GOBIERNO DE LA CIUDAD DE BUENOS AIRES

DIRECCION GENERAL DE ESTADISTICA Y CENSOS

ESTIMADORES MUESTRALES DE LA ENCUESTA ANUAL DE HOGARES (EAH) DE 2004

Junio 2005

Índice de contenido

ESTIMADORES MAESTRALES DE LA ENCUESTA ANUAL DE HOGARES (EAH) DE 2004

ENCUESTA ANUAL DE HOGARES (EAH) DE LA DIRECCIÓN GENERAL DE ESTADÍSTICA Y CENSOS DEL GOBIERNO DE LA CIUDAD DE BUENOS AIRES

ESTIMADORES MAESTRALES DE LA ENCUESTA ANUAL DE HOGARES (EAH) DE 2004

La encuesta EAH es un estudio por muestreo de la población residente en las viviendas particulares de la ciudad de Buenos Aires, por lo que no incluye a la población de viviendas colectivas, ni tampoco a las personas que viven en la vía pública por su movilidad.
La Encuesta Anual de Hogares, como su nombre lo indica, se caracteriza por realizar un solo relevamiento en el año, que se lleva a cabo durante los meses de octubre a diciembre.
La encuesta anual se inició en el año 2002 y acumula ya tres relevamientos, 2002/3/4.

I- El procedimiento de muestreo y los dominios de resultados o de análisis de la encuesta

I-1- Descripción general del procedimiento de muestreo empleado

La muestra de la encuesta básicamente consiste en un muestreo de viviendas estratificado en 16 grandes estratos (o poblaciones) que son los 16 Centros de Gestión y Participación (CGP) en que halla organizada la ciudad, con subdivisión en cada uno de ellos según tipos de vivienda. Las viviendas de tipo "Inquilinato-Hotel-Pensión-Casa Tomada", las viviendas en "Villas de Emergencia" y el resto, designado como viviendas "Residenciales Generales", son las tres grandes categorías que definen las subpoblaciones de los CGP.

La subpoblación conformada por las viviendas "Particulares Generales" que comprende a las viviendas tipo casa, departamento, etc, cuantitativamente es por mucho el mayor de los tres, con más del 90% del total de las viviendas de cada CGP.

El relevamiento del 2004, que conceptualmente es semejante a los de 2002 y 2003, desde el punto de vista del muestreo ha recibido algunos cambios de tipo metodológico. Por una parte se ha modificado la unidad primaria de muestreo, que en los dos primeros años fue el segmento censal del CNP01, por unidades primarias mayores (con mayor cantidad de viviendas) y más estrechamente vinculadas con la división natural de la ciudad en manzanas. Cambio que alcanzó sólo a las viviendas particulares generales. Por otra parte se ha agregado, en cada CGP, un nuevo estrato, conformado por las viviendas (piezas, habitaciones) de inquilinato, hotel, pensión y casa tomada, con la finalidad de poder brindar un nuevo dominio de análisis, para la ciudad.

Estos cambios finalmente dieron lugar, en cada CGP, a cuatro marcos de muestreo. El marco 1, compuesto por las viviendas "particulares generales" cuyas unidades primarias son conglomerados de una manzanas o parte de manzana. El marco 2 de muestreo fue un listado de domicilios o casas de inquilinato, hoteles familiares, pensiones y casas tomadas, conformado con datos de diversas fuentes entre las cuales, obviamente, se hallaron los últimos dos censos de población. Un desprendimiento de unas muy pocas unidades primarias del marco 2 (grandes hoteles familiares) por facilidad de manejo constituyeron lo que se llamó el marco 3. Y finalmente el/las área/s de villas de emergencia de cada CGP, que conformaron el marco 4 de muestreo. Es de señalar que no todos los CGP tienen marcos 3 y 4.

El procedimiento de muestreo en el marco 1 consistió en un muestreo replicado, con seis replicas iguales e independientes. En cada replica la muestra fue seleccionada en dos etapas, donde las unidades de primera etapa se seleccionaron con probabilidad igual a la cantidad de viviendas, según el CNP01, con reposición, en tanto que la segunda etapa fue una selección de viviendas con igual probabilidad y sin reposición.

En el marco 2 la selección se efectuó también a dos etapas, pero sin replicación. La unidade de primera etapa fue el domicilio o casa de "inquilinato, hotel particular, pensión o casa tomada", las que se seleccionaron con probabilidad proporcional a la cantidad de viviendas (aproximadas) del mismo. La unidad de segunda etapa fue la vivienda ocupada, al momento de la encuesta. El listado de viviendas ocupadas (piezas) de cada unidad primaria seleccionada lo confeccionó, in situ, el mismo encuestador en su primera visita, quien también seleccionó, al azar y sin reposición, las viviendas a encuestar, para finalmente encuestar las mismas.

El marco 3, estuvo compuesto por unas pocas unidades de inclusión forzosa, desprendidas del marco 2, tal como se señaló arriba.

