ANEXO METODOLÓGICO 4

MÉTODOS Y TÉCNICAS DE ANÁLISIS ESTADÍSTICO UTILIZADOS EN EL INFORME

1. MEDIDAS DE DESARROLLO HUMANO Y SOCIAL

Los Índices de Desarrollo Humano y Social se expresan en una escala de calificación de 0 a 10 puntos, donde 0 representa la máxima distancia a los umbrales normativos mínimos y 10 expresa el acceso a las condiciones establecidas por estos umbrales. Los índices se exponen mediante el cálculo de promedios de estas calificaciones, para los totales y las categorías de las variables de corte seleccionadas.

Como medida de las diferencias existentes entre los grupos determinados por cada variable de corte, se calcularon diferencias de medias. Se trata de la diferencia absoluta, obtenida mediante la sustracción de los valores promedio del índice para cada categoría respecto de la categoría seleccionada como comparación. Poniendo como ejemplo el corte según estrato socioeconómico, la diferencia se realiza restando el valor promedio del índice en el estrato muy bajo menos el del medio alto (categoría de referencia). Al tratarse de una diferencia absoluta, está expresada en la misma unidad de medida que los Índices y, por lo tanto, puede interpretarse como puntos de diferencia entre calificaciones. En el ejemplo, esta medida será igual a cero (DM=0) cuando el promedio en la categoría de la variable de corte y su categorías de referencia sea el mismo, lo cual implica una ausencia de asociación entre la estructura socioeconómica y el índice analizado. Valores mayores que cero (DM>0) sugieren una asociación directa entre el estrato socioeconómico y el índice respectivo, mientras que valores menores que cero (DM<0) suponen una relación inversa entre el estrato socioeconómico y el índice.

Con el objetivo de determinar la importancia de estas diferencias se aplicaron pruebas de significatividad. La prueba elegida para la evaluación de la significatividad de las diferencias es la prueba de diferencia de medias para muestras independientes. Se trata de muestras independientes puesto que las categorías de los grupos utilizados para la clasificación, tanto de los estratos socioeconómicos, de los conglomerados metropolitanos, como de las demás variables de corte, son independientes entre sí.

A continuación se describe brevemente la prueba de diferencia de medias para muestras independientes. Si disponemos de dos muestras aleatorias independientes, la primera consta de n_x observaciones de una población con una media de μ_x y la media muestral resultante es $\overline{X}$. La segunda tiene n_y observaciones de una población cuya media es μ_y y la media muestral resultante es $\overline{Y}$. El objetivo es contrastar la hipótesis de que las medias poblacionales μ_x y μ_y son iguales. Es decir, las hipótesis nula y alternativa son:

H_0) $\mu_x = \mu_y$
H_1) $\mu_x \neq \mu_y$.

De acuerdo a H_0, el estadístico de prueba:

$$z = \frac{(\overline{X} - \overline{Y})}{\sqrt{\frac{\sigma_X^2}{n_x} + \frac{\sigma_Y^2}{n_y}}}$$

tiene distribución normal estándar. Sin embargo, las varianzas poblacionales son desconocidas. Si se utilizan las varianzas muestrales, esta variable aleatoria sigue una distribución t de Student con ($n_x + n_y$ - 2) grados del libertad. La regla de decisión es rechazar H_0 a favor de H_1 toda vez que la probabilidad asociada a este estadístico sea menor que el nivel de significación elegido. Cuando los tamaños muestrales son grandes, la aproximación normal es buena incluso al sustituir las varianzas poblacionales por las muestrales (Newbold, 1997).

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

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

Donde k es la cantidad de comparaciones realizadas.

2. INDICADORES DE DÉFICIT

El estudio de cada dimensión del Desarrollo Humano y Social se encaró a través de una serie de indicadores elaborados a partir de las preguntas del cuestionario de la EDSA. Estos indicadores se expresaron de manera cualitativa como variables dicotómicas 0-1, donde la unidad representa la presencia de privaciones en el contenido al que refiere el indicador.

