APÉNDICE II

ANÁLISIS ESTADÍSTICO APLICADO AL MONITOREO DE LA DEUDA SOCIAL

El presente apéndice ha sido elaborado por Jimena M. Macció

II.1. Elaboración de los indicadores

Con el fin de brindar evidencia empírica acerca de la segregación socioeconómica residencial -definida a partir de criterios de estratificación socioeducativa- en la distribución de fuentes de bienestar y logros en el espacio del desarrollo humano (Salvia, 2005), se construyó un conjunto de indicadores de privación que efectuado reflejar la realidad socioeconómica de nuestro país a partir de tres mediciones semestrales realizadas entre junio de 2004 y junio de 2005. El análisis estadístico efectuado sobre la Encuesta de la Deuda Social Argentina (EDSA) está constituido por la aplicación de diversos métodos estadísticos que intentaron comprobar la significatividad de las diferencias en los indicadores elaborados según espacio residencial socioeducativo.

Los indicadores construidos a partir de los datos de la encuesta para las distintas dimensiones se pueden agrupar en dos clases principales. Por un lado, se confeccionaron indicadores de déficit, que se concentran en la ausencia de determinada capacidad, el acceso deficiente a un determinado servicio, la posesión insuficiente de un recurso específico, la posibilidad de ejercer en forma incompleta un derecho dado, etc. Por otro lado, se elaboraron indicadores de florecimiento. Estos se concentran principalmente en la capacidad, en la posibilidad de acceder, de ejercer, de brindar, etc. Mientras en el caso del déficit se espera una relación inversa con el espacio residencial socioeducativo, en cuanto a que la proporción de personas u hogares con déficit tendería a disminuir en forma significativa a medida que se pasa del ERS 1 al ERS 4, en los indicadores de florecimiento se esperaría una relación directa.

En su mayoría, los indicadores fueron construidos de manera dicotómica. Es decir, adquieren dos valores: uno referido a la situación de privación y el otro constituido por los casos que no la padecen. Mediante estos indicadores dicotómicos, el tratamiento de los datos permite calcular las proporciones o porcentajes de personas u hogares en cada una de las situaciones, según espacio residencial socioeducativo. Por lo tanto, la exploración estadística que se describe a continuación paso a paso se concentra sobre el análisis de estimadores de proporción.

Los datos utilizados para construir estos indicadores de privación fueron los de la EDSA en sus tres mediciones. Estos datos fueron consistidos, buscándose y resolviéndose tanto errores de carga como de respuesta, teniendo especial consideración con los casos faltantes. En particular, el tratamiento de los datos de ingreso resultó un aspecto de importancia. Es usual entre las encuestas de hogares que no todos los individuos entrevistados respondan a las preguntas de ingresos. Este fenómeno puede sesgar las estimaciones de desigualdad si, por un lado la no respuesta depende del ingreso, y por otro, si el porcentaje de no respuesta varía en el tiempo (Gasparini y Sosa Escudero, 2001). Está última situación se puede dar como efecto de cambios metodológicos en el instrumento, su método de aplicación o en la codificación de las variables, así como también la estructura de la muestra o el operativo de relevamiento de información (Salvia y Donza, 1999). En este caso, se pueden confirmar una clara relación entre el ingreso y la no respuesta así como también una diferente proporción de no respuestas según la onda. Sin embargo, el instrumento de medición y su aplicación permanecen constantes, pudiéndose asociar las diferencias al cambio en la estructura de la muestra, al cual se hará referencia más adelante.

