

LAS TÉCNICAS RDD EN LA METODOLOGÍA DE ENCUESTAS TELEFÓNICAS Y SU IMPLEMENTACIÓN EN ANDALUCÍA

Diego Iglesias Espinosa
Instituto de Estadística de Andalucía

Manuel Ortega Ramírez
Universidad de Granada

RESUMEN

Este documento esboza las líneas de actuación de los procesos de encuestación telefónica mediante el método de marcación al azar (RDD, *Random Digit Dialing*) en dos etapas de Mitofsky-Waskberg. Para ello, se ha llevado a cabo un estudio exhaustivo del sistema telefónico español y de su estructura de distritos en la Comunidad Autónoma de Andalucía.

A partir de la escasa información disponible sobre este tipo de datos, se procede a: describir el marco muestral, establecer las etapas del diseño muestral y determinar los tamaños muestrales y los estimadores básicos.

Palabras clave: *encuesta telefónica, técnicas RDD, marcos en Andalucía.*

Introducción

Los métodos RDD han alcanzado un gran desarrollo en Estados Unidos debido a la gran cobertura que tiene el país en cuanto a la cantidad de hogares que disponen de teléfono y al creciente volumen de números telefónicos que desaparecen de los listados o guías telefónicas por deseo expreso de los suscriptores. Esta situación ha dado lugar, desde hace décadas, al desarrollo de métodos de marcación al azar. Vamos a ver cómo funcionan estos métodos y cómo se llevaría a cabo su aplicación en la comunidad andaluza, concretamente con el método RDD en dos etapas de Mitofsky-Waskberg.

Los métodos RDD están basados en un marco que contiene todos los números posibles. Así se elimina el problema de “no cobertura” siempre que el marco se encuentre actualizado. El método RDD básico consiste en marcar un número al azar y, si el número está en uso y corresponde a un hogar, intentar hacer la entrevista. Este proceso da lugar a demasiadas marcaciones infructuosas por lo que se han desarrollado múltiples modificaciones para mejorarlo.

El método RDD en dos etapas de Mitofsky-Waskberg (1978) se enfrenta al problema de los números improductivos de una forma menos directa. Emplea un procedimiento en dos etapas que incrementa la proporción de números residenciales en la segunda etapa. La primera consiste en seleccionar al azar una unidad primaria que llamaremos PSU (Primary Sample Unit), que será un conjunto de 100 números telefónicos correlativos. El concepto viene de considerar a las unidades primarias como grupos formados por los 7 primeros dígitos que componen el número telefónico, de forma que quedan 2 dígitos por marcar para formar un número telefónico completo. Si tenemos 2 dígitos por completar y hay 10 posibilidades (del 0 al 9), queda el conjunto de 100 números telefónicos. Una vez seleccionada la PSU, se marca un número telefónico completo aleatoriamente dentro de esa PSU, si el número marcado está en uso, se considera que la PSU es válida y se selecciona para el submuestreo de la segunda etapa; si no es así, se deshecha sustituyéndola por otra hasta tener la cantidad de PSUs que habíamos considerado en la planificación de nuestro estudio. En la segunda etapa, marcamos aleatoriamente una cantidad fija de números telefónicos dentro de cada PSU seleccionada con anterioridad. La cantidad fija recomendada es $k = 8$, en realidad $k+1 = 9$, ya que la primera llamada que hacíamos en la primera etapa también se aprovechaba para intentar completar una entrevista. Si alguno de estos números no aporta una entrevista, se genera otro para sustituirlo.

El proceso de dos etapas utiliza probabilidades de selección proporcionales al tamaño para las PSUs y probabilidades iguales para las unidades elegibles como números residenciales. La proporción de números que son elegibles (asignados a hogares), es mucho mayor en la segunda etapa. Por ejemplo, en el diseño en dos etapas de Mitofsky-Waskberg, la proporción de números de teléfono que son hogares puede incrementarse desde aproximadamente el 25% en la primera etapa a más de un 60% en la segunda. Debemos aclarar que estos porcentajes provienen de experiencias de Estados Unidos, en España los porcentajes serán más bajos por las razones que se darán más adelante.

