

Recibido: 18 de mayo de 2020

Aceptado: 04 de febrero de 2021

**estudios
sociológicos**
de El Colegio de México

Primero en línea: 16 de agosto de 2021

2021, 39(117), sept.-dic., 773-792

Artículo

Investigar la incorporación de hijos de migrantes a través de Respondent Driven Sampling (RDS)

*Researching the Incorporation of Children
of Migrants through Respondent
Driven Sampling (RDS)*

Pablo Sebastián Gómez

Facultad de Ciencias Sociales

Consejo Nacional de Investigaciones Científicas y Técnicas

Universidad Nacional de Córdoba. Universidad Siglo 21

Córdoba, Argentina

<https://orcid.org/0000-0002-8364-7592>

pablosgomez@unc.edu.ar

Maximiliano Iglesias

Instituto de Estadística y Demografía

Facultad de Ciencias Económicas

Universidad Nacional de Córdoba

Córdoba, Argentina

<https://orcid.org/0000-0002-8957-8321>

miglesias@unc.edu.ar



Esta obra está bajo una licencia de Creative Commons.

Reconocimiento-No Comercial-Sin Obra Derivada 4.0 Internacional.

Resumen: En el campo de los estudios migratorios uno de los temas de investigación de larga tradición está constituido por las “segundas generaciones” o “hijos de migrantes”. En este artículo presentamos los problemas metodológicos derivados de la generación de datos primarios sobre el tópico: en primer lugar, la población es de tamaño pequeño y las encuestas tradicionales no capturan adecuadamente la información. En segundo lugar, la falta de un marco muestral adecuado, dado que no contamos con una lista exhaustiva de los miembros. Aquí desarrollamos una posible alternativa para solucionar estos problemas: el muestreo Respondent Driven Sampling (RDS) para el estudio de los procesos de incorporación de hijos de migrantes en general e hijos de migrantes sur-surens particular.

Palabras clave: hijos de migrantes; sur global; respondent driven sampling (RDS); Argentina.

Abstract: *In the field of migration studies, one long-standing research topic is the study of the “second generations” or “children of migrants.” In this paper, we present the methodological problems derived from the generation of primary data on the issue. First, the population is small and traditional surveys do not capture the information. Second, since we do not have an exhaustive list of member, the sampling frame is inadequate. Here we develop a viable alternative to solve these problems: Respondent-Driven Sampling (RDS) for the study of the processes of incorporating the children of migrants in general and the children of South-South migrants in particular.*

Keywords: *children of migrants; global south; respondent-driven sampling (RDS); Argentina.*

Los procesos de incorporación de los migrantes y de las “segundas generaciones” o “hijos de migrantes” es un campo de investigación con una extensa tradición e implicancias en términos del discurso público y político, fundamentalmente en países del norte global (Alba, & Nee, 2003;

Bohrt, & Itzigsohn, 2015; Itzigsohn, 2009; Levitt, & Waters, 2006; Portes, Aparicio, Gómez, & Haller, 2016; Portes, & Rumbaut, 2006; Portes, & Zhou, 1993). A pesar de esta extensa tradición, las direcciones, magnitudes e implicancias tanto demográficas como culturales y políticas no están claras. Diferentes contextos llevan a que, en Estados Unidos, por ejemplo, se hable de “segunda generación”, “generación 1.5” (Portes, & Rumbaut, 2006) e incluso “tercera generación”; sin embargo, en otros contextos, como el francés, se señala lo problemático de estas categorías (Simon, 2015).

En ese contexto, queremos dirigir nuestra atención a lo que ocurre en un país del sur global como Argentina en general y de manera específica en la Ciudad de Córdoba. El contexto actual de estudio de los mecanismos de incorporación de los hijos de migrantes / “segundas generaciones” se produce cuando existe cierto consenso en reconocer la importancia de las migraciones sur-sur y las particularidades que las distinguen en relación con los flujos norte-sur. Tanto en la composición de clase de los mismos como en los contextos de recepción (Gómez, 2019; Hujo, & Piper, 2010).

En esta nota de investigación presentamos los aspectos metodológicos de un trabajo de investigación en desarrollo en la Ciudad de Córdoba, Argentina, cuyo objetivo es analizar los procesos de incorporación de los hijos de migrantes bolivianos y peruanos. Específicamente, presentamos los problemas metodológicos derivados de la generación de datos primarios sobre el tópico y el desarrollo de la técnica Responden Driven Sampling (RDS).