En el marco 4, compuesto por viviendas en villas de emergencia, también se utilizó un muestreo a dos etapas donde la unidad primaria fue la "unidad de relevamiento a cargo de un guía en el CNP01". La muestra se seleccionó con probabilidad proporcional a la cantidad de viviendas en cada unidad primaria, con reposición. La cantidad de viviendas seleccionadas en cada unidad primaria se mantuvo constante entre unidades primarias, excepto uno pocos casos particulares.

Finalmente corresponde señalar que en el marco 1, pese al esfuerzo realizado para extraer o separar del mismo los "inquilinatos, hoteles familiares, pensiones y casas tomadas" y pasarlos al marco 2, quedó en aquel una cierta cantidad de estos tipos de viviendas. De modo similar, pero a la inversa, ocurrió en los marcos 2 y 3 donde quedaron viviendas que deberían formar parte del marco 1, es decir, que siendo viviendas "particulares generales", no pudieron ser identificadas como tales y finalmente quedaron en el marco 2.
Esta aclaración apunta a comunicar que para efectuar las estimaciones específicamente referidas a las viviendas "particulares generales" se recurrió a la utilización de los marcos 1, 2 y 3 y, de igual modo, para efectuar las estimaciones específicas sobre inquilinatos, hoteles, pensiones y casas tomadas, también se combinaron las estimaciones de los tres marcos.

Las selecciones fueron mutuamente independientes entre marcos y, obviamente, entre CGP.

I-2- Los dominios de análisis de la EAH-2004, por CGP y para la ciudad en su conjunto

A partir de los cuatro marcos de muestreo mencionados, de los procedimientos de selección utilizados, de los estimadores empleados y obviamente de la información relevada será posible obtener resultados para el análisis de los cinco dominios o subpoblaciones de análisis siguientes:

<u>Dominio 1</u>- Suministra estimaciones o resultados para la totalidad de las viviendas particulares del CGP. Incluye viviendas "particulares generales", "viviendas de I-H-P-CT" y viviendas en villas de emergencia. Esto es, comprende todos los tipos de viviendas del CGP. Involucra por lo tanto a todos los tipos de viviendas de los cuatro marcos de muestreo arriba citados, (marcos 1+2+3+4).

<u>Dominio 2</u>- Da resultados para el conjunto de viviendas particulares del CGP, excluidas las viviendas de "villas de emergencia". Comprende todos los tipos de viviendas existentes en los marcos 1,2 y 3. Sólo se excluye el marco 4 de viviendas en villas.

<u>Dominio 3</u>- Brinda resultados para viviendas "particulares generales" del CGP, excluidas las viviendas de "I-H-P-CT" y de "villas de emergencia". Por lo tanto comprende todos los tipos de viviendas de los marcos 1,2 y 3 con exclusión de las viviendas de los tipos "I-H-P-CT".

<u>Dominio 4</u>- Aporta resultados para viviendas particulares de "I-H-P-CT", exclusivamente. Comprende únicamente a las viviendas de los tipos "I-H-P-CT", de los marcos 1, 2, 3.

<u>Dominio 5</u>- Suministra resultados referidos a las viviendas de villas de emergencia, exclusivamente. Comprende todos los tipos de viviendas existentes en las villas de emergencia, únicamente. Involucra sólo el marco 4.

Sin embargo los tamaños de muestra empleados por la EAH 2004 para relevar información sobre los dominios 4 y 5 en cada CGP son demasiado pequeñas como para intentar realizar, con la información resultante, algún análisis por CGP, por lo que los cálculos que se realicen con esta desagregación deben ser considerados sólo cálculos intermedios. El análisis de dichos dominios cobran sentido desde el punto de vista estadístico cuando se refieren a la totalidad de la ciudad. Por lo señalado, los dominios de análisis que puede ofrecer la EAH2004 para la ciudad y para cada CGP son los siguientes.

Dominios de análisis disponibles

Ciudad (total)	1, 2, 3, 4, 5
CGP	1, 2, 3

En total son 53 dominios los que puede brindar la EAH de la ciudad: 5 para el total ciudad, y 48 por CGP, a razón de 3 por cada uno de los 16 CGP en que se subdivide la ciudad.

Es posible además producir estimaciones para dominios especiales, como por ejemplo para un subconjunto de Centros de Gestión y Participación, o bien para áreas no necesariamente compuestas por CGPs completos. Sin embargo por lo señalado más arriba debe quedar excluido de esta posibilidad la producción de estimaciones que involucren los dominios 4 y 5.
Es importante señalar que esta posibilidades es una necesidad para la DGEyC a fin de brindar servicios de información, con los datos primarios de la misma EAH, para áreas especiales de la ciudad, en forma rápida, económica, y con una razonable precisión, para distintas necesidades del gobierno local.
En el punto III de este informe se detalla con mayor amplitud el tratamiento de los "dominios especiales".