Aunque este método parece operacionalmente muy sencillo y con un bajo coste, hay muchas consideraciones a tener en cuenta antes de ponerlo en práctica. Se trata de un

tipo de muestreo muy especial y pueden surgir problemas: duplicidad en la selección (es decir, hogares con más de una línea telefónica), falta de respuesta, números telefónicos de los que se desconoce su situación (si están en uso o no), y muchos otros aspectos particulares para los que se recomienda acudir a la extensa bibliografía disponible.

Cobertura

Lo primero que hay que tener en cuenta para ver si resulta factible realizar una encuesta telefónica es conocer qué proporción de la población tiene teléfono en su hogar. Según la *Encuesta de Tecnologías de la Información en los Hogares* de 2003 realizada por el Instituto Nacional de Estadística español (INE), el porcentaje de hogares andaluces con teléfono fijo es del 83,3 %, frente al 88,1 % del territorio nacional. Este dato nos hace ver que una encuesta llevada a cabo sólo por teléfono dejaría a una parte importante de la población fuera del estudio. Sobre todo si tenemos en cuenta que la posesión de teléfono está altamente correlacionada con los ingresos familiares. También influyen otros aspectos como las características geográficas (rural o urbano), tamaño del hogar, estado laboral, edad del cabeza de familia, tipo de residencia, estado civil, ocupación, raza u origen étnico. Todo esto hace que podamos considerar que las personas excluidas del estudio serían las más desfavorecidas o las más jóvenes, sobre todo si tenemos en cuenta el enorme auge del teléfono móvil, que está pasando de ser un complemento a sustituir al teléfono fijo. Esto puede subsanarse complementando el estudio mediante otros tipos de recogida de datos.

Marco muestral

El marco muestral, como ya hemos mencionado, está constituido por todos los números telefónicos posibles. El sistema telefónico español establece números de 9 dígitos con la siguiente notación:

NXY AB MCDU

donde NXY denota el prefijo y AB indica el bloque de 10.000 números. De esta forma, las PSUs que vamos a considerar estarán determinadas por NXY AB MC, dejando DU para completar el número telefónico en la segunda etapa. Pero dentro de este sistema hay que tratar de conocer, de la mejor forma posible, cómo es la estructura telefónica en nuestro territorio de estudio. En este caso hemos considerado Andalucía.

Para cualquier organización no gubernamental o empresa de estudios mediante encuestas, resulta casi imposible conseguir información sobre los clientes de telefonía. La causa es la feroz competencia entre operadores. Cualquier tipo de dato, incluso sobre la cantidad de líneas en uso que tiene cada uno de ellos, podría ser utilizada por sus competidores.

La información a la que se tiene acceso a través de Internet, y que usamos en este documento, corresponde a los números que forman los distritos telefónicos (obtenidos de la página web de la *Comisión del Mercado de las Telecomunicaciones*, CMT) y los municipios que cubre cada distrito (obtenidos de la página del *Ministerio de Ciencia y Tele-*

comunicaciones). De la CMT se obtienen una serie de archivos del que destaca uno donde se especifican los siguientes puntos:

<i>IND</i>	<i>Dígitos NXY</i>
<i>Bloque</i>	<i>Dígitos AB (bloques de 10.000 números)</i>
<i>Sub-bloque</i>	<i>Dígitos M,C,D</i>
<i>Provincia</i>	<i>Provincia telefónica</i>
<i>Nombre de distrito</i>	
<i>Código de distrito</i>	
<i>Operador</i>	
<i>Fecha de notificación</i>	

Se computa un total de 76 distritos, de forma que los datos sobre cada uno de ellos estaban actualizados a fecha de 6 Mayo de 2003.

Si asociamos los dos conjuntos de datos de que disponemos, obtenemos suficiente información como para dirigir la marcación aleatoria dentro del ámbito que estamos considerando. La Tabla 1 muestra los datos del distrito de Abla para presentar de una forma más clara qué es exactamente lo que podemos saber de cada distrito.

