

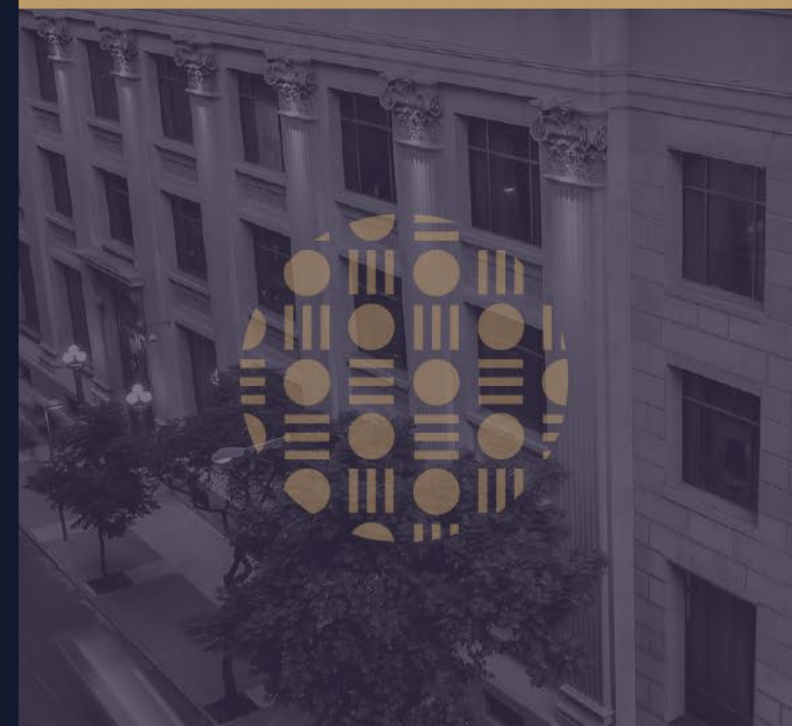


LA CONFIANZA  
ES CENTRAL

# Taller Aplicado Encuesta Financiera de Hogares

Javiera Vásquez, División de Estadísticas y Datos

Banco Central de Chile, mayo, 2025



# Antecedentes



- La Encuesta Financiera de Hogares (EFH) es una encuesta que tiene como objetivo recabar información detallada sobre la situación financiera de los hogares chilenos, utilizando como fuente principal a los propios hogares.
- En concreto, la encuesta pregunta sobre el estado de las familias en relación con sus ahorros, deudas, ingresos, activos financieros, uso de medios de pago, entre otros.
- Desde 2007, y cada tres años, el Banco Central de Chile (BCCh) realiza la EFH, entrevistando una muestra representativa de aproximadamente 4.500 hogares distribuidos en todo el territorio nacional urbano.
- A diferencia de otras encuestas del Banco, este producto ha sido diseñado completamente por el BCCh (muestra, cuestionario, imputaciones, factores de expansión, etc.), solo el trabajo de campo se externaliza mediante un proceso de licitación pública.





