Métodos de ponderación en encuestas complejas: recomendaciones para una buena práctica.

Juan Pablo Ferreira[1], Guillermo Zoppolo[2]

RESUMEN

Los ponderadores son un componente clave para la producción de estimaciones de distintas cantidades poblacionales cuando se utilizan datos provenientes de una encuesta por muestreo. Por ejemplo, un estimador para el total de una variable de interés y adopta la forma $\hat{t}_y = \sum_s w_k y_k$, donde w_k es el ponderador del elemento k observado en la muestra s.

En un primer paso, los ponderadores responden a que los datos de la muestra, deben representar al total de la población de interés U y dependen de la estrategia de selección de la muestra, es decir, el diseño de muestreo. En este caso, los ponderadores se definen como el inverso de la probabilidad de selección en la muestra π_k, o sea, $w_k = 1/\pi_k$.

Posteriormente, los ponderadores son sometidos a sucesivos ajustes hasta obtener los ponderadores finales. Dichos ajustes responden a motivos de diversa índole. Imperfecciones en el marco muestral (fundamentalmente sobre cobertura), ajustes debidos a la existencia de no respuesta y ajustes para lograr consistencia con totales poblacionales conocidos.

En este documento se presentan distintos métodos para llevar a cabo dichos ajustes. Se hace énfasis en los aspectos prácticos para encuestas de hogares y personas y a la implementación de los procedimientos en el software de distribución libre R.

Palabras clave: *Diseños muestrales complejos, Estimadores calibrados, Inferencia basada en el diseño, No respuesta, Ponderadores muestrales.*

[1]IESTA, DMMCC - FCEA, INE, UCU
[2]IESTA, DMMCC - FCEA

1. Introducción

Las encuestas por muestreo están sujetas a dos tipos de error. Por una parte, siempre existe un error de muestreo, inherente al hecho que solamente un subconjunto de la población de interés, y no su totalidad, es observado. Por otra, siempre hay presentes errores no muestrales debidos a distintas causas, entre las que se destacan: a) deficiencias en el marco muestral; la existencia de elementos seleccionados originalmente para los cuales no se obtiene respuesta; y c) problemas en la recolección de los datos.

Dichos errores llevan a que los datos finalmente obtenidos, sin ser debidamente tratados, reflejen solamente las características particulares de la muestra efectivamente observada y no a la población objetivo. Los estimadores basados simplemente en las cantidades observadas tendrán sesgos considerables.

Es usual que el diseño muestral, el mecanismo aleatorio por medio del cual se selecciona un subconjunto de la población para ser observado, determine distintas probabilidades de inclusión para los individuos. éstas distintas probabilidades se justifican por razones de eficiencia, entendida como menor varianza de los estimadores y/o menor costo, para un mismo tamaño de muestra y/o tasas de muestreo distintas para ciertos dominios de estimación de la encuesta. De esta manera, aún en el caso que no existiera otro tipo de errores, las distintas probabilidades de selección llevan a que los datos observados deban ser corregidos de alguna manera a la hora de su procesamiento, es decir, deben ser ponderados por algún factor, de manera que las estimaciones que se obtengan para cantidades poblacionales desconocidas sean aproximadamente insesgadas.

Por su parte, los errores no muestrales, terminan modificando, involuntariamente, las distintas probabilidades de inclusión planificadas en el diseño de la muestra. Dichas modificaciones suelen ser no menores y deben tenerse en cuenta en la construcción de los ponderadores.

Adicionalmente, existen otras causas que llevan a ajustar los ponderadores. Es común que algunas cantidades poblacionales sean conocidas o se encuentren estimadas con una muy alta precisión y por lo tanto se desee que al estimar dichas cantidades con los datos ponderados las estimaciones coincidan con los valores conocidos. De esta manera se logra coherencia y comparabilidad con los datos de otras fuentes de información, generalmente consideradas más precisas, y, adicionalmente, se logra controlar alguna fluctuación producto del azar en la selección de la muestra.

Por último, cuando se dispone de información auxiliar como la descrita en el párrafo anterior y las variables de interés están correlacionadas con dichas variables auxiliares, es razonable aprovechar esta correlación para reducir la varianza de los estimadores. Se agrega así, un motivo más para el ajuste de los ponderadores.

1.1. Objetivos

El principal objetivo de este documento es describir los procedimientos de ponderación utilizados más frecuentemente a la hora de atender los requerimientos anteriores y lograr estimadores que sean aproximadamente insesgados y consistentes. Si bien es generalmente aceptado que los datos deben ser ponderados el procedimiento por el que se calculan los ponderadores es poco conocido y visto como una caja negra para muchos de los usuarios de este tipo de datos.

Un segundo objetivo es ilustrar, a través de un ejemplo, la implementación de los procedimientos de ponderación analizados. A pesar de que existe *software* específico para el tratamiento de datos provenientes de encuestas por muestreo, en la actualidad, casi cualquier paquete estadístico incorpora rutinas que permiten su tratamiento y en particular realizar los procedimientos de ajuste que se van a presentar en este documento. En particular se opta por el lenguaje `R` (R Core Team, 2017) y dentro de éste el *package survey* (Lumley, 2017).

Algunas de las razones para esta elección son: i) la distribución del `R` es libre lo que está en consonancia con los objetivos de divulgación del documento ya que está al alcance de usuarios no especializados; ii) el `R` posee una muy buena documentación, en especial el *survey*, que además cuenta con un libro dedicado a su uso y cuyo autor es el mismo que se encarga de mantener al paquete *survey* (Lumley,2010); iii) el *survey* y en general el R están en permanente actualización incorporando los avances en la estadística en general como en el área específica de muestreo de poblaciones finitas; por último, iv) el `R` tiene un uso muy extendido, tanto a nivel académico como de empresas públicas y privadas.

El ejemplo y el código usado para la implementación de los procedimientos de ponderación que se presentan son objeto de la sección 7.

1.2. Tipo de encuestas a tratar

Los métodos de ponderación son distintos según las razones que persiga y del tipo de encuesta de que se trate. En lo que sigue se supone que se trata de una encuesta de gran escala, dirigidas a hogares y personas, donde se releva información sobre muchas variables, existe la necesidad de obtener estimaciones rápidamente, y, posiblemente, la información sea sensible desde el punto de vista político. Lo usual es que este tipo de encuestas sea de forma presencial, lo que implica que las muestras sean seleccionadas bajos diseños muestrales complejos, estratificados y con varias etapas de selección de forma de minimizar la dispersión geográfica de la muestra. Adicionalmente, si bien pueden usarse distintos ponderadores según la variable que se analice, en principio, se busca un único sistema o juego de ponderadores para ser aplicados a todas las variables de interés.

Si bien la descripción anterior encaja perfectamente en el tipo de encuestas que se procesan en las oficinas de producción estatal el interés recae fundamentalmente en las encuestas que realizan empresas privadas o institutos de investigación cuya actividad principal no es la elaboración de estadísticas oficiales si no que tienen objetivos más modestos y no cuentan con un personal especializado para su tratamiento. En este sentido el trabajo que se presenta es más de divulgación que de carácter estrictamente técnico debido a que cada vez es más frecuente la producción de encuestas a medida.

El énfasis que se pone en este tipo de encuestas determina que por motivos de imparcialidad, transparencia y facilidad de comunicación se adopte un enfoque de inferencia no basada en modelos, o en todo caso, asistido por modelos, que en la práctica es el más usado para este tipo de encuestas. Aunque siempre es discutible cuando algunos supuestos constituyen o no un modelo, en lo que sigue se entiende que asistido por modelos refiere al caso en que los estimadores usando los ponderadores finalmente obtenidos sean aproximadamente insesgados y consistentes en el diseño (Särndal et al., 1992). Puede considerarse una excepción el caso de los ajustes por no respuesta.

Adicionalmente, en lo que sigue, se supone que el fin último de la ponderación es descriptivo. Se busca, a través de las estimaciones de medias, totales y razones, para las variables de interés, caracterizar lo más fielmente posible a una población concreta. Por tanto se sigue el enfoque inferencial tradicional en muestreo y se considera que las variables de interés son valores fijos asociados a cada individuo de la población. En muchos casos, se requiere dar un paso más y se plantean estudios analíticos que suponen un modelo que relaciona algunas de las variables observadas. En estos casos el énfasis está puesto en la estimación de algunos parámetros que caracterizan al modelo propuesto y describen una super población que tuvo una realización particular en la población de la que se extrajo la muestra.

Si bien la ponderación juega un papel muy importante el los estudios analíticos, estos temas no serán tratados en lo que sigue.

1.3. Procedimiento general para la construcción de ponderadores

En general, la construcción de los ponderadores se realiza siguiendo una serie de etapas que se analizarán detalladamente en las siguientes secciones (Vaillant et al., 2013).

El punto de partida para los procedimientos de ponderación es la construcción de los ponderadores originales, o ponderadores base, w_{0k}. Estos deben expandir los datos de la muestra de forma que representen a la totalidad de la población objetivo, tomando en cuenta las distintas probabilidades de inclusión de las unidades del marco en la muestra. El procedimiento es bastante directo, y aunque depende de la complejidad del diseño muestral, la regla para obtener estimadores insesgados (en ausencia de otro tipo de error) consiste en multiplicar los datos obtenidos por el inverso de su probabilidad de selección. Es decir, $w_{0k} = 1/\pi_k$ donde π_k es la probabilidad de el elemento k sea seleccionado. En otras palabras, para estimar un total poblacional, los estimadores del tipo Horvitz - Thompson, $\sum_s y_k/\pi_k = \sum_s w_{0k} y_k$ son insesgados para $t_y = \sum_U y_k$ independientemente del diseño, en la medida que se trate de un diseño aleatorio. En la sección 2 se reseña brevemente la teoría de la estimación insesgada.

En un segundo paso, los ponderadores w_{0k} son usualmente ajustados debido a que algunos de los elementos seleccionados en la muestra no logran ser contactados y no existe evidencia de su elegibilidad, es decir, no se puede determinar si los mismos pertenecen o no a la población objetivo. Como resultado de esta etapa se obtienen los ponderadores w_{1k}. La propia naturaleza de estos elementos, su elegibilidad desconocida, determina que los procedimientos de ajuste suelan ser muy sencillos debido a que no se cuenta con información auxiliar. La construcción de los ponderadores w_{1k} se analiza en la sección 3.

En una tercer etapa, los ponderadores se corrigen a causa de la no respuesta, o sea se busca que los ponderadores correspondientes a las unidades efectivamente observadas representen también a las unidades elegibles de las que no se obtuvo información. Si bien existen distintos enfoques para el tratamiento de los problemas de no respuesta, generalmente se opta por el ajuste en celdas, que se basa en supuestos más débiles que los enfoques basados en modelos y requiere menos información auxiliar. La construcción de los ponderadores ajustados por la no respuesta tiene como resultado un vector $w_{2k} = nr_k w_{1k}$ y es el objeto de la sección 4.

Por último, los ponderadores w_{2k} se someten a un procedimiento de calibración, que persigue los objetivos de coherencia con la información auxiliar sobre totales poblacionales conocidos, la reducción de varianza en los estimadores finales y subsanar posibles problemas de subcobertura en el marco muestral. En esta etapa se parte de los ponderadores w_{2k} y se busca modificarlos lo “menos posible”, según alguna función de distancia, de forma que los ponderadores finales, w_{3k}, cumplan las llamadas ecuaciones de calibración. O sea, $\sum_s w_{3k}\mathbf{x}_k = \sum_U \mathbf{x}_k$, donde $\mathbf{x}_k$ es un vector de variable auxiliares conocidas. Este procedimiento se analiza en la sección 5.

