



Ventajas y Limitaciones de la Metodología de Inferencia Ecológica propuesta por G. King. Aplicaciones al análisis del triunfo de Perón en las elecciones presidenciales argentinas de 1946

Javier Balsa (jjbalsa@isis.unlp.edu.ar) UNLP-CONICET (Argentina)

Abstract

The ecological inference tries to estimate the individual behavior from aggregate data. In this article, King's new methodology for dealing with ecological inference will be applied to estimated the proportion of workers who voted for Perón in 1946. Data of the national census of 1947 about the Federal District and Buenos Aires Province are used. In a second section, I point out some limitations of King's methodology when data distribution does not help to make acceptable estimations. In the last section, I show one of the greatest advantages of this methodology. Since it provides estimations not only for the whole region but also for each of the districts which conform the region, it is possible to study the covariations between the estimated variable and other explanatory variables (in this example, the growing population rates of the districts).

Resumen

La inferencia ecológica intenta estimar la conducta individual desde datos agregados. En este artículo, la nueva metodología de King para tratar con la inferencia ecológica se aplicará para estimar la proporción de obreros que votaron por Perón en 1946. Se usarán los datos del censo nacional de 1947 sobre el Distrito Federal y la Provincia de Buenos Aires. En una segunda sección, señalo algunas limitaciones de la metodología de King cuando la distribución de los datos no ayuda para hacer estimaciones aceptables. En la última sección, muestro una de las más grandes ventajas de esta metodología. Desde que no sólo mantiene estimaciones sobre la región entera sino también para cada uno de los distritos que conforman la región, es posible estudiar las covariaciones entre la variable estimada y otras variables explicativas (en este ejemplo, las proporciones de la población crecientes de los distritos).

Introducción

En las Ciencias Sociales, muchas veces nos encontramos con el problema de no contar con los datos de las unidades de análisis que nos interesa analizar, sino tan sólo con la información agregada de las características que presentan los agrupamientos geográficos de tales unidades. En estos casos, no se cuenta con el cruce de las variables de interés, y sólo se tienen las distribuciones de cada una de las variables (los marginales de las tablas de contingencia) para las diferentes zonas.

La causa de esta falta de información suele ser de dos tipos. En algunos casos, se han perdido los datos registrados originalmente a nivel individual (por ejemplo, las fichas censales), entonces solamente se conservan los datos editados sobre la cantidad de casos, en cada unidad geográfica, según algunos valores o intervalos escogidos de



cada variable (o, más bien, tan sólo de las variables que finalmente se han decidido editar). En el segundo tipo de situaciones, el dato individual ha sido producido secretamente (como las conductas electorales) y nunca pudo, ni podrá, conocerse, por lo cual únicamente contamos con la información agregada por, en el mejor de los casos, las mesas electorales.

Si el objetivo de la investigación es de tipo descriptivo, habitualmente alcanza con esta información agregada, pues nos permite conocer las distribuciones univariadas. Sin embargo, si lo que interesa es analizar las relaciones entre dos o más variables (por ejemplo, entre los distintos sectores sociales y las conductas electorales) deberemos introducirnos en el problema de la inferencia ecológica: tratar de derivar las conductas individuales a partir de los datos agrupados. Partiendo de los marginales de una tabla de contingencia se intenta estimar los valores de las distribuciones condicionales (las celdas interiores de dicha tabla). Lamentablemente, esta inferencia ha presentado habitualmente graves dificultades para ser realizada, y, por lo tanto, una serie de cuestiones no han encontrado aún solución por la imposibilidad de obtener estimaciones de los datos individuales (desconocidos) a partir de los datos colectivos (1).

Recientemente se ha propuesto una nueva metodología para abordar la inferencia ecológica (King, 1997) y en esta ponencia intentaremos ejemplificar su uso, a fin de llamar la atención de los científicos sociales sobre sus potencialidades, aunque también señalaremos algunas limitaciones que se presentan en su aplicación (2). Para desarrollar estos objetivos trabajaremos a partir de un ejemplo principal, y algunos ejemplos menores, todos centrados en la misma cuestión. En principio, si conocemos la proporción de algún sector social en un conjunto de unidades geográficas dadas, y, al mismo tiempo, sabemos la proporción de votantes y no-votantes a determinada fuerza política (los ejemplos "clásicos" son los que relacionan a los obreros con el comunismo o el socialismo, para los casos europeos, y con el peronismo, para el caso argentino), podemos intentar inferir la proporción de los miembros de dicho sector social que votaron a ese partido. De contarse con esta estimación, podría analizarse, entre otras cuestiones, hasta qué punto la conducta de este sector permite explicar un determinado resultado electoral. Para ejemplificar este problema, escogimos un tema que ya ha suscitado el empleo de la inferencia ecológica con anterioridad: los comportamientos de los distintos sectores sociales en las elecciones de 1946 (3). Precisando más aun el ejemplo, proponemos relacionar las conductas de los sectores obreros con el voto a Juan Domingo Perón en 1946. Las unidades geográficas mínimas (i), en este ejemplo, serán las circunscripciones electorales de la Capital Federal, y los partidos de la Provincia de Buenos Aires. Denominaremos *distritos* a estas unidades, y *región* al total de los distritos. Los datos conocidos para cada uno de los i distritos son: (X_i) la proporción de obreros (3) (4) y desocupados (5) varones (6) mayores de 18 años (7) registrados en el Censo Nacional de 1947; ($1 - X_i$) la proporción de no-obreros varones mayores de 18 años; (T_i) la proporción de votantes a la candidatura de Perón en 1946 sobre el total de votantes (8); ($1 - T_i$) la proporción de no-votantes a Perón, y (N_i) la cantidad total de votantes en dicha elección (9). Estos datos constituyen los marginales de nuestra tabla de contingencia. Los datos que desconocemos y deseamos estimar (las distribuciones condicionales o celdas interiores) son la proporción de obreros en cada distrito que votó a Perón (β_i^o) y la proporción de no-obreros que votó a Perón (β_i^n). Obviamente, si conociéramos β_i^o , sabríamos la proporción de obreros que no votó por Perón ($1 - \beta_i^o$), y podríamos calcular también β_i^n y ($1 - \beta_i^n$).

Entonces, para cada uno de los i distritos, tendríamos la siguiente tabla de contingencia:

Sector social	Voto a Perón	No voto a Perón	
Obreros y desoc.	β_i^o	$(1 - \beta_i^o)$	X_i
No obreros	β_i^n	$(1 - \beta_i^n)$	$(1 - X_i)$
	T_i	$(1 - T_i)$	

Datos conocidos (celdas sombreadas)



Pero, seguramente, también queremos estimar las proporciones de interés a nivel regional (Capital y Buenos Aires), a las que denominamos B^o (proporción de obreros y desocupados que votaron por Perón) y B^n (proporción de no-obreros que votaron por Perón). Estos son, en general, los parámetros que han intentado estimarse en las investigaciones realizadas (10).

La evolución de las técnicas de inferencia

El problema de la inferencia ecológica prácticamente nació con la sociología, en *El Suicidio* de E. Durkheim (11). Pero quien planteó el problema con rigurosidad por primera vez fue W. Robinson (1950), con uno de los artículos de mayor repercusión en la historia de la metodología de las ciencias sociales. Su punto de vista fue que, con las técnicas disponibles en su época, era imposible realizar inferencias ecológicas. En los manuales de metodología puede observarse el impacto de este artículo. Su negativa conclusión aparece, por ejemplo, en Galtung (1966: 45-49, donde cita explícitamente el trabajo de Robinson) y en Seltz y otros (1976: 439-440). Es también bajo la denominación de "falacia ecológica" que la cuestión de la inferencia de lo colectivo a lo individual aparece en el manual de estadística para historiadores de C. Dollar y R. Jensen (1971: 97-104). Ellos acuerdan con que el uso de la correlación ecológica como una aproximación a la correlación individual no tiene ninguna justificación y sería un clásico ejemplo de falacia ecológica. Ante esta grave situación para la resolución de muchas cuestiones historiográficas recomiendan, o bien, mantenerse en el nivel ecológico (predicando sobre los agregados geográficos), o bien, analizar sólo áreas homogéneas (pero advierten que no siempre existen zonas homogéneas para todos los grupos a analizar, y que, además, cuando se encuentra a estas zonas, ellas pueden ser muy atípicas). Como última opción, proponen estimar los valores de las celdas interiores a partir de los datos agregados, pero aclarando que esto no resulta demasiado confiable (12).

A pesar del fuerte impacto de la posición crítica de W. Robinson, rápidamente habían surgido dos vías de avance sobre el problema: las propuestas de L. Goodman (1953) y de O. D. Duncan y B. Davis (1953).

Estos dos últimos autores proponían explorar los límites que, para cada una de las unidades geográficas mínimas, los datos marginales imponían a la variabilidad de los valores de las celdas interiores. Realizando los cálculos para todas las unidades mínimas, se obtienen, muchas veces, rangos de variaciones posibles mucho más acotados para las cantidades de interés en el conjunto del área en estudio. Los β_i 's correspondientes a cada distrito deben caer al interior de estos límites determinísticos, y en la práctica ellos casi siempre son más estrechos que el rango (0,1). Luego, utilizando los límites determinísticos de cada uno de los distritos y ponderando las cantidades posibles, se pueden calcular los límites de las proporciones a nivel regional. Además, estos autores han demostrado que estos ajustes resultan más precisos (rangos de variaciones posibles más acotados) cuando los datos se encuentran más desagregados, es decir, en unidades geográficas menores (ver tabla 2 en Duncan y Davis, 1953: 666). Por último, destacan que los rangos no son intervalos de confianza, sino, mejor aun, los límites absolutos de los valores y, por lo tanto, no requieren supuestos sobre el tipo de distribución de los mismos.

