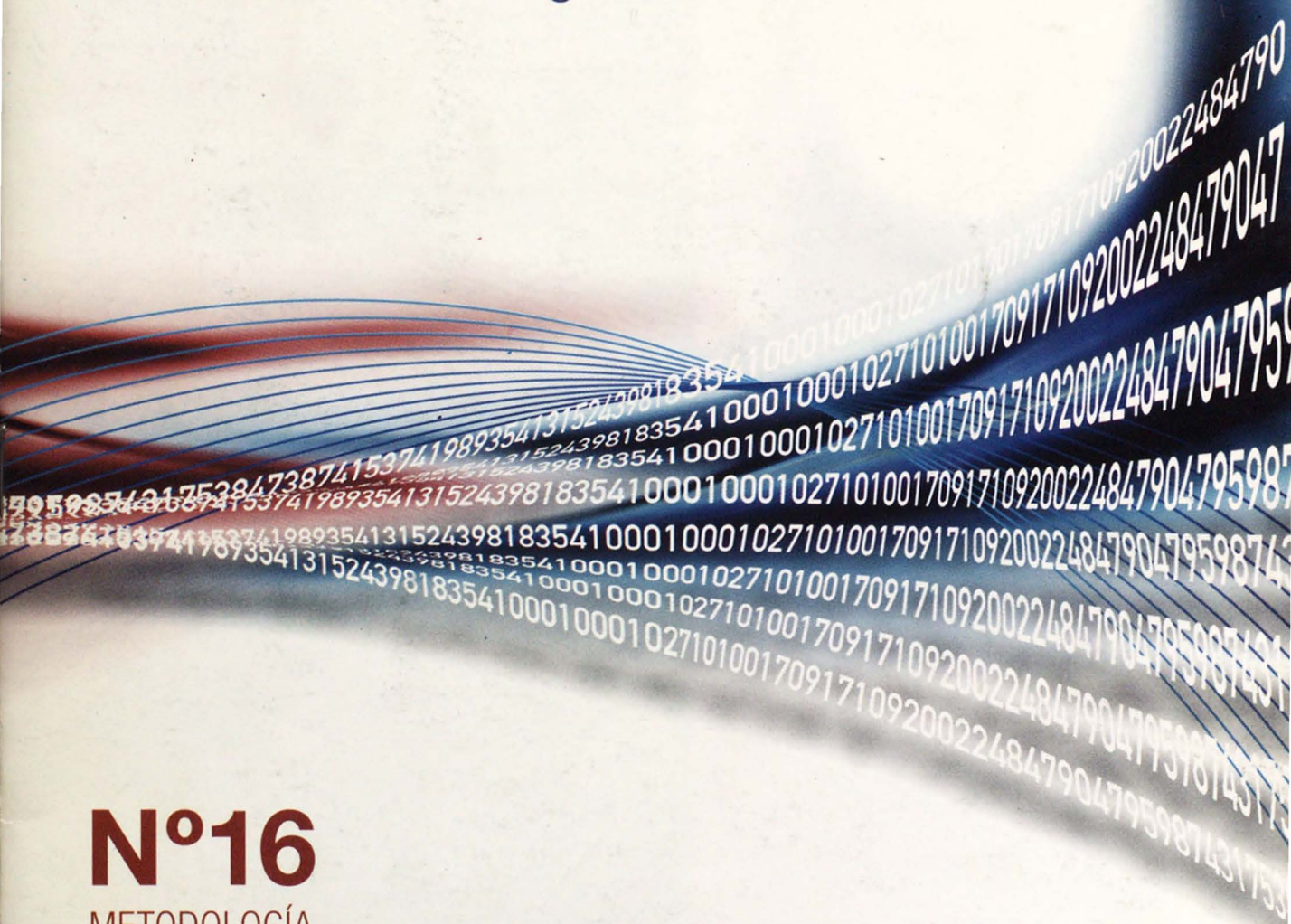


Diseño muestral y factores de expansión en encuestas a viviendas, hogares y población: el caso de la Encuesta Nacional de Factores de Riesgo 2009



N°16

METODOLOGÍA



INDEC

DIRECCIÓN NACIONAL
DE ESTADÍSTICAS DE
CONDICIONES DE VIDA
COORDINACIÓN DE
ENCUESTAS ESPECIALES

Diseño muestral y factores de expansión en encuestas a viviendas, hogares y población: el caso de la Encuesta Nacional de Factores de Riesgo 2009



Diseño muestral y factores de expansión en encuestas a viviendas, hogares y población: el caso de la Encuesta Nacional de Factores de Riesgo 2009

La presente publicación fue desarrollada por la Coordinación de Encuestas Especiales a cargo del Lic. Alejandro A. Moyano.

Participaron en su realización Leonardo Parodi, el Lic. Mariano Zaragoza, la Lic. Luciana Ureta y Juan Pablo Pilorget, bajo la coordinación de la Lic. Dolores Ondarsuhu.

Signos convencionales utilizados por el INDEC

Para la sustitución o complementación de un dato numérico, el INDEC utiliza los siguientes signos, según el caso:

- * Dato provisorio
- Dato igual a cero
- Dato ínfimo, menos de la mitad del último dígito mostrado
- . Dato no registrado
- ... Dato no disponible a la fecha de presentación de los resultados
- /// Dato que no corresponde presentar debido a la naturaleza de las cosas o del cálculo
- s Dato confidencial por aplicación de las reglas del secreto estadístico
- e Dato estimado por extrapolación, proyección o imputación

Diseño muestral y factores de expansión en encuestas a viviendas, hogares y población : el caso de la encuesta nacional de factores de riesgo 2009.

- 1a ed. - Buenos Aires : Inst. Nacional de Estadística y Censos - INDEC, 2010.

36 p. ; 19x25 cm. - (Metodología; 16)

ISBN 978-950-896-404-5

1. Estadísticas. 2. Metodología Estadística.

CDD 310

Fecha de catalogación: 15/09/2010

© Queda hecho el depósito que fija la Ley N° 11.723. Buenos Aires 2010

Permitida la reproducción parcial con mención de la fuente

Responsable de la edición: Lic. Ana María Edwin

Diseño de Tapa: DG Esteban Solari

Revisión a cargo de : Silvia Marano

Publicaciones del INDEC

Las publicaciones editadas por el Instituto Nacional de Estadística y Censos se encuentran a la venta en INDEC, Centro Estadístico de Servicios, Av. Presidente Julio A. Roca 609, C1067ABB, Ciudad Autónoma de Buenos Aires, Argentina. El horario de atención al público es de 9:30 a 16:00. También puede comunicarse a los teléfonos (011) 4349-9652/54/62, al fax (011) 4349-9621, o a través del correo electrónico: ces@indec.mecon.gov.ar, o a la página de Internet: <http://www.indec.gov.ar>

PERMITIDA LA REPRODUCCIÓN PARCIAL CON MENCIÓN DE LA FUENTE

Contenido

1. Presentación
2. Diseño muestral
 - 2.1 Población objetivo y Dominios de estimación
 - 2.2 El Marco de Muestreo Nacional de Viviendas (MMNV)
 - 2.3 Etapas de muestreo
 - 2.4 Factores de expansión
3. Características de la muestra
 - 3.1 Rendimiento de la muestra
4. La No Respuesta
 - 4.1 Tratamiento de la No Respuesta
 - 4.2 Tasas de No Respuesta
 - 4.3 Corrección de la No Respuesta
5. Expansión de la muestra y calibración de los factores de expansión
6. Bibliografía

1

Presentación

El desarrollo de los proyectos de Encuestas Especiales implica una serie de tareas y actividades simultáneas y/o encadenadas que se enmarcan en la praxis general del campo de la Metodología de la Investigación (Hernández Sampieri, 2003; Sabino, 1996; Samaja, 1997; Vieytes, 2004).

Nos detendremos en este informe técnico en algunas de esas actividades que desembocan en la utilización de datos poblacionales obtenidos a partir de una muestra aleatoria. Nos referimos aquí al diseño muestral y el cálculo de los factores de expansión.

Las temáticas que abordaremos en el siguiente documento estarán referidas específicamente a lo realizado en la Encuesta Nacional de Factores de Riesgo 2009 (ENFR-2009), pero puede servir también de base para cualquier Encuesta Especial desarrollada en el INDEC y demás estudios similares.

2

Diseño muestral

Las Encuestas Especiales surgen de requerimientos realizados por un Organismo Nacional, a partir del cual el INDEC desarrolla un proyecto de investigación.

Una de las primeras actividades en la preparación del proyecto es el diseño muestral. Esta actividad se realiza conjuntamente entre el Organismo solicitante y el INDEC. Participan por el Instituto la Coordinación de Encuestas Especiales y la Dirección de Metodología Estadística, siendo esta última la responsable de la selección muestral a partir del Marco de Muestreo vigente.

En estudios anteriores se trabajó con el Marco de Muestreo Nacional Urbano, con una cobertura de localidades urbanas de 2.000 y más habitantes. En el caso de la ENFR-2009, esta selección se realizó a partir del Marco de Muestreo Nacional de Viviendas (marco con una cobertura nacional urbana y rural). Esta selección incluyó un total de 47.159 viviendas en localidades de 5.000 y más habitantes de la República Argentina⁽¹⁾. En ambos casos, la clasificación de los tamaños de las localidades, así como el recorte geográfico, se corresponden con los datos del Censo Nacional de Población y Viviendas 1991.

La muestra se realizó con sobremuestreo para compensar la posible pérdida por No Respuesta. Esto implica que una vivienda seleccionada en la que no se obtiene la información buscada no es reemplazada por otra.

2.1 Población objetivo y dominios de estimación

Este proyecto tuvo una duración de doce meses, desde julio de 2009 hasta junio de 2010. La población bajo estudio fue aquella de 18 años y más que reside en hogares particulares en localidades de 5.000 o más habitantes.

La ENFR-2009 tiene una muestra cuyo tamaño total es de 47.159 viviendas, obteniéndose información para distintos dominios de estimación:

⁽¹⁾ El diseño original era de 47.170 viviendas seleccionadas, pero contenía un error de duplicación de áreas, por lo que se procedió a recalcular los factores de expansión y ajustar levemente el tamaño muestral.