En la sección 6 se analiza el control de algunos efectos negativos que puede arrojar el uso de ponderadores muy variables. Finalmente, en la sección 8 se presenta una serie de conclusiones y recomendaciones.

2. Cálculo de los ponderadores base, w_{0k}

El punto de partida de cualquier encuesta por muestreo es la definición del conjunto de las unidades de las que se quiere obtener información. Dicho conjunto se denomina población objetivo (población de interés o simplemente población), se anota U y su tamaño, posiblemente desconocido, es N. El conjunto U debe estar definido sin ambigüedades, y suele estar acotado en tiempo y espacio. Por ejemplo, personas residentes en hogares particulares del departamento de Montevideo durante los meses en que se realiza la encuesta. El objetivo final es obtener estimaciones para los totales, medias y ratios de algunas variables definidas, ya sea sobre los elementos de la población total o para distintos subconjuntos de la población, denominados dominios. Es una buena práctica que dichas estimaciones sean acompañadas de algún tipo de medida de su variabilidad, como ser, el error estándar, coeficiente de variación o intervalos de confianza.

Para lo que sigue son útiles algunas definiciones que esencialmente están tomadas de Särndal et al. (Särndal et al., 1992).

En principio una muestra, s es cualquier subconjunto de la población U. En la práctica, la selección los elementos de s se realiza a partir de un listado denominado marco. El marco muestral se define como el conjunto de unidades, procedimientos y mecanismos que identifican, distinguen y permiten acceder a los elementos de la población U. Los elementos del marco se denominan "unidades" para distinguirlos de los "elementos" de la población. Los elementos de la población que están incluidos en el marco constituyen lo que se denomina población marco. Cuando el marco coincide con U, se dice que se realiza un muestreo directo de elementos.

Para el tipo de encuestas en las que se está pensando, lo anterior no es posible. La selección física de la muestra suele realizarse a partir de un listado de grupos de elementos y que, en general, por diversas razones (desactualizaciones, duplicaciones, errores, etc.) no contiene ni a todos, ni solamente los elementos de la población objetivo. Siguiendo con el ejemplo anterior, para una encuesta de personas es común que se utilice el listado de viviendas del último censo disponible. Luego, es a través de la selección de viviendas que se encuentren ocupadas al momento de la realización de la encuesta que se llega a la selección de personas.

Se dice que la muestra es probabilística, o aleatoria, cuando se cumplen las siguientes condiciones:

1. El conjunto de todas las muestras posibles, $\mathscr{S}$, la colección de los subconjuntos de U que pueden ser seleccionados, es conocida.
2. Cada muestra posible tiene asignada una probabilidad de ser seleccionada $p(s)$; dicha asignación debe asegurar que cada elemento k de la población tiene una probabilidad no nula de pertenecer a s, $\pi_k = P(k \in s)$.
3. De $\mathscr{S}$ se selecciona una única s con un mecanismo aleatorio que asegura que cada s tenga probabilidad de ser seleccionada $p(s)$.

   La función $p(s)$ define una distribución de probabilidades en $\mathscr{S}$, o sea, cumple

$$p\,(s) \geqslant 0 \;\forall\; s \in \mathscr{S} \quad \text{y} \quad \sum_{s \in \,\mathscr{S}} p(s) = 1.$$

y es denominada diseño muestral.

Se define el vector $\mathbf{I} = (I_1, \ldots, I_k, \ldots, I_N)'$ donde cada $I_k : \mathscr{S} \longrightarrow \{0, 1\}$ tal que

$$I_k(s) = \begin{cases} 1 \text{ si } k \in \; s \\ 0 \text{ si } k \notin \; s \end{cases}$$

De esta manera, la distribución conjunta de $\mathbf{I}$ determina el diseño muestral. O sea, la probabilidad de extraer la muestra particular s es $p(s) = P(I_1 = i_1, \ldots, I_k = i_k, \ldots, I_N = i_N)$ donde i_k vale 1 si el elemento k pertenece a la muestra s y cero en caso contrario.

Con los elementos anteriores, si la población y el marco coinciden y la variable y es medida sin error para todos los elementos de la muestras se tiene que

$$E\left(\sum\nolimits_s y_k/\pi_k\right) = E\left(\sum\nolimits_U I_k y_k/\pi_k\right) = \sum\nolimits_U y_k = t_y,$$

y por tanto, el estimador $\sum_s y_k/\pi_k$ es insesgado para estimar el total en la población de la variable y, t_y. Así, tomando $w_{0k} = 1/\pi_k$ resuelve el primer paso para la construcción de los ponderadores.

Para terminar esta sección, vale la pena mencionar el caso en que la muestra sea seleccionada con reposición el principio anterior se generaliza y lo correcto es considerar $w_{0k} = 1/E(R_k)$ con R_k el número de veces que k puede ser elegido para participar en la muestra.

2.1. Problemas de marco

El supuesto de que la población objetivo y la población marco coinciden, es casi siempre, irreal. Por definición la población objetivo es básicamente una construcción téorica mientras que la segunda es el mejor listado que se dispone para realizar la selección de la muestra.

Para el tipo de encuestas analizadas es fácil ver el tipo de dificultades que aparecen. En el ejemplo, el listado de viviendas data del momento del último censo, y en término medio, suponiendo censos de viviendas decenales, va a tener una desactualización de cinco años. Más allá que los listados de viviendas censales están sujetos a error, siempre existirán viviendas de construcción reciente posterior a la realización del listado y viviendas desocupas o demolidas que estaban listadas al momento de la realización del censo que originan problemas de subcobertura y sobrecobertura. Dichos problemas pueden sesgar considerablemente las estimaciones, por ejemplo, las viviendas nuevas presentan un patrón de crecimiento donde el mayor crecimiento de viviendas se da en asentamientos irregulares y/o en zonas con poder adquisitivo alto. Adicionalmente, las direcciones, la forma de llegar hasta las viviendas, pueden tener errores o haber cambiado, lo que implica que no puedan ser ubicables.

Otro problema común en los marcos es la multiplicidad de unidades que ocurre cuando existen viviendas que están listada más de una vez, aunque para el caso del ejemplo esto parece ser un problema menor.

3. Cálculo de los ponderadores ajustados por elegibilidad desconocida, w_{1k}

En esta sección se tratan los problemas asociados a la elegibilidad desconocida. Sin importar los esfuerzos que se realicen para la conformación del marco de muestreo van a existir unidades en el marco y en la muestra, cuya elegibilidad es desconocida. Por ejemplo, en una encuesta a hogares, no se tiene información si la vivienda es ocupada. Por variadas razones, puede ocurrir que no sea posible determinar la elegibilidad de la vivienda. Por ejemplo, no es posible ubicar la vivienda en campo (dirección no especificada) o luego de varios intentos no fue posible contactar a la vivienda y no existe evidencia alguna para determinar su elegibilidad.

Si hay unidades en la muestra que son detectadas como no elegibles, es razonable suponer que existen otras unidades no elegibles dentro de las unidades de la muestra con eligibilidad desconocida como en la parte del marco no muestreada. Esto implica que sea necesario algún tipo de ajuste.

Siguiendo la clasificación propuesta en Vaillant et al. (Vaillant et al., 2013) se considera los siguientes subconjuntos de la muestra seleccionada

$s_{ED} =$ subconjunto de la muestra con elegibilidad desconocida

$s_E =$ subconjunto de la muestra que son elegibles

$s_I =$ subconjunto de la muestra que son inelegibles

$s_{ER} =$ subconjunto de la muestra que son elegibles respondentes

$s_{ENR} =$ subconjunto de la muestra que son elegibles pero no responden

$s_{EC} =$ subconjunto de la muestra que su elegibilidad es conocida.

Los ajustes por elegibilidad desconocida suelen ser bastante sencillos dada la poca información auxiliar con que se cuenta, por otra parte, dado que el ajuste por no respuesta se considera de mayor

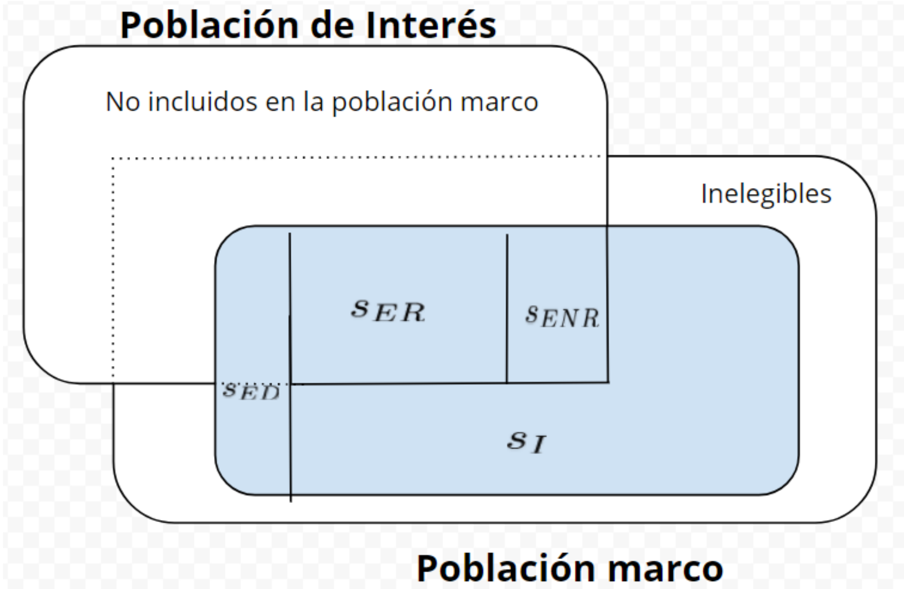

Figura 1: Población objetivo, población marco y subconjuntos de la muestra.

relevancia es natural que el mayor esfuerzo se concentre en ese tipo de ajuste. El método más sencillo consiste en distribuir la suma de los ponderadores de las unidades sobre las que se desconoce su elegibilidad entre aquellas unidades cuya elegibilidad es conocida. Lo más común es adoptar una estrategia que distinga subgrupos identificables con la información disponible a nivel de la muestra en su conjunto, tales que se pueda suponer que la falta de información sobre su condición de elegibilidad es al azar, o ignorable. Un enfoque similar puede ser usado para el tratamiento de la no respuesta y a menos que se cuente con distinta información auxiliar los grupos o clases consideradas pueden ser los mismos.

