

Tratamiento de la no respuesta en encuestas de
panel en el caso de poblaciones finitas

Margarita Antía
Ana Coimbra

Octubre 2009

Serie DT (09/02)
ISSN : 1688-6453

Tratamiento de la no respuesta en encuestas de panel en el caso de poblaciones finitas.

Margarita Antía - Ana Coimbra

Este trabajo pretende mostrar cómo tratar la no respuesta en un caso particular de las encuestas por muestreo como son las encuestas de panel. Éstas refieren a estudios basados en observaciones repetidas efectuadas sobre las mismas unidades de muestreo: personas, hogares, empresas, etc. La muestra es extraída por única vez al inicio del estudio y todas las unidades seleccionadas serán visitadas a lo largo de la duración del panel.

Los distintos momentos del tiempo en los que las encuestas son llevadas a cabo se denominan “olas”; la duración del panel y el período entre olas son definidos en la etapa del diseño de la encuesta.

La medición periódica de elementos permite realizar un seguimiento de la población objetivo, logrando captar su dinámica en el tiempo. Los resultados particulares en cada instancia de medición (estimaciones transversales) pueden ser obtenidos sin perjuicio de lo anterior y, aunque no sea el objetivo principal de las encuestas de panel suelen ser de interés en sí mismos.

Un problema usual en las encuestas de panel es la mortalidad de unidades a lo largo del tiempo. En términos generales, esta dificultad puede pensarse como un problema de no respuesta, definida como la imposibilidad de obtener toda o alguna información para una o más de las unidades seleccionadas en la muestra. Este es un fenómeno presente en la mayoría de las encuestas por muestreo y es imprescindible su tratamiento para evitar sesgos en las estimaciones. La inclusión del factor “tiempo” en las encuestas de panel provoca un agravamiento del problema de no respuesta respecto a las encuestas cross-section, reflejado en reducciones considerables en el “tamaño de muestra” período a período debido a la movilidad,

el fallecimiento y otros factores (como la pérdida de cooperación de unidades) que resultan en el “agotamiento” del panel. Otro efecto causado por la inclusión del factor “tiempo” es la potencial pérdida de representatividad de la muestra para inferir resultados transversales en olas posteriores a la primera.

1. Estimador de cambio bajo condiciones ideales

Las condiciones ideales en cualquier tipo de encuesta por muestreo están regidas por la obtención de respuesta completa de las unidades muestreadas a partir de un marco muestral perfecto acorde a la población objetivo, sin presencia de errores de medición. En las encuestas por panel, dichas condiciones requieren el supuesto adicional de que la población objetivo sea fija en el tiempo. Adicionalmente se requiere obtener respuesta completa de todas las unidades en todas las olas. Esto es, s es una muestra aleatoria de la población U de N individuos, tomada bajo un diseño $p(s)$ ¹ de tamaño n_s el cual genera probabilidades de inclusión π_k .

Bajo estos supuestos, el estimador de cambios de la variable de interés es :

$$\begin{aligned}\hat{A}_{j,j+h} &= (\hat{t}_{j+h} - \hat{t}_j) \\ &= \sum_s \left(\frac{y_{(j+h)k} - y_{jk}}{\pi_k} \right) \\ &= \sum_s \frac{a_{(j,j+h)k}}{\pi_k}\end{aligned}\quad (1)$$

donde $a_{(j,j+h)k}$ es cambio individual del elemento k en la variable de interés entre las olas j y $j+h$.

Los estimadores $\hat{A}_{j,j+h}$ tienen la forma de estimadores π (Horvitz-Thompson), por lo tanto, también comparten sus propiedades. La varianza de $\hat{A}_{j,j+h}$ también puede expresarse como

$$Var_{p(s)}(\hat{A}_{j,j+h}) = Var(\hat{t}_j) + Var(\hat{t}_{j+h}) - 2Cov(\hat{t}_j, \hat{t}_{j+h}) \quad (2)$$

y dado que se estos totales están calculados en base a las mismas unidades, se espera los totales estimados en olas sucesivas estén correlacionados positivamente resultando en una varianza pequeña del estimador de diferencias.

¹La elección del diseño a utilizar depende del objeto de estudio y no de la utilización de paneles.

Una alternativa a la utilización de paneles es la estimación de diferencias de la variable de interés utilizando totales estimados mediante encuestas cross-section en los momentos j y $j + h$. En este caso el estimador de diferencias también será insesgado pero su varianza será mayor que en el caso anterior, ya que la independencia entre muestras en una y otra instancia determina que el tercer término de (2) sea cero.

Este es un aspecto importante en la justificación de utilización de paneles frente a encuestas cross-section para medir cambios.

2. Calibración como tratamiento de la no respuesta

El estimador propuesto tiene la cualidad de ser insesgado y de sencillo cálculo, ya que es un estimador π . Para su desarrollo, se partió de supuestos muy restrictivos rara vez presentes en la práctica, a saber: la existencia de *respuesta perfecta* y población fija en el tiempo, reflejados en ponderadores constantes en el tiempo para cada elemento e iguales al inverso de su probabilidad de inclusión en la muestra.

El levantamiento del supuesto de respuesta perfecta plantea la necesidad de estudiar cuales son los efectos que provoca la imposibilidad de obtener toda o alguna información para todas o algunas unidades seleccionadas en la muestra.

Si la no respuesta se presentara de manera completamente aleatoria, el único inconveniente al que se enfrenta el investigador es la reducción del tamaño de muestra y un consiguiente aumento en la varianza de las estimaciones, que podría ser fácilmente contrarrestado mediante un “sobremuestreo” (fijando un tamaño de muestra mayor en la etapa de diseño). De esta manera el único efecto negativo de la no respuesta sería el incremento en la carga administrativa y los costos de recolección de datos. En la práctica, la situación anterior sería una “feliz” casualidad. Las unidades que no contestan “normalmente” difieren en algunas características de aquellas que sí lo hacen, y el sesgo introducido en las estimaciones por esta causa constituye el obstáculo más importante por corregir. Frente a la pérdida del insesgamiento de los estimadores, el incremento de su varianza es un disturbio menor: en presencia de sesgo significativo, un intervalo de confianza calculado estará centrado en un valor erróneo y no se logra el nivel de confianza requerido.

Sea \hat{t}_y el estimador de t_y cuando hay respuesta completa, o sea cuando el conjunto de respondentes r coincide con la muestra s ; sea \hat{t}_{yNR} el estimador de t_y en presencia

de no respuesta. El *error de estimación* de \hat{t}_{yNR} puede expresarse mediante un término que representa el *error muestral* y otro que representa el *error por no respuesta*.

$$Error = \hat{t}_{yNR} - t_y = (\hat{t}_y - t_y) + (\hat{t}_{yNR} - \hat{t}_y) \quad (3)$$

Siendo

- $\hat{t}_y - t_y$ el error muestral (el error que surge por elegir y observar una muestra, en vez de observar toda la población).
- $\hat{t}_{yNR} - \hat{t}_y$ el error por no respuesta (error que surge por la no existencia de respuesta completa).

