

# Metodología de ponderadores de la ELPS ola 2

Juan Pablo Ferreira

## Introducción

La ELPS es una encuesta de panel, es decir, un estudio basado en mediciones repetidas efectuadas sobre las mismas unidades de muestreo, en este caso, personas y hogares. La medición a través del tiempo sobre las mismas unidades permite realizar un seguimiento del universo de elegibles, logrando captar su dinámica en el tiempo. El objetivo principal, como en toda encuesta de panel, es poder estimar cambios netos y brutos en la población. Si bien los cambios netos pueden ser llevados a cabo por medio de encuestas cross section (muestras independientes en cada momento de tiempo), estas tienden a ser menos eficientes (mayores errores estándar  $<SE>$ ) respecto a las encuestas de panel.

A su vez, los cambios brutos (construcción de matrices de transición) únicamente pueden ser llevadas a cabo por medio de encuestas de panel. Por otra parte, las estimaciones en cada ola (estimaciones transversales) pueden ser llevadas a cabo con la construcción de un sistema de ponderadores específicos que refleje dicha población en un momento del tiempo dado (e.g. tiempo de relevamiento de la ola 2) aunque no sea el objetivo principal de las encuestas de panel.

Para la ola 2 de la ELPS se construyeron ponderadores longitudinales (que permiten la construcción de estimaciones del cambio bruto) y ponderadores transversales para las estimaciones de la población en el momento de la realización de la ola 2. Al ser la ELPS una encuesta a personas que releva información sobre el hogar (en donde una persona elegible fue seleccionada por hogar) fue necesario calcular los ponderadores mencionados anteriormente tanto para personas como hogares.

Los ponderadores longitudinales reflejan la cohorte seleccionada en el año base (ola 1) en el momento de la ola 2 y por ende reflejan los cambios que sufrió dicha cohorte a lo largo del tiempo, es por esto que son dichos ponderadores los que deben ser utilizados para el cálculo de las matrices de transición. Lo anterior implica que los ponderadores longitudinales “expanden” a la estructura de la población de personas por edad, sexo y región, así como a la cantidad de hogares en el momento del inicio del panel de la ELPS. En tanto, los ponderadores transversales (si bien no es el objetivo principal de la ELPS) son construidos para realizar estimaciones del nivel de la población de elegibles en la ola 2. Al ser un panel puro en donde no existe un sorteo de nacimientos, dichos ponderadores transversales representan a la población de 16 años o más al momento de la realización de la ola 2.

Independientemente del tipo de ponderador, la construcción general de los mismos se realizó en dos etapas. Se partieron de los ponderadores de las personas respondientes en la ola 1 de la ELPS, luego los ponderadores de los elegibles que también respondieron en la ola 2 fueron ajustados teniendo en cuenta la no respuesta (atracción) típicas de las encuestas de panel. En una última etapa los ponderadores ajustados por no respuesta y finalmente ajustados (calibrados) a conteos demográficos poblacionales conocidos (sexo, tramo edad y región) tanto para los ponderadores longitudinales como transversales, en donde, la diferencia radica en la información auxiliar utilizada (momento del tiempo) en cada uno de ellos.

## 1. Diseño muestral y cálculo de los ponderadores de la ola 1

El diseño muestral de la ELPS para la primera ola fue aleatorio, estratificado y en varias etapas de selección utilizando para la selección de la muestra el marco de viviendas del Censo 2011 realizado por el Instituto Nacional de Estadística (INE). En base a la estrategia de selección (diseño muestral) se calcularon los ponderadores originales tanto de las personas como de los hogares en base a las probabilidades de selección.

Luego, los ponderadores fueron ajustados por no respuesta creando clases (post-estratos) en base a los departamento y por urbanicidad (ver metodología ola 1). Finalmente, los ponderadores ajustados por no respuesta fueron calibrados a conteos demográficos provenientes de las proyecciones de población del año 2013 y de la ECH. De aquí en más a los ponderadores finales de la ola 1 se los denotará como  $d_i^{(1)}$  para el caso de los hogares y como  $d_{ip}^{(1)}$  para el caso del informante.

## 2. Ajuste por no respuesta (atracción)

En las encuestas de panel, los ponderadores deben ser ajustados para tratar el problema de atracción, esto es, cuando una observación (informante u hogar) que fue seleccionado en la primera ola pero no pudo ser localizado o el mismo se niega a responder en la segunda ola.

El ajuste utilizado para el tratamiento de la no respuesta fue basado en un modelo de grupos de respuesta homogéneos, el cual consiste en clasificar a las unidades elegibles de la muestra de la ola 2 en  $C$  grupos para los cuales se supone que la probabilidad de responder son similares (e.i. presentan las mismas características). Los grupos fueron construidos en base a la clasificación cruzada del estrato del diseño muestral y el tamaño del hogar. Esta última variable se incluyó en el ajuste dado que los hogares menos numerosos fueron menos propensos a responder o ser localizados en la ola 2.

