

# Encuesta Anual de Hogares

## Diseño muestral

**Sergio Passamonti**

## Diseño Muestral

Dominios de estimación:

- ❖ Estimaciones Total CABA
- ❖ Estimaciones Total CABA - Inquilinatos, Hoteles familiares, Pensiones, Casas (inmuebles) usurpadas/tomadas
- ❖ Estimaciones Total CABA - Villas de Emergencia
- ❖ Estimaciones de cada una de las 15 Comunas

## Diseño Muestral (cont.)

La CABA esta organizada administrativamente en 15 Comunas.

Para satisfacer la demanda de información según los dominios de estimación, en cada Comuna se definieron 3 Marcos de muestreo:

- ❖ Viviendas particulares generales (Marco 1)
- ❖ Viviendas en IHPCT (Marco 2)
- ❖ Viviendas en villas de emergencia (Marco 3, solo presente en 6 Comunas)

## Diseño Muestral (cont.)

El diseño muestral es el mismo en cada uno de los 3 marcos en cada una de las Comunas.

Es un diseño muestral en 2 etapas:

- ❖ Primera Etapa: Selección de Unidades Primarias de muestreo (UP). La definición de UP difiere de un Marco al otro.
- ❖ Segunda Etapa: Dentro de cada UP seleccionada, se seleccionan las unidades secundarias (US), definiéndose como tal a las viviendas.

## **Diseño Muestral (cont.)**

### Viviendas Particulares Generales – Marco 1

Unidades Primarias (UP): Conjunto de Viviendas contiguas donde se trata que cada una de ellas tenga el mismo tamaño.

Las UP están agrupadas en 6 réplicas, donde año a año rotan en la muestra, las viviendas correspondientes a 2 de ellas es decir el 33 %.

Las UP fueron seleccionadas con probabilidad proporcional al tamaño siendo la medida de tamaño el total de viviendas (US) que las componen.

Las US se seleccionan mediante un muestreo sistemático.

## **Diseño Muestral (cont.)**

### Viviendas en IHPCT – Marco 2

Unidades Primarias (UP): listado de casas de “Inquilinatos, Hoteles familiares, Pensiones y Casas (inmuebles) usurpadas/tomadas” (IHPCT), considerándose cada una de ellas como UP

Año a año las UP son seleccionadas con probabilidad proporcional al tamaño siendo la medida de tamaño el total de habitaciones (US) de cada una de ellas.

Dentro de cada UP seleccionada, los encuestadores seleccionan las viviendas mediante un muestreo simple al azar sin reposición (con una tabla de números al azar), a partir del listado de “viviendas ocupadas”, actualizado por el mismo en el momento previo a la realización de la encuesta.

## **Diseño Muestral (cont.)**

### Viviendas en Villa de Emergencia – Marco 3

Debido a la imposibilidad -por razones de seguridad- de efectuar listados de actualización de sus viviendas, el marco está definido por la información del Censo Nacional de Población, Hogares y Vivienda 2010 (CNPHV2010).

Las UP son las manzanas de los “guías de censistas” del CNPHV2010. En cada comuna se selecciona una muestra de UP con probabilidad proporcional al total de viviendas de 2010.

En cada UP muestreada las viviendas son seleccionadas en campo por el coordinador que conduce a los encuestadores en el interior de las mismas, mediante un procedimiento aleatorio, debido a la necesidad de limitar el tiempo de permanencia del grupo de trabajo en cada UP.

## Diseño Muestral (cont.)

Marco de Muestreo	Tamaños Muestrales	
	UP	US
Marco 1	858	8580
Marco 2	274	550
Marco 3	45	450
Total	1177	9580

## Expansión de los Datos Muestrales

Una etapa crítica en toda encuesta por muestreo es la expansión de la información. Una selección adecuada, y el posterior desarrollo, de una metodología para ponderar los datos muestrales es importante, ya que debe atender simultáneamente varias necesidades

Una dificultad que se presenta una vez desarrollado el trabajo de campo es la existencia de algunos subgrupos de la población objetivos sub o sobre-representados. Las características de un diseño muestral complejo, como es el caso de la EAH, no permite muchas veces controlar esta situación ya que involucra probabilidades de inclusión distintas para las unidades de muestreo sumado a que las muestras son no autoponderadas.