- Total del país
- 4 agrupamientos según tamaños de localidades (por cantidad de habitantes):
 - de 5.000 a 100.000
 - de 100.001 a 500.000
 - de 500.001 a 1.500.000
 - más de 1.500.000
- 6 regiones estadísticas:
 - Gran Buenos Aires
 - Pampeana
 - Noroeste
 - Noreste
 - Cuyo
 - Patagónica
- División política:
 - 24 jurisdicciones del país
 - 7 aglomerados urbanos de más de 500.000 habitantes:
 - Gran Buenos Aires
 - Gran Rosario
 - Gran Córdoba
 - Gran La Plata
 - Mar del Plata-Batán
 - Gran Mendoza
 - Gran Tucumán – Tafí Viejo

2.2 El Marco de Muestreo Nacional de Viviendas (MMNV)

El Marco de Muestreo Nacional de Viviendas es una muestra probabilística maestra de áreas y departamentos, cuyo dominio es el total urbano-rural del país. La selección de las áreas fue realizada mediante muestreo estratificado y polietápico y el listado de las viviendas de éstas se encuentra ingresado en soporte informático.

Esto permite luego seleccionar muestras de viviendas según diferentes criterios en lo que hace a tamaño y dominios de estimación. Partimos del supuesto de que el listado de las viviendas de estas áreas se mantiene actualizado, por lo que las estimaciones correspondientes no deberían presentar sesgos.

El Marco anteriormente utilizado para las encuestas a hogares es el Marco de Muestreo Nacional Urbano (MMNU) del que se dispone una publicación con información exhaustiva (INDEC, Marco de Muestreo Nacional Urbano para Encuestas a Hogares, Metodologías nº 12, INDEC, Buenos Aires, 1999).

Una limitación que implica la utilización del MMNV tiene que ver con su actualización: esta no supera los dos años en localidades y/o aglomerados urbanos relevados por la Encuesta Permanente de Hogares (EPH), cumpliendo la tarea de actualización paralelamente al relevamiento. Esto significa que las localidades que en el Censo Nacional de Población y Viviendas de 1991 tenían 100.000 y más habitantes y las capitales de provincia están comprendidas en esta situación. Dicho de otra manera, la población cubierta en estas condiciones es de alrededor del 73% del total del país.

2.3 Etapas de muestreo

Las etapas de selección fueron cuatro. Las tres primeras corresponden al Marco de Muestreo Nacional de Viviendas (MMNV): la primera, el muestreo de unidades primarias (departamento o partidos); la segunda, el muestreo de unidades secundarias (radios censales o conjuntos de radios censales contiguos); la tercera, selección de viviendas. La cuarta, que atañe a la selección de la persona a entrevistar, surgió a partir de la utilización en campo de la tabla de Kish.

Los factores de expansión para cada una de esas etapas de selección se presentan a continuación.

2.4 Factores de expansión

Factor de expansión 1: por departamento

Cuadro 1: Frecuencia de valores del factor de expansión 1 (por departamento)

Valores del factor de expansión 1 agrupados	Cantidad de departamentos	Porcentaje de departamentos
Total	152	100
Valor 1	73	48,03
de 1,0001 a 1,9999	19	12,50
de 2 a 2,9999	28	18,42
de 3 a 3,9999	11	7,24
de 4 a 12,6601	21	13,82

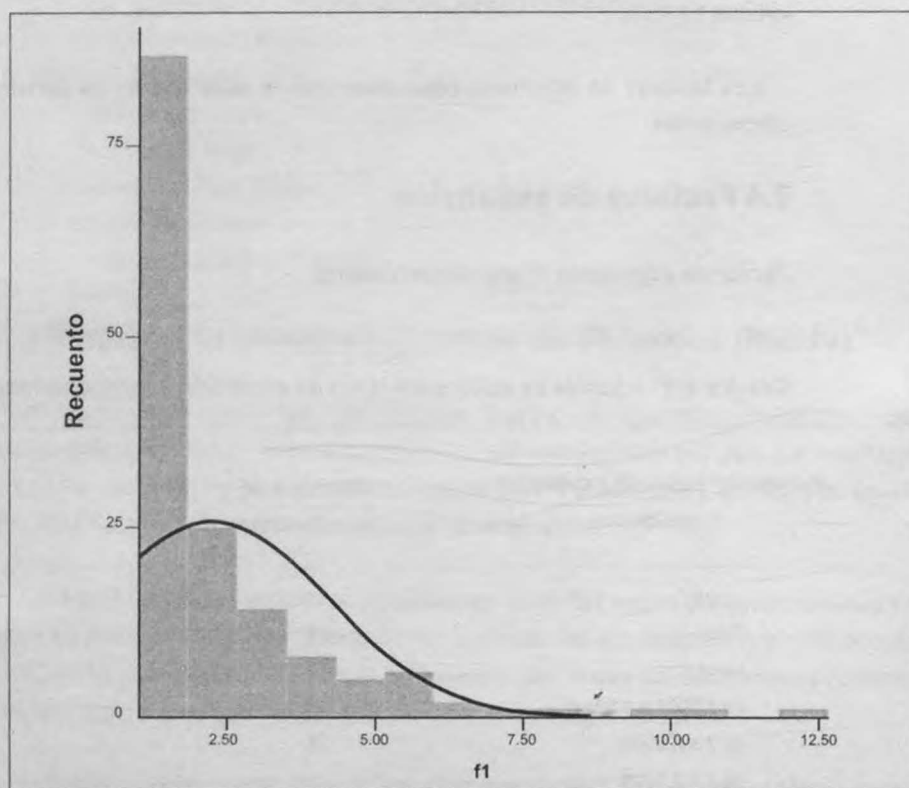
Fuente: INDEC. Encuesta Nacional de Factores de Riesgo 2009 (ENFR-2009).

Cuadro 2: Medidas de dispersión y posición

Media	2,27	
Desvío Estándar	1,96	
Varianza	3,86	
Percéntiles	25	1,00
	50	1,58
	75	2,79
Valor Mínimo	1,00	
Valor Máximo	12,66	

Fuente: INDEC. Encuesta Nacional de Factores de Riesgo 2009 (ENFR-2009).

Gráfico 1: Histograma de frecuencias del factor de expansión 1. Incluye curva Normal



Fuente: INDEC. Encuesta Nacional de Factores de Riesgo 2009 (ENFR-2009).

Para contrastar cada hipótesis de semejanza entre una distribución empírica y otra teórica, utilizamos la prueba de bondad de ajuste de Kolmogorov – Smirnof. Los valores significativos menores a 0,05 (con un 95 % de confianza) rechazan la semejanza.

Cuadro 3: Prueba de bondad de ajuste de Kolmogorov - Smirnof

Distribución de contraste	Valor Z de K-S	Significancia
Normal	3,2	0,00
Exponencial	4,4	0,00
Poisson (*)	1,95	0,01
Uniforme	7,71	0,00

(*) A fin de establecer la comparación con esta distribución, se transformó la variable original continua en discreta.

Fuente: INDEC. Encuesta Nacional de Factores de Riesgo 2009 (ENFR-2009).

La distribución del factor 1 no se asemeja a ninguna de las cuatro distribuciones, sin embargo, podemos decir que la distribución que mejor ajusta de las propuestas es la Poisson, ya que su valor de Z es el más bajo.

Factor de expansión 2: por radios censales.

Cuadro 4: Frecuencia de valores del factor de expansión 2 por radios censales

Valores del factor de expansión 1 agrupados	Cantidad de radios censales	Porcentaje de radios censales
Total	3.235	100,00
de 1 a 4,99	1.261	38,98
de 5 a 9,99	1.040	32,15
de 10 a 14,99	402	12,43
de 15 a 19,99	293	9,06
de 20 en adelante	239	7,39

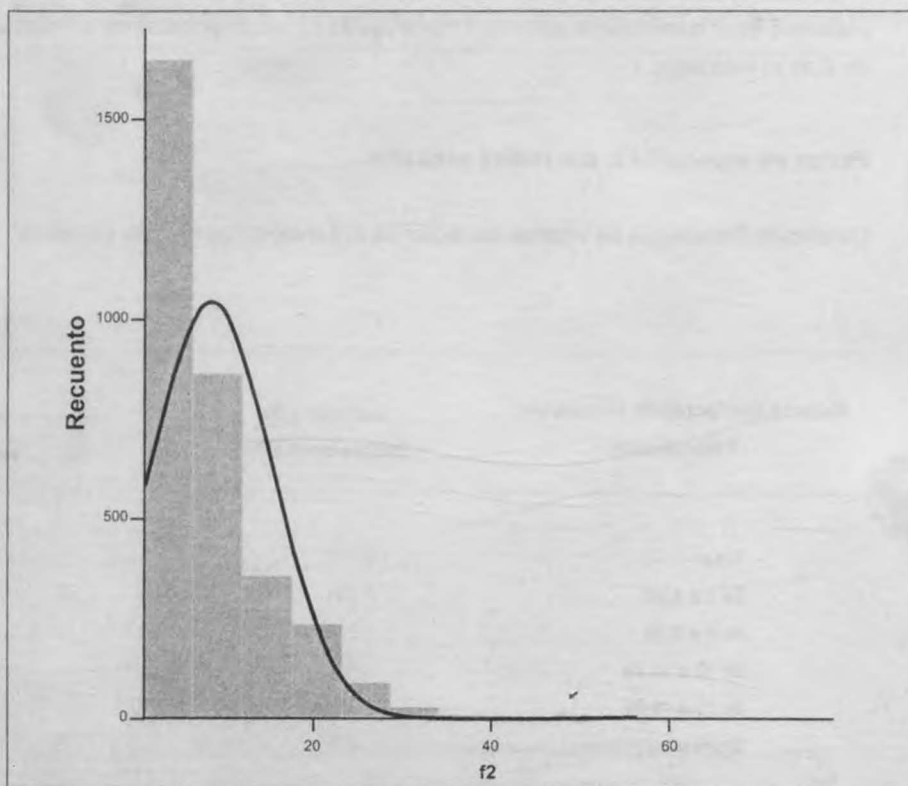
Fuente: INDEC. Encuesta Nacional de Factores de Riesgo 2009 (ENFR-2009).

Cuadro 5: Medidas de dispersión y posición

Media	8,47
Desvío Estándar	6,86
Variánza	47,86
Percentiles	
25	3,55
50	6,35
75	11,08
Valor Mínimo	1,00
Valor Máximo	78,63

Fuente: INDEC. Encuesta Nacional de Factores de Riesgo 2009 (ENFR-2009).

Gráfico 2: Histograma de frecuencias del factor de expansión 2. Incluye curva Normal



Fuente: INDEC. Encuesta Nacional de Factores de Riesgo 2009 (ENFR-2009).