# Aspectos metodológicos

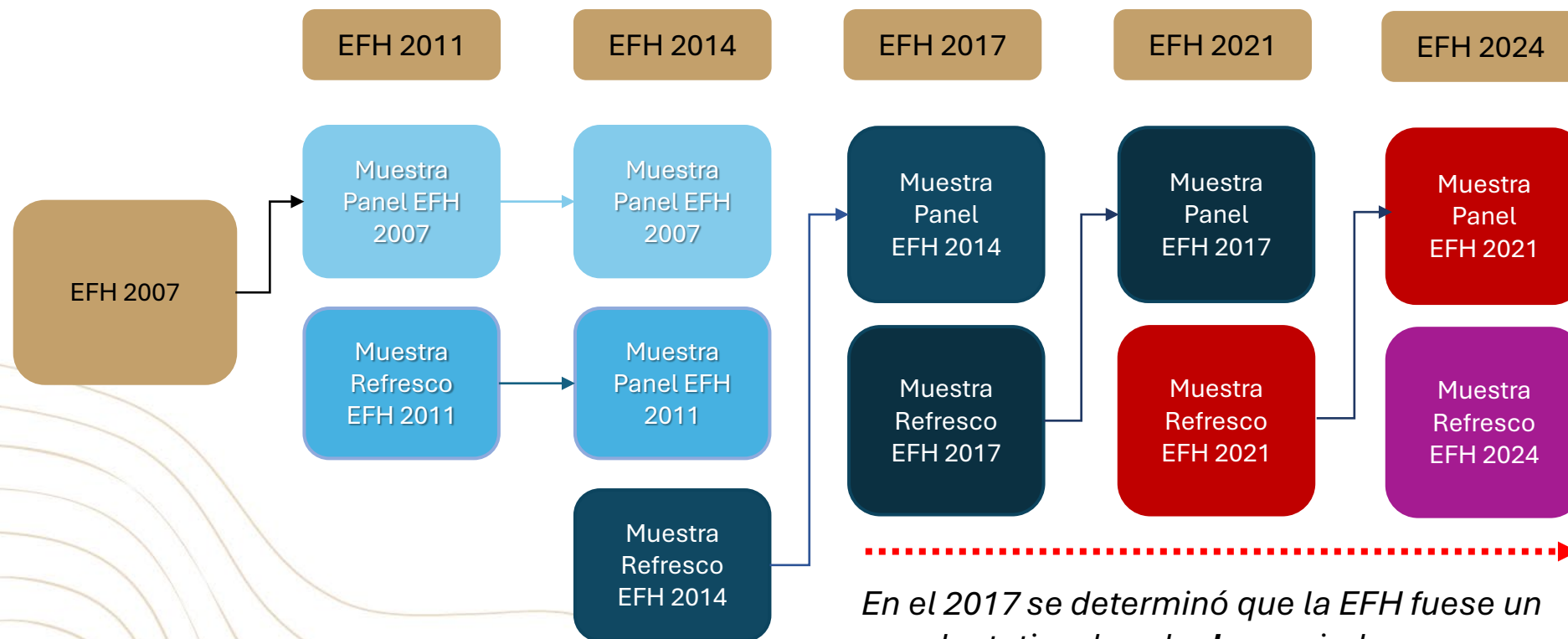


# Principales aspectos metodológicos

- La **población objetivo** de la EFH corresponde a los hogares urbanos que residen en viviendas particulares en el territorio nacional.
- La **cobertura** es nacional urbana, con representatividad a nivel de [macrozonas](#) y [estratos de riqueza](#).
- El diseño del estudio es del tipo panel rotativo, es decir, en cada ronda del estudio existe una parte de la muestra que fue entrevistada en ronda anterior (panel) y una parte de la muestra que va a ser entrevistada por primera vez (refresco).
- El **cuestionario** consta de 14 módulos temáticos con aproximadamente 400 preguntas. Su aplicación, tiene una duración promedio de 35 minutos.
- La **modalidad de levantamiento** ha sido históricamente cara a cara (presencial) con uso de dispositivo electrónico para la recolección de respuestas (CAPI), a excepción del último levantamiento que por motivo de la pandemia se efectuó de manera telefónica (CATI), luego de un precontacto presencial.
- La **muestra objetivo** es de 4.500 hogares, aproximadamente 2.800 hogares de la muestra refresco y 1.700 hogares de la muestra panel.
- La unidad de muestreo y de análisis en la EFH corresponde al **hogar**.
- Sólo una persona por hogar, el [entrevistado](#), responde la encuesta por él y el resto de los miembros del hogar.



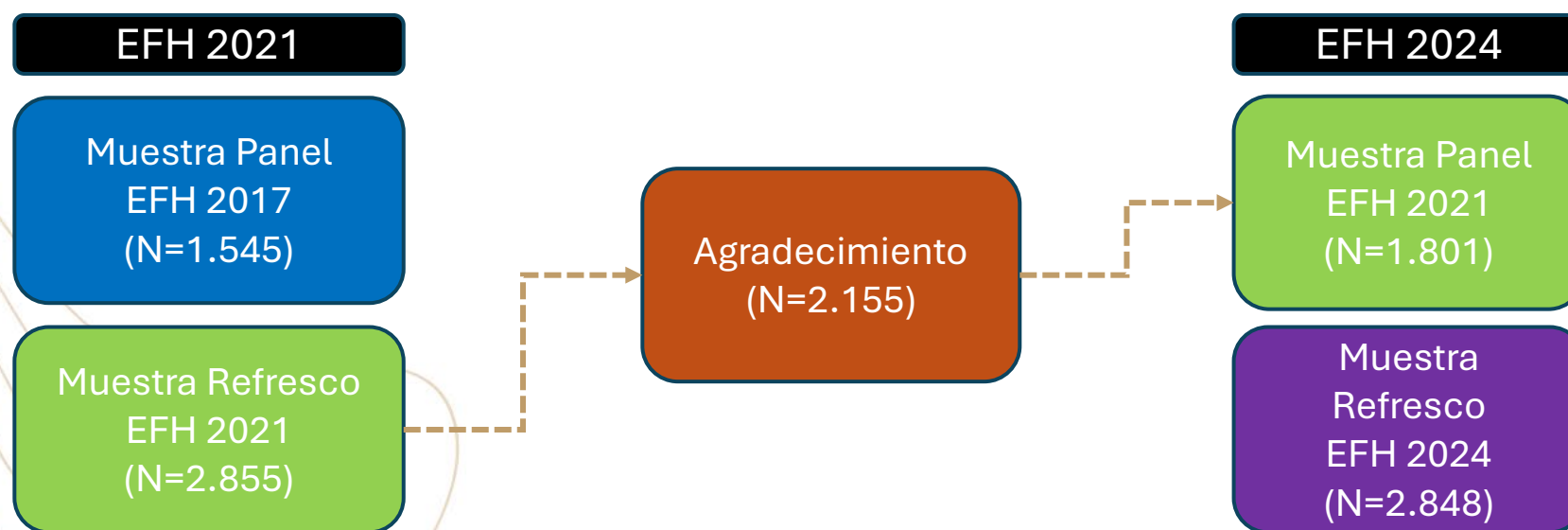
# Diseño del estudio



# Muestra

La muestra de la EFH está compuesta por una parte panel, es decir, personas que fueron entrevistadas en el año 2021, y una muestra de refresco que será entrevistada por primera vez este año.

La muestra está distribuida en 77 comunas a lo largo del territorio nacional.



# Diseño muestral y factores de expansión



- La muestra de la EFH es probabilística, seleccionada a través de un diseño muestral complejo el que se caracteriza por ser estratificado y con tres etapas de selección.
- Los estratos están definidos por la intersección de las 4 macrozonas geográficas y los 3 estratos de riqueza.
- La Unidad Primaria de Muestreo (UPM) es la comuna, luego, en una segunda etapa se seleccionan manzanas (USM) al interior de cada comuna, y finalmente, la Unidad Tercera de Muestreo (UTM) son las viviendas al interior de cada manzana.
- Los factores de expansión corresponden al inverso de la probabilidad de selección, y por lo tanto corresponde a la capacidad que tiene cada hogar seleccionado de representar la población objetivo.
- Para obtener estadísticas de posicionamiento tales como la media, mediana o distribuciones de frecuencias, que sean representativas de la población objetivo, es necesario ocupar los factores de expansión, y así indicar al software estadístico que se utilice cuántos hogares representa cada una de las unidades muestrales, y ponderando los resultados muestrales para obtener un estimador de la media, mediana o frecuencias de la población.