Sin embargo, esta técnica muchas veces no ha permitido establecer rangos acotados y, por lo tanto, útiles para las investigaciones. Para nuestro ejemplo de la Capital Federal y la provincia de Buenos Aires, los límites determinísticos resultaron de una amplitud tal que carecen de interés. La proporción regional en que los obreros votaron a Perón (B^o) oscila entre 0,1242 y 0,9592; mientras que la proporción en que lo eligieron los no obreros (B^n) lo hace entre 0,1584 y 0,9233. Es decir que ambas proporciones pueden tomar casi todos los valores posibles.

En cambio, la propuesta elaborada por L. Goodman provee, en general, de estimadores más acotados y ha tenido una difusión mayor. Su núcleo consiste en plantear que "en circunstancias muy especiales" el análisis de regresión entre variables ecológicas puede ser utilizado para hacer inferencias sobre las conductas individuales. En cada distrito se verifica la identidad básica $\{T_i = \beta_i^o X_i + \beta_i^n (1-X_i)\}$. Esta identidad (13) presenta dos parámetros desconocidos (β_i^o y β_i^n) y dos observaciones por distrito (T_i y X_i). Para solucionar esta indeterminación, Goodman propone el supuesto extremo de que las proporciones (en nuestro caso, β_i^o y β_i^n) se mantuviesen constantes en todos los distritos ($\beta_i^o = B^o$ y $\beta_i^n = B^n$, para todos los i). La nueva ecuación resulta manejable, ya que tiene sólo dos parámetros y numerosas (i) observaciones. Entonces, si se cumpliera el supuesto, la estimación por mínimos



cuadrados ordinarios de T_i en X_i daría un estimador insesgado de los parámetros B^0 y B^n . En la ecuación obtenida ($T_i = a + b X_i$), B^n será igual a la ordenada al origen (14), y B^0 a la suma de la ordenada y la pendiente (15).

En nuestro ejemplo realizado con los distritos de Capital y Buenos Aires, la regresión lineal por mínimos cuadrados presentó una ordenada de 0,4765 y una pendiente de 0,0332. Por lo tanto, con el método de Goodman, se estima que los obreros y desocupados votaron a Perón en una proporción del 0,5097; mientras que los no-obreros lo hicieron en una proporción del 0,4765. El Gráfico 1 muestra la recta de regresión estimada con esta metodología.

Boudon ha analizado críticamente la propuesta de Goodman, explicitando la teoría social que se encuentra implícita en el supuesto requerido para solucionar el problema de la indeterminación: la propensión de los obreros a votar por determinado partido, lo mismo que la propensión de los no obreros, son constantes. En particular, estas propensiones no dependen ni de la proporción de obreros, ni de la proporción de votos a esa fuerza política en una circunscripción determinada"; generalizando, "no depende del contorno del elector". Para este autor, "la teoría sociológica subyacente representa en efecto una simplificación sin duda excesiva de la realidad" (Boudon, 1963: 265-266). Sin embargo, la propuesta de Boudon se mantiene dentro de la misma técnica planteada por Goodman, introduciendo tan sólo la utilización de modelos no lineales y sugiriendo el análisis de las diferentes capacidades de dichos modelos a ajustarse a los datos (16).

Además de la crítica teórica sobre el requisito del supuesto de la constancia de los parámetros β_i^0 y β_i^n , surgen otros puntos débiles en la propuesta de Goodman: en primer lugar, el método requiere de una fuerte relación lineal (o de algún otro tipo, en las variantes de Boudon) entre las dos variables, y, en segundo término, si los parámetros varían de modo correlacionado con X_i (en nuestro ejemplo, si la proporción de los obreros a votar a Perón se hubiese incrementado a medida que en los distritos aumentase la presencia de los obreros), la regresión ordinaria no producirá buenos estimadores del promedio de estos parámetros (problema del sesgo de agregación).

Como consecuencia de estos problemas, las aplicaciones empíricas del modelo frecuentemente han dado resultados desatinados o incluso imposibles. Esto último es, además, posible ya que no tiene en cuenta la información del método de los límites de Duncan y Davis, y a menudo dan estimadores que los exceden, pronosticando que más del 100% de un grupo de individuos tuvieron una determinada conducta.

La propuesta de G. King

Ha sido a partir de la enumeración de estas dificultades que Gary King desarrolló una nueva propuesta que promete ser la solución al problema de la inferencia ecológica. La misma utiliza las dos vías anteriormente comentadas: la información determinística contenida en los límites, y un modelo estadístico que permite realizar una estimación al interior de dichos límites. Para ello, se mueve sucesivamente del nivel distrital al regional y viceversa. Primero, al tratar cada distrito por separado, el método utiliza toda la información disponible para dar un rango de *posibles* valores para estas cantidades de interés a nivel distrital (17). Luego, a fin de aproximarse a una respuesta más acotada, el modelo estadístico "toma fuerza" de todos los datos de los otros distritos para dar la *probable* ubicación de cada verdadera cantidad de interés dentro de los conocidos límites determinísticos. Finalmente, se calculan, en base a las estimaciones distritales, los parámetros regionales (18).

Aunque se basa en las técnicas anteriores, el método de King no consiste en una simple agregación de las mismas, tal como veremos a continuación, cuando desarrollemos las particularidades de los componentes determinísticos y estadísticos del método. Pero antes de entrar en sus complejidades, conviene aclarar, para evitar falsas ilusiones, que debido a que el problema está causado por la falta de información a nivel individual, ningún método producirá estimaciones precisas en todas las instancias (tal como lo adelanta King en las primeras páginas de su trabajo). Sin embargo, este método sortea varias dificultades y permite diagnosticar la pertinencia de los supuestos y la confianza en los estimadores obtenidos. De este modo, permite que se obtenga información sobre cuestiones que eran hasta ahora inabordables.



El componente determinístico

El método comienza con la información determinística sobre cada una de las cantidades de interés a nivel distrital pero destaca que los límites no representan toda la información determinística existente sobre los parámetros. En cada distrito, los β_i^o y β_i^n están relacionados linealmente: ante unos marginales dados (X_i y T_i), si la proporción inferida de obreros que votan a Perón aumenta, debe disminuir la proporción de los no-obreros que lo hace. Se determina, entonces, una línea sobre la cuál se ubica el punto (desconocido) que grafica las verdaderas e ignoradas proporciones en que obreros y no obreros votaron por Perón (19). Se procede, entonces, a calcular y a graficar, tal como puede observarse en el Gráfico 2, toda la información determinística sobre β_i^o y β_i^n para todos los distritos que conforman la región estudiada (20). Estos cómputos incrementan sustancialmente nuestro conocimiento acerca de los parámetros de interés, en la medida en que esto reduce la búsqueda de las estimaciones de los verdaderos β_i^o y β_i^n a un punto (dos coordenadas) en una simple línea y no a un punto en la totalidad de la superficie del cuadrado del gráfico.

El modelo estadístico

Llegados a este momento, debemos tratar de estimar en qué punto de las líneas determinísticas se ubican β_i^o y β_i^n . Para ello se modela la distribución conjunta de los β_i^o 's y β_i^n 's de todos los distritos de la región. Para poder estimar esta distribución se proponen dos supuestos: los β_i^o presentan una distribución normal, pero que no puede tomar valores negativos ni superiores a la unidad, y los β_i^n poseen una distribución con las mismas características. Se vinculan ambas distribuciones, relacionándolas según un determinado coeficiente de correlación. Es decir que el modelo requiere de un supuesto clave: que β_i^o y β_i^n presenten una distribución normal bivariada truncada. Los parámetros (γ) de esta distribución que interesa estimar son las medias, las desviaciones estándar y la correlación de β_i^o y β_i^n a través del conjunto de los distritos (21).

Las medias de esta distribución bivariada normal truncada pueden ser estimadas a partir del valor esperado de T_i dado X_i , con una regresión (22); las desviaciones estándar y la correlación son estimadas a partir del patrón de heterocedasticidad alrededor del valor lineal esperado, a través del método de máxima verosimilitud (23).

Una vez obtenidas las estimaciones de los parámetros (24), se procede a representar la particular distribución bivariada normal truncada, como surge de los cinco parámetros estimados, en el mismo cuadrado de β_i^o por β_i^n . El gráfico apropiado es de tres dimensiones, con ejes constituidos por β_i^o , β_i^n y, en el eje de la altura, la densidad de probabilidad (en el ejemplo analizado se obtuvo el Gráfico 3). Para poder graficar la distribución en dos dimensiones se utilizan elipses de contorno, que son la proyección bidimensional de las densidades tridimensionales, en donde el punto grafica la moda, el contorno interno delimita al 50% del volumen de la distribución y el contorno externo al 95% (ver Gráfico 4).