Para poder resolver el problema de no respuestas se realizó la estimación de un modelo de regresión que permite efectuar la imputación de ingresos a los no respondentes a partir de los ingresos de las personas en condiciones laborales, demográficas y socioeconómicas similares. De hecho, se realizaron dos estimaciones separadas, una para el ingreso laboral y otra para el ingreso de los hogares. Particularmente, el análisis de regresión para el ingreso laboral tomó en cuenta variables de índole demográfica (sexo, grupos de edad, situación conyugal) y socioeconómica (nivel educativo, situación ocupacional, ocupación principal, jefatura de hogar, etc.), además de tenerse en cuenta los espacios residenciales socioeducativos. El análisis para el ingreso del hogar incluyó otras variables como cantidad de componentes, población económicamente activa dentro del hogar, tipo de familia, ciclo vital del hogar, clima educativo, características del jefe de hogar, además de tenerse en cuenta la recepción de asistencia en forma monetaria o no monetaria por parte de organismos públicos y privados.

Es importante agregar, por otro lado, que el análisis estadístico se realizó tomando en cuenta la naturaleza compleja de la muestra, descripta previamente en el Apéndice I, utilizando el módulo de muestras complejas del SPSS en su versión 13 (1).

II.2. Estrategia de análisis de los datos

La estrategia de análisis estadístico adoptada combina el análisis estático de *stock* con el análisis dinámico de flujos, considerando que las estadísticas de flujo son necesarias para complementar los datos de *stock*, en especial cuando se intenta evaluar los cambios brutos y sus determinantes (Mehran, 1999). De esta manera, el análisis estático permite establecer la incidencia relativa de las privaciones en cada uno de los espacios considerados mediante el cálculo de tasas de recuento, que contabilizan el porcentaje de unidades con déficit sobre el total de las unidades observadas. Pero al concentrar el análisis en

la tasa de recuento se deja de lado el proceso dinámico que lo genera, de modo que con el análisis de flujos se busca discriminar los efectos que explican las tasas de recuento o incidencia desde un punto de vista dinámico, en especial la entrada y la permanencia en la situación deficitaria. (Salvia, 2005)

Las sucesivas mediciones de la EDSA procuraron reiterar la encuesta a los mismos sujetos. Si bien esta era la instrucción y la prioridad de los encuestadores, este objetivo se logró en parte de los casos. Del total de sujetos que respondieron la última onda de la EDSA (Junio de 2005), aproximadamente 60% respondieron las tres ondas (panel ondas 1, 2 y 3), 15% respondieron dos de las ondas (5% las ondas 1 y 3, 10% las 2 y 3), mientras que un cuarto de la muestra total representan casos nuevos con una única medición.

Estos porcentajes muestran que en total hay un 60% de los casos que se solapan completamente, mientras que el resto ha sufrido algún tipo de reemplazo a lo largo de las 3 ondas. Esta característica de la muestra de la EDSA tiene importantes consecuencias, ya que permite realizar sobre los datos dos tipos de análisis que fueron aplicados en forma complementaria y que a continuación se describen.

Incidencia de las privaciones en el Desarrollo Humano

En primer lugar resulta necesario conocer cómo inciden las privaciones, tanto en forma general como de manera diferencial para cada espacio residencial socioeducativo. En función de esto se calcularon tasas de incidencia del déficit (florecimiento), que simplemente consideran la cantidad de casos con déficit (florecimiento) sobre el total de casos observados.

Operativamente, el hecho de que la estructura de la encuesta se haya mantenido fundamentalmente constante (reiteración de preguntas formuladas de la misma manera) a lo largo de las tres mediciones permite considerar los datos recabados de manera conjunta. En este sentido, se elaboró una base de datos que consideraba a los 1100 sujetos encuestados en cada medición en forma agregada, resultando un total de 3300 casos.

"Cuando se combinan muestras independientes, se aumenta en forma directa la precisión a medida que las varianzas disminuyen proporcionalmente a los tamaños de las muestras. Sin embargo, con muestras solapadas como en este caso, la situación se vuelve más complicada, y las ventajas de las combinaciones tienden a disminuir en proporción a las correlaciones entre las variables medidas en cada instancia y el nivel de solapamiento (Kish, 1995). (Muiños, 2005)". Para evaluar las ventajas y desventajas de este procedimiento se calcularon las varianzas como abajo se consigna.