El *sesgo total de \hat{t}_{yNR}* se obtiene calculando el valor esperado bajo los mecanismos de selección y de respuesta de los dos componentes de error previamente definidos:

$$\begin{aligned} B_{pq}(\hat{t}_{yNR}) &= B_{SAM} + B_{NR} \\ &= E(\hat{t}_{yNR} - \hat{t}_y) + E(\hat{t}_y - t_y) \end{aligned} \quad (4)$$

El término B_{SAM} (sesgo muestral) es cero o irrelevante para la mayoría de los propósitos prácticos, por lo tanto el sesgo de \hat{t}_{yNR} se convierte casi enteramente en el sesgo por no respuesta. Esto evidencia la necesidad de realizar algún tipo de tratamiento en la etapa de estimación.

En general se distinguen dos tipos de no respuesta: no respuesta al ítem, que refiere a faltantes en la respuesta para un ítem en particular del formulario debido a omisión (tanto del entrevistador como del entrevistado) o negativa del encuestado a contestar; y no respuesta de la unidad, que se da cuando la unidad seleccionada para ser entrevistada no es encontrada o se rehúsa a participar en la encuesta.

Las técnicas dominantes en la literatura actual para el tratamiento de la no respuesta son la calibración y la imputación (Särndal y Lundström (2005)). Usualmente la calibración predomina en el tratamiento para el caso de no respuesta de unidades; mientras que la imputación es más extensamente aplicada en los problemas de no respuesta de ítems. La primera es una estrategia global, tratando todas las variables de forma simultánea, mientras que la segunda es particular, específica de cada variable. La decisión acerca de utilizar una u otra no es obvia, y depende de distintos factores como lo son: la cantidad y el número de olas, el tipo de análisis

a llevarse a cabo, la disponibilidad de variables auxiliares con poder predictivo de los valores faltantes y el costo de implementar los procedimientos.

No obstante lo anterior, se han ensayado soluciones que aplican ambas técnicas en conjunto (Deville y Särndal (1994)).

La imputación es el procedimiento a través del cual los valores faltantes en una o más variables de estudio se completan con sustitutos. Los valores perdidos en la base de datos se reemplazan por los valores “plausibles” dando como resultado una matriz completa de valores. Existen varios métodos de imputación que básicamente difieren en como definen “plausible”, pero la mayoría coinciden en la necesidad de utilización de información auxiliar.

La calibración, o más precisamente el uso de estimadores calibrados, se basa fuertemente en el uso de información auxiliar tanto a nivel poblacional como a nivel de la muestra original. Su creciente popularidad puede explicarse porque no se basa en la especificación de un modelo de no respuesta, brinda un enfoque unificado dentro de la teoría del muestreo de poblaciones finitas, es computacionalmente sencilla de implementar y generaliza otras técnicas del tratamiento de la no respuesta como la post-estratificación, el raking y algunos casos de los ajustes basados en la teoría del muestreo en dos fases. El enfoque de calibración abarca a una familia de estimadores \hat{t}_{yw} cuyos miembros corresponden a diferentes inputs de información.

Es deseable que los estimadores afectados por la no respuesta sean útiles para estimar los totales de las variables de interés, que no estén sesgados y que tengan varianza reducida. El sistema de ponderadores que surge del enfoque de calibración verifica también, que al ser aplicado a las variables auxiliares reproduce el input de información auxiliar. La idea central de los estimadores calibrados es sencilla, consiste en modificar los ponderadores originales de la muestra minimizando alguna función de distancia entre dichos ponderadores y los ponderadores finales (o calibrados) y de manera que estos últimos estimen sin error algunas cantidades conocidas con los datos de los respondentes.

Sea w_k el peso calibrado para $k \in r$, luego, el estimador de $t_y = \sum_U y_k$ es:

$$\hat{t}_{yw} = \sum_r w_k y_k \quad (5)$$

Se busca el conjunto de valores w_k para todo $k \in r$ que satisfaga la *ecuación de calibración*:

$$\mathbf{X} = \sum_r w_k \mathbf{x}_k \quad (6)$$

Se dice que estos pesos w_k están calibrados al input de información \mathbf{X} , ya que cuando se aplican al vector auxiliar \mathbf{x}_k reproducen exactamente la información dada en \mathbf{X} .

Como resultado de la selección de la muestra, a cada elemento k le corresponde el peso $d_k = \frac{1}{\pi_k}$. En presencia de no respuesta, $\sum_r d_k y_k$ subestima $\sum_U y_k$, en una magnitud $\sum_{s=r} d_k y_k$ (en el caso que la variable de interés tome solamente valores positivos). Es por esto que los d_k deben ser modificados. Se buscarán nuevos pesos que sean mayores que d_k al menos para la mayoría de los respondentes, de manera de compensar la pérdida de unidades. Los nuevos ponderadores $w_k = d_k v_k$ se obtienen “aumentando” los pesos originales mediante el factor v_k , que reflejará las características individuales conocidas de los elementos $k \in r$ (resumidas en el vector \mathbf{x}_k), y puede pensarse como una función del vector auxiliar $v_k = F(\lambda' \mathbf{x}_k)$, donde λ es un vector de la misma dimensión que \mathbf{x}_k y se determinará para que se verifique la ecuación de calibración.

La clave para una calibración exitosa es el uso de información auxiliar poderosa; permitiendo reducir tanto sesgo como la varianza. La efectividad del estimador de calibración para controlar el sesgo ocasionado por la no respuesta dependerá de propiedades del vector auxiliar. Särndal y Lundström (2005) realizaron un estudio de Simulación Monte Carlo a través del cual se obtiene evidencia empírica de la fuerte relación existente entre el sesgo del estimador de calibración que proviene de la no respuesta y la información auxiliar utilizada para calibrar. Éste será menor cuanto más estrecha sea la relación entre la información auxiliar y la probabilidad de respuesta o la variable de interés. Si el sesgo es modesto el intervalo de confianza será válido y la probabilidad de cobertura será cercana al nivel de confianza requerido. Limitar el sesgo de las estimaciones en presencia de no respuesta se tornará en la mayor preocupación, la minimización de la varianza pasará a segundo plano ya que de nada sirve que un estimador presente varianza chica cuando está fuertemente sesgado.

3. Estimadores calibrados en encuestas de panel

La no respuesta de unidades en una ola es una forma de no respuesta parcial particular al muestreo por paneles, generando distintos patrones de respuesta a lo largo del estudio, como muestra el siguiente cuadro. Algunos miembros de la muestra pueden abandonar la encuesta en cierta ola y perderse para el resto del estudio (desertores); mientras que otros pueden perderse en una ola, y volver al panel en alguna de las siguientes (respondentes episódicos).

| Patrón | Estado de respuesta | Ola 1 | Ola 2 | Ola 3 | Ola 4 | Ola 5 |
|--------|---------------------|-------|-------|-------|-------|-------|
| 1 | Respondentes | x | x | x | x | x |
| 2 | No | x | x | x | x | - |
| 3 | Respondentes | x | x | x | - | - |
| 4 | por | x | x | - | - | - |
| 5 | Desgaste | x | - | - | - | - |
| 6 | No | x | x | - | x | x |
| 7 | Respondentes | x | - | - | x | x |
| 8 | Episódicos | x | - | - | - | x |

Ref: x: respuesta, -: no respuesta

La no respuesta en panel (manifestada bajo los patrones de *no respuesta por desgaste* y *no respuesta episódica*) genera en cada ola, un conjunto de respondentes r_i , todos incluidos en la muestra s . Esto requerirá el cálculo de ponderadores calibrados particulares a cada individuo respondente en cada ola.