I-3- Los datos de la encuesta EAH 2004 y su relación con la omisión censal de 2001

Marco muestral 1: las cifras que produce la EAH corrige la omisión de hogares y de personas que pudieron deslizarse en el CNP2001, pero no puede aportar corrección alguna sobre las viviendas omitidas ya que la selección de viviendas se efectúa sobre los listados censales del 2001. Recién cuando se disponga de nuevos listados de viviendas por unidad primaria las estimaciones de la muestra estará corrigiendo la eventual omisión censal de viviendas, que agregada a la corrección de hogares y población que ya está haciendo la EAH se estaría salvando la omisión censal del CNP01 en la ciudad.

Marcos muestrales 2 y 3: Estos marcos son listados de unidades primarias (UP) claramente definidas por lo que en ellos, por construcción, no existe la omisión de (UP), pudiendo si existir omisión censal de viviendas en el interior de las UP, pero como las mismas han sido listadas completamente, en el mismo momento de realización de las encuestas, se ha corregido así la omisión de viviendas ocupadas. En cuanto a la omisión censal de hogares y de personas en las viviendas ocupadas lo estaría corrigiendo la encuesta de viviendas, tal como ocurre en el marco 1. Por lo tanto, sobre las estimaciones que se producen con estos dos marcos no estarían afectadas por las omisiones que pudieran haberse producido en el CNP.

Marco 4 (villas de emergencia): es una situación similar a la del marco 1. La EAH04 corrige la omisión de hogares y de personas de las viviendas censadas, pero no puede corregir la omisión de viviendas del CNP01.

Seguidamente se explicitan los estimadores empleados para la onda 2004 de la EAH.

II- Estimadores maestrales de la Encuesta Anual de Hogares de 2004 (EAH-04)

Nomenclatura básica utilizada

Y_{jk} : valor que asume la variable Y en la jk-ésima US, o unidad de listado

${}_1Y_{jk}, {}_2Y_{jk}, {}_3Y_{jk}, {}_4Y_{jk}$: valor que toma la variable Y en la jk-ésima unidad de segunda etapa o de listado (vivienda), en los distintos marcos de muestreo (1, 2, 3, 4)

${}_1N$: total de unidades de primera etapa en el marco 1, similar para los marcos 2, 3, 4

${}_1M_j$: total de unidades de segunda etapa, en la j-ésima UP, del marco 1, y similar para los tres marcos restantes

${}_1M_o = \sum_{j}^{{}_1N} {}_oM_j$: total de unidades de listado (viviendas) en el marco 1. Similar en los tres marcos restantes.

${}_1Y^T = \sum_{J}^{{}_1N} \sum_{k}^{{}_1M_j} {}_1Y_{jk}$ agregado total de la variable en el marco 1. De modo similar ${}_2Y^T, {}_3Y^T, {}_4Y^T$ para los marcos 2, 3 y 4

${}_1n$: tamaño de la muestra seleccionada de ${}_1N$. Similar para los marcos 2 y 4

${}_1m_j$: tamaño de la muestra seleccionada de ${}_1M_j$. Similar para los marcos 2, 3 y 4

${}_1y_{jk}, {}_3y_{jk}, {}_3y_{jk}, {}_4y_{jk}$ valor que toma la variable Y en la jk-ésima unidad de la muestra, en los distintos marcos

${}_1\hat{Y}^T$: estimador muestral del agregado poblacional ${}_1Y^T$. Semejante par los marcos 2, 3 y 4

$s^2_{{}_1\hat{Y}_T}$, $\hat{s}^2_{{}_1\hat{Y}_T}$: variancia del estimador del TOTAL, y estimador muestral de dicha variancia. Igual para los demás marcos de muestreo

a : indicador de viviendas tipo "particular general": casa, departamento, etc. (ver aclaración en punto 1.1)

b : indicador de viviendas de "inquilinato, hotel familiar, pensión" (ver aclaración en punto 1.1)

c : indicador de viviendas de "casa tomada" (ver aclaración en punto 1.1)

${}_1Y^T_{(abc)}$ agregado de la variable Y en las viviendas de los tipos *a,b,c* del marco 1. De forma semejante se emplea para indicar los distintos tipos de viviendas, y en los distintos marcos.

II-1- Estimación del agregado TOTAL (Y^T) para la variable Y, en un CGP cualquiera, incluyendo las viviendas en villas de emergencia

Este total general del CGP, del cual participan también las viviendas en villas, se forma con la suma de las estimaciones de los cuatro marcos de muestreo 1, 2, 3, 4 .
La magnitud a estimar es

$$Y^T = {}_1Y^T + {}_2Y^T + {}_3Y^T + {}_4Y^T = {}_1Y^T_{(abc)} + {}_2Y^T_{(abc)} + {}_3Y^T_{(abc)} + {}_4Y^T_{(abc)} = Y^T_{(abc)}$$

Siendo los cuatro marcos poblaciones independientes con muestras también independientes, básicamente los estimadores del total Y^T serán

$$\hat{Y}^T = {}_1\hat{Y}^T + {}_2\hat{Y}^T + {}_3\hat{Y}^T + {}_4\hat{Y}^T \quad ; \quad \hat{s}^2_{\hat{Y}^T} = \hat{s}^2_{{}_1\hat{Y}^T} + \hat{s}^2_{{}_2\hat{Y}^T} + \hat{s}^2_{{}_3\hat{Y}^T} + \hat{s}^2_{{}_4\hat{Y}^T} \quad ; \quad \hat{CV} = \frac{\hat{s}^2_{\hat{Y}^T}}{\hat{Y}^T} * 100$$