"La varianza de la media combinada de dos muestras iguales sería, sin solapamiento, $S^2/2$. Con solapamiento parcial P, [la varianza] sería $\frac{(1+PR)S^2}{2}$, siendo R la correlación entre las dos mediciones

D Capitulo APE2 5/5/06 8:09 PM Page 426

de la variable en cuestión. En forma general, y para medias basadas en J muestras solapadas, la expresión general para la varianza es la siguiente.

$$P_{t/1} = \frac{R_s 6}{2P_e}$$

En el caso de esta investigación, con J=3 y suponiendo desvíos uniformes en las tres muestras, un nivel de solapamiento del 70% y una correlación del 90%, debería obtenerse una reducción en la varianza de los estimadores del orden del 46%." (Muiños, 2005).

Dados estos resultados teóricos y considerando los porcentajes de solapamiento consignados anteriormente, se procedió a determinar la reducción aproximada en la varianza para los datos propiamente dichos. "La utilización del módulo de muestras complejas del SPSS versión 13 motivó algunas limitaciones en el tratamiento de las muestras apiladas. No hubo complicación para calcular los nuevos ponderadores pero sí para indicar tanto el nivel de solapamiento como la correlación entre las variables estimadas. Según el resultado teórico consignado en el párrafo anterior se hicieron algunas pruebas para determinar el nivel de disminución de la varianza de los estimadores utilizando este software y se obtuvieron resultados similares a los teóricos por lo que se procedió a utilizarlo para todas las estimaciones." (Muiños, 2005)

Tras la calibración del software a partir de las mencionadas peculiaridades de la muestra, se utilizó esta base de datos apilados para realizar las estimaciones de los indicadores de incidencia. El objetivo de este análisis consiste en poder brindar una estimación más estructural de la prevalencia o incidencia de déficit o florecimientos por espacio, con ventajas tales como la disminución de las variaciones estacionales surgidas en las ondas, así como también de los errores de muestreo.

En concreto, se realizó la estimación de los porcentajes de déficit para el total de los casos, evaluándose la magnitud de las diferencias por espacio. Dado el objetivo principal de este estudio, en términos de demostrar las desigualdades existentes entre espacios residenciales socioeducativos, se incluyeron en esta presentación tres medidas de diferencia, en forma de ratios (razones o cocientes) entre espacios. Estas medidas de desigualdad marcan la amplitud de la brecha existente entre los distintos espacios para cada indicador elaborado. Se seleccionaron tres medidas principales:

- Ratio VLD / ERS 4: marca la diferencia relativa entre la proporción de déficit del grupo de vulnerables (ERS 1 + ERS + ERS3) y del grupo de control. Este cociente da idea de la segmentación existente entre los espacios vulnerados y el grupo de control según el indicador analizado. Se obtiene simplemente mediante el cociente entre ambos porcentajes, lo cual redunda en una medida relativa de la brecha entre los espacios mencionados.

D Capitulo APE2 5/5/06 8:09 PM Page 427

- Ratio ERS 1 / ERS 4: marca la diferencia relativa entre la proporción de déficit del grupo Muy Bajo y del grupo de control. Este cociente permite evaluar el grado de polarización entre el espacio más vulnerable y el grupo de control. Se calcula de la misma manera que el anterior.

- Ratio ERS 1 / ERS 3: marca la diferencia relativa entre la proporción de déficit del grupo Muy Bajo y del grupo Medio Bajo. Esta brecha muestra la diferenciación entre los dos espacios extremos al interior de los vulnerables, es decir, es una medida de polarización dentro de la vulnerabilidad. Se calcula de la misma manera que el anterior (2).