Estimación transversal

Para la estimación transversal se calcularán los ponderadores de las unidades respondentes en cada ola. El estimador calibrado del total correspondiente a la ola i , \hat{t}_{ywi} se define por:

$$\hat{t}_{ywi} = \sum_{r_i} w_{k_i} y_{k_i} \quad (7)$$

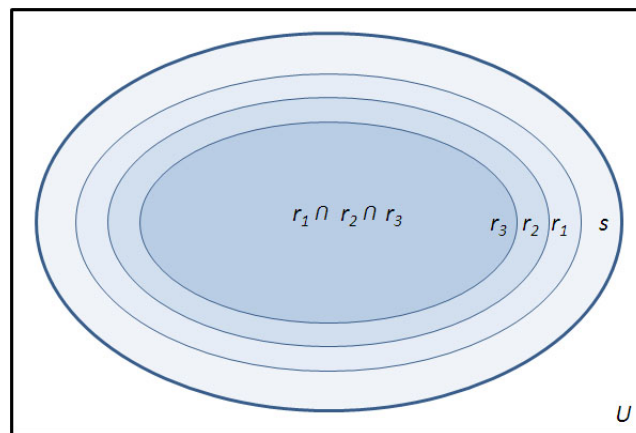
siendo w_{k_i} los ponderadores calibrados del elemento k respondente en la ola i , y y_{k_i} el valor de la variable de interés para este individuo en dicha ola.

Estimación longitudinal

Adicionalmente, para las estimaciones longitudinales será necesario el cálculo de un nuevo conjunto de ponderadores aplicables únicamente a las unidades respondientes en todas las instancias de las que se quiere medir el cambio: *respondentes simultáneos*.

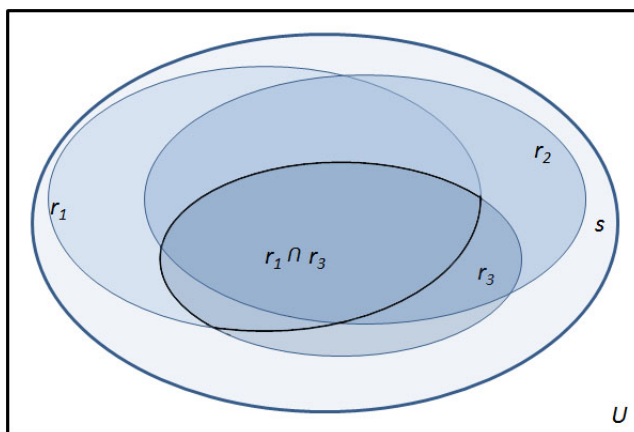
Cuando el patrón de respuesta admite únicamente no respuesta por desgaste, los respondientes de i -ésima ola también fueron respondientes en las olas anteriores $i - 1, i - 2, \dots, 1$, por lo tanto, los cambios solamente podrán ser medidos para las unidades respondientes en la ola más reciente. En la siguiente figura puede verse la representación gráfica de un estudio de panel de tres olas con patrón de respuesta por desgaste. A modo de ejemplo, si el interés radica en la estimación de cambios de una variable entre las olas 1 y 3, se calibrarán los cambios individuales para cada respondiente de la ola 3.

Figura 1: Conjunto de respondientes con patrón de respuesta: desgaste



Frente a un patrón de respuesta episódica, los cambios entre dos olas también serán medidos en los respondientes simultáneos, pero en este caso, este conjunto no necesariamente coincide con la ola más reciente de las sujetas a medición. Siguiendo el mismo ejemplo, la estimación de cambios entre la primera y la tercera ola se realizará en base al conjunto marcado en la figura (2).

Figura 2: Conjunto de respondentes con patrón de respuesta: episódica



Luego de definir el conjunto de respondentes simultáneos a las olas j y $j + h$, la estimación de las diferencias de la variable y entre dichas olas se obtiene mediante la siguiente fórmula:

$$\begin{aligned} \hat{A}w_{j,j+h} &= \sum_{r_j \cap r_{j+h}} (y_{(j+h)k} - y_{jk}) w_{(j,j+h)k}; \\ &= \sum_{r_j \cap r_{j+h}} a_{(j,j+h)k} w_{(j,j+h)k} \end{aligned} \quad (8)$$

Donde:

- $w_{(j,j+h)k}$ son los ponderadores obtenidos mediante algún método de calibración para las unidades respondentes en las olas j y $j + h$, y
- $a_{(j,j+h)k} = y_{(j+h)k} - y_{jk}$ es el cambio en la variable de interés y entre estas olas para cada elemento respondente $k \in \{r_j \cap r_{j+h}\}$.

La aplicación de la fórmula (8) en los patrones de respuesta episódicos y por desgaste difiere únicamente en la definición del conjunto de respondentes simultáneos $r_j \cap r_{j+h}$.

4. Aplicación: Las damas perdidas

Una aplicación concreta del uso de ponderadores calibrados para encuestas de panel se realizó para la “Encuesta sobre Situaciones Familiares y Desempeños Sociales en Montevideo y Área Metropolitana”, llevada a cabo por un equipo de investigadores de la Universidad de la República (Facultad de Ciencias Económicas y de Administración - Instituto de Economía y de la Facultad de Ciencias Sociales - Departamento de Economía y Programa de Población).

Este panel consiste en dos olas, la primera fue realizada entre marzo y octubre de 2001 a una muestra de 1806 mujeres y la segunda ola se realizó en el año 2008, logrando recontactar a 828 de ellas. El conjunto de respondentes simultáneos para estimaciones longitudinales está entonces compuesto por las 828 mujeres recontactadas en 2008.

El método de calibración aplicado fue el *raking* o *calibración en las marginales*. Para el cálculo de los ponderadores se utilizó el programa R Development Core Team (2008) con el paquete *survey* (T. Lumley (2009)) y la función *rake*. Los insumos necesarios para calcular estos nuevos ponderadores son el diseño que genera los expansores originales (estratificado por nivel socioeconómico) y los totales marginales sobre los cuales se calibrará, que en esta aplicación corresponden a la Edad y Nivel Educativo estimados a partir de la Encuesta Nacional de Hogares Ampliada del año 2006 (ENHA 2006).

Obtención de ponderadores

Los ponderadores a ser utilizados en la estimación longitudinal de cambios entre olas y en la estimación transversal de totales no podrán ser los mismos, ya que estas estimaciones estarán basadas en distintas unidades. Los cambios longitudinales solamente podrán ser estimados utilizando las mediciones efectivas sobre las 828 mujeres que fueron encuestadas en las dos olas del estudio, mientras que para las estimaciones transversales se cuenta con información sobre el total de mujeres entrevistadas (1229 mujeres). Por este motivo se obtendrán dos conjuntos de ponderadores, utilizando el *raking* como técnica de calibración.

