

Minuta sobre el uso del factor de expansión longitudinal
Encuesta Social COVID-19 Panel r1-r3

1. Antecedentes

El Ministerio de Desarrollo Social y Familia (MDSF) pone a disposición del público general un panel conformado por los integrantes de los hogares encuestados en las rondas 1, 2 y 3 de la Encuesta Social COVID-19 (ESC-19). Esto permitirá realizar análisis longitudinales adicionales a los análisis de corte transversal que se pueden efectuar a partir de las bases publicadas para cada ronda.

El objetivo del panel de la ESC-19 es comprender fenómenos longitudinales asociados a la pandemia a partir del seguimiento de una muestra de hogares en el tiempo. Principalmente, aquellos fenómenos relativos a las consecuencias sociales y económicas del COVID-19 en las condiciones de vida de los hogares del país.

El objetivo de este documento es proporcionar una breve descripción del factor de expansión longitudinal, así como instrucciones para su uso y vinculación con las bases de datos de la ronda 1, ronda 2 y ronda 3.

2. Descripción del factor de expansión longitudinal

El factor de expansión longitudinal fue elaborado por el Instituto Nacional de Estadísticas (INE) de acuerdo con el diseño muestral de la Encuesta Social COVID-19¹. Este factor considera a los hogares y personas de la ronda 1, ronda 2 y ronda 3; y su elaboración incluye las siguientes etapas:

- Ajuste por no repuesta del panel: tiene por objetivo lograr que las unidades que no respondieron el panel de la encuesta puedan ser representadas por aquellas que sí lo hicieron. El método utilizado para compensar la no respuesta fue *propensity score*, que consiste en modelar la probabilidad de responder de las viviendas del panel utilizando un grupo de covariables. Posteriormente, se ordenan de menor a mayor las unidades de acuerdo con su probabilidad de responder y se generan veintiles para el ajuste.
- Suavizamiento: el método utilizado para el suavizamiento del factor de expansión longitudinal ajustado por no respuesta es un método mixto conformado por el método de contracción a la media y el método R-K.
- Calibración: es el último ajuste que se realiza al factor de expansión longitudinal y tiene como objetivo contar con estimaciones concordantes con fuentes externas, como son las proyecciones de población. El método utilizado para esta etapa fue *raking*, que utiliza la información de proyección de población, según sexo y edad.

3. Vinculación de las bases de datos de las rondas 1, 2 y 3

El factor de expansión longitudinal (exp_longitudinal) se encuentra disponible en la base de datos "fexp_longitudinal_r1-r2-r3" que además contiene las variables asociadas al diseño complejo de la encuesta (varunit_l y varstrat_l) y la llave para vincular el factor longitudinal con las bases de datos

¹ La metodología completa del diseño muestral, incluyendo lo referente a factores de expansión, se encontrará en la página <http://observatorio.ministeriodesarrollosocial.gob.cl/encuesta-social-covid19> durante octubre de 2021.

de las rondas 1, 2 y 3. En la tabla a continuación se muestran los nombres de las bases de datos y sus *links* de descarga.

Ronda	Nombre base de datos	Link
Factor longitudinal	fexp_longitudinal_r1-r2-r3	http://observatorio.ministeriodesarrollosocial.gob.c/encuesta-social-covid19-panel
Ronda 1	Encuesta Social COVID-19	http://observatorio.ministeriodesarrollosocial.gob.c/encuesta-social-covid19-primera-ronda
Ronda 2	Encuesta Social COVID-19 II	http://observatorio.ministeriodesarrollosocial.gob.c/encuesta-social-covid19-segunda-ronda
Ronda 3	Encuesta Social COVID-19 III	http://observatorio.ministeriodesarrollosocial.gob.c/encuesta-social-covid19-tercera-ronda

Para hacer análisis longitudinales es necesario vincular las variables de interés para el análisis de las bases de datos de las rondas 1, 2 y 3 con la base de datos que contiene el factor de expansión longitudinal, utilizando las siguientes variables para cada ronda:

fexp_longitudinal_r1-r2-r3	Encuesta Social COVID-19	Encuesta Social COVID-19 II	Encuesta Social COVID-19 III	Descripción
id_vivienda	id_vivienda	id_vivienda	id_vivienda	Identificador de la vivienda
n_hogar	n_hogar	n_hogar	n_hogar	Número de hogar entrevistado en la vivienda
lista_hogar__id_r1	lista_hogar__id			Identificador de personas en el hogar ronda 1
lista_hogar__id_r2		lista_hogar__id		Identificador de personas en el hogar ronda 2
ids			ids ²	Identificador de personas en el hogar para el análisis longitudinal

Las variables descritas en la tabla anterior permiten construir una llave en cada ronda, compuesta por el identificador de la vivienda (id_vivienda), el identificador del número de hogar entrevistado en la vivienda (n_hogar) y por el identificador de personas en el hogar para obtener el factor longitudinal desde la base “fexp_longitudinal_r1-r2-r3” (Ver Anexo 1).

Luego de realizar la vinculación se sugiere verificar que los universos de la base de datos resultante correspondan a lo especificado a continuación:

Unidad	N
Personas	4.892
Hogares	1.788

² La variable “ids” es un identificador único de personas a lo largo del tiempo que no necesariamente se encuentra en un orden correlativo dentro de cada hogar. A diferencia de la variable “lista_hogar__id” que sí presenta un orden correlativo dentro de cada hogar.

4. Consideraciones para el análisis longitudinal:

- Selección de variables: los cuestionarios de las distintas rondas de la ESC-19 se componen de un grupo de preguntas centrales, que se encuentran en todos los cuestionarios; de un grupo de preguntas rotativo, que están presentes en al menos dos rondas; y de un grupo de preguntas variables, que solo se encuentran en una ronda. Por ello, para hacer análisis longitudinales se sugiere revisar el documento “Comparativo preguntas r1, r2 y r3”, donde se comparan a nivel de pregunta los cuestionarios de las tres rondas disponibles.