# Réplicas de los factores de expansión

- Para estudiar la significancia estadística de la media, por ejemplo, o para la obtención de intervalos de confianza de un parámetro poblacional, necesitamos calcular el error estándar de la estimación.
- En el caso de un diseño muestral complejo, como la EFH, las fórmulas de un muestreo aleatorio simple no pueden ser utilizadas, y en este caso, debemos indicar al software estadístico las características del diseño muestral complejo: estratos y unidades de muestreo, para que pueda hacer el cálculo del error estándar considerando el efecto diseño.
- Sin embargo, en el caso de la EFH, por temas de confidencialidad y resguardo de la identidad de los hogares entrevistados este tipo de información no es pública.
- Así, para poder calcular los errores estándar considerando el efecto diseño se utilizan una técnica no paramétrica llamada bootstrap weights. Así, el usuario de la EFH debe indicar al software las 1000 réplicas de los factores de expansión que deben ser utilizadas para simular los errores estándar y así poder obtener los correspondientes intervalos de confianza.

# Imputaciones



- Al momento de levantar una encuesta, los hogares pueden decidir no entregar información para algunas preguntas. Entre algunos de los factores que explican la no respuesta están la falta de comprensión de la pregunta, el desconocimiento de la respuesta o la negación a entregar información que para el hogar es sensible.
- El problema de la no respuesta tiende a ser mayor en preguntas de ingresos, deudas y activos, por lo cual en la EFH este es un tema recurrente.
- Dentro del proceso de generación de las bases de datos de la EFH se hacen imputaciones de las variables de monto.
- Se aplica el método de Hotdeck y de Imputación Múltiple Encadenada.
- El usuario dispone de 30 bases de imputaciones que permiten calcular las medias, medianas y errores estándar considerando todos los hogares, ya que aquellos que no responden son imputados.





# Bases de datos



# Bases de Datos

- Las bases de datos de la EFH y todo el material asociado, puede ser descargado en la página web del Banco Central de Chile ([www.bcentral.cl](http://www.bcentral.cl)) en la sección de Encuestas Económicas:

## [Encuesta Financiera de Hogares \(EFH\) - Banco Central de Chile](#)

- En la sección “Documentos para investigadores” se puede acceder a las bases de datos luego de un breve registro:

### Documentos para investigadores

En este sitio puedes encontrar las bases de datos de los distintos levantamientos de la EFH, los cuestionarios y las principales investigaciones que utilizan los datos de esta encuesta.

→ **Base de datos EFH**

→ **Documentación encuestas anteriores**

**Documentos EFH 2021**



**Estudios e investigación**



# Bases de Datos

- Las bases de datos de la EFH y todo el material asociado, puede ser descargado en la página web del Banco Central de Chile ([www.bcentral.cl](http://www.bcentral.cl)) en la sección de Encuestas Económicas:

## [Encuesta Financiera de Hogares \(EFH\) - Banco Central de Chile](#)

- En la sección “Documentos para investigadores” se puede acceder a las bases de datos luego de un breve registro:

### Documentos para inve

En este sitio puedes encontrar las bases de datos de esta encuesta.

→ **Base de datos EFH**

**Documentos EFH 2021**

**Estudios e investigación**

## Usuarios Registrados

Correo electrónico (\*)

jvasquez@bcentral.cl

Contraseña (\*)

....



No soy un robot



reCAPTCHA  
Privacidad - Términos

# Bases de Datos No Imputada

2021

 Base Factores Replicados EFH 2021

 Base imputada EFH 2021

 Base no imputada EFH 2021

 Llave Panel EFH

1

La base no imputada corresponde a la base bruta de la EFH, sin considerar imputaciones. Esta base contiene todas las variables y toda la información recolectada por la encuesta, exceptuando la información que permite individualizar los hogares. Cada hogar es identificado con un id y cada miembro del hogar a través de la variable ord.

# Bases de Datos No Imputada

Preguntas a nivel individual (para cada miembro del hogar)

|    | id    | ord          | genero_ent | tr_edad_ent        | p6_r7                | genero | edad | c1a | d1 |
|----|-------|--------------|------------|--------------------|----------------------|--------|------|-----|----|
| 1  | 10002 | Entrevistado | Mujer      | entre 35 y 49      | Jefe(a) de Hogar     | Mujer  | 37   | Si  | No |
| 2  | 10002 | 2            | .          | .                  | Cónyugue o pareja    | Hombre | 43   | Si  | No |
| 3  | 10002 | 3            | .          | .                  | Hijo(a), hijastro(a) | Hombre | 18   | No  | No |
| 4  | 10002 | 4            | .          | .                  | Hijo(a), hijastro(a) | Hombre | 16   | No  | No |
| 5  | 10002 | 5            | .          | .                  | Hijo(a), hijastro(a) | Mujer  | 2    | .   | No |
| 6  | 10003 | Entrevistado | Hombre     | igual o mayor a 65 | Jefe(a) de Hogar     | Hombre | 76   | No  | No |
| 7  | 10003 | 2            | .          | .                  | Cónyugue o pareja    | Mujer  | 74   | No  | No |
| 8  | 10004 | Entrevistado | Hombre     | igual o mayor a 65 | Jefe(a) de Hogar     | Hombre | 69   | No  | No |
| 9  | 10004 | 2            | .          | .                  | Cónyugue o pareja    | Mujer  | 63   | No  | No |
| 10 | 10004 | 3            | .          | .                  | Hijo(a), hijastro(a) | Hombre | 22   | No  | No |
| 11 | 10005 | Entrevistado | Mujer      | entre 35 y 49      | Jefe(a) de Hogar     | Mujer  | 45   | Si  | No |
| 12 | 10005 | 2            | .          | .                  | Cónyugue o pareja    | Hombre | 46   | Si  | No |
| 13 | 10005 | 3            | .          | .                  | Hijo(a), hijastro(a) | Hombre | 20   | No  | No |
| 14 | 10005 | 4            | .          | .                  | Hijo(a), hijastro(a) | Hombre | 13   | .   | No |
| 15 | 10005 | 5            | .          | .                  | Hijo(a), hijastro(a) | Mujer  | 26   | No  | No |
| 16 | 10006 | Entrevistado | Hombre     | entre 50 y 64      | Jefe(a) de Hogar     | Hombre | 60   | Si  | No |
| 17 | 10006 | 2            | .          | .                  | Cónyugue o pareja    | Mujer  | 48   | Si  | No |
| 18 | 10006 | 3            | .          | .                  | Hijo(a), hijastro(a) | Mujer  | 21   | No  | No |
| 19 | 10006 | 4            | .          | .                  | Hijo(a), hijastro(a) | Hombre | 12   | .   | No |
| 20 | 10006 | 5            | .          | .                  | Hijo(a), hijastro(a) | Hombre | 14   | .   | No |
| 21 | 10006 | 6            | .          | .                  | Hijo(a), hijastro(a) | Mujer  | 25   | No  | No |
| 22 | 10007 | Entrevistado | Mujer      | igual o mayor a 65 | Jefe(a) de Hogar     | Mujer  | 74   | No  | No |
| 23 | 10007 | 2            | .          | .                  | Hijo(a), hijastro(a) | Hombre | 56   | Si  | No |
| 24 | 10008 | Entrevistado | Mujer      | igual o mayor a 65 | Jefe(a) de Hogar     | Mujer  | 65   | No  | No |
| 25 | 10008 | 2            | .          | .                  | Hijo(a), hijastro(a) | Mujer  | 21   | No  | No |