A partir de estos indicadores se calcularon tasas de recuento, que se computan de manera muy simple como el cociente entre la cantidad de casos que poseen determinado atributo y el total de casos observados. Se analizó la incidencia de los indicadores, tanto en forma general como de manera diferencial para cada criterio de clasificación (estratos socioeconómicos, conglomerados urbanos y demás características seleccionadas).

Como una medida de las diferencias existentes entre las categorías de las variables de corte, se calculó el Riesgo Relativo de que el hogar/persona estuviera en situación de privación en relación con la categoría de referencia elegida. El riesgo relativo es un ratio (razón o cociente) entre las incidencias de cada indicador para cada grupo. Es una medida relativa de interpretación directa, que indica cuántas veces mayor es la incidencia de determinado indicador en un grupo respecto del otro. Su magnitud da

idea de la desigualdad existente respecto del indicador entre los grupos comparados. Como ejemplo, se puede comparar el estrato socioeconómico Muy bajo con el Medio alto. Cuando el riesgo relativo es igual a uno (RR=1) la incidencia de un indicador en ambos grupos es la misma, lo cual sería signo (en el ejemplo) de la falta de asociación entre la estructura socioeconómica y el indicador analizado. Valores mayores que 1 (RR>1), suponen una asociación directa (positiva) entre la estructura socioeconómica y el indicador analizado, mientras que valores menores que 1 (RR<1), sugieren una relación inversa (negativa) entre la estratificación socioeconómica y el indicador medido.[30]

Al mismo tiempo, con el objetivo de determinar la importancia de las diferencias entre las incidencias según estratos socioeconómicos o conglomerado metropolitanos, se aplicaron pruebas de significatividad. La prueba elegida para la evaluación de la significatividad de las diferencias es la prueba de diferencia de proporciones para muestras independientes. La selección de este método se basó en que los indicadores de la EDSA son elaborados como tasas de recuento y, por lo tanto, constituyen proporciones. Además, las categorías de los grupos utilizados para la clasificación, estratos socioeconómicos, conglomerados metropolitanos y demás variables de corte, son independientes entre sí. En consecuencia, una prueba de diferencia de proporciones de este tipo permiten determinar si las incidencias de cada indicador resultan o no significativamente diferentes para los diferentes grupos.

30 Por tratarse de una medida relativa, el riesgo relativo no está expresado en las unidades de medida de las variables que compara. Esto constituye una ventaja para su interpretación, dado que lo hace comparable entre variables o indicadores. Es decir, si un indicador cualquiera muestra un riesgo relativo alto entre los grupos mencionados esto es signo de desigualdad entre ellos. Otro indicador que tenga un riesgo relativo menor a aquel en el mismo momento, mostrará menor desigualdad, no importa lo diferentes que sean las dimensiones que estos indicadores midan.

A continuación se brinda una breve descripción de la prueba de diferencia de proporciones para muestras independientes. Si disponemos de dos muestras aleatorias independientes, la primera consta de n_x observaciones de una población con una proporción de p_x y la proporción muestral resultante es $\hat{p}_x$. La segunda tiene n_y observaciones de una población cuya proporción es p_y y la proporción muestral resultante es $\hat{p}_y$. 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 , el estadístico de prueba es:

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

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. Nuevamente se aplicó la corrección de Bonferroni, seleccionando un nivel de sigificatividad del 5% (Newbold, 1997, op.cit.).

3. VARIABLES DE CORTE SELECCIONADAS

Tanto las medidas de desarrollo humano y social como los indicadores de déficit utilizados para su construcción fueron analizados de acuerdo a su distribución según ciertas variables de corte, seleccionadas debido a su carácter estructural y a su importancia en la explicatividad de los resultados.

A continuación se presenta una breve descripción de las variables elegidas.

