

UNA MOSTRA ESBIAXADA.
ESTIMACIÓ RESULTATS ELECTORALS.
METODOLOGIA.

J.M. CANALS

Es descriu un mètode de selecció de mostres utilitzant l'Anàlisi en Components Principals Normades. S'aplica al cas de selecció de Meses Electorals - per tal de realitzar estimacions ràpides al començament de l'escrutini. S'estudien les condicions que determinen quina informació utilitzar de cara a definir el biaix. S'estudia també el comportament o rendiment de la mostra efectivament utilitzada en el recompte del Referèndum Constitucional de Desembre del 1978 per al Municipi de Barcelona.

1. INTRODUCCIÓ

El 6 de Desembre del 1978 per primera vegada l'Ajuntament de Barcelona va informatitzar l'escrutini de resultats d'un Referèndum o Eleccions¹.

Els objectius d'una operació d'aquest estil són:

- Rapidesa en la disponibilitat de resultats
- Qualitat de la informació que s'emmagatzema, és a dir seguretat
- Riquesa i disponibilitat de la informació

de cara a la seva anàlisi.

La seguretat s'aconsegueix amb un sofisticat sistema de filtres i controls. De cara a la riquesa de la informació oferta es disposava en cada moment a la pantalla de teleprocés (vegeu fig.1) de l'estat de l'escrutini per a cada Districte, i, amb un cert desfasament es disposava de llistats, resums de control i mapes (vegeu fig. 2). L'endemà, amb la informació emmagatzemada es varen poder fer mapes a nivell de secció censal i anàlisis estadístiques descriptives. Al mateix temps, i sense un esforç suplementari, es poden aconseguir



Fig. 1.
Resultats del Districte 3

- J.M. Canals. Ajuntament de Barcelona. Subdepartament d'Estadística. Portal del Angel, 8. Barcelona - 2.
- Article rebut el Març del 1979.

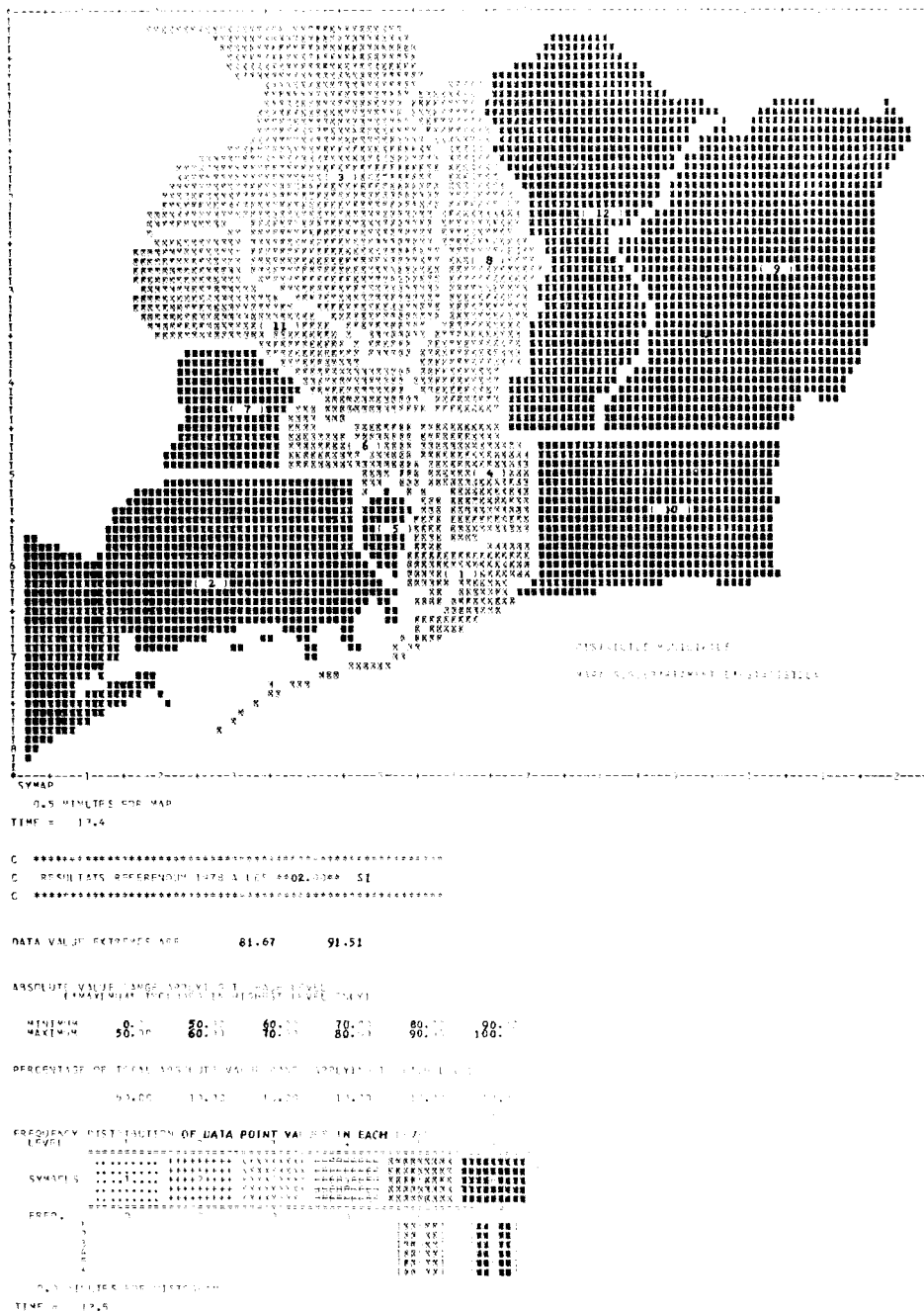


Fig. 2.
Mapa de Barcelona per Districtes

seguir avui o en qualsevol moment les dades més desagregades i en suport informàtic.

La rapidesa i la seguretat són objectius concrets. Es pot rebaixar l'hora en què es disposa dels resultats finals eliminant filtres i controls. En concret, calculem que sense filtres s'hauria pogut escurçar l'operació en quasi dues hores.

El primer resultat va entrar a quarts de 10 de la nit i l'operació es va acabar totalment a quarts de 5, les 5 de la matinada.

De tota manera l'objectiu "seguretat" es va considerar prioritari, i això tant per el pes específic d'ell mateix, com per les possibilitats que ofería el càlcul d'estimacions. Es a dir, si es calculen estimacions fiables dels resultats, es pot ajornar una mica la consecució dels resultats definitius. Cal remarcar, a més a més, que un Referèndum és l'operació d'aquest estil més curta i senzilla.

Les meses electorals que componen la mostra van rebre un tractament organitzatiu di-

Taula 1.
Resultats i estimació Barcelona

	SI	NO	BLANC	NUL	ABSTEN
RESULTATS	88,53	6,63	4,24	0,57	36,41
σ	3,07	2,67	0,43	0,08	3,39
ESTIMACIÓ	88,57	6,52	4,29	0,60	36,07
σ	3,69	3,15	0,63	0,23	3,82

ferent: el resultat es va comunicar per telèfon. A un quart d'onze del vespre les estimacions ja eren disponibles. La seva qualitat es pot veure a la Taula 1. En aquell moment, els resultats acumulats es separaven de l'estimació en més de cinc punts, en tant per cent. L'arribada de resultats a l'oficina d'escrutini depèn de la seva situació geogràfica. Volem parlar ara amb més detall de la metodologia utilitzada per escollir les Meses que componien la mostra i després analitzarem quines són les mesures més adequades per valorar la qualitat d'aquest tipus de mostres.