Cuanto más cercanos a uno (1) sean estos ratios, menores serán las diferencias apreciadas entre los espacios. Por otro lado, cuanto más se alejen de la unidad, mayor relación tendrá el espacio con el indicador analizado. Ahora bien, la importancia de la diferencia entre las distintas incidencias de déficit según espacio no siempre resultó evidente en esta medida tomada por sí misma. Fue necesario realizar un análisis adicional que valide los resultados obtenidos a partir de la determinación de la significación por espacio.

Considerando las limitaciones del software disponible se efectuó un análisis de diferencia de proporciones para muestras independientes. Esto fue posible dado que los espacios son resultantes del diseño muestral y, por lo tanto, independientes entre sí. En consecuencia, una prueba de diferencia de proporciones de este tipo permitió determinar si las proporciones de déficit o florecimiento resultaban o no significativamente diferentes por espacio. El software proveyó la estimación de las proporciones y sus correspondientes desvíos estándar. Con estos datos se construyó en una planilla de cálculos la prueba propiamente dicha y se realizó el contraste de las hipótesis. A continuación se describe brevemente la prueba.

Disponemos de dos muestras aleatorias independientes, la primera consta de n_x observaciones de una población con una proporción de déficit de p_x y la proporción muestral de déficit resultante es p_x. La segunda tiene n_y observaciones de una población cuya proporción de déficit es p_y y la proporción muestral resultante es $\hat{p}_y$. Si el tamaño de las muestras es grande se puede considerar que la siguiente variable aleatoria tiene distribución normal estándar, de acuerdo al Teorema Central del Límite (Newbold, 1997).

$$z = \frac{(\hat{p}_x - \hat{p}_y) - (p_x - p_y)}{\sqrt{\frac{p_x(1-p_x)}{n_x} + \frac{p_y(1-p_y)}{n_y}}}$$

El objetivo es contrastar la hipótesis de que las proporciones poblacionales p_x y p_y son iguales. Es decir, las hipótesis nula y alternativa son

H_0) $p_x - p_y = 0$
H_1) $p_x - p_y \neq 0$

De acuerdo a H_0 y si el valor común de las proporciones es p_0, este estadístico es:

$$z = \frac{(\hat{p}_x - \hat{p}_y)}{\sqrt{\frac{p_0(1-p_0)}{n_x} + \frac{p_0(1-p_0)}{n_y}}}$$

La regla de decisión es rechazar H_0 a favor de H_1 toda vez que la probabilidad asociada a este estadístico con distribución normal estándar aproximada, a dos colas, sea menor que el nivel de significación elegido. Habiendo seleccionado un nivel de significación del 5% se optó por utilizar la corrección de Bonferroni para comparaciones múltiples. "Se decidió utilizar la técnica de Bonferroni para fijar el nivel de significación a utilizar, a los efectos de disminuir la probabilidad de rechazar hipótesis nulas por azar, debido a la realización de múltiples comparaciones de a dos grupos. Asimismo, antes de realizar estas comparaciones de a pares, se sugirió utilizar una comparación global que permitiera evaluar la significación global del problema. Si este test encontrara diferencias significativas globales, es adecuado determinar entre qué grupos estas diferencias se manifiestan." (Muiños, 2005). Por ende, las comparaciones que se llevan a cabo son cuatro, tres de las cuales se han mencionado anteriormente en términos de los ratios, y una que es previa:

1. Comparación entre los cuatro espacios al mismo tiempo: ERS 1 vs. ERS 2 vs. ERS 3 vs. ERS 4. Esta primera comparación da lugar a las siguientes.
2. Comparación entre VLD (ERS 1 + ERS 2 + ERS 3) y el ERS 4
3. Comparación entre el ERS 1 y el ERS 4
4. Comparación entre el ERS 1 y el ERS 3

Consecuentemente, el verdadero nivel de significación de acuerdo a la corrección de Bonferroni es:
Alfa de Bonferroni

$$\text{Alfa de Bonferroni} = \frac{\alpha}{4} = \frac{0{,}05}{4} = 0{,}0125$$

según se consigna en los cuadros correspondientes.

