulades i en traurem conclusions (Mètode de Montercarlo (MM)).

Necessitem els resultats de les 5863 taules, disponibles en `www.comad.es`, i un bon software. La potència de càlcul dels actuals PC permet fer un grapat de simulacions en no res.

Resultats del MM amb 1000 simulacions

	escons			
cens escrutat	**PP**	**PSOE**	**IU**	**majoria**
25 %	55 (*50*)	46 (*51*)	10 (*10*)	PSOE + IU
50 %	56 (*52*)	45 (*49*)	10 (*10*)	PP (*PSOE + IU*)
75 %	56 (*54*)	45 (*47*)	10 (*10*)	PP (*PSOE + IU*)
90 %	56 (*55*)	45 (*46*)	10 (*10*)	PP (*PSOE + IU*)
100 %	57 (*57*)	45 (*45*)	9 (*9*)	PP

Escrutini més desfavorable al PP (58 de 1000). Entre parèntesi l'escrutini original. Les conclusions ...

L'altra hipòtesi: tot era normal

Les altres autors mantenen que l'escrutini va ser normal perquè. *"La primera hipótesis que cualquiera se haría sería: se tardará más en contar los votos en aquellas mesas donde hayan votado más personas."* Per a recolzar la seua hipòtesi ordenen les taules de menor a major, fan l'escrutini i obtenen un resultat pràcticament igual que l'original.

Un crítica inmediata a aquesta proposta en la manca absoluta d'aleatorietat. Cal acceptar que els resultats arriben ordenadets, l'un darrere l'altre per ordre de cens de la taula? Cap problema? Cap incidència?

Una variant: ordre aleatori ponderat

El determinisme d'una ordenació per cens de la taula és difícil de pair. És poc creïble que no hi haja un factor aleatori que altere l'orde. Aceptar-lo tal qual equival quasi a admetre la manipulació de l'escrutini.

Però aquesta idea d'un ordre que pare compte de la grandària de la taula no és cap bogeria. Podem fer-la compatible amb l'existència d'un factor aleatori. Com? Assignant a cada taula un pes inversament proporcional al seu i ordenant després aleatòriament d'acord amb estos pesos (permutació aleatòria ponderada de les 5863 taules). La resta serà aplicar el Mètode de Montecarlo.

Resultats del MM ponderat amb 1000 simulacions

	escons			
cens escrutat	**PP**	**PSOE**	**IU**	**majoria**
25 %	54 (*50*)	47 (*51*)	10 (*10*)	PSOE + IU
50 %	55 (*52*)	46 (*49*)	10 (*10*)	PSOE + IU
75 %	56 (*54*)	46 (*47*)	9 (*10*)	PP (*PSOE + IU*)
90 %	56 (*55*)	45 (*46*)	10 (*10*)	PP (*PSOE + IU*)
100 %	57 (*57*)	45 (*45*)	9 (*9*)	PP

Escrutini més desfavorable al PP (65 de 1000). Entre parèntesi l'escrutini original. Les conclusions ...

Després de les eleccions

... els escons

Circumscripcions i escons

Caldria començar dient alguna cosa sobre les escons assignats a cada circumscripció. El criteri de distribució d'escons és, principalment, polític i sol seguir la regla general de procurar contentar a tots. El resultat és, quasi sempre, injust i desequilibrat com posa de manifest les diferències en el *preu* de l'escó segons circumscripcions.

Les eleccions al Parlament de Catalunya del propassat dia 16 de novembre i els escons assignats a les quatre circumscripcions catalanes, ens poden servir d'exemple per a il·lustrar el que diem.