La estrategia de ajuste puede sintetizarse como sigue (Vaillant et al. 2013)

i. Se parte de particionar la muestra seleccionada, s, en $1, \ldots, b, \ldots, B$ clases o celdas con la información disponible a nivel del marco. En principio, las clases no tienen que coincidir con los estratos ni con las unidades primarias de muestreo (UPMs), aunque en la práctica es común que coincidan. El objetivo es lograr que en las clases formadas el desconocimiento sobre la elegibilidad de una unidad pueda considerarse al azar. Por lo tanto cuanto más se correlacionen las variables usadas para formar las clases con las causas de la no identificación de la elegibilidad las celdas formadas cumplirán mejor su objetivo. Este tipo de razonamiento se utiliza también a la hora de tratar la no respuesta y será analizado más en profundidad.

ii. Para cada clase formada, se define s_b como el subconjunto de unidades en la muestra pertenecientes a la clase b sin importar su condición de elegibilidad ni su estatus de respondente.

iii. El ajuste por elegibilidad desconocida de las unidades en la clase b es entonces

$$de_k = \begin{cases} \sum\limits_{s_b} w_{0k} \;\; / \sum\limits_{s_b \cap s_{EC}} w_{0k} & \text{si } k \in \; s_b \cap s_{EC} \\ \\ 0 & \text{si } k \in \; s_b \cap s_{ED} \end{cases}$$

iv. Por último se obtiene

$$w_{1k} = de_k \times w_{0k} \;\; \forall \; k \in s.$$

Debido a que los nuevos ponderadores w_{1k} son igual a cero para los elementos con elegibilidad desconocida, dichos individuos son desechados de aquí en más para los ajustes de no respuesta y calibración.

4. Cálculo de los ajustados por no respuesta w_{2k}

En esta sección se analizan los ajustes por no respuesta. Parte de las unidades seleccionadas son elegibles pero que no terminan brindado toda la información requerida. Los ajustes a los ponderadores w_{1k} serán más o menos sofisticados dependiendo de la cantidad y la calidad de la información auxiliar disponible a nivel de respondentes y no respondentes.

Por más extremas que sean las medidas que se adopten, la no repuesta es casi imposible de ser eliminada y se hace necesario su tratamiento. La no respuesta ocasiona dos problemas: 1) reducción en la precisión (mayor varianza) de las estimaciones producto de una reducción en el tamaño de muestra efectivo y ii) un aumento en el sesgo de las estimaciones. En la práctica los tamaños de muestra son incrementados teniendo en cuenta la tasa de respuesta esperada de forma de minimizar el problema i), pero esta solución suele incrementar en ocasiones el sesgo. Un error común es pretender remediar el problema de la no respuesta únicamente aumentando el tamaño de muestra originalmente planeado. En pocas palabras, se remplazan no respondentes por respondentes con lo que no sólo no se enfrenta el problema si no que se incrementa la precisión del error que se pudiera estar cometiendo. Muchas veces el remplazo queda avalado porque las unidades sustitutas se eligen de manera de que ciertas características observables coinciden o por lo menos son similares. Se olvida, entonces, que una de esas características, por su parte definitoria del problema, es que los que responden, precisamente responden y los que no lo hacen, precisamente no lo hacen.

La no respuesta debe ser tratada, y su tratamiento implica suponer que existe un mecanismo por el cual se genera y luego hacer algunos supuestos que implican modelizar dicho mecanismo. Lamentablemente, dichos supuestos no pueden ser sometidos a prueba.

4.1. Generalidades de la no respuesta

En todas las encuestas por muestreo existe algún nivel de no respuesta, entendida como la falla en la obtención de la totalidad de los datos que se planificó recolectar para todas las unidades elegibles seleccionadas en la muestra. Este fenómeno es ajeno al diseño muestral, incluso para el caso de un censo, donde la muestra intenta abarcar a toda la población, la no respuesta estará presente.

En la literatura sobre el tema existen dos distinciones recurrentes. La primera, refiere al tipo de no respuesta. En este caso se distingue entre no respuesta de unidades y no respuesta de ítems. La no respuesta de unidades ocurre cuando algunas de las unidades seleccionada y elegibles no proporcionan ninguna información útil (se rehúsa a participar o luego de varias visitas no se logra establecer contacto). La no respuesta de ítems ocurre cuando, se obtiene solamente una parte la información requerida. Si bien ambos casos constituyen casos de no respuesta presentan grandes diferencias.

Generalmente, la no respuesta de unidades es considerada más grave. La no respuesta de una unidad seleccionada tiene dos efectos, uno equivalente a la reducción en el tamaño de la muestra final, el cual repercute en las precisiones de las estimaciones de los distintos parámetros producto de la reducción del tamaño de muestra, y otro, posiblemente más perjudicial, se debe a que el caso perdido no puede considerarse perdido al azar. En cambio, la no repuesta de ítems generalmente afecta a algunas y no todas las variables y distorsiona la correlación estimada entre variables. En ambos casos, la no respuesta genera sesgos más o menos graves.

La segunda distinción, refiere a la forma de tratar el problema. Por un lado las estrategias de prevención y por otro las estrategias de tratamiento.

Las estrategias de prevención, tanto para la no respuesta de unidades como de ítems, son variadas y en general tratan aspectos de la planificación de la encuesta, el entrenamiento de los encuestadores, el diseño de estrategias de revisitas *callbacks* y técnicas para aumentar la motivación a responder. En general, estos aspectos no son tratados con ajustes en la ponderación y son ajenos a este trabajo, con lo que no serán analizados. No obstante, es importante destacar que todos estos esfuerzos deben quedar rigurosamente registrados, de la misma manera que deben registrarse detalladamente los casos en que finalmente no se logra respuesta. Esta información puede ser muy útil a la hora del tratamiento de la no respuesta.

Dentro de las estrategias de tratamiento, tampoco se analizarán aquellas que implican un esfuerzo adicional a la hora de conseguir la respuesta, por ejemplo, el submuestreo de no respondentes.

Respecto a las estrategias de tratamiento que buscan corregir el problema en base al uso de información auxiliar, si bien no es la regla general, se puede decir que los casos de no respuesta de ítems se tratan con técnicas de imputación y la no respuesta de unidades con ajustes en la ponderación. Sin embargo hay técnicas que combinan ambas estrategias.

Las técnicas de imputación, están más o menos basadas en algún tipo de modelo, suelen incorporar solamente parte de la información que la unidad proporcionó correctamente y no involucran ajustes en los ponderadores y tampoco son objeto de este trabajo.

Respecto a las formas de tratar la no respuesta vía el ajuste de los ponderadores, como ya se dijo, existen procedimientos más o menos elaborados, según el tipo de la información auxiliar disponible. En principio, dichos procedimientos pueden basarse tanto en información auxiliar disponible solamente a nivel de las unidades seleccionadas, tanto respondentes como no respondentes, como en información auxiliar a nivel de toda la población. Existen diversos enfoques y matices según las posibles combinaciones de información auxiliar existente.

Por razones expositivas y por el tipo de encuestas analizadas, aunque con cierto riesgo de seguir un enfoque demasiado esquemático, se opta por distinguir tres casos: i) la información disponible para usar en los ajustes de no respuesta es a nivel de muestra pero no suficientemente detallada como para modelizar la no respuesta a nivel de individuos; ii) como en el caso anterior solamente se utiliza información a nivel de muestra, pero en este caso es suficientemente detallada para permitir una modelización a nivel de individuos; y iii) se dispone de información no suficientemente detallada pero dicha información está disponible tanto a nivel de muestra como de la población marco.

En iii) es de destacar un enfoque relativamente reciente y con bastante popularidad que plantean Särndal y Lundsröm (Särndal y Lundsröm, 2005). Este enfoque, pretende enmaracar el tratamiento de la no respuesta en un enfoque unificado del uso de estimadores calibrados dentro de la teoría del muestro de poblaciones finitas, es computacionalmente sencilla de implementar y generaliza otras técnicas del tratamiento de la no respuesta como la postestratificación, el raking (postestratificación incompleta) y algunos casos de los ajustes basados en la teoría del muestreo en dos fases.

En el caso ii) las técnicas de ajuste se basan en la estimación de la probabilidad o propensión de responder para cada unidad de la muestra (*Propensity score*). Si bien no se hará énfasis en este enfoque, se presentará de manera sucinta dado que puede ser la base para la construcción de clases o grupos de no respuesta que pueden ser útiles en el contexto del caso iii). Estas técnicas suelen ser más aplicables a otro tipo de encuestas, por ejemplo a empresas.

En lo que sigue se pondrá énfasis en el caso i) ya que tiene menos requerimientos, es apropiado pare el tipo de encuestas de hogares y personas y tiene un nivel de complejidad intermedio.

En el tipo de encuesta que se analizan la no respuesta es especialmente grave. En las encuestas oficiales suele existir la obligatoriedad de responder aunque la regla es que el ejercicio de esta norma no sea más que contraproducente. De cualquier manera, que la institución patrocinante de la encuesta sea el propio estado suele ser una ventaja frente a otro tipo de patrocinadores.

Trabajos relativamente recientes sobre el tema coinciden en subrayar que la no respuesta, sobre todo en encuestas a personas u hogares en el campo de las ciencias sociales, es un problema cada vez más relevante y con cierta tendencia a agudizarse (Groves, 2006, Olson, 2006 y Särndal y Lundström, 2005).

Otra aspecto clave para el correcto tratamiento de la no respuesta es la meticulosa codificación de las causas por las que no se obtuvo la información deseada: rechazo, no contacto, etc.

La no respuesta puede conceptualizarse bajo dos paradigmas

i. No repuesta determinística. En el marco existen dos grupos independientemente de la muestra que sea seleccionada (pueden pensarse como estratos) uno donde las unidades a ser seleccionadas responden y otro, donde no responden.

ii. No repuesta estocástica. Todas la unidades del marco tienen asociada un probabilidad de responder, que se denomina distribución de respuesta. Cuando una unidad es seleccionada en la muestra la unidad decide si responde o no de a cuerdo a dicha probabilidad.

En el enfoque estocástico la respuesta puede pensarse como un segundo mecanismo de selección que genera la muestra efectiva de respondentes, una vez tomada la muestra. En otra palabras, se supone que, condicional a la muestra observada, la respuesta sigue una distribución desconocida. Conceptualmente esto es equivalente a pensar que sobre los conjuntos s se define una distribución de la misma manera que sobre sobre los subconjuntos de U se define el diseño muestral. Mientras que el diseño determina las probabilidades de inclusión en la muestra, le distribución de la respuesta determina las probabilidades de que cada unidad seleccionada responda. El mecanismo final por el que se llega a obtener la respuesta de un elemento de la población es análogo a un diseño en dos fases. La diferencia fundamental es que mientras el diseño muestral, que sería el diseño de primera fase, es conocido, la distribución de la respuesta, el diseño de segunda fase es desconocido. Si las probabilidades de responder fueran conocidas el problema estaría resuelto bajo la teoría del muestreo en dos fases. El paso siguiente consiste en suponer un modelo que pueda estimarse en base a la información auxiliar disponible y que representa correctamente a la distribución de respuesta desconocida.

Antes de analizar los procedimientos de ajuste por no respuesta puede ser útil presentar una clasificación comúnmente utilizada de distintos tipos de no repuesta que pueden aplicarse tanto a la no respuesta de unidades como de ítems (Little y Rubin, 2002).

i. Respuesta totalmente al azar (MCAR, *Missing Completely at Random*). Es el caso en que la probabilidad de responder ϕ_k no depende de las variables de interés ni de variables auxiliares. Si puede suponerse que la no respuesta es MCAR, el único efecto de ésta será la reducción en el tamaño de muestra final y el ajuste necesario consiste en distribuir la suma de los ponderadores de los no respondentes de manera uniforma entre los respondentes.

ii. Respuesta aleatoria (MAR, *Missing at Random*). Cuando el estatus de repondente no depende de las variables de interés pero queda completamente explicada por algunas de las variables auxiliares se dice que los datos faltantes son MAR, o sea $\phi_k = \phi(\mathbf{x}_k)$ donde $\mathbf{x}_k$ es el vector de información auxiliar para el individuo k. En este caso, es posible construir un modelo de respuesta basado en la información auxiliar, que es conocida para respondentes y no respondentes de forma de obtener una buena representación de la distribución de la respuesta y solucionar satisfactoriamente el problema. Por lo anterior es común que en esos casos se diga que la no respuesta es ignorable.

iii. No repuesta no ignorable (NINR, Nonignorable Nonresponse). Si la distribución de la respuesta depende de algunas de las variables de interés, y se dice que la no respuesta es NINR. En este caso no es posible construir un modelo de respuesta apropiado ya que no se dispone de las variables de interés para las unidades que no responden.