CARACTERÍSTICAS DE LA PERSONA	
SEXO	Se refiere al sexo del entrevistado. Las categorías posibles son varón y mujer. Se utiliza la primera como categoría de referencia para las comparaciones.
EDAD	Se refiere a la edad del entrevistado. Se ha organizado en tres grupos de edad: de 18 a 34 años, de 35 a 59 años, 60 años y más. Se toma como categoría de referencia a las edades adultas.
NIVEL DE EDUCACIÓN	Se refiere a la educación del entrevistado. Se diferencian cuatro categorías: niveles menores al primario completo, escuela primaria completa, escuela secundaria completa, educación superior completa. Esta última opera como categoría de comparación.
POSICIÓN EN EL HOGAR	Toma el valor Jefe cuando el entrevistado es el jefe de hogar, o No jefe cuando el entrevistado es cualquier otro componente. La primera es la categoría de comparación.
CARACTERÍSTICAS DEL HOGAR	
TIPO DE HOGAR	Se considera hogar familiar no monoparental a aquellos hogares compuestos por un núcleo conyugal completo, ya sea con o sin hijos. Se considera hogar familiar monoparental a aquellos hogares compuestos por un núcleo conyugal incompleto, con la presencia de hijos solteros. Finalmente, son hogares no familiares los hogares unipersonales o multipersonales compuestos por personas parientes o no parientes. La primera categoría se utiliza para la comparación.
SEXO DEL JEFE DEL HOGAR	Se refiere al sexo del miembro del hogar que el entrevistado declara como jefe. Las categorías son varón y mujer, siendo la primera la categoría de comparación.
ESTRATO SOCIOECONÓMICO	Se consideran cuatro estratos socioeconómicos que representan niveles crecientes de capital educativo y de acceso a tecnologías de información y comunicación (TIC): Muy bajo, Bajo, Medio bajo, Medio alto. Este último estrato se utiliza como categoría de comparación. La construcción de esta variable se considera con detalle en el apartado 4 de este mismo anexo.
CARACTERÍSTICAS DEL CONGLOMERADO	
CONGLOMERADO URBANO	Se toman dos grupos de aglomerados principales, de los cuales la EDSA es representativa a lo largo de todo el período analizado, Gran Buenos Aires y ciudades del interior. Adicionalmente, en 2007 se desagregan las grandes ciudades del interior: Rosario, Córdoba y Mendoza. La categoría de comparación es GBA.

4. CLASIFICACIÓN DE LOS CASOS PARA LA PRESENTACIÓN DE LOS RESULTADOS

El análisis de los datos se realizó a partir de una clasificación de los datos en términos del conjunto de variables de corte descriptas en el apartado anterior. Sin embargo, dos de los criterios de clasificación se consideran principales. Por un lado, el criterio regional, que clasifica los casos según el conglomerado urbano donde estos fueron relevados. En particular, se subdividen los resultados entre el Gran Buenos Aires (GBA) y las grandes Ciudades del Interior del país.

El segundo criterio de clasificación de importancia es el Estrato Socioeconómico del Hogar encuestado. Este criterio considera los activos principales del hogar en dos niveles principales: atributos propios del hogar y atributos del entorno residencial donde éste se encuentra localizado. Para el nivel del hogar se utilizó la información de la EDSA, empleándose los datos brindados por el encuestado acerca de los demás miembros del hogar y de la vivienda en la que habitan. A nivel del entorno residencial, se optó por utilizar datos del Censo de Población de 2001 para incorporar el perfil del vecindario en donde se ubica el hogar mediante la

identificación de las características del radio censal al que éste pertenece.

La construcción del estrato socioeconómico implicó una serie de decisiones metodológicas. En particular, el esquema clasificatorio propuesto procuró integrar los atributos del hogar y del vecindario en materia de dotación de capital educativo y de acceso a las tecnologías de información y comunicación (TIC). Las variables utilizadas para la definición de estos aspectos fueron:

• **Capital educativo del vecindario:** definido operativamente como el porcentaje de jefes de hogar con secundaria completa o más en el radio censal de localización del hogar.