Cuadro 6: Prueba de bondad de ajuste de Kolmogorov - Smirnov

Distribución de contraste	Valor Z de K-S	Significancia
Normal	7,94	0,00
Exponencial	7,68	0,00
Poisson (!)	15,9	0,01
Uniforme	38,78	0,00

(!) A fin de establecer la comparación con esta distribución, se transformó la variable original continua en discreta.

Fuente: INDEC. Encuesta Nacional de Factores de Riesgo 2009 (ENFR-2009).

La distribución del factor 2 no se asemeja a ninguna de las cuatro distribuciones, sin embargo, podemos decir que la distribución que mejor ajusta de las propuestas es la Exponencial, ya que su valor de Z es el más bajo.

Factor de expansión 3: por viviendas.

Cuadro 7: Frecuencia de valores del factor de expansión 3 por viviendas

Valores del factor de expansión 1 agrupados	Cantidad de radios censales	Porcentaje de radios censales
Total	47.159	100,00
de 1 a 9,9999	2.581	5,47
de 10 a 19,9999	18.083	38,34
de 20 a 29,9999	12.804	27,15
de 30 a 99,9999	13.562	28,76
de 100 a 261,63	129	0,27

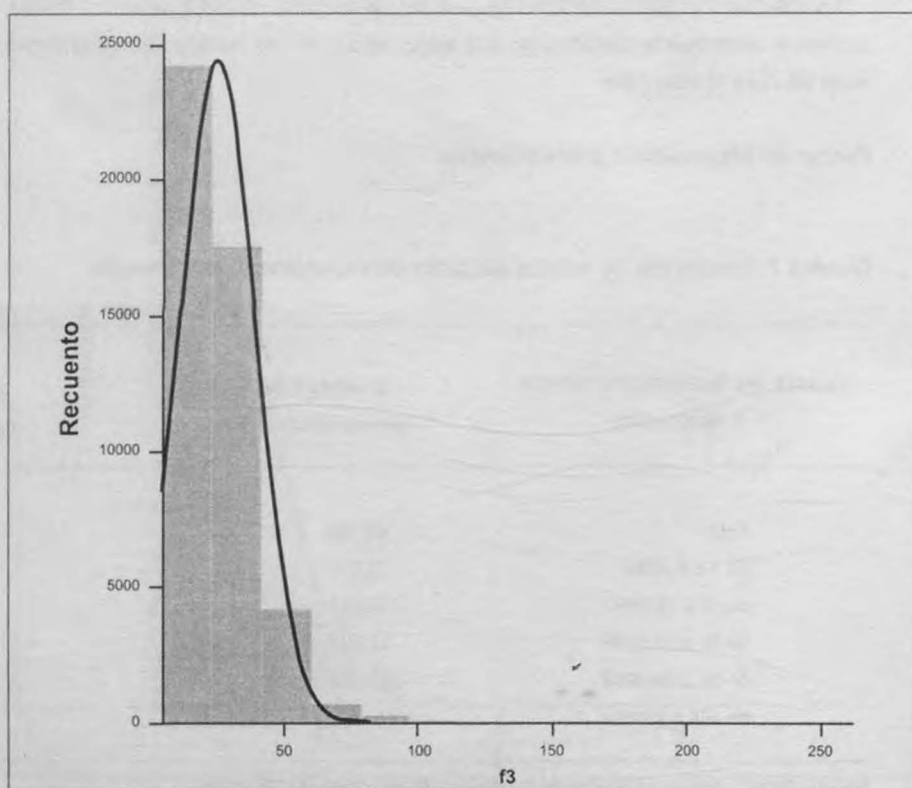
Fuente: INDEC. Encuesta Nacional de Factores de Riesgo 2009 (ENFR-2009).

Cuadro 8: Medidas de dispersión y posición

Media	25,03
Desvío Estándar	14,16
Varianza	200,42
Percentiles	
25	15,21
50	21,84
75	31,81
Valor Mínimo	4,10
Valor Máximo	261,63

Fuente: INDEC. Encuesta Nacional de Factores de Riesgo 2009 (ENFR-2009).

Gráfico 3: Histograma de frecuencias del factor de expansión 3. Incluye curva Normal



Fuente: INDEC. Encuesta Nacional de Factores de Riesgo 2009 (ENFR-2009).

Cuadro 9: Prueba de bondad de ajuste de Kolmogorov - Smirnof

Distribución de contraste	Valor Z de K-S	Significancia
Normal	22,3	0,00
Exponencial	60,01	0,00
Poisson (!)	63,81	0,01
Uniforme	168,96	0,00

(!) A fin de establecer la comparación con esta distribución, se transformó la variable original continua en discreta.

Fuente: INDEC. Encuesta Nacional de Factores de Riesgo 2009 (ENFR-2009).

La distribución del factor 3 no se asemeja a ninguna de las cuatro distribuciones, sin embargo, podemos decir que la distribución que mejor se ajusta de las propuestas es la Normal, ya que su valor de Z es el más bajo.

3

Características de la muestra

Si bien en la ENFR realizada en 2005 la selección fue obtenida del Marco de Muestreo Nacional Urbano (MMNU), las encuestas son plenamente comparables, al ser iguales los dominios representados.

En la distribución de la ENFR-2005, que se obtuvo a partir de estudios previos y sobre la base de estimaciones de prevalencias para diversos factores de riesgo, Buenos Aires tenía asignado un tamaño de 4.998 viviendas (2.076 en GBA y 2.922 en Resto de Buenos Aires). Para la ENFR-2009 se decidió aumentar el tamaño de la muestra en esta jurisdicción hasta 9.134 viviendas (5.062 en GBA y 4.072 en Resto de Buenos Aires), compensando con una reducción en el resto de las provincias. Este ajuste se realizó con el objetivo de disminuir la dispersión de los factores de expansión. También se decidió aumentar el tamaño de la muestra en Ciudad de Buenos Aires y en los grandes aglomerados.

En las provincias pequeñas, la reducción de muestra respetó un tamaño mínimo por provincia y dominio de estimación. Dichas reducciones se hicieron considerando los requisitos planteados que la muestra debe cumplir. Con estos ajustes se intentó llegar a un número total de viviendas similar al obtenido en la anterior encuesta.

Dentro de cada provincia se distribuyó la muestra por área para lograr un diseño aproximadamente autoponderado. Decimos aproximadamente ya que se fijaron límites inferiores y superiores al número de viviendas seleccionadas por área. Los límites fueron establecidos por razones de eficiencia con respecto a problemas de organización de campo: no encuestar menos de 8 ni menos de 24 viviendas por área en cada salida a campo.

3.1 Rendimiento de la muestra

Resultaron encuestables 42.188 viviendas, representando el 89,5% del total de viviendas seleccionadas. A partir de este conjunto se obtuvo respuesta en 34.934 viviendas, por lo tanto el rendimiento efectivo de la muestra, medido en viviendas, fue de 82,8%.

Jurisdicción	Rendimiento efectivo de la muestra
Total del país	82,8
Jujuy	96,2
Salta	95,0
Misiones	94,6
Formosa	93,7
La Rioja	93,7
Catamarca	93,0
Santa Cruz	91,6
San Luis	91,1
Río Negro	90,3
Chaco	89,3
Entre Ríos	88,7
Santiago del Estero	88,7
Tucumán	88,2
San Juan	86,7
Córdoba	86,4
La Pampa	86,1
Tierra del Fuego	84,4
Corrientes	84,3
Chubut	84,1
Mendoza	82,5
Santa Fe	80,2
Resto Pcia. de Bs. As.	77,0
Neuquén	75,0
24 PCB	60,9
Ciudad de Bs. As.	59,3

Fuente: INDEC. Encuesta Nacional de Factores de Riesgo 2009 (ENFR-2009).

En esas viviendas se encontraron 36.080 hogares, es decir que en el 2,6% de las viviendas hay hogares que la comparten.

Para responder el bloque individual del cuestionario, se seleccionó en cada hogar, de manera aleatoria, una persona entre las que tenían 18 años y más. Se obtuvo respuesta en 34.732 hogares correspondientes a 33.655 viviendas. De esta manera, del total de viviendas encuestables de la muestra de origen (42.188), se logró información para la unidad de análisis, en un 79,8%, que llamamos rendimiento específico de la muestra de viviendas.

Provincia	Rendimiento efectivo de la muestra
Total del país	79,8
Jujuy	94,1
Salta	92,4
Formosa	92,4
Misiones	92,2
La Rioja	90,8
Catamarca	90,1
San Luis	89,2
Santa Cruz	88,9
Río Negro	88,4
Entre Ríos	87,9
Chaco	86,8
San Juan	85,9
Santiago del Estero	85,4
Córdoba	85,4
La Pampa	84,3
Tucumán	83,7
Chubut	82,5
Tierra del Fuego	81,9
Corrientes	79,2
Santa Fe	77,4
Mendoza	76,9
Resto Pcia. de Bs. As.	74,1
Neuquén	70,8
Ciudad de Bs. As.	56,1
24 PCB	54,6

Fuente: INDEC. Encuesta Nacional de Factores de Riesgo 2009 (ENFR-2009).

4

La No Respuesta

Se entiende por No Respuesta a la ausencia de datos en los formularios a partir de la visita a una vivienda seleccionada. Esta No Respuesta puede ser Efectiva o No Efectiva.

La No Respuesta No Efectiva es cuando no se puede obtener el dato buscado por alguna de las siguientes causas: la vivienda estaba deshabitada, demolida, estaba siendo utilizada como fin de semana, estaba en construcción, era una vivienda utilizada como establecimiento o presentaba variaciones en el listado. Todas estas causas implican que en la vivienda seleccionada no había un hogar residiendo en la semana de referencia del operativo.

La No Respuesta Efectiva se produce cuando se detecta la presencia de un hogar residiendo en dicha dirección y no se puede obtener el dato debido a que los habitantes estaban ausentes, rechazaron ser encuestados u otras causas, como ser duelo, alcoholismo, discapacidad, idioma extranjero, problemas de seguridad o por problemas de accesibilidad climáticos u otros.

Es posible, entonces que una vivienda seleccionada contenga un solo hogar y que, tanto la vivienda seleccionada como el hogar que reside en ella, sea Respuesta o No Respuesta, en este último caso tanto Efectiva como No Efectiva. Pero también es posible que en una vivienda se detecten más de un hogar, siendo al menos uno de ellos Respuesta y al menos uno de los otros No Respuesta Efectiva.