Ponderadores para Estimaciones Longitudinales

Entre la instancia inicial en 2001 y la segunda ola en 2008, la imposibilidad de contacto de algunas mujeres (fallecimiento, movilidad, etc.) y la negativa de la mujer seleccionada a seguir participando genera la pérdida de 978 mujeres, provocando el desgaste del panel. Las 1806 mujeres entrevistadas en la primera instancia representan bien a la población de mujeres del 2001 con las características ya mencionadas. Si fuera posible medir los cambios entre olas en ciertas variables en cada una de estas mujeres, los mismos también serían representativos de la población objetivo 2008. La no respuesta imposibilita dicha medición para todas las unidades de la muestra, y debe compensarse. Este objetivo se puede lograr siguiendo distintas estrategias: calculando ponderadores calibrados para las 828 mujeres respondentes para que re- presenten a las 1806 iniciales (que a su vez, representan la población objetivo 2001), o calibrar los ponderadores de las 828 respondentes para que representen a la población objetivo 2008 de forma directa (las mujeres de 32 a 61 años de Gran Montevideo). Se opta por seguir la segunda estrategia.

Como ya se dijo, el procedimiento de calibración utilizado para el cálculo de estos ponderadores es el raking. Del pool de variables disponibles se deben seleccionar aquellas que serán utilizadas como información auxiliar para el cálculo de los nuevos ponderadores. Algunas de las características deseables para las variables auxiliares son su estabilidad en el tiempo y que permitan una buena caracterización de la población objeto de estudio. Para el caso particular de este estudio de panel, se descarta la utilización de variables provenientes del cuestionario de 2001 como variables auxiliares, ya que las mismas serán objeto de la medición de cambios longitudinales. Las variables finalmente elegidas para la calibración son la Edad y Nivel Educativo, al entenderse que verifican las condiciones mencionadas. Los totales poblacionales correspondientes se estiman a partir de la Encuesta Nacional de Hogares Ampliada del año 2006 (ENHA 2006).

Raking

Para el cálculo de los ponderadores se utilizó el programa R Development Core Team (2008) con el paquete *survey* (T. Lumley (2009)) y la función *rake*. Los insumos necesarios para calcular estos nuevos ponderadores son el diseño que genera los expansores originales y los totales marginales sobre los cuales se calibrará. En los cuadros siguientes se presenta dicha información.

Cuadro 1: Totales Poblacionales, Muestrales y Expansores originales por Estrato

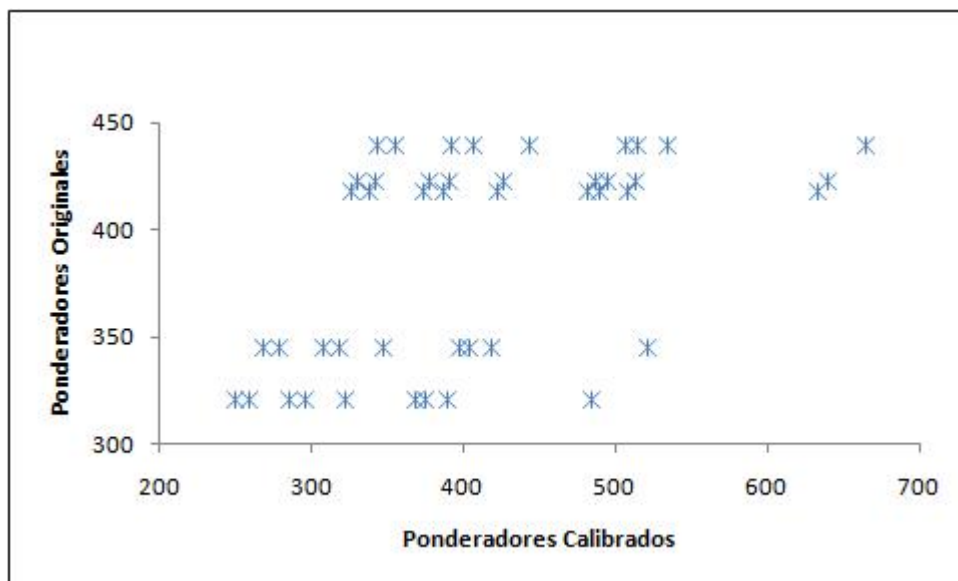
| Estrato | Tot. Poblacionales | Tot. Muestrales | Expansor Original |
|----------------|--------------------|-----------------|-------------------|
| MVD Bajo | 46924 | 111 | 533,738 |
| MVD Medio Bajo | 68596 | 199 | 344,703 |
| MVD Medio Alto | 76267 | 238 | 320,449 |
| MVD Alto | 57736 | 138 | 418,377 |
| Periferia | 62418 | 142 | 439,563 |
| Total | 311941 | 828 | |

Cuadro 2: Totales Poblacionales de las Variables Auxiliares

| Nivel Educativo | Frecuencia |
|-----------------|---------------|
| Primaria | 83112 |
| Secundaria | 147627 |
| Terciaria | 81202 |
| Total | 311941 |

| Edad | Frecuencia |
|--------------|---------------|
| 32 a 40 años | 94225 |
| 41 a 50 años | 112040 |
| 51 a 61 años | 105676 |
| Total | 311941 |

En el siguiente gráfico se presenta la modificación de los expansores originales obtenidos a partir del raking.



En el gráfico puede notarse que los ponderadores dentro de cada estrato (representados por los conjuntos de puntos alineados de forma paralela al eje de las abscisas) dejan de ser iguales. Dentro de cada estrato, hay nueve pesos diferentes, determinados por la interacción entre tramo de edad y nivel educativo, utilizados en el raking. En el siguiente cuadro se presentan los totales muestrales que dan origen a 45 nuevos ponderadores.

Cuadro 3: Totales muestrales según Edad por Estrato y Nivel Educativo

| Estrato y Nivel Educativo/Edad | 32 a 40 | 41 a 50 | 51 a 61 |
|--------------------------------|---------|---------|---------|
| Mvd Bajo - Primaria | 9 | 17 | 20 |
| Mvd Bajo - Secundaria | 22 | 18 | 14 |
| Mvd Bajo - Terciaria | 4 | 2 | 5 |
| Mvd Medio Bajo - Primaria | 8 | 13 | 22 |
| Mvd Medio Bajo - Secundaria | 33 | 46 | 38 |
| Mvd Medio Bajo - Terciaria | 13 | 11 | 15 |
| Mvd Medio Alto - Primaria | 2 | 5 | 16 |
| Mvd Medio Alto - Secundaria | 27 | 49 | 45 |
| Mvd Medio Alto - Terciaria | 26 | 45 | 23 |
| Mvd Alto - Primaria | 1 | 1 | 4 |
| Mvd Alto - Secundaria | 9 | 17 | 19 |
| Mvd Alto - Terciaria | 17 | 36 | 34 |
| Periferia - Primaria | 10 | 20 | 21 |
| Periferia - Secundaria | 22 | 26 | 20 |
| Periferia - Terciaria | 11 | 10 | 2 |

Del cuadro surge la justificación de la utilización del raking frente a la post estratificación, ya que existen celdas de la clasificación estrato, nivel educativo y edad con muy pocas observaciones. Frente a la alternativa de colapsar categorías para luego aplicar la post estratificación para la obtención de ponderadores, se opta por la aplicación directa del raking.