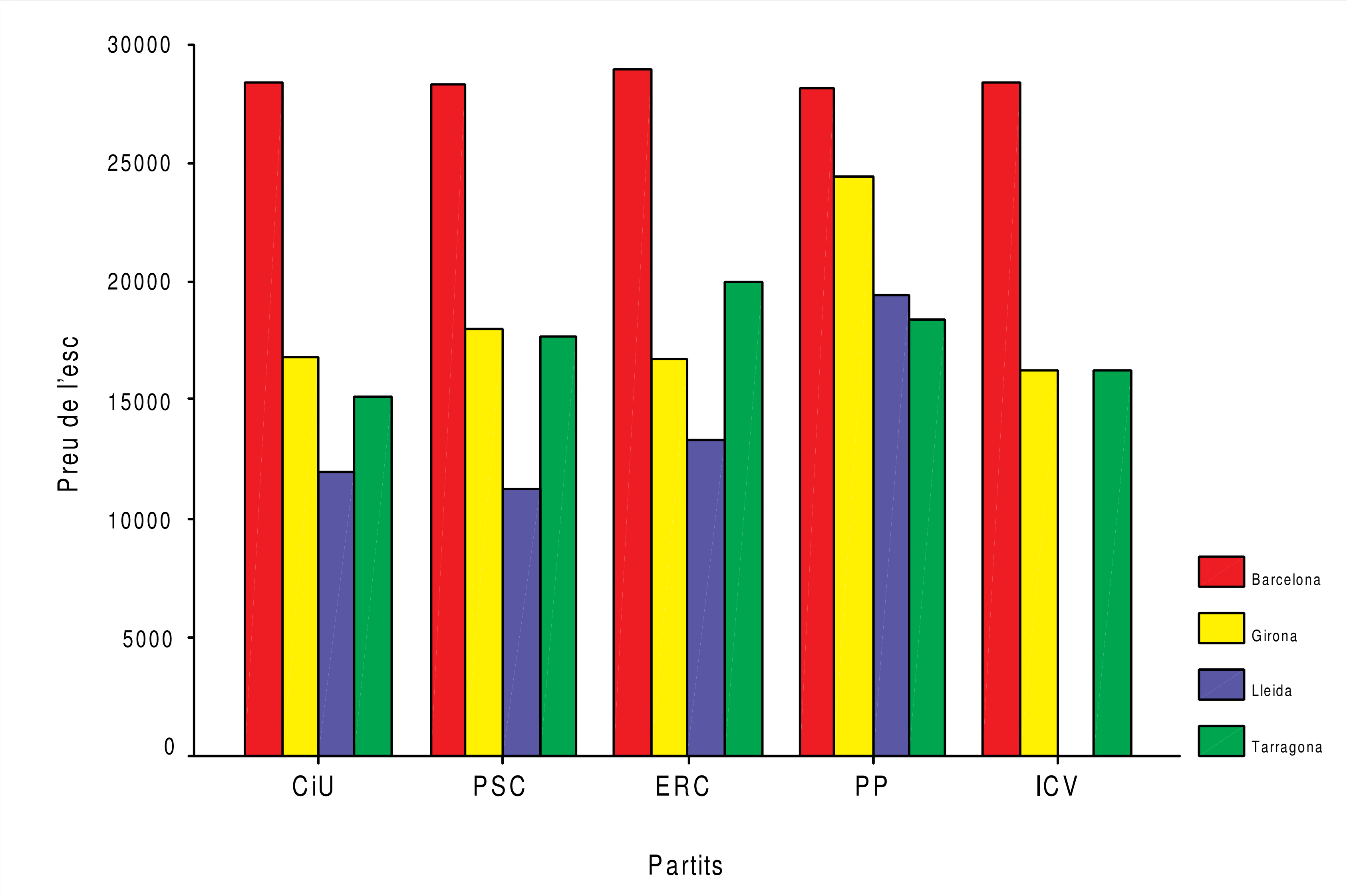
Resultats de les eleccions al Parlament de Catalunya 2003

	Barcelona		Girona		Lleida		Tarragona	
	vots	es	vots	es	vots	es	vots	es
CiU	710.549	25	117.799	7	83.636	7	106.131	7
PSC-CpC	820.153	29	71.965	4	45.214	4	88.698	5
ERC	375.222	13	66.827	4	40.131	3	59.865	3
PP	309.806	11	24.487	1	19.446	1	36.911	2
ICV-EA	199.114	7	16.218	1	8.750		16.276	1

