



## ANÁLISIS DE DATOS ELECTORALES HISTÓRICOS

**Lila Ricci<sup>1</sup>, Marcela Ferrari<sup>2</sup>, Marcela Natal<sup>1</sup>, Pamela Marcote<sup>3</sup>**

<sup>1</sup> *Depto. de Matemática, Facultad de Ciencias Exactas y Naturales, Universidad Nacional de Mar del Plata. [lricci@mdp.edu.ar](mailto:lricci@mdp.edu.ar)*

<sup>2</sup> *Depto. de Historia, Facultad de Humanidades, UNMDP*

<sup>3</sup> *Bibliotecología, Facultad de Humanidades, UNMDP*

### RESUMEN

Para explorar y analizar la percepción de los adultos mayores, sobre el proceso político y social argentino entre los años 1945 y 2002, se analizaron 130 entrevistas. Las mismas fueron tomadas a personas mayores de 75 años, combinando técnicas de ciencias sociales y de la historia oral. En particular, se detalla el tratamiento de los datos hecho en función del siguiente objetivo: correlacionar variables socio-económicas con opción de voto para establecer correspondencias entre extracción social y voto partidario. Primero fue necesario representar las respuestas a preguntas referidas a características socioeconómicas en un grupo reducido de variables numéricas. Aplicando Análisis Factorial de Correspondencias seguido de Clasificaciones Jerárquicas se logró definir grupos de individuos, que quedaron caracterizados por ciertas modalidades. Se analizó

la relación de cada nueva variable con las opciones de voto, para descartar aquellas que no asociadas. Se resumieron las variables socioeconómicas retenidas, en un número reducido de factores principales aplicando el método de Análisis de Correspondencias Múltiples. Las opciones de voto fueron agrupadas en un número reducido de modalidades para cada una de las votaciones presidenciales efectuadas entre los años 1946 y 1999. Se representaron estas votaciones, consideradas como variables ilustrativas, en ejes definidos por los principales factores.

**PALABRAS CLAVES:** votaciones - análisis de correspondencias múltiples - clasificación jerárquica.

### 1. INTRODUCCIÓN

Este trabajo forma parte de un proyecto que procura explorar y analizar la percepción de los adultos mayores sobre el proceso político y social argentino entre los años 1945 y 2002. Es de destacar que la historia política argentina de la segunda mitad del siglo XX se ha incorporado como objeto de estudio al campo de la historia en los últimos veinte años. Con respecto a los actores políticos, la atención privilegiada de las ciencias sociales se ha concertado en los sindicatos, los partidos políticos y el ejército. Poco se sabe en cambio de la percepción de los ciudadanos

comunes, aquellos que no tienen acceso a construir un discurso público, con respecto a ese pasado. Partiendo de esta premisa, se planteó como objetivo general del proyecto observar de qué manera la gente común, “los sin voz”, vivieron el período 1945-2002. Se aplicaron técnicas de la historia oral, que consisten en recoger testimonios de quienes han vivido los acontecimientos. Estos testimonios fueron recolectados en forma de encuestas, posibilitando así el tratamiento estadístico de datos provenientes de una investigación histórica.

Uno de los objetivos particulares del proyecto consistió en explorar las correspondencias entre la extracción social de los encuestados y sus opciones de voto en las elecciones presidenciales de ese período. En este artículo se detalla el tratamiento estadístico realizado para aportar a ese fin.

Además de su posible influencia sobre la opción de voto, cabe esperar que las características socio-económicas estén interrelacionadas entre sí, esto sugirió que la información contenida podía ser representada a través de un número reducido de factores, aplicando un análisis factorial de correspondencias múltiples (AFCM). Este método tiene la ventaja de proporcionar representaciones gráficas de variables ilustrativas (en este caso las opciones de voto) en los planos óptimos, estos gráficos facilitan la visualización de la información contenida en los datos. Existen varios antecedentes de la representación de votaciones usando esos métodos, pueden verse por ejemplo el trabajo de Salvo (2004) en el que se aplican técnicas multivariadas a datos electorales en España y el de Chiche (2000) que trabajó sobre el electorado francés.

El presente artículo está organizado en cinco secciones: en la Sección 2 se detalla la obtención de ejes factoriales a partir de las variables socio-económicas; en la Sección 3 se describe la clasificación jerárquica de los encuestados de acuerdo a sus proximidades en el espacio factorial y la partición de la muestra en clases homogéneas; la Sección 4, integrando los resultados anteriores, exhibe la representación en los planos factoriales de las opciones de voto, conjuntamente con los centros de las clases; finalmente la Sección 5 contiene algunas conclusiones.