• **Acceso a tecnologías en el vecindario:** definido operativamente como el promedio de los porcentajes de hogares en el barrio con acceso a medios de comunicación e información. Los bienes considerados son tres: televisión por cable, teléfono celular e Internet.

• **Capital educativo del hogar:** definido operativamente como el total de años de educación del jefe de hogar, como miembro más representativo del mismo.

• **Acceso a tecnologías en el hogar:** definido operativamente como un indicador compuesto del acceso que logra el hogar a cada uno de los bienes tecnológicos mencionados. Estos tres aspectos se combinaron para formar un índice, logrado como la suma simple o inventario de bienes que este posee. De esta manera, los hogares que posean sendos bienes tendrán un puntaje de 3, mientras que aquellos que no posean ninguno tendrán un puntaje de cero.

Una vez identificados, se combinaron ambos atributos del hogar y ambos atributos del vecindario en dos índices, cada uno logrado a partir de un promedio ponderado que otorga mayor peso al capital educativo (75%) por sobre el acceso a tecnologías (25%). Estos dos índices correspondientes al vecindario y al hogar fueron luego combinados entre sí, nuevamente mediante la realización de un promedio aritmético ponderado, otorgándole un 45% de peso al factor residencial y un 50% al factor del hogar. El 5% restante se asignó a una tercera variable (vivienda en villa de emergencia o asentamiento precario), representativa de un nivel intermedio referente al entorno próximo al hogar, con la intención de corregir parte de la heterogeneidad presente en el nivel barrial.

En un último paso, se generaron los cuartiles de este índice. Estos cuatro grupos ordenados se caracterizan por representar niveles crecientes de capital educativo y de acceso a tecnologías de información y comunicación, incluyendo tanto el capital individual como el capital del entorno residencial donde habitan los individuos y hogares que forman parte del estudio[31] (DII-ODSA, 2007). Los cuatro grupos resultantes quedaron clasificados en: 1) Estrato Muy Bajo, 2) Estrato Bajo, 3) Estrato Medio Bajo y 4) Estrato Medio Alto (ver cuadro AM4.1).

31 La decisión de definir los grupos para la presentación tabular utilizando cuartiles supone una clasificación relativa, razón por la cual la ubicación de cada individuo en la escala depende de la posición de los demás individuos encuestados.

CUADRO AM 4.1
Variables utilizadas para la clasificación según estrato socioenconómico.

	ATRIBUTOS DEL HOGAR		ATRIBUTOS DEL VECINDARIO					
	CAPITAL EDUCATIVO	ACCESO A TIC	CAPITAL EDUCATIVO	ACCESO A TIC				
	Años de educación del jefe (promedio)	Inventario de bienes tecnológicos (promedio)	Jefes con secundaria completa en el vecindario (%)	Hogares con teléfono celular en el vecindario (%)	Hogares con TV con cable en el vecindario (%)	Hogares con internet en el vecindario (%)	Promedio (%)	Hogares ubicados en villa de emergencia
2004								
• MUY BAJO	5,1	0,3	10,8	13,0	33,2	2,1	16,1	27,8
• BAJO	7,9	0,5	21,7	21,5	42,2	5,7	23,1	9,3
• MEDIO BAJO	11,5	0,9	30,0	26,8	48,7	8,7	28,0	2,6
• MEDIO ALTO	14,0	1,8	60,5	46,6	70,6	31,3	49,4	0,0
2005								
• MUY BAJO	5,3	0,4	11,3	13,4	33,4	2,2	16,3	28,7
• BAJO	8,0	0,7	21,3	21,2	42,6	5,7	23,2	12,8
• MEDIO BAJO	11,4	1,2	29,8	26,7	48,1	8,7	27,8	2,3
• MEDIO ALTO	14,1	1,9	60,4	46,4	70,0	31,1	49,1	0,0
2006								
• MUY BAJO	5,8	0,4	11,8	13,8	32,6	2,5	16,3	22,7
• BAJO	8,1	1,0	23,4	22,4	45,1	6,6	24,7	9,8
• MEDIO BAJO	10,6	1,4	39,1	32,0	54,3	14,8	33,6	1,1
• MEDIO ALTO	13,9	2,3	59,4	44,2	68,6	29,8	47,5	0,5
2007								
• MUY BAJO	5,5	0,6	12,5	14,7	34,5	2,6	17,3	19,4
• BAJO	8,3	1,1	23,1	22,8	44,7	6,5	24,6	9,5
• MEDIO BAJO	10,5	1,8	37,7	31,2	52,6	13,7	32,5	0,5
• MEDIO ALTO	14,2	2,3	59,9	44,5	68,0	29,6	47,3	0,0