La generación de datos primarios

Cuando buscamos generar datos primarios sobre hijos de migrantes o “segunda generación” emerge una serie de complejidades que dificultan la utilización de muestreos tradicionales: en *primer lugar*, la falta de un marco muestral adecuado, es decir, una lista exhaustiva de todos los miembros de la población desde la cual generar nuestra muestra. En *segundo lugar*, la propia construcción de un marco muestral es compleja ya que la población suele ser de un tamaño tan pequeño que capturarla por medio de una encuesta entre la población general sería altamente prohibitivo en

terminos de costos. En *tercer lugar*, la falta de información completa y la baja calidad de la disponible (en los datos secundarios) no permiten definir de manera adecuada la magnitud y la sectorización de este grupo, por lo que se dificulta su acceso mediante los métodos de muestreo de probabilidad estándar. Desde nuestra perspectiva, estos grupos poblacionales pueden ser tratados, entonces, como “*hidden population*” (poblaciones escondidas o “*hard-to-reach*”) en términos estadísticos.

Las técnicas más frecuentes para el abordaje muestral de este tipo de poblaciones son el muestreo por bola de nieve (Goodman, 1961), el enfoque de informante clave (Deaux, & Callaghan, 1985) y el muestreo dirigido por mapas etnográficos (Watters, & Biernacki, 1989). Sin embargo, las principales críticas a este tipo de muestreo radican en el sesgo resultante y la validez de los resultados obtenidos, ya que no se trata de un muestreo aleatorio.

Ante este contexto, la técnica de muestreo Respondent-Driven Sampling (en adelante RDS) (Goel, & Salganik, 2010; Heckathorn, 1997; Salganik, & Heckathorn, 2004; Wejnert, & Heckathorn, 2011) tiene como objetivo superar las deficiencias metodológicas mencionadas utilizando elementos conceptuales tanto de la teoría de redes como de la cadena de Markov, junto con un sistema de incentivos estructurado que permite controlar los problemas de sobrerepresentación por referencia. El RDS supera estos problemas al aplicar un modelo matemático que pondera la muestra para compensar el hecho de que no fue obtenida de manera aleatoria (Salganik, & Heckathorn, 2004).

El Respondent-Driven Sampling (RDS) tiene una relativa extensión en diferentes campos, a saber, y sin pretensión de exhaustividad, como el estudio de los músicos de *jazz* en Estados Unidos (Heckathorn, & Jeffri, 2001), el estudio de adicciones (Abdul-Quader; Heckathorn; McKnight, *et al.*, 2006), temáticas de género (Ramirez-Valles; Garcia, Campbell, *et al.*, 2008) e incluso estudios sobre temáticas migratorias (Qiu; Yang; Ma, *et al.*, 2012; Tyldum, & Johnston, 2014).

El muestreo RDS “accede” a lo que consideramos población oculta mediante las redes sociales de sus miembros, empleando diversas bolas de nieve o muestreos por “cadenas de referencia”. En ese sentido, el estudio comienza por seleccionar un grupo de personas consideradas como “semillas”. Estas semillas reclutan a sus contactos, amigos o familiares que califican para ser incluidos en el estudio bajo la forma de primera

“onda”. La primera onda recluta posteriormente a la “segunda onda”, ésta a la “tercera onda” y así sucesivamente.

Aspectos operacionales

La elección de las semillas (“seeds”) o nodos iniciales se determina mediante un criterio basado en la afinidad entre pares. Después de que una persona completa la encuesta se le solicita que reclute a un número específico de personas para ser encuestadas, y éstas subsecuentemente hacen lo mismo hasta llegar al tamaño óptimo de la muestra. En este sentido, la semilla puede no pertenecer a la población objetivo, pero debe garantizar el contacto entre el encuestador y las personas pertenecientes a esta población. Cada eslabón de la cadena (a excepción de la inicial) desempeña entonces un doble rol, el de reclutador y reclutado. Esto permite un mejor acceso debido a la identificación recíproca entre las partes. Dado que la composición del muestreo RDS no depende de las características de las semillas, no es necesario que las mismas sean seleccionadas aleatoriamente. Sin embargo, contar con diversas y heterogéneas semillas incrementa la eficacia del RDS debido a la velocidad a la que la composición de la muestra se vuelve independiente de las características de los primeros encuestados.

Las cadenas de reclutamiento crecen sólo si las semillas tienen un rol activo, por lo tanto, es importante que sean actores con altos niveles de contactos sociales y estén comprometidos con el estudio. En los muestreos tradicionales por bola de nieve a esas semillas de “voluntarios” se les considera una fuente de sesgo. Por el contrario, en el RDS las características previas de las semillas pierden relevancia como fuente de sesgo en la medida en que éstas logren generar cadenas de reclutamiento expansivas. Su alta centralidad en términos de individuos con múltiples conexiones no aumenta el sesgo, sino que lo reduce al acelerar los procesos de reclutamiento. Las semillas, además, no son reclutadas aleatoriamente, de otra manera no estaríamos en presencia de un tipo de población “escondida”, sino que, por el contrario, son personas con las cuales ya tuvimos acceso previamente.