Luego, a pesar de que el modelo tiene sólo cinco parámetros a ser estimados (los elementos de γ), él provee (no independientemente) información sobre los dos parámetros de interés β_i^o y β_i^n para cada uno de los i distritos (en nuestro ejemplo, determinó los puntos que pueden observarse en el Gráfico 5). Podría decirse que el conocimiento sobre γ es resumido en una etapa de estimación, a partir de la cual β_i^o y β_i^n son calculados en una etapa de predicción (25). Si tomáramos un corte en el Gráfico 3 a lo largo de la línea de uno de los distritos (i), nos quedaría una distribución univariada que representaría la información prestada sobre la probabilidad de la ubicación de β_i^o y β_i^n en dicho distrito. Esta distribución univariada es la distribución posterior de las cantidades de interés en el distrito i , condicional a la información aportada por los otros distritos.

En síntesis, a fin de conocer más sobre dónde en cada línea (y entonces dónde al interior de los límites) caen los verdaderos β_i^o y β_i^n , el método saca fuerza estadísticamente del modelo de la distribución bivariada normal truncada de todo los distritos. Esto es, la distribución posterior de un distrito es exactamente la distribución univariada rebanada del contorno por su línea en los Gráficos 3 o 4 y proyectada sobre los ejes vertical y horizontal. Según el modelo, la localización de los verdaderos β_i^o y β_i^n es más posible que caiga en la porción de la línea que pasa más cerca del centro del contorno de la distribución bivariada normal truncada, estimado a partir de todas las



líneas (26). Entonces, una vez obtenida una estimación de los parámetros de esta distribución, simulamos salidas de valores de β_i^o , y de este modo estimamos la media y el desvío estándar de β_i^o para cada distrito, como así también intervalos de confianza (que surgen de los percentiles correspondientes) (27).

En el caso del ejemplo que estamos desarrollando, los valores medios estimados de los β_i^o y de los β_i^n son los que se observan en el Gráfico 5. Podemos observar que las proporciones estimadas en que los obreros votaron a Perón no son, de ningún modo, constantes (supuesto requerido por el método de regresión), sino que variarían entre los distintos distritos. La proporción de los no obreros votantes a Perón, en cambio, resultaría más estable.

Como última etapa, producimos una estimación final de los parámetros a nivel regional, a través de simulación a partir de las estimaciones de las distribuciones distritales y de su ponderación. En el ejemplo desarrollado las estimaciones de los parámetros fueron las siguientes:

$$B^o = 0.7872 \quad B^n = 0.3088$$

$$S(B^o) = 0.0264 \quad S(B^n) = 0.0250$$

Es decir, que el 78,72% de los obreros y desocupados de Capital Federal y la provincia de Buenos Aires habrían votado por Perón en 1946, mientras que un 30.88% de los no-obreros habrían tenido esta conducta electoral. Los reducidos desvíos estándar que presentan ambas estimaciones, como así también las distribuciones simuladas de B^o y B^n (la primera de las cuáles se observa en el Gráfico 6), brindan una elevada confianza en que los parámetros verdaderos muy probablemente no difieran demasiado de estas estimaciones puntuales. Es más, el programa desarrollado por King permite simular una serie de estimaciones de los parámetros regionales, de modo que es posible construir intervalos de confianza a partir de su distribución. En nuestro caso, utilizando los percentiles 10 y 90, obtenemos un 80% de confianza en que el verdadero parámetro B^o se encuentre entre 0.755 y 0.822, y un 95% de confianza para el rango 0.741 y 0.832 (28).

Algunas limitaciones del método de King

No en todos los casos el método de King funciona adecuadamente. El mismo King ha señalado tres problemas importantes que pueden obstaculizar su aplicación: supuestos incorrectos acerca de la distribución, sesgos de agregación, y dependencia espacial. El supuesto central del modelo es que β_i^o y β_i^n tienen una distribución normal truncada bivariada. De este modo, como debe haber una única moda, debería ser posible observar en la vista tomográfica un lugar donde todas las líneas se cruzasen. La mayoría de las "emisiones" deberían partir desde allí. En nuestro ejemplo es posible encontrar esta área, pero no es sencillo. Entonces, los datos no confirman nuestro supuesto, pero tampoco lo rechazan (en el "tomography plot" no se observan evidencias de la presencia de una moda adicional que haya quedado fuera de los contornos producidos por el modelo). Según King, el modelo es bastante robusto frente a desviaciones del supuesto de distribución (29). A pesar de esta ventaja, debemos reconocer que la distribución existente en nuestro ejemplo no es la más apropiada.

Utilizaremos otro ejemplo para mostrar una inferencia en la cual el supuesto resulta mucho más plausible: la estimación de la proporción de ciudadanos residentes en áreas urbanas o rurales que fueron a votar en la elección de 1946 (30). Por lo tanto, en este caso, X_i es la proporción de adultos masculinos que vivían en ciudades (en Argentina, es considerado como "población urbana" todo residente en localidades de más de 2.000 habitantes), T_i es la proporción de votantes sobre el total de empadronados para votar, y N_i es el número de empadronados (β_i^u es la proporción de ciudadanos urbanos que votaron y β_i^r es la misma proporción para los ciudadanos rurales). Como puede observarse en el Gráfico 7, es posible de localizar claramente un centro emisor de las líneas. En consecuencia, en este caso, obtenemos estimadores más precisos de nuestras cantidades de interés ($B^u = 0.855$, y $SE(B^u) = 0.0053$; $B^r = 0.7853$, y $SE(B^r) = 0.0095$).

Desafortunadamente, son los problemas los que nos aportan los datos y sus distribuciones, y éstas no son, entonces, escogidas por los investigadores. En el último caso presentado, si bien tenemos datos excelentes para



correr el modelo propuesto por King, sólo nos permiten saber que los ciudadanos residentes en las ciudades fueron a votar en una proporción algo superior a los que vivían en el campo, y esta diferencia resulta usual debido a los problemas de traslado que tenían los habitantes rurales (31).

Otro problema que puede presentar el modelo es la presencia de un sesgo de agregación: el parámetro puede estar relacionado con X_i . Entonces, no produce un estimador apropiado de la distribución. Afortunadamente, es posible testear si existe un sesgo de agregación (producido por una correlación entre β_i^0 y X_i) a través de un gráfico de dispersión (32). De este modo podemos probar que en el ejemplo principal que hemos desarrollado no presenta problemas de sesgo de agregación: en el Gráfico 8 puede observarse que ambas variables no presentan relación. El cuadrado del coeficiente de correlación es sólo 0,050. Una vez más, el nivel de votación de los ciudadanos urbanos es un mejor ejemplo de la ausencia de sesgo de agregación (ver Gráfico 9).

El tercer problema señalado por King es la existencia de dependencia espacial. Hemos graficado β_i^0 en un mapa, y, si bien no encontramos fuertes relaciones entre las proporciones estimadas de obreros votantes a Perón entre distritos vecinos, tampoco podemos afirmar que sea nula. En todo caso, el único problema serio surge cuando la dependencia espacial está relacionada con X_i (King, 1997: 166), y, como ya hemos visto, esto no ocurre en nuestro ejemplo.

Además de estos tres problemas, otras dificultades puede aparecer. A continuación comentaremos las que surgieron trabajando en torno a la elección de 1946. En primer lugar, pueden presentarse valores bajos de X_i en todos los distritos. En estos casos casi todas las líneas en el "tomography graph" serán prácticamente horizontales. Esto significa que, según la información determinística representada por las líneas, la baja proporción de X_i permite a β_i^0 a tomar todos los valores posibles. Al mismo tiempo, estos distritos están casi por completo conformados por sujetos que no presentan la característica en cuestión y, por lo tanto, se estrecha sustancialmente los posibles valores de β_i^0 , tal como las líneas lo demuestran. Por dar otro ejemplo, si intentamos estimar la proporción en la que los obreros (manuales) del sector servicios votaron a Perón (B^s), encontraremos que los datos ofrecen poca información determinística y, asimismo, que la distribución no parece ser normal bivariada. Los Gráficos 10 y 11 muestran este problema. Por lo tanto, obtendremos un error estándar extremadamente alto para la proporción estimada (B^s es estimado en 0,4953, y SE (B^s) en 0,2249). La distribución posterior de B^s (mostrada en el Gráfico 11) confirma la importancia de esta dificultad, en comparación con la distribución de B^0 (la proporción de obreros que votaron por Perón) en el ejemplo principal (Gráfico 6). El origen del problema es que había muy pocos obreros en el sector servicios en casi todos los distritos. Entonces, los límites determinísticos para sus conductas electorales son demasiado amplios (33). Asimismo, en estos casos es muy difícil de localizar la moda a partir de las líneas, especialmente en la dirección horizontal. Entonces, serán muy inciertas las inferencias acerca del tipo de distribución que ajusta a estos datos. Un problema similar surge si queremos estimar la proporción de obreros industriales que votaron a Perón (34).

Este no es el caso de nuestro ejemplo principal, en la que la proporción de obreros en cada distrito oscila entre 0,32 y 0,64. Este tipo de distribución no provee de límites estrechos para ninguna de las proporciones estimadas (como ya hemos visto), pero permite al modelo estadístico correr adecuadamente.