Preu de l'escó a les circumscripcions catalanes (Taula)

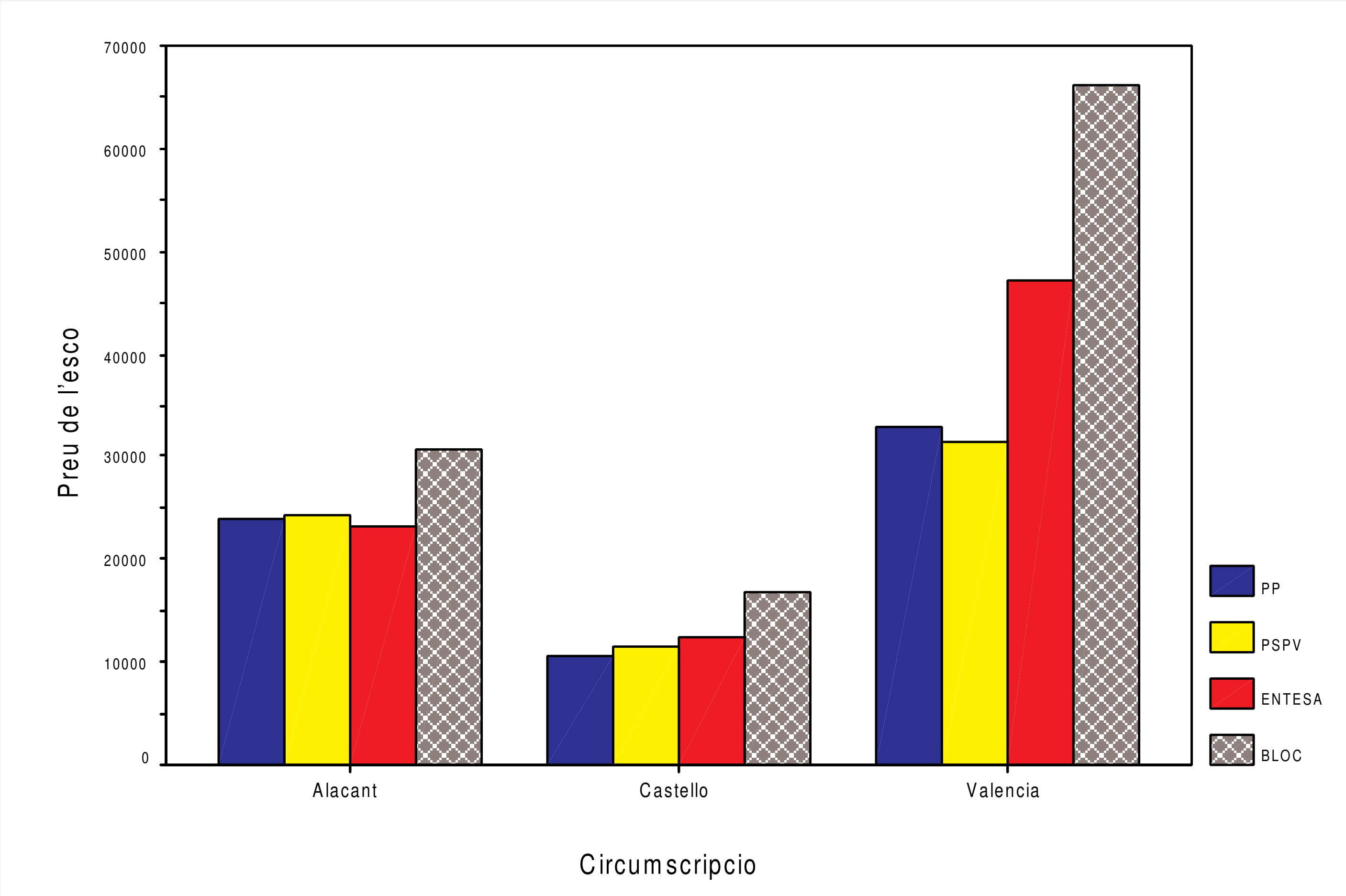
	Barcelona	**Girona**	**Lleida**	**Tarragona**
CiU	28422	16828	11948	15162
PSC-CpC	28281	17991	11304	17740
ERC	28863	16707	13377	19955
PP	28164	24487	19446	18456
ICV-EA	28445	16218		16276
mitjana	28435	18446	14019	17518
desv.tip	264,89	3438,97	3720,54	1868,31

Preu de l'escó a les circumscripcions catalanes (Gràfic)



Preu de l'escó a les circumscripcions del PV (Gràfic)

Però la cosa pot encara empitjorar ...