1.1 Los estimadores ${}_1\hat{Y}^T$ y $\hat{s}^2_{{}_1\hat{Y}^T}$ (o también ${}_1\hat{Y}^T_{(abc)}$ y $\hat{s}^2_{{}_1\hat{Y}^T_{(abc)}}$)

En el marco 1 se realizó un muestreo replicado, con seis replicaciones independientes entre sí. En cada réplica se practicó un muestreo a dos etapas de selección. En la primera se trabajó con una selección con reposición y con probabilidad variable entre unidades primarias. El submuestreo consistió en una selección con igual probabilidad entre unidades secundarias (viviendas) y sin reposición.

Básicamente los estimadores para este procedimiento de selección son los usuales

$${}_1\hat{Y}^T = \frac{1}{l}\sum_l^L {}_1\hat{Y}^T_l = \frac{1}{6}\sum_l^6 ({}_1M_o \frac{1}{{}_1n}\sum_j^{{}_1n} \frac{1}{{}_1m_{lj}} \sum_k^{{}_1m_{lj}} {}_1y_{ljk}) = \frac{1}{6}\sum_l^6 (\frac{1}{{}_1n}\sum_j^{{}_1n} {}_1M_o \frac{1}{{}_1m_{lj}} \sum_k^{{}_1m_j} {}_1y_{jk}) =$$

$$= \frac{1}{6}\sum_l^6 (\frac{1}{{}_1n_1}\sum_j^{{}_1n} {}_1\hat{Y}^T_{l(j)}) = \frac{1}{6}\sum_l^6 {}_1\hat{Y}^T_l$$

donde

l : indicador de réplica l=1,...,6

${}_1M_o$ y ${}_1n$ son las mismas para las seis réplicas

La variancia muestral del estimador resulta

$$\hat{s}^2_{{}_1\hat{Y}^T} = \frac{1}{6^2}\sum_l^6 \hat{s}^2_{{}_1\hat{Y}^T_l} = \frac{1}{6^2}\sum_l^6 (\frac{1}{{}_1n_1(n-1)}\sum_j^{{}_1n} ({}_1\hat{Y}^T_{l(j)} - \frac{\sum_j^{{}_1n} {}_1\hat{Y}^T_{l(j)}}{{}_1n})^2)$$

A fin de poder llevar adelante las distintas estimaciones planteadas por la EAH 2004, en las cuales participan las unidades de este marco 1, resulta necesario distinguir cuatro categorías o tipos de unidades de listado, a saber:

a : US (viviendas) residenciales que no son: I-H-P-CT, viviendas colectivas ni viviendas móviles. Más precisamente, comprende: viviendas tipo 1,2, 5, 6, 8 (excluido casas tomadas).

b : viviendas de tipos: inquilinato, hotel o pensión, sin incluir viviendas usurpadas o tomadas.

c : viviendas usurpada, de cualquier tipo de vivienda

d : US que no son viviendas o que siéndolo se hallan excluidas de los objetivos de estudio de la EAH, como son: colectivas y móviles

Estas cuatro categorías son exhaustivas y excluyentes en el marco 1, por lo que una US pertenece a una y sólo una categoría, verificando en toda UP la igualdad

$${}_1m_j = {}_1m_{aj} + {}_1m_{bj} + {}_1m_{cj} + {}_1m_{dj}$$

Dado que las viviendas de la muestra pertenecientes a las categorías a y c fueron encuestadas, en tanto que aquellas pertenecientes a la b (I-H-P) no se encuestaron (sólo se registró la cantidad de viviendas ocupadas), es necesario introducir el siguiente ajuste al estimador del total

$${}_1\hat{Y}^T = {}_1\hat{Y}^T_{(abc)} = \frac{1}{6}\sum_l^6 \left(\frac{1}{{}_1n_l}\sum_j^{{}_1n_l} {}_1M_o \frac{1}{{}_1m_{lj}}\left(\sum_k^{{}_1m_{laj}} {}_1y_{lajk} + \sum_k^{{}_1m_{lbj}} {}_1y_{lbjk} + \sum_k^{{}_1m_{lcj}} {}_1y_{lcjk}\right) = \right.$$

$$= \frac{1}{6}\sum_l^6 \left(\frac{1}{{}_1n_l}\sum_j^{{}_1n_l} {}_1M_o \frac{1}{{}_1m_{lj}}\left(\sum_k^{{}_1m_{laj}} {}_1y_{lajk} + {}_1m_{lbj}\,{}_1\bar{y}_{lbj} + \sum_k^{{}_1m_{lcj}} {}_1y_{lcjk}\right)\right) = \frac{1}{6}\sum_1^6 \left(\frac{1}{{}_1n_l}\sum_j^{{}_1n_l} {}_1\hat{Y}^T_{l(j)}\right) \qquad \text{donde}$$