## **2. OBTENCIÓN DE LOS PLANOS FACTORIALES**

### **2.1. Las encuestas**

Se trabajó con 130 encuestas administradas a personas que estaban habilitadas para votar a partir de 1946, en el caso de los varones y de 1952 en el de las mujeres. El cuestionario utilizado recabó entre otros datos, información sobre antecedentes familiares e historia personal y memoria de las votaciones durante el período 1946-2001. Fue imposible, dados nuestros recursos, construir una muestra aleatoria. La desactualización de los padrones, el rechazo de las personas ubicadas por ese medio a acceder a la consulta por parte de desconocidos en un contexto de inse-

guridad social y la negativa de muchos individuos a responder al saber que debían emitir opiniones políticas, nos llevó a construir el universo de entrevistados siguiendo el criterio conocido como “bola de nieve”. De manera que no es posible –ni es nuestra intención– generalizar los resultados obtenidos al conjunto de la sociedad.

## 2.2. Asociación entre opciones de voto y variables socio-económicas

Se consideraron para cada votación las siguientes opciones, que son las que aparecieron con frecuencias mayores que cinco:

- 1951: *peronista, radical, socialista* y "no participó"
- 1954: *peronista, radical* y "no participó"
- 1958: *UCRI, UCRP, izquierda* y "no participó"
- 1963: *UCRI, UCRP, izquierda* y "no participó"
- 1973, marzo: *FREJULI, UCR, izquierda* y "no participó"
- 1973, septiembre: *FREJULI, UCR, izquierda* y "no participó"
- 1983: *justicialista, UCR, izquierda* y "no participó"
- 1989: *justicialista, UCR, izquierda* y "no participó"
- 1995: *justicialista, UCR, izquierda* y "no participó"
- 1999: *justicialista, Alianza, izquierda* y "no participó"

En cuanto a las preguntas referidas a la situación socio-económica de los encuestados, una primera selección de las que serían tomadas en cuenta para el análisis factorial se basó en la asociación que presentaba cada una de ellas con las opciones de voto. Para medir estas asociaciones se calculó la significación correspondiente a la prueba de independencia de chi-cuadrado y –siguiendo el criterio de Mickey y Greenland (1989)– fueron retenidas aquellas variables socio-económicas que presentaban, para al menos una votación, un *valor-p* menor que 0.25. Los resultados pueden verse en la Tabla 1.

**Tabla 1.** Significación de las asociaciones entre variables socio-económicas y votaciones.

Variables socio-económicas	Votaciones				
	1951	1954	1958	1963	III/1973
Sexo	0.136	0.002	0.011	0.178	0.005
Calificación laboral	0.011	0.019	0.018	0.113	0.110
Estado civil	0.611	0.998	0.673	0.513	0.575
Posesiones a los 10 años	0.345	0.672	0.852	0.253	0.996
Posesiones al casarse	0.280	0.003	0.231	0.578	0.594
Nivel de educación	0.007	0.607	0.399	0.114	0.179
Participación en clubes	0.815	0.764	0.729	0.657	0.804
Participación en iglesias	0.005	0.289	0.275	0.242	0.336
Participación en sindicatos	0.008	0.000	0.459	0.474	0.000
Participación en sociedades de fomento	0.590	0.071	0.456	0.032	0.283
Participación en bibliotecas	0.011	0.042	0.011	0.077	0.119
Participación en cooperadoras	0.657	0.071	0.432	0.976	0.536

**Tabla 1 (continuación).** Significación de las asociaciones entre variables socio-económicas y votaciones.

Variables socio-económicas	Votaciones				
	IX/1973	1983	1989	1995	1999
Sexo	0.070	0.059	0.104	0.230	0.120
Calificación laboral	0.186	0.126	0.408	0.759	0.274
Estado civil	0.210	0.911	0.770	0.281	0.888
Posesiones a los 10 años	0.630	0.678	0.815	0.722	0.970
Posesiones al casarse	0.287	0.399	0.225	0.094	0.359
Nivel de educación	0.586	0.535	0.216	0.878	0.613
Participación en clubes	0.909	0.659	0.717	0.435	0.157
Participación en iglesias	0.268	0.286	0.060	0.225	0.313
Participación en sindicatos	0.173	0.004	0.041	0.017	0.158
Participación en sociedades de fomento	0.774	0.308	0.208	0.571	0.555
Participación en bibliotecas	0.214	0.000	0.032	0.001	0.000
Participación en cooperadoras	0.519	0.506	0.505	0.201	0.795

Luego de agrupar aquellas categorías con frecuencias demasiado bajas, las variables socio-económicas seleccionadas quedaron constituidas así:

- sexo: *mujer; varón*
- calificación laboral: *profesional, técnica/operativa, no calificado, ama de casa*
- estado civil: *casado, soltero*
- posesiones a los 10 años: *con bienes(10), sin bienes(10)*
- posesiones al casarse: *con bienes, sin bienes*
- participación en clubes: *sí, no*
- participación en iglesias: *sí, no*
- participación en sindicatos: *sí, no*
- participación en sociedades de fomento: *sí, no*
- participación en bibliotecas: *sí, no*
- participación en cooperadoras: *sí, no*