2. LA SELECCIÓ DE LA MOSTRA

La hipòtesi de la nostra argumentació és la següent: Existeixen Seccions que tenen un comportament electoral semblant al de la demarcació a la qual pertanyen. Es poden saber (estimar) els resultats d'aquesta darrera, coneixent els del conjunt d'aquestes Seccions escollides.

S'han de remarcar els següents extrems:

1.- La hipòtesi tradicional per seleccionar mostres és l'aleatorietat del mètode d'extracció, de cara a fer que hi quedin reflectits, i amb el mateix pes, tots els individus de l'univers. En el nostre cas, es busquen explícitament la no aleatorietat, aprofitant al màxim la informació sobre el fenomen (comportament electoral) de què es disposa. Vegem que això és possible a causa de la utilització que es vol fer de la mostra, és a dir, sols es vol estimar un fenomen, el mateix que ha donat peu a la selecció, i no un ventall de fenòmens que no han intervingut en els criteris de selecció, com és habitual. Actuar en el nostre cas utilitzant un mètode de selecció aleatori seria

desaprofitar la informació de què es disposa, actuar simulant un estat desfavorable, penalitzar a priori els resultats².

2.- Que el comportament de determinades seccions sigui substituïble pel de la demarcació que les conté -o bé a la inversa- no és pas un pur atzar, més aviat és el fruit de la barreja de classes, de la composició d'edats, dels hàbits de vot, etc., etc. És a dir, del conglomerat de molts fenòmens sociològics. La nostra argumentació es basa en l'existència de tots aquests fenòmens, però no els mesura. Busca tan sols els resultats, i deixa els components formant una "capsa negra". Ara bé, s'ha de distingir entre l'afirmació que la semblança entre comportaments no és un fruit de l'atzar, i la possibilitat que en determinats casos (només mesurant el resultat) la semblança serà realment un pur fruit de l'atzar. Per tal d'eliminar aquestes seccions que presenten un comportament semblant a la seva demarcació superior, però sols com a resultat de l'atzar, utilitzarem el criteri d'estabilitat del comportament, és a dir concomitància al llarg de diferents consultes electorals.

3.- El risc d'utilitzar la nostra hipòtesi -prové en primer lloc del canvi, en un moment donat, de la composició dels electors en una secció de la mostra, per exemple, degut a la remodelació de la zona. En segon lloc, per l'existència, en una consulta determinada, d'un fenomen electoral nou i que no segueix les pautes dels altres. En el cas de pautes semblants, la substituïbilitat funciona. Per exemple, en estimar correctament els vots en blanc al Referèndum de 1.978, que encara que era un fenomen nou, va seguir les pautes de vot establertes el 15 de Juny de 1.977.

Evidentment les desviacions entre els resultats i l'estimació provenen de canvis en el comportament de les seccions, i -- també de corriments i canvis en les pautes. Aquestes són veritablement les fonts de riscos que no es poden eliminar en considerar com una capsula negra les causes -- últimes del comportament electoral. D'altra banda, els canvis soferts per cada secció quant a altes i baixes, si que -- són coneguts, i, per tant, un refinament del mètode podria consistir en el fet de tenir en compte aquestes variacions.

- 4.- El problema estadístic, doncs, consisteix a buscar aquelles seccions que tenen un comportament electoral semblant al del seu districte. En el nostre cas, hem utilitzat l'Anàlisi en Components Principals Normades³ (ACPN) /1/, /2/ com a eina per mesurar una tal semblança. Si bé hem trobat referències a la hipòtesi de substitubilitat en la literatura sobre qüestions electorals, no hem pogut trobar -- cap descripció del mètode de selecció, i menys encara de les mesures de qualitat emprades.

Les fases del nostre treball quant a selecció de la mostra són:

- I.- Realitzar una ACPN per a cada Districte (a Barcelona hi ha 12 districtes administratius i 1.058 Seccions Censals). La matriu de dades posseeix en rengles les -- Seccions Censals del Districte i en columnes els resultats del 15 de Juny de 1.977⁴. Existeix una matriu per als resultats del Congrés i una altra per al Senat. En total, doncs, 24 anàlisis.
- II.- Determinació de l'índex de semblança i de les seves cotes. L'origen dels eixos resultats de l'ACPN d'un Districte coincideix amb el seu centre de gravetat, és a dir, constitueix el perfil del seu resultat electoral. Per tant la proximitat a aquest centre és l'índex que busquem -- per a valorar la semblança entre cada -- Secció del Districte i el Districte mateix.

S'han de fer les següents anotacions:

- a) La inèrcia captada pels primers factors és força important. Vàrem considerar que

els fenòmens representats pels factors -- més enllà del quart no aportaven més que soroll de fons de cara al comportament -- global. En realitat l'ACPN es va fer sobre freqüències i el primer factor resultant no es va tenir en compte, ja que és un factor grandària. (Vegeu fig. 3). Remarquem com les variables estan situades aproximadament en la mateixa coordenada d'aquest factor. En tots els Districtes passa igual. Recordem que l'ACPN sols -- normalitza les variables i no pas les observacions. En posteriors anàlisis s'ha deixat de banda aquesta dificultat, utilitzant com a dades, no pas les freqüències sinó els tants per cent sobre el -- nombre de vots en cada Secció -- el que en diem "perfil de la Secció". S'ha de remarcar, encara, que les seccions tenen -- grandàries molt diferents, o, si més no, prou diferents per a ser detectats per -- l'ACPN com a factor principal. A banda -- aquest factor de grandària, els altres -- factors més utilitzats es poden descriure de la següent manera:

2^{on}: oposició esquerra - dreta

3^{er}: importància de les abstencions

Les inèrcies són semblants en tots els -- Districtes i les situacions relatives de les diferents candidatures varien, encara que no pas substancialment.

- b) La primera extracció de la mostra (abans del Desembre de 1.978) sols va tenir en compte el comportament al llarg del temps per tal de detectar la seva estabilitat, i això s'aconsegueix esbrinant a cada -- consulta la proximitat de la Secció al centre de gravetat del Districte, tenint en compte els primers factors de l'ACPN -- corresponent.
- c) L'índex efectivament utilitzat és la coordenada de l'observació sobre els primers factors. Degut a la jerarquia dels factors, es va desestimar el càlcul de la distància al centre de gravetat en l'hiperplà -- determinat per aquests factors.