Los resultados de la ENFR-2009 han sido los siguientes:

Cuadro 10: Viviendas seleccionadas, Respuesta y No Respuesta por causas. Encuesta Nacional de Factores de Riesgo 2009

Jurisdicción	Viviendas No respuesta							Total No Respuesta No Efectiva	% de Total No Respuesta No Efectiva
	Total de viviendas y hogares en la muestra ⁽¹⁾	Causa 1	Causa 2	Causa 3	Causa 4	Causa 5	Causa 6		
	A	B	C	D	E	F	G		
Total	48.533	2.240	256	621	381	538	1.006	5.042	10,39
Jujuy	1.710	59	5	22	12	12	12	122	7,13
Chaco	1.571	50	10	6	6	16	28	116	7,38
Chubut	1.487	60	3	6	10	22	22	123	8,27
24 PCB	5.174	181	20	47	51	30	113	442	8,54
Santa Cruz	1.565	53	2	6	12	23	38	134	8,56
Entre Ríos	1.651	69	8	11	9	19	27	143	8,66
Tucumán	1.797	83	10	9	11	6	38	157	8,74
Córdoba	2.698	107	10	30	11	26	64	248	9,19
Mendoza	1.588	69	9	11	11	18	28	146	9,19
Santa Fe	2.734	107	6	39	18	37	47	254	9,29
Catamarca	1.563	77	6	14	14	28	8	147	9,40
San Juan	1.399	72	15	7	4	12	22	132	9,44
Río Negro	1.658	82	9	7	13	15	40	166	10,01
Neuquén	1.598	47	3	27	11	24	52	164	10,26
Ciudad de Bs. As.	2.287	139	5	25	21	40	20	250	10,93
Formosa	1.546	84	12	23	12	20	19	170	11,00
San Luis	1.629	70	6	29	15	17	44	181	11,11
Salta	1.632	72	11	38	17	17	32	187	11,46
Buenos Aires	9.338	441	37	207	95	74	237	1.091	11,68
Tierra del Fuego	1.437	62	14	1	13	22	60	172	11,97
La Rioja	1.630	119	3	24	18	20	12	196	12,02
Misiones	1.457	73	51	15	13	11	13	176	12,08
Santiago del Estero	1.289	95	8	9	10	14	20	156	12,10
Corrientes	1.658	66	8	12	14	15	91	206	12,42
La Pampa	1.611	84	5	43	11	30	32	205	12,73
Resto Pcia. Bs. As.	4.164	260	17	160	44	44	124	649	15,59

(continúa)

(¹) Si bien las viviendas seleccionadas fueron 47.159, hubo viviendas donde se detectó más de un hogar.

Fuente: INDEC. Encuesta Nacional de Factores de Riesgo 2009 (ENFR-2009).

Cuadro 10: (conclusión)

Viviendas No respuesta									
Jurisdicción	Causa 7	Causa 8	Causa 9	Total No	% de	Total No	% de	Total No	% de
				Respuesta Efectiva	Total No Respuesta Efectiva	Respuesta Efectiva	Total No Respuesta Efectiva	Respuesta Efectiva	Total No Respuesta Efectiva
	J	K	L	M=J+K+L	N=M/A*100	O=H+M	P=O/A*100	Q	R=Q/A*100
Total	3.020	3893	498	7411	15,27	12.453	25,66	36.080	74,34
Jujuy	23	33	6	62	3,63	184	10,76	1.526	89,24
Chaco	73	62	19	154	9,80	270	17,19	1.301	82,81
Chubut	116	87	10	213	14,32	336	22,60	1.151	77,40
24 PCB	604	1.023	218	1.845	35,66	2.287	44,20	2.887	55,80
Santa Cruz	65	49	6	120	7,67	254	16,23	1.311	83,77
Entre Ríos	50	94	22	166	10,05	309	18,72	1.342	81,28
Tucumán	79	102	6	187	10,41	344	19,14	1.453	80,86
Córdoba	166	141	23	330	12,23	578	21,42	2.120	78,58
Mendoza	87	143	24	254	15,99	400	25,19	1.188	74,81
Santa Fe	191	243	48	482	17,63	736	26,92	1.998	73,08
Catamarca	52	46	3	101	6,46	248	15,8	1.315	84,13
San Juan	71	90	1	162	11,58	294	21,02	1.105	78,98
Río Negro	68	72	6	146	8,81	312	18,82	1.346	81,18
Neuquén	197	153	5	355	22,22	519	32,48	1.079	67,52
Ciudad de Bs. As.	356	464	8	828	36,20	1.078	47,14	1.209	52,86
Formosa	51	33	2	86	5,56	256	16,56	1.290	83,44
San Luis	40	74	12	126	7,73	307	18,85	1.322	81,15
Salta	20	50	1	71	4,35	258	15,81	1.374	84,19
Buenos Aires	934	1.463	256	2.653	28,41	3.744	40,09	5.594	59,91
Tierra del Fuego	71	121	12	204	14,2	376	26,17	1.061	73,83
La Rioja	38	49	2	89	5,46	285	17,48	1.345	82,52
Misiones	23	44	3	70	4,80	246	16,88	1.211	83,12
Santiago del Estero	63	56	7	126	9,78	282	21,88	1.007	78,12
Corrientes	94	121	15	230	13,87	436	26,30	1.222	73,70
La Pampa	92	103	1	196	12,17	401	24,89	1.210	75,11
Resto Pcia. Bs. As.	330	440	38	808	19,40	1.457	34,99	2.707	65,01

Como se puede apreciar, la respuesta fue del 74,34% de las viviendas seleccionadas/hogares detectados.

La jurisdicción que presentó el mayor nivel de Respuesta fue la provincia de Jujuy con un 89,24%; mientras que la más baja Respuesta se ubicó en la Ciudad de Buenos Aires, con tan solo el 52,86% y la provincia de Buenos Aires con el 59,91%.

La No Respuesta fue del 25,66%, compuesta por un 10,39% de No Respuesta No Efectiva y un 15,27% de No Respuesta Efectiva.

Las jurisdicciones con mayores niveles de No Respuesta No Efectiva fueron La Pampa y Corrientes, con 12,73% y 12,42% respectivamente; mientras que los valores más bajos se encontraron en las de Jujuy y Chaco, con 7,13% y 7,38% respectivamente.

En cuanto a la No Respuesta Efectiva, los porcentajes más altos se encontraron en la provincia de Buenos Aires 28,41% y en la Ciudad de Buenos Aires 36,20%; mientras que los valores más bajos fueron para Jujuy 3,63% y Misiones 4,80%.

4.1 Tratamiento de la No Respuesta

El análisis de No Respuesta tiene dos objetivos principales: el primero, en campo, controlar y minimizar la No Respuesta (mediante el control de encuestadores, la supervisión y el fortalecimiento del trabajo, entre otras tareas) y el segundo, posterior, la calibración de los factores de expansión iniciales, con la corrección de la ponderación.

En relación con el primero, la comparación con los datos obtenidos en la Encuesta Nacional de Factores de Riesgo 2005 (ENFR-2005) y con la Encuesta Nacional sobre Prevalencias de Consumo de Sustancias Psicoactivas 2008 (ENPreCoSP-2008), nos sirvieron de guía para saber si el trabajo realizado se encuentra dentro de parámetros correctos.

Cuadro 11: Viviendas seleccionadas, Respuesta y No Respuesta por causas

Encuesta	No Respuesta Total	No Respuesta No Efectiva	No Respuesta Efectiva
ENFR-2005	18,9%	9,5%	9,4%
ENPreCoSP-2008	23,9%	10,6%	13,3%
ENFR-2009	25,9%	10,5%	15,4%

Fuente: INDEC. Encuesta Nacional de Factores de Riesgo 2005 (ENFR-2005), Encuesta Nacional sobre Prevalencia de Consumo de Sustancias Psicoactivas 2008 (ENPreCoSP-2008), Encuesta Nacional de Factores de Riesgo 2009 (ENFR-2009).

Los datos de No Respuesta No Efectiva de la ENFR-2009 no presentan modificaciones considerables respecto de la ENPreCoSP-2008 y la ENFR-2005.

En cambio, la No Respuesta Efectiva presenta un aumento sistemático. Tres jurisdicciones han contribuido fuertemente en este proceso: la Ciudad de Buenos Aires de 18,2% a 36,2%, la provincia de Neuquén de 16,2% a 22,2% y la provincia de Buenos Aires, fundamentalmente los 24 Partidos el Conurbano Bonaerense, donde la No Respuesta Efectiva llegó al 35,7%.

Una vez concluido el trabajo de campo e ingreso de datos se procede a la revisión de todos los cuestionarios ingresados como No Respuesta para verificar la exactitud de la información suministrada en los campos destinados a "Razón de No Entrevista" en función de las causas y motivos señalados anteriormente. Esta tarea se realiza analizando conjuntamente lo señalado como causa y motivo de No Entrevista y lo anotado por el personal de campo que corresponda en el campo "Observaciones" según se establece en el Manual de Procedimientos. En caso de ser necesario se solicita la revisión del cuestionario utilizado para garantizar la calidad de la información.