Un obstáculo parecido se presenta cuando existe escasa variabilidad en la variable X_i . Cuanto menor es la variación en X_i , mayor será la imprecisión que se deberá asociar a los resultados del modelo. En estos casos, se podrán ver líneas paralelas en los "tomography graph", porque la pendiente de estas líneas es $-X_i/(1-X_i)$. Debido a que el modelo es condicional a X_i , esta imprecisión no siempre se indica en los errores estándar estimados. Por ejemplo, si quisiéramos analizar la relación entre la edad de los votantes y el voto a Perón en 1946, nos encontraríamos con este tipo de problema. La complicación es que la proporción de gente de 20 a 39 años (X_i) es casi la misma en todos los distritos (ver Gráfico 12). Entonces, todas las líneas determinísticas son casi paralelas, y es prácticamente imposible de estimar dónde se ubica la moda (Gráfico 13). Sin embargo, estos problemas no se reflejan en los parámetros estimados. El error estándar no es excesivamente grande: SE (B^y) es 0,046, y SE (B^n) es 0,048. En cambio, en las líneas tomográficas es posible detectar el error en ajuste del modelo. Por lo tanto, los problemas de ajuste deben ser cuidadosamente considerados. En conclusión, se debe ignorar la estimación anómala de la



proporción de ciudadanos de 20 a 39 años que habrían votado a Perón (0,289) y, también, la de los mayores a 40 (0,803).

De los estudios descriptivos a los explicativos

Además de la ventaja principal del método de King (brindarnos precisas estimaciones regionales) (35), esta propuesta tiene un segundo beneficio. No sólo no es necesario sostener el supuesto del valor constante de los parámetros a lo largo de todos los distritos (como era el caso en el método de Goodman) sino que, incluso, se estiman los parámetros para cada distrito. Entonces, a partir de estas estimaciones puede avanzarse en una segunda fase de análisis, incorporando terceras variables. Es posible contrastar la utilidad de variables explicativas para dar cuenta de la variabilidad de los parámetros estimados a nivel distrital.

Una hipótesis importante en torno al triunfo de Perón ha asociado este fenómeno con la activación política de una nueva clase obrera. La idea central es que el núcleo de su victoria provino de aquellos que habían arribado recientemente al Gran Buenos Aires desde el Interior del país, atraídos por el desarrollo industrial del decenio previo a 1946 (36). No todos los historiadores acuerdan con esta hipótesis y, desafortunadamente, no existen datos directos para resolver este asunto. El debate entre Smith (1972 y 1974) y Germani (1973) finalmente no produjo mucha evidencia para resolver la controversia.

He utilizado alguna información censal extraviada para mejorar, junto con la técnica de King, el conocimiento sobre este tema. En 1938 la provincia de Buenos Aires realizó un censo demográfico. Sin embargo, el gobierno lo anuló, justo cuando ya estaban saliendo los resultados de los procesamientos preliminares (37). Tan sólo quedó el total de la población para cada distrito (publicada en 1942) (38). A partir de esta información hemos podido calcular la tasa anual de crecimiento poblacional para cada distrito de la provincia de Buenos Aires entre el 18 de diciembre de 1938 y el 10 de mayo de 1947. Para la ciudad de Buenos Aires contamos con información más detallada publicada en el censo de población de 1936 (39) (realizado el 22 de octubre). De este modo, pudimos calcular la misma tasa anual para un período bastante similar.

Si los obreros recientemente arribados hubieran sido el núcleo del triunfo de Perón, debería haber una relación fuerte entre la tasa de crecimiento poblacional de los distritos y la proporción estimada en la que los obreros habrían votado a Perón (β_i^0) (40).

El gráfico de dispersión (Gráfico 14) muestra que no existe una fuerte relación entre estas dos variables. A pesar de cierta apariencia de dependencia, no pueden sostenerse relaciones lineales (R^2 ajustada es 0,088) ni de tipo logarítmicas. Aunque todos los distritos con altas tasas de crecimiento demográfico presentan una alta proporción estimada de obreros votando por Perón, existen numerosos distritos con un crecimiento reducido o negativo que también tienen una alta proporción de obreros peronistas. Hemos probado con una serie de transformaciones de ambas variables pero fue imposible de mejorar la relación entre las variables.

Entonces, no es posible establecer una clara relación entre antigüedad de la clase obrera y diferentes tasas de apoyo a la candidatura de Perón. No sólo la nueva clase obrera habría votado en una alta proporción a Perón; en algunos distritos, la antigua clase obrera también lo habría preferido en porcentajes similares. A pesar de que no surgieron relaciones sistemáticas, es posible afirmar que casi toda la nueva clase obrera habría votado por Perón, y que la vieja clase obrera presentó una gran variabilidad en su conducta electoral. Sin duda, se deberían llevar adelante trabajos mucho más detallados para mejorar nuestra comprensión de estos fenómenos.

He tratado de mostrar cuán útil resulta esta metodología para analizar relaciones entre variables. De este modo, posibilita el pasaje de estudios descriptivos (basados en la información univariable disponible) a trabajos explicativos. Es por ello que el método de King tiene un gran potencial para abrir nuevas aproximaciones hacia un gran número de problemas que habían sido dejados de lado debido a que no había posibilidades de estimar las distribuciones condicionales. Por esta razón, creo que es una técnica que posibilitará revolucionar importantes áreas del conocimiento de nuestro pasado.



Notas

(1) Frente a un problema de este tipo existe siempre el riesgo de cometer una "falacia ecológica": inferir erróneamente la conducta de un sector a partir del análisis de la covarianza entre la importancia de dicho sector social y la magnitud de esta determinada conducta conjuntos geográficos. Muchas veces resulta matemáticamente posible que la mayor proporción de votantes a determinada fuerza política (por continuar con el mismo ejemplo), covarie positivamente con la importancia de determinado sector (los obreros), pero que sea el resto de los sectores sociales los que estén votando a este partido (observar el ejemplo que plantea Boudon, 1963: 255).

(2) No nos proponemos corroborar la validez de la técnica, cuestión que nos excede, y que además el propio King ha desarrollado, testeando su método con una gran cantidad de casos reales, para los cuales contaba con el conocimiento de las verdaderas proporciones. Asimismo, ha desarrollado las técnicas para detectar, a partir de los datos agregados, la violación de los supuestos necesarios para su funcionamiento, y los efectos que estas violaciones conllevan sobre la calidad de las estimaciones.

(3) Ver, por ejemplo, Cantón, Jorrot y Juárez (1976) y los trabajos contenidos en Mora y Araujo y Llorente (1980).

(4) Hemos respetado la clasificación del Censo que consigna como obreros, no sólo a los obreros de la industria manufacturera, sino también a los del sector primario y a los de servicios. Sin embargo, a partir de las planillas de resumen inéditas, existentes en el INDEC es posible diferenciar a los trabajadores manuales por ramas y realizar otras estimaciones.

(5) Asimilamos a la condición obrera a los desocupados y a los no ocupados sin ingresos.

(6) Lamentablemente la información sobre la cantidad de obreros discriminados por sexo se ha perdido en el INDEC; contamos sí con la cantidad de obreros por ramas productivas, y la discriminación de ocupados según sexo por tales ramas para cada uno de los distritos analizados (datos inéditos del Censo Nacional de 1947). Hemos supuesto que todos los obreros del sector primario eran varones, y hemos extrapolado para los otros dos sectores, la proporción de varones y de mujeres entre los obreros de las ramas secundarias y terciarias de la Capital Federal, según cálculos efectuados en base a los datos de ocupados por sexo y rama y a la cantidad total de obreros varones que se encuentra consignada en el artículo de Cantón, Jorrot y Juárez (1976: 412).

(7) Como tampoco se encontró esta información, se utilizó el coeficiente de ocupados varones mayores de 18 años sobre el total de ocupados varones mayores de 14 años disponible para cada uno de los distritos analizados (datos inéditos del Censo Nacional de 1947).

(8) Información consignada en Cantón (1968).

(9) X_i y T_i presentan bases (N_i) distintas. Hemos trabajado con N_i igual a la cantidad de votantes, obviando la cuestión de los inscriptos que no votaron, como así también la de las nacionalidades de los varones mayores de 18 años. Por lo tanto, se ha trabajado bajo el supuesto de que no habría sesgos por nacionalidad ni por no votantes. No es que nos resulten supuestos empíricamente ciertos, sino que no contamos con la información de nacionalidad en el Censo de 1947, sino tan sólo el país de origen, sin los nacionalizados (y tememos que el sesgo resulte mayor). Tampoco contamos, obviamente, con la proporción de asistencia al comicio discriminada por sectores sociales. Esto último es factible de inferir con la metodología de King (y lo hemos realizado), pero su inclusión complicaría bastante la exposición en las limitadas páginas de la presente ponencia.

(10) Cabe diferenciar a B^o y B^n (que son promedios ponderados por la significación numérica de los distritos, en este caso según la cantidad de votantes) de los promedios simples de β_i^o y β_i^n , que aquí designamos con el uso de la cursiva: B^o y B^n .

(11) Al respecto puede consultarse el análisis de Selvin (1958)

(12) Según estos autores, se requeriría que la correlación ecológica sea relativamente elevada (al menos +0.7), y que se realice el supuesto fuerte de que cada grupo se comporta de la misma manera, no importa quienes sean sus vecinos inmediatos (la característica de su zona). En estos casos, sugieren realizar una regresión lineal y sostienen que la pendiente (b) es similar al coeficiente de correlación ecológico (r).