Hogar 10002 de 5 personas, se entrevistó a la jefa de hogar de 37 años.

Pregunta a nivel de hogar, se repite la información para todos los miembros del hogar

ord=1 corresponde al entrevistado

# Bases de Datos Imputada



Contiene observaciones a nivel de hogar, los que se identifican a través de su id, para cada hogar está la información original ( $imp=0$ ), más las 30 imputaciones de las variables que pasaron por este proceso ( $imp=1, \dots, 30$ ).

# Bases de Datos Imputada

Información sin imputar a nivel de hogar

| id    | imp | genero_ent | tr_edad_ent   | ytotef    | t_dhip |
|-------|-----|------------|---------------|-----------|--------|
| 10056 | 0   | Mujer      | entre 50 y 64 | .         | 1      |
| 10056 | 1   | Mujer      | entre 50 y 64 | 2065250   | 1      |
| 10056 | 2   | Mujer      | entre 50 y 64 | 2131666.7 | 1      |
| 10056 | 3   | Mujer      | entre 50 y 64 | 2085833.3 | 1      |
| 10056 | 4   | Mujer      | entre 50 y 64 | 2148333.3 | 1      |
| 10056 | 5   | Mujer      | entre 50 y 64 | 2173333.3 | 1      |
| 10056 | 6   | Mujer      | entre 50 y 64 | 2098333.3 | 1      |
| 10056 | 7   | Mujer      | entre 50 y 64 | 2090000   | 1      |
| 10056 | 8   | Mujer      | entre 50 y 64 | 2068833.3 | 1      |
| 10056 | 9   | Mujer      | entre 50 y 64 | 2081666.7 | 1      |
| 10056 | 10  | Mujer      | entre 50 y 64 | 2315000   | 1      |
| 10056 | 11  | Mujer      | entre 50 y 64 | 2075833.3 | 1      |
| 10056 | 12  | Mujer      | entre 50 y 64 | 2072500   | 1      |
| 10056 | 13  | Mujer      | entre 50 y 64 | 2090000   | 1      |
| 10056 | 14  | Mujer      | entre 50 y 64 | 2148333.3 | 1      |
| 10056 | 15  | Mujer      | entre 50 y 64 | 2098333.3 | 1      |
| 10056 | 16  | Mujer      | entre 50 y 64 | 2094166.7 | 1      |
| 10056 | 17  | Mujer      | entre 50 y 64 | 2106666.7 | 1      |
| 10056 | 18  | Mujer      | entre 50 y 64 | 2190000   | 1      |
| 10056 | 19  | Mujer      | entre 50 y 64 | 2071666.7 | 1      |
| 10056 | 20  | Mujer      | entre 50 y 64 | 2073333.3 | 1      |
| 10056 | 21  | Mujer      | entre 50 y 64 | 2173333.3 | 1      |
| 10056 | 22  | Mujer      | entre 50 y 64 | 2070833.3 | 1      |
| 10056 | 23  | Mujer      | entre 50 y 64 | 2140000   | 1      |
| 10056 | 24  | Mujer      | entre 50 y 64 | 2106666.7 | 1      |
| 10056 | 25  | Mujer      | entre 50 y 64 | 2081666.7 | 1      |
| 10056 | 26  | Mujer      | entre 50 y 64 | 2081666.7 | 1      |
| 10056 | 27  | Mujer      | entre 50 y 64 | 2071666.7 | 1      |
| 10056 | 28  | Mujer      | entre 50 y 64 | 2165000   | 1      |
| 10056 | 29  | Mujer      | entre 50 y 64 | 2073333.3 | 1      |
| 10056 | 30  | Mujer      | entre 50 y 64 | 2100833.3 | 1      |