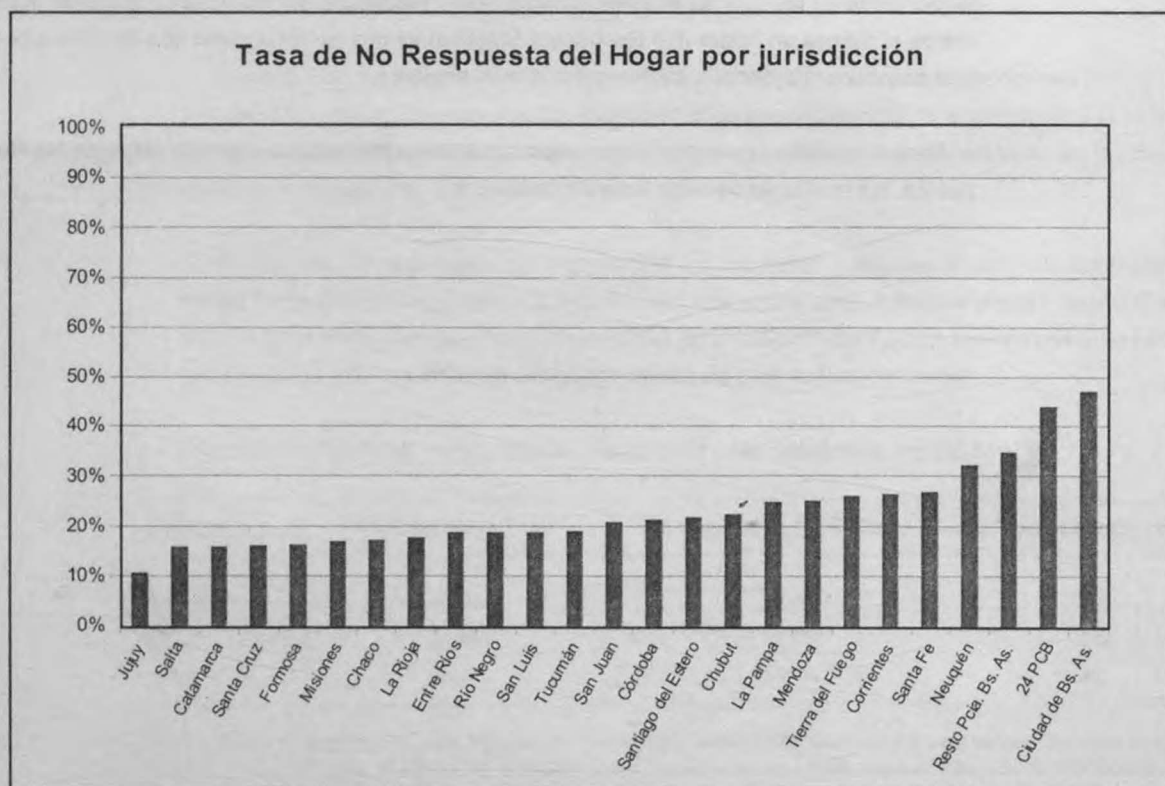
Esto es fundamental a efectos de realizar una correcta separación de las viviendas seleccionadas en las que no vive ningún hogar (No Respuesta No Efectiva) de aquellas donde reside al menos un hogar (No Respuesta Efectiva) ya que el tratamiento que se dará a cada uno es totalmente distinto, como veremos más adelante.

Para el análisis y posterior elaboración del informe se calculan distintas tasas de No Respuesta, las cuales se detallan a continuación:

4.2 Tasas de No Respuesta

- **Tasa de No Respuesta del Hogar:** es el cociente entre la cantidad de cuestionarios No Respuesta sobre la cantidad total de cuestionarios.

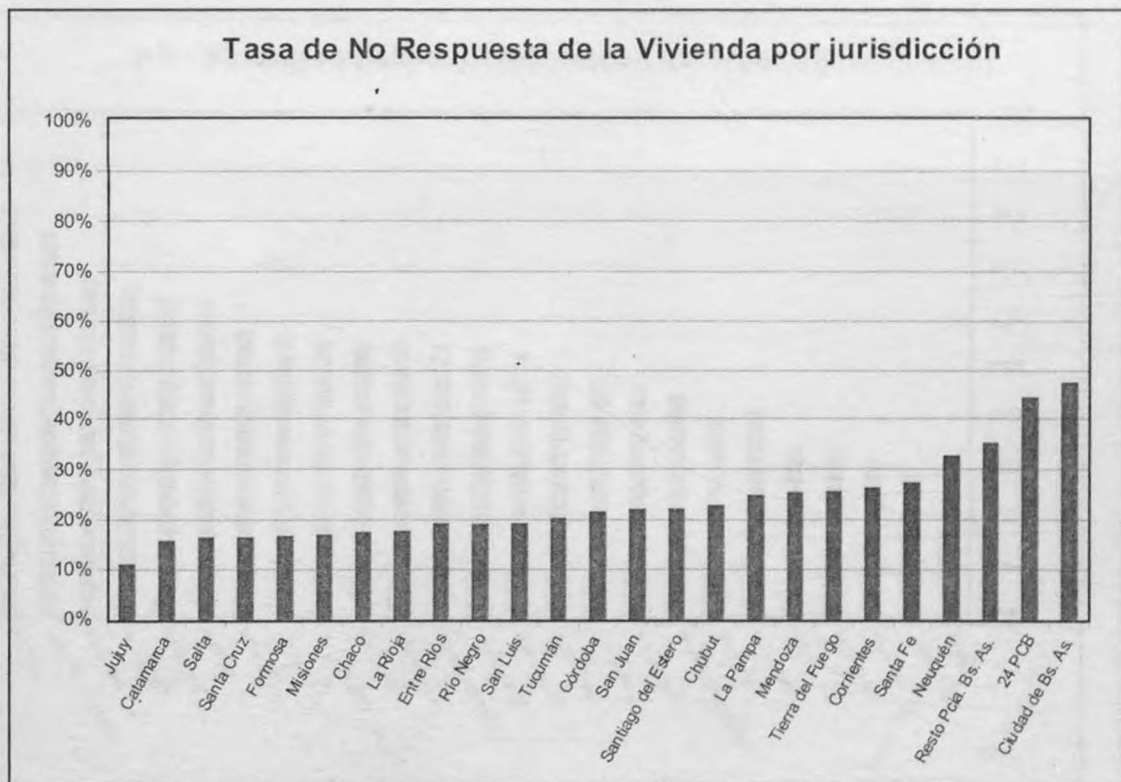
$$\frac{\sum_{i=1}^n C_{NRi}}{\sum_{i=1}^n C_i} = 0.257$$



Fuente: INDEC. Encuesta Nacional de Factores de Riesgo 2009 (ENFR-2009).

- **Tasa de No Respuesta de la Vivienda:** es el cociente entre la cantidad de viviendas No Respuesta sobre la cantidad total de viviendas seleccionadas.

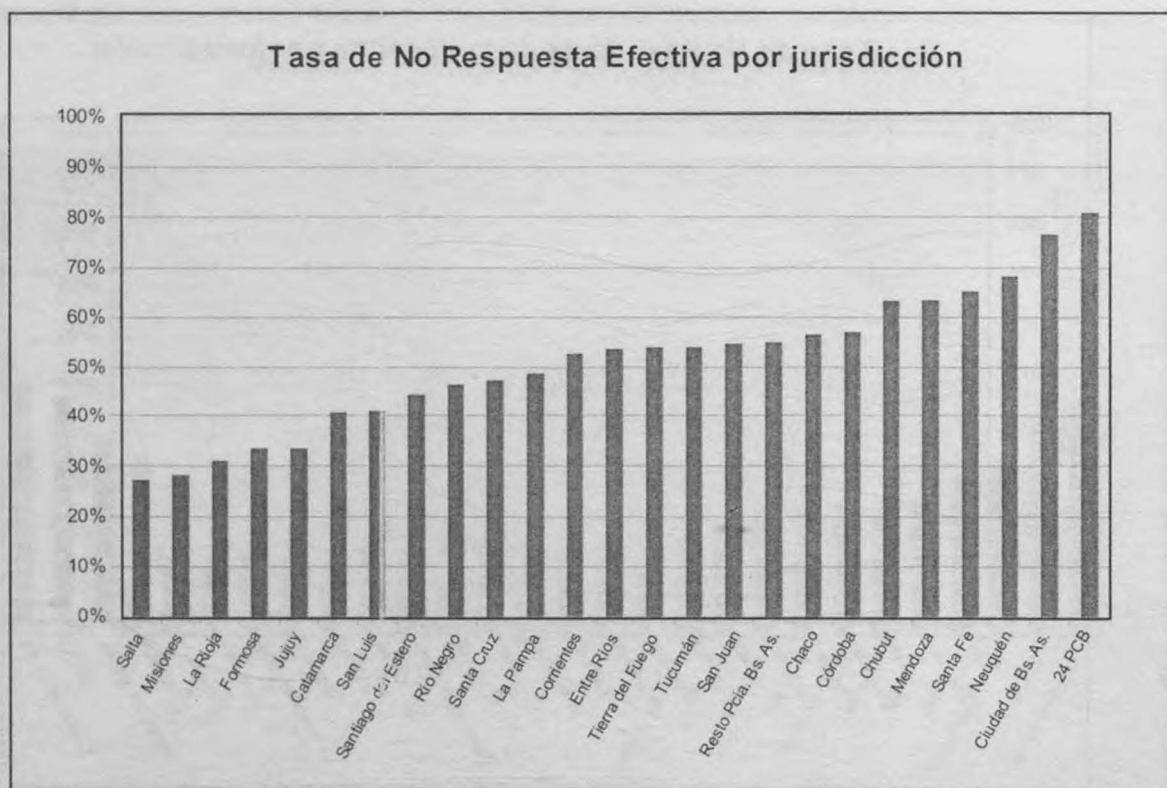
$$\frac{\sum_{i=1}^n V_{NRi}}{\sum_{i=1}^n V_i} = 0.259$$



Fuente: INDEC. Encuesta Nacional de Factores de Riesgo 2009 (ENFR-2009).

- **Tasa de No Respuesta Efectiva:** es el cociente entre la cantidad de cuestionarios No Respuesta Efectiva sobre la cantidad total de cuestionarios No Respuesta.

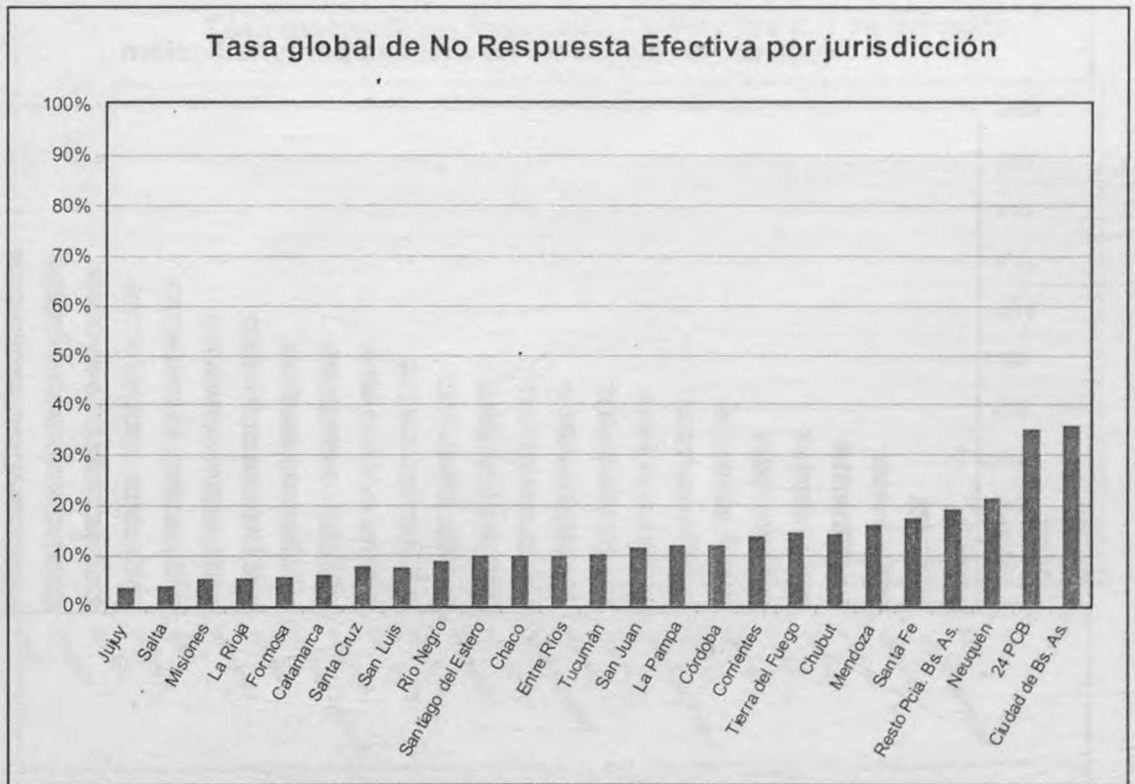
$$\frac{\sum_{i=1}^n C_{NRei}}{\sum_{i=1}^n C_{NRi}} = 0.595$$



Fuente: INDEC. Encuesta Nacional de Factores de Riesgo 2009 (ENFR-2009).