Luego se completó el análisis calculando la incidencia del déficit (florecimiento) de acuerdo con una serie de variables de corte, propuestas para intentar caracterizar a los sujetos u hogares que padecen las privaciones. Las variables tomadas caracterizan, en primer lugar, al encuestado. Se trata de variables principalmente de carácter socio-demográfico, como el sexo, grupos de edad, situación conyugal, nivel de educación, situación laboral; y de índole psicológica, como la comprensión verbal o el malestar psicológico.

Por otro lado, se consideraron variables descriptivas del hogar de los encuestados, como la cantidad de miembros, el ciclo de vida de las familias, la situación del núcleo conyugal y el clima educativo. Finalmente, se caracterizaron los conglomerados de pertenencia. En este sentido se evaluó el aglomerado de residencia y la homogeneidad del conglomerado barrial, a partir de datos educativos. En todos los casos, el análisis se realizó según el espacio residencial socioeducativo.

Evolución temporal de los indicadores de privación

Una vez conocida la incidencia general de las privaciones sobre los espacios residenciales socioeducativos, se pudo estudiar la evolución temporal de la misma, según fue medida a partir de los indicadores elaborados.

Con este fin, se presentaron los *stock* de déficit (florecimiento) en cada momento del tiempo. Se dio prioridad a la evolución anual, junio de 2004 a junio de 2005, toda vez que los datos lo hicieran posible. Cuando este no era el caso, se presentó la evolución semestral, diciembre de 2004 a junio de 2005. Para determinar la importancia de los cambios temporales en estos *stock* se calculó la diferencia absoluta entre las proporciones de déficit (florecimiento) en cada momento, según el espacio residencial socioeducativo.

Con el propósito de determinar la importancia de las diferencias halladas se recurrió a una prueba de significatividad. Si bien el proceso coherente con el resto del desarrollo del análisis hubiese sido la realización de una prueba de diferencia de proporciones para muestras relacionadas (3), las limitaciones del software utilizado en términos del cálculo de la covarianza entre las dos series de datos a partir del módulo de muestras complejas, impidió que se llevara a cabo esta prueba. En su reemplazo se utilizó la prueba *chi cuadrado* entre la variable dicotómica de déficit (florecimiento) y una variable indicativa de la onda de la encuesta.

La prueba de *chi cuadrado* compara a través de un estadístico las posibles diferencias entre las frecuencias observadas en la distribución de una variable y las esperadas de acuerdo a determinada hipótesis (Visauta Vinacua, 1997). En este caso, la prueba se utiliza para determinar si dos variables son independientes (hipótesis nula), o bien si están relacionadas. Es decir, si las proporciones de déficit están relacionadas con el momento del tiempo, se puede decir que existe una diferencia significativa entre los *stock* de déficit en cada período.

El valor del estadístico de prueba se obtiene de la siguiente manera:

$$\chi^2 = \sum_{i=1}^{N} \frac{(O_i - E_i)^2}{E_i}$$

donde o_i son las frecuencias observadas, E_i son las frecuencias esperadas y K el número de categorías de la variable. Cuando la muestra es grande (n>30), el estadístico se distribuye de acuerdo a una distribución χ^2 con K-1 grados de libertad, es decir, el número de categorías de la variable menos uno.

Las frecuencias esperadas equivalen a las que debería tener la tabla de contingencia si se cumpliera la hipótesis nula. Cuando se realiza la prueba para comprobar la independencia estadística, las frecuencias esperadas son aquellas que ocurrirían si las variables fueran independientes. Cuanto más parecidas sean la frecuencia esperada y la observada, el numerador será más pequeño, derivando en un estadístico más bajo, que estará asociado a una probabilidad más alta. Esto llevará a que no se pueda rechazar la hipótesis nula de independencia estadística. Si las frecuencias esperada y observada son muy diferentes, el denominador será mayor, por lo tanto, también lo será el estadístico, la probabilidad asociada será baja y esto llevará a que se rechace la hipótesis nula de independencia estadística.