- educación: *primaria incompleta, primaria, secundaria, universitaria*

### 2.3. Factores principales

Para obtener los ejes factoriales se partió de una matriz  $\mathbf{X}$  de 130 individuos caracterizados por 12 variables activas, que son las variables socio-económicas descritas en la Sección 2.1. Esta matriz se expandió a otra matriz  $\mathbf{Z}$ , formada por unos y ceros de acuerdo a que el individuo pertenezca o no a cada modalidad. Resumiendo se tiene:

$$\begin{aligned} X &= (x_j^i) \text{ matriz de datos; } 1 \leq i \leq 130; 1 \leq j \leq 12 \\ m(j) &: \text{modalidades de la variable } j \\ Z &= (z_j^{im}) \text{ matriz expandida binaria,} \\ z_j^{im} &= \begin{cases} 1 & \text{si } x_j^i = m \\ 0 & \text{en caso contrario} \end{cases} \end{aligned}$$

Como es sabido (véase Escofier, 1998), el problema consiste en representar simultáneamente en un mismo espacio los individuos y las modalidades de forma tal que la inercia (varianza) entre las modalidades sea máxima. Es decir que se busca una solución para la expresión:

$$\max_Y \frac{\frac{1}{12} \sum_{j=1}^{12} \sum_{m=1}^{m(j)} p_{j,m} \|\mathbf{m}_{j,m} - \mathbf{g}\|^2}{\|\mathbf{Y}^t \mathbf{Y}\|}$$

donde  $\mathbf{Y}$  es la representación de los individuos en la nueva base,  $p_{j,m}$  es la frecuencia o peso de la modalidad  $m$  en la variable  $j$ ,  $\mathbf{m}_{j,m}$  es el centro de gravedad de dicha modalidad y  $\mathbf{g}$  es el centro de gravedad del conjunto formado por los ciento treinta individuos. La matriz  $\mathbf{Y}$  debe ser tal que la media de cada columna sea cero y que sus columnas sean ortogonales entre sí.

El problema planteado equivale a encontrar una base ortonormal  $\{\mathbf{u}_k\}_{k=1}^q$  que maximice la expresión

$$\frac{\sum_{k=1}^q \sigma_k^2 \mathbf{u}_k^t \mathbf{A} \mathbf{u}_k}{\sqrt{\sum_{k=1}^q \sigma_k^4}}$$

donde  $\sigma_k^2$  es la varianza de  $y_k$

y la matriz  $\mathbf{A}$  es:

$$A = \frac{1}{12} \sum_{j=1}^{12} A_j; A_j = z_j \begin{bmatrix} p_{j1} & 0 & \dots & 0 \\ 0 & p_{j2} & \dots & 0 \\ & & \ddots & \\ 0 & 0 & & p_{j=m(i)} \end{bmatrix}^{-1} z_j^t$$

La solución óptima está dada por los autovectores de la matriz A; por su parte los autovalores representan el porcentaje de inercia explicada por cada eje. Las coordenadas de una modalidad en la nueva base indican la asociación entre esa modalidad y el factor.

Las Figuras 1, 2 y 3 ilustran el resultado del AFCM aplicado a las variables socio-económicas descritas en la sección anterior. Cada modalidad está representada en los planos factoriales por el punto medio de todos los individuos que tienen ese atributo; la proximidad entre modalidades se interpreta entonces como mayor coincidencia entre los individuos aludidos, es decir que hay muchos encuestados que presentan ambas características simultáneamente.