Los encuestados, por su parte, reclutan a sus conocidos, amigos y/o familiares, que generalmente tienden a mostrar perfiles sociodemográficos

y económicos similares (en términos de ingresos, educación, religión, etc.). Este principio de “homofilia” produce que el reclutamiento no sea aleatorio sino conformado por la red social que conecta al encuestado con la población objetivo (donde encuestados con mayor densidad están sobrerepresentados y aquellos con menor densidad subrepresentados). Por consiguiente, la muestra refleja estos patrones.

En ese sentido, el RDS utiliza el modelo de las cadenas de Markov para aproximar este proceso. El mismo se basa en dos elementos (Hekkathorn, 2002): en primer lugar, si las cadenas de reclutamiento son lo suficientemente grandes, el equilibrio se alcanza y la composición de la muestra es independiente de las semillas iniciales. En segundo lugar, la información recolectada durante el proceso de muestreo puede utilizarse para medir el sesgo.

La técnica considera a este mecanismo de selección un *proceso de Markov de primer orden*. Las cadenas de Markov tienen la propiedad de que la probabilidad de que se asuma un estado depende solamente de su estado inmediatamente anterior en el proceso. En consecuencia, el sistema de selección asume la propiedad de que las características de cada informante $X_{(i),j}$ sólo dependen de las características del informante inmediatamente anterior $X_{(i),j-1}$, es decir, de su reclutador, pero no de quien reclutó a este último. El primer subíndice $i=1, \dots, p$ hace referencia a la variable en estudio, mientras que el segundo $j=1, \dots, n$ indica la posición del participante en el secuencia de eslabones.

En consecuencia, la probabilidad de que $X_{(i),j} = s$ dependa del estado inmediatamente anterior en la cadena $X_{(i),j-1}$, al ser homogéneas en cada eslabón es: $P(X_{(i),j} = s | X_{(i),j-1} = t)$

Esta perspectiva posibilita el cálculo de las “probabilidades de transición” (p_{st}) entre los “m” posibles estados o categorías correspondientes a cada una de las características en estudio y la construcción de las “matrices de transición” (T).

La cadena está conformada por un proceso, en donde en cada eslabón existe la posibilidad de que ocurra alguno de los “m” sucesos mutuamente excluyentes al admitir la vinculación entre “s” y “t” mediante la siguiente notación:

$$p_{st} = P(X_{(i),j} = s | X_{(i),j-1} = t)$$

Los valores que asumen p_{st} se denominan “*probabilidades de transición*”. En donde $s, t = 1, 2, \dots, m$. y $p_{st} > 0$. A la vez, se cumple que para categoría “s” se satisface la condición $\sum_t p_{st} = 1$

Por su parte, la combinación de los valores que asumen las p_{st} configuran una matriz de dimensión $m \times m$ denominada “*matrices de transición*” (T).

$$T = \begin{bmatrix} p_{11} & p_{12} & \dots & p_{1m} \\ p_{21} & p_{22} & \dots & p_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ p_{m1} & p_{m2} & \dots & p_{mm} \end{bmatrix}$$

Estas matrices permiten posteriormente determinar matemáticamente la muestra teórica de equilibrio y los niveles de “tolerancia” para la validez estadística.

Suponiendo un proceso $X_{(i)}$ que en cada etapa sólo puede asumir alguno de los dos estados mutuamente excluyentes “s” y “t”, resulta una matriz de transición $T_{2 \times 2}$. Los valores de la muestra teórica de equilibrio (P^E) satisface el siguiente sistema.

$$P_s^E + P_t^E = 1$$

$$P_s^E = p_{ss}P_s^E + p_{st}P_t^E$$

En donde P_s^E y P_t^E son los valores de las muestras teóricas de equilibrio para los grupos “s” y “t”, respectivamente.