Esta prueba se utiliza para variables categóricas, ordinales o nominales. Por tratarse de una prueba no paramétrica, no es necesario realizar ningún supuesto sobre la distribución subyacente. Los datos se suponen provenientes de una muestra aleatoria. La frecuencia esperada de cada categoría debería ser de por lo menos 1; no más del 20% de las categorías debería tener una frecuencia esperada menor a 5. En forma adicional, para cada momento del tiempo se realizó el análisis de las brechas entre espacios, a partir de los ratios antes descriptos.

Análisis de las trayectorias de cambio en las situaciones de privación

Para completar el análisis es necesario estudiar los flujos de cambio en la incidencia de las situaciones de privación. Este tipo de exploración requiere la construcción de una base de datos de panel, conformada por aquellos individuos que respondieron más de una de las EDSA realizadas. En términos de la estructura de la encuesta se analizaron dos conjuntos principales de casos. En primer lugar, se analizaron los datos de los individuos que habían contestado en junio de 2004 y en junio de 2005. Este panel estuvo compuesto por un total de 715 sujetos que respondieron, por lo menos, la primera y tercera encuestas (4). Un segundo panel estuvo conformado por 765 individuos, quienes respondieron, por lo menos, la segunda y la tercera encuesta (5).

Los indicadores que se elaboraron en esta instancia tienen la categoría de trayectorias temporales. Básicamente, se trabajó con trayectorias que muestran el cambio entre dos momentos del tiempo. Siempre que los datos estuvieran disponibles, éstas se definieron para representar el flujo interanual de los individuos entre situaciones deficitarias (de florecimiento) o no deficitarias (de no florecimiento). La construcción anual permitió, por un lado, evitar posibles problemas de estacionalidad en los datos, además de admitir el transcurso de un período suficiente para observar cambios significativos. Si los datos no estaban disponibles, como puede ser el caso de una pregunta que fuera reiterada en las ondas 2 y 3 de la encuesta, la trayectoria se elaboró en forma semestral. En todos los casos, se clasificó a las

D Capitulo APE2 5/5/06 8:09 PM Page 431

unidades en las siguientes trayectorias conforme a la situación observada en el momento de inicio (junio de 2004) y en el momento de llegada (junio de 2005):

(a) **se mantuvo en la situación no deficitaria:** el sujeto no se encontró en situación de déficit en ninguno de los dos momentos,
(b) **salió de la situación deficitaria:** se hallaba en la situación deficitaria en junio de 2004 y en la situación no deficitaria en junio de 2005,
(c) **entró en la situación deficitaria:** no tenía déficit inicialmente, mientras que estaba en situación deficitaria en junio de 2005,
(d) **se mantuvo en la situación deficitaria,** se halló en situación deficitaria en los dos momentos.

Figura AII.1 Construcción de las trayectorias de cambio en las situaciones de privación

		Junio de 2005	
		Déficit	No déficit
Junio de 2004	Déficit	Se mantuvo en la situación deficitaria	Salió de la situación deficitaria
	No déficit	Entró en la situación deficitaria	Se mantuvo en la situación no deficitaria

Fuente: EDSA, Observatorio de la Deuda Social. UCA.

Para determinar si cualquiera de estas situaciones ocurría de manera diferencial en alguno de los espacios se aplicaron pruebas de diferencia de proporciones para muestras independientes, de la misma manera que la descripta anteriormente.