D'aquesta manera, a més, es pot realitzar un filtre total a partir de la projecció sobre el primer eix significatiu (que recull l'efecte esquerra-dreta fonamentalment).

d) El valor de l'índex utilitzat com a cota, depèn del nombre de consultes que es tenen en compte, ja que es considera que la grandària de la mostra és fixa (10%). És a dir, si només es disposa d'una consulta, pot ser més petita (en el nostre cas era un límit de 0.09 per al segon i el tercer factor); en la segona mostra, es disposarà de dos resultats i, per tant, de dues anàlisis. La informació conjunta (producte) serà molt més rica, encara que es relaxi la cota superior (en el nostre cas a 0.29). Existeixen diferències entre els Districtes, i quan són més estables en la votació existeixen més Seccions en cotes més baixes; evidentment, del conjunt de Seccions que compleixen el criteri, es seleccionen aquelles que el compleixen millor. És a dir, posseeixen combinacions de coordenades més baixes.

Una consideració addicional. La coordenada d'una observació sobre un factor, mesura la seva correlació. D'aquesta forma, doncs, es poden valorar la significació de les cotes utilitzades. En qualsevol cas, cal fer notar la irrelevança de les cotes (0.09, 0.29) utilitzades. A tall d'exemple, vegeu a la Taula 2 les Seccions de la mostra del Districte 3 i les seves coordenades sobre els eixos 2 i 3 el 15 de Juny.

3. LA QUALITAT DE LES ESTIMACIONS

Un cop coneguts els resultats finals, podem valorar l'apropament de les estimacions. Aquesta valoració es realitza utilitzant la potencialitat de la desviació standard i a un altre nivell utilitzant l'ACPN.

0.799E 00	-----	-----	-----	-----	-----	-----
0.779E 00	I	I	NU-V	*	I	*
0.761E 00	I	I		*	I	*
0.742E 00	I	I		*	I	*
0.724E 00	I	I		*	I	*
0.706E 00	I	I		*	I	*
0.687E 00	I	I	AB-R	*	I	*
0.669E 00	I	I		*	I	*
0.651E 00	I	I		*	I	*
0.632E 00	I	I		*	I	*
0.614E 00	I	I		*	I	*
0.595E 00	I	I		*	I	*
0.577E 00	I	I		*	I	*
0.559E 00	I	I		*	I	*
0.540E 00	I	I		*	I	*
0.522E 00	I	I		*	I	*
0.503E 00	I	I		*	I	*
0.485E 00	I	I		*	I	*
0.467E 00	I	I		*	I	*
0.448E 00	I	I		*	I	*
0.430E 00	I	I		*	I	*
0.412E 00	I	I		*	I	*
0.393E 00	I	I		*	I	*
0.375E 00	I	I		*	I	*
0.356E 00	I	I		*	I	*
0.338E 00	I	I		*	I	*
0.320E 00	I	I		*	I	*
0.301E 00	I	I		*	I	*
0.283E 00	I	I		*	I	*
0.265E 00	I	I		*	I	*
0.246E 00	I	I		*	I	*
0.228E 00	I	I		*	I	*
0.209E 00	I	I		*	I	*
0.191E 00	I	I		*	I	*
0.173E 00	I	I		*	I	*
0.154E 00	I	I		*	I	NU-V
0.136E 00	I	I		*	I	*
0.117E 00	I	I		*	I	*
0.91E-01	I	I		*	I	*
0.807E-01	I	I		*	I	*
0.624E-01	I	I		*	I	DL-V
0.440E-01	I	I		*	I	*
0.256E-01	I	I		*	I	*
0.721E-02	I	I		*	I	*
-0.112E-01	I	I		*	I	*
-0.296E-01	I	I		*	I	*
-0.479E-01	I	I		*	I	*
-0.663E-01	I	I		*	I	*
-0.847E-01	I	I		*	I	*
-0.103E 00	I	I		*	I	*
-0.121E 00	I	I		*	I	*
-0.140E 00	I	I		*	I	*
-0.158E 00	I	I		*	I	*
-0.177E 00	I	I		*	I	*
-0.195E 00	I	I		*	I	*
-0.213E 00	SI-V	I		*	I	*
-0.962E 00						
		-0.507E 00		-0.521E-01		0.403E 00
ABSCISSES	EIX 1	VALORS EXTREMS	-0.96220	0.92290	UNITAT ESCALA	0.06500
ORDENADES	EIX 2	VALORS EXTREMS	-0.23174	0.79757	UNITAT ESCALA	0.01030

Fig. 3.

ACPN-6 Dec Referendum Per Cent Sobre Vots Sec Cen Dist 3

Taula 2.
Coordenades seccions mostra-1 dist. 3

	F1	F2
8	0,05	0,03
13	0,08	0,12
50	0,04	0,03
54	0,07	0,01
59	0,004	0,10
69	0,03	0,10
70	0,05	0,07
72	0,05	0,15

3.1 Aproximació estimació resultats

Amb un cop d'ull a les taules 1 i 3 podem veure la forta aproximació aconseguida. Amb termes quantitius, podem veure també que no hi han estimacions que s'apartin en més de $\pm 0.6\sigma$ (vegeu línies ES1). No pas tots els Districtes ni totes les variables són igualment ben estimades. Es poden fer les següents remarques:

- En general, els Districtes 3, 11, 6 i 1 són els més mal estimats. En tots ells es dona el fenomen de gran variabilitat (vegeu desviacions) i importància del vot NO i BLANC. Aquesta primera pista serà més tard confirmada: la forta proximitat entre resultats i estimacions es deu també a la concentració que genera l'alt tant per cent del SÍ.
- En general també, les desviacions de la mostra són més petites que les del seu univers. Això confirma la hipòtesi en el sentit que l'univers de la mostra és més homogeni. Intervé però, també l'efecte del seu més petit nombre d'observacions. Les excepcions a aquesta regla general són també significatives, i amb el mateix sentit que dèiem en el paràgraf anterior. Vegeu les ABST del districte 3, del 5 i del 2. Aquesta variable presenta també dificultats per a ser estimada.

3.2 Una mesura no vàlida

Cal remarcar que el mètode de selecció emprat no vol substituir tot l'univers de partida. És a dir, sols busca substituir el centre de gravetat d'aquest univers; amb unes altres

paraules, tracta de realitzar l'estimació -- del perfil de votació en una demarcació en -- tants per cent. Tot això s'ha de dir per tal de posar en guàrdia contra utilitzacions no vàlides de les estimacions.

Amb els mètodes de selecció aleatoris es busca la substitució de l'univers de partida -- per un univers restringit que posseeixi les mateixes característiques. Una forma de controlar aquestes semblances consisteix en el fet de comparar, per exemple, les ACPN de -- l'univers i de la mostra. L'estructura de la informació continguda en ambdós "corpus" ha d'ésser semblant perquè la qualitat de la -- mostra sigui suficient. Noteu que precisament la semblança de perfils és la hipòtesi bàsica per tal de poder estimar diferents fenòmens utilitzant la mateixa mostra. Aquest no és pas el nostre cas.

En el mètode que descrivim en aquest article s'esbiaixa voluntàriament la selecció. S'eliminen sistemàticament tots els casos extrems i, fins i tot, aquelles observacions que formen normalment el paquet de presència més freqüent. Per altra banda, els criteris de selecció no inclouen -sota cap punt de vista explícit ni implícit- la posició relativa -- dels elements de la mostra sobre els factors. Vegem-ho:

Els criteris utilitzats tenen a veure amb el valor absolut de la coordenada sobre els factors. Per tant, no ens fixem:

- ni en el signe de tal coordenada
- ni en la taxa de representativitat de cada element de la mostra sobre els seus semblants, que serien aquells que li són pròxims, o bé que formen amb ell un "tipus"⁵.