Estimador de cambios longitudinales

Las diferencias para la variable y entre 2001 y 2008 se estiman por $\hat{A}_{W_{2001,2008}}$:

$$\begin{aligned}\hat{A}_{W_{2001,2008}} &= \sum_{k=1}^{828} (y_{2008k} - y_{2001k}) w_{Ck}; \\ &= \sum_{k=1}^{828} a_{(2001,2008)k} w_{Ck}\end{aligned}\tag{9}$$

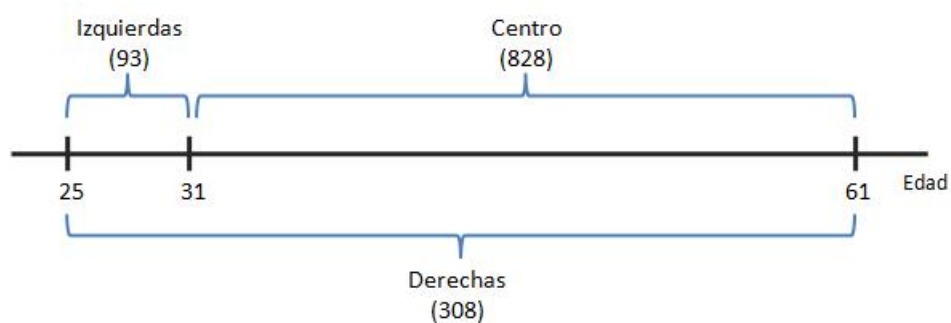
donde $a_{(2001,2008)k} = (y_{2008k} - y_{2001k})$ representa la diferencia de los valores de la variable de interés y medidos en 2008 y 2001 para las 828 mujeres de la muestra original entrevistadas en 2001 y en 2008, y w_{Ck} el ponderador que surge de la aplicación del método de calibración raking para este grupo de mujeres.

Ponderadores para Estimaciones Transversales

Las estimaciones transversales para el 2008 se calcularán en base a las respuestas dadas por las 1229 mujeres respondentes: 828 provenientes del panel original, 93 mujeres de edades entre 25 y 31 años (las “rejuvenecedoras” del panel) y 308 de 25 a 61 años de edad (las “ampliadoras” del panel). El objetivo de lograr que las 1229 mujeres representen bien a las mujeres de Gran Montevideo con edades entre 25 y 61 años en 2008 requiere el cálculo por separado de ponderadores en cada uno de estos tres grupos, ya que las mujeres que integran cada grupo provienen de muestras distintas. La libertad de elección de ponderadores iniciales $d_{\alpha k}$ mencionada en la sección (5.5.1) no contempla el caso en que las unidades provengan de muestras distintas. Es por esto que en cada uno de estos grupos se calcularán los nuevos ponderadores con el método raking para luego combinar los resultados. Las variables auxiliares a ser utilizadas en el procedimiento de calibración seleccionado para cada uno de estos grupos serán edad y nivel educativo, de igual manera que en la sección anterior.

En la siguiente figura se representa el rango de edad de las 1229 mujeres relevadas en 2008 según la muestra de la que provienen: centro, derechas e izquierdas.

Figura 3: Representación gráfica de las edades de las 1229 mujeres relevadas en 2008 según muestra de origen



Raking Centro

Los nuevos ponderadores para estas 828 mujeres centro son los mismos que fueron calculados en la parte anterior.

Raking Izquierdas

En este caso la única variable auxiliar para calibrar es nivel educativo dado que todas estas mujeres tienen entre 25 a 31 años, que corresponde a una única categoría de la variable edad. Los ponderadores originales y los totales poblacionales se presentan en el cuadro 6.7 y el cuadro 6.8.

Cuadro 5: Totales Poblacionales, Muestrales y Expansores Originales por Estrato

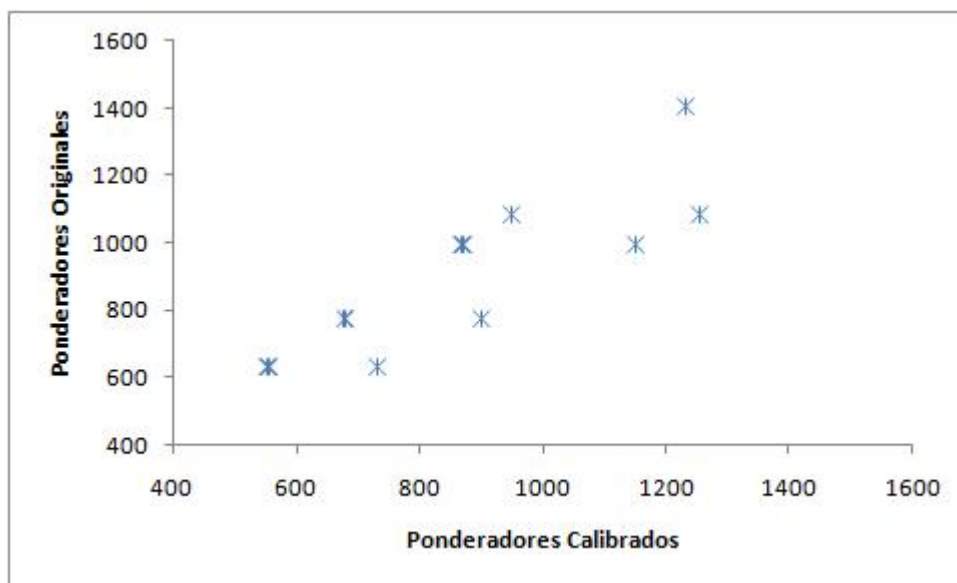
| Estrato | Tot.Poblacionales | Tot. Muestrales | Expansor Original |
|----------------|-------------------|-----------------|-------------------|
| MVD Bajo | 13931 | 14 | 995,071 |
| MVD Medio Bajo | 18622 | 24 | 775,917 |
| MVD Medio Alto | 21677 | 20 | 1083,850 |
| MVD Alto | 14050 | 10 | 1405,000 |
| Periferia | 15804 | 25 | 632,160 |
| Total | 84084 | 93 | |

Cuadro 6: Totales Poblacionales de la Variable Auxiliar

| Nivel Educativo | Frecuencia |
|-----------------|--------------|
| Primaria | 12804 |
| Secundaria | 42852 |
| Terciaria | 28428 |
| Total | 84084 |

En este caso el método raking genera los mismos resultados que el método de post estratificación, ya que calibrar en las marginales de una única variable auxiliar es equivalente a calibrar en las celdas.

En el siguiente gráfico se presenta la modificación de los expansores originales obtenidos a partir del raking.



En este caso hay tres ponderadores por estrato, las tres categorías de nivel educativo, excepto para los estratos Montevideo Medio Alto y Alto; esto último se debe a que en esta muestra no hay mujeres con solamente primaria completa en estos estratos, como se muestra en el siguiente cuadro.

Cuadro 7: Totales Muestrales por Nivel Educativo según Estrato

| Estrato/Nivel Educativo | Primaria | Secundaria | Terciaria | Total |
|-------------------------|-----------|------------|-----------|-----------|
| MVD Bajo | 4 | 9 | 1 | 14 |
| MVD Medio Bajo | 4 | 14 | 6 | 24 |
| MVD Medio Alto | 0 | 8 | 12 | 20 |
| MVD Alto | 0 | 2 | 8 | 10 |
| Periferia | 12 | 9 | 4 | 25 |
| Total | 20 | 42 | 31 | 93 |

Raking Derechas

Los insumos necesarios para calcular estos nuevos ponderadores son el diseño que genera los expansores originales y los totales marginales sobre los cuales se calibrará, de igual manera a las partes anteriores, pero con la excepción de que la variable auxiliar edad tiene ahora cuatro categorías.