Se generan 30 imputaciones para esta variable que no tenía información en la base original

| id    | imp | genero_ent | tr_edad_ent   | ytotef  | t_dhip |
|-------|-----|------------|---------------|---------|--------|
| 10002 | 0   | Mujer      | entre 35 y 49 | 1480000 | 1      |
| 10002 | 1   | Mujer      | entre 35 y 49 | 1480000 | 1      |
| 10002 | 2   | Mujer      | entre 35 y 49 | 1480000 | 1      |
| 10002 | 3   | Mujer      | entre 35 y 49 | 1480000 | 1      |
| 10002 | 4   | Mujer      | entre 35 y 49 | 1480000 | 1      |
| 10002 | 5   | Mujer      | entre 35 y 49 | 1480000 | 1      |
| 10002 | 6   | Mujer      | entre 35 y 49 | 1480000 | 1      |
| 10002 | 7   | Mujer      | entre 35 y 49 | 1480000 | 1      |
| 10002 | 8   | Mujer      | entre 35 y 49 | 1480000 | 1      |
| 10002 | 9   | Mujer      | entre 35 y 49 | 1480000 | 1      |
| 10002 | 10  | Mujer      | entre 35 y 49 | 1480000 | 1      |
| 10002 | 11  | Mujer      | entre 35 y 49 | 1480000 | 1      |
| 10002 | 12  | Mujer      | entre 35 y 49 | 1480000 | 1      |
| 10002 | 13  | Mujer      | entre 35 y 49 | 1480000 | 1      |
| 10002 | 14  | Mujer      | entre 35 y 49 | 1480000 | 1      |
| 10002 | 15  | Mujer      | entre 35 y 49 | 1480000 | 1      |
| 10002 | 16  | Mujer      | entre 35 y 49 | 1480000 | 1      |
| 10002 | 17  | Mujer      | entre 35 y 49 | 1480000 | 1      |
| 10002 | 18  | Mujer      | entre 35 y 49 | 1480000 | 1      |
| 10002 | 19  | Mujer      | entre 35 y 49 | 1480000 | 1      |
| 10002 | 20  | Mujer      | entre 35 y 49 | 1480000 | 1      |
| 10002 | 21  | Mujer      | entre 35 y 49 | 1480000 | 1      |
| 10002 | 22  | Mujer      | entre 35 y 49 | 1480000 | 1      |
| 10002 | 23  | Mujer      | entre 35 y 49 | 1480000 | 1      |
| 10002 | 24  | Mujer      | entre 35 y 49 | 1480000 | 1      |
| 10002 | 25  | Mujer      | entre 35 y 49 | 1480000 | 1      |
| 10002 | 26  | Mujer      | entre 35 y 49 | 1480000 | 1      |
| 10002 | 27  | Mujer      | entre 35 y 49 | 1480000 | 1      |
| 10002 | 28  | Mujer      | entre 35 y 49 | 1480000 | 1      |
| 10002 | 29  | Mujer      | entre 35 y 49 | 1480000 | 1      |
| 10002 | 30  | Mujer      | entre 35 y 49 | 1480000 | 1      |

Este hogar tiene información para esta variable, no se generan imputaciones, se repite el valor original

# Bases de Factores Replicados

2021

DTA

Base Factores Replicados EFH 2021

3

>

DTA

Base imputada EFH 2021

>

DTA

Base no imputada EFH 2021

>

DTA

Llave Panel EFH

>

Para cada hogar (id) contiene el factor de expansión más las 1.000 replicas necesarias para el cálculo de los errores estándar mediante *Bootstrap weights*.



# Distribución de frecuencias



# Distribuciones de frecuencias

- Comenzaremos abriendo la base no imputada:

```
01_TallerUsoEFH X
1 *****TALLER USO DE EFH EN STATA*****
2
3 use "C:\Taller_EFH_STATA\02_Bases\Base no imputada EFH 2021.dta" ,clear
4
```

- Recordar que la EFH nos permite obtener estadísticas a nivel de hogar, así en el caso de la base no imputada que viene a nivel de individuos repitiendo la información del hogar para cada uno de sus miembros, debemos filtrar las estadísticas para `ord=1`.
- Además, recordar que debemos utilizar el factor de expansión para que estas frecuencias sean representativas de la población objetivo.
- Por ejemplo, obtengamos la siguiente tabla:

`tab e1 macrozona if ord==1 [w=factor], col nofreq`  
(frequency weights assumed)

*Existen diferencias importantes entre macrozonas*

| ¿Qué tipo de vivienda ocupa el hogar? | Macrozona |        |        |        | Total  |
|---------------------------------------|-----------|--------|--------|--------|--------|
|                                       | Norte     | Centro | Sur    | RM     |        |
| Casa                                  | 73.01     | 75.07  | 90.77  | 58.74  | 69.28  |
| Casa en cité                          | 0.19      | 0.12   | 0.12   | 1.15   | 0.59   |
| Casa en condominio                    | 3.99      | 2.66   | 1.34   | 4.36   | 3.43   |
| Departamento en edifi                 | 19.15     | 20.53  | 6.56   | 33.80  | 24.73  |
| Mediagua, mejora u ot                 | 2.17      | 1.19   | 0.54   | 0.23   | 0.80   |
| pieza(s) en casa o d                  | 0.87      | 0.19   | 0.53   | 0.78   | 0.58   |
| Pieza en casca antigu                 | 0.62      | 0.24   | 0.14   | 0.94   | 0.59   |
| Total                                 | 100.00    | 100.00 | 100.00 | 100.00 | 100.00 |

Filtra al entrevistado, de esta forma se calcula la distribución de frecuencia para los hogares.

**Un 69.3% de los hogares reside en casa y un 24.7% en departamento.**

# Distribuciones de frecuencias

- Otro ejemplo, es la pregunta G1, como el hogar califica su nivel de endeudamiento:

```
. tab g1 estrato if ord==1 [w=factor], col nofreq
(frequency weights assumed)
```

|                                                                                                            |                                                                  |           |           |        |
|------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------|-----------|-----------|--------|
| Tomando en<br>cuenta<br>todas las<br>deudas de<br>su hogar,<br>¿cómo<br>calificaría<br>a el nivel<br>de en | <i>Existen diferencias importantes según estrato de ingresos</i> |           |           |        |
|                                                                                                            | Estrato de ingreso total del hogar                               |           |           |        |
|                                                                                                            | Deciles 1                                                        | Deciles 6 | Deciles 9 | Total  |
| Excesivo                                                                                                   | 10.70                                                            | 8.01      | 7.30      | 8.87   |
| Alto                                                                                                       | 25.57                                                            | 19.28     | 20.98     | 22.21  |
| Moderado                                                                                                   | 41.99                                                            | 46.92     | 40.35     | 43.20  |
| Bajo                                                                                                       | 21.63                                                            | 25.77     | 31.24     | 25.63  |
| No responde                                                                                                | 0.05                                                             | 0.02      | 0.13      | 0.06   |
| No sabe                                                                                                    | 0.05                                                             | 0.00      | 0.00      | 0.02   |
| Total                                                                                                      | 100.00                                                           | 100.00    | 100.00    | 100.00 |