${}_1\bar{y}_{lbj}$ es el promedio aritmético simple de la variable, por unidad de listado, observado en las ${}_1m_{lbj}$ unidades secundarias de la muestra, perteneciente a la j-ésima UP, de la réplica l. Sin embargo dado que no se han relevado los valores de la variable en la categoría b, en lo sucesivo se sustituirá ${}_1\bar{y}_{lbj}$ por el promedio de la misma variable y categoría (b) del marco 2, del mismo CGP. Esto es, se lo sustituirá por el promedio el aritmético simple ${}_2\bar{y}_b$ (en rigor por ${}_2\bar{y}_{bc}$) del mismo CGP (para todo l), lo cual significa que se dará a las unidades b del marco 1 el mismo comportamiento de b en el marco 2 del mismo CGP.
Como sustituto del término ${}_1\bar{y}_{lbj}$ en rigor se utilizará el promedio por vivienda de las categorías (bc) (I-H-P-CT) del marco 2 donde prácticamente no existen casas tomadas por lo que ${}_2\bar{y}_{bc}$ prácticamente es igual a ${}_2\bar{y}_b$. Se lo calculará:

$${}_2\bar{y}_{bc} = \frac{1}{{}_2n^{\bullet}_{bc}}\sum_j^{{}_2n^{\bullet}_{bc}} \frac{1}{{}_2m_{(bc)j}}\sum_k^{{}_2m_{(bc)j}} {}_2y_{(bc)jk}$$

que es el promedio aritmético puro de la variable y en todas las viviendas encuestadas que pertenecen a las categorías (bc).

Además, dado que las viviendas de categoría b, en algunos casos, en realidad se encontraron más de una vivienda del tipo I-H-P, se introducirá una variable auxiliar ${}_1\bar{x}_{lbj}$ que se definirá

${}_1\bar{x}_{lbj}$ = cantidad promedio (por unidad de listado, seleccionadas y de reserva) de viviendas de I-H-P halladas en la unidad primaria correspondiente. Es decir: ${}_1\bar{x}_{lbj} = \frac{1}{{}_1DR_{lbj}} \sum_{k}^{{}_1DR_{lbj}} {}_1x_{lbjk}$

Finalmente, para dar participación en las estimaciones a las viviendas de reserva se introducirán los conceptos ya utilizados de

${}_1VO_{lj}$: cantidad de viviendas (ocupadas) tipos 1, 2, 3, 4, 5, 6, 7, 8 de la muestra seleccionada (únicamente); cuya razón de "no entrevista"/"entrevista" fuesen 7, 8, 9, 10 (viviendas ocupadas). Deberá verificarse que

$${}_1VO_{lj} = {}_1VO_{laj} + {}_1VO_{lbj} + {}_1VO_{lcj}$$

${}_1DR_{lj}$: cantidad de viviendas realizadas (seleccionadas y de reserva) tipos 1,2, 3, 4, 5, 6, 7, 8, cuya razón de "entrevista" fuese 10. Aquí también se verificará

$${}_1DR_{lj} = {}_1DR_{laj} + {}_1DR_{lbj} + {}_1DR_{lcj}$$

Con lo que el estimador tomaría la forma

$${}_1\hat{Y}^T = \frac{1}{6}\sum_{l}^{6}\left(\frac{1}{{}_1n_l}\sum_{j}^{{}_1n_l} {}_1M_o \frac{1}{{}_1m_{lj}}\left(\frac{{}_1VO_{laj}}{{}_1DR_{laj}}\sum_{k}^{{}_1DR_{laj}} {}_1y_{lajk} + {}_1m_{lbj}\,{}_1\bar{x}_{lbj}\,{}_2\bar{y}_b + \frac{{}_1VO_{lcj}}{{}_1DR_{lcj}}\sum_{k}^{{}_1DR_{lcj}} {}_1y_{lcjk}\right)\right) =$$

$$= \frac{1}{6}\sum_{l}^{6}\left(\frac{1}{{}_1n_l}\sum_{j}^{{}_1n_l}\left(\left({}_1M_o \frac{1}{{}_1m_{lj}}\frac{{}_1VO_{laj}}{{}_1DR_{laj}}\sum_{k}^{{}_1DR_{laj}} {}_1y_{lajk}\right) + \left({}_1M_o \frac{1}{{}_1m_{lj}}\,{}_1m_{lbj}\,{}_1\bar{x}_{lbj}\,{}_2\bar{y}_b\right) + \left({}_1M_o \frac{1}{{}_1m_{lj}}\frac{{}_1VO_{lcj}}{{}_1DR_{lcj}}\sum_{k}^{{}_1DR_{lcj}} {}_1y_{lcjk}\right)\right)\right) =$$

$$= \frac{1}{6}\sum_{1}^{6}\left(\frac{1}{{}_1n_l}\sum_{j}^{{}_1n_l}\left({}_1\hat{Y}^T_{la(j)} + \hat{Y}^T_{lb(j)} + \hat{Y}^T_{lc(j)}\right)\right) = \frac{1}{6}\sum_{1}^{6}\left(\frac{1}{{}_1n}\sum_{j}^{{}_1n_l} {}_1\hat{Y}^T_{l(abc)(j)}\right) \qquad (1)$$