**Figura 1.** Plano definido por los factores 1 y 2.

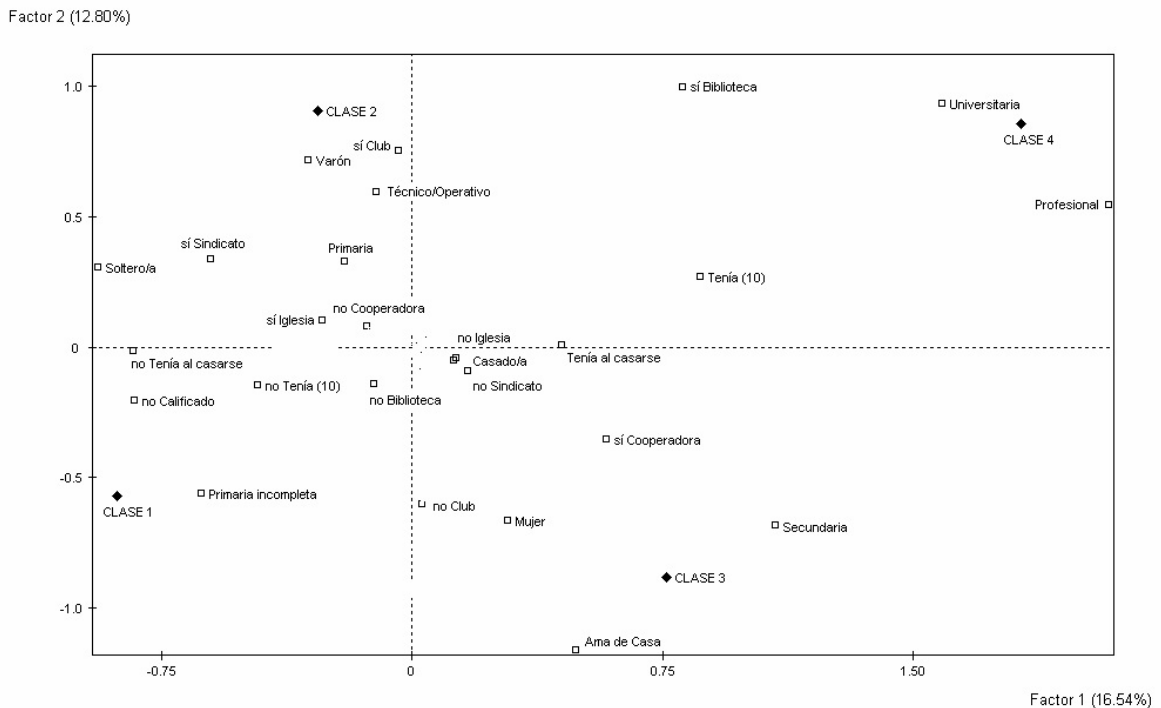


Figura 2. Plano definido por los factores 1 y 3.

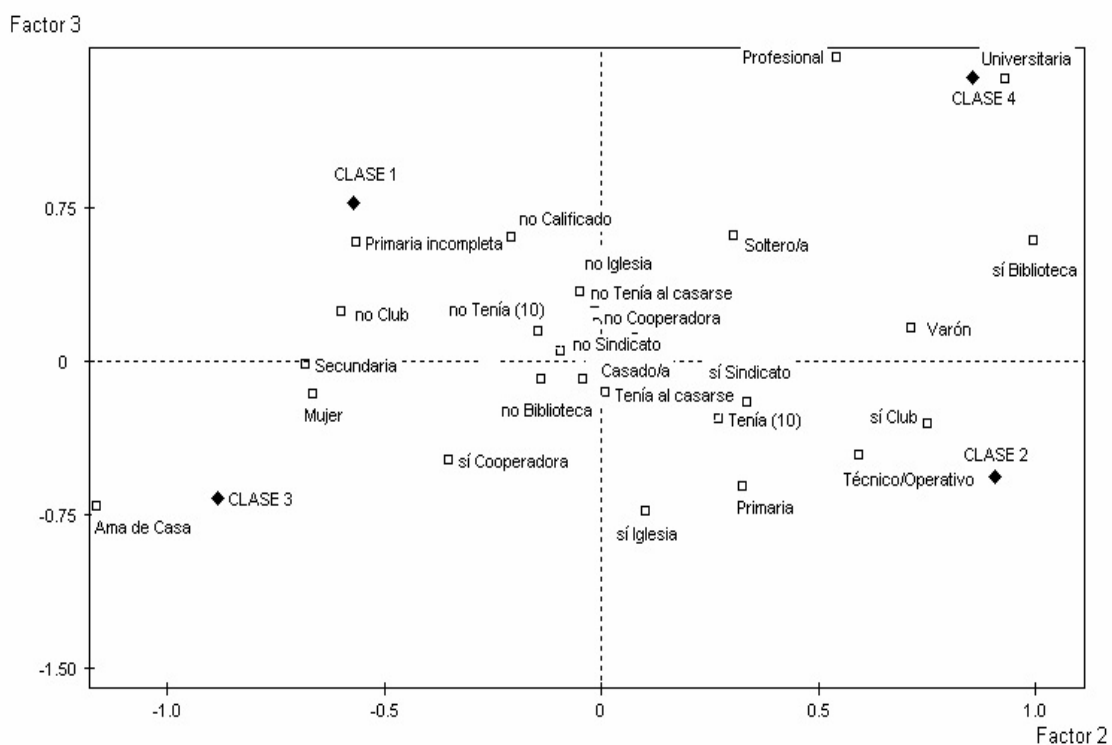


Figura 3. Plano definido por los factores 2 y 3

La interpretación del significado de estos factores se basó en el aporte de las modalidades más influyentes. Las coordenadas de cada modalidad en la nueva base dan una medida de su contribución, para que sean comparables se trabajó con el llamado "valor-test" que es la coordenada de la modalidad en ese factor, convenientemente estandarizada para que tenga media cero y varianza unitaria. El cálculo efectuado fue el siguiente:

$\bar{y}_q^k$  : k-sima coordenada del centro de la modalidad q

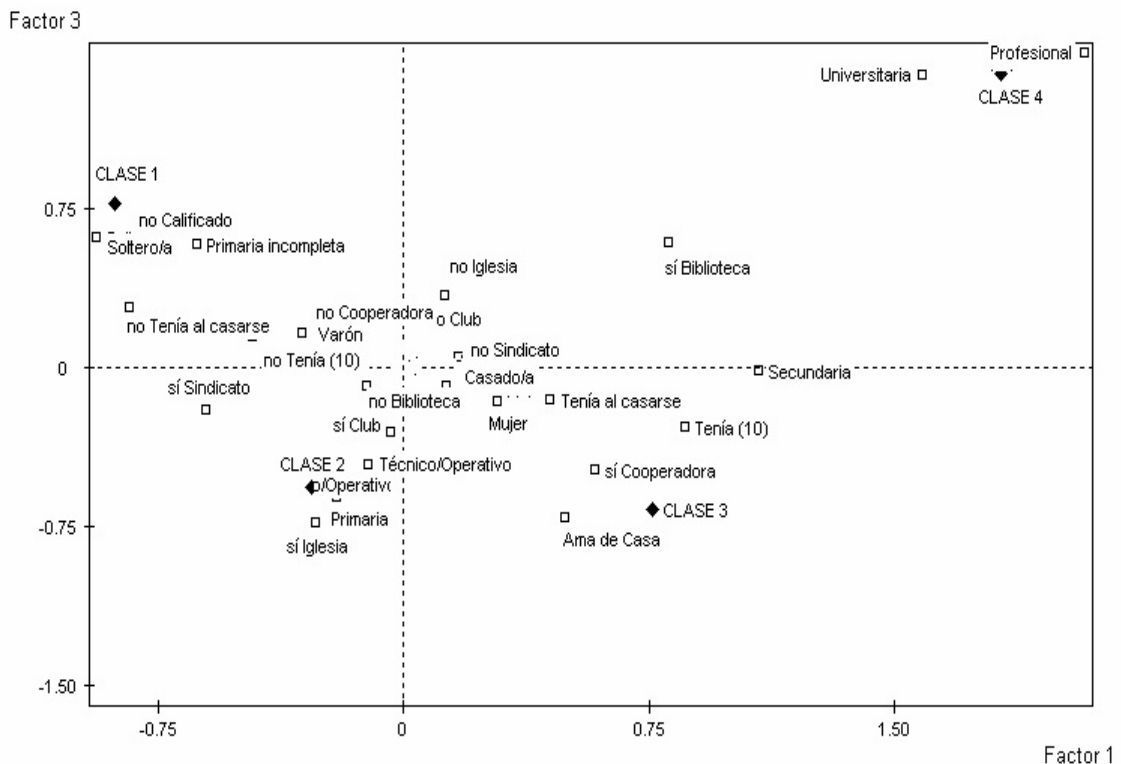
$$s^2 = \frac{\sum_{i=1}^{130} (y_i^k)^2}{130} : \text{varianza total}$$

$$s_{\bar{y}_q}^2 = \frac{s^2}{p_q} \times \frac{130 - p_q}{129} : \text{varianza de la modalidad q}$$

$$\frac{\bar{y}_q^k}{s_{\bar{y}_q}} : \text{valor-test}$$

En las Tablas 2 a 4 se detallan los valores-test de cada modalidad en los primeros tres factores y sus pesos, definidos como el número de encuestados que presentaron la modalidad correspondiente.

**Tabla 2:** Interpretación del FACTOR 1.





Valores negativos			Valores positivos		
modalidad	coord. est.	peso	modalidad	coord. est.	peso
Sin bienes	-6.18	37	Profesional	6.81	10
Sin bienes (10)	-6.11	69	Con bienes	6.18	68
No calificado	-5.41	32	Con bienes (10)	6.11	36

**Tabla 3:** Interpretación del FACTOR 2.

Valores negativos			Valores positivos		
modalidad	coord. est.	peso	modalidad	coord. est.	peso
No club	-6.94	59	Sí club	6.94	46
Mujer	-6.87	54	Varón	6.87	51
ama de casa	-5.58	21	Técnico/operativo	4.83	42

**Tabla 4:** Interpretación del FACTOR 3.

Valores negativos			Valores positivos		
modalidad	coord. est.	peso	modalidad	coord. est.	peso
No iglesia	-5.52	71	primaria	5.72	6
Profesional	-4.73	10	Sí iglesia	5.25	34
Universitaria	-4.60	10	Ama de casa	3.89	21

Vistos estos resultados, una posible interpretación de los tres primeros factores, es la siguiente:

- FACTOR 1 opone a personas sin posesiones o con baja calificación situadas en un extremo con personas con alto nivel de educación y/o con posesiones en el otro extremo
- FACTOR 2 opone a mujeres amas de casa con varones de calificación intermedia y/o participación en clubes
- FACTOR 3 opone a encuestados con calificación laboral y nivel de educación altos y/o que no concurren a la iglesia por un lado, con amas de casa, personas con educación primaria y quienes sí participaron en las iglesias por otro lado.