Como se dijo, el principal problema de la no respuesta es el potencial sesgo que puede introducir en los estimadores. Es útil analizar el sesgo del estimador de un total bajo distintos supuestos sobre la distribución de la respuesta.

Bajo el supuesto de que la no respuesta puede considerarse MCAR, y siendo s_{ER} al subconjunto de los respondentes se tiene la distribución de la respuesta se tiene que

$$P(k \in s_{ER}|s) = \phi \; \forall \; k \in s_{ER} \cup s_{ENR}$$

por tanto, el estimador del total de la variable y basado en la muestra de respondentes viene dado por

$$\hat{t}_r = N\hat{\bar{y}}_r = N\frac{\sum_{s_{ER}} y_k/\phi\pi_k}{\sum_{s_{ER}} 1/\phi\pi_k} = N\frac{\sum_{s_{ER}} y_k/\pi_k}{\sum_{s_{ER}} 1/\pi_k}$$

y es aproximadamente insesgado.

En un contexto de respuesta determinística se supone que la población U está particionada en dos subpoblaciones U_1 y U_2 de tamaños respectivos N_1 y N_2. Los elementos de U_1 responden con probabilidad 1, mientras que los elementos en U_2 no responden. Sean $\bar{Y}_1$ y $\bar{Y}_2$ las medias de la variable y en cada una de las subpoblaciones. Entonces, el sesgo del estimador del total de la variable y basado en la muestra de respondentes viene dado por

$$E(\hat{t}_r - t_y) = N_2(\bar{Y}_1 - \bar{Y}_2)$$

es decir, el sesgo depende positivamente del tamaño de la población de los no respondentes y de la diferencia en el promedio de la variable de interés entre los que responden y no responden. Si respondentes y no respondentes tienen el mismo promedio e sesgo es nulo.

Bajo el modelo estocástico sencillo también es fácil analizar el sesgo del estimador anterior. Bajo la distribución de respuesta dada por

$$P(k \in s_{ER}|s) = \phi_k \; \forall \; k \in s_{ER} \cup s_{ENR}$$

y se obtiene

$$E(\hat{t}_r - t_y) \doteq N \frac{\sum_U y_k \phi_k}{\sum_U \phi_k} - t_y$$

lo que implica un sesgo relativo dado por

$$\frac{E(\hat{t}_r - t_y)}{t_y} \doteq R_{y\phi U} cv_{yU} cv_{\phi U}$$

donde $R_{y\phi U}$ es la correlación en la población entre la probabilidad de respuesta, ϕ, y la variable de interés y, cv_{yU} es el coeficiente de variación de la variable y en U y $cv_{\phi U}$ el coeficiente de variación de las probabilidades de respuesta en la población. Así, el sesgo relativo del estimador $\hat{t}_r$ depende de la correlación entre la variable de interés y la probabilidad de respuesta.

4.2. Ajustes basados en grupos de respuesta homogéneos

Un modelo sencillo y comúnmente usado es el modelo de grupos de respuesta homogéneos que consiste en clasificar a los individuos elegibles de la muestra en C grupos, exhaustivos y excluyentes, $s_1, \ldots, s_c, \ldots, s_C$ para los cuales se supone que las probabilidades de responder son similares y que los individuos responden de manera independiente (Sa̋rndal, et al 1992). Se supone entonces,

$$P(k \in s_{ER}|s) = \phi \; \forall \; k \in s_c$$

Los grupos deben crearse de manera tal que las probabilidades de responder dentro de cada grupo se mantengan aproximadamente constantes y/o que tengan aproximadamente los mismos valores en las variables de interés. De esta manera la mejor partición en grupos será la que mejor describa el estatus de respondente y los distintos niveles de los valores de las variables de interés. En la práctica, dado que los valores de las variables de interés son desconocidos para los no respondentes, por lo tanto los grupos formados pueden no ser igualmente efectivos en su descripción, con lo cual, la formación de los grupos suele basarse únicamente en las tasas de respuesta. Enfoques algo más complejos incorporan información auxiliar a nivel poblacional en la construcción de los grupos de forma que estos sean buenos predictores de las variables de interés y de esta forma de permitir mejores ajustes.

En el caso más sencillo, el ajuste por no respuesta, nr, con la notación que se viene usando vienen dado por

$$nr_k = \begin{cases} \sum\limits_{s_c \cap s_E} w_{1_k} \;/\; \sum\limits_{s_c \cap s_{ER}} w_{1k} & \text{si } k \in \; s_c \cap s_{ER} \\ \\ 1 & \text{si } k \in \; s_c \cap s_{EI} \\ \\ 0 & \text{si } k \in \; s_{ED} \cap s_{ENR} \end{cases}$$

o sea, el ajuste nr_k es el inverso de la tasa de respuesta ponderada en la clase a la que pertenece el individuo k.

Otra opción puede ser no considerar en el cálculo de de los nr los ponderadores w_{1k} de forma que cuando nr no sea cero o uno su valor sea $\#s_c/\#(s_c \cap s_{EC})$. Si las probabilidades de respuesta son efectivamente muy similares dentro de cada grupo esta opción hace más estables a los ajustes nr y será útil en la medida que los w_{1k} sean muy variables.

Por último, se tiene $w_{2k} = nr_k \times w_{1k} = nr_k \times de_k \times w_{0k}$.

4.3. Otras estrategias de ajuste

Como se mencionó anteriormente, si las probabilidades de responder, $P(k \in s_{ER}|s) = \phi_k$ son conocidas entonces, considerando $w_{2k} = 1/\pi_k\phi_k$ se otiene un estimador insesgado. Luego si $\phi_k = \phi(\mathbf{x}_k)$ es razonable intentar construir un modelo para las probabilidades de respuesta siempre y cuando el vector $\mathbf{x}_k$ esté disponible para todas las unidades elegibles de la muestra. Lo anterior implica que la respuesta sea ignorable condicional a la información auxiliar con que se cuenta. En última instancia, desde un punto de vista pragmático, sin este supuesto hay muy poco para hacer.

Una vez que se asume $P(k \in s_{ER}|s) = \phi(\mathbf{x}_k)$ el paso natural que sigue es proponer un modelo de regresión binario para predecir la respuesta. Rosenbaum y Rubin denominan a las cantidades $\phi(\mathbf{x}_k)$ *Propensity Scores* (Rosenbaum y Rubin, 1983).

Sin entrar en un análisis detallado de la estrategia para estimar los *propensity scores* vale la pena señalar que involucra los aspectos usuales en la estimación de este tipo de modelos. La selección de variables puede ser un tema delicado, aunque aquí el interés es lograr un buen nivel explicativo y se puede correr el riesgo de agregar variables o interacciones entre variables a cuenta de mejorar el ajuste. Un tema adicional es si el modelo debe o no ser estimado considerando los ponderadores w_{1k}.

Una vez estimadas las probabilidades de respuesta hay varias estrategias alternativas para su uso. El más directo es construir los ponderadores $w_{2k} = nr_k w_{0k}$ con $nr_k = \hat{\phi}(\mathbf{x}_k)$. Los estimadores basados en éstas cantidades serán a lo sumo aproximadamente insesgados aunque los ajustes individuales puede dar lugar a ponderadores muy extremos. Otras alternativas se basan en la formación de grupos de respuesta basados en el *propensity scores* estimados con la lógica de hacer más estables las estimaciones finales de la respuesta. Una vez formados los grupos, no muchos más que cinco, se utilizan las medias o las medianas de los valores $\hat{\phi}(\mathbf{x}_k)$ como representantes de cada grupo.

Otro método para la formación de grupos de respuesta para el ajuste por no respuesta es vía algoritmos de clasificación, por ejemplo, CART, *classification and regression trees* o CHAID, *chisquare automatic interaction detection*. Sin entrar en detalle, estas técnicas buscan clasificar a las unidades muestreadas en respondentes y no respondentes en basa a las covariables disponibles. Los algoritmos de clasificación pueden presentar ventajas frente a la modelización del *propensity score* ya que se resuelven muchos problemas de manera automática, por ejemplo, que variables entran en el proceso, la forma en que interactúan y la formación de categorías a partir de variables continuas.

5. Cálculo de los ponderadores calibrados w_{3k}

El último paso en la construcción de los ponderadores consiste en usar alguna técnica de calibración, o más precisamente, el uso de estimadores calibrados (Deville y Särndal, 1992). Los estimadores calibrados abarcan una clase muy amplia de estimadores basados en el uso de información auxiliar. Otros procedimientos más antiguos que incorporan información auxiliar, son casos particulares de la calibración. En particular, las técnicas de *raking* y los estimadores postestratificados, pertenecen a dicha clase. (Deville et al., 1993). Otra caso particular de la clase estimadores calibrados son los estimadores regresión generalizados (GREG, *general regression estimator*) (Fuller, 2002).

La calibración se justifica por varios motivos. En primer instancia, los estimadores calibrados se proponen con el objeto de reducir la varianza del estimador de un total. A diferencia de los estimadores $GREG$, aquí no se parte de suponer ningún modelo que describa a la población de interés. Simplemente se considera un vector de ponderadores iniciales, d_k, tales que $\hat{t}_y = \sum_s d_k y_k$ es aproximadamente insesgado y consistente para t_y. Se busca entonces, un nuevo juego de ponderadores, w_k, que cumplan dos condiciones, que modifiquen 'lo menor posible' los ponderadores iniciales y que cumplan $\sum_s w_k \mathbf{x}_k = \sum_U \mathbf{x}_k$.

Si la primera condición se cumple, la modificación que se realiza a los ponderadores de partida es pequeña y se preservaran, por lo menos aproximadamente, las buenas propiedades que estos tuvieran. La segunda condición, lleva a que si los nuevos ponderadores son aplicados para estimar los totales de las variables auxiliares, se obtendrá, independientemente de la muestra obtenida, estimaciones que coinciden con los totales, y verdaderos, totales poblacionales. De esta manera, si las variables auxiliares se correlacionan con las variables de interés es de esperar que el error nulo en la estimación de los totales poblacionales en las variables $\mathbf{x}$ de alguna forma se traduzca en un error de muestreo pequeño en las estimaciones de las variable de interés, y.

La segunda condición tiene un atractivo adicional ya que las estimaciones que se realicen con los ponderadores obtenidos de esta manera serán coherentes con datos poblacionales conocidos. Así se obtiene una segunda justificación al uso de los estimadores calibrados.

Por último, existe una razón adicional para el uso de estimadores calibrados, al ajustarse las cantidades estimadas en base a la muestra a los totales poblacionales conocidos, esto contribuirá a subsanar potenciales problemas de subcobertura en el marco muestral.