Hem de concloure, doncs, que el perfil de la nostra mostra no té perquè assemblar-se al perfil de l'univers de referència.

A la pràctica se'ns apareixen casos de semblança i de divergència. A l'engròs podem dir que els casos de semblança posseeixen una d'aquestes característiques:

- informació resumida, per exemple, centres de gravetat de conjunts de seccions.
- demarcacions (Districtes) amb més estabilitat o homogeneïtat de vot, per exemple, --

Taula 3.
Resultats, estimacions amb la mostra-1 i amb la mostra-2,
i les seves desviacions per a cada districte

	SI	NO	BLANC	NULS	ABSTN		SI	NO	BLANC	NULS	ABSTN
1 RES	89,4	5,9	4,0	0,7	39,3	7	90,7	4,6	4,0	0,5	33,3
σ	3,2	2,6	0,8	0,3	6,1		1,8	1,3	0,8	0,7	3,4
ES1	89,09	5,88	4,18	0,85	37,71		91,34	4,18	3,87	0,59	33,82
σ	1,3	1,04	0,61	0,23	2,77		1,85	0,86	0,95	0,63	3,25
ES2	89,4	5,81	4,01	0,78	38,6		91,06	4,92	3,82	0,19	33,52
σ	0,8	0,9	0,40	0,25	1,83		0,46	0,65	0,34	0,15	1,81
2 RES	90,9	4,6	3,7	0,5	35,7	8	88,8	6,2	4,2	0,6	35,6
σ	2,3	1,7	0,9	0,8	6,8		3,1	2,6	0,8	0,7	4,9
ES1	91,89	3,26	3,76	1,08	33,53		89,45	5,89	4,07	0,57	34,06
σ	1,31	0,68	0,75	1,60	7,08		2,50	1,90	1,10	0,88	3,19
ES2	90,92	4,59	3,93	0,54	35,42		89,08	5,73	4,65	0,52	39,12
σ	0,33	0,92	0,70	0,42	5,5		0,65	0,66	0,49	0,42	9,20
3 RES	82,0	12,3	5,1	0,5	39,1	9	91,3	4,3	3,7	0,5	32,8
σ	5,6	4,9	1,4	0,6	5,9		2,0	1,3	0,8	0,7	6,2
ES1	80,41	12,66	6,04	0,88	40,58		91,59	3,97	3,62	0,43	30,53
σ	3,46	3,88	1,38	0,95	9,66		1,06	0,70	0,65	0,67	5,62
ES2	82,27	12,50	4,94	0,27	38,88		91,53	4,09	3,75	0,62	32,44
σ	0,83	1,20	0,48	0,32	2,32		0,40	0,57	0,58	0,65	3,12
4 RES	86,4	8,2	4,7	0,5	38,3	10	90,5	4,8	4,0	0,6	29,8
σ	4,0	3,3	1,2	0,7	4,3		1,9	1,5	0,7	0,9	4,6
ES1	86,76	7,96	4,73	0,54	35,42		91,07	4,18	4,26	0,47	30,89
σ	1,77	1,59	0,84	0,45	2,46		1,36	1,01	1,01	0,49	3,9
ES2	87,45	7,38	4,27	0,88	35,40		90,58	4,89	3,95	0,56	29,32
σ	1,53	1,2	0,69	1,03	3,46		0,29	0,39	0,25	0,36	3,37
5 RES	90,8	4,5	4,0	0,5	44,3	11	84,1	10,7	4,6	0,4	35,7
σ	2,1	1,4	1,1	0,7	6,8		5,4	4,3	1,3	0,6	2,8
ES1	91,46	4,08	3,93	0,56	44,50		81,94	12,92	4,94	0,18	37,33
σ	1,36	0,90	1,24	0,66	7,70		3,56	2,70	1,15	0,25	2,03
ES2	91,04	4,73	3,85	0,36	44,32		84,61	10,43	4,68	0,27	36,00
σ	0,50	1,26	0,59	0,27	4,68		0,78	1,13	0,50	0,19	3,28
6 RES	85,3	8,1	4,3	0,5	36,4	12	90,8	4,6	3,9	0,49	38,4
σ	11,4	2,9	1,0	0,8	6,1		2,0	1,5	0,8	0,6	5,3
ES1	87,11	8,01	4,03	0,83	35,95		89,89	5,22	4,08	0,78	38,52
σ	3,09	1,89	0,7	1,14	1,9		2,15	1,95	0,93	0,79	4,97
ES2	87,51	7,33	4,35	0,79	35,13		90,78	4,60	4,02	0,58	39,23
σ	1,6	0,86	0,69	1,03	1,43		0,45	0,70	0,62	0,54	3,46

els districtes 9 i 10.

Aquestes consideracions no invaliden pas les afirmacions fetes més amunt; tan sols les volen matisar. És a dir, en el cas d'existència de situacions extraordinàriament favorables, pot ésser, que fins i tot els perfils de la mostra i de l'univers s'assemblin, encara que no hi ha esperança que això passi.

Només amb ànim exemplificador veiem dos casos. En primer lloc la comparació d'ACPN a nivell Districte (figures 4 i 5). En ambdós hi han 12 observacions, una per Districte, i les mateixes variables. En el primer (A), -- les dades són els tants per cent resultants definitius, i en el segon (B), els tants per cent estimats amb la mostra.

Les inèrcies respectives són semblants (en %):

	F1	F2
A	60	19
B	61	21

Els dos primers eixos són iguals en A i B. - El primer oposa les respostes SÍ a les respostes NO i BLANC. Aquest és l'efecte més important al interpretar els resultats d'aquest Referèndum. El segon, també realitza la mateixa classificació i ordenació de les variables en els dos anàlisis. Per tant, respecta les variables, i tenint en compte, la inèrcia acumulada, podem dir que la mostra també reflecteix l'estructura de l'univers. Les diferències provenen de la situació de cada Districte. Sobre el primer factor fonamentalment, tots queden ben classificats. Sobre el segon però, sols podem dir que estan en el mateix lloc: 5, 6, 7, 9 i 10. Els altres no queden en el mateix lloc en comparar les ACPN. Fins i tot, malgrat la mateixa estructura de les variables, el perfil relatiu de

cada Districte respecte a la informació global no és igual. És a dir, encara que un resultat i la seva estimació siguin iguals, o molt semblants, pot ser que respecte a la informació global -totes les demarcacions- ocupin llocs diferents. Això significa, amb altres paraules, que el perfil de la informació global no és el mateix. Veiem ara un altre cas. En la figura 6 es grafien els dos primers eixos de l'ACPN en què les observacions són les 1.058 seccions censals de la Ciutat (C) (Degut al seu elevat nombre preferim no grafiar-les. En la figura 7, mostrem l'ACPN de totes les Seccions censals (108) que componen la mostra (D). Les inèrcies respectives en tant per cent són:

	F1	F2
C	42	20
D	48	22

Ara sols ens interessen les situacions de --

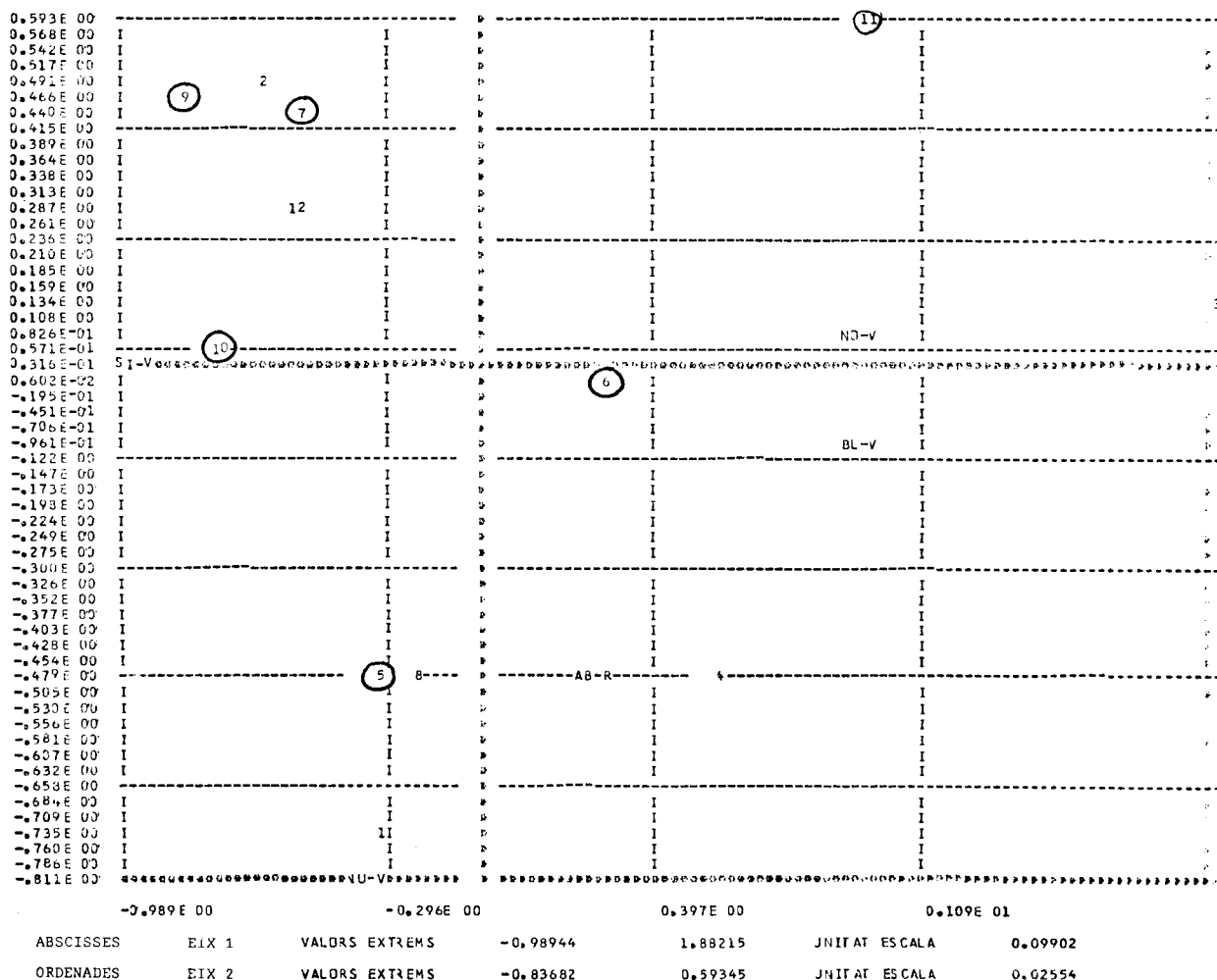


Fig. 4.

ACPN-6 Dec Referendum Per Cent Sobre Vots Districtes

les variables. El primer factor és el mateix en les dues anàlisis. Val a dir que en totes les anàlisis realitzades, aquest primer factor apareix constantment. Es pot notar, a nivell interpretatiu, que la proximitat entre les respostes BLANC i NO es deu més aviat a l'oposició al SÍ, més que no pas a la seva - concomitància. El segon factor de "C" és difícil de trobar-lo tan clar en qualsevol altra anàlisi. Evidentment no té res a veure - amb el segon factor de D. Una possible interpretació d'aquest factor consisteix a valorar l'efecte del gran nombre d'observacions (1.058). Les abstencions, que dins de cada - Districte tenen una variació més aviat erràtica però petita, en analitzar tot Barcelona, els contrastos es fan més grans i per tant - la variable capta més inèrica del núvol. Tor- - nant al nostre fil, doncs, el segon factor - no és el mateix. El tercer i quart factors - (no grafats degut a l'espai) són prou sem- - blants en ambdues anàlisis.

Com a resum d'aquest paràgraf podríem dir -- que:

- La qualitat global de les estimacions quantitatives és força alta.
- Les variables més ben estimades a nivell - quantitatiu i qualitatiu són, per aquest - ordre, SI, NO i BLANC.
- La forta polarització sobre el SI ajuda -- força a minimitzar les desviacions⁶.
- Una prova realitzada amb una mostra aleatòria dóna uns resultats més dolents.

