

Ajustes metodológicos factor de expansión EOD marzo 2020

En el contexto de la pandemia mundial, la Encuesta de Ocupación y Desocupación (EOD) de marzo de 2020 se realizó de manera telefónica, usando los datos de contacto para los cuartos pertenecientes al panel de hogares de levantamientos anteriores. El Centro de Microdatos (CMD), incorporó ajustes a los factores de expansión por posibles sesgos que puede inducir el cambio de modalidad, de acuerdo a las recomendaciones de CEPAL (2020).

En particular, se realizaron ajustes que consideraron dos aspectos: a) correcciones por ajustes de no respuesta telefónica y b) calibración en dos etapas para ajustar las características de la muestra a la encuesta panel y luego a las poblacionales. A continuación, se detallan ambos ajustes de acuerdo a las recomendaciones de CEPAL (2020)

Ajuste por *Propensity Score* para la no respuesta telefónica

El enfoque propuesto por Rosenbaum y Rubin (1993) es útil para esclarecer la estructura de no respuesta y, por lo tanto, para corregir el sesgo de selección (Lensvelt-Mulders, Lugtig and Hubregtse, 2009), que puede surgir por la no respuesta telefónica derivado del cambio de modalidad de la EOD.

Para la estimación de la no respuesta, se debe considerar las variables I_k y D_k , que indican si el hogar pertenece a la muestra original y si ha respondido la encuesta telefónica, respectivamente. Luego, asumiendo que la distribución de respuestas puede ser estimada, el *propensity score* o propensión de respuesta telefónica de un hogar en la muestra está dado por:

$$\Phi_h = \Pr(D_h = 1 | I_h = 1)$$

Es decir, la probabilidad de cada hogar de responder la encuesta. Esta probabilidad o propensión de respuesta telefónica es diferente para cada hogar y puede ser estimado utilizando variables de la composición socio-demográfica de los hogares del panel de hogares de la EOD, por medio de regresiones logísticas.

En la misma línea, Kim y Riddles (2012) muestran que es posible usar un modelo basado en el ajuste del *propensity score* de la muestra telefónica usando la siguiente expresión:

$$\text{logit}(\hat{\Phi}_h) = x_h \hat{\beta}$$

Es decir, se puede estimar la propensión de respuesta telefónica (*propensity score*), desde realizar una estimación de la probabilidad de responder dados ciertos observables x_h de los hogares y el resultado del contacto, con esto se puede recuperar los parámetros $\hat{\beta}$ y estimar $\text{logit}(\hat{\Phi}_h)$.

Para el caso de la EOD, $\hat{\beta}$ -el vector de los coeficientes estimados- fue recuperado mediante estimar la siguiente regresión logística:

$$\text{Responde}_h = \beta_0 + \beta_1 GSE_h + \beta_2 \text{cuarto}_h + \beta_3 \text{sexo}_h + \beta_4 \text{educacion}_h + \beta_5 \text{Ing. Fam}_h + e_h$$

Donde, cada variable es recuperada del panel original y son construidas como:

Responde_h : Variable dicotómica que toma valor 1 si el hogar responde.

GSE_h : Grupo socioeconómico del hogar.

Cuarto_h : Cuarto de la muerte al que pertenece el hogar.

Sexo_h : Sexo de la persona que deja su contacto en el hogar.

Educación_h : Nivel educacional de la persona que deja su contacto en el hogar.

Ing. Fam_h : Ingreso Familiar del hogar.