5.1. Derivación del Estimador Calibrado

El objetivo es estimar $t_y = \sum_U y_k$. Sea $\mathbf{x}_k$ un vector de variables auxiliares conocidas $\forall\ k \in U$ y que no existen problemas de elegibilidad desconocida ni de no respuesta con lo que se cuenta con un vector de ponderadores d_k de manera que $\hat{t}_y = \sum_s d_k y_k$ es aproximadamente insesgado para t_y. Se busca entonces un nuevo juego de ponderadores, w_k que cumplan las siguientes restricciones: i) la modificación debe ser 'lo menor posible', o sea, debe minimizar una medida de distancia entre los vectores $(d_1, \ldots, d_k, \ldots, d_n)'$ y $(w_1, \ldots, w_k, \ldots, w_n)'$; ii) se debe cumplir que $\sum_s w_k \mathbf{x}_k = \sum_U \mathbf{x}_k$, o sea, los nuevos ponderadores permiten estimar sin error los totales poblacionales conocidos de las variables auxiliares.

Para la función de distancia se considera una función G de argumento $x = w_k/d_k$ con las siguientes propiedades

1. G es positiva y estrictamente convexa
2. $G(1) = G'(1) = 0$
3. $G''(1) = 1$

De esta manera, $G(w_k/d_k)$ mide la distancia entre los ponderadores originales d_k y los ponderadores nuevos w_k. Luego, $\sum_s d_k G(w_k/d_k)$ es la distancia total estimada para la muestra s. Se busca que el nuevo vector de ponderadores minimice la distancia total estimada y cumpla las llamadas ecuaciones de calibración

$$\sum\nolimits_s w_k \mathbf{x}_k = \sum\nolimits_U \mathbf{x}_k$$

Se tiene entonces el siguiente problema de optimización

$$\min_{w_k} \sum\nolimits_s d_k G(w_k/d_k) - \boldsymbol{\lambda}'(\sum\nolimits_s w_k \mathbf{x}_k - \sum\nolimits_U \mathbf{x}_k)$$

donde $\boldsymbol{\lambda} = (\lambda_1, ..., \lambda_j, ..., \lambda_J)'$ es el vector de multiplicadores de Lagrange.

Derivando respecto a w_k, y llamando $g(x) = G'(x)$ se obtiene

$$g\left(\frac{w_k}{d_k}\right) - \boldsymbol{\lambda}'\mathbf{x}_k = 0$$

finalmente, despejando

$$w_k = d_k F(\mathbf{x}_k' \boldsymbol{\lambda})$$

donde $F(u) = g^{-1}(u)$.

Los supuestos hechos sobre G garantizan que su función inversa, F, exista y el problema de optimización tenga solución única . Para calcular los nuevos ponderadores se debe, determinar el valor de $\boldsymbol{\lambda}$. Esto se consigue resolviendo las ecuaciones de calibración, que se obtienen sustituyendo los w_k que se obtienen de la ecuación anterior en la restricción dada por la ecuaciones de calibración, así

$$\sum\nolimits_s d_k F(\mathbf{x}_k' \boldsymbol{\lambda}) \mathbf{x}_k' = \sum\nolimits_U \mathbf{x}_k'$$

donde $\boldsymbol{\lambda}$ es desconocido. Para obtener $\boldsymbol{\lambda}$ del sistema anterior generalmente se utilizan métodos numéricos. Una vez que se obtiene está determinado, se calculan los ponderadores calibrados $w_k = d_k F(\mathbf{x}_k' \boldsymbol{\lambda})$ y el estimador calibrado de t_y viene dado por

$$\hat{t}_{ycal} = \sum\nolimits_s w_k y_k = \sum\nolimits_s d_k F(\mathbf{x}_k' \boldsymbol{\lambda}) y_k$$

De lo anterior queda claro que el cálculo de los w_k requiere el conocimiento de los totales poblacionales y los valores de las variables utilizadas para los elementos de la elegibles y que responden de la muestra. De esta manera, los ponderadores calibrados pueden construirse en base al conocimiento de los totales poblacionales de algunas variables, en la medida en la encuesta se recojan, de la misma manera en que éstos totales hayan sido medidos, las variables auxiliares en cuestión.

En la práctica, alguna de las funciones de de distancia G y las funciones F asociadas son:

1. *lineal*: $G(x) = (1/2)(x-1)^2 \ \forall\, x \in \mathbf{R};\ F(u) = 1 + u$.
2. El modelo *multiplicativo (o raking)*: $G(x) = x \ln x - x + 1 \ \forall\, x \in \mathbf{R}^+;\ F(u) = \exp(u)$.
3. El modelo *logit*(L, U): Dadas dos constantes L y U tales que $L < 1 < U$ y $A = (U-L)/\{(1-L)(U-1)\}$. Entonces se define

   $$G(x) = \frac{1}{A}\left[(x-L)\log\left(\frac{x-L}{1-L}\right) + (U-x)\log\left(\frac{U-x}{U-1}\right)\right]$$

   si $L < x < U$ y $G(x) = \infty$ en otro caso; la función F correspondiente es

   $$F(u) = \frac{L(U-1) + U(1-L)\exp(Au)}{U - 1 + (1-L)\exp(Au)}$$

   que toma valores en (L, U).
4. El modelo *lineal truncado*(L, U): Dadas dos constantes L y U tales que $l < 1 < U$ se define $G(x) = (1/2)(x-1)^2$ si $L < x < U$ y $G(x) = \infty$ en otro caso; la función F correspondiente es, $F(u) = 1 + u$ si $u \in [L-1, U-1]$, $F(u) = L$ si $u < L - 1$ y $F(u) = U$ si $u > U - 1$.

Los modelos 3. y 4. son útiles porque permiten controlar el rango de variación de los nuevos ponderadores.

El modelo 1. conduce a los estimadores GREG, donde el modelo implícito es

$$E(y_k) = \mathbf{x}'_k\boldsymbol{\beta} \text{ y } V(y_k) = \sigma_k^2$$

Si la información auxiliar simplemente indica la pertenencia o no a una partición de la población en $1, \ldots, g, \ldots, G$ celdas, se obtiene el estimador postestratificado, sin importar la función de distancia. El modelo implícito es

$$E(y_k) = \beta_g \text{ y } V(y_k) = \sigma_g^2$$

Una alternativa a la postestratificación es el *raking*, que utiliza la información de las distribuciones marginales de una tabla de contingencia. Bajo el modelo 2. se obtiene el *raking* tradicional, que resulta en ponderadores estrictamente positivos. Con otros modelos se otendrán otros estimadores, denominados como de *raking* generalizado. Estas opciones requieren menos información que el estimador postestratificado, lo que reduce problemas de estabilidad que pudieran aparecer cuando se pretende ajustar a todas las celdas de la tabla y en algunos casos hay pocas observaciones. Adicionalmente permite combinar distintas fuentes de información. El modelo implícito en el *raking* es un ANOVA a dos vías, o sea

$$E(y_k) = \mu + \alpha_i + \beta_j \text{ y } V(y_k) = \sigma^2$$

Algunos autores suelen enfatizar las diferencias en el razonamiento que hay detrás de los estimadores GREG y la calibración (Särndal, 2007). Si bien en ambos casos se incorpora información auxiliar para obtener estimadores con mejores propiedades, mientras que los estimadores GREG parten de la especificación de un modelo que asista a la estimación, la calibración persigue un objetivo de coherencia expresado en las ecuaciones de calibración.

A pesar de lo anterior pensar en el modelo subyacente puede ser útil para facilitar la comprensión. En este sentido vale la pena ensayar distintos modelos, seleccionar distintas las variables, analizar posibles transformaciones de éstas y realizar análisis de diagnostico del modelo que finalmente se vaya a utilizar.

Por último, en la práctica se trabaja únicamente con la muestra de elegibles respondentes, por lo cual, los ponderadores d_k son sustituidos por los ponderadores ajustados por no respuesta w_{2k} y se obtiene

$$w_{3k} = g_k \times w_{2k} \text{ si } k \in s_{ER}$$

donde $g_k = F(\mathbf{x}'_k\boldsymbol{\lambda})$, $\mathbf{x}_k$ el vector de variables auxiliares y $\boldsymbol{\lambda}$ es el vector de multiplicadores de Lagrange.

Vale la pena resaltar que los ponderadores gk, si bien tienen variabilidad muestral, no dependen de los valores y_k con lo que los ponderadores obtenidos bajo un modelo pueden utilizarse para todas o muchas de las variables de interés.

6. Variabilidad de los ponderadores

Es normal que los ponderadores finales sean muy variables. Dicha variabilidad responde a distintas fuentes que se asocian a las distintas etapas en los procedimientos de ajuste analizados. En primer lugar, el diseño muestral puede considerar probabilidades de inclusión muy distintas. Etapas de selección proporcional al tamaño, estratos con tasas de muestreo muy diferentes,PSUs con tamaños muy variables y sobremuestreo de algunos dominios de particular relevancia son frecuentes en las encuestas de hogares y personas. En segundo lugar, las tasas de respuesta y/o de elegibilidad desconocida distintas por grupos incrementan la variabilidad. En tercer lugar, la imposición del cumplimiento de las ecuaciones de calibracı'on puede forzar a los ponderadores a variaciones de importancia y/o a situaciones indeseadas, por ejemplo, ponderadores negativos.

En principio, la variabilidad en los ponderadores no es necesariamente un problema. Los distintos ponderadores implicados por el diseño pueden ser muy eficientes si reflejan la importancia relativa en los valores de las variables de interés entre las distintas unidades de la población. Lo anterior es común en las encuestas a empresas en donde las tasas de muestreo en los estratos conformados por las empresas grandes son cercanos a uno y en tanto los estratos de las empresas pequeñas cercanos a cero. Por otro lado, en ocasiones la variabilidad en los ponderadores que induce el diseño vienen dadas por limitaciones en el marco y para reducir costos de relevamiento y ser, por tanto una variabilidad perjudicial para la precisión de los estimadores. Comentarios similares caben para los ajustes por no respuesta, si los supuestos realizados son correctos la variabilidad que conllevan se justifica en su capacidad de reducir los sesgos por este motivo. Los ajustes debidos a la calibración para lograr ajustes a totales poblacionales correctos serán beneficiosos y tendrán la capacidad de reducir los sesgos de subcobertura.

De cualquier manera, intuitivamente, a mayor variabilidad en los ponderadores se tendrá mayor variabilidad en los estimadores. Esto no es más que otra manifestación de la disyuntiva habitual entre sesgo y varianza. Los ajustes realizados a los ponderadores tiene por objetivo principal la reducción del sesgo. Luego de dichos ajustes, cabe preguntarse si no se torció desmedidamente la balanza y se está incurriendo en indeseados aumentos en la variabilidad del los estimadores.

Es por esto que en cada etapa de los ajustes de los ponderadores se debe controlar que no existan ajustes muy influyentes que aumenten la variabilidad de los ponderadores.

Un primer paso obvio consiste en verificar que los ponderadores originales sean mayores o iguales a uno. Otros controles de rutina consisten en verificar es que la suma de los ponderadores estime razonablemente el total de la población objetivo. Por otra parte, si existe calibración a totales poblacionales se debe revisar que las estimaciones obtenidas de los totales de las variables auxiliares coincidan con los totales verdaderos.