Tabla 1: *distrito de Abla, 461 (Almería)*

INP	Bloque	Sub-blq	PSUs	Operador
950	2	0	10	RETEVISIÓN I, S.A.U.
950	4	0	10	UNI2 TELECOMUNICACIONES, S.A.U.
950	10	2	10	AIRTEL MÓVIL, S.A.
950	16	1	10	TELEFÓNICA DE ESPAÑA, S.A.U.
950	35		100	TELEFÓNICA DE ESPAÑA, S.A.U.
950	52	1	10	TELEFÓNICA DE ESPAÑA, S.A.U.
950	54	0	10	TELEFÓNICA DE ESPAÑA, S.A.U.
950	63	2	10	TELEFÓNICA DE ESPAÑA, S.A.U.
950	75	8	10	BT IGNITE ESPAÑA, S.A.U.
950	95	9	10	JAZZ TELECOM, S.A.U.
950	98	5	10	TELEFÓNICA DATA ESPAÑA, S.A.
		Total	200	
Municipios: ABLA, ABRUCENA, FIÑANA, JERGAL, NACIMIENTO, OLULA DE CASTRO, LAS TRES VILLAS.				

En este caso, el distrito de Abla cubre 7 municipios. Como se puede observar en la tabla, en algunos casos podemos conocer los sub-bloques, y así restringir mucho más la búsqueda. Además, al especificar el operador que administra el bloque o sub-bloque, es posible llamar a dicho operador si surgen dudas sobre el estado del número al que se está llamando. Mediante la observación de los municipios de todos los distritos, hemos observado que se producen solapamientos entre distritos adyacentes, es decir, hay varios distritos que aportan cobertura a un mismo municipio. Por ejemplo, el municipio de

Nijar es el que presenta mayor solapamiento ya que está cubierto por tres distritos telefónicos distintos (Almería, Nijar y Vera); el resto son 33 municipios que están cubiertos por parejas de distritos. Esto puede ser un problema si el estudio requiere una identificación a nivel local y el municipio aparece en varios distritos, aunque siempre se podría preguntar al entrevistado sobre su situación geográfica. De todas maneras este tipo de métodos están indicados sobre todo para estudios a grandes poblaciones, para estudios locales es más fácil y práctico acudir a las guías telefónicas.

Tratando todos los distritos, como el que hemos mostrado de ejemplo, podemos saber cuántas PSUs hay en cada uno de ellos. La cantidad varía mucho, como puede observarse en la Tabla 2, sobre todo en torno a las capitales. Mostramos sólo los datos de dos de las ocho provincias andaluzas, a modo de representación.

Tabla 2: *distritos Telefónicos de Andalucía y las PSUs de las que están compuestos.*

Distrito	Provincia	PSUs	Distrito	Provincia	PSUs
Almería	Almería	3210	Abla	Almería	200
Adra	Almería	780	Albox	Almería	480
Canjajar	Almería	200	Nijar	Almería	320
Tijola	Almería	190	Velez Rubio	Almería	190
Vera	Almería	700			
...
Sevilla	Sevilla	11.220	El Arahál	Sevilla	380
Cantillana	Sevilla	370	Carmona	Sevilla	700
Constantina	Sevilla	280	Ecija	Sevilla	340
Lora del Río	Sevilla	280	Morón de la Frontera	Sevilla	300
Osuna	Sevilla	580	Pilas	Sevilla	280
Utrera	Sevilla	870			