- Presencia de atrición: el carácter de panel de la ESC-19 origina atrición, lo que ocurre cuando un hogar deja de participar en alguna ronda de la encuesta³. El panel conformado por las rondas 1, 2 y 3 está compuesto por 1.788 hogares y 4.892 personas, por lo que se recomienda revisar los coeficientes de variación de las estimaciones al efectuar análisis desagregados considerando que su calidad podría verse afectada por la disminución del tamaño muestral⁴.

- Diferencias entre las tres rondas: al vincular las tres bases de datos, es posible observar diferencias para algunas observaciones en variables como “sexo” y “edad”⁵ que se pueden explicar, principalmente, por lo siguiente:
 - La ESC-19 es una encuesta telefónica donde las respuestas son entregadas por un informante idóneo que responde tanto por sí mismo como por todos los miembros del hogar. Este informante puede cambiar entre una ronda y otra, dependiendo de quién estaba disponible en el hogar para responder la encuesta, lo que puede generar variaciones en las respuestas entre las rondas.
 - Errores no muestrales característicos de toda encuesta que se evidencian en estudios longitudinales, y que pueden responder a dificultades por parte del informante para recordar ciertos datos con precisión, a errores de tipeo al registrar las respuestas en el *software* diseñado para ello, entre otras.

- Jefatura de hogar: al vincular las bases de datos de las tres rondas publicadas, es posible observar que en algunos hogares la jefatura de hogar varía entre una ronda y otra. Es decir, en algunos hogares del panel la persona identificada como jefe(a) de hogar no es la misma a lo largo de las tres rondas. Además, existen 38 hogares que al conformar el panel no tienen un jefe(a) de hogar identificado según los datos de la ronda 3, debido a que este no formaba parte del hogar en al menos una de las rondas anteriores.

³ La atrición es abordada a través del ajuste por no respuesta.

⁴ Se sugiere revisar los estándares de calidad definidos por el INE: <https://www.ine.cl/docs/default-source/buenas-practicas/estandares/documento/estandar-evaluacion-de-calidad-de-estimaciones.pdf>

⁵ Los resultados longitudinales publicados por el Ministerio de Desarrollo Social y Familia en el sitio web del Observatorio Social, utilizan la información de “sexo” y “edad” de la tercera ronda de la encuesta.

Anexo 1. Código en Stata para la vinculación de las bases de datos con el factor longitudinal

```

clear all
set more off
global input "asignar_directorio_de_trabajo_1"
global output "asignar_directorio_de_trabajo_2"
use "$input\fexp_longitudinal_r1-r2-r3.dta", clear

***** Ronda 1 (Encuesta Social COVID-19)
rename lista_hogar__id_r1 lista_hogar__id
merge 1:1 id_vivienda n_hogar lista_hogar__id using "$input\Encuesta Social COVID-19.dta"
keep if _merge==3
*Guarda las variables necesarias para el análisis longitudinal
*Ejemplo variables de interés: ie17_1 ie17_2 ie17_3 ie17_4
keep id_vivienda n_hogar ids varunit_l varstrat_l exp_longitudinal lista_hogar__id lista_hogar__id_r2 ie17_1
ie17_2 ie17_3 ie17_4
*Renombra variables para identificar ronda de origen
rename (ie17_1 ie17_2 ie17_3 ie17_4) (ie17_1_r1 ie17_2_r1 ie17_3_r1 ie17_4_r1)
rename lista_hogar__id lista_hogar__id_r1

***** Ronda 2 (Encuesta Social COVID-19 II)
rename lista_hogar__id_r2 lista_hogar__id
merge 1:1 id_vivienda n_hogar lista_hogar__id using "$input\Encuesta Social COVID-19 II.dta"
keep if _merge==3
*Guarda las variables necesarias para el análisis longitudinal
*Ejemplo variables de interés: ie19_1 ie19_2 ie19_3 ie19_5
keep id_vivienda n_hogar ids varunit_l varstrat_l exp_longitudinal lista_hogar__id_r1 ///
lista_hogar__id ie17_1_r1 ie17_2_r1 ie17_3_r1 ie17_4_r1 ie19_1 ie19_2 ie19_3 ie19_5
*Renombra variables para identificar ronda de origen
rename (ie19_1 ie19_2 ie19_3 ie19_5) (ie19_1_r2 ie19_2_r2 ie19_3_r2 ie19_5_r2)
rename lista_hogar__id lista_hogar__id_r2

***** Ronda 3 (Encuesta Social COVID-19 III)
merge 1:1 id_vivienda n_hogar ids using "$input\Encuesta Social COVID-19 III.dta"
keep if _merge==3
*Guarda las variables necesarias para el análisis longitudinal
*Ejemplo variables de interés: ie24_1 ie24_2 ie24_3 ie24_5
keep id_vivienda n_hogar ids varunit_l varstrat_l exp_longitudinal ///
lista_hogar__id_r1 ie17_1_r1 ie17_2_r1 ie17_3_r1 ie17_4_r1 ///
lista_hogar__id_r2 ie19_1_r2 ie19_2_r2 ie19_3_r2 ie19_5_r2 ///
ids ie24_1 ie24_2 ie24_3 ie24_5 macrozona
order macrozona, after(ids)
*Renombra variables para identificar ronda de origen
rename (ie24_1 ie24_2 ie24_3 ie24_5) (ie24_1_r3 ie24_2_r3 ie24_3_r3 ie24_5_r3)

*Guarda BBDD vinculada
save "$output\nombre_panel_r1_r2_r3.dta"

```