

Ajustes metodológicos factor de expansión EOD Junio 2020

En el contexto de la pandemia mundial, la Encuesta de Ocupación y Desocupación (EOD) en el trimestre de junio del 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, siguiendo las recomendaciones de CEPAL (2020).

En particular, se realizaron dos ajustes que consideraron 1) correcciones por ajustes de no respuesta telefónico y 2) calibración en dos etapas para ajustar las características de la muestra a la encuesta panel y luego a las poblacionales. Ambos ajustes se detallan a continuación, siguiendo lo expuesto en 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_h y D_h , 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 es 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 línea de lo anterior, 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 dado 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 desde estimar la siguiente regresión logística:

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

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.

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

Ahora si consideramos 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 es calculado como:

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

Método de calibración de dos pasos

Según indica la literatura (CEPAL, 2020) la incorporación de métodos de calibración son deseados, dado que estos reducen el error de muestreo y el sesgo de selección inducidos por el cambio de metodología telefónico. 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) .

Reforzando lo ya expuesto, 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 tal que la muestra obtenida desde la encuesta telefónica EOD es calibrada por sexo, edad y grupo socioeconómico a la muestra original (panel EOD).

De esta forma, ambos ajustes están considerado en la variable fact_ajus2 de la base EOD para junio del 2020.

Calculo proyecciones trimestrales de la población

Para actualizar los factores de expansión de la Encuesta de Ocupación y Desocupación (EOD), se utilizaron las estimaciones y proyecciones anuales del INE sobre la población a nivel comunal de 2002 a 2035¹. Como es de esperarse, en países en desarrollo, el crecimiento de la población no es constante ni sigue un comportamiento lineal², por lo que el uso de modelos geométricos o exponenciales es recomendado para el cálculo de proyecciones intercensales (United Nations, 2018).

La idea de un crecimiento geométrico es tomar en cuenta la población inicial, pero también el incremento experimentado hasta entonces. Así, el aumento de la población siempre será relativo al tamaño de la población al empezar cada trimestre sin olvidar el valor final de la población al finalizar el año analizado. Un crecimiento acumulado entonces, implica que la misma tasa se va aplicando a cantidades progresivamente mayores, si es positiva, o a cantidades progresivamente menores, si es negativa. Por tanto, el cálculo acumulado de la población en un trimestre t en el año q determinado se calcula como:

$$Pob_t = Pob_{t-1} * (1 + r_{0-q})^n$$

Donde Pob_t es la población proyectada para el trimestre t y Pob_{t-1} es la población acumulada en el trimestre anterior. Mientras que r_{0-q} es la tasa de variación de la población del año anterior, 0, con el actual, q, es decir:

$$r_{0-q} = \left(\left(\frac{Pob_q}{Pob_0} \right)^{\frac{1}{n}} - 1 \right)$$

¹ Ver proyecciones anuales en INE (2019).

² Se proyecta un crecimiento de la población hasta más allá de 2050 para Chile. Más información en United Nations (2015).

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
- Instituto Nacional de Estadísticas (2019), *Estimaciones y Proyecciones de la Población de Chile 2002-2035 a Nivel Comunal*. Obtenido de https://www.inec.cl/docs/default-source/proyecciones-de-poblacion/metodologia/proyecci%C3%B3n-base-2017/estimaciones-y-proyecciones-2002-2035-comunas-metodolog%C3%ADa.pdf?sfvrsn=9459d1b0_4
- 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
- United Nations, Department of Economic and Social Affairs, Population Division (2015). *World Population Prospects: The 2015 Revision, Key Findings and Advance Tables*. Working Paper No. ESA/P/WP.241.
- United Nations, Department of Economic and Social Affairs, Population Division (2018). *World Urbanization Prospects: The 2018 Revision, Methodology*. Working Paper No. ESA/P/WP.252. New York: United Nations.
- 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.