## **Expansión de los Datos Muestrales (cont.)**

Como consecuencia de ello en la práctica es habitual corregir o calibrar los pesos o factores de expansión iniciales (las inversas de las probabilidades de selección) con la ayuda de información auxiliar conocida o preestablecida a partir de registros o fuentes externas.

Desde el año 2011 en la EAH se optó por llevar a cabo un ajuste a los pesos iniciales a través de la técnica de “calibración por marginales fijos” siguiendo la metodología desarrollada por Deville y Särndal [1992].

## Expansión de los Datos Muestrales (cont.)

El empleo adecuado de la técnica de calibración permite:

- ❖ Reducir el error cuadrático medio de las estimaciones.
- ❖ Corregir posibles sesgos en la etapa de selección.
- ❖ Dar una solución al problema de “no-respuesta”.
- ❖ Lograr, a partir de las restricciones impuestas, una deseable consistencia, o concordancia, entre los totales conocidos y los obtenidos a través de la muestra, de las variables auxiliares.

## **Expansión de los Datos Muestrales (cont.)**

La calibración se efectúa en forma independiente en cada combinación Comuna – Dominio, siendo éstos “Viviendas Particulares”, “Viviendas en IHPCT” y “Viviendas en Villa”. Este último en aquellas Comunas donde estén presentes.

Con respecto a la información auxiliar empleada para la calibración se recurre a totales de población obtenidos a partir de Proyecciones de Población generadas por el Departamento de Análisis Demográfico de la DGEyC.

## Expansión de los Datos Muestrales (cont.)

Se tuvieron en cuenta los siguientes totales:

- ❖ Total de personas por sexo
- ❖ Total de personas por grupos de edad, considerando los grupos: de 0 a 9 años, de 10 a 24 años, de 25 a 49 años, de 50 a 64 años, 65 años y más
- ❖ Totales de viviendas y hogares por Comuna, estimado a partir de los factores de expansión corregidos por no-respuesta.

En total se realizan 36 procesos de calibración

## Errores de Muestreo

Resumiendo, la EAH tiene las siguientes características:

- ❖ Diseño muestral complejo
- ❖ Los factores de expansión surgen de un ajuste a marginales fijos conocidos a través del empleo de estimadores por regresión generalizada de compleja formulación.

Ambas características son tenidas en cuenta a la hora del cálculo de los errores de muestreo.

## **Errores de Muestreo (cont.)**

En virtud de la innumerable cantidad de estimaciones que surgen del cruce de las variables o características de los distintos formularios, la presentación de los errores para cada una de ellas es prácticamente imposible aún cuando se está en condiciones de calcularlas o de disponerlas.

Esto llevó a buscar una manera que resuma y facilite a los usuarios de los datos, disponer de una medida del error muestral para estimaciones originadas de los datos de la encuesta empleando pocos recursos y no obligándolos a efectuar operaciones complejas, especialmente para aquellos con pocos conocimientos en estadística.

## Errores de Muestreo (cont.)

Para satisfacer los requerimientos de facilidad de cálculo, pocos recursos y síntesis, entre otros, se buscó aproximar los Coeficientes de Variación (CV) para un grupo importante de estimaciones de totales a partir del cruce de distintas variables a partir del empleo de una **F**unción **G**eneralizadora de **V**arianza (FGV).

La FGV es un modelo que relaciona en forma aproximada las estimaciones con sus errores estándar asociados. Como consecuencia el modelo permite, dada una estimación, obtener una aproximación a su CV y como alternativa construir tablas que permitan buscar estos valores.

## Errores de Muestreo (cont.)

Total estimado	CV	Total estimado	CV	Total estimado	CV
1000000	1,7	625000	2,1	60000	6,3
975000	1,7	600000	2,1	40000	7,6
950000	1,7	575000	2,1	20000	10,6
925000	1,7	550000	2,2	150000	4,1
900000	1,7	525000	2,2	100000	4,9
875000	1,8	500000	2,3	80000	5,5
850000	1,8	450000	2,4	60000	6,3
825000	1,8	400000	2,6	40000	7,6
800000	1,8	350000	2,7	20000	10,6
775000	1,9	300000	2,9	15000	12,1
750000	1,9	250000	3,2	10000	14,7
725000	1,9	200000	3,5	5000	20,5
700000	2,0	150000	4,1	4000	22,7
675000	2,0	100000	4,9	3000	26,1
650000	2,0	80000	5,5	2000	31,6