### 3. CLASIFICACIÓN DE LOS ENCUESTADOS

#### 3.1. Partición del conjunto

Una vez definidos los planos factoriales se procedió a hacer una clasificación jerárquica de los encuestados, que puso en evidencia la posibilidad de definir una partición de la muestra en cuatro subconjuntos. Se aplicó el proceso de clasifica-

ción jerárquica conocido como Algoritmo de Ward (véase Escofier, 1998) que parte de los individuos aislados y los va uniendo de acuerdo a sus proximidades, hasta que todos constituyan un único conjunto. Para medir proximidades se usó la distancia euclídea entre la representación factorial de dos individuos  $i_1$  e  $i_2$  definida como:

$$d(i_1, i_2) = \|y_{i_1} - y_{i_2}\|$$

Una vez formados los grupos, la distancia entre dos de ellos,  $g_1$  y  $g_2$ , es

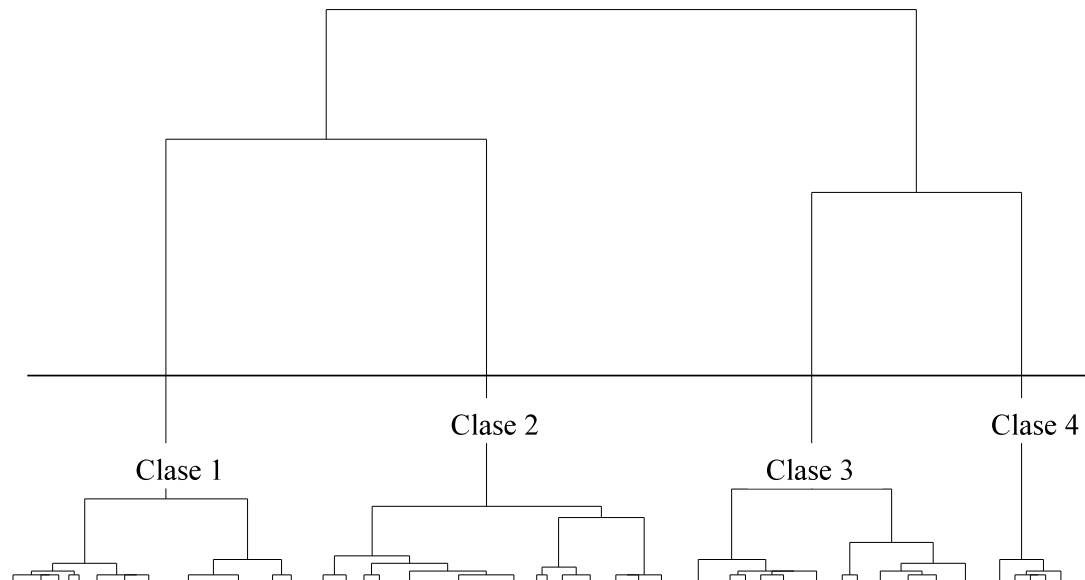
$$\delta(g_1, g_2) = \frac{p_1 p_2}{p_1 + p_2} \|\bar{y}_1 - \bar{y}_2\|$$

$\bar{y}_i$  : centro de gravedad del grupo

$p_i$  ; número de individuos en el grupo  $i$

El resultado de la clasificación se ilustra en la Figura 4, como puede apreciarse en ese esquema hay cuatro clases bien diferenciadas. Así, de acuerdo a sus atributos –modalidad de respuesta a las preguntas seleccionadas– los encuestados fueron agrupados en subconjuntos de personas con características homogéneas. Cabe aclarar que no se trata de clases sociales, sino de conjuntos de encuestados que coincidieron en la mayoría de sus respuestas a las preguntas formuladas.

Figura 4. Clasificación jerárquica de los individuos encuestados



Para poder caracterizar las clases que se formaron, se analizó por un lado la prevalencia de respuestas de determinado tipo o modalidades, entre los individuos

pertenecientes a esa clase (mod/cla) y por el otro lado el grado de inclusión de la clase en la modalidad (cla/mod). Los resultados pueden verse en las Tablas 4 a 7.