En el caso, del grupo correspondiente a la categoría o estado “s”, se satisface $P_s^E = p_{ss}P_s^E + p_{st}(1 - P_s^E)$ o de manera equivalente $P_s^E = \frac{p_{st}}{(1 - p_{ss} + p_{st})}$. En forma análoga, se despeja la muestra de equilibrio para “t” (P_t^E) al resultar $P_t^E = \frac{1 - p_{ss}}{(1 - p_{ss} + p_{st})}$

Se detalla la obtención de los valores de equilibrio:

$$P_s^E + P_t^E = 1 \quad (1)$$

$$P_s^E = p_{ss}P_s^E + p_{st}P_t^E \quad (2)$$

En (1) despejamos P_t^E al resultar:

$$P_t^E = 1 - P_s^E \quad (3)$$

Reemplazado a P_t^E en el segundo miembro de la ecuación (2) por su equivalente obtenido en (3), tenemos:

$$\Rightarrow P_s^E = p_{ss}P_s^E + p_{st}(1 - P_s^E)$$

$$P_s^E = p_{ss}P_s^E + p_{st} - p_{st}P_s^E$$

$$P_s^E = (p_{ss} - p_{st})P_s^E + p_{st}$$

$$P_s^E - [1 - (p_{ss} - p_{st})] = p_{st}$$

$$P_s^E - [1 - p_{ss} + p_{st}] = p_{st}$$

$$\therefore P_s^E = \frac{p_{st}}{1 - p_{ss} + p_{st}} \quad (4)$$

La muestra de equilibrio para “t” (P_t^E) se puede obtener a partir de las ecuaciones (1) y (2), por lo que se obtiene:

$$p_{st}P_t^E = P_s^E - p_{ss}P_s^E$$

$$\Rightarrow p_{st}P_t^E = (1 - P_s^E) - p_{ss}(1 - P_s^E)$$

$$p_{st}P_t^E = (1 - P_s^E) - p_{ss}(1 - P_s^E)$$

$$p_{st}P_t^E = 1 - P_s^E - p_{ss} + p_{ss}P_s^E$$

$$p_{st}P_t^E = 1 - p_{ss} - (1 - p_{ss})P_s^E$$

$$p_{st}P_t^E + (1 - p_{ss})P_t^E = 1 - p_{ss}$$

$$[p_{st} + (1 - p_{ss})]P_t^E = 1 - p_{ss}$$

$$[1 - p_{ss} + p_{st}]P_t^E = 1 - p_{ss}$$

$$\therefore P_t^E = \frac{1 - p_{ss}}{1 + p_{st} - p_{ss}} \quad (5)$$

En forma alternativa se puede alcanzar la ecuación (5), reemplazando en la ecuación (1) el valor de la muestra de equilibrio de “s” (P_s^E) obtenida en (4).

$$\Rightarrow \left(\frac{p_{st}}{1 - p_{ss} + p_{st}} \right) + P_t^E = 1$$

$$P_t^E = 1 - \left(\frac{p_{st}}{1 - p_{ss} + p_{st}} \right)$$

$$P_t^E = \left(\frac{1 - p_{ss} + p_{st} - p_{st}}{1 - p_{ss} + p_{st}} \right)$$

$$P_t^E = \left(\frac{1 - p_{ss}}{1 - p_{ss} + p_{st}} \right)$$

Mientras que, $p_{ss} = P(X_{(i),j} = s | X_{(i),j-1} = s)$ y $p_{st} = P(X_{(i),j} = s | X_{(i),j-1} = t)$ indican las probabilidades de transición de que el encuestado cuya característica es “s” haya sido precedido por un encuestado con característica “s” y “t”, respectivamente.

El nivel de tolerancia (ε_s) indica la diferencia entre el valor de la muestra en equilibrio y la distribución real de la muestra (observada). Esto es

$$\varepsilon_s = |P_s^E - P_s|$$

Las distribuciones muestrales de “s” y “t” se definen como

$P_s = \frac{n_s}{n}$ y $P_t = \frac{n_t}{n}$ donde bajo los supuestos del proceso $X_{(i)}$ se cumple, además, que $n = n_s + n_t$.

Nivel de saturación de la muestra

El nivel saturación del muestreo se encuentra a partir de las ondas necesarias para alcanzar estabilidad en las proporciones entre las categorías de las características definidas para el presente análisis. Las categorías que componen a cada una de esta serie de características deben ser mutuamente excluyentes y permitir la conformación de grupos en la población que sean relevantes para la investigación (Mantecón; Montse; Calafat, *et al.*, 2008).

En relación con esto último, se seleccionarán para la delimitación de los grupos y el posterior análisis de saturación y “tolerancia” muestral las siguientes tres variables dicotómicas:

$X_{(1)}$: “Género”, en donde $X_{(1)} = 1$ “Masculino” y $X_{(1)} = 2$ “Femenino”.

$X_{(2)}$: “Grupo etario”, en donde $X_{(2)} = 1$ “Hasta 25 años” y $X_{(2)} = 2$ “Más de 25 años”.

$X_{(3)}$: “Generación migrante”, en donde $X_{(3)} = 1$ “Generación 1.5” y $X_{(3)} = 2$ “Generación 2.0”.