Por último, el paso más delicado consiste en evaluar la variabilidad global de los ponderadores obtenidos. Una medida global frecuentemente utilizada conocida como efecto diseño debido a la ponderación o efecto de Kish (Kish, 1965, 1992) y representa el incremento en la variabilidad de los estimadores por usar ponderadores distintos respecto al uso de el mismo ponderador para todos los caso y se define como uno más la varianza relativa de los ponderadores, w, o sea

$$deff_w = 1 + \frac{1}{n}\frac{\sum_s (w_k - \bar{w})^2}{\bar{w}^2}$$

donde $\bar{w} = n^{-1}\sum_s w_k$ es el promedio de los ponderadores.

El $deff_w$ debe ser calculado en los ponderadores resultantes de cada una de las etapas de ajuste mencionadas en este documento. En la práctica $deff_w > 1,5$ indican que los ajustes realizados deben ser revisados por la posible presencia de valores extremos que repercutan en los ponderadores finales. Por ejemplo, si $deff_w > 1,5$ en la etapa de calibración, el proceso debería repetirse utilizando algún procedimiento de calibración truncada, de forma de que los ajustes g_k puedan ser acotados a un límite inferior y superior. Si realizado lo anterior, el problema no fuera solucionado, los ponderadores finales pueden ser recortados posteriormente, aunque dejarán de cumplir las ecuaciones de calibración. Si esto ocurre, se debe revisar que tan lejos están las estimaciones de los totales de las variables auxiliares con los ponderadores recortados. En los ajustes por elegibilidad y no respuesta, en la medida que se hayan usados ajustes por grupos o celdas una opción habitual es colapsar algunas categorías que presenten ajustes muy extremos.

7. Ejemplo de aplicación en `R`

De forma de aplicar los distintos ajustes a los ponderadores originales de la muestra w_{0k} vistos en este documento, se realiza un ejemplo práctico, utilizando el software `R`. Para dicha ilustración se utilizan datos provenientes de la Encuesta Continua de Hogares (ECH) del año 2016, la cual se encuentra disponible en la pagina web del INE. Se trabaja únicamente con los casos correspondientes al departamento de Montevideo. Dichos datos (sin expandir) se utilizan como población y población marco para la realización del ejemplo.

Para este ejemplo, son elegibles para la muestra las personas entre 14 y 65 años (población objetivo, U). El objetivo es seleccionar una muestra de $n = 600$ personas bajo un diseño complejo similar al de la ECH.

Una vez seleccionada la muestra los ponderadores originales (w_{0k}) que van a depender del diseño de muestreo aplicado, son ajustados por no respuesta y posteriormente, calibrados a totales poblacionales conocidos. En cada paso de los ajustes de los ponderadores se calcula el efecto de diseño de kish para evaluar el incremento en la variabilidad de los estimadores producto del ajuste realizado.

El marco F contiene $N_F = 17777$ hogares. En F, como sucede habitualmente en la práctica, hay unidades que no son elegibles (no pertenecen a U). En este caso, son inelegibles aquellos hogares en donde no existe ningún integrante entre 14 y 65 años y los mismos ascienden a 3421 hogares. Esto implica que dentro de los 600 hogares que se seleccionen en la muestra original, se espera, que un $20\,\% = 3421/1777$ aproximadamente sean inelegibles. Cabe aclarar, que si el objetivo de la encuesta fuera obtener aproximadamente 600 hogares elegibles, el tamaño de muestra original ($n = 600$) debería ser incrementado en un 20 % aproximadamente.

7.1. Diseño muestral

Las n personas son seleccionadas bajo un diseño aleatorio, estratificado y en tres etapas de selección. Las unidades primarias de muestreo (UPM) corresponden a los segmentos censales. Dichas UPM son estratificadas en cinco estratos socioeconómicos (Bajo, Medio Bajo, Medio, Medio Alto y Alto). Las UPM, son seleccionadas bajo un diseño sistemático proporcional al tamaño utilizando como medida de tamaño (MOS) la cantidad de hogares en el segmento. En una segunda etapa, se seleccionan, dentro de cada UPM incluida en la muestra de la primera etapa, cuatro hogares (unidades secundarias de muestreo, USM) con igual probabilidad de selección bajo un diseño simple. Finalmente, dentro de cada hogar incluido en la muestra, se selecciona una persona elegible con igual probabilidad de selección.

El tamaño de muestra se asigna en los estratos de forma proporcional en base a la cantidad de hogares por estrato. Este tipo de asignación en conjunto con el método de selección utilizado para la selección de las UPM y USM busca tener un diseño aproximadamente autoponderado (mismos ponderadores originales para cada hogar). En el siguiente cuadro se presentan las cantidades de hogares en F por estrato N_h, las UPM seleccionadas en la muestra m_h y finalmente la cantidad de hogares en la muestra ($n_h = 4 \times m_h$)

h	Descripción estrato	N_h	m_h	n_h
1	Bajo	2775	23	92
2	Medio Bajo	3574	30	120
3	Medio	4710	40	160
4	Medio Alto	4461	38	152
5	Alto	2257	19	76
Total		17777	150	600

7.2. Supuestos

Se harán una serie de supuestos para emular las condiciones que se dan habitualmente en la práctica:

1. El marco de muestreo F únicamente contiene los hogares, las UPM y los estratos del diseño. Esto implica que la selección de la muestra “en oficina”, se realizará hasta el nivel del sorteo del hogar.
2. Se asume que no existen unidades con elegibilidad desconocida. Es posible determinar en campo que hogares son elegibles (hay al menos un integrante del hogar que cumple los criterios de elegibilidad de la encuesta) y cuales son inelegibles.

3. Se simula la probabilidad que una persona tiene de responder a la encuesta. Se supone que la probabilidad de responder depende del estrato del diseño (los del estrato bajo son más proclives a responder que los del estrato alto) y de la edad de los integrantes del hogar (los jóvenes son menos proclives a contestar respecto a los de mayor edad).

   La probabilidad de las personas de responder según el estrato (ϕ_h) se fijan en $\phi_1 = 0{,}95$, $\phi_2 = 0{,}90$, $\phi_3 = 0{,}90$, $\phi_4 = 0{,}85$ y $\phi_5 = 0{,}70$.

   Por otra parte, la probabilidad de responder se define a nivel de cuatro grupos de edad (ϕ_g). Los grupos son 1- 14 a 24 años, 2- 25 a 39 años, 3- 40 a 49 años y 5 - 50 a 65 años, y sus respectivas probabilidades de responder son $\phi_1 = 0{,}80$, $\phi_2 = 0{,}85$, $\phi_3 = 0{,}90$ y $\phi_4 = 0{,}95$.

   Se asume que las probabilidad de responder son independientes del estrato y de la edad de la persona, es decir, la probabilidad de responder de la persona k perteneciente al estrato h y que tiene una edad comprendida en el grupo g queda definida como $\phi_k = \phi_h \times \phi_g$.

   Para determinar que persona es respondente o no, se genera un número aleatorio entre 0 y 1 (RN) previamente a la selección de la muestra. Si RN es menor a ϕ_k se clasifica a la persona como respondente (R) y en el otro caso no respondente (NR).

   Finalmente, se asume que la no respuesta es únicamente a nivel de hogar. Es decir, si en el hogar existe al menos una persona clasificada como NR, el hogar también es clasificado como NR.

4. Se asume, al igual que sucede en la práctica, que la única información auxiliar que se tiene disponible sobre los hogares NR es la que se encuentra incluida en F (estratos y UPM).

5. Dado que en la práctica el sorteo de la persona dentro del hogar se realiza *in situ*, únicamente se dispondrá de información de los integrantes del hogar que fueron catalogados como elegibles respondentes.

6. La única información disponible a nivel de toda la población son los conteos del total de personas por edad y sexo. Dicha información es conocida únicamente para los hogares elegibles respondentes. En la práctica cabe recordar que dicha información generalmente proviene de las proyecciones de población o del último censo reciente.

7.3. La muestra

De los $n = 600$ hogares seleccionados inicialmente, 120 resultaron ser inelegibles, y dentro de los hogares elegibles, 365 son respondentes. En la siguiente tabla se presentan la cantidad de hogares elegibles, la cantidad de hogares elegibles respondentes (ER) y la tasa de respuesta desagregada por estrato

h	Descripción estrato	n	n_E	n_{ER}	Tasa respuesta
1	Bajo	92	80	62	77.5 %
2	Medio Bajo	120	96	60	62.5 %
3	Medio	160	132	93	70.5 %
4	Medio Alto	152	117	76	65.0 %
5	Alto	76	55	36	65.5 %
Total		600	480	327	68.1 %

7.4. Determinación de los ponderadores

Como primer paso se determinan los ponderadores originales de los hogares en la muestra (w_{0k}) los cuales dependen del diseño de muestreo. Debido a que se asume que no hay hogares con elegibilidad desconocida no se realiza ninguna ajuste, es decir, $w_{0k} = w_{1k}$. Posteriormente, se calculan los ponderadores ajustados por no respuesta (w_{2k}) utilizando el ajustes basados en grupos de respuesta homogéneos, en donde los grupos se definen únicamente utilizando información auxiliar contenida en F. Finalmente, los ponderadores (w_{2k}) son calibrados (ajustados) en base a la información auxiliar disponible (tramos de edad y sexo) de forma de obtener los ponderadores finales w_{3k}. Para este último paso en el ajuste de los ponderadores se evalúan las distintas estrategias de calibración vistas en este documento: postestratificación completa e incompleta (*raking*) y el *raking* truncado (limitando la variabilidad de los ajustes g_k).

7.4.1. Determinación de los ponderadores originales

Para la determinación de los ponderadores originales de los hogares en la muestra es necesario determinar las probabilidad de selección de los mismos, El cálculo de las probabilidades de selección (π_k) dependen del diseño muestral implementado y deben ser calculadas, como ya se dijo, en el momento de la selección de la muestra. El ponderador original del hogar k perteneciente al estrato del diseño h es:

$$w_{0k} = \frac{1}{\pi_k} = \frac{N_h}{n_h}$$

Por el tipo de diseño implementado los ponderadores son iguales para cada uno de los hogares pertenecientes al mismo estrato y aproximadamente iguales para toda la muestra producto de la asignación proporcional utilizada para asignar los tamaños de muestra por estrato.

h	Descripción estrato	N_h	n_h	$w_k = N_h/n_h$
1	Bajo	2775	92	30.16
2	Medio Bajo	3574	120	29.78
3	Medio	4710	160	29.44
4	Medio Alto	4461	152	29.35
5	Alto	2257	76	29.70

En ausencia de no respuesta, el ponderador original para la persona sorteada en el hogar quedaría determinado como $w_{0kp} = w_{0k} \times P_k$ donde P_k es la cantidad de personas elegibles (entre 14 y 65 años) en el hogar, en donde P_k es generalmente obtenido en campo. Debido a que existe no respuesta a nivel de hogar, el ponderador de la persona elegible respondente perteneciente al hogar k queda determinado como $w_{2kp} = w_{2k} \times P_k$, donde w_{2k} es el ponderador ajustado por no respuesta.

7.4.2. Ajuste por no respuesta

Una vez definidos los ponderadores originales w_{0k} para todos los hogares incluidos en la muestra original, se descartan los hogares de la muestra que son inelegibles, es decir, se trabaja únicamente con los hogares elegibles independientemente de que los mismos sean respondentes o no ($s_E = s_{ER} \cup s_{ENR}$).

Posteriormente, teniendo en cuenta que las tasas de respuestas obtenidas varían por estrato, y que la única información disponible es aquella que está contenida en el marco de muestreo, se decide ajustar a los ponderadores utilizando el ajuste basado en grupos de respuesta homogéneos en donde los grupos, son definidos en base a los estratos del diseño.