Con esta información ya sabemos cuántas PSUs hay en cada distrito y los prefijos, bloques y sub-bloques a los que hay que llamar para alcanzar un hogar de la población que es objeto de estudio. En este caso, la estructura tiene una organización casi idílica para este tipo de metodología. Decimos esto porque no hay distritos andaluces que cubran municipios de las comunidades colindantes, como ocurre en otros lugares de España. Sí que surge el problema contrario (es decir, hay municipios o parte de ellos que no pueden ser alcanzados si no es a través de distritos de una comunidad vecina). En nuestro caso, sólo se da esta situación en dos municipios de Jaén: Benatae y Torres de Albánchez; que están parcialmente cubiertos por el distrito de Alcazar, que pertenece a Albacete. Otra pequeña excepción, de distinto tipo, se da en el municipio jienense de Lopera, que es cubierto parcialmente por el distrito de Montoro, que pertenece a Córdoba. En este caso, si estuviéramos llevando a cabo un estudio sólo sobre la población de Jaén, habría que pensar en la forma de acceder a esta parte de la población que se queda fuera de nuestro marco inicial, o corregir esa subcobertura de alguna forma a posteriori.

Diseño muestral

La cobertura la proporcionan 76 distritos telefónicos con un total de 73.110 PSUs. Conocemos la distribución de las PSUs en cada distrito. Partimos de un muestreo estratificado con 76 estratos de la siguiente forma:

$$N = 73.110 \quad \left| \begin{array}{l} N_1 = 3.210 \\ N_2 = 200 \\ \dots \\ N_{76} = 870 \end{array} \right. \begin{array}{l} \rightarrow n_1 \\ \rightarrow n_2 \\ \dots \\ \rightarrow n_{76} \end{array}$$

Si determinamos un tamaño muestral de, por ejemplo, $n_1 = 25$, significa que hemos escogido 25 unidades primarias en el estrato 1, es decir, en el distrito de Almería. Al no tratarse de unidades elementales, necesitamos un submuestreo dentro de cada una de ellas. Su tamaño lo conocemos: 100 unidades elementales dentro de cada PSU; y el tamaño de la submuestra lo vamos a tomar como $k + 1 = 9$, debido a que en los textos especializados se señala en multitud de ocasiones como el tamaño más conveniente en base a la experiencia. Se trata de aplicar un muestreo bietápico dentro de cada estrato. Por ejemplo, para el primero de ellos:

$$N_1 = 3.210 \rightarrow n_1 = 25 \quad \left| \begin{array}{l} M_1 = \bar{M} = 100 \\ M_2 = \bar{M} = 100 \\ M_3 = \bar{M} = 100 \\ \dots \\ M_{25} = \bar{M} = 100 \end{array} \right. \begin{array}{l} \rightarrow M_1 = \bar{m} = 9 \\ \rightarrow M_2 = \bar{m} = 9 \\ \rightarrow M_3 = \bar{m} = 9 \\ \dots \\ \rightarrow m_{25} = \bar{m} = 9 \end{array}$$

Tendríamos un tamaño muestral para el estrato 1 (distrito de Almería) de 225 unidades elementales.

Tamaño muestral

Antes de tratar el tema de la determinación del tamaño muestral de forma estricta, debemos notar que en las encuestas telefónicas habrá que llamar a muchos números generados al azar antes de encontrar uno en uso, que pertenezca a un hogar y que acepte ser entrevistado. Para obtener una muestra final que tenga un tamaño cercano al deseado, podemos considerar la siguiente expresión:

$$\text{Tamaño muestral} = \frac{\text{Entrevistas Completas}}{(\text{Razón de teléfonos en uso} \times \text{Incidencia} \times \text{Razón de Realización})}$$

donde,

- *Razón de teléfonos en uso (Working Phones Rate, WPR)*

Según *Survey Sampling, Inc.*, la WPR puede variar, generalmente, de un 65 % a un 40 %. Sin embargo, hemos comprobado que en España las razones son aún más bajas. Esto puede explicarse por la gran cantidad de operadores que han irrumpido en el mercado, reservando bloques de números para su uso que no ocupan.