Una vez estimados los parámetros $\hat{\beta}$ se estima $\hat{\Phi}_h$, donde es importante notar que $\hat{\beta} = [\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \hat{\beta}_4, \hat{\beta}_5]$ y $x_h = [1, GSE_h, \text{cuarto}_h, \text{sexo}_h, \text{educacion}_h, \text{Ing. Fam}_h]$

Al considerar los ponderadores estándar de la EOD como w_h^{eod} , y habiendo estimado $\hat{\Phi}_h$, entonces el factor de expansión ajustado bajo esta primera corrección se calcula así:

$$w_h = \frac{w_h^{eod}}{\hat{\Phi}_h}$$

Método de calibración de dos pasos

En concordancia con las indicaciones de CEPAL (2020), la incorporación de métodos de calibración es deseable porque reduce tanto el error de muestreo como el sesgo de selección inducidos por el cambio de metodología telefónica. El ajuste de calibración consiste en una primera etapa en conseguir ponderadores que ajusten la distribución en variables claves de la muestra original a la

poblacional (por ejemplo, sexo, edad). Mientras en una segunda instancia conseguir ponderadores que ajusten la muestra telefónica a la muestra original, en variables que deberían ser las mismas utilizadas en la primera etapa más otras variables claves (por ejemplo, nivel socioeconómico).

Asimismo, Särndal y Lundström (2006) afirman que cuando las muestras de encuestas se ven afectadas por no respuesta es deseable tener un sistema de ponderación que reproduzca información auxiliar disponible y es eficiente cuando estima cualquier característica de interés en un estudio de multipropósito. Estimadores de calibración (Deville y Särndal, 1992) satisfacen estas condiciones y pueden ser fácilmente ajustados para mitigar el sesgo generado por el cambio en la modalidad de recolección.

Para aplicar esta corrección son necesarias dos fuentes de información. Por un lado, es necesaria la información de ajuste población usualmente utilizada en la EOD en base a las proyecciones poblacionales del INE (denotada x_{1k}). Por otro lado, las variables que fueron medidas en la encuesta original (panel de hogares de la EOD) (denotadas x_{2k}).

La primera etapa consiste entonces en encontrar el set de ponderaciones calibradas sujeta a la siguiente restricción (Särndal, 2007):

$$\sum_{s_m} w_{1k} x_{1k} = \sum_U x_{1k}$$

Este primer ajuste de calibración es el estándar en la EOD, donde se ajusta la muestra obtenida por sexo y edad a los niveles poblacionales.

En la segunda etapa, las ponderaciones intermedias w_{1k} deben ser utilizadas para calcular los ponderadores calibrados finales w_k de la muestra telefónica, sujeto a la siguiente restricción:

$$\sum_{s_t} w_k x_{2k} = \sum_{s_m} w_{1k} x_{1k} = \left(\begin{array}{c} \sum_U x_{1k} \\ \sum_{s_t} w_{1k} x_{2k} \end{array} \right)$$

Este segundo ajuste de calibración, es incorporado, de manera tal que la muestra obtenida desde la encuesta telefónica EOD esté calibrada por sexo, edad y grupo socioeconómico con la muestra original (panel EOD).

Bibliografía

- CEPAL (2020), “Recommendations for eliminating selection bias in household surveys during the coronavirus disease (COVID-19) pandemic”, mayo [online]
https://repositorio.cepal.org/bitstream/handle/11362/45553/S2000315_en.pdf?sequence=1&isAllowed=y
- Deville, J. C. and C. E. Särndal (1992), “Calibration estimators in survey sampling”, *Journal of the American Statistical Association*, vol. 87, No. 418.
- Gelman, A. and T. Little (1997), “Poststratification into many categories using hierarchical logistic regression”, *Survey Methodology*, vol. 27, No. 2
- Kim, J. K. and M. K. Riddles (2012), “Some theory for propensity-score-adjustment estimators in survey sampling”, *Survey Methodology*, vol. 38, No. 2
- Lensvelt-Mulders, G., P. Lugtig and M. Hubregtse (2009), “Separating selection bias and non-coverage in Internet panels using propensity matching”, *Survey Practice*, 2, No. 6.
- Rosenbaum, P. R. and D. B. Rubin (1983), “The central role of the propensity score in observational studies for causal effects”, *Biometrika*, vol. 70, No. 1
- Särndal, C. E. (2007), “The calibration approach in survey theory and practice”, *Survey Methodology*, vol. 33, No. 2
- Särndal, C. E. and S. Lundström (2006), “Estimation in surveys with nonresponse”, *Wiley Series in Survey Methodology*, Wiley.