(13) Goodman presenta esta identidad como un supuesto (1953:664), sin embargo, como aclara King (1997: 39), esta identidad es verdadera por definición, ya que, en cada distrito, la cantidad total de votantes a Perón, en nuestro ejemplo, es igual a la suma de la cantidad de obreros que lo votaron ($\beta_i^o X_i$) y de la cantidad de no-obreros que también lo hicieron [$\beta_i^n (1-X_i)$].

(14) Si $T_i = a + b X_i$, cuando no hay obreros ($X_i=0$), $T_i = a = B^n$.

(15) Cuando son todos obreros ($X_i=1$), $T_i = a + b$.

(16) En esta misma tendencia, puede consultarse el análisis realizado por Klatzmann (1956) sobre la proporción de votos comunistas debidas a los obreros.



(17) Al utilizar la información determinística que surge del método de los límites para las cantidades de interés en cada uno de los distritos se logra un sustancial avance y se hacen inferencias especialmente robustas frente al sesgo de agregación. Como los límites son conocidos con certeza, este procedimiento suma una gran cantidad de información al modelo estadístico.

(18) King ha desarrollado dos programas que permiten utilizar su propuesta. Los ejemplos que se exponen a continuación fueron realizados con el Ezi, A(n easy) Program for Ecological Inference, elaborado por King y Ken Benoit. Este programa, que no requiere ningún software adicional (a diferencia del EI), puede obtenerse gratuitamente en <http://GKing.Harvard.Edu>.

(19) Podemos ver esto rearmando la identidad básica $\{T_i = \beta_i^o X_i + \beta_i^n (1-X_i)\}$, con una incógnita (β_i^n) expresada como función de la otra (β_i^o): $\beta_i^n = \{[T_i / (1-X_i)] - [X_i / (1-X_i)] \beta_i^o\}$.

(20) Como los conceptos utilizados en esta figura son matemáticamente equivalentes a una versión idealizada del problema "tomográfico", tal como existe en los rayos X y en la tomografía computada, King denominó a esta representación gráfica como "tomography plot".

(21) Según King, este es un supuesto bastante razonable. Si los distritos tienen algo en común (el mismo gobierno, similares características, nos referimos a la misma elección, etc) es posible que β_i^o y β_i^n , si no son constantes en todos los distritos como en el modelo de Goodman, se agrupen, al menos, en torno a una única moda, aunque sea con una amplia varianza. Esta es una de las posibles motivaciones para el supuesto de modelado de una distribución normal bivariada truncada.

(22) Se parte de discriminar los parámetros distritales en el parámetro regional y un término de error: $\beta_i^o = B^o + \varepsilon_i^o$ y $\beta_i^n = B^n + \varepsilon_i^n$ donde los términos de error, ε_i^o y ε_i^n , tienen media cero, porque ellos están definidos como los desvíos de sus correspondientes valores esperados condicionales. $E(\beta_i^o) = B^o$. Si sustituimos en la ecuación $T_i = \beta_i^o X_i + \beta_i^n (1 - X_i)$, obtenemos $T_i = (B^o + \varepsilon_i^o) X_i + (B^n + \varepsilon_i^n) (1-X_i) = B^o X_i + B^n (1-X_i) + \varepsilon_i$. A partir de esto, la función media de la variable de salida es: $E(T_i | X_i) = B^o X_i + B^n (1-X_i)$. Esta ecuación muestra que B^o y B^n pueden ser estimados por algo similar a la regresión de Goodman, a pesar de que este procedimiento es mejorado en nuestro caso. La estimación se realiza primero a través de los parámetros de una distribución normal bivariada no truncada y luego se la trunca.

(23) Las estimaciones de máxima verosimilitud consisten, sintéticamente, en encontrar un valor para un parámetro que produzca salidas, en su aplicación, lo más cercanas a las salidas conocidas, para las cuales se desea estimar este parámetro. Un desarrollo de los estimadores de regresión por máxima verosimilitud puede encontrarse en Hanushek y Jackson (1977: 344-348).

(24) En el ejemplo que estamos desarrollando, las estimaciones de los parámetros de la distribución bivariada normal truncada (ψ) son los siguientes: Medias: $B^o = 0.6928$ $B^n = 0.3009$ Desvíos Estándar: $S(B^o) = 0.1875$ $S(B^n) = 0.0503$ Coeficiente de Correlación: $\rho^{on} = 0.2312$

(25) Incluso si supiéramos con certeza los valores de ψ , esto no implicaría conocimiento perfecto de β_i^o y β_i^n .

(26) Este último paso lo realiza por simulación (a partir de la distribución estimada), ya que las integrales requeridas son a menudo muy difíciles o imposibles de resolver.

(27) Por ejemplo, una vez que podamos tomar números aleatorios de la distribución posterior $P(\beta_i^o | T)$, podremos realizar cualquier juicio deseado sobre β_i^o , a partir de la información presente en los datos agrupados. Si quisiéramos conocer nuestra mejor suposición (estimación puntual) de β_i^o podríamos tomar la media de un gran número de salidas aleatorias. Su error estándar podría calcularse computando la desviación estándar de las salidas aleatorias. La completa distribución posterior de β_i^o podría aproximarse dibujando un histograma de esos números aleatorios.

(28) Para B^n , el intervalo al 80% es de 0.276 a 0.339, y al 95%, de 0.266 a 0.352.

(29) Esto se debe a que sólo las distribuciones univariadas -coratadas de la distribución normal bivariada por las líneas tomográficas- son usadas para computar las cantidades de interés, y estas líneas tomográficas son conocidas con certeza (King, 1997: 184-185).

(30) En este caso, solo se utilizaron los distritos de la provincia de Buenos Aires.

(31) Tal vez lo interesante sea la alta participación de la población rural en esta elección. Perón trabajó especialmente para ganar el apoyo de los trabajadores rurales y de los pequeños agricultores (Lattuada, 1986). Sin embargo, para avanzar en esta hipótesis sería necesario realizar estudios comparativos con otras elecciones argentinas.

(32) A pesar de que el modelo utilizó el supuesto de la no existencia de sesgo de agregación, al incorporar los límites como parte del cálculo, posibilita que los estimadores a nivel distrital puedan estar correlacionados con X_i (King, 1997: 183).

(33) Al mismo tiempo, con distritos casi sin obreros del sector servicios, el método ofrece estimaciones muy precisas para la conducta electoral de los ciudadanos que no eran obreros de este sector ($SE(B^n) = 0,0388$).

(34) La misma dificultad aparece si quisiéramos analizar la relación entre creencias no católicas y opciones electorales en 1946, tal como alguien me sugirió realizar en un congreso de historiadores. Como casi toda la población declaraba en el censo ser católica, todas las estimaciones sobre la conducta electoral de los no-católicos resultan extremadamente inciertas.



(35) Tal como es testeado a partir de datos individuales conocidos en una serie de ejemplos transcritos en King (1997).

(36) Entre 1935 y 1947, casi un millón de personas migraron al Gran Buenos Aires desde las provincias menos desarrolladas (especialmente desde la región noroeste) y desde las zonas rurales de la región pampeana. Germani (1973) destaca la importancia de los primeros, y Lattes (1978) enfatiza el papel de los segundos.

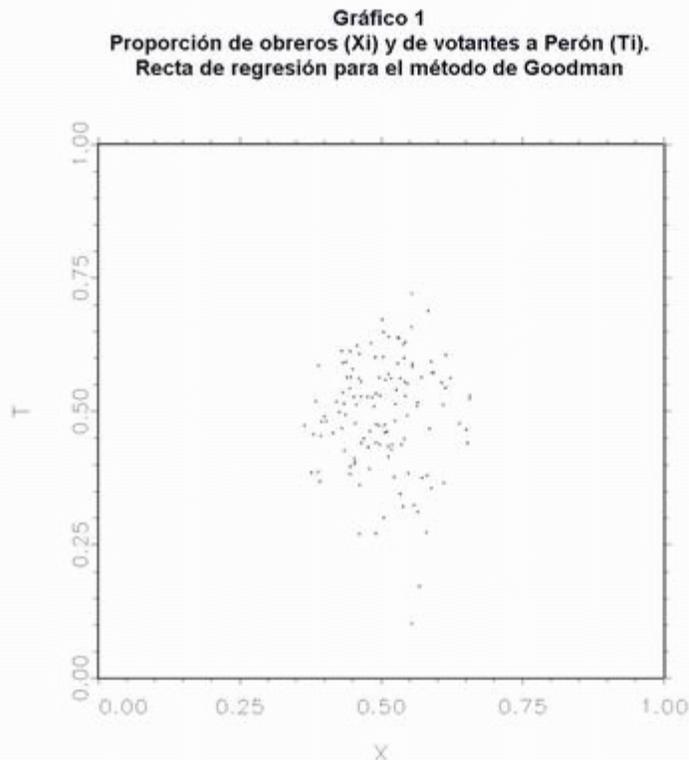
(37) No pude encontrar ninguna información sobre esta decisión, pero probablemente el gobierno conservador no quiso reconocer el crecimiento poblacional de los suburbios de Buenos Aires. La población de los mismos era todavía estimada a partir de los datos censales de 1914 y, por lo tanto, subestimada. Estas estimaciones tenían consecuencias sobre la importancia numérica de la representación política de estos distritos con predominio de la clase obrera.