Si hay 2.431.805 hogares en Andalucía (INE, *Censos de Población y Viviendas 2001*) y el 83,3 % tienen teléfono, obtenemos 2.025.694 hogares con teléfono. Si la cantidad de PSUs era 73.110, eso da lugar a 7.311.000 líneas telefónicas teóricas. Suponiendo que cada hogar tiene una sola línea telefónica fija, y excluyendo a los teléfonos correspondientes a negocios (que en principio no nos interesan) y a los escasos hogares de los municipios que eran cubiertos por el distrito de Albacete (que nos interesan pero que no se pueden identificar entre las PSUs del distrito albaceteño, salvo quizás mediante la tediosa observación de la guía telefónica correspondiente), la razón de teléfonos en uso sería del 27,7 %.

- *Incidencia*

La incidencia es el porcentaje existente en la población completa del tipo de respondiente que se desea para la entrevista. En nuestro caso, si se desea muestrear a toda la población de hogares, es del 100 %. Si no hubiéramos excluido a los teléfonos de negocios del cálculo de la razón de teléfonos en uso, ahora tendríamos que averiguar qué porcentaje de números en uso pertenecen a hogares y cuáles a negocios, lo que se antoja bastante complicado. Al considerar los teléfonos de negocios como en desuso incurrimos en un error de concepto pero evita el problema de discriminar sobre la naturaleza del número telefónico y no causa ningún inconveniente al objetivo de estimar el tamaño muestral.

- *Razón de realización*

La longitud del estudio y la sensibilidad de la materia afectarán al porcentaje de respondientes que completen la encuesta. En base a la extensa experiencia de *Survey Sampling, Inc.* en este campo, se obtuvieron algunos valores de este porcentaje en función de esos dos factores, que pueden observarse en la Tabla 3.

Tabla 3: *razón de respondientes según la longitud y la sensibilidad de la encuesta.*

Longitud de entrevista	No-sensible	Sensible
Corta (menos de 5 min)	0.35	0.30
Media (6-20 min)	0.33	0.30
Larga (20-30 min)	0.30	0.28
Muy larga (más de 30 min)	0.25	0.23

Las razones de realización de la Tabla 3 tuvieron en cuenta la no respuesta, contestación de máquinas, negativas inmediatas y señales de ocupado, asumiendo tres esfuerzos de llamada para cada número telefónico.

Por ejemplo, si se necesitan 1.000 entrevistas completas, y la razón de teléfonos en uso es 0.277, la incidencia es 1.00, y la razón de realización se estima en 0.33; entonces se requeriría llamar a 10.940 números ($1.000/(0.277 \times 1.00 \times 0.33)$).

De manera más formal, en la Tabla 4 se detallan los tamaños muestrales para distintos niveles de significación mediante probabilidades proporcionales al tamaño, sólo para los municipios de las provincias de Almería y Sevilla. Se han considerado las fórmulas de la varianza para la proporción mediante un muestreo estratificado simple. La finalidad de esta consideración ha sido simplificar los cálculos, ya que se complicaban en exceso al considerar un submuestreo en los estratos. También se han detallado estas fórmulas y los valores tomados para los cálculos.

Tabla 4: *Estimación del tamaño muestral mediante muestreo estratificado*

Distritos	PSUs (N_h)	n_h				
		$e = 0'1$	$e = 0'05$	$e = 0'01$	$e = 0'005$	$e = 0'001$
Almería	3210	5	17	374	1107	2984
Adra	780	2	5	91	269	725
Canjayar	200	1	2	24	69	186
Tijola	190	1	1	23	66	177
Vera	700	1	4	82	242	651
Abla	200	1	2	24	69	186
Albox	480	1	3	56	166	447
Nijar	320	1	2	38	111	298
Velez Rubio	190	1	1	23	66	177
...
Sevilla	11220	15	59	1304	3868	10428
Cantillana	370	1	2	44	128	344
Constantina	280	1	2	33	97	261
Lora del Rio	280	1	2	33	97	261
Osuna	580	1	4	68	200	540
Utrera	870	2	5	102	300	809
El Arahal	380	1	2	45	131	354
Carmona	700	1	4	82	242	651
Ecija	340	1	2	40	118	316
Morón de la Frontera	300	1	2	35	104	279
Pilas	280	1	2	33	97	261
Total	73110	136	417	8534	25242	67981