Para la determinación de las probabilidades de transición entre las distintas situaciones se calcularon tasas específicas de cambio, relacionadas a los procesos de entrada y salida de la situación de deficitaria. En todos los casos las diferencias entre los espacios residenciales de vulnerabilidad (ERS 1 + ERS 2 + ERS 3) y los espacios medios altos (ERS 4) se midieron mediante los mismos ratios antes mencionados, cuya significación se estimó con pruebas de diferencias de proporciones. La identificación de los determinantes de las trayectorias fue efectuada mediante la aplicación de la técnica de regresión logística multinomial, por medio de la cual se buscó establecer el efecto neto de los espacios residenciales socioeducativos (ERS) sobre los cambios ocurridos. A continuación se presentan con mayor detalle las características del método de regresión empleado (Salvia, 2005).

D Capitulo APE2 5/5/06 8:09 PM Page 432

Estimación de las probabilidades de cambio en las situaciones de privación a partir de modelos de regresión logística multinomial

Cuando se requiere estimar la presencia o ausencia de una característica o resultado según los valores de un conjunto de variables explicativas se puede recurrir a los modelos de elección. Estos modelos son básicamente de dos tipos: (i) Modelos de elección binaria, cuando la variable dependiente tiene dos categorías, en general asociadas a la presencia o ausencia de algún atributo; (ii) Modelos de elección múltiple, cuando la variable dependiente es categórica, pero posee más de dos alternativas posibles que pueden estar ordenadas o no ordenadas. "Para los casos de estudio de la movilidad social se ha sugerido utilizar el modelo de regresión logística múltiple que permitirá calcular probabilidades de movilidad de sujetos de acuerdo a distintos escenarios provistos por las propias variables utilizadas en el estudio." (Muiños, 2005).

Los modelos de elección múltiple (Greene,1998) implican la elección de una sola alternativa entre varias. Una virtud reconocida de estos modelos es la capacidad de predecir la fuerza y el sentido explicativo de cada una de las independientes consideradas en el modelo sobre las categorías de la variable dependiente, manteniendo constante el resto de los factores (Aldrich y Forrest, 1984; Agreste, 1990).

El objetivo principal de la estimación de este tipo de modelos consiste en representar las probabilidades de que un individuo seleccione una de esas alternativas, posea ese atributo o pertenezca a esa categoría, según sea la variable dependiente. Para estimar estas probabilidades es necesario realizar algún supuesto sobre su distribución, que no debería permitir valores negativos o mayores que uno. Por esta razón, en general se seleccionan distribuciones logísticas o normales, derivando en modelos logísticos o probit. El modelo probit es de muy difícil tratamiento, razón por la cual se utilizó el modelo logístico, en concordancia con la mayor parte de la bibliografía relacionada.

En tal sentido, el modelo logístico multinomial permite determinar las probabilidades de clasificación de los individuos en función de los valores de una serie de variables predictoras cuando la variable dependiente comporta más de dos categorías. Se trata de un modelo similar al modelo de regresión logística binomial, pero la variable a explicar, en lugar de ser dicotómica, puede adquirir más de dos valores no ordenados. Entre las variables explicativas podemos contar con un grupo de variables que describan las características del individuo, y que por lo tanto sean iguales para todas las respuestas posibles.

Podemos expresar el modelo logístico multinomial como a continuación se detalla. La probabilidad de la realización j es:

$$P(Y = j) = \frac{e^{\beta_j x_i}}{1 + \sum_{k=1}^{J} e^{\beta_k x_i}}$$

D Capitulo APE2 5/5/06 8:09 PM Page 433

Mientras que la probabilidad de que no sea j es:

$$P(Y=0)=\frac{1}{1+\sum_{k=1}^{J} e^{\beta' z_k}}$$

donde z_{ij} es en ambos casos la matriz de variables explicativas utilizadas en el modelo.