(38) Provincia de Buenos Aires, Ministerio de Gobierno, *Censo de 1938 y cálculo a 1942*. La Plata: Taller de Impresiones Oficiales, 1942.

(39) República Argentina, Municipalidad de la Ciudad de Buenos Aires, *Cuarto Censo General, 1936, Población*, Buenos Aires, 1939.

(40) El supuesto realizado es que los nuevos trabajadores eran más importantes en aquellos distritos donde la población había crecido más en la última década. Esto resulta muy plausible pues la bibliografía especializada sugiere que el fenómeno migracional estuvo muy asociado con el desarrollo industrial. En este período, es muy difícil vincular la tasa de crecimiento demográfico con otro sector social.

Gráficos



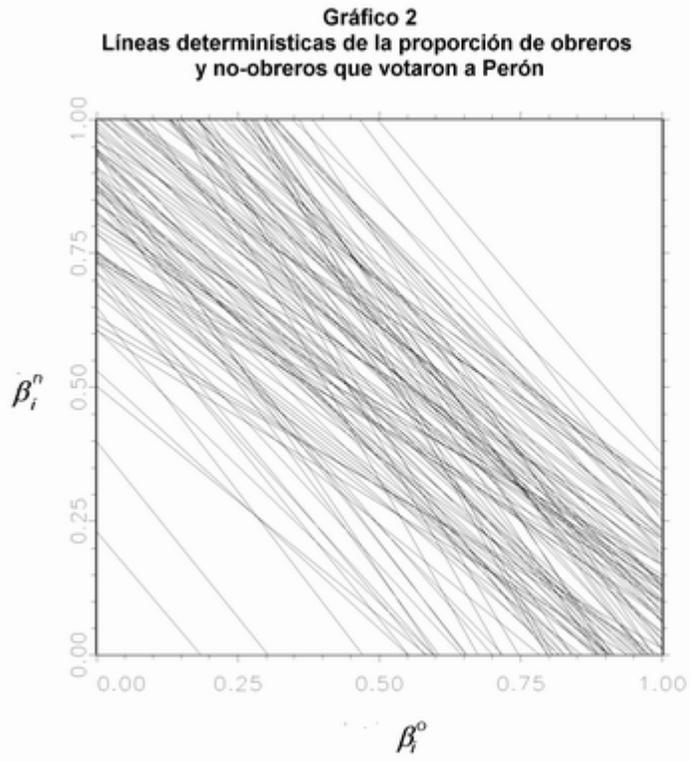




Gráfico 3
Distribución normal bivariada truncada

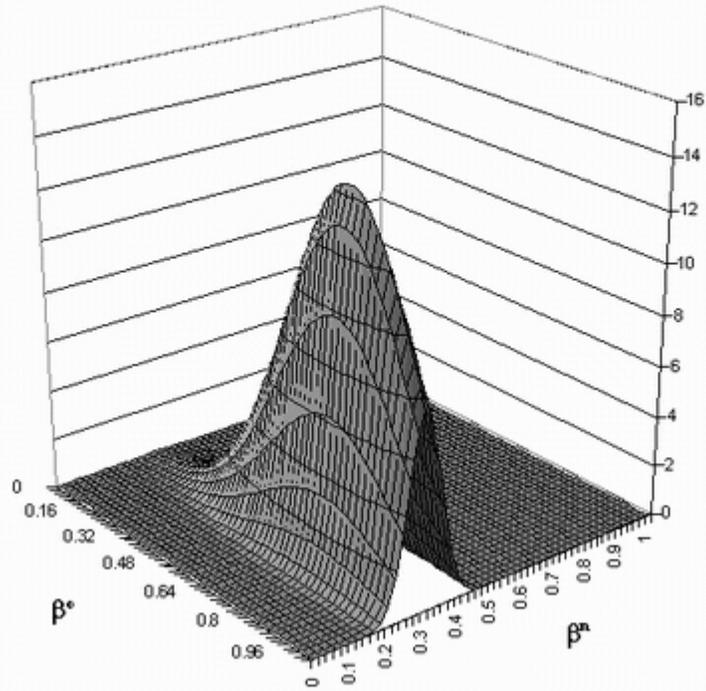




Gráfico 4
"Tomography plot" de la proporción de obreros y no obreros que votaron a Perón con elipses de contorno del 50% y el 95%

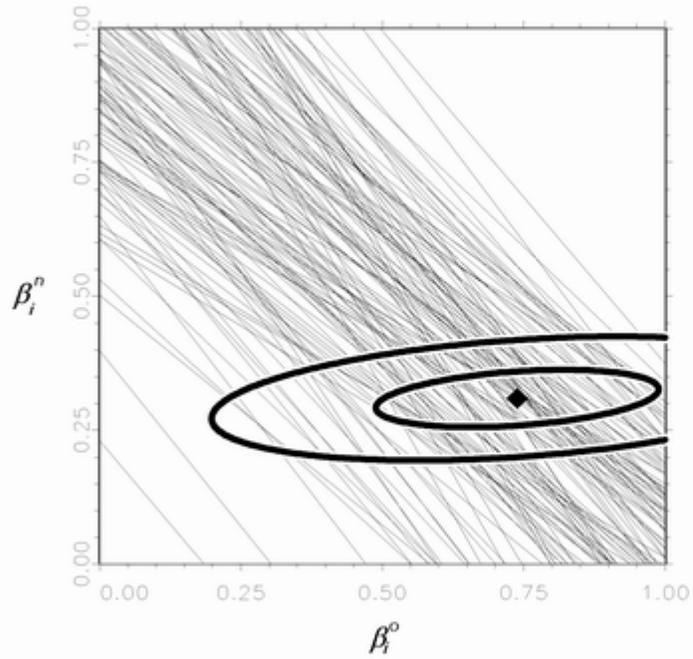




Gráfico 5
Proporciones estimadas de obreros
y no-obreros que votaron a Perón

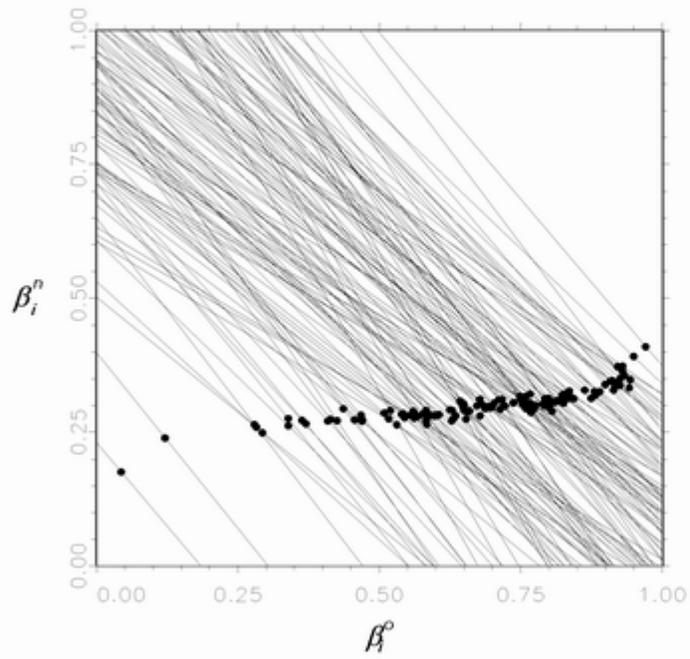




Gráfico 6
Distribución posterior de la proporción
regional de obreros que votaron por Perón

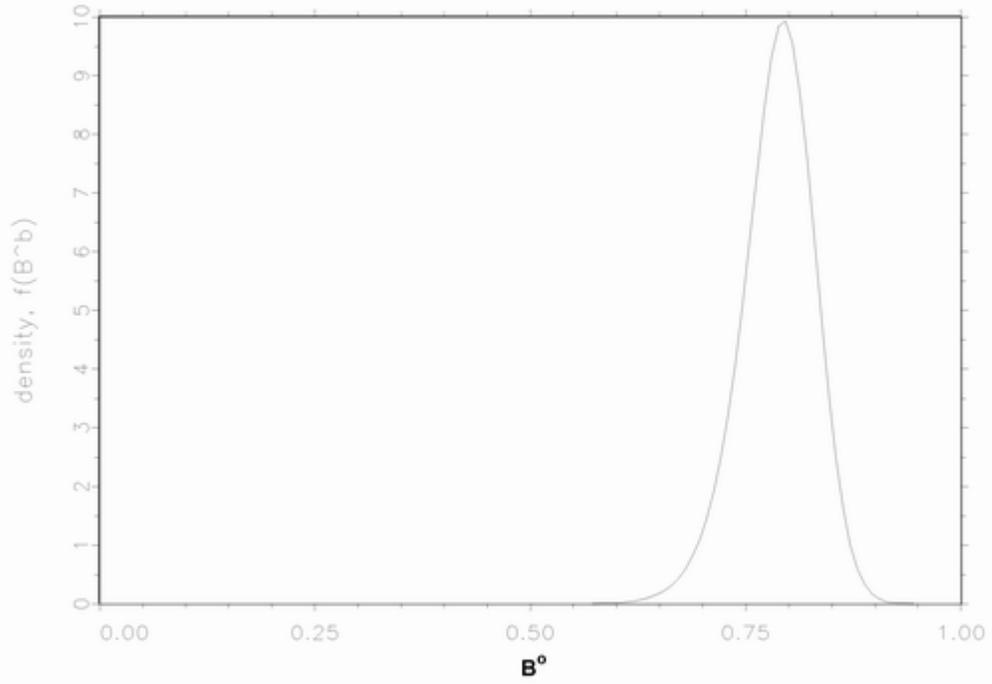




Gráfico 7
"Tomography plot" de la proporción de ciudadanos urbanos y rurales que fueron a votar