Mètodes de repartiment dels escons

La manera més inmediata de repartir una quantitat entre un grup de subjectes prenent com a criteri alguna característica numèrica, de forma que quan major siga aquesta més li corresponga, és el repartiment proporcional. Si a més a més emprem una proporcionalitat exacta, no tindrem cap problema. El problema es presenta quan hi han restriccions, com ara que "la solució deu estar formada per per números enters no negatius, que sumen una quantitat fixa".

A nosaltres ens interessa repartir escons entre partits d'acord amb els vots rebut en una elecció. La restricció és doncs inevitable.

Mètodes de repartiment proporcional d'escons

Les solucions al problema, en plural perquè no és única, les seues propietats i tot el que vulgueu saber sobre les matemàtiques que hi ha al darrere podeu trobar-ho en l'article de V. Ramírez (2002), *Matemática electoral*[a].

Parlem-ne de dos dels mètodes més coneguts: el Hamilton o dels Residus Majors i el dels divisors o Llei d'Hondt.

---

[a] Article publicat en "El Lenguaje de la Matemáticas en sus Aplicaciones", Secretaría General de Formación del Profesorado, Ministerio de Educación, Cultura y Deporte, Madrid, ISBN: 84-369-3535-7

Mètode de Hamilton o dels Residus Majors

Si hem repartir m escons, si v_i designa el número de vots del partir i $v = \sum_{i=1}^{n} v_i$, la quota que li correspon al partit i és

$$q_i = m\frac{v_i}{v}.$$

El repartiment es fa en dues pases:

1. A cada partit li assignem $\lfloor q_i \rfloor$ escons. Aquest repartiment no esgotara els n, excepte si totes les q_i son enteres. Suposem que en queden k.

2. Ordenem els residus $r_i = q_i - \lfloor q_i \rfloor$ de major a menor i assignem un escó adicional als k primers.

Mètode dels divisors o Llei d'Hondt

Consisteix en dividir els vots de partit per els naturals $1, 2, 3, 4 \ldots$ i ordenar conjuntamente els quocients de tots els partits. Els n escons s'assignen als n primers.

El mètode rep el nom del matemàtic belga Victor d'Hondt que el va proposar l'any 1832, tot creient que l'havia inventat, però allò cert és que al 1790 els escons de la Cambra dels EEUU ja es varen distribuir entre els estats de la Unió amb aquest mètode, a proposta de Thomas Jefferson.

Aplicació als resultats de la Comunitat de Madrid

				escons	
	vots	**quota**	**residu**	**RM**	**d'Hondt**
IU	233.942	9,84	0,84	10	9
PP	1.333.498	56,07	0,07	56	57
PSOE	1.072.340	45,09	0,09	45	45
Total	2.639.780			111	111

Aplicació als resultats de Catalunya

	Barcelona	Girona	Lleida	Tarragona	
	RM-d'Hondt	RM-d'Hondt	RM-d'Hondt	RM	d'Hondt
CiU	25	7	7	6	7
PSC-CpC	29	4	4	5	5
ERC	13	4	3	4	3
PP	11	1	1	2	2
ICV-EA	7	1		1	1
Total	85	17	15	18	18

Això és tot

Fora de programa, algunes adreces sobre Probabilitat

- http://davidmlane.com/hyperstat/
- http://members.aol.com/johnp71/javastat.html
- http://www.math.utep.edu/Faculty/mleung/probabilityandstatistics/pslinks.html
- http://www.math.uah.edu/stat/
- http://www-stat.stanford.edu/~susan/surprise/