El ajuste por no respuesta para el hogar elegible respondente k perteneciente al estrato h queda definido como $nr_k = \sum_{s_{h,E}} w_{0k} / \sum_{s_{h,ER}} w_{0k}$, es decir, los ponderadores son ajustados por el inverso de la tasa de respuesta ponderada por estrato utilizando los ponderadores w_{0k}. En el siguiente cuadro se presentan los ajuste por no respuesta y los ponderadores ajustados por no respuesta ($w_{2k} = nr_k \times w_{0k}$).

h	$w_{0k} = \frac{N_h}{n_h}$	$\sum_{s_{h,E}} w_{0k}$	$\sum_{s_{h,ER}} w_{0k}$	$nr_k = \frac{\sum_{s_{h,E}} w_{0k}}{\sum_{s_{h,ER}} w_{0k}}$	$w_{2k} = nr_k \times nr_k$
1	30.16	5097.55	4494.29	1.134	34.21
2	29.78	6105.58	4467.50	1.367	40.70
3	29.43	8154.19	6123.00	1.332	39.20
4	29.34	6427.36	5047.97	1.273	37.37
5	29.69	4068.54	2553.97	1.593	47.31

Otra opción para la conformación de los grupos para el ajuste de la no respuesta son las UPM. El uso de las UPM como grupos de no respuesta puede provocar ajustes extremadamente altos provocando así ponderadores muy influyentes, lo que se traduce en un aumento considerable de las varianzas de los estimadores, desembocando así, en estimaciones inestables de las distintas cantidades poblacionales de interés de la encuesta.

7.4.3. Calibración de los ponderadores

El último paso para la construcción de los ponderadores es el ajuste o calibración a totales poblacionales. El objetivo es encontrar los ponderadores $w_{3kp} = g_k \times w_{2kp}$ para las personas elegibles respondentes. Se evalúan distintas formas de calibrar (calcular los ajustes g_k) utilizando la misma información auxiliar junto con su implementación en `R` utilizando en paquete `survey`.

La información auxiliar utilizada para la calibración es total de personas por cuatro tramos de edad (14 a 24 años, 25 a 39 años, 40 a 49 años y 50 a 65 años) y sexo de las personas. Dicha información en la práctica proviene del último censo o de las proyecciones de población. A su vez, como ya se dijo, dicha información (edad y sexo) debe ser conocida para los elegibles respondentes de la encuesta, e.i, estar incluida en el formulario de la encuesta.

En la siguiente tabla se presentan los conteos de la cantidad de personas por tramo de edad y sexo (N_g); y las estimaciones de los mismos ($\hat{N}_g$), obtenidos con los ponderadores de las personas ajustados por no respuesta $w_{2kp} = P_k \times w_{2k}$.

	N_g		$\hat{N}_g$	
Edad	H	M	H	M
14 a 24	3206	3430	3505	2944
25 a 39	4349	4952	4181	5107
40 a 49	2520	2923	2011	2295
50 a 65	3789	4798	3323	6487

Por otra parte, de forma de mostrar las sentencias o comandos necesarios para calibrar los ponderadores en `R` se presenta la estructura del data frame con la muestra de elegibles respondentes.

```
ST   UPM         USM  UUM  SEXO  EDAD  TRAMO.EDAD        w2p
 1  17102  2016040927   3    1    21          T1   102.63532
 1  99167  2016002881   4    2    15          T1   136.84710
 1  09026  2016007315   1    1    41          T3   102.63532
 1  11092  2016030638   3    1    15          T1   102.63532
 1  17341  2016033385   2    1    23          T1    68.42355
 1  11001  2016026431   1    1    41          T3   102.63532
```

`ST`= estrato de la encuesta (1- bajo, 2- medio bajo, 3 -medio, 4 - medio alto, 5-alto)

`UPM`=etiquetas de las unidades primarias de muestreo (UPM=segmento censal)

`USM`= etiquetas de las unidades secundarias de muestreo (USM=hogar)

`UUM`= etiquetas de las unidades últimas de muestreo (UUM=persona)

`SEXO`= sexo de la persona sorteada (1- hombre, 2- mujer)

`EDAD`= edad de la persona sorteada (en años cumplidos)

`TRAMO.EDAD`= tramo de edad (T1=14 a 24, T2=25 a 39, T3=40 a 49 y T4=50 a 65)

`w2p`= ponderadores de las personas ajustados por no respuesta

Postestratificación

El ajuste proveniente de la postestratificación para la persona k perteneciente a un postestrato cualquiera g viene dado por $g_k = N_g/\hat{N}_g$, donde N_g es el conteo poblacional del postestrato o celda de la tabla g y $\hat{N}_g = \sum_{s_{g,ER}} w_{2kp}$ es la estimación delpost estrato g utilizando los ponderadores de las personas ajustados por no respuesta.

g_k postestratificación

	H	M
14 a 24	0.915	1.165
25 a 39	1.040	0.970
40 a 49	1.253	1.273
50 a 65	1.140	0.740

$deff_w = 1{,}291$

Observando la tabla anterior, se puede apreciar que, ajustes g_k menores a 1, indican que existe una sub estimación para la cantidad de personas en la celda. De forma análoga ajustes g_k mayores a 1 indican que existe una sobre estimación. Lo anterior se puede deber al simple hecho de que se está trabajando con una muestra y/o también por la presencia de no respuesta (en donde la probabilidad de una persona de responder depende de la edad).

Para realizar la postestratificación en `R` se utiliza la función `postStratify`. La función toma tres argumentos: `design` (el objeto donde se encuentra guardado el diseño), `strata` (variable en la muestra que identifica la variable utilizada para la postestratificación y `population` (tabla o data frame con los conteos poblacionales de los postestratos)

```
post=postStratify(design=p,
                  strata=~PS,
                  population=N.P)
```

La función `postStratify` acepta una única variable (llamada PS en este caso) para definir los postestratos (celdas). Es decir, en este caso, la variable PS es una combinación de las variables `TRAMO.EDAD` y `SEXO`. El objeto `p` es creado utilizando la función `svydesign` la cual tiene información acerca del diseño muestral (ponderadores, estratos y UPM). En este caso,

```
p=svydesign(id=~UPM,strata=~ST,weights=~w2p, data=R)
```

donde `w2p` es una variable con los ponderadores de las personas ajustados por no respuesta $w_{2kp} = P_k \times w_{2k}$ y `R` es el data frame en donde se encuentra guardada la muestra de respondentes.

Una vez realizada la postestratificación los ponderadores finales w_{3kp} pueden ser extraídos utilizando la función `weights(post)`.

Raking - Postestratificación incompleta

Otra alternativa a la postestratificación es el *raking*, el cual puede ser implementado utilizando más de una variable auxiliar. En el ejemplo, los postestratos son construidos utilizando la clasificación cruzada del `TRAMO.EDAD` y `SEXO`, lo cual implica que los conteos poblacionales para cada celda (N_g) deben ser conocidos, lo cual, en la práctica, puede ser que no se encuentren disponibles si las variables provienen de distintas fuentes de información. Adicionalmente, se requiere que el tamaño de muestra en la celda (post-estrato) sea lo suficientemente grande para que $\hat{N}_g$ no sea inestable. En cambio, para el *raking* solo es necesario las marginales de las variables `TRAMO.EDAD` y `SEXO`. Al no ser necesario realizar la clasificación cruzada, permite utilizar mayor cantidad de variables que sean importantes como predictores de las variables de interés, para trabajar la no respuesta y/o problemas de cobertura en el marco de muestreo.

g_k raking		
	H	M
14 a 24	1.069	0.981
25 a 39	1.049	0.963
40 a 49	1.322	1.213
50 a 65	0.925	0.850
$deff_w = 1{,}275$		

El raking en R se puede realizar utilizando la función `calibrate` o `rake`. En el siguiente código se presenta el *raking* utilizando la función `rake`

```
raking=rake(design=p,
            sample.margins=list(~TRAMO.EDAD,~SEXO),
            population.margins=list(POP.TRAMO,POP.SEXO))
```

La función `rake` se llama de forma iterativa a la función `postStratify`, postestratificando una variable a la vez hasta llegar a la convergencia.

Al igual que la función `postStratify` utiliza tres argumentos: el objeto con el diseño (`p`), `sample.margins` en donde se ponen como lista el nombre de las variables de las marginales del raking y finalmente `population.margins` la lista de las tablas o *data frames* con las frecuencias poblacionales de las marginales por separado de cada una de las variables utilizadas para el ajuste.

Distancia lineal

Otra forma de calcular los ajustes g_k es por medio de la calibración utilizando una función de distancia lineal. En este caso, dado que las variables son categóricas, la calibración es equivalente a un estimador de regresión lineal, donde el modelo que lo asiste al estimador es un modelo de efectos fijos (tramo de edad y sexo).

g_k lineal		
	H	M
14 a 24	1.068	0.982
25 a 39	1.049	0.963
40 a 49	1.310	1.224
50 a 65	0.932	0.846
$deff_w = 1{,}274$		

Para realizar la calibración utilizando la distancia lineal se debe utilizar la funci 'on `calibrate`.

```
lineal=calibrate(design=p,
                 formula=~as.factor(TRAMO.EDAD)+as.factor(SEXO),
                 population=CONTEOS,
                 calfun='linear')
```

en donde `CONTEOS` es un vector con el total de la población (N) y con los conteos de las categorías de cada una de las variables utilizadas en la calibración, las cuales deben seguir el orden de la fórmula y en donde se excluye la primer categoría de cada una de las variables, las cuál son tomadas como de referencia.

La función `calibrate` permite realizar la calibración utilizando distintas funciones de distancias (lineal, *raking* y *logit*). A su vez, permite realizar la calibración truncada, es decir, limitar los ajustes g_k fijando los límites inferior y superior de dichos ajustes. Esto implica que el *raking* calculado anteriormente pueda ser modificando limitando los g_k de forma de evitar ajustes extremos que luego produzcan ponderadores finales grandes w_{3k} que condicionen el análisis.

La función `calibrate` acepta varios argumentos, algunos son:

`design=` objeto con el diseño muestral

`formula=`modelo con las variables auxiliares para la calibración

`population=` Variable (vector) con los totales poblacionales

`calfun=` función de distancia utilizada para la calibración (Lineal, Raking, logit).

Raking truncado

Cuando se calibran los ponderadores bajo las tres estrategias vistas anteriormente puede ocurrir:

- Obtener ponderadores negativos (con la distancia lineal). Esto es poco probable que pase si las probabilidad de inclusión son pequeñas y la cantidad de variables auxiliares no son demasiadas para lo que la muestra puede soportar.
- Obtener ponderadores muy grande. Esto puede ocurrir si las tasas de repuesta o elegibilidad obtenidas en algunos dominios son pequeñas y si la cantidad de variables de auxiliares son excesivas para lo que la muestra puede soportar.