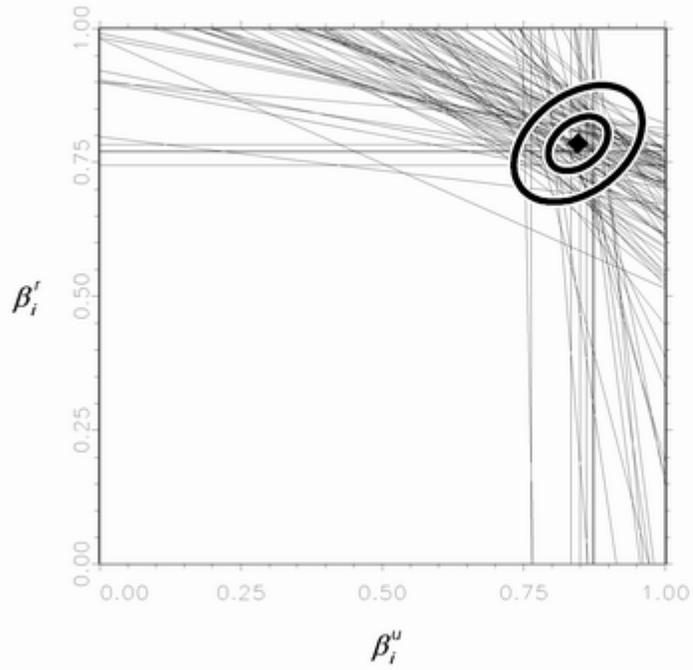




Gráfico 8
Proporción de obreros y proporción estimada
de obreros que votaron a Perón

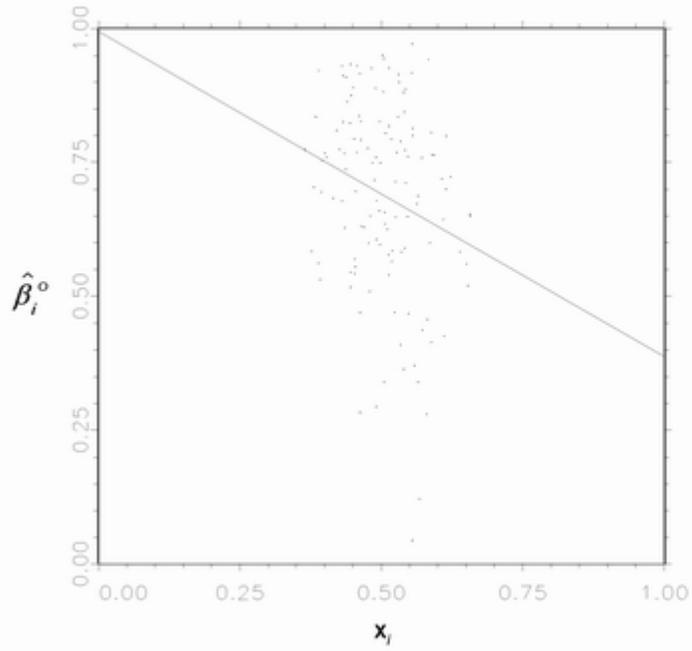




Gráfico 9
Proporción de ciudadanos urbanos y proporción
estimada de votantes urbanos

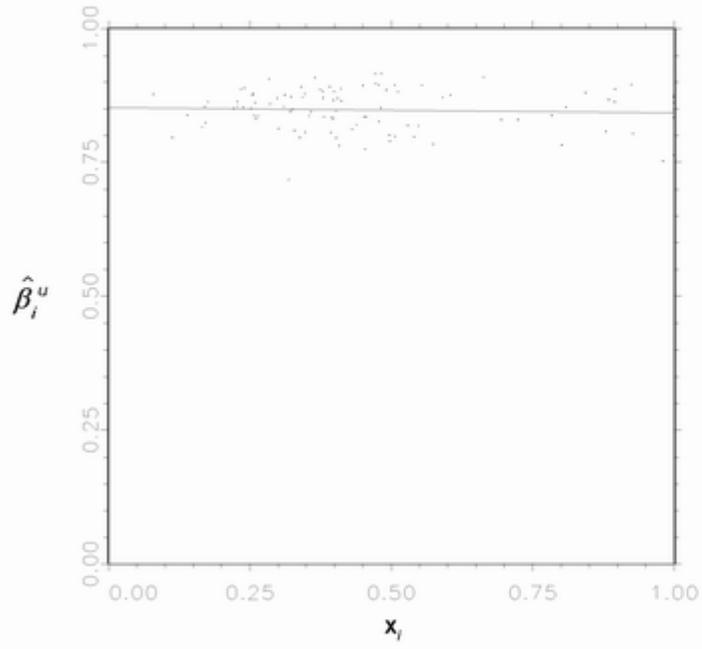




Gráfico 10
Líneas determinísticas de las proporciones de obreros del sector servicios y del resto de los ciudadanos que votaron a Perón

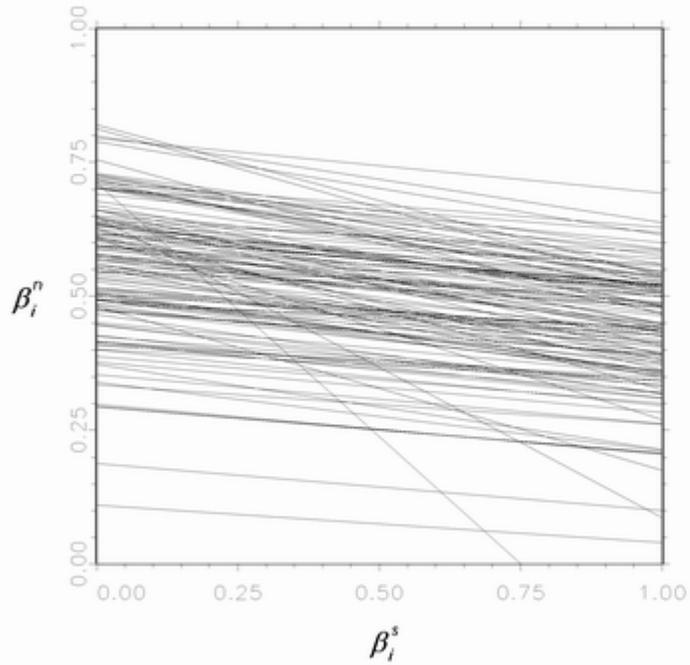




Gráfico 11
Distribución posterior de la proporción regional
de los obreros del sector servicios que votaron a Perón

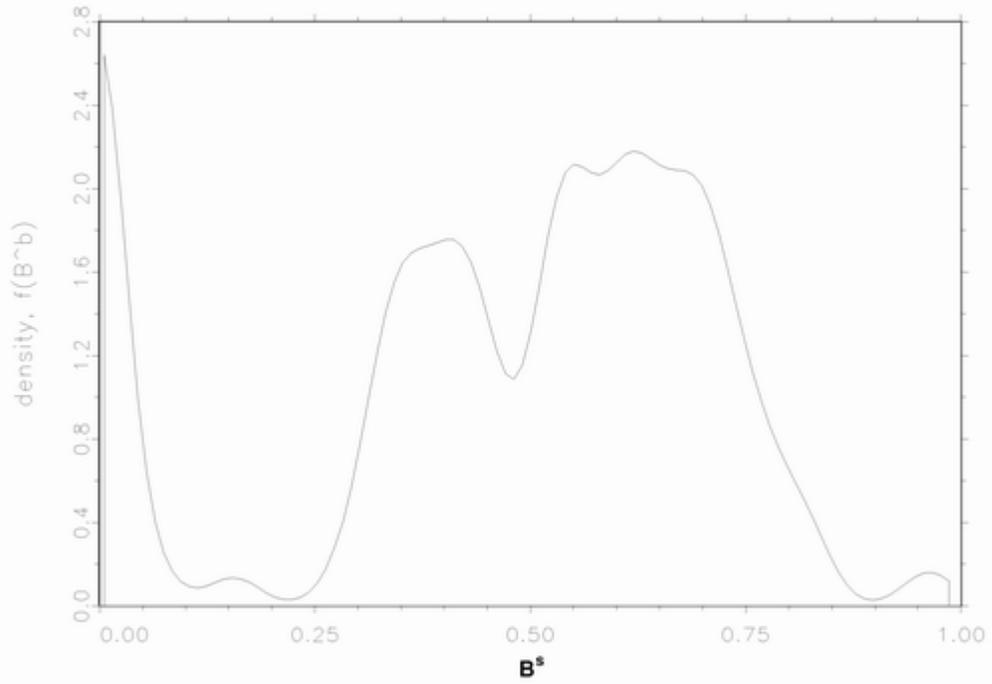




Gráfico 12
Proporción de hombres de 20 a 39 años (Xi)
y proporción de votantes a Perón (Ti)