*Un 43.2% de los hogares califica su nivel de endeudamiento como moderado, y un 8.9% como excesivo.*

# Distribuciones de frecuencias

- Otro ejemplo, es la pregunta i1\_2, tenencia de inversión en fondos mutuos:

```
. tab i1_2 estrato if ord==1 [w=factor], col nofreq
(frequency weights assumed)
```

*Existen diferencias importantes según estrato de ingresos*

| Fondos<br>mutuos | Estrato de ingreso total del<br>hogar |           |           | Total  |
|------------------|---------------------------------------|-----------|-----------|--------|
|                  | Deciles 1                             | Deciles 6 | Deciles 9 |        |
| Si               | 2.54                                  | 7.98      | 21.38     | 7.93   |
| No               | 97.29                                 | 91.66     | 78.27     | 91.81  |
| No responde      | 0.08                                  | 0.10      | 0.13      | 0.10   |
| No sabe          | 0.09                                  | 0.25      | 0.22      | 0.17   |
| Total            | 100.00                                | 100.00    | 100.00    | 100.00 |

**Un 7.9% de los hogares posee inversiones en fondos mutuos.**

# Distribuciones de frecuencias

- En el caso de estar utilizando la base imputada que es a nivel de hogar, no es necesario ocupar el filtro de `ord=1`, sin embargo, recordar que esta base viene con 30 bases de imputaciones más la base original (`imp=0`), por lo cual, para obtener las estadísticas de frecuencias de variables no imputadas, por ejemplo, de tenencia de crédito hipotecario, debemos utilizar el siguiente código:
- Primero abriremos la base imputada:

```
use "C:\Taller_EFH_STATA\02_Bases\Base imputada EFH 2021.dta", clear
```

- Luego tabulamos la variable binaria “Tendencia de deuda hipotecaria” para la base original y utilizando el factor de expansión:

```
. tab t_dhip estrato if imp==0 [w=factor], col nofreq
(frequency weights assumed)
```

*Existen diferencias importantes según estrato de ingresos*

| Tenencia de deuda hipotecaria a total | Estrato de ingreso total del hogar |           |           | Total  |
|---------------------------------------|------------------------------------|-----------|-----------|--------|
|                                       | Deciles 1                          | Deciles 6 | Deciles 9 |        |
| 0                                     | 92.01                              | 80.02     | 53.03     | 80.64  |
| 1                                     | 7.99                               | 19.98     | 46.97     | 19.36  |
| Total                                 | 100.00                             | 100.00    | 100.00    | 100.00 |

Filtra la base original

*Un 19.4% de los hogares tiene deuda hipotecaria*

# Media de variables continuas sin imputar



# Media de variables sin imputaciones

- La mayor parte de las variables continuas asociadas a ingresos, deuda y activos pasan por un proceso de imputación, así, para obtener estadísticas representativas de la población, no sesgadas por la no respuesta, se deben utilizar las bases de imputación para obtener las estadísticas. Esto lo veremos en la siguiente sección.
- A continuación, se presentan la media de variables continuas que no se imputan.
- Por ejemplo, en la base imputada tenemos la variable `numh` que corresponde al tamaño del hogar:

```
. sum numh if imp==0 [w=factor]
(analytic weights assumed)
```

| Variable | Obs   | Weight  | Mean     | Std. dev. | Min | Max |
|----------|-------|---------|----------|-----------|-----|-----|
| numh     | 4,400 | 5616257 | 3.104823 | 1.476931  | 1   | 11  |

- También el número de ocupados en el hogar:

```
. sum ocuph if imp==0 [w=factor]
(analytic weights assumed)
```

| Variable | Obs   | Weight  | Mean     | Std. dev. | Min | Max |
|----------|-------|---------|----------|-----------|-----|-----|
| ocuph    | 4,400 | 5616257 | 1.487278 | .9633451  | 0   | 6   |



# Media de variables continuas con imputaciones



# Media de variables con imputaciones

- Por ejemplo, en el caso de la variable ingreso efectivo del hogar ( $y_{tote\text{f}}$ ), un 12% de los hogares no disponen de información para esta variable debido a la no respuesta en algunas de las variables utilizadas para construir el ingreso efectivo del hogar.
- Para esta variable se disponen de 30 imputaciones en caso de no respuesta del hogar.
- Para obtener la media de  $y_{tote\text{f}}$  con imputaciones se debe calcular el promedio de la variable (utilizando el factor de expansión) para cada una de las 30 bases de imputaciones ( $imp=1, \dots, 30$ ), y luego calcular el promedio simple de estos 30 promedios:

$$\overline{y_{tote\text{f}}}^{IMP} = \frac{1}{30} \sum_{i=1}^{30} \overline{y_{tote\text{f}}}_{imp=i}$$

- A continuación, se presenta el código para obtener la media imputada del ingreso efectivo del hogar.

# Media de variables con imputaciones

```
forvalues i=1(1)30 {  
    sum ytoteif if imp==`i' [w=factor]  
    g prom_ytoteif_`i'=r(mean)  
}  
  
egen prom_imp_ytoteif=rowmean(prom_ytoteif_1-prom_ytoteif_30)
```

```
. sum prom_imp_ytoteif
```

| Variable     | Obs     | Mean    | Std. dev. | Min     | Max     |
|--------------|---------|---------|-----------|---------|---------|
| prom_imp_y~f | 136,400 | 1612392 | 0         | 1612392 | 1612392 |

- El promedio del ingreso efectivo mensual del hogar es \$1.612.392.