Para el análisis de las estimaciones se pueden comparar las probabilidades, considerando Y = 0 como la categoría de comparación, con el fin de determinar qué alternativa es más probable. Es posible, entonces, tomar el cociente de las probabilidades antes presentadas:

$$\frac{P_{ij}}{P_{i0}}=e^{\beta' x_i}$$

Si se aplican logaritmos a esta expresión, se obtiene:

$$\ln\left[\frac{P_{ij}}{P_{i0}}\right]=\beta'_j x_i$$

El análisis de estos cocientes de probabilidades para cada categoría de las variables explicativas es ilustrativo de cómo cada variable contribuye al incremento de las probabilidades de cada alternativa.

En este documento, la regresión logística multinomial se utilizó con el objetivo de indagar sobre los determinantes de las trayectorias de déficit (florecimiento) según fueron definidas anteriormente. Por lo tanto, se analizó la probabilidad de que un sujeto o un hogar transcurra por alguna de las trayectorias, de acuerdo a una serie de atributos que los caracterizan. Las variables que fueron seleccionadas se presentan en la Figura AII.2.

Luego de realizada la estimación, que se llevó a cabo para cada espacio en forma separada (si bien considerando los mismos atributos antes mencionados), se efectuó el cálculo de las probabilidades de cada trayectoria, según espacio residencial socioeducativo y para cada atributo. Las probabilidades estimadas se presentaron en forma gráfica para facilitar su interpretación. Este tipo de análisis permitió evaluar cuál es el impacto diferencial de cada espacio sobre las probabilidades de transcurrir una u otra trayectoria de privación, controlando por los efectos que pudieran ejercer los demás atributos.

D Capitulo APE2 5/5/06 8:09 PM Page 434

Figura AII.2 Atributos seleccionados para la estimación de las probabilidades de cambio en las situaciones de privación

Atributos Individuales

Variable	Categoría
Sexo	*Varón* Mujer
Grupos de edad	De 18 a 34 años De 35 a 54 años *55 años y más*
Situación conyugal	*Soltero* Casado o unido de hecho Separado, divorciado o viudo
Nivel educativo	Años de escolaridad
Malestar Psicológico	Déficit *No déficit*
Comprensión verbal	Déficit *No déficit*
Antigüedad en el barrio	Hasta 5 años De 6 a 15 años *Más de 15 años*
Condición de migración	Migrante interno o externo *No migrante*

Atributos del hogar

Variable	Categoría
Tamaño	Unipersonales 2 a 4 componentes *5 o más componentes*
Ciclo vital del hogar	Familias en etapa inicial o con hijos pequeños Familias con hijos adolescentes o mayores Familias nido vacío *Hogar no familiar*
Tipo de núcleo conyugal	Familias monoparentales Núcleo completo *Resto de los hogares*
Clima educativo del hogar	Promedio de los años de educación de los miembros del hogar mayores de 19 años

Atributos del contexto residencial

Variable	Categoría
Homogeneidad del conglomerado barrial	Índice
Regiones metropolitanas	AMBA *Ciudades del Interior*

Nota: Las categorías utilizadas como comparación fueron marcadas en cursivas.

Fuente: EDSA, Observatorio de la Deuda Social, UCA.

Notas del capítulo

(1) Este software resultó muy útil en ciertos aspectos, mientras que en otros presentó ciertas restricciones que obligaron a cambiar marginal o significativamente la estrategia de análisis.

(2) Debe tenerse en cuenta que estos ratios se encuentran así definidos para los análisis de indicadores de déficit. En el caso de los indicadores de florecimiento, la comparación es la inversa, queriéndose distinguir cuánto mejor es la situación del grupo de control por sobre los demás ERS. Por lo tanto, los ratios son: ERS 4 / VLD, ERS 4 / ERS 1, ERS 3 / ERS 1.

(3) Las muestras son relacionadas porque se evalúa los mismos sujetos en dos momentos del tiempo.

(4) Se trata de 662 individuos que respondieron las tres ondas, sumados a 53 personas que respondieron la primera y la tercera.

(5) Son 662 individuos que respondieron las tres ondas, sumados a 103 sujetos que contestaron en la segunda y tercera medición.