

Estudio Reinserción, Desistimiento y Reincidencia en Mujeres *Evaluación de Sesgo en Línea Base**

Sebastián Daza¹ y Pilar Larroulet²

¹University of Wisconsin-Madison

²Pontificia Universidad Católica de Chile

1. Introducción

En este documento examinamos el sesgo de no-respuesta de la *Línea Base* del estudio *Reinserción, Desistimiento y Reincidencia en Mujeres Privadas de Libertad en Chile* (ver más detalles en Larroulet y col., 2019). En primer lugar, describimos el diseño muestral del estudio y estimamos la tasa de respuesta de la línea base. Luego examinamos el balance entre aquellos que participan en la línea base y los que no usando once variables. Finalmente, calculamos el sesgo relativo y estandarizado, y comparamos los valores observados con muestras simuladas del mismo tamaño. Concluimos identificando los principales sesgos en la muestra y posible soluciones.

2. Diseño muestral

El estudio *Reinserción, Desistimiento y Reincidencia en mujeres privadas de libertad en Chile* tiene por objetivo describir el proceso de reinserción que vive la mujer una vez que sale en libertad, identificando factores que inciden en su desistimiento o reincidencia. Para ello, seguimos a una cohorte de mujeres que egresaron luego de haber cumplido una pena privativa de libertad, a quienes se les aplicó una encuesta de línea base semanas antes del egreso, y cuatro mediciones fuera de la cárcel.

La población objetivo consiste en mujeres chilenas privadas de libertad en el Centro Penitenciario Femenino de San Joaquín (CPF) o Centro de Estudio y Trabajo Talita Kum, de nacionalidad chilena, que hayan tenido una condena por 30 días o más y que salieran en libertad entre el día 14 de septiembre de 2016 y el 31 de marzo de 2017. Se consideran solo mujeres que egresan por cumplimiento de condena, libertad condicional y libertad vigilada.¹

*Versión: **26 de abril de 2019**. Centro de Estudios Justicia y Sociedad, Instituto de Sociología, Pontificia Universidad Católica de Chile. Este estudio fue realizado con el apoyo de Fundación San Carlos de Maipo y Fundación Colunga, y aportes adicionales del Banco Interamericano de Desarrollo (BID) y la *Division on Women and Crime* de la Asociación Norteamericana de Criminología (DWC-ASC). **Citar como:** Daza, S. y Larroulet, P. (2019). *Evaluación de Sesgo en Línea Base: Estudio Reinserción, Desistimiento y Reincidencia en Mujeres Privadas de Libertad en Chile*. Centro de Estudios Justicia y Sociedad del Instituto de Sociología, Pontificia Universidad Católica de Chile.

¹Las mujeres que egresan por orden del tribunal y/o por Ley 19.856 de rebaja de condena se eliminaron de la población objetivo por la dificultad para anticipar sus fechas de egreso.