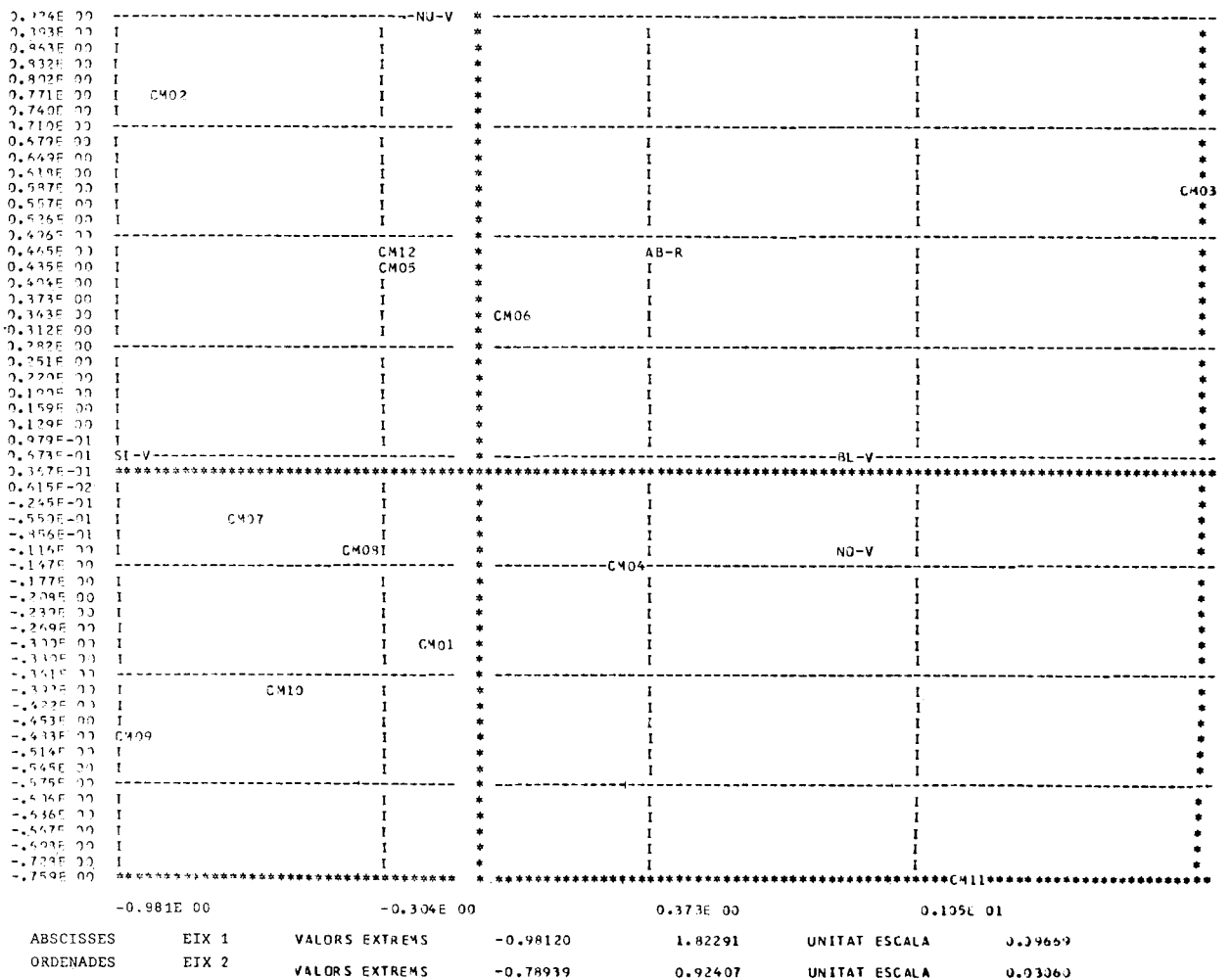


Fig. 5. ACPN-6 Dec Referendum Per Cent Centre Grav Per Distr Secc Mostra 1

4. SELECCIÓ D'UNA NOVA MOSTRA

4.1 Informació disponible

Una vegada realitzat l'escrutini del Referèndum, es disposava dels resultats efectius⁷. Es va efectuar la mateixa bateria d'anàlisis per cada Districte, per tot Barcelona a nivell de Secció Censal i de Districte, i les Seccions de la mostra, és a dir, un total de 15 anàlisis. Un dels objectius és aconseguir la nova mostra a utilitzar en les Eleccions Generals de l'1 de Març de 1.979. A aquesta nova mostra en direm Mostra-2.

Aplicant la manera de fer, descrita més amunt, s'han seleccionat aquelles seccions que compleixin els criteris esmentats. Els ACPN que s'han utilitzat tenen com a dades d'entrada els percentatges que en cada secció representen les candidatures al Congrés (15 de Juny)

i les respostes al Referèndum (6 de Desembre). Com que l'estimació es vol per Districtes -- s'utilitza un ACPN per a cada un d'ells.

S'han seleccionat aquelles seccions en ambdues anàlisis (Congrés, Referèndum) no ultra passen la cota de 0.29 en la coordenada sobre el primer factor. Per eliminar les sobbrants s'utilitza el criteri adicional d'escollir aquelles que tenen una coordenada sobre el segon factor (dels dos anàlisis) més petita.

En cas de semblança entre Seccions es considera més favorable per entrar en la mostra -- una coordenada més baixa en el ACPN dels resultats del Congrés (vegeu taula 4).

La grandària de la mostra l'hem considerat -- fixa, pels volts del 10%. Amb les desviacions conegudes dels nostres universos i amb el mètode significativament esbiaixat que hem uti

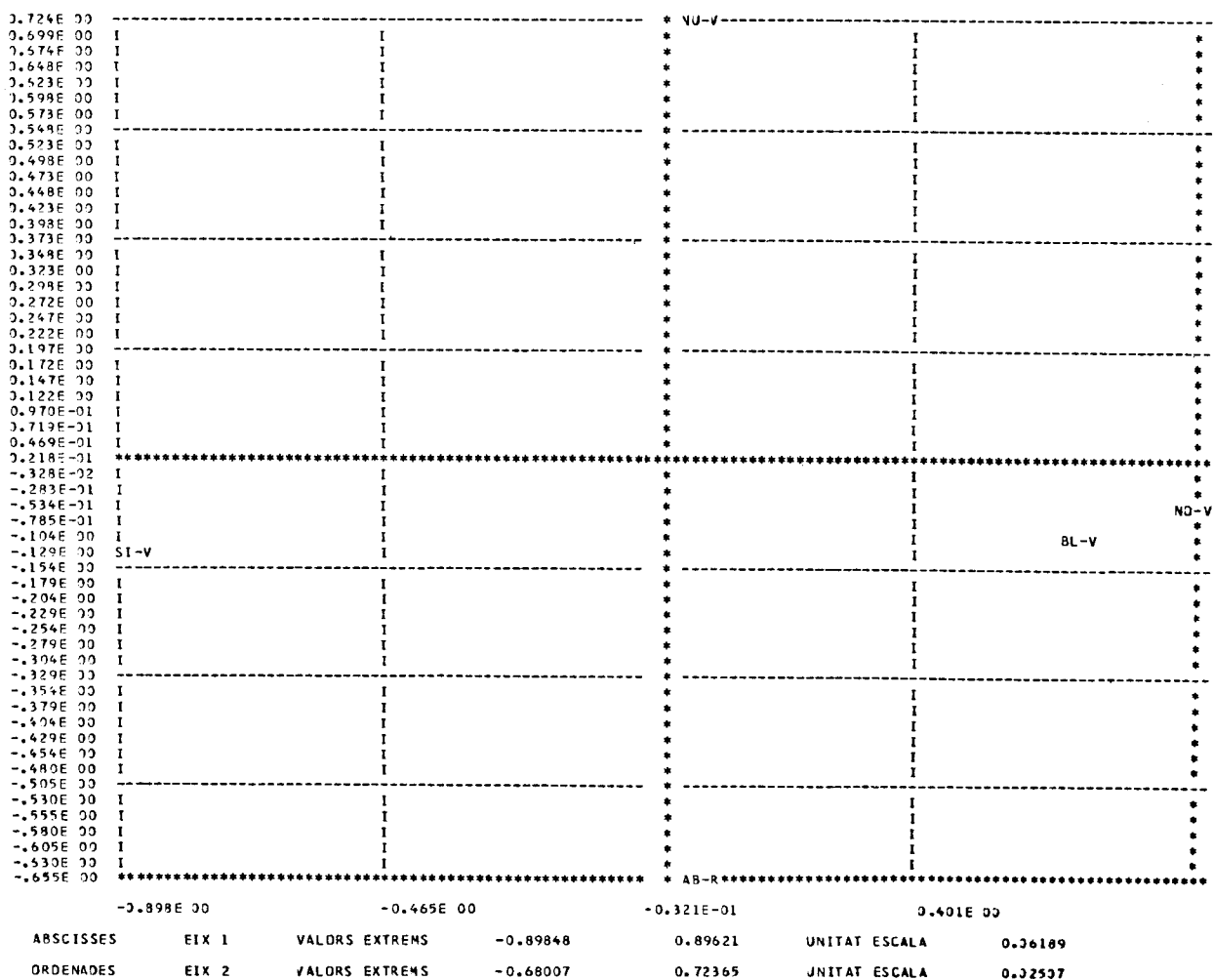


Fig. 6.