Una vez calculados los valores ${}_1\hat{Y}_{l(j)}$ se calculará la variancia de cada réplica ($\hat{s}^2_{{}_1\hat{Y}^T_l}$), con los cuales a su vez, y del modo usual se calculará la variancia del total

$$\hat{s}^2_{{}_1\hat{Y}^T} = \frac{1}{6^2}\sum_{l}^{6}\hat{s}^2_{{}_1\hat{Y}^T_l} = \frac{1}{6^2}\sum_{l}^{6}\left(\frac{1}{{}_1n_l\left({}_1n_l - 1\right)}\sum_{j}^{{}_2n_l}\left({}_1\hat{Y}^T_{l(abc)(j)} - \frac{\sum_{j}^{{}_2n} {}_1\hat{Y}^T_{l(abc)(j)}}{{}_1n_l}\right)^2\right) \qquad (2)$$

1.2 Los estimadores ${}_2\hat{Y}^T$ y $\hat{s}^2_{{}_2\hat{Y}^T}$ (o también ${}_2\hat{Y}^T_{(abc)}$ y $\hat{s}^2_{{}_2\hat{Y}^T_{(abc)}}$)

En el marco 2 se llevó a cabo un muestreo similar al del marco 1, pero sin las replicaciones del primero, por lo que los estimadores esencialmente son semejantes a los precedentes. Esto es

$$ {}_2\hat{Y}^T = \frac{1}{{}_2n}\sum_{j}^{{}_2n}\frac{{}_2\hat{Y}_j^T}{{}_2P_j} = \frac{1}{{}_2n}\sum_{j}^{{}_2n}\frac{{}_2M_j^*\,{}_2\bar{y}_j}{{}_2P_j} = \frac{1}{{}_2n}\sum_{j}^{{}_2n}\frac{{}_2M_j^*}{{}_2P_j}\frac{1}{{}_2m_j}(\sum_{k}^{{}_2m_j}{}_2y_{jk}) = \frac{1}{{}_2n}\sum_{j}^{{}_2n}{}_2\hat{Y}_j^T = $$

$$ = \frac{1}{{}_2n}\sum_{j}^{{}_2n}\frac{{}_2M_j^*}{{}_2P_j}\frac{1}{{}_2m_j}(\sum_{k}^{{}_2m_{aj}}{}_2y_{ajk} + \sum_{k}^{{}_2m_{bj}}{}_2y_{bjk} + \sum_{k}^{{}_2m_{cj}}{}_2y_{cjk}) \qquad (1) $$

donde ${}_2M_j^*$ es el total de viviendas que se ha obtenido por recuento total de las mismas en cada UP de la muestra y también de las reservas. En esta estimación ${}_2n$ representa la muestra seleccionada en el Marco 2, que incluye también las UP de reserva utilizadas.

Se distinguen las mismas categorías anteriores. La categoría a es de escasa presencia en la muestra y dichas unidades no fueron encuestadas, por lo que no se dispone de sus datos.
Las categorías b y c fueron muestreadas en conjunto, en tanto que la d no corresponde a viviendas de interés para la EAH, o bien no son viviendas (comercios, inmuebles en demolición, etc).

Uniendo las categorías b y c y considerando la no encuesta de las viviendas de la categoría a, el estimador queda

$$ {}_2\hat{Y}^T = \frac{1}{{}_2n}\sum_{j}^{{}_2n}\frac{{}_2M_j^*}{{}_2P_j}\frac{1}{{}_2m_j}({}_2m_{aj}\,{}_2\bar{y}_{aj} + \sum_{k}^{{}_2m_{(bc)j}}{}_2y_{(bc)jk}) \qquad (2) $$

notar que una UP de este marco generalmente tendrá viviendas que pertenecerán a las categorías bc o, de no ser así, toda la UP será una sóla vivienda de la categoría a, en muy raras ocasiones la UP contendrá viviendas de las tres categorías. Cuando toda UP sea una sola vivienda de la categoría a entonces, en ese caso, la vivienda se la incluirá en el proceso de estimación como si la misma hubiese sido realizada (se la considerará en ${}_2n$) dándole el valor de ${}_2M_j^{\bullet} = 1$, el valor de ${}_2m_{aj} = 1$, y el valor de la variable, tal como se detalla abajo, será ${}_2\bar{y}_{aj} = {}_1\bar{y}_{aj}$, en tanto que ${}_2P_j$ es el valor original usado en el proceso de selección de la muestra.