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

CUADRO AM 4.2
Distribución de los casos de la EDSA según estrato socioeconómico. *Años 2004/2007*

	Serie histórica				Muestra ampliada[2]
	Año 2004	Año 2005	Año 2006	Año 2007[1]	Año 2007
CANTIDAD DE CASOS					
TOTAL	**1.063**	**1.063**	**1.449**	**1.740**	**2.517**
• MUY BAJO	279	300	295	356	524
• BAJO	308	280	344	411	545
• MEDIO BAJO	223	228	372	426	580
• MEDIO ALTO	253	255	438	547	868
DISTRIBUCIÓN PONDERADA DE LOS CASOS (sin ajustar - en porcentaje)					
TOTAL	**100**	**100**	**100**	**100**	**100**
• MUY BAJO	34	34	25	25	25
• BAJO	32	31	25	25	25
• MEDIO BAJO	20	21	25	25	25
• MEDIO ALTO	14	14	25	25	25

[1] Los resultados no incluyen las ciudades de Paraná y Rosario.
[2] Los resultados incluyen las ciudades de Paraná y Rosario.
Fuente: EDSA, Observatorio de la Deuda Social. UCA.

Este mismo procedimiento se repitió para cada medición. Sin embargo, el factor para el cual se deben obtener los cuatro grupos no es estrictamente comparable.[32] En particular, esto se debe a que los casos agregados con la ampliación de la muestra llevada a cabo en junio de 2006 y de 2007 presentan –por definición– capitales educativos y accesos tecnológicos (a nivel del hogar y del entorno residencial) superiores a las medias observadas de las mediciones anteriores. Para superar esta dificultad, se optó por calcular el factor de ordenamiento socioeconómico de los hogares para las mediciones anteriores, agregando los casos adicionados en junio de 2007, manteniendo constantes los valores observados en las variables de clasificación. Es decir, los valores exactos de junio de 2007 de capital educativo y de acceso a tecnologías del vecindario y del hogar fueron replicados para las mediciones anteriores. Este procedimiento permitió disponer de 1740 casos (considerando únicamente los aglomerados comparables) en todas las ondas. Luego de realizar la reclasificación de los casos de todas las mediciones, definiendo cuatro grupos a partir del cálculo de cuartiles, se eliminaron del análisis las observaciones agregadas.

Este tipo de procedimiento supone que los grupos construidos constituyen cuartiles, en tanto que dividen a la muestra (ponderada) en cuatro partes iguales, únicamente para los dos últimos relevamientos. En las mediciones iniciales existe una menor representación, en términos de casos relevados, de los estratos socioeconómicos más altos (ver cuadro AM4.2). Este problema fue encarado mediante una modificación adicional sobre los ponderadores, logrando que los casos correspondientes a estratos más altos tuvieran un peso mayor y los correspondientes a estratos más bajos un peso menor, hasta que cada estrato representara un cuarto del total de la muestra.