$L = 76$
 $N = 73110$
 $Ph = Qh = 0.5$
 $T = 1,96$
 $T^2 = 3,8416$
 $n_h = k * N_h$
 $Sum = 3,4231E-06$

$$e^2 = T^2 \left(\frac{1}{k} - 1 \right) \sum_{1 \leq h \leq L} \frac{W_h^2}{N_h - 1} P_h Q_h = T^2 \left(\frac{1}{k} - 1 \right) \cdot Sum$$

$$k = \frac{T^2 \sum_{1 \leq h \leq L} \frac{W_h^2}{N_h - 1} P_h Q_h}{e^2 + \left(T^2 \sum_{1 \leq h \leq L} \frac{W_h^2}{N_h - 1} P_h Q_h \right)} = \frac{T^2 \cdot Sum}{e^2 + (T^2 \cdot Sum)}$$

$$Sum = \sum_{1 \leq h \leq L} \frac{W_h^2}{N_h - 1} P_h Q_h$$

De forma que los tamaños totales se presentan en la Tabla 5. Hay que tener en cuenta que en la práctica es casi imposible conseguir un tamaño de muestra que se acerque al que se había fijado en un principio, sobre todo por las altas razones de no respuesta. Por ese motivo se ha introducido la fórmula anterior que tenía en cuenta este aspecto y que, aunque sencilla, puede dar una idea de los resultados prácticos.

Tabla 5: *tamaños muestrales finales.*

e	K	N = K*N	n final
0,1	0,00131328	96,0138271	136
0,05	0,0052325	382,548131	417
0,01	0,11621788	8496,68893	8534
0,005	0,34469303	25200,5074	25242
0,001	0,92932892	67943,2373	67981

Estimadores para el total, media y proporción

Notación

N_h , cantidad de PSUs en el h-ésimo estrato o distrito; $h = 1, \dots, 76$.

n_h , cantidad de PSUs de la muestra en el h-ésimo estrato o distrito; $h = 1, \dots, 76$.

$W_h = \frac{N_h}{N}$, peso del h-ésimo estrato; $h = 1, \dots, 76$.

\bar{x}_h , media muestral del h-ésimo estrato; $h = 1, \dots, 76$.

s_h^2 , cuasivarianza muestral del h-ésimo estrato; $h = 1, \dots, 76$.

\bar{x}_{jh} , media muestral de la j-ésima unidad primaria (PSU) en el h-ésimo estrato o distrito;
 $h = 1, \dots, 76, j = 1, \dots, n_h$.

$\bar{M} = 100$, tamaño de cada PSU, que será igual para todos.

$\bar{m} = 9$, tamaño de la muestra dentro de cada PSU seleccionado, que será igual para todos salvo que se seleccione una PSU tan escaso en números de hogares que resulte imposible obtener el tamaño requerido.

x_{ijh} = valor de la i -ésima observación en la j -ésimo PSU seleccionado del h -ésimo estrato; $h = 1, \dots, 76, j = 1, \dots, n_h, i = 1, \dots, \bar{m}$.

\hat{P}_h , proporción de unidades con la característica requerida en el h -ésimo estrato; $h = 1, \dots, 76$.

$A_{ijh} = 1$, si la i -ésima unidad elemental de la j -ésima PSU en el estrato h presenta la característica requerida, siendo 0 en caso contrario; $h = 1, \dots, 76, j = 1, \dots, n_h, i = 1, \dots, \bar{m}$.