El ponderador del hogar  $i$  en la ola 2 perteneciente a una clase cualquiera  $c$  ajustado por no respuesta queda definido como

$$d_{i_{nr}}^{(2)} = \left( \sum_{s_{c,E}} d_i^{(1)} \right) \times \left( \sum_{s_{c,ER}} d_i^{(1)} \right)^{-1} \times d_i^{(1)}$$

Finalmente, el ponderador del informante ajustado por no respuesta en la ola 2 se define como

$$d_{ip_{nr}}^{(2)} = d_{i_{nr}}^{(2)} \times E_i$$

donde  $E_i$  es la cantidad de miembros elegibles del hogar para participar en la ELPS. La determinación de la cantidad de miembros elegibles varía según si el ponderador ajustado por no respuesta es utilizado para la calibración de los ponderadores longitudinales o transversales. En el caso de los ponderadores longitudinales se tiene en cuenta el número de personas elegibles en el hogar al momento de la realización de la ola 1, es decir, no se admiten miembros nuevos. Para el caso de los ponderadores transversales,  $E_i$  puede contener miembros elegibles que pertenecían a otros hogares en la ola 1.

### 3. Calibración de los ponderadores

El objetivo de la calibración es obtener un nuevo sistema de ponderadores  $w_i$  que se encuentren “próximos” de los ponderadores ajustados por no respuesta. Pero, cuando los ponderadores  $w_i$  son usados para estimar los conteos de las variables control reproducen exactamente dichos conteos. La idea de mantener cerca los ponderadores es para pedir prestada cualquiera propiedad buena de estimación que los ponderadores tengan.

Los principales beneficios de la calibración son:

1. Reducción de los errores estándar (SE) de las estimaciones. Esto ocurre si las variables de control de alguna forma explican la variabilidad en las variables de interés de la encuesta, i.e. se encuentran correlacionadas.
2. Posible reducción del sesgo por problemas de cobertura.
3. Reducción del Sesgo ocasionado por la no respuesta. Esto ocurre si las variables explican de alguna forma la probabilidad de responder de la persona.
4. Comparabilidad. Calibrar produce las mismas estimaciones para las variables de control cruzando con otras fuentes, e.g. proyecciones de población del INE y la Encuesta Continua de hogares (ECH).

Formalmente, el problema es encontrar un nuevo juego de ponderadores  $w_i = g_i \times d_i^{(2)}$  que minimicen una medida de distancia,

$$\text{cambio ponderadores} = L(w_i, d_i^{(2)})$$

Y que cumplan con la ecuación de calibración

$$\sum_{i \in s} w_i \times \mathbf{x}_i = \sum_{i \in U} \mathbf{x}_i$$

donde  $\mathbf{x}_i$  es el set de variables de control.

Los conteos utilizados para el ajuste por calibración provienen de las proyecciones de población y de estimaciones provenientes de la ECH (para el caso de los conteos de hogares por departamento).

#### **4.1. Calibración de los Ponderadores longitudinales.**

Debido a que el objetivo de los ponderadores longitudinales es reflejar la población de la primera ola de la ELPS (año 2013) los conteos utilizados para la calibración de los ponderadores de hogares y personas son los mismos que fueron utilizados en la ola 1, es decir, para los ponderadores longitudinales de las personas los conteos poblacionales son el total de personas por departamento, sexo y ocho tramos de edad: 14 a 19 años, 20 a 29 años, 30 a 49 años, 50 a 59 años, 60 a 64 años, 65 a 69 años, 70 a 74 años y 75 y más años.

En tanto, para calibración de los ponderadores longitudinales de los hogares, los conteos poblacionales son el número de hogares por departamento provenientes de la ECH del año 2013 y el total de personas desagregado por departamento y sexo provenientes de las proyecciones de población para el mismo año.

#### **4.2. Calibración de los Ponderadores transversales**

Debido a que el objetivo de los ponderadores transversales es reflejar la población al momento de la realización de la ola 2 (año 2016) los conteos utilizados para la calibración de los ponderadores de hogares y personas son los mismos que fueron utilizados para los ponderadores longitudinales pero utilizando información de las proyecciones de población como de la ECH para el año 2016. Cabe aclarar que los tramos de edad utilizados son iguales a diferencia del primero que pasa a ser de 16 a 19 años. Lo anterior se debe a que la ELPS es un panel puro (no hay sorteo de nacimientos) por lo cual no puede reflejar la población de 14 años o más al año 2016.

## 4.2. Método de calibración

Una vez definido el set de variables auxiliares para la calibración y definido los conteos poblacionales, se optó por la distancia raking ratio

$$L(w_i, d_i^{(2)}) = \sum_{i \in s} [w_i \log(w_i/d_i^{(2)}) - w_i + d_i^{(2)}]$$

Adicionalmente se le agregó una restricción a los cambios relativos de los ponderadores  $(w_i/d_i^{(2)})$  permitiendo únicamente cambios que se encuentren en el intervalo 0.5 y 1.8.