5. VARIACIONES TEMPORALES

Con el fin de estudiar la evolución temporal de los Índices de Desarrollo Humano y Social y de los indicadores utilizados para su construcción, se ana-

32 El problema de la comparabilidad no reside en el diferente número de casos de las diversas mediciones, sino que surge porque la muestra original no cubría todo el espectro socioeconómico, pues dejaba fuera un estrato ("clase media").

lizaron los cambios netos a lo largo del tiempo. El cálculo de estos cambios se realizó para la serie histórica de valores comparables[33], correspondientes a los valores de 2004 a 2007, sin incluir los aglomerados Gran Rosario y Paraná, que fueran incorporados durante la última medición.

Para determinar la importancia de los cambios temporales en cada serie de resultados se calcularon dos tipos de diferencias. En el caso de los Índices de Desarrollo Humano y Social, se calcularon diferencias relativas.[34] Para los indicadores utilizados en la construcción de estos índices, se aplicaron diferencias absolutas[35] entre las proporciones observadas en cada momento. En cada caso, se analizaron dos tipos de variaciones en el tiempo:

• **Variaciones interanuales**: muestra la variación (aumento o disminución) en el valor entre un año y el siguiente, a lo largo de todo el período analizado.

• **Variaciones respecto del año base**: Junio de 2004 es la medición utilizada como Línea de Base para la EDSA. Por este motivo, se calculan las variaciones ocurridas entre este año y cada año posterior.

Con el propósito de determinar la importancia de las variaciones temporales en los niveles de incidencia se recurrió a una prueba de significatividad. En el caso de los Índices de Desarrollo Humano y Social se aplicaron diferencias de medias. En el caso de los indicadores de privación, se utilizaron diferencias de proporciones. Si bien el proceso coherente con el resto del desarrollo del análisis hubiera sido la realización de pruebas para muestras relacionadas[36], 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 utilizaron las pruebas para muestras independientes descriptas anteriormente.

6. CURVAS DE INCIDENCIA DE LAS CALIFICACIONES DE LOS ÍNDICES AGREGADOS DE DESARROLLO HUMANO Y SOCIAL

Ya se ha hecho referencia a este análisis en el Recuadro 1.1. En cada uno de los gráficos presentados, el eje vertical representa la medida de incidencia correspondiente a cada calificación del Índice respectivo. De manera más concreta, dado un puntaje cualquiera del índice I_X (dentro del rango posible, 0-10) cada curva representa el porcentaje de personas u hogares que posee una calificación igual o menor a I_X. De esta manera se puede observar de forma inmediata la incidencia relativa a todos y cada uno de los puntajes, sin la necesidad de establecer un umbral determinado. Asimismo, dado un umbral de privación definido por algún criterio (teórico o empírico), este tipo de gráficos permiten realizar la evaluación crítica de la efectividad del valor seleccionado.

Las curvas de incidencia permiten, a su vez, la comparación de los niveles de incidencia de privaciones relativas en dominios establecidos; en el caso de este informe: períodos, estratos socioeconómicos y aglomerados urbanos. La condición de dominancia de primer orden indica, para un rango determinado de calificaciones, que la privación relativa en un dominio es más elevada (o más baja) si su curva de incidencia se halla siempre por arriba (o por debajo) de la curva de otro dominio.

33 Ver punto 4 del Apéndice Metodológico 1: La Encuesta de la Deuda Social Argentina (2004-2007) de este volumen.

34 Las diferencias relativas en el tiempo se calculan como $[(Valor_t / Valor_{t-k}) - 1] \cdot 100$, donde t representa el período actual y k la cantidad de períodos transcurridos entre los dos momentos del tiempo que se quieren evaluar.

35 Las diferencias absolutas en el tiempo se calculan como $Valor_t - Valor_{t-k}$.

36 Las muestras son relacionadas porque se evalúan, en cierta proporción, los mismos sujetos u hogares en dos momentos del tiempo.