3. Tasa de respuesta Línea Base

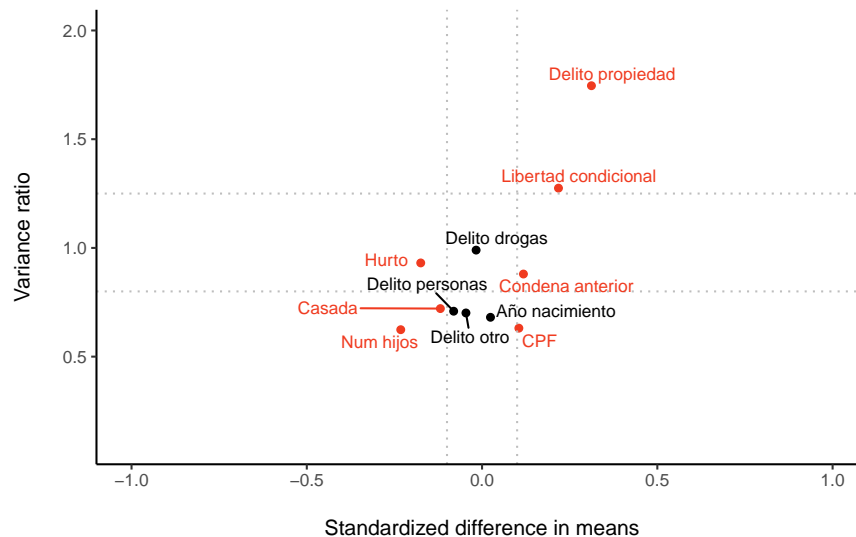
La tasa de respuesta corresponde a la fracción entre el número de entrevistas realizadas y el total de mujeres elegibles (que cumplen con los criterios de la población objetivo). De un total de 277 mujeres elegibles, 225 participaron en la línea base (81 %). Un total de 22 mujeres rechazaron participar. Entre las razones para no participar están la falta de motivación, no querer recordar la experiencia en la cárcel, poder enfocarse a sus hijos y familia, y no estar dispuestas a que se consulten sus datos administrativos. Del total de 277 mujeres, 30 fueron incluidas en el marco muestral de manera retrospectiva, por lo que no pudieron ser contactadas para la aplicación de la encuesta de *línea de base*.

4. Balance participantes / no-participantes

Debido a la selección no aleatoria de las mujeres que no participaron en el estudio, es posible que existan sesgos sistemáticos en la muestra que pueden distorsionar las inferencias de la población. Para evaluar la existencia de sesgos, estimamos el balance (qué tan similares son) entre las mujeres que participaron ($n=225$) y las que no lo hicieron ($n=52$) usando once variables.²

Usamos dos medidas para evaluar balance: diferencia estandarizada de medias estandarizadas y razón de varianza. La Figura 1 muestra once variables y sus respectivas diferencias y razones de varianza. En rojo aparecen las variables con diferencias estandarizadas mayores que 0.1.³

Figura 1: Balance Muestra / No participantes



La variable *delito de propiedad* es aproximadamente 0.3 desviaciones estándares más alta en la muestra que entre las mujeres que no responden. La desviación es menor respecto a la proporción de mujeres con

²El código usado para realizar este análisis y figuras se encuentra en:

https://github.com/sdaza/reentry-chile/blob/master/reports/bias_baseline.ipynb.

³Para las mujeres participantes se utilizó información obtenida de Gendarmería y, cuando estos no estaban disponibles, de su propio autoreporte en la línea base. En contraste, los datos de quienes no participaron en el estudio fueron obtenidos únicamente a través de Gendarmería. Esto podría explicar, en parte, las diferencias entre ambos grupos, principalmente en términos de número de hijos y estado civil.

libertad condicional, condena anterior y en la unidad CPF. Las mujeres que no participaron, por el contrario, poseen mayor número de hijos, hurtos, delitos a personas y son en mayor proporción casadas.

Como es de esperar, debido a las diferencias en el número de casos de la muestra y las mujeres que no participaron, las razones de varianza tienden a ser menores que 1 (la varianza es más alta entre las mujeres que no participaron), a excepción de *delitos de propiedad*.

5. Simulación

Podemos estimar el sesgo respondiendo la siguiente pregunta hipotética: cómo hubiese sido la distribución de las variables si la muestra de la línea base fuese completamente aleatoria. Esta aproximación nos permite tomar en cuenta tanto el sesgo como la variabilidad y error asociado al tamaño de la muestra (225).