- **Tasa global de No Respuesta Efectiva:** es el cociente entre la cantidad de cuestionarios No Respuesta Efectiva sobre la cantidad total de cuestionarios.

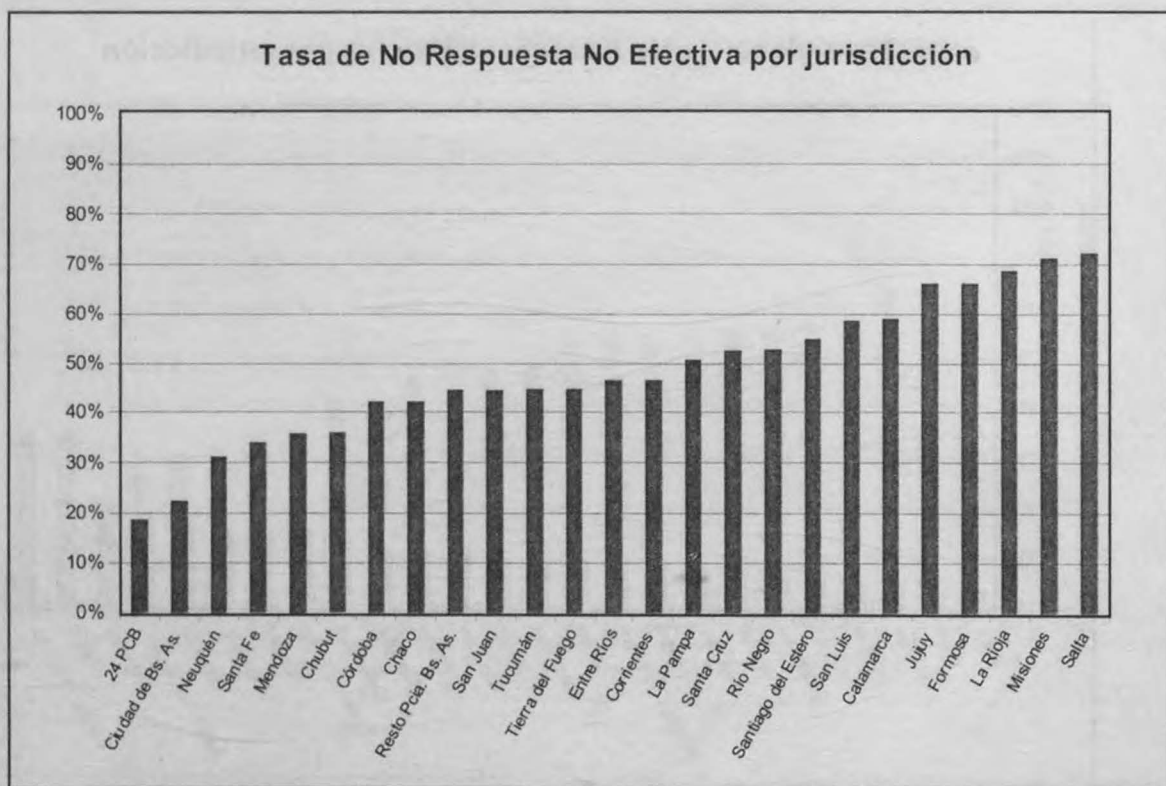
$$\frac{\sum_{i=1}^n C_{NRei}}{\sum_{i=1}^n C_i} = 0.153$$



Fuente: INDEC. Encuesta Nacional de Factores de Riesgo 2009 (ENFR-2009).

- **Tasa de No Respuesta No Efectiva:** es el cociente entre la cantidad de cuestionarios No Respuesta No Efectiva sobre la cantidad total de cuestionarios No Respuesta.

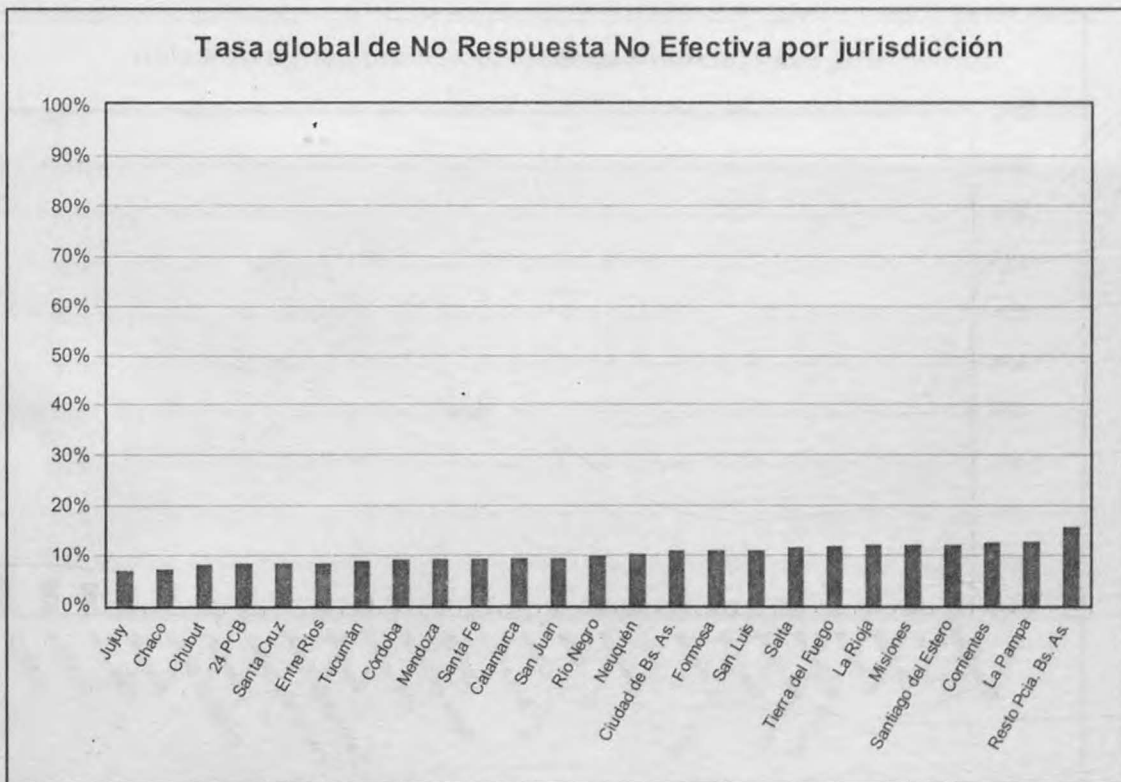
$$\frac{\sum_{i=1}^n C_{NRnei}}{\sum_{i=1}^n C_{NRi}} = 0.405$$



Fuente: INDEC. Encuesta Nacional de Factores de Riesgo 2009 (ENFR-2009).

- **Tasa global de No Respuesta No Efectiva:** es el cociente entre la cantidad de cuestionarios No Respuesta No Efectiva sobre la cantidad total de cuestionarios.

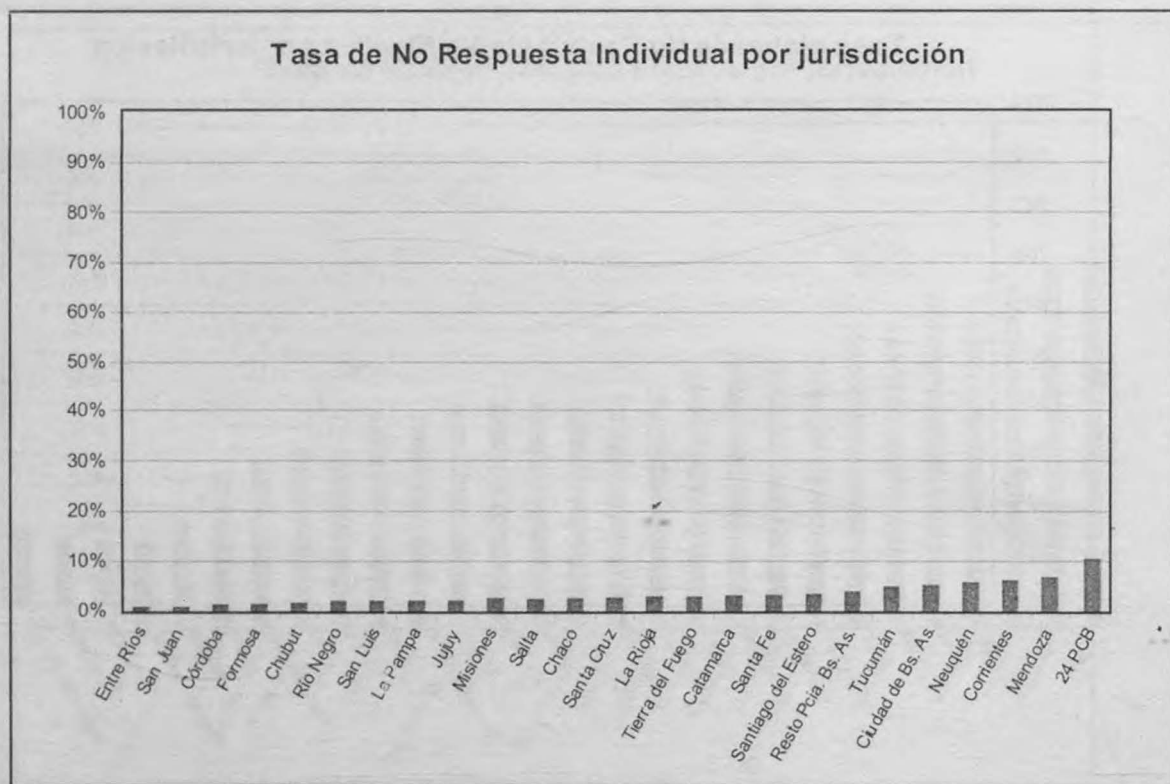
$$\frac{\sum_{i=1}^n C_{NRnei}}{\sum_{i=1}^n C_i} = 0.104$$



Fuente: INDEC. Encuesta Nacional de Factores de Riesgo 2009 (ENFR-2009).