# Media de variables con imputaciones

- También podemos calcular la media de la deuda hipotecaria asociada a la vivienda principal:

```
*Promedio de la deuda hipotecaria asociada a la vivienda principal (imputado)*
forvalues i=1(1)30 {
    sum d_vp if imp==`i' & d_vp>0 [w=factor]
    g prom_d_vp_`i'=r(mean)
}

egen prom_imp_d_vp=rowmean(prom_d_vp_1-prom_d_vp_30)

. sum prom_imp_d_vp
```

| Variable     | Obs     | Mean     | Std. dev. | Min      | Max      |
|--------------|---------|----------|-----------|----------|----------|
| prom_imp_d~p | 136,400 | 4.87e+07 | 0         | 4.87e+07 | 4.87e+07 |

```
. display r(mean)
48675448
```

- El valor promedio de la deuda hipotecaria asociada a la vivienda principal es de \$48.675.448.

# Mediana de variables continuas con imputaciones



# Mediana de variables con imputaciones

- Utilizando el mismo procedimiento presentado para la media, se puede obtener la mediana imputada.
- Es decir, calculamos la mediana de la variable para cada una de las 30 bases, y luego promediamos las 30 medianas para obtener la mediana imputada de la variable.

```

forvalues i=1(1)30 {
    sum ytoteif if imp==`i' [w=factor],d
    g mediana_ytoteif_`i'=r(p50)
}

egen mediana_imp_ytoteif=rowmean(mediana_ytoteif_1-mediana_ytoteif_30)

. sum mediana_imp_ytoteif

```

| Variable     | Obs     | Mean    | Std. dev. | Min     | Max     |
|--------------|---------|---------|-----------|---------|---------|
| mediana_im~f | 136,400 | 1138860 | 0         | 1138860 | 1138860 |

- La mediana del ingreso efectivo del hogar es \$1.138.860.

# Mediana de variables con imputaciones

- También podemos calcular la mediana de la deuda hipotecaria asociada a la vivienda principal

```

forvalues i=1(1)30 {
    sum d_vp if imp==`i' & d_vp>0 [w=factor],d
    g mediana_d_vp_`i'=r(p50)
}

egen mediana_imp_d_vp=rowmean(mediana_d_vp_1-mediana_d_vp_30)

. sum mediana_imp_d_vp

    Variable |      Obs      Mean   Std. dev.      Min      Max
-----+-----+-----+-----+-----+-----
 mediana_imp_d_vp |    136,400    3.37e+07         0    3.37e+07    3.37e+07

. display r(mean)
33659076

```

- La mediana es \$33.659.076.



# Error estándar



# Errores estándar en diseños complejos

- En un contexto de un diseño muestral complejo, como es el caso de la EFH, el cálculo del error estándar de los estimadores no se puede obtener bajo el supuesto de un muestreo aleatorio simple. Es necesario definir las variables del diseño (estratos y unidades de muestreo) para aplicar distintos métodos de estimaciones paramétricas.
- En el caso de la EFH, por temas de confidencialidad de la información muestral y resguardo de la identidad de los hogares entrevistados, no se publican las variables de diseño.
- Así, se utilizan técnicas no paramétricas, mediante réplicas de los factores de expansión, para la obtención de los errores estándar.
- Por lo que, dentro de la información publicada de la encuesta, no solo se pone a disposición del usuario el factor de expansión, sino 1000 réplicas de estos, los que son necesarios para el cálculo de los errores estándar mediante la técnica de *Bootstrap Weights*.

# Definición del diseño complejo (svyset)

- Primero, debemos pegar las 1.000 réplicas de los factores a la base imputada o la base no imputada, dependiendo de cuál estemos utilizando:

```
merge m:1 id using "C:\Taller_EFH_STATATA\02_Bases\Base Factores Replicados EFH 2021.dta"
```

- Luego, especificamos el diseño complejo con las réplicas de los factores:

```
. svyset [pw=factor], vce(bootstrap) bsrweight(factor_*)  
  
Sampling weights: factor  
                  VCE: bootstrap  
                  MSE: off  
Bootstrap weights: factor_1 .. factor_1000  
                  Single unit: missing  
                  Strata 1: <one>  
                  Sampling unit 1: <observations>  
                  FPC 1: <zero>
```

# Error estándar: variables no imputadas

- En el caso de variables que no tienen imputaciones, el error estándar de la media lo podemos obtener fácilmente utilizando los comandos `svy` de STATA, por ejemplo, si queremos calcular el error estándar del promedio del tamaño del hogar:

```
. svy: mean numh if imp==0
(running mean on estimation sample)
```

Bootstrap replications (1,000): .....10.....20.....30.....40.....50.....60.....70.....80.....90.....100.....110.....120.....130.....  
> ..140.....150.....160.....170.....180.....190.....200.....210.....220.....230.....240.....250.....260.....270.....280.....29  
> 0.....300.....310.....320.....330.....340.....350.....360.....370.....380.....390.....400.....410.....420.....430.....440...  
> .....450.....460.....470.....480.....490.....500.....510.....520.....530.....540.....550.....560.....570.....580.....590.....  
> ..600.....610.....620.....630.....640.....650.....660.....670.....680.....690.....700.....710.....720.....730.....740.....75  
> 0.....760.....770.....780.....790.....800.....810.....820.....830.....840.....850.....860.....870.....880.....890.....900...  
> .....910.....920.....930.....940.....950.....960.....970.....980.....990.....1,000 done

Survey: Mean estimation                      Number of obs    =    4,400  
                                                 Population size = 5,616,257  
                                                 Replications    =    1,000

|      | Observed<br>mean | Bootstrap<br>std. err. | Normal-based<br>[95% conf. interval] |          |
|------|------------------|------------------------|--------------------------------------|----------|
| numh | 3.104823         | .0277846               | 3.050366                             | 3.159279 |

medidas de tendencia central o posicionamiento, sino sólo el error estándar.