Siguiendo a Groves, 2006, usamos dos medidas de sesgo, sesgo relativo (r) y estandarizado (s):

$$r = \frac{\bar{y}_r - \bar{y}_n}{\bar{y}_n}$$
$$s = \bar{y}_r^{std} - \bar{y}_n^{std}$$

Donde \bar{y}_n representa el promedio de la población, \bar{y}_r el promedio de la muestra, y std corresponde a $\frac{y_i - \bar{y}_n}{\sigma_y}$. Luego, simulamos 1000 muestras aleatorias de tamaño 225 y estimamos el sesgo relativo y estandarizado. Obtenemos, de este modo, una *distribución de sesgos* en base a 1000 muestras aleatorias de tamaño 225 que comparamos con el sesgo observado en la muestra de línea base. Si el sesgo de la muestra observada es considerablemente diferente de los valores de la distribución, podemos concluir la presencia de sesgo sistemático.

La Figura 2, por ejemplo, revela que los sesgos relativos más importantes son delitos de propiedad, número de hijos y hurto. El resultado es similar cuando usamos sesgo estandarizado.

Figura 2: Distribución Sesgo Relativo 1000 muestras y valor observado

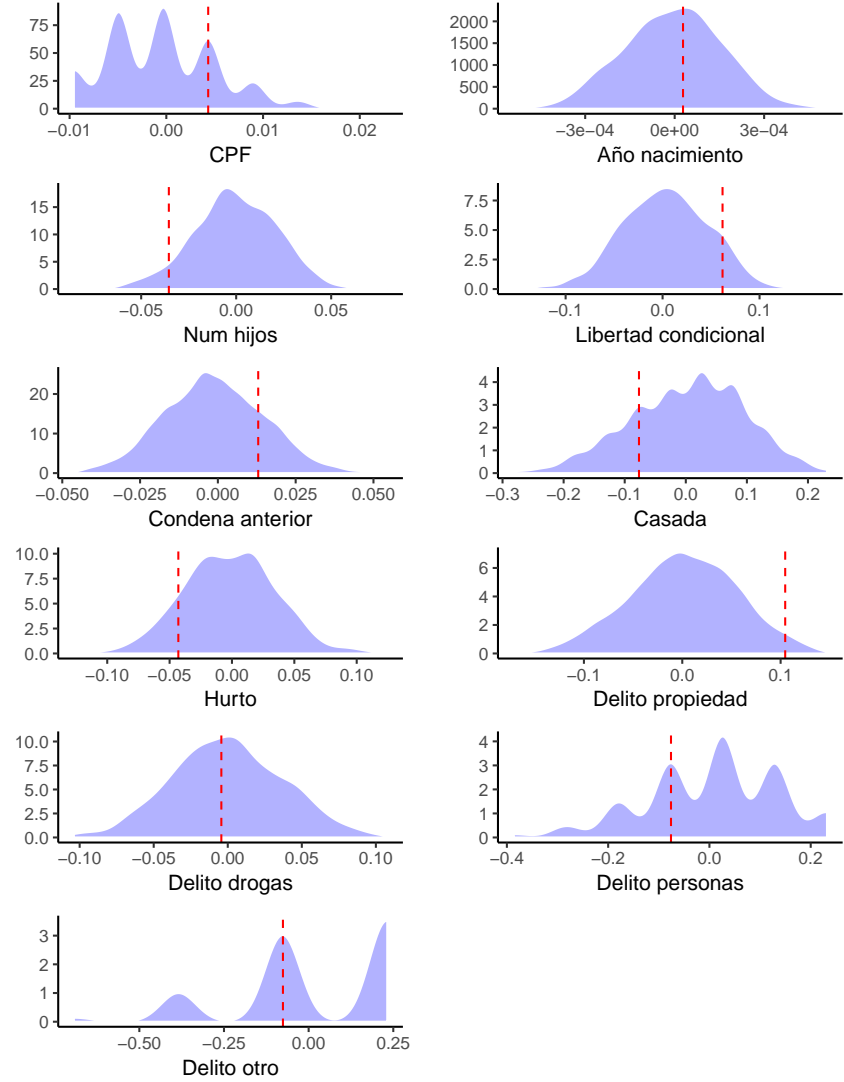
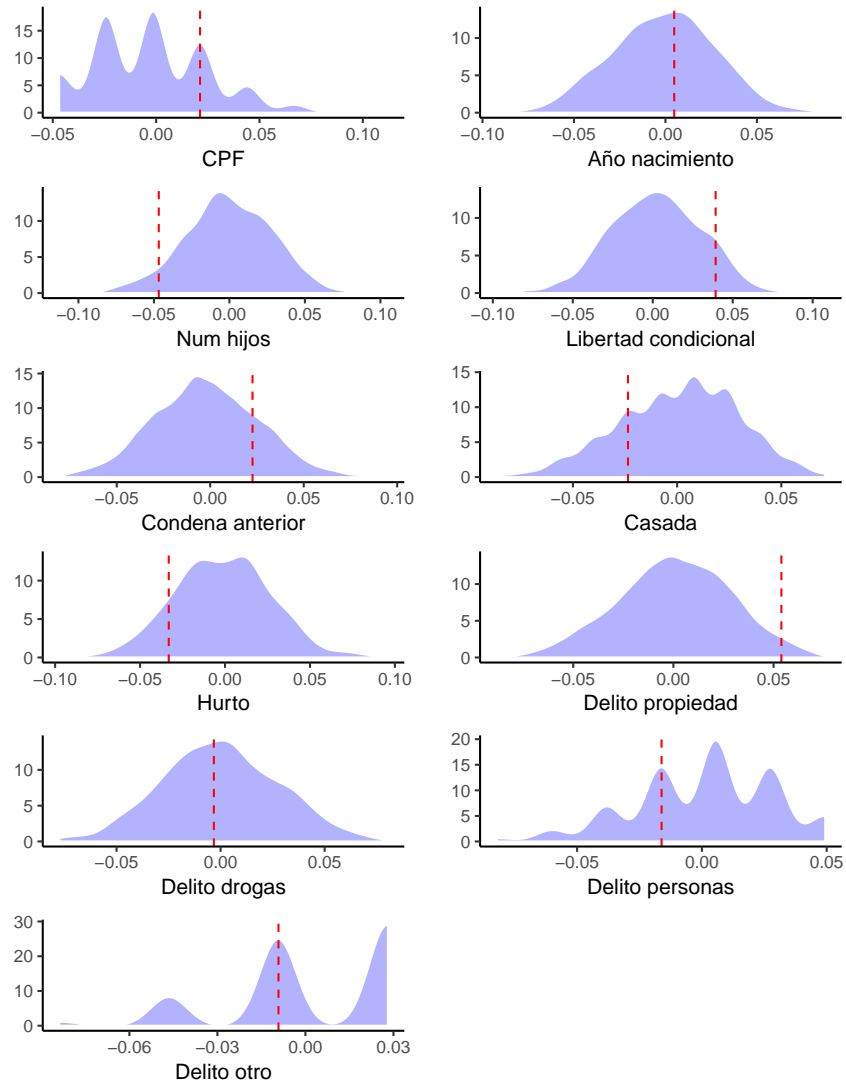