- **Tasa de No Respuesta Individual:** es el cociente entre la cantidad de cuestionarios No Respuesta del Bloque Individual sobre la cantidad total de cuestionarios Respuesta del Bloque del Hogar.

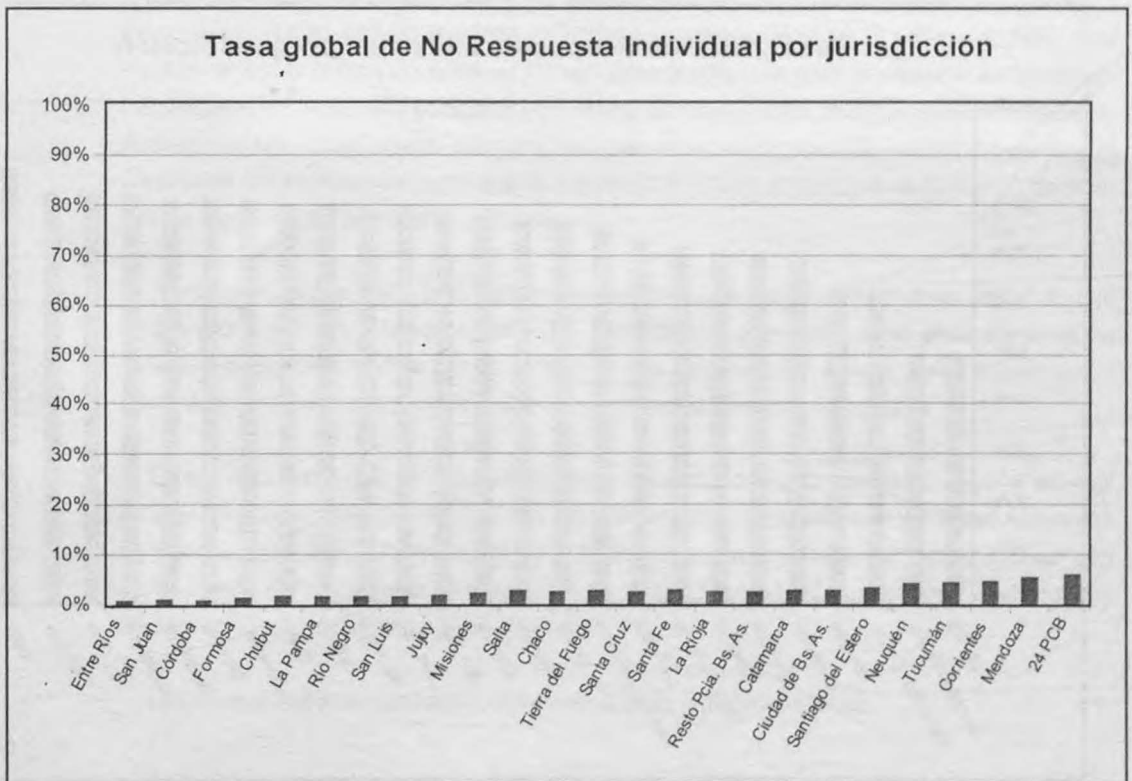
$$\frac{\sum_{i=1}^n C_{NRIi}}{\sum_{i=1}^n C_{RHi}} = 0.037$$



Fuente: INDEC. Encuesta Nacional de Factores de Riesgo 2009 (ENFR-2009).

- **Tasa global de No Respuesta Individual:** es el cociente entre la cantidad de cuestionarios No Respuesta del Bloque Individual sobre la cantidad total de cuestionarios.

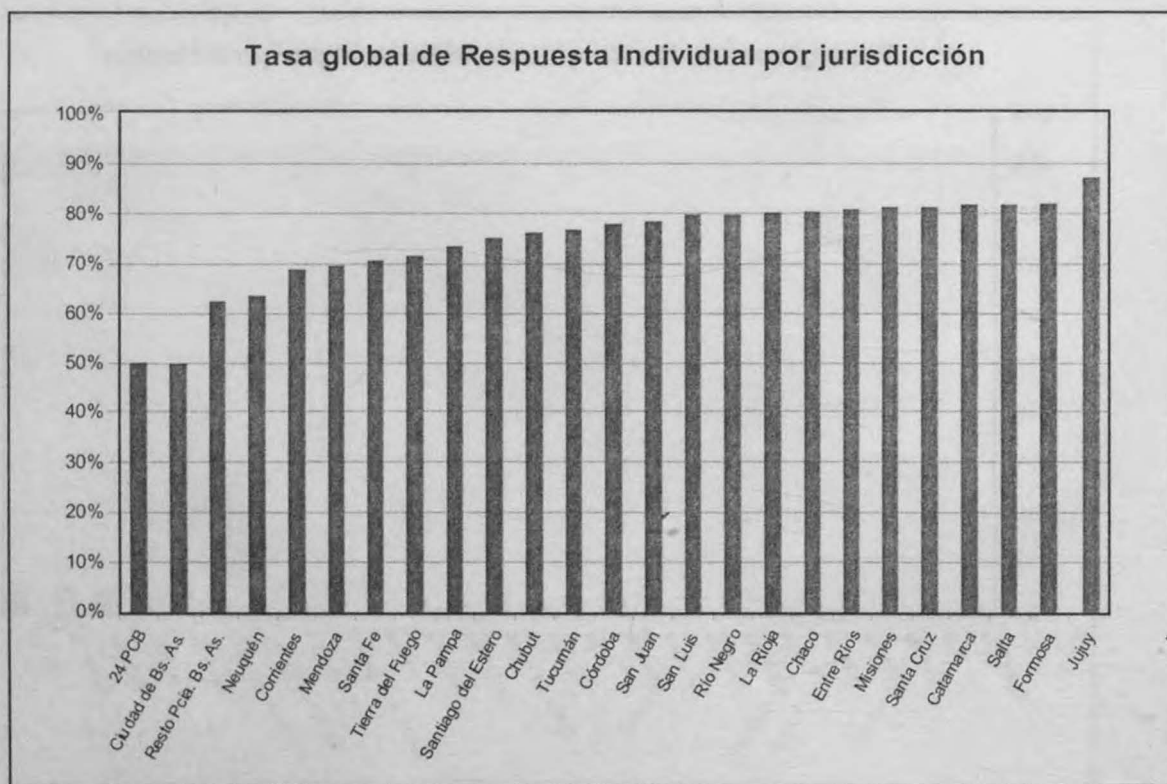
$$\frac{\sum_{i=1}^n C_{NRIi}}{\sum_{i=1}^n C_i} = 0.027$$



Fuente: INDEC. Encuesta Nacional de Factores de Riesgo 2009 (ENFR-2009).

- **Tasa global de Respuesta Individual:** es el cociente entre la cantidad de cuestionarios Respuesta del Bloque Individual sobre la cantidad total de cuestionarios.

$$\frac{\sum_{i=1}^n C_{RIi}}{\sum_{i=1}^n C_i} = 0.716$$



Fuente: INDEC. Encuesta Nacional de Factores de Riesgo 2009 (ENFR-2009).

También se cruzan, para realizar los análisis de No Respuesta, estos valores en función de variables de categorización (sexo, grupo de edad, etcétera) a fin de observar, de acuerdo con los distintos estratos considerados, el comportamiento de la No Respuesta, para tenerlo en cuenta en futuras encuestas.

4.3 Corrección de la No Respuesta

Existe un primer tratamiento de la No Respuesta a efectos de ajustar los factores iniciales o teóricos de expansión determinados al momento de la selección de las viviendas. Como mencionamos, en la ENFR-2009 los factores de expansión se aplicarán a cada etapa de selección de acuerdo al nuevo Marco de Muestreo Nacional de Viviendas (MMNV).

El factor de corrección por No Respuesta fue definido como el cociente entre las viviendas que responden, más las que no responden teniendo al menos un hogar, dividido por las viviendas respondidas.

En ENFR-2009 la No Respuesta se corrigió dentro de cada estrato por unidad primaria de selección (departamento). Hay que tener en cuenta que este nuevo marco consta de 12 estratos detallados en letras (A-L), cuyos puntos de corte varían entre las unidades primarias de selección (UPS). Debido a la falta de información existente sobre el criterio utilizado para determinar estos puntos de corte en las variables empleadas para la creación de los distintos estratos, no es posible comparar una misma letra de estrato en departamentos distintos. A pesar de ello, cabe aclarar que este inconveniente no perjudica la estimación, ya que la corrección por No Respuesta se realizó siguiendo el mismo criterio, es decir, dentro de cada estrato y por unidad primaria de selección.

Para llevar a cabo dicha tarea es importante tratar de disminuir la chance del error, que podría producirse en el resultado obtenido del factor de corrección al no encontrar ninguna vivienda respuesta dentro de alguno de los estratos de cada UPS. La chance de encontrarnos con este problema crece a medida que aumentamos la cantidad de estratos.

Es por esta razón que se propuso disminuir la cantidad de estratos hasta llegar a obtener 6. Se propuso esta y no otra cantidad de estratos teniendo en cuenta que la Ciudad Autónoma de Buenos Aires, los 24 Partidos del Conurbano Bonaerense y Tierra del Fuego cuentan con los 6 estratos definidos por la EPH. De esta manera, todas las UPS cuentan con 6 estratos (no comparables y con distintos puntos de corte).

Los nuevos estratos quedaron conformados de la siguiente manera:

Cuadro 12: Remodificación de estratos

Estratos originales	Estratos recodificados
A	
B	A
C	
D	B
E	
F	C
G	
H	D
I	
J	E
K	
L	F

Nota: La letra "A" representa el estrato mas bajo y la letra "L" (en la recodificación letra "F") representa el estrato mas alto.