Cuadro 8: Totales Poblacionales, Muestrales y Expansores Originales por Estrato

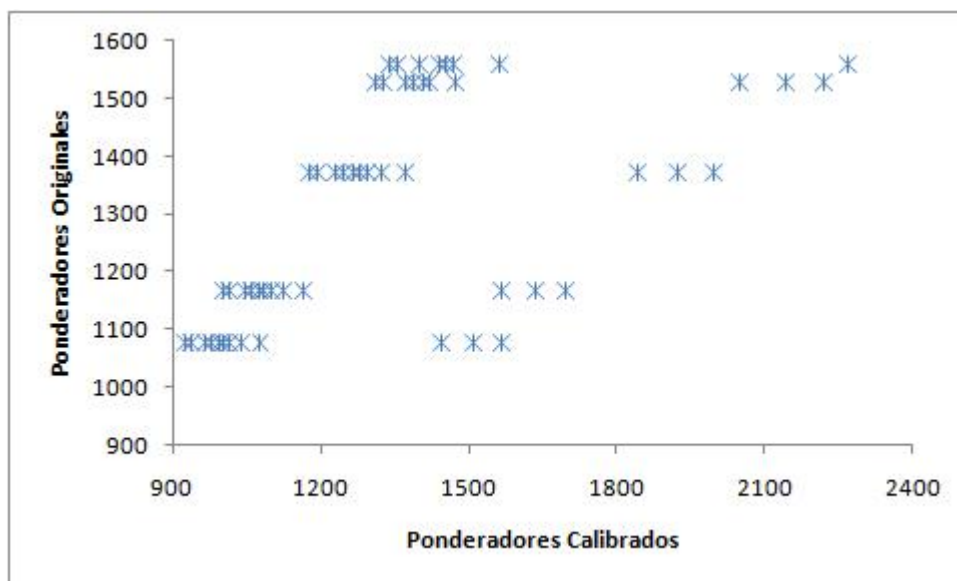
| Estrato | Tot. Poblacionales | Tot. Muestrales | Expansor Original |
|----------------|--------------------|-----------------|-------------------|
| MVD Bajo | 60855 | 39 | 1560,385 |
| MVD Medio Bajo | 87218 | 81 | 1076,765 |
| MVD Medio Alto | 97944 | 84 | 1166,000 |
| MVD Alto | 71786 | 47 | 1527,362 |
| Periferia | 78222 | 57 | 1372,316 |
| Total | 396025 | 308 | |

Cuadro 9: Totales Poblacionales de las Variables Auxiliares

| Nivel Educativo | Frecuencia |
|-----------------|------------|
| Primaria | 95916 |
| Secundaria | 190479 |
| Terciaria | 109630 |
| Total | 396025 |

| Edad | Frecuencia |
|--------------|------------|
| 25 a 31 años | 84084 |
| 32 a 40 años | 94225 |
| 41 a 50 años | 112040 |
| 51 a 61 años | 105676 |
| Total | 396025 |

En el siguiente gráfico se presenta la modificación de los expansores originales obtenidos a partir del raking.



La interacción entre las variables nivel educativo (3 categorías) y edad (4 categorías) para cada uno de los cinco estratos debería generar 60 pesos diferentes. En los hechos, se pueden distinguir 54 ponderadores diferentes debido a que no todas las celdas correspondientes a la interacción de las variables auxiliares contienen observaciones, como se puede ver en el cuadro 6.12.

Cuadro 10: Totales Muestrales según Edad por Estrato y Nivel Educativo

| Estrato y Nivel Educativo/Edad | 25 a 31 | 32 a 40 | 41 a 50 | 51 a 61 |
|--------------------------------|---------|---------|---------|---------|
| Mvd Bajo - Primaria | 4 | 6 | 4 | 3 |
| Mvd Bajo - Secundaria | 4 | 7 | 10 | 0 |
| Mvd Bajo - Terciaria | 0 | 1 | 0 | 0 |
| Mvd Medio Bajo - Primaria | 5 | 5 | 4 | 6 |
| Mvd Medio Bajo - Secundaria | 15 | 14 | 9 | 7 |
| Mvd Medio Bajo - Terciaria | 4 | 5 | 4 | 3 |
| Mvd Medio Alto - Primaria | 1 | 3 | 1 | 4 |
| Mvd Medio Alto - Secundaria | 8 | 7 | 21 | 7 |
| Mvd Medio Alto - Terciaria | 7 | 11 | 8 | 6 |
| Mvd Alto - Primaria | 0 | 2 | 0 | 2 |
| Mvd Alto - Secundaria | 4 | 5 | 6 | 2 |
| Mvd Alto - Terciaria | 6 | 1 | 8 | 11 |
| Periferia - Primaria | 2 | 4 | 10 | 2 |
| Periferia - Secundaria | 9 | 10 | 9 | 5 |
| Periferia - Terciaria | 1 | 2 | 2 | 1 |

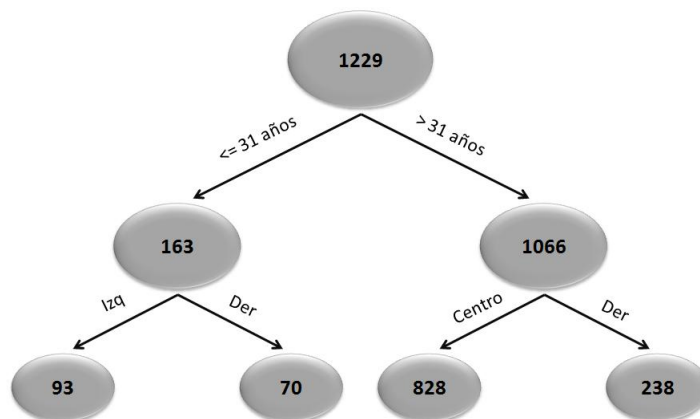
Ponderadores Combinados

Para cada una de las 1229 mujeres se calculó su respectivo ponderador, en relación a la submuestra a la que pertenece. Las 828 mujeres del centro representan al total de mujeres de Gran Montevideo con edades entre 32 y 61 años (311941 mujeres); las 93 mujeres de la izquierda representan a las 84084 mujeres dentro de la franja etaria de 25 a 31 años; y las 308 de la derecha a aquellas de edades entre 25 y 61 años, cuyo total asciende a 396025 mujeres. Si para realizar cálculos de totales

se utilizaran los ponderadores obtenidos en cada uno de los grupos de manera directa, dichos totales se estarían sobreestimando. De hecho, se estaría estimando el total correspondiente a una población compuesta por el doble de las mujeres existentes en Gran Montevideo de edades entre 25 y 61 años. Es por este motivo que los ponderadores calculados en las partes anteriores deben utilizarse de forma combinada, para lograr la estimación sobre el total efectivo de mujeres de dichas características: 396025 mujeres.

Las 1229 mujeres entrevistadas en 2008 se clasifican en cuatro subpoblaciones en relación a su edad (mayor o menor a 31 años) y muestra de procedencia (centro, izquierdas, derechas), como muestra la Figura 6.2.