Figura 3: Distribución Sesgo Estandarizado 1000 muestras y valor observado



6. Pesos para ajustar no respuesta

Una manera de solucionar el problema del sesgo es ponderar la muestra de modo que la distribución de las once variables sea igual a la de la población. Usando las once variables con las que evaluamos el sesgo de no respuesta en la muestra de línea base, elaboramos pesos de modo que la distribución de las variables en dicha muestra se comporte como la población. Para ello, utilizamos un modelo *random forest* para estimar la probabilidad de que una mujer del marco muestral participe en el estudio (Valliant y col., 2018). Estas probabilidades las utilizamos luego para calcular pesos con esta fórmula:

$$\frac{1}{\text{prob}_i}$$

Figura 4: Distribución Sesgo Relativo 1000 muestras y valor observado con ponderador

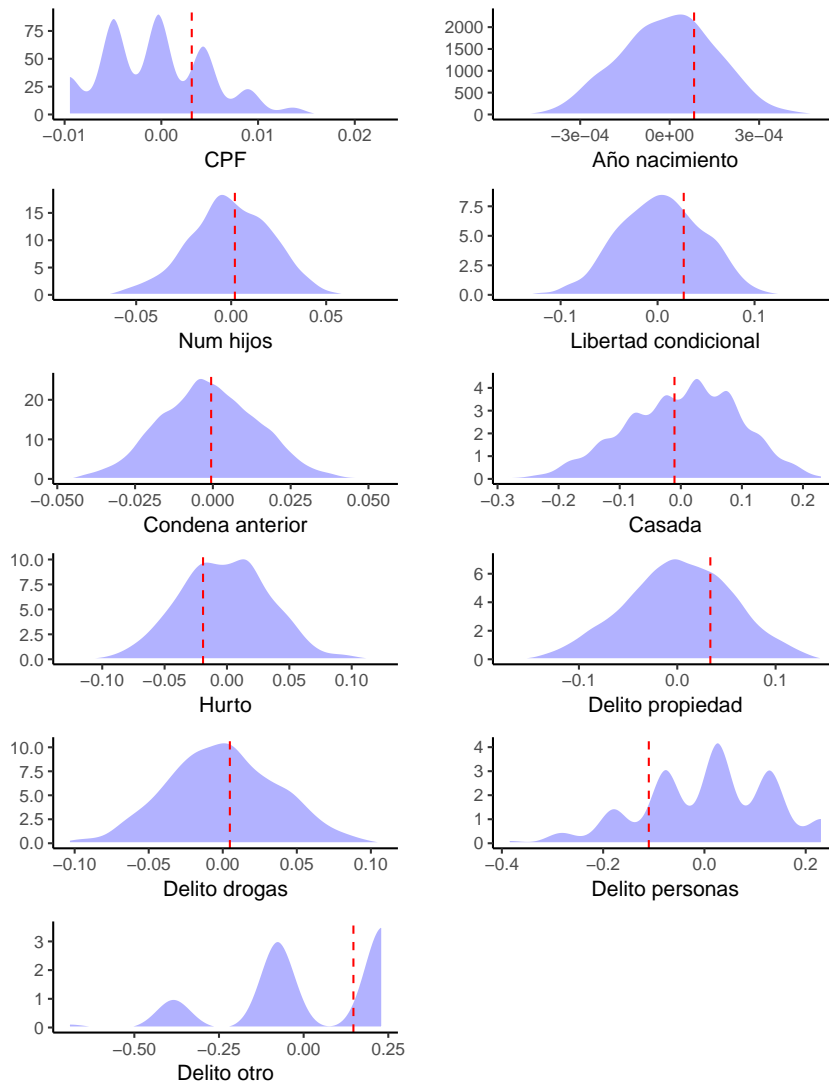
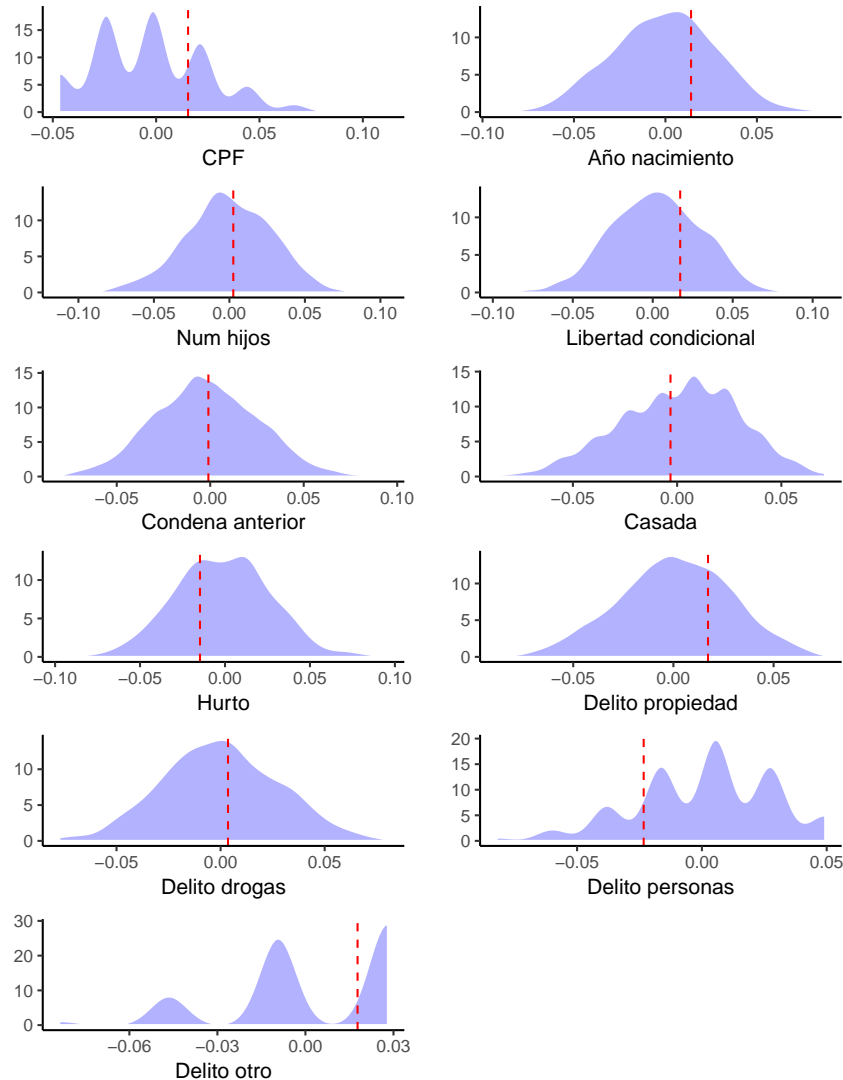