Lo anterior se puede prevenir o restringir añadiendo una restricción a los ajustes g_k, es decir, utilizando la calibración truncada. Por ejemplo, en el *raking* anterior, se puede realizar pero restringiendo a los ajustes, fijando por ejemplo como límite inferior a los ajustes $g_k = 0{,}86$ y como límite superior $g_k = 1{,}3$.

g_k raking truncado

	H	M
14 a 24	1.082	0.966
25 a 39	1.064	0.950
40 a 49	1.300	1.234
50 a 65	0.905	0.866

$deff_w = 1{,}275$

Como ya se dijó la calibración truncada en `R` se realiza utilizando la función `calibrate` y usando el argumento `bounds`

```
raking.trunc=calibrate(design=p,
                       formula=~as.factor(TRAMO.EDAD)+as.factor(SEXO),
                       population=CONTEOS,
                       calfun='raking',
                       bounds=c(0.86,1.3))
```

En algunos casos puede ocurrir que la convergencia en la calibración no se alcance con los límites fijados mínimos y máximos fijados para los g_k. En estos casos los pasos a seguir (jerárquicos o no) son:

- Si la convergencia no es alcanzada con los valores por defecto, aumentar el número de iteraciones `maxit`.
- Aumentar el nivel de tolerancia (`epsilon`)
- Forzar el cálculo de los ponderadores (`Force=TRUE`), es decir, que devuelva unos ajustes g_k aunque la solución no haya sido alcanzada.
- Si `Force=TRUE` se debe revisar que tan lejos están de cumplirse las ecuaciones de calibración.

GREG

Finalmente, se calibran los ponderadores utilizando el GREG con las mismas variables auxiliares utilizadas en el lineal pero agregando una variable cuantitativa (la edad de las personas en años cumplidos). Para ello, en el vector de `CONTEOS` se debe agregar el total de años en la población. Esto implica que los ponderadores calibrados obtenidos van a estimar también, sin error, la edad promedio en la población.

```
GREG=calibrate(p,
               ~as.factor(TRAMO.EDAD)+as.factor(SEXO)+EDAD,
               CONTEOS,
               calfun='linear')
```

En la siguiente tabla se presentan los ajustes g_k promedios por celda, dado que al agregar una variable numérica en la calibración habrá un ajuste para cada una de las edades simples dentro de la celda.

g_k GREG		
	H	M
14 a 24	1.071	0.978
25 a 39	1.047	0.970
40 a 49	1.317	1.220
50 a 65	0.933	0.861
$deff_w = 1{,}270$		

7.4.4. Recorte y distribución de los ponderadores finales

En la práctica aunque se tenga extremadamente cuidado en los ajustes a los ponderadores realizados en cada una de las etapas, puede igual ocurrir que existan ponderadores extremadamente grandes o chicos. En estos casos se pueden fijar límites mínimos y máximos para los ponderadores finales w_{3k}. Nótese que no se está hablando de limitar los ajustes g_k sino limitar los ponderadores finales w_{3k} una vez concluidos todos los ajustes mencionados anteriomente.

La idea es fijar un límite inferior (L) y un límite superior (U), los cuales son arbitrarios y dependen de la distribución de los ponderadores finales obtenidos. Los pasos a seguir son:

- Definir L y U
- Cualquier ponderador fuera de los límites es recortado, es decir,

$$w_{trim,k} = \begin{cases} U & \text{si } w_{3k} \geq U \\ w_{3k} & \text{si } L < w_{3k} < U \\ L & \text{si } w_{3k} \leq L \end{cases}$$

- Determinar la suma $K = \sum_{s_{ER}} |w_{3k} - w_{trim,k}|$, es decir, la cantidad de ponderadores perdida por el recorte.
- Distribuir K entre los ponderadores no recortados.
- Repetir los pasos anteriores hasta que no haya más ponderadores por fuera de los límites.

El recorte o *trimming* de los ponderadores en `R` puede ser llevado a cabo utilizando la siguiente función

```
trimWeights(GREG,
            upper=U,
            lower=L,
            strict=TRUE)
```

Si `strict=TRUE` la función se llama así misma de forma recursiva hasta que los limites son satisfechos.

8. Síntesis y recomendaciones

Este documento persigue un propósito de divulgación. Está dirigido a usuarios no especializados de encuestas de hogares y personas que necesitan saber para qué y cómo se construyen los ponderadores. Se se intentó una presentación sencilla y que estuviera apoyada por un ejemplo concreto. En el ejemplo se utilizó el *package survey* del *sofware* R

Los datos provenientes de encuestas por muestreo se obtienen para un subconjunto de la población de interés. La elección de dicho subconjunto se realiza de forma aleatoria y, generalmente, con distintas probabilidades de inclusión. Esto implica que, para lograr estimadores insesgados o aproximadamente insesgados, de totales y medias poblacionales, los datos deben usarse de manera ponderada. Incluso en diseños autoponderados, donde el ponderador es el mismo para todos los individuos, existirán razones para el uso de ponderadores distintos. En última instancia, no ponderar implica optar por el mismo ponderador para todos los casos, por lo tanto será mejor utilizar ponderadores elegidos de una manera consciente y justificadamente más eficaz. Esto siempre es posible en la medida que exista alguna información auxiliar relevante que pueda ser usada para mejorar los estimadores.

Desde la construcción de los ponderadores base, que se deducen del diseño muestral, hasta la obtención de los ponderadores que se usarán finalmente para las estimaciones, se transita por distintas etapas. En cada una de estas etapas se aborda un problema distinto y los ponderadores de partida son ajustados dando lugar a los ponderadores de iniciales de la etapa siguiente.

Una vez terminado el trabajo de campo de la encuesta, existirán algunos elementos seleccionados de los cuales no se pudo determinar su pertenencia a la población objetivo, y otros, que se saben elegibles, pero de los que no se obtiene la información deseada. Esto da lugar a los ajustes por elegibilidad desconocida y por no respuesta. Para que los procedimientos de ajuste de estas etapas sean exitosos es imprescindible contar con un registro minucioso y suficientemente desagregado de los motivos por los cuales una unidad seleccionada no termina brindando información. En general, estos aspectos dependen mucho del tipo de encuesta y la forma de recolectar la información que se utilice. Debe planificarse desde un principio como van a clasificarse los casos de no y que información adicional de puede recoger en estos casos, ya que será parte de la información auxiliar que se utilice a la hora de realizar los ajustes. Adicionalmente, las tasas de respuesta y sus distintas causales son indicadores fundamentales, pero poco usados, de la calidad de una encuesta.

De los dos ajustes anteriores, la no respuesta suelen ser de mayor importancia. La no respuesta nunca es MAR. En los casos en que el supuesto de no respuesta MCAR es sostenibles debe plantearse un modelo que permita su tratamiento ya que bajo este supuesto, teóricamente, es posible una solución satisfactoria. Si bien la validez de los ajustes queda sujeta a la validez del modelo, y esto se aleja de las prácticas comunes en el tipo de encuestas que se analizaron, no hay otra forma posible para su tratamiento.

En la última etapa los ponderadores son sometidos a un proceso de calibración. Siempre existe algún tipo de información auxiliar que es conocido a nivel poblacional y que debe ser usada para mejorar la estimaciones. Como se analizó, la calibración es mecanismo poderoso y multipropósito que permite incorporar distintos tipos de información auxiliar. Nuevamente, esto implica un ajuste a los ponderadores. Si bien los procedimientos de calibración pueden pensarse como no basados en un modelo, en muchos casos el ajuste al que se llega puede verse como el resultado de plantear un modelo superpoblacional que describa a la población de interés. En estos casos es, es útil reflexionar sobre el modelo a usar y vale la pena dedicar tiempo a los procedimientos de selección de variables, diagnóstico y bondad de ajuste.

El objetivo principal de la ponderación es la reducción del sesgo. Como es habitual, existe la disyuntiva entre sesgo y varianza, ponderadores excesivamente variables y/o extremos pueden aumentar injustificadamente, la varianza de los estimadores. Por tanto, una vez calculados los ponderadores finales, debe analizarse la necesidad de suavizar valores extremos y reducir su variabilidad general.

Si bien son necesarios, los procedimientos de ponderación suelen ser engorrosos. La disponibilidad creciente de *software* que permite implementar los procedimientos considerados, junto con la disponibilidad, también creciente, de información auxiliar determinan que los ajustes de los ponderadores sean cada vez más sofisticados y que su uso esté cada vez más generalizado.

Un tema no tratado, aunque de mucha importancia a la hora de producir estimaciones es la necesidad de acompañarlas de medidas que reflejen su precisión, usualmente el desvío o el coeficiente de variación. Si bien es un tema complejo, dichas estimaciones deben tomar en cuenta que los ponderadores dependen de la muestra observada y por lo tanto tienen variabilidad muestral y que esta afecta, a su vez, la variabilidad las estimaciones. En este sentido, vale la pena señalar que una forma práctica, si no la única, de atacar este problema es con el uso de técnicas de remuestreo.

Otro tema que queda en el tintero es el uso de ponderadores cuando los datos provenientes de una encuesta son usados para estimar los parámetros de un modelo analítico. Si bien parece obvio que es necesario su uso, no es la práctica más común, dadas las complicaciones teóricas y prácticas que ello origina. De hecho, es relativamente reciente el desarrollo de *software* que permita su incorporación. La tendencia actual es a no ignorar el problema y por lo menos analizar que tan perjudicial puede ser no utilizar los ponderadores en la estimación de modelos (Bollen, et al. 2016; Fuller, 2009).

9. Bibliografía

Bollen, K.A., Biemer, P.P., Karr, A.F., Tueller, S. y Berzofsky M.E. (2016) Are Survey Weights Needed? A Review of Diagnostic Tests in Regression Analysis. Annual Review of Statistics and Its Application. 3:375-92.

Deville, J.C., and Särndal, C.E. (1992). Calibration estimators in survey sampling. Journal of the American Statistical Association, 87

Deville, J.C., Särndal, C.E. and Sautory, O. (1993). Generalized raking procedures in survey sampling. Journal of the American Statistical Association, 88

Fuller, W.A. (2009) Sampling Statistics, John Wiley & Sons.

Fuller, W.A. (2002). Regression estimation for survey samples. Survey Methodology, 28, 523.

Groves, R. M. (2006) Nonresponse rates and nonresponse bias in household surveys, Public Opinion Quarterly. 70, No. 5, 646-675

Kish L. (1965). Survey Sampling. John Wiley & Sons, Inc., New York

Kish L. (1992). Weighting for unequal pi. Journal of Official Statistics 8(2): 183-200

Little R.J.A., Rubin D.B. (2002). Statistical Analysis with Missing Data. John Wiley & Sons, Inc., New Jersey

Lumley, T. (2010). Complex Surveys. John Wiley & Sons, Inc., New York

Lumley, T. (2017) ”survey: analysis of complex survey samples”. R package version 3.32

Lumley, T. (2004) Analysis of complex survey samples. Journal of Statistical Software 9(1): 1-19

Olson, K. (2006) Survey participation, nonresponse bias, measurement error bias, and total bias. Public Opinion Quarterly. 70, No. 5, 737-75

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

Rosenbaum P., Rubin D.B. (1983). The central role of the propensity score in observational studies for causal effects. Biometrika 70:41-55

Särndal, C.E. (2007). The calibration approach in survey theory and practice. Survey Methodology 33(2):99-119.

Särndal, C.E., and Lundström, S. (2005). Estimation in Surveys with Nonresponse. New York: John Wiley & Sons, Inc.

Särndal, C.E., Swensson, B., Wretman, J.H. (1992) Model Assisted Survey Sampling. New York: Springer-Verlag

Valliant, R., Dever, J.A., Kreuter, F. (2013) Practical Tools for Designing and Weighting Survey Samples. New York: Springer-Verlag