Media

$$\hat{x}_{st} = \frac{1}{N} \sum_{h=1}^{76} N_h \bar{x}_h \quad \bar{x}_h = \frac{1}{n_h} \sum_{j=1}^{n_h} \bar{x}_{jh} \quad \bar{x}_{jh} = \frac{1}{\bar{m}} \sum_{i=1}^{\bar{m}} x_{ijh}$$

Total

$$\hat{x}_{st} = N^* \hat{x}_{st}$$

Proporción

$$\hat{P}_{st} = \frac{1}{N} \sum_{h=1}^{76} N_h \hat{P}_h \quad \hat{P}_h = \frac{1}{\bar{m} n_h} \sum_{j=1}^{n_h} \sum_{i=1}^{\bar{m}} A_{ijh}$$

El estimador del total presenta un problema inherente al diseño que se ha considerado. Esto se debe a que los N_h y N considerados no se refieren a totales de la población, si no a totales de PSUs de tamaño 100 que en la mayoría de los casos estarán en desuso o no corresponderán a hogares. De hecho, si todos los números de los distritos estuvieran en uso, tendríamos un total de 7.311.000 líneas telefónicas, cuando la comunidad andaluza tenía 7.357.558 habitantes (Censo del 2001). Casi habría un teléfono fijo por habitante cuando habíamos visto que solo el 83.3 % de los hogares andaluces tenían teléfono. Los estimadores de la media y la proporción no se ven afectados por esta consideración debido a que los N_h se dividen entre N , obteniéndose un W_h que es proporcional al que se obtendría si consideráramos datos de la población. El problema con el estimador del total es que si multiplicamos por N la media estimada, tendremos un valor estimado del total muy diferente del valor real. Este inconveniente se puede resolver sustituyendo N por lo que hemos denotado por N^* , que sería el total de hogares o de personas en Andalucía, dependiendo de si estamos midiendo características del hogar o personales.

Estimación de las varianzas

$$\text{Media} \quad \text{Var}(\hat{x}_{st}) = \sum_{h=1}^{76} W_h^2 \text{Var}(\bar{x}_h) = \sum_{h=1}^{76} W_h^2 \left[(1 - f_{1h}) \frac{\hat{s}_{1h}^2}{n_h} + (1 - f_{2h}) \frac{\hat{s}_{2h}^2}{n_h \bar{m}} \right]$$

donde,

$$f_{1h} = \frac{n_h}{N_h}$$

$$f_{2h} = \frac{\bar{m}}{M} = \frac{9}{100} = 0.09$$

$$\hat{s}_{1h} = \sum_{j=1}^{n_h} \frac{(\bar{x}_{jh} - \bar{x}_h)^2}{n_h - 1}$$

$$\hat{s}_{2h} = \frac{1}{n_h} \sum_{j=1}^{n_h} \sum_{i=1}^{\bar{m}} \frac{(x_{ijh} - \bar{x}_{jh})^2}{\bar{m} - 1}$$

Total

$$Var(\hat{x}_{st}) = M^2 Var(\hat{\bar{x}}_{st})$$

Donde M^2 es el total de unidades secundarias. Como ocurría en el estimador del total, esta expresión se aplica cuando todas las unidades secundarias son elegibles o existen. El valor de M que deberíamos considerar es el total de líneas telefónicas en uso que son objeto de estudio.

$$Var(\hat{P}_{st}) = \sum_{h=1}^{76} W_h^2 Var(\hat{P}_h) =$$

Proporción

$$\sum_{h=1}^{76} W_h^2 \left[\frac{1-f_{1h}}{n_h} \sum_{j=1}^{n_h} \frac{(P_{jh} - \hat{P}_h)^2}{n_h - 1} + \frac{f_{1h}(1-f_{2h})}{n_h^2(\bar{m}-1)} \sum_{j=1}^{n_h} P_{jh}(1-P_{jh}) \right]$$

donde,

$$P_{jh} = \sum_{i=1}^{\bar{m}} \frac{A_{ijh}}{\bar{m}} = \sum_{i=1}^9 \frac{A_{ijh}}{9}, \text{ proporción muestral de unidades en la } j\text{-ésima PSU del } h\text{-ésimo estrato que presentan la característica requerida.}$$

Conclusión

Aunque este tipo de métodos lleva usándose y desarrollándose desde hace décadas en países como Estados Unidos (prueba de ello es la gran cantidad de modificaciones que han ido aportando distintos autores a lo largo de los años), el nivel de cobertura que teníamos en España hacía inviable su aplicación en nuestro país. Este problema se ha superado debido al desarrollo experimentado en los últimos años.