ACPN-6 Dec Referendum Per Cent Sobre Vots Totes Sec Cen de Ciutat

litzat, la grandària de la mostra és important. La capacitat organitzativa, però, i la desproporció entre resultats de les diferents candidatures, han aconsellat no reduir el nombre de seccions en la mostra (108).

4.2 Comparació de mostres

Una vegada seleccionada la nova mostra, hem volgut copsar el seu rendiment, i la millora que suposa respecte a la Mostra-1. A continuació descriurem aquest guany. Les dades més importants es poden consultar en les taules 3 i 4.

A la taula 4 es donen les distribucions que presenta cada mostra (M1, M2), en cada Districte, segons les coordenades sobre els dos

primers factors en el corresponent ACPN de cada Districte. Per exemple, en el Districte 2, hi han dues seccions per a la Mostra-2, i 1 secció per la Mostra-1, que poseeixen una coordenada sobre el primer factor (F1) entre 0.1 i 0.19, i sobre el segon factor (F2) entre 0.0 i 0.09.

El tret comú en la comparació per a cada districte és la menor dispersió de M2 i la concentració en els valors més petits. Recordem ara la qualitat de la Mostra-1 i la seva relativa concentració de freqüències en els valors més alts de F1. Per tant, és d'esperar una més gran qualitat en les estimacions realitzades amb M2. Això no és pas una tautologia -en funció del mètode de selecció- sinó, al contrari, la contrastació de la hipòtesi que la proximitat al centre de gravetat supo

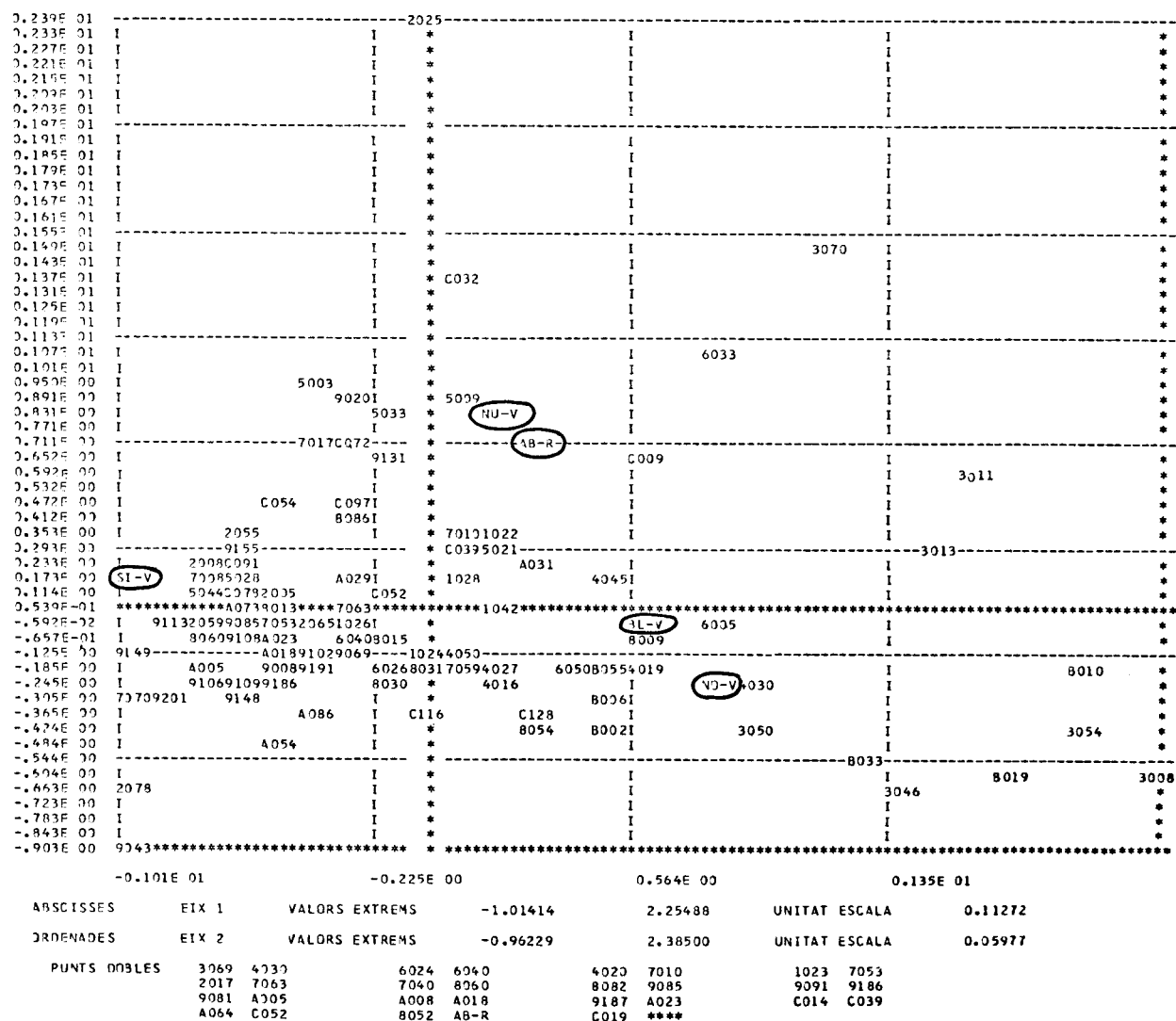


Fig. 7.

ACPN-6 Dec Referendum Per Cent Sobre Vots Sec Cen de la Mostra 1

QÜESTIÓ - v.3, n.º 1 (març 1979)

Taula 4.
Classificació de les Seccions d'Ambdues
mostres segons les coordenades sobre 2 Factors

F1	0 - 0,09			0,1 - 0,19			0,2 - 0,29			0,3 -		
	0-0,09	0,1-0,19	0,2-0,29	0,3-	0-0,09	0,1-0,19	0,2-0,29	0,3-	0-0,09	0,1-0,19	0,2-0,29	0,3-
M1	1				1				1			
M2	1		1		1						1	
M1		1								3		1
M2	1	1		3								
M1			2						1			2
M2	2	3	1	2								
M1	1		1		1				1			
M2			1		2					2		
M1				3								1
M2												4
M1		1			1							
M2	3	1			1							
M1		1										
M2			1		2				1			
M1		1										
M2	2	1	1						2			2
M1		1										
M2	2	1	3		1					4		2
M1	1	3	1									
M2	1	2	3		3				4	4		1
M1												
M2	3				2				4	2		1
M1		1										
M2	1	1	1							2		1
M1		1										
M2	1		1									
M1				5								
M2	2	3							2	1	3	4

sa una substituïbilitat més gran.