La variable “Generación migrante” hace referencia a aquellos hijos de migrantes “Sur-Sur” que no nacieron en el país, pero migraron antes de los 11 años (“Generación 1.5”) y a aquellos hijos de migrantes que sí nacieron en el país (“Generación 2.0”). En ese sentido, el muestro RDS posee ciertos requisitos particulares de los datos, ya que no sólo necesitamos información sobre las variables de interés sino también dos elementos adicionales de información que sirven para proporcionar el marco de muestreo a partir del cual se calculan los pesos posteriores:

- El reclutamiento entre grupos (por ejemplo, en nuestro estudio, el reclutamiento proporcional de hijos de migrantes varones realizados por hijos de migrante mujer y el reclutamiento de hijos de migrantes mujeres realizado por hijos de migrantes varones).
- El tamaño medio estimado de la red (por ejemplo, a cuántas personas conoce que sean hijos de migrante o que llegaron antes de los 11 años a la Argentina).

La razón por la cual debemos hacer un seguimiento de quién reclutó a quién se halla en que podamos calcular estas proporciones de reclutamiento entre grupos. Preguntamos, asimismo, el tamaño de la red personal de cada individuo para poder calcular el grado medio estimado.

En cada una de las ondas de relevamiento se medirá las proporciones de estos atributos para observar sus valores muestrales. Como señala Mantecón *et al.* (2008), la cantidad de ondas es variable y depende de la investigación. Los autores mencionados, por ejemplo, llegaron a 12 ondas hasta estabilizar la muestra. A pesar de que existe la posibilidad de que la muestra logre estabilidad con rapidez, establecemos como objetivo inicial las 12 ondas o más, tal cual sugiere la literatura (Heckathorn, 2002).

Sistema de reclutamiento

El sistema de incentivos dual es característico de este enfoque y fundamental para lograr el tamaño muestral necesario en el tiempo de relevamiento estimado. En ese sentido, los incentivos funcionan como el sistema de reclutamiento.

Este régimen de incentivos se basa en una doble “recompensa” para cada individuo en función de los dos roles que asume: el de encuestado y reclutador. En consecuencia, se bonifica a cada participante en cada uno de estos momentos para motivar su compromiso.

La estructuración de tal régimen se pensó a partir de la entrega de un cupón en cada instancia, por su participación como encuestado y por cada reclutamiento en que se garantice la encuesta. En cada uno de estos cupones se estipula que participarán en un sorteo por definir.

Como señalamos previamente, el sesgo en el muestreo es minimizado en el RDS cuando logramos mayores cadenas de referencia. En ese sentido, para estimular cadenas de referencias mayores y promover mayor profundidad sociométrica se utilizan cuotas (cantidad máxima de cupones por participante) de reclutamiento para limitar la capacidad de que los miembros de la población con grandes redes personales dominen la muestra (Heckathorn, 1997).

Como señalan Wejnert, & Heckathorn (2011), si un encuestado recluta a 10 personas, cada uno de los cuales recluta a 10 más, el tamaño de la muestra crecería rápidamente. Por ejemplo, se comienza con una sola semilla (onda 0) a 10, luego a 100, a 1 000 y 10 000 en la onda 4. En contraste, si cada encuestado sólo recluta a 2 personas, el crecimiento será mucho más lento, desde la semilla inicial a 2, luego a 4, luego 8 y finalmente 16 en la onda 4. Es decir, cuotas de reclutamiento más restrictivas producen cadenas de reclutamiento con más ondas y consiguen que los reclutamientos sean escasos y valiosos como para perderlos.

Siguiendo con el argumento de Wejnert, & Heckathorn (2011), elegir el tamaño de la cuota de reclutamiento implica un *trade-off*. Si la cuota es demasiado pequeña, el reclutamiento puede desaparecer porque algunos sujetos no van a lograr reclutar la cantidad potencial que podrían, además de que al ser tan restrictiva la cuota se invalida el potencial que podrían alcanzar sujetos con alta densidad de redes sociales. La recomendación es que las cuotas sean pequeñas (aquí es importante un conocimiento previo del campo de estudio). Los antecedentes señalan que las cuotas se han establecido entre 3 y 4 por reclutador. En ese sentido, en nuestro estudio se tiene previsto limitar a un máximo de 4 la cantidad de reclutados por participante.

Estimación de los valores poblacionales e intervalos de confianza

Los valores poblacionales de cada atributo pueden estimarse mediante el conocimiento de la población objetivo que posee el encuestado y la matriz de transición de probabilidades resultante de la muestra.