Figura 5: Distribución Sesgo Estandarizado 1000 muestras y valor observado con ponderador



Al replicar las Figuras 2 y 3, pero esta vez usando el ponderador, se aprecia que el sesgo observado (ponderado) se ubica cerca de cero (ver Figuras 4 y 5).

7. Conclusión

Utilizando once variables podemos concluir que existen diferencias sistemáticas entre la muestra y la población de mujeres egresadas entre 2016 y 2017. La diferencia más importante consiste en la proporción declarada de delitos de propiedad. Las mujeres en la muestra parecen tener un mayor involucramiento en delitos de propiedad distintos de hurto y previo contacto con el sistema judicial (condenas anteriores), estar en menor proporción casadas y tener un menor número de hijos.

¿De qué manera estas diferencias sesgan las conclusiones de estudio? Los descriptivos del estudio se pueden ver levemente distorsionados por las diferencias entre la muestra y aquellas mujeres que no par-

ticiparon. Un manera de solucionar este problema es ponderar la muestra. Aún así, pueden haber otras diferencias en variables que no consideramos y observamos. Sin embargo, las once variables que evaluamos (delito, edad, familia, condenas previas) son pertinentes para corregir diferencias relacionadas con reinserción, reincidencia y desistimiento.

Referencias

Groves, R. M. (2006). Nonresponse Rates and Nonresponse Bias in Household Surveys. *Public Opinion Quarterly*, 70(5), 646-675.

Larroulet, P., Daza, S., Del Villar, P., Droppelmann, C., Figueroa, A. & Valenzuela, E. (2019). *Estudio Reinserción, Desistimiento y Reincidencia en Mujeres: Diseño, Metodología y Caracterización de la Muestra*. Centro de Estudios Justicia y Sociedad del Instituto de Sociología, Pontificia Universidad Católica de Chile. Santiago, Chile.

Valliant, R., Dever, J. A. & Kreuter, F. (2018). *Practical tools for designing and weighting survey samples*. Springer.