Sin embargo, aunque el marco ha pasado a ser apropiado, el cambio ha sido tan rápido que se ha llegado a la situación a la que se enfrentan los investigadores de los países más experimentados en estas técnicas. Esto se explica por el hecho de la enorme expansión del teléfono móvil, que puede llegar a ser, si no lo es ya, un sustituto del teléfono fijo (como ocurre en otros países). El Ministerio de Ciencia y Tecnología presenta datos estadísticos que pueden reflejar este cambio. En 2002, los hogares españoles con teléfo-

no fijo llegaban al 90,2%, mientras que en 2003 bajaron a un 88,1%. El caso contrario al de los teléfonos móviles, que han pasado de un 65% a un 73,7% para el mismo período. Ya se han realizado encuestas a través de teléfono móvil pero más como complemento que como medio único.

Teóricamente, los métodos RDD podrían extenderse a los teléfonos móviles. En la práctica, la dificultad puede deberse a que los teléfonos móviles tienen una naturaleza de aparato personal, en lugar de corresponder a un hogar. El muestreo de personas dentro de un hogar, contactadas mediante el teléfono móvil de uno de los miembros, es muy complicado. Además, las personas contactadas podrían estar en cualquier parte. Las fuentes consultadas para la realización de este documento no aportan mucha información sobre la situación de la telefonía móvil, solo aportan los siguientes datos:

Dígitos	Dígitos NXY
Tipo	Tecnología del servicio móvil
Estado	Asignado, reservado, Situación especial
Operador	
Fecha estado	Fecha de la asignación o reserva

Lo que no representa ninguna información de tipo geográfico, un gran inconveniente para delimitar el ámbito de la población en estudio.

En una encuesta realizada en 1999 por Perone, Matrondola y Soverini, se encontró que las razones de rechazo no excedían de las encontradas en las encuestas a teléfonos fijos. Sin embargo, las razones de "no contacto" eran más altas, principalmente debido a encontrarse con teléfonos apagados o fuera de cobertura. En otros estudios se ha observado que la duración de las entrevistas suele ser mayor cuando se realiza a través del móvil.

Pese a los nuevos retos que plantea la metodología RDD, no deja de ser interesante conocer y experimentar con estos métodos para ver qué resultados pueden aportar, y estar preparados para los nuevos avances en el campo de las encuestas telefónicas mediante el conocimiento de los pasos previos.

Referencias

- Lepkowski, J. M. (1988). Telephone sampling methods in the United States. En R.M. Groves y otros: *Telephone Survey Methodology*. New York: Wiley, 73-98.
- Massey, J. T. (1988). An overview of telephone coverage. En R.M. Groves y otros: *Telephone Survey Methodology*. New York: Wiley, 3-8.
- Nathan, G. (2001). Telesurvey methodologies for household surveys – a review and some thoughts for the future. *Survey Methodology* 27, 7-31.
- Noble, I., Moon, N. and McVey, D. (1998). *Using RDD telephone methods for large scale social policy and opinion research in the UK*. MRS Annual Conference, 25-27 March 1998.
- Rueda García, M. Del M. y Arcos Cebrián, A. (1998). Problemas de muestreo en poblaciones finitas. Grupo Editorial Universitario.

- Sebold, J. (1988). Survey period length, unanswered numbers, and nonresponse in telephone surveys. En R.M. Groves y otros: *Telephone Survey Methodology*. New York: Wiley, 247-256.
- Thornberry, O.T. Jr. and Massey, J. T. (1988). Trends in United States telephone coverage across time and subgroups. En R.M. Groves y otros: *Telephone Survey Methodology*. New York: Wiley, 25-49.
- Trewin, D. and Lee, G. (1988). International comparisons of telephone coverage. En R.M. Groves y otros: *Telephone Survey Methodology*. New York: Wiley, 3-24.
- Waksberg, J. (1978). Sampling Methods for Random Digit Dialing. *Journal of the American Statistical Association*, 73 (361), Application Section.