La densidad de la población (D) se abordará mediante la incorporación en la encuesta de una pregunta que indague sobre la cantidad de personas de la población objetivo que el participante conoce en cada una

de las “m” categorías de las variables analizadas ($X_{(1)}$, $X_{(2)}$, $X_{(3)}$). Esto nos permite tener una primera idea del tamaño medio de la población (Mantecón *et al.*, 2008). A modo de ilustración, consideremos la variable $X_{(1)}$ que indica el género del hijo o hija de migrantes sur-sur, en donde la primera categoría (“s”) refiere a la categoría “varón” y la segunda (“t”) a la condición “mujer”. Por lo tanto, cada individuo “j” nos revela la cantidad de personas que conoce en cada una de las subpoblaciones mencionadas (D_{js} y D_{jt}). En consecuencia, a partir de los valores obtenidos en la muestra podemos estimar tanto el tamaño promedio de la red general (\bar{D}) como el tamaño medio de la red para cada grupo (\bar{D}_s y \bar{D}_t). En donde $\bar{D}_s = \frac{\sum_j^n D_{js}}{n}$, $\bar{D}_t = \frac{\sum_j^n D_{jt}}{n}$ y $\bar{D} = \bar{D}_t + \bar{D}_s$.

Uno de los aportes más importante que realiza esta metodología es la posibilidad de construcción de intervalos de confianza para los estimadores obtenidos. Esto se logra mediante la generación de nuevas muestras de igual tamaño construidas a partir de un proceso de selección con reemplazo (“bootstrap”) realizado sobre la muestra original obtenida.

El método partitiona la muestra al ubicar a cada individuo “j” en función del grupo al que corresponde el individuo “j-1” que lo reclutó. Estos grupos son los conformados por las categorías de las características de relevancia seleccionadas ($X_{(i)}$). En consecuencia, al ser “m” la cantidad de categorías o estados posibles que puede asumir cada observación en una variable específica ($X_{(i)}$) se consideran “m” particiones sobre la muestra original (n). En cada sub-muestra (n_k) se ubicarán aquellas personas que fueron reclutadas por personas pertenecientes al mismo grupo. Por lo que se cumple que $\sum_{k=1}^m n_k = n$

Posteriormente se selecciona un elemento de la muestra (n) y se observa el valor que asume la característica de estudio ($X_{(i)} = k$). Es decir, se observa a qué grupo de la población objetivo pertenece el encuestado y en función de dicha respuesta se selecciona un nuevo elemento en la submuestra conformada por encuestados reclutados por personas pertenecientes a este grupo.

El proceso de selección continúa con esta lógica hasta que la nueva muestra (n') obtenida logra el tamaño de la muestra inicial (n) y se comienza un nuevo proceso de selección para alcanzar una nueva réplica (n'') de igual tamaño.

A partir de las réplicas obtenidas se calculan las proporciones muestrales y se construyen los intervalos de confianza para dicho estimador. Por lo tanto, es importante en el RDS realizar un seguimiento estricto de quién reclutó a quién para calcular las proporciones correspondientes.

El estimador de la proporción poblacional para el grupo “s” (\hat{P}_s) utilizado por el RDS bajo el enfoque de estimación de la varianza mediante “bootstrap” se puede expresar con la siguiente fórmula:

Estimador Poblacional RDS

$$\hat{P}_s = \frac{p_{st}\bar{D}_t}{p_{st}\bar{D}_t + p_{ts}\bar{D}_s}$$

p_{st} proporción (muestral) del grupo “s” seleccionado por el grupo “t”, p_{ts} proporción (muestral) del grupo “t” seleccionado por el grupo “s”, (\bar{D}) tamaño de red promedio general, \bar{D}_s tamaño medio de la red (grado) para el grupo “s”, \bar{D}_t tamaño de red promedio (grado) para el grupo “t”. Además, bajo los supuestos establecidos se cumple la condición de que $\hat{P}_s + \hat{P}_t = 1$. Por lo tanto, despejando se puede obtener el estimador para la categoría complementaria: $\hat{P}_t = \frac{p_{ts}\bar{D}_s}{p_{st}\bar{D}_t + p_{ts}\bar{D}_s}$

Se detalla la obtención de los valores de equilibrio:

$$\hat{P}_s = \frac{p_{st}\bar{D}_t}{p_{st}\bar{D}_t + p_{ts}\bar{D}_s} \quad (6)$$

$$\hat{P}_s + \hat{P}_t = 1 \quad (7)$$

Despejando en (2)

$$\hat{P}_t = 1 - \hat{P}_s \quad (8)$$

Reemplazando en el segundo miembro de (3) el valor del estimador \hat{P}_s definido en (1)

$$\Rightarrow \hat{P}_t = 1 - \left(\frac{p_{st}\bar{D}_t}{p_{st}\bar{D}_t + p_{ts}\bar{D}_s} \right)$$

$$\hat{P}_t = \frac{p_{st}\bar{D}_t + p_{ts}\bar{D} - p_{st}\bar{D}_t}{p_{st}\bar{D}_t + p_{ts}\bar{D}_s}$$

$$\therefore \hat{P}_t = \frac{p_{ts}\bar{D}}{p_{st}\bar{D}_t + p_{ts}\bar{D}_s} \quad (9)$$