**Tabla 4.** Caracterización de la CLASE 1.

Modalidad	cla/mod	mod/cla	peso
Primaria inc	74.19	76.67	31
Sin bienes(10)	43.48	100.00	69
No club	47.46	93.33	59
No calificado	59.38	63.33	32
Sin bienes	54.05	66.67	37

**Tabla 5.** Caracterización de la CLASE 2.

Modalidad	cla/mod	mod/cla	peso
Sí club	69.57	84.21	46
Varón	64.71	86.84	51
Primaria	60.78	81.58	51
Técnico/operativo	59.52	65.79	42
Sí Soc. de fomento	100.00	15.79	6

**Tabla 6.** Caracterización de la CLASE 3.

Modalidad	cla/mod	mod/cla	peso
Mujer	51.85	96.55	54
Ama de casa	76.19	55.17	21
Con bienes	42.65	100.00	68
Secundario	69.23	31.03	13
Con bienes(10)	47.22	58.62	36

**Tabla 7.** Caracterización de la CLASE 4.

Modalidad	cla/mod	mod/cla	peso
Profesionales	80	100	10
Universitarios	70	87.5	10

Puede decirse que la CLASE 1 está compuesta mayoritariamente por personas con pocas posesiones y no calificadas, en la CLASE 2 predominan los varones con una calificación intermedia que participan en clubes y en sociedades de fomento, en la CLASE 3 predominan las amas de casa y personas con bienes, tanto a los 10 años como al casarse y con formación secundaria mientras que, en la CLASE 4 prevalecen individuos profesionales con formación universitaria completa.

En las Figuras 1 a 3 pueden verse las cuatro clases proyectadas en los tres primeros planos factoriales. Cada clase está representada por la media de todos los individuos que la componen, es decir por su centro de gravedad.

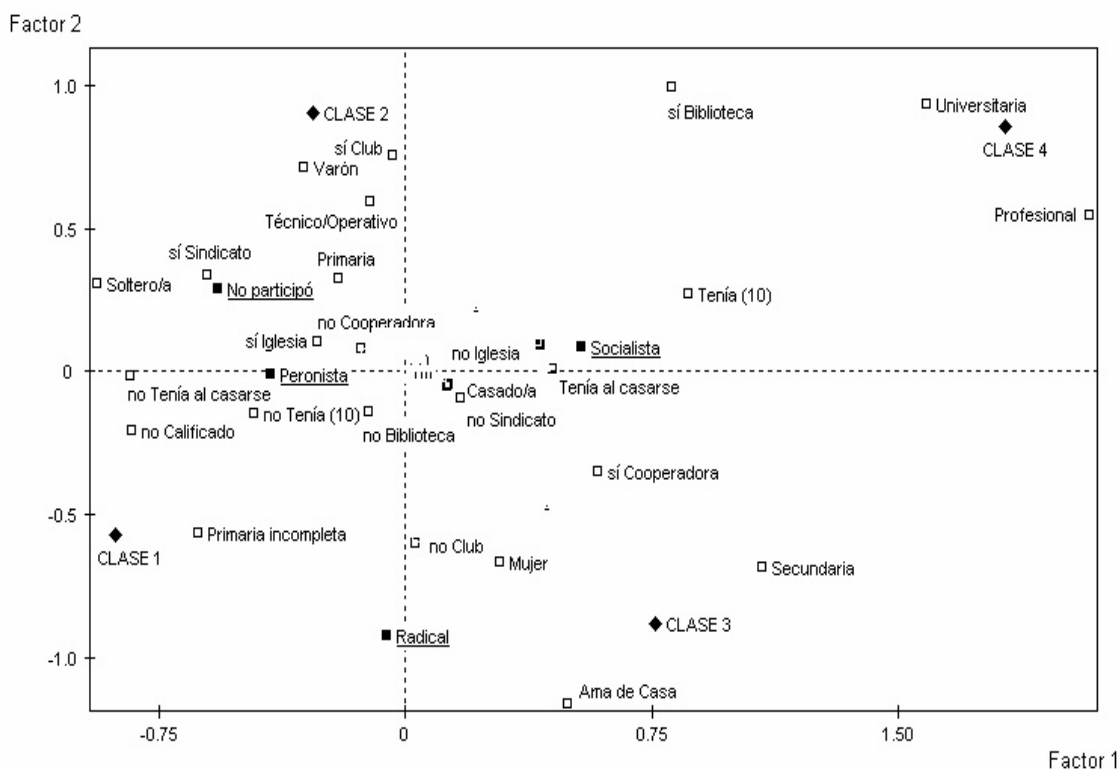
#### 4. REPRESENTACIÓN DE LAS VOTACIONES EN LOS PLANOS FACTORIALES

Finalmente se representó cada una de las votaciones en el plano formado por los dos primeros factores principales. A través de estos gráficos puede apreciarse la asociación entre las distintas opciones de voto, las modalidades de respuesta y los subconjuntos en los que se agruparon los individuos encuestados. A modo de ejemplo, en las Figuras 5 a 7, se exhiben los resultados correspondientes a las votaciones de los años 1958, septiembre de 1973 y 1983.