Figura 4: Clasificación de las mujeres entrevistadas en 2008



El primer nivel de división (corte por edad) permite identificar dos grandes grupos: 163 mujeres de edades entre 25 y 31 años que deberán representar a 84084 mujeres, y 1066 mayores a 31 años, que deberán representar a 311941 mujeres existentes en dicha franja etaria. El segundo corte queda determinado por la muestra de la que provienen cada una de las mujeres: las 163 mujeres menores a 32 años pueden pertenecer a la muestra de izquierdas o derechas, y las 1066 mayores de 31 pueden ser parte del panel original o de la muestra de derechas.

La combinación de resultados sigue la misma lógica que la figura: en primer lugar, se particiona la muestra total de acuerdo a la edad (mayores o menores de 31 años) formando dos grandes grupos. Dentro de cada grupo los nuevos ponderadores combinados tendrán en cuenta la muestra de la que provienen a través

de la relación que existe entre la cantidad de mujeres en esta última partición y la cantidad de mujeres en el grupo etario que corresponda. Por ejemplo, los pesos obtenidos a través del raking para las 828 mujeres provenientes del panel original se multiplicarán por 828/1066, la proporción de mujeres del panel que corresponde a la franja etaria establecida. Considerando que las varianzas de los estimadores son del orden de $1/n_i$ siendo n_i el tamaño de cada una de las cuatro particiones, se combinan los ponderadores resultantes del raking de manera que las mujeres provenientes de subgrupos más grandes adquieran mayor importancia en el análisis.

Los ponderadores combinados a ser utilizados en las estimaciones transversales resultan ser:

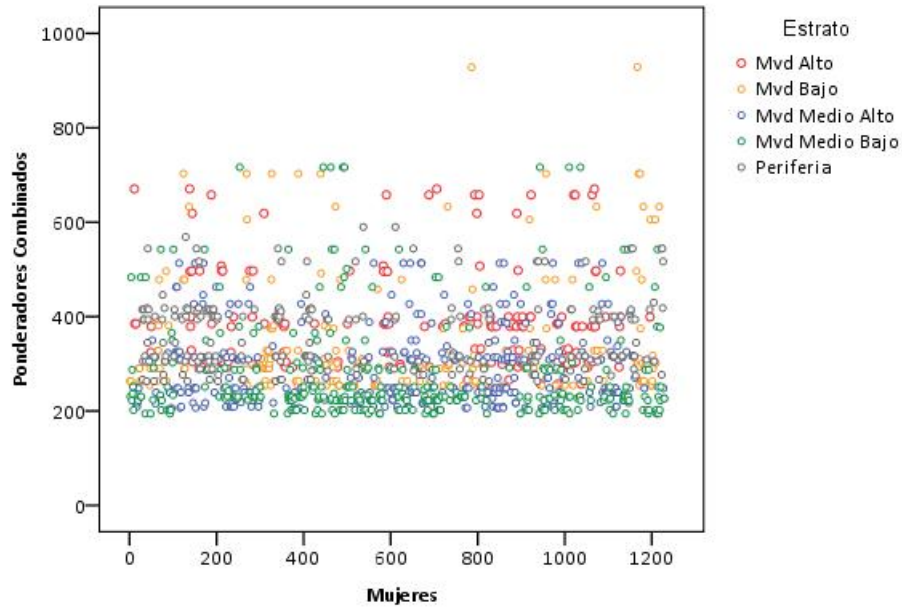
$$w_{comb_k} = \begin{cases} \frac{93}{163} w_{Ik} & k \in \text{muestra de izquierdas} \\ \frac{70}{163} w_{Dk} & k \in \text{muestra de derechas con edades entre 25 y 31 años} \\ \frac{238}{1066} w_{Dk} & k \in \text{muestra de derechas con edades entre 32 y 61 años} \\ \frac{828}{1066} w_{Ck} & k \in \text{panel original} \end{cases}$$

Estos ponderadores combinados estiman sin error la cantidad total de mujeres pertenecientes a la población objetivo estimada por la ENHA 2006, que se considera equivalente a la población 2008.

$$\begin{aligned} \hat{N} &= \sum_{k=1}^{1229} w_{comb_k} = \\ &= \sum_{k=1}^{93} \frac{93}{163} w_{Ik} + \sum_{k=1}^{70} \frac{70}{163} w_{Dk} + \sum_{k=1}^{238} \frac{238}{1066} w_{Dk} + \sum_{k=1}^{828} \frac{828}{1066} w_{Ck} \\ &= 47974,3 + 36109,7 + 69645,4 + 242295,6 \\ &= 396025 = N \end{aligned}$$

La siguiente figura muestra los ponderadores combinados para las 1229 mujeres según el estrato al que pertenecen.

Figura 5: Ponderadores Combinados por Estrato



Estimación de totales transversales

La estimación transversal de los totales de interés debe realizarse utilizando los ponderadores combinados de acuerdo a la siguiente fórmula:

$$\begin{aligned}
 \hat{t}_{y_{w_{comb}}} &= \sum_{k=1}^{1229} w_{comb_k} y_k \\
 &= \sum_{k=1}^{93} \frac{93}{163} w_{I_k} y_k + \sum_{k=1}^{70} \frac{70}{163} w_{D_k} y_k + \sum_{k=1}^{238} \frac{238}{1066} w_{D_k} y_k + \sum_{k=1}^{828} \frac{828}{1066} w_{C_k} y_k \quad (10)
 \end{aligned}$$

5. Conclusiones

La medición de cambios entre distintas instancias en el tiempo es el principal objetivo de las encuestas de panel. Se presentaron estimadores del cambio de las variables de interés que, bajo el supuesto de respuesta perfecta, son insesgados para estimar el cambio total en la población objetivo. Para su desarrollo, se partió de supuestos muy restrictivos rara vez presentes en la práctica: la existencia de respuesta perfecta y población fija en el tiempo, reflejados en ponderadores constantes en el tiempo para cada elemento e iguales al inverso de su probabilidad de inclusión en la muestra.

La no respuesta es un fenómeno presente en la mayoría de las encuestas por muestreo y es necesario su tratamiento para evitar sesgos en las estimaciones. En las encuestas de panel la inclusión del factor tiempo provoca un agravamiento del problema de no respuesta, reflejado en reducciones considerables en el tamaño de muestra período a período. Si la no respuesta se presentara de manera completamente aleatoria, el único inconveniente al que se enfrenta el investigador resulta en la reducción del tamaño de muestra y su respectivo aumento en la varianza de las estimaciones, pero las unidades que no contestan suelen diferir de aquellas que sí lo hacen, y el sesgo introducido en las estimaciones por esta causa constituye el obstáculo más importante por corregir.

Aún cuando se invierta esfuerzo en intentar alcanzar la mayor tasa de respuesta posible, la no respuesta existe y es deber de los investigadores realizar algún tratamiento en la etapa de análisis de datos para controlar el sesgo introducido a las estimaciones por su causa. Como método de tratamiento se descartan la sustitución y submuestreo de no respondentes y los métodos basados en la quasi-randomization por entender que no son las opciones preferibles en encuestas de panel. La imputación como forma de tratar la no respuesta de unidades en la ola también se descarta por generar una fabricación masiva de datos, que pueden distorsionar las asociaciones entre las variables que representan el principal objetivo del panel. Es preferible entonces utilizar una estrategia global como es la calibración cuando es la unidad la que no provee de respuesta. De todas maneras la imputación es el tratamiento elegido como manera de compensar la no respuesta en los ítems, previo a la calibración por la no respuesta de unidades.