El estimador de \hat{P}_s es calculado sobre la muestra original (n). Sin embargo, para la construcción de su intervalo de confianza mediante este enfoque es necesario el cálculo de la estimación de la proporción poblacional de cada una de las categorías de la variable en estudio $X_{(i)}$ en cada una de las muestras réplicas obtenidas anteriormente (n' , n'' , ...). Esto último nos permite estimar la desviación estándar del estimador *E.E.* (\hat{P}_s) y la obtención del intervalo de confianza para \hat{P}_s .

Intervalo de Confianza del Estimador Poblacional RSD

$$IC(\hat{P}): \hat{P}_s \pm E.E.(\hat{P}_s)$$

Especificamente para el procesamiento de los datos resultado del muestreo RDS se tiene previsto trabajar con el *software* R (R Core Team, 2020) y el paquete RDS (Handcock; Fellows, & Gile, 2012).

Con la generación de datos primarios se tiene previsto contribuir a la discusión sobre las interrelaciones entre los procesos migratorios y los mecanismos de producción y reproducción de desigualdades en los países del sur global. Se espera profundizar este debate y superar las limitaciones propias de los datos secundarios disponibles. En efecto, en las estadísticas disponibles generadas por el aparato estadístico estatal, como las encuestas periódicas de hogares, no es frecuente tener información sobre la nacionalidad de los padres. Y cuando ésta existe, la población de interés relevada es de tamaño pequeño (n).

Referencias

Abdul-Quader; Abu, S.; Heckathorn, Douglas D.; McKnight, Courtney; Bramson, Heidi; Nemeth, Chris; Sabin, Keith; Gallagher, Kathleen, & Des Jarlais, Don (2006). Effectiveness of Respondent-Driven Sampling for Recruiting Drug Users in New York City: Findings from a Pilot Study. *Journal of Urban Health*, 83(3), 459-476.

- Alba, Richard, & Nee, Victor (2003). *Remaking the American Mainstream: Assimilation and Contemporary Immigration*. Cambridge, Massachusetts: Harvard University Press.
- Bohrt, Marcelo A., & Itzigsohn, José (2015). Class, Race, and The Incorporation of Latinos/as Testing the Stratified Ethnoracial Incorporation Approach. *Sociology of Race and Ethnicity*, 1(3), 360-377.
- Deaux, Edward, & Callaghan, John W. (1985). Key Informant Versus Self-Report Estimates of Health-Risk Behavior. *Evaluation Review*, 9(3).
- Goel, Sharad, & Salganik, Matthew J. (2010). Assessing Respondent-Driven Sampling. *Proceedings of the National Academy of Sciences*, 107(15), 6743-6747.
- Gómez, Pablo S. (2019). Perspectivas teóricas en el estudio de la incorporación de migrantes: modelos analíticos, entramados jerárquicos y racialización. *Tabula Rasa*, 30.
- Goodman, Leo A. (1961). Snowball Sampling. *The Annals of Mathematical Statistics*, 32(1), 148-170.
- Handcock, Mark; Fellows, Ian, & Gile, Krista (2012). RDS: Respondent-Driven Sampling, Version 0.9-2. Recuperado en mayo de 2020 de <https://CRAN.R-project.org/package=RDS>
- Heckathorn, Douglas D. (1997). Respondent-Driven Sampling: A New Approach to the Study of Hidden Populations. *Social Problems*, 44(2), 174-199.
- Heckathorn, Douglas D. (2002). Respondent-Driven Sampling II: Deriving Valid Population Estimates from Chain-Referral Samples of Hidden Populations. *Social Problems*, 49(1), 11-34.
- Heckathorn, Douglas, D. & Jeffri, Joan (2001). Finding the beat: Using Respondent-Driven Sampling to Study Jazz Musicians. *Poetics*, 28(4), 307-329.
- Hujo, Katja, & Piper, Nicola (2010). *South-South Migration: Implications for Social Policy and Development*. London: Palgrave Macmillan.
- Itzigsohn, Jose (2009). *Encountering American Faultlines: Race, Class, and Dominican Experience in Providence*. New York: Russell Sage Foundation.
- Levitt, Peggy, & Waters, Mary C. (2006). *The Changing Face of Home: The Transnational Lives of the Second Generation*. New York: Russell Sage Foundation.