En la taula 3 es poden veure les estimacions i desviacions de M1 i M2. En gran majoria -- dels casos la desviació es redueix extraordinàriament, fins i tot en el cas de les Abstencions. El cas del Districte 1, que resulta ara el més difícil d'estimar, es deu a la forta polarització i petit nombre de Seccions. Existeixen dos comportaments molt diferents i identificables en el ACPN corresponent. La variabilitat de les respostes és -- molt gran, sobretot en la Abstenció. De tota manera, es pot concloure que la M2 posseeix molta més capacitat predictiva que la M1. -- Els Districtes 11 i 3 hi guanyen fonamentalment, pel fet de recollir la M2 directament un fenomen que no havia aparegut explícitament el 15 de Juny: resposta Blanc.

La comparació de ACPN tal com hem fet abans ens dóna els següents resultats:

- L'ACPN de totes les seccions de la Mostra-2 és sensiblement igual al de la Fig. 7. -- La diferència més important és l'allunyament entre les respostes Abstenció i Nul -- sobre el segon factor.
- L'ACPN dels centres de gravetat de les seccions de la Mostra-2 es diferencia del representat en la Fig. 5. D'aquestes diferències, les més importants són:

- . Tots els districtes queden ben classificats respecte a l'ACPN de resultats dels districtes, exceptuant el Districte 1.
- . El segon factor és el que hem esmentat -- en parlar de l'ACPN de totes les seccions de la Ciutat.

Podem concloure doncs que, tant des d'un -- punt de vista quantitatiu, com qualitatiu, -- la capacitat estimativa de la Mostra-2 és -- més gran.

5. SOBRE L'ARGUMENTACIÓ

En acabar, sembla interessant recordar un -- punt crucial de tota la nostra argumentació. Aquesta és la hipòtesi sobre l'estabilitat -- del comportament electoral. Si no existeixen seccions que tinguin un comportament estable,

i a més, proper al centre de gravetat de la seva demarcació, tota la nostra problemàtica no té sentit. Es a dir, no hi ha informació per aprofitar, i cal un mètode aleatori.

Sembla però, que les següents consideracions s'han de tenir en compte:

- La població electoral que compon les seccions no varia radicalment.
- Amb el temps, les opinions polítiques sembla que s'aniran sedimentant, i això afavoreix també aquest mètode.
- Amb el temps també, cada nova consulta serà més consistent quant a informació incorporada, ja que es disposarà de més informació.
- Es pot arribar a formular una hipòtesi prèvia amb el següent contingut: el canvi de comportament electoral es més improbable -- com més radical és. Com a corol·lari, podem afirmar, doncs, que la proximitat al -- centre de gravetat, en qualsevol consulta, és un criteri que permet esperar les estimacions més ajustades. De moment, la pràctica no invalida pas el nostre plantejament, tot al contrari.

6. CONCLUSIONS

De vegades hom disposa d'informació que pot ser útil per a seleccionar una mostra, aleshores resulta interessant esbiaixar el mètode de selecció aprofitant aquella informació. En el nostre cas, aquesta informació és el -- comportament electoral en anterior consultes de les seccions censals de la Ciutat.

En funció del comportament estable de les -- seccions i seleccionant aquelles que poseixen un comportament més semblant al de la seva zona o demarcació, s'ha establert el mètode de selecció. (La hipòtesi d'estabilitat s'ha anat confirmant en successives consultes electorals).

La forma emprada de trobar les seccions més semblants al conjunt del Districte aprofita els resultats de l'ACPN.

El rendiment de la mostra així seleccionada,

d'acord amb els resultats obtinguts relatius al recompte a Barcelona del Referèndum Constitucional del 1978, és molt alt, tant per l'aproximació als resultats globals, com per la seva petita desviació.

7. BIBLIOGRAFIA

- /1/ "Recompte de resultats del Referèndum -- Constitucional al municipi de Barcelona". NOVATICA, Novembre-Desembre 1978, pp. 52-54.
- /2/ LEBART, L., MORINEAU, A., i TABARD, N. - "Techniques de la description statistique". Dunod, Paris, 1977.
- /3/ CANALS, J.M. "Técnicas de Análisis de Datos Multidimensionales". Tesi de Llicenciatura. Universitat Autònoma de Barcelona. Facultat d'Econòmiques. 1977.

8. APÈNDIX: L'APLICACIÓ DE L'ANÀLISI EN COMPONENTS PRINCIPALS NORMADES

L'Anàlisi en Components Principals Normades és una tècnica de la bateria de l'Anàlisi -- Multidimensional de Dades. Consisteix a trobar uns eixos o factors que -per ordre: F1, F2, F3, ...- expliquen de més a menys la -- inèrcia del núvol de punts. S'utilitza com a índex de semblança o distància al coeficient de correlació linial. Per tant, si dos punts- variable sobre el gràfic de les distribucions de sengles seccions (vegeu figures 3 a 7 amb eixos resultats de l'anàlisi) estan a prop, vol dir que les distribucions s'assemblen, -- és a dir, quan en una de les seccions puja -- una variable, a l'altre també; i a l'inrevés, quan p.e. el SI és gran en una secció, aleshores el NO és petit.

La propietat que aprofitem en aquest treball és que el centre de la nova base, es a dir -- l'origen de coordenades dels eixos o factors correspon exactament al centre de gravetat -- del núvol --en el nostre cas als resultats en tant per cent del Districte. Per aquesta raó s'escullen aquelles seccions que estan a -- prop del origen o centre de gravetat del seu Districte.

9. NOTES

- ¹La Guàrdia Urbana es va encarregar de recollir els documents. El Centre de Procés de Dades i el Subdepartament d'Estadística van dissenyar el sistema i van tractar la informació.
- ²Encara que pel mimetisme imperant en l'aplicació de l'Estadística sigui pràcticament -- la norma no adequar el mètode a la informació disponible sobre el fenomen.
- ³Els programes utilitzats són els del Subdepartament d'Estadística de l'Ajuntament de Barcelona.
- ⁴Aquests resultats varen ésser oferts per -- l'Equip de Sociologia Electoral de la Universitat Autònoma (Facultat de Dret) de Barcelona, el qual va realitzar un important -- treball de depuració.
- ⁵Aquesta consideració obre un camí nou de selecció de mostres estratificades. Es basa -- en la elaboració prèvia d'una tipologia. En el nostre cas, es podia realitzar un mètode consistent a seleccionar de cada tipus un -- nombre proporcional al cardinal del tipus, i precisament aquells que estan més a prop del centre de gravetat.
- ⁶Les Seccions tenen una coordenada alta -- en l'ACPN del seu Districte- difícilment superen la cota de ± 0.756 al resultat mitjà del Districte.
- ⁷El Subdepartament d'Estadística disposa del fons documental dels resultats per Secció -- de les consultes fetes des del 15 de Juny -- de 1977. Al mateix temps són consultables la bateria de mapes i ACPN realitzats. Es -- pensa incorporar la mateixa informació per cada nova consulta.