En la expresión anterior se reemplazará el promedio por unidad de listado ${}_2\bar{y}_{aj}$ por el valor promedio de la misma variable, categoría y CGP $({}_1\bar{y}_a)$, valor que se lo calculará:

$$ {}_1\bar{y}_a = \frac{1}{6}\sum_{l}^{6}(\frac{1}{{}_1n_l}\sum_{j}^{{}_1n_l}\frac{1}{{}_1DR_{laj}}\sum_{k}^{{}_1DR_{laj}}{}_1y_{lajk}) = \frac{1}{6}\sum_{l}^{6}\frac{1}{{}_1n_l}\sum_{j}^{{}_1n_l}{}_1\bar{y}_{laj} $$

Corresponde también considerar que de la muestra de unidades primarias (${}_2n$), una parte serán UP del tipo IH-P (${}_2n_{(bc)}$), de las cuales a su vez se habrán encuestado efectivamente (${}_2n^*_{(bc)}$). Para corregir la falta de las viviendas "encuestables" que no han sido encuestadas en la estimación se les imputará a las mismas un valor de variable igual al promedio de las efectivamente encuestadas, en el mismo CGP. Introduciendo este ajuste la expresión (2) finalmente quedará

$$ {}_2\hat{Y}^T = \frac{1}{{}_2n}\sum_{j}^{{}_2n}\frac{{}_2M_j^*}{{}_2P_j}\frac{1}{{}_2m_j}({}_2m_{aj}\,{}_2\bar{y}_{aj}) + \frac{1}{{}_2n}\frac{{}_2n_{bc}}{{}_2n_{bc}^*}\sum_{j'}^{{}_2n_{bc}^*}\frac{{}_2M_{j'}^*}{{}_2P_{j'}}\frac{1}{{}_2m_{j'}}\sum_{k}^{{}_2m_{(bc)j}}{}_2y_{(bc)jk} = $$

$$ = \frac{1}{{}_2n}\sum_{j}^{{}_2n}\frac{{}_2M_j^*}{{}_2P_j}\frac{1}{{}_2m_j}({}_2m_{aj}\,{}_2\bar{y}_{aj}) + \frac{1}{{}_2n}\frac{{}_2n_{bc}}{{}_2n_{bc}^*}\sum_{j}^{{}_2n} z_j\frac{{}_2M_j^*}{{}_2P_j}\frac{1}{{}_2m_j}\sum_{k}^{{}_2m_{(bc)j}}{}_2y_{(bc)jk} = $$

donde

${}_2n_{(bc)}$: cantidad de UP encuestables, pertenecientes a las categorías I-H-P

${}_2n_{(bc)}^*$: cantidad de UP encuestables, pertenecientes a las categorías I-H-P, efectivamente encuestadas, y

z_j : variable auxiliar, con valores: $z_j = 1$ si j $\in {}_2n_{(bc)}^*$, y $z_j = 0$ si j $\notin {}_2n_{(bc)}^*$ para j=1,...,${}_2$n

Con lo cual la expresión anterior puede volverse a escribir

$$ = \frac{1}{{}_2n}\sum_{j}^{{}_2n}\frac{{}_2M_j^*}{{}_2P_j}\frac{1}{{}_2m_j}\left({}_2m_{aj}\,{}_2\bar{y}_{aj} + \frac{{}_2n_{bc}}{{}_2n_{bc}^*} z_j \sum_{k}^{{}_2m_{(bc)j}}{}_2y_{(bc)jk}\right) = \frac{1}{{}_2n}\sum_{j}^{{}_2n}{}_2\hat{Y}^T_{(abc)(j)} $$

Procediendo de este modo una vez calculados los valores ${}_2\hat{Y}_{(j)} = {}_2\hat{Y}^T_{(abc)(j)}$ se aproxima la estimación de la variancia empleando la fórmula usual

$$ \hat{s}^2_{{}_2\hat{Y}^T} = \frac{1}{{}_2n({}_2n-1)}\sum_{j}^{{}_2n}\left({}_2\hat{Y}^T_{(j)} - \frac{\sum_{j}^{{}_2n}{}_2\hat{Y}^T_{(j)}}{{}_2n}\right)^2 $$

NOTA: en el marco 2 el total de unidades que participaron de la muestra (seleccionadas y de reserva) se las considerará como seleccionadas, es decir formando parte de ${}_2n$. Las UP no realizadas con código de no entrevista 7, 8, 9 no se las incluye en ${}_2n$.

1.3 Los estimadores ${}_3\hat{Y}^T$ y $\hat{s}^2_{{}_3\hat{Y}^T}$ (o también ${}_3\hat{Y}^T_{(abc)}$ y $\hat{s}^2_{{}_3\hat{Y}^T_{(abc)}}$)

Este marco 3 es un derivado del marco 2 y está compuesto sólo por seis unidades primarias autorrepresentadas, distribuidas en 5 CGP, cuyas unidades de segunda etapa fueron submuestreadas con igual probabilidad y en una selección sin reposición.