- Mantecón, Alejandro; Montse, Juan; Calafat, Amador; Becoña, Elisardo, & Román, Encarna (2008). Respondent-Driven Sampling: un nuevo método de muestreo para el estudio de poblaciones visibles y ocultas. *Adicciones*, 20(2), 161-170.
- Portes, Alejandro; Aparicio Gómez, Rosa, & Haller, William (2016). *Spanish Legacies: The Coming of Age of the Second Generation*. Oakland, California: University of California Press.
- Portes, Alejandro, & Rumbaut, Rubén G. (2006). *Immigrant America: a portrait*. Oakland, California: University of California Press.
- Portes, Alejandro, & Zhou, Min (1993). The New Second Generation: Segmented Assimilation and its Variants. *The ANNALS of the American Academy of Political and Social Science*, 530(1), 74-96.
- Qiu, Peiyuan; Yang, Yang; Ma, Xiao; Wu, Fang; Yuan, Ping; Liu, Qiaolan, & Caine, Eric (2012). Respondent-Driven Sampling to Recruit In-Country Migrant Workers in China: a Methodological Assessment. *Scandinavian Journal of Public Health*, 40(1), 92-101.
- R Core Team (2020). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. Recuperado de <https://www.R-project.org/>
- Ramirez-Valles, Jesus; Garcia, Dalia; Campbell, Richard T; Diaz, Rafael, & Heckathorn, Douglas (2008). HIV Infection, Sexual Risk Behavior, and Substance Use Among Latino Gay and Bisexual Men and Transgender Persons. *American Journal of Public Health*, 98(6), 1036-1042.
- Salganik, Matthew J., & Heckathorn, Douglas D. (2004). Sampling and Estimation in Hidden Populations Using Respondent-Driven Sampling. *Sociological Methodology*, 34(1), 193-240
- Simon, Patrick (2015). The Choice of Ignorance: The Debate on Ethnic and Racial Statistics in France. En Simon, Patrick; Piché, Victor; Gagnon, Amélie (eds.), *Social Statistics and Ethnic Diversity: Cross-National Perspectives in Classifications and Identity Politics* (pp. 65-87). New York: Springer Open.
- Tyldum, Guri, & Johnston, Lisa (2014). *Applying Respondent Driven Sampling to Migrant Populations: Lessons from the Field*. London: Palgrave Macmillan.
- Watters, John K., & Biernacki, Patrick (1989). Targeted Sampling: Options for the Study of Hidden Populations. *Social Problems*, 36(4), 416-430.

Wejnert, Cyprian, & Heckathorn, Douglas (2011). Respondent-Driven Sampling: Operational Procedures, Evolution of Estimators, and Topics for Future Research. En Williams, Malcolm, & Vogt, Paul, *The SAGE Handbook of Innovation in Social Research Methods* (pp. 473-497). London, United Kingdom: SAGE Publications.

Acerca de los autores

Pablo Sebastián Gómez es investigador del Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) y docente de la Facultad de Ciencias Sociales de la Universidad Nacional de Córdoba, Argentina. Es doctor en demografía por la Facultad de Ciencias Económicas de la Universidad Nacional de Córdoba. Sus principales áreas de interés comprenden: migraciones internacionales, estratificación social, métodos cuantitativos, métodos mixtos, interacciones género/trabajo/racialización. Sus artículos más recientes son:

En coautoría con Maximiliano Iglesias (2021). *Incorporación segmentada de migrantes e hijos de migrantes sur-sur en Argentina: intersecciones de género, generaciones y trabajo. Revista Brasileira de Estudos de População* (en prensa).

En coautoría con Virginia Bolatti (2021). *La participación de varones y mujeres jóvenes en la toma de decisiones sobre salud sexual y reproductiva en Argentina: género y asimetrías socioeconómicas. Población y salud en Mesoamérica. 2021.*

Maximiliano Iglesias es licenciado en Economía por la Universidad Nacional de Córdoba (UNC). Maestrando en estadística aplicada (UNC) y candidato a doctor en demografía por la Facultad de Ciencias Económicas de la Universidad Nacional de Córdoba, Argentina. En la actualidad es docente del Departamento de Matemática y Estadística de la Facultad de Ciencias Económicas, investigador del Instituto de Estadística y Demografía y becario de la Secretaría de Ciencia y Técnica de la misma universidad. Sus principales áreas de interés comprenden: mercados de

trabajo, econometría, métodos longitudinales, estratificación social y migraciones internacionales. Su artículo más reciente es:

Pablo Sebastián Gómez (2021). *Incorporación segmentada de migrantes e hijos de migrantes sur-sur en Argentina: intersecciones de género, generaciones y trabajo. Revista Brasileira de Estudos de População* (en prensa).