- En este caso el error estándar de la media del tamaño del hogar es 0.0278, con un intervalo de confianza que va entre 3.05 y 3.16.

# Error estándar: variables no imputadas

- También podemos calcular el error estándar de una proporción, por ejemplo, de la tendencia de deuda hipotecaria, variable que tampoco es imputada:

```
. svy: proportion t_dhip if imp==0
(running proportion on estimation sample)

Bootstrap replications (1,000): .....10.....20.....30.....40.....50.....60.....70.....80.....90.....100.....110.....120.....130.....
> ..140.....150.....160.....170.....180.....190.....200.....210.....220.....230.....240.....250.....260.....270.....280.....29
> 0.....300.....310.....320.....330.....340.....350.....360.....370.....380.....390.....400.....410.....420.....430.....440...
> .....450.....460.....470.....480.....490.....500.....510.....520.....530.....540.....550.....560.....570.....580.....590.....
> ..600.....610.....620.....630.....640.....650.....660.....670.....680.....690.....700.....710.....720.....730.....740.....75
> 0.....760.....770.....780.....790.....800.....810.....820.....830.....840.....850.....860.....870.....880.....890.....900...
> .....910.....920.....930.....940.....950.....960.....970.....980.....990.....1,000 done

Survey: Proportion estimation      Number of obs   =    4,400
                                   Population size = 5,616,257
                                   Replications   =    1,000
```

|             | Observed<br>proportion | Bootstrap<br>std. err. | Logit<br>[95% conf. interval] |          |
|-------------|------------------------|------------------------|-------------------------------|----------|
| t_dhip<br>0 | .8064497               | .0073087               | .791721                       | .8203732 |
| 1           | .1935503               | .0073087               | .1796268                      | .208279  |

- La tenencia de deuda hipotecaria es de 19.4, con un error estándar de 0.7. Así, el intervalo de confianza va desde 18.0 a 20.8.

# Error estándar: variables imputadas

- Recordemos que la media, la mediana o cualquier otra estadística de una variable imputada, se calcula como el promedio de las 30 estadísticas, cada una de ellas calculadas en cada una de las bases de imputaciones.
- Una vez que hemos calculado el estimador (por ejemplo, la media), debemos calcular su varianza (y su error estándar) de la siguiente forma:

$$T = W + \frac{31}{30} B$$

- Donde  $W$  es el componente *within* de la varianza y se obtiene promediando las 30 varianzas obtenidas para el estimador en cada una de las 30 bases:

$$W = \frac{1}{30} \sum_{i=1}^{30} \hat{V}_i$$

- Y donde  $B$  es el componente *between* de la varianza, y se obtiene calculando las desviaciones medias del estimador respecto a su promedio:

$$B = \frac{1}{29} \sum_{i=1}^{30} (\hat{Q}_i - \bar{Q})^2$$

# Error estándar: variables imputadas

```

matrix RES=J(30,2,0)

forvalues i=1(1)30 {
  svy: mean ytoteif if imp==`i'
  matrix R`i'=r(table)
  matrix RES[`i',1]=R`i'[1,1]
  matrix RES[`i',2]=R`i'[2,1]*R`i'[2,1]
}

clear
svmat RES

sum RES1
g double mean_ytoteif=r(mean)
sum RES2
g double W=r(mean)
g double Bi=(RES1-mean_ytoteif)^2
sum Bi
g double B=r(mean)*(30/29)
g double T=W+(31/30)*B
sum T
g double se_ytoteif=sqrt(r(mean))

```

Se define una matriz (RES) que guardara para cada una de las bases de imputaciones (filas), la medida de ytoteif (columna 1) y la varianza de la media de yototeif (columna 2), los que serán calculados utilizando las réplicas de los factores de expansión.

Para cada una de las bases (i=1 a 30) se utiliza la opción svy para calcular el promedio y el error estándar de ytoteif utilizando las 1.000 réplicas de los factores de expansión. Este comando guarda internamente la información en r(table), generamos la matriz R con todos estos resultados y rescatamos la media R[1,1] y el error estándar R[2,1], este último lo elevamos al cuadrado para obtener la varianza.

Se limpia la data y la matriz RES se transforma en una base de datos.

El promedio de RES1, corresponde a la media utilizando las imputaciones, el promedio de RES2 corresponde al promedio de las varianzas (componente within). Luego, se calculan las desviaciones de cada media respecto al promedio y se elevan al cuadrado. El promedio de estas se multiplica por 30/29 ya que en el promedio se pierde un grado de libertad. Este es el componente between. Finalmente se calcula la varianza y se obtiene la raíz cuadrada para obtener el error estándar.

. sum se\_ytoteif

| Variable   | Obs | Mean     | Std. dev. | Min      | Max      |
|------------|-----|----------|-----------|----------|----------|
| se_ytoteif | 30  | 24991.56 | 0         | 24991.56 | 24991.56 |

La ejecución de este código toma aproximadamente 11 minutos



# Anexos



# Macrozonas



## Estratos de Riqueza



# Unidad de Análisis y Entrevistado

- La unidad de muestreo y de análisis en la EFH corresponde al hogar
- En la EFH se define el hogar como:

*Un grupo de personas que viven en una misma vivienda y que comparten un presupuesto de alimentación.*

- Sólo una persona por hogar, el entrevistado, responde la encuesta por él y el resto de los miembros del hogar.
- Para determinar la persona a ser entrevistada se sigue un protocolo, por lo general resulta en que la persona entrevistada corresponde a aquel miembro del hogar de 18 años o más que se reconoce como la que tiene mayor conocimiento sobre las finanzas del hogar.