La idea central de los estimadores calibrados es sencilla, en base a información auxiliar se modifican los ponderadores originales de la muestra minimizando alguna función de distancia entre dichos ponderadores y los ponderadores finales

(o calibrados) y de manera que estos últimos estimen sin error totales poblacionales conocidos de las variables auxiliares que asisten al procedimiento de calibración. El estimador calibrado corresponde en realidad a una familia entera de estimadores que dependen de formulaciones diferentes del vector auxiliar y de la función de distancia.

La efectividad del estimador de calibración para controlar el sesgo ocasionado por la no respuesta dependerá de propiedades del vector auxiliar. Se obtiene una expresión del sesgo aproximado que demuestra que éste será menor cuanto más estrecha sea la relación entre la información auxiliar y la probabilidad de respuesta o la variable de interés. También se propone una formulación para la estimación de la varianza del estimador calibrado que podrá ser utilizada en la construcción de intervalos de confianza. Si el sesgo es modesto el intervalo de confianza será válido y la probabilidad de cobertura será cercana al nivel de confianza especificado.

El estimador de cambios entre olas propuesto para el caso de respuesta perfecta se extiende utilizando los ponderadores calibrados calculados para las unidades respondientes. De esta manera, los cambios entre las olas de interés medido en los respondientes simultáneos ponderado por los pesos calibrados representarán los cambios de la población objetivo.

Referencias

- [1] Bailar, B.A. (1975). “The Effects of Rotation Group Bias on Estimates from Panel Surveys”. *Journal of the American Statistical Association*, 70(349): 23–30.
- [2] Chhikara, R.S.; and Deng, L.Y. (1992). “Estimation Using Multiyear Rotation Design Sampling in Agricultural Surveys”. *Journal of the American Statistical Association*, 87(420): pp 924–932.
- [3] Copeland, K.R. (2004). “Nonresponse Adjustment in the Current Employment Statistics Survey”. U.S. Bureau of Labour Statistics, Washington DC.
- [4] De Leeuw, E.D.; Hox, J.; and Huisman, M. (2003). “Prevention and Treatment of Item Nonresponse”. *Journal of Official Statistics*, 19 (2): 153–176.
- [5] De Leeuw, E.D. (2006). “Introduction to Survey Nonresponse”. Summer Institute in Survey Research Techniques. *Survey Methodology* 988.223.
- [6] Dennis, J.M. and Li, R. (2003). “Effects of Panel Attrition on Survey Estimates”. For Presentation at the 2003 Annual Meeting of the American Association for Public Opinion Research in Nashville, Tennessee.
- [7] Deville, J.C. and Särndal, C.E. (1992). “Calibration Estimators in Survey Sampling”. *Journal of the American Statistical Association*, 87(418).
- [8] Deville, J.C.; Särndal, C.E.; and Sautory, O. (1993). “Generalized Raking Procedures in Survey Sampling”. *Journal of the American Statistical Association*, 88(423).
- [9] Deville, J.C. and Särndal, C.E. (1994). “Variance estimation for the regression imputed Horvitz – Thompson estimator”. *Journal of Official Statistics*. 10: 381–394.
- [10] Ernst, L.R (1986). “Weighting Issues for Longitudinal Household and Family Estimates”. Bureau of Census, SRD Research Report Number: CENSUS/SRD/RR-86/23.
- [11] Fuller, W.A. and Breidt, F.J. (1999). “Estimation for Supplemented Panels”. *The Indian Journal of Statistics Special Issue on Sample Surveys*, 61: 58–70.

- [12] Gross Sobol, M. (1959). "Panel Mortality and Panel Bias". *Journal of the American Statistical Association*, 54(285): 52 - 68.
- [13] Groves, R.M. and Couper M.P (1998). *Nonresponse in Household Interview Surveys*. New York: John Wiley & Sons.
- [14] Kalton, G. and Brick, J.M . "Weighting Schemes for Household Panel Surveys: The Survey of Income and Program Participation". U.S. Department of Commerce Bureau of Census.
- [15] Kish, Leslie (1995). *Diseño Estadístico para la Investigación*. España: Siglo XXI de España Editores SA.
- [16] Little, R.J.A and David, H.M. (1983). "Weighting Adjustment for Nonresponse in Panel Surveys". Working Paper: U.S. Bureau of Census, Washington DC.
- [17] Little, R.J.A and Rubin, D.B. (1987). *Statistical Analysis with Missing Data*. New York: Wiley & Sons.
- [18] Lumley, T. (2004) Analysis of complex survey samples. *Journal of Statistical Software* 9(1): 1-19
- [19] Lumley, T. (2009) "survey: analysis of complex survey samples". R package version 3.11-2.
- [20] Lynn, P.; Buck, N.; Burton, J.,; Jäckle, A.; and Laurie, H. (2005). "A Review of Methodological Research Pertinent to Longitudinal Survey Design and Data Collection". Working Papers of the Institute for Social and Economic Research, paper 2005 – 29. Colchester: University of Essex.
- [21] Rubin, Donald B. (1987). *Multiple Imputation for Nonresponse in Surveys*. New York: Wiley & Sons.
- [22] R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- [23] Särndal, C.E.; Swensson, B.; and Wretman, J. (1992). *Model Assisted Survey Sampling*. New York: Springer-Verlag.

- [24] Särndal, C.E. and Lundström, S. (1999). “Calibration as a Standard Method for Treatment of Nonresponse”. *Journal of Official Statistics*. 15(2): 305–327.
- [25] Särndal, C.E. and Lundström, S. (2005). *Estimation in Surveys with Nonresponse*. New York: John Wiley and Sons, Ltd.
- [26] Särndal, C.E. and Lundström, S. (2008). “Assessing auxiliary vectors for control of nonresponse bias in the calibration estimator”. *Journal of Official Statistics*, 24 (2):167–191.
- [27] Sastry, N; Ghosh-Dastidar, B; Adams, J; and Pebley, A (2000). “The Design of a Multilevel Longitudinal Survey of Children, Families, and Communities: The Los Angeles Family and Neighborhood Survey”. Labour and Population Program, Working Paper Series 00-18.
- [28] Sharot, T. (1991). “Attrition and Rotation in Panel Surveys”. *The Statistician*, 40(3): 325–331.
- [29] Simard, M. (2002). “Generation and Gender Survey Sample Design Guidelines”. Statistics Canada.
- [30] Taylor, B.; Brook, L.; and Lynn, P. (1997). “Incentives Information and Number of Contacts: Testing the Effects of these Factors of Response to a Panel Survey”. *Survey Methods Centre Newsletter*, 17(3): 712.
- [31] Wadsworth, R.N. (1952). “The experience of a User of a Consumer Panel”. *Applied Statistics*, 1(3): 169–178.
- [32] Williams, W.H. and Mallows, C.L. (1970). “Systematic Biases in Panel Surveys Due to Differential Nonresponse”. *Journal of the American Statistical Association*, 65(331).
- [33] Woodruff, R.S. (1963). “The Use of Rotating Samples on the Census Bureau’s Monthly Surveys”. *Journal of the American Statistical Association*, 58(302): 454–467.