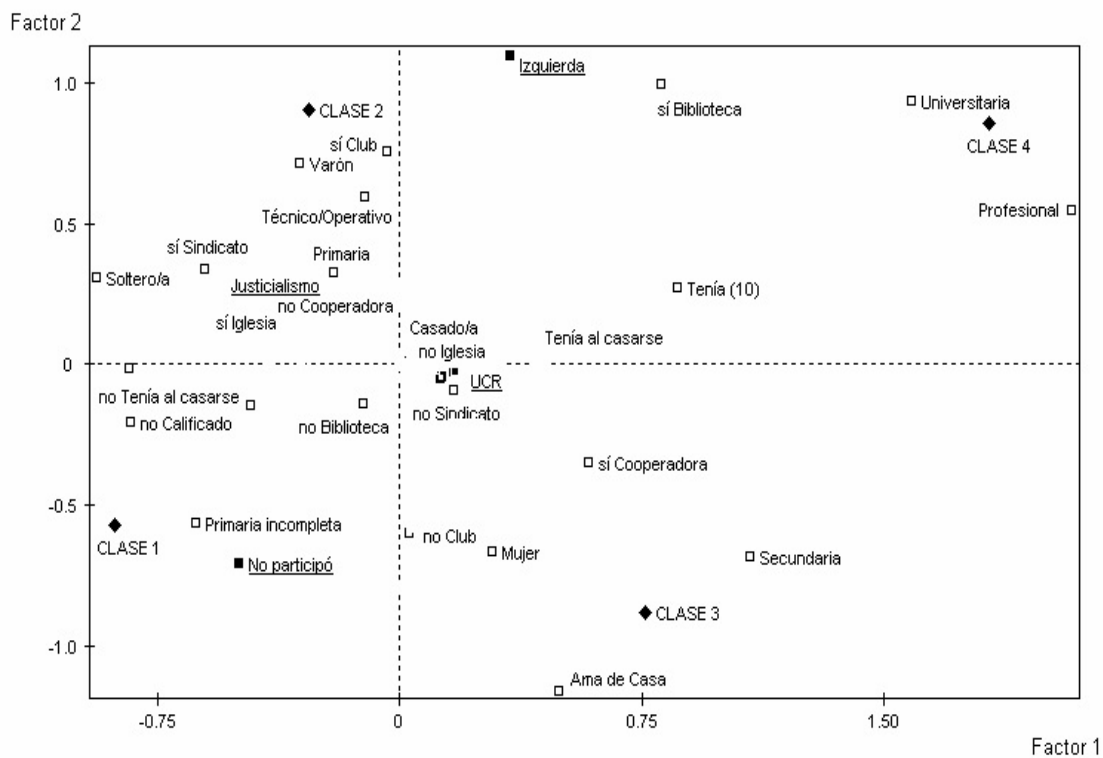
**Figura 5.** Votación de 1958.



**Figura 6.** Votación de septiembre de 1973.



**Figura 7.** Votación de 1983.



Con respecto a la evolución de las votaciones, a partir de la observación de estas figuras puede señalarse que:

- El voto peronista aparece en el semiplano de las Clases 1 y 2, salvo en la elección de septiembre de 1973 cuando Perón fue el candidato
- El voto de la izquierda aparece próximo a las Clases 2 y 4, acercándose más a la Clase 2 a partir de 1983
- El voto radical es el más cercano a la Clase 3

## 5. DISCUSIÓN

La interpretación de resultados provenientes de un análisis estadístico multivariado suele ser engorrosa para investigadores no familiarizados con la estadística, como es el caso de los historiadores. El AFCM complementa los resultados numéricos con representaciones gráficas en espacios óptimos en el sentido de mostrar la mejor "cara" –la más informativa– de los datos. En este caso la representación simultánea de las clases en que se agruparon los encuestados y de las opciones de voto, en los planos definidos por las variables socio-económicas resultó una síntesis muy informativa, comprensible y de fácil interpretación del contenido de las encuestas. Resultó así una herramienta facilitadora del diálogo interdisciplinario.

## REFERENCIAS

- Escofier, B. y Pagés, J. (1998) *Analyses Factorielles Simples et Multiples*, ed. Dunod
- Lebart, L.; Morineau, A. y Piron, M. (1998) *Statistique Exploratoire Multidimensionnelle*, ed. Dunod
- Chiche, J.; Le Roux, B.; Perrineau, P. y Rouanet, H. (2000) L'espace politique des électeurs français a la fin des années 1990. *Revue Française de Sciences Politiques*, 50, 463-487.
- Mickey, J. y Greenland, S. (1989) A study of the impact of confounder-selection criteria on effect estimation, *American Journal of Epidemiology*, 129, 125-137.
- Salvo, S.; Galindo, P. y Martin, J. (2004) An application of nonsymmetric correspondence analysis based on TUCKALS algorithm to electoral marketig data, *Actas del VI Coloquio Latinoamericano de Sociedades de Estadística*, Concepción, Chile.