Cabe aclarar que la transformación anteriormente detallada no elimina la posibilidad de que se produzca un error en el cálculo del factor de corrección por No Respuesta, aunque sí la disminuye considerablemente.

5

Expansión de la muestra y calibración de los factores de expansión

Entre los problemas más comunes en las encuestas complejas de gran envergadura, que utilizan diseños multietápicos, con conglomerados, distintos niveles de estratificación, selección de unidades controladas, o con probabilidades de selección proporcionales al tamaño, etcétera, se halla la introducción de sesgos en la etapa de estimación. A estos problemas se le agregan los originados por desactualización del marco muestral empleado para la selección, el no relevamiento total o parcial de algunas unidades de muestreo, errores que se introducen en la etapa de la recolección de datos, ingreso o codificación y la imputación de valores faltantes en ciertas variables de interés.

Todas estas dificultades dependiendo de la magnitud de los errores- pueden dar como resultado sesgos pronunciados, impidiendo una interpretación correcta de las estimaciones o imposibilitando la comparación de los resultados con fuentes alternativas. Como consecuencia, en la práctica es habitual corregir o calibrar los pesos o factores de expansión iniciales (las inversas de las probabilidades de selección), con ayuda de información auxiliar conocida o preestablecida a partir de registros o fuentes externas.

Naturalmente, la corrección se debe valer de argumentos estadísticos sólidos que determinen pesos óptimos en algún sentido. Es decir, es importante que en las estimaciones no se introduzcan sesgos a causa de la metodología a emplear (con una magnitud tal que lleven a producir importantes crecimientos en el Error Cuadrático Medio del estimador, la habitual medida de precisión en estas circunstancias).

Para la Encuesta Nacional de Factores de Riesgo 2009 se optó por trabajar con la macro CALMAR, que es una macro SAS programada por el INSEE (Instituto de Estadística Francés) para la finalidad mencionada anteriormente, con la ventaja principal de que funciona en el entorno SAS PC. Esta macro, por métodos iterativos calcula los pesos finales para cada uno de los casos, según sus características, teniendo en cuenta los pesos iniciales.

Para la calibración se plantean las ecuaciones que deben satisfacer los pesos nuevos, de acuerdo con las marginales introducidas, con la condición de que sean lo más próximo posibles a los pesos iniciales.

Desde el punto de vista de las estimaciones, el hecho de contar con una estrategia de ajuste o calibración a través del empleo de información auxiliar, definen estimadores que en general llevan a ganar precisión en los resultados finales. Ésta ganancia se acentúa cuando la correlación entre las variables de estudio y las auxiliares empleadas para el ajuste son importantes.

Para la ENFR-2009 se emplearon controles para evitar la generación de pesos extremos, y también para evitar pesos negativos, ya que puede ocurrir que se generen dichos pesos, los cuales serían perjudiciales en cuanto al análisis.

Con respecto a la información auxiliar empleada para la calibración, se recurrió a proyecciones de población calculadas por el INDEC para el periodo de referencia de la Encuesta (a nivel provincial y a nivel de aglomerado según la Encuesta Permanente de Hogares), a las cuales se les aplicó la estructura interna de la muestra para las variables sexo (varón-mujer), grupo de edad (0 a 17 años, 18 a 24 años, 25 a 34 años, 35 a 49 años, 50 a 65 años, y mayor de 65 años), nivel de instrucción y para una variable que indicaba si la persona era o no jefe de hogar.

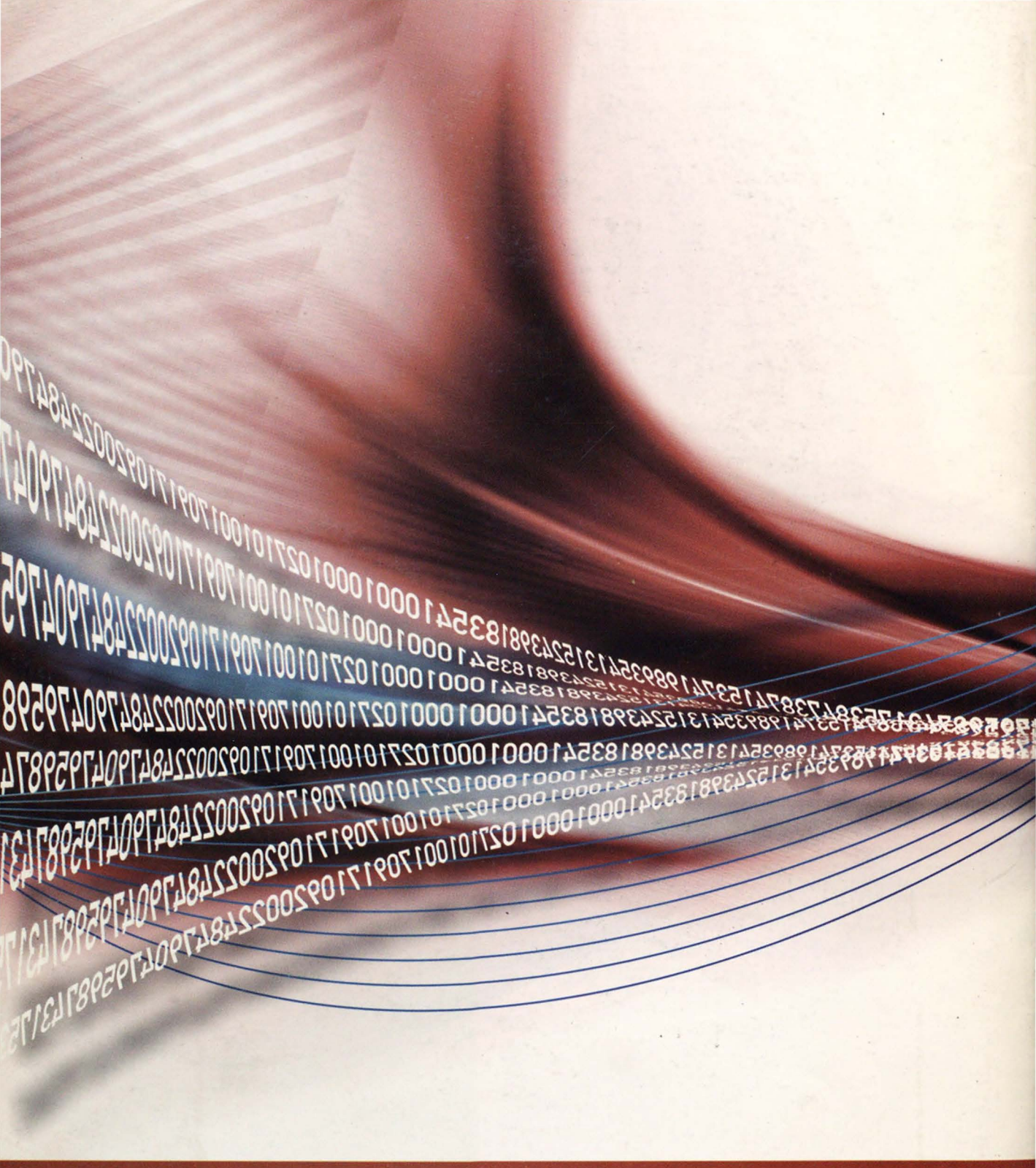
Con el fin de salvar posibles inconsistencias al momento de volcar esta información en cuadros para publicar, los pesos calibrados fueron tratados por un algoritmo de redondeo para eliminar la componente decimal con los que salen del proceso iterativo, sin destruir las concordancias alcanzadas.

6

Bibliografía

- Deville, J. y Särndal, C.E., Calibration Estimators in Survey Sampling. *JASA*, 87, 376-382, 1992.
- Deville, J., Särndal, C.E., Sautory O., Generalized Raking Procedures in Survey Sampling. *JASA*, 88, 1013-1020, 1993.
- Hernández Sampieri, Roberto (Coord.), Metodología de la Investigación. Ed. Mc Graw Hill. México, D.F. 2003.
- Huang, E. y Fuller, W., Nonnegative Regression Estimation in Sample Survey Data. Proceedings of the Social Statistics Section, American Statistical Association, 1978.
- INDEC, Manual del Encuestador de la Encuesta Nacional de Factores de Riesgo 2009, INDEC, Buenos Aires, 2009.
- INDEC, Marco de Muestreo Nacional Urbano para Encuestas a Hogares, Metodologías n° 12. INDEC, Buenos Aires, 1999.
- Jayasuriya, B. y Valliant, R., An Application of Restricted Regression Estimation in a Household Survey, *Survey Methodology*, Vol. 22, N° 2, pp. 127-137. 1996.
- Mitas, G., Estimaciones a través del Método de Regresión Generalizada Restringida en la ENGH, (Documento Técnico Interno). INDEC, Buenos Aires, 1998.
- Sabino, Carlos, El proceso de investigación. Lumen Humanitas, Buenos Aires, 1996.
- Samaja, Juan. Epistemología y Metodología. Ed. EUDEBA, Bs. As. 1997.
- Särndal, C.E, Swesson, B y Wretman, J.H, Model Assisted Survey Sampling. Springer-Verlag, New York, 1992.
- SAS Institute, SAS/STAT User's Guide Release 8.2 Edition, SAS Institute Inc., Cary, NC, USA.
- Sautory, O., La Macro CALMAR: Redressement d'un Echantillon par Calage sur Marges, Document de travail de la Direction des Statistiques Demographiques et Sociales, N° F9310, 1993.
- Vieytes, Rut. Metodología de la investigación en organizaciones, mercado y sociedad. Epistemología y técnicas. Ed. de las Ciencias, Buenos Aires, 2004.

ESTE VOLUMEN
METODOLOGÍA N°16
SE TERMINÓ DE IMPRIMIR
EN EL MES DE SEPTIEMBRE DE 2010,
CON UNA TIRADA DE 1.000 EJEMPLARES,
EN EL DEPARTAMENTO PUBLICACIONES DEL INDEC
AV. PRESIDENTE JULIO A. ROCA 609
C1067ABB CIUDAD AUTÓNOMA DE BUENOS AIRES
REPÚBLICA ARGENTINA



REPÚBLICA ARGENTINA
MINISTERIO DE ECONOMÍA Y FINANZAS PÚBLICAS
INSTITUTO NACIONAL DE ESTADÍSTICA Y CENSOS