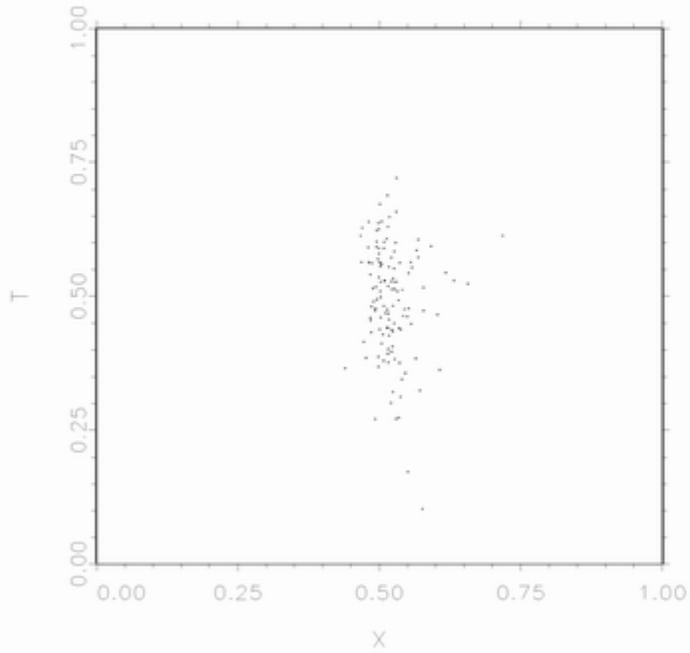




Gráfico 13
Lineas determinísticas de la proporción de hombres
de 20 a 39 años y mayores de 40 que votaron por Perón

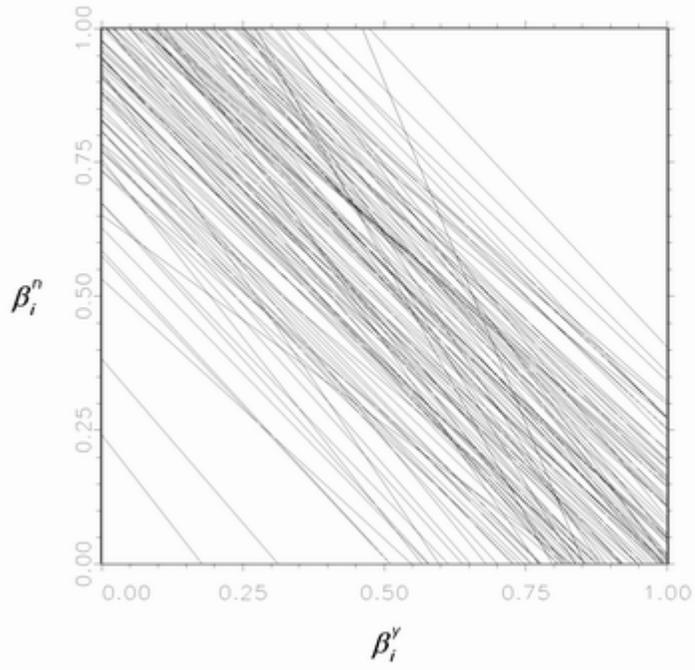
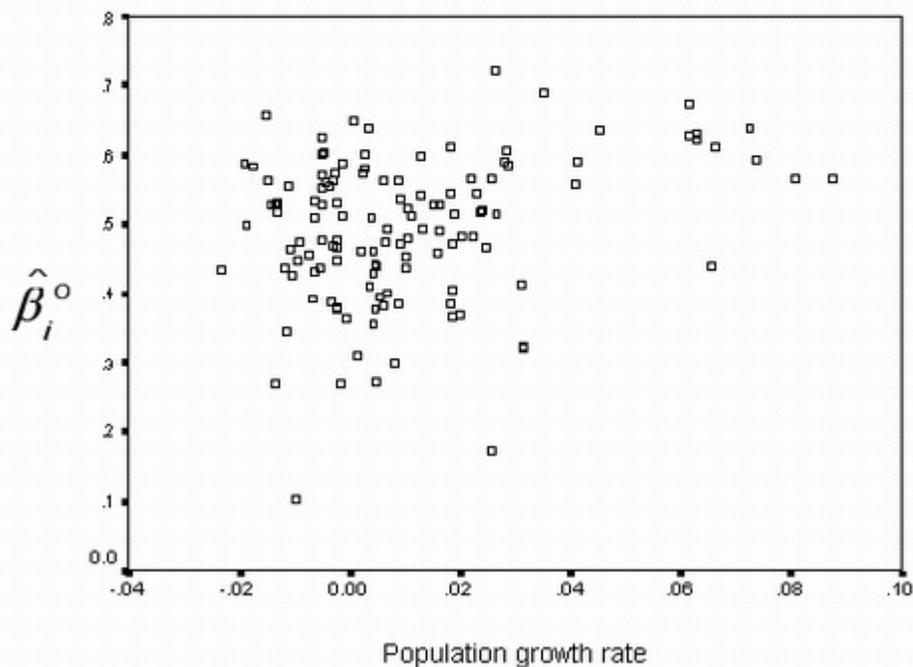




Gráfico 14
Tasa de crecimiento poblacional y
proporción estimada de obreros que votaron a Perón



Bibliografía

Boudon, Raymond (1963). "Propiedades individuales y propiedades colectivas; un problema de análisis ecológico", en R. Boudon y P. Lazarsfeld, *Metodología de las Ciencias Sociales*. Volúmen II. Barcelona, Laia, 1974; p.p. 247-284.

Cantón, Darío (1968). *Materiales para el estudio de la sociología política en la Argentina*. Buenos Aires, Editorial del Instituto.

Canton, Darío, Jorge Raúl Jorrot y Eduardo Juárez (1976). "Un intento de estimación de las celdas interiores de una tabla de contingencia basado en el análisis de regresión. El caso de las elecciones presidenciales de 1946 y marzo de 1973", *Desarrollo Económico*, 63; p.p. 395-417.

Canton, Darío y Jorge Raúl Jorrot (1998). "Buenos Aires en tiempos del voto venal: elecciones y partidos entre 1904 y 1910", mimeo.

Dollar, Charles M. y Richard J. Jensen (1971). *Historian's Guide to Statistics*. Quantitative Analysis and Historical Research. New York, Holt, Rinehart and Winston.

Duncan, Otis Dudley y Beverly Davis (1953). "An Alternative to Ecological Correlation", *American Sociological Review*, 18; p.p. 665-666.

Galtung, Johan (1966). *Teoría y métodos de la investigación social*. Tomo I. Buenos Aires, EUDEBA, 1978.



-
- Germani, Gino (1955). *Estructura social de la Argentina*. Buenos Aires, Raigal.
- Germani, Gino (1973). "El surgimiento del peronismo: el rol de los obreros y de los migrantes internos", *Desarrollo Económico*, 51: 435-488.
- Goodman, Leo A. (1953). "Ecological Regressions and Behavior of Individuals", *American Sociological Review*, 18; p.p. 663-664.
- Hanushek, Eric y John Jackson (1977). *Statistical Methods for Social Scientists*. San Diego, Academic Press.
- King, Gary (1997). *A Solution to the Ecological Inference Problem*. Reconstructing Individual Behavior from Aggregate Data. Princeton, Princeton University Press.
- Klatzmann, Joseph (1956). "Comportamiento electoral y clase social", en R. Boudon y P. Lazarsfeld, *Metodología de las Ciencias Sociales*. Volúmen II. Barcelona, Laia, 1974; p.p. 285-299.
- Lattes, A. (1978). *La dinámica de la población rural en la Argentina entre 1870 y 1970*. Buenos Aires, CENEP.
- Lattuada, Mario (1986). *La política agraria peronista, 1943-1983*. Buenos Aires, Centro Editor de América Latina.
- Llorente, Ignacio. (1977). "Alianzas políticas en el surgimiento del peronismo: el caso de la provincia de Buenos Aires". *Desarrollo Económico*, 65: 61-68.
- Mora y Araujo, Manuel e Ignacio Llorente, comp. (1980). *El voto peronista*. Ensayos de sociología electoral argentina. Buenos Aires, Sudamericana.
- Robinson, William S (1950). "Ecological Correlation and the Behavior of Individuals", *American Sociological Review*, 15, p.p. 351-357.
- Selltiz, Claire, L. Wrightsman y S. Cook (1976). *Research Methods in Social Relation*. Tercera edición. New York, Holt, Rinehart and Winston.
- Selvin, Hanan (1958). "Aspectos metodológicos del *Suicide*", en R. Boudon y P. Lazarsfeld, *Metodología de las Ciencias Sociales*. Volúmen II. Barcelona, Laia, 1974; p.p.353-372.
- Smith, Peter (1972). "The social base of peronism", *Hispanic American Historical Review*, 52: 55-73.
- Smith. Peter (1974). "Las elecciones de 1946 y las inferencias ecológicas". *Desarrollo Económico*, 54: 385-398.