Bajo estas condiciones los estimadores son

$$ {}_3\hat{Y}^T = \sum_{j}^{{}_3N_a} {}_3M^*_{aj} \frac{1}{{}_3m_{aj}} \sum_{k}^{{}_3m_{aj}} {}_3y_{ajk} + \sum_{j}^{{}_3N_b} {}_3M^*_{bj} \frac{1}{{}_3m_{bj}} \sum_{k}^{{}_3m_{bj}} {}_3y_{bjk} + \sum_{j}^{{}_3N_c} {}_3M^*_{cj} \frac{1}{{}_3m_{cj}} \sum_{k}^{{}_3m_{cj}} {}_3y_{cjk} = $$

$$ = \sum_{j}^{{}_3N_a} {}_3M^*_{aj}\, {}_3\bar{y}_{ajk} + \sum_{j}^{{}_3N_b} {}_3M^*_{bj}\, {}_3\bar{y}_{bjk} + \sum_{j}^{{}_3N_c} {}_3M^*_{cj}\, {}_3\bar{y}_{cjk} \qquad (1) $$

donde

${}_3N_a$ indica el total de UP autorrepresentadas de la categoría <u>a</u> en el marco 3 del CGP
$N = {}_3N_a + {}_3N_b + {}_3N_c$ total de UP autorrepresentadas en el marco 3, según categoría de vivienda
${}_3M^*_{aj}$ es el total de unidades de listado, también obtenidas por recuento en la UP aj-ésima

la misma nomenclatura se aplica a las categorías <u>b</u> y <u>c</u> .

Siendo cada UP una unidad en la cual la selección de las unidades de listado de la muestra se realizó en forma independiente de las restantes la variancia muestral del estimador será

$$ \hat{s}^2_{{}_3\hat{Y}^T} = \sum_{j}^{{}_3N_a} {}_3M^{*2}_{aj} \frac{({}_3M^*_{aj} - {}_3m_{aj})}{{}_3M^*_{aj}\, {}_3m_{aj}} \hat{S}^2_{{}_3y_{ajk}} + \sum_{j}^{{}_3N_b} {}_3M^{*2}_{bj} \frac{({}_3M^*_{bj} - {}_3m_{bj})}{{}_3M^*_{bj}\, {}_3m_{bj}} \hat{S}^2_{{}_3y_{bjk}} + \sum_{j}^{{}_3N_c} {}_3M^{*2}_{cj} \frac{({}_3M^*_{cj} - {}_3m_{cj})}{{}_3M^*_{cj}\, {}_3m_{cj}} \hat{S}^2_{{}_3y_{cjk}} $$

En la práctica cada CGP tiene uno o a lo sumo dos UP de este marco, con lo cual las viviendas pertenecerán a una sola categoría, por lo que la variancia se reducirá a un solo término.

Donde

$$ \hat{S}^2_{{}_3y_{ajk}} = \frac{1}{({}_3m_{aj} - 1)} \sum_{k}^{{}_3m_{aj}} \left({}_3y_{ajk} - \frac{\sum_{k}^{{}_3m_{aj}} {}_3y_{ajk}}{{}_3m_{aj}} \right)^2 $$ y semejante para $\hat{S}^2_{{}_3y_{bjk}}$ y $\hat{S}^2_{{}_3y_{cjk}}$

1.4 Los estimadores ${}_4\hat{Y}^T$ **y** $\hat{s}^2_{{}_4\hat{Y}^T}$ (o también ${}_4\hat{Y}^T_{(abc)}$ y $\hat{s}^2_{{}_4\hat{Y}^T_{(abc)}}$)

El marco de muestreo 4 de un CGP contiene sólo las viviendas situadas en villas de emergencia de dicho CGP. Dentro de este marco no se realizan estimaciones separadas para las categorías <u>a</u> y <u>bc</u> como en los tres marcos antes tratados dado que la pertenencia a villas de emergencia conforma en sí misma una categoría especial, más allá que en las villas también existen distintos tipos de viviendas. Por lo tanto, la magnitud o agregado a estimar alcanzará a todos los tipos de viviendas, esto es

$$ {}_4Y^T = \sum_{j}^{{}_4N} \sum_{k}^{{}_4M_j} {}_4Y_{jk} = \sum_{j}^{{}_4N} \sum_{k}^{{}_4M_j} {}_4Y_{(abc)jk} = {}_4Y^T_{(acb)} $$

En este dominio el procedimiento de muestreo empleado a sido un muestreo a dos etapas con probabilidades variables de selección de las unidades primarias (UP) y con restitución, y un muestreo simple al azar sin reposición de unidades de segunda etapa (US) (viviendas).

Se tomaron como UP a las áreas de trabajo de los "guías de relevamiento del CNP01", y las probabilidades de selección de las mismas fueron proporcionales a la cantidad de hogares censados en cada guía en el 2001.

El estimador del total bajo las condiciones descriptas resulta

$$ {}_4\hat{Y}^T = {}_4M_o \frac{1}{{}_4n}\sum_j^{{}_4n}\frac{1}{{}_4m_j}\sum_k^{{}_4m_{lj}} {}_4y_{jk} = \frac{1}{{}_4n}\sum_j^{{}_4n}\frac{{}_4M_o}{{}_4m_j}\sum_k^{{}_4m_{lj}} {}_4y_{jk} = \frac{1}{{}_4n}\sum_j^{{}_4n} {}_4\hat{Y}^T_{(j)} \quad \text{o también} $$

$$ {}_4\hat{Y}^T_{(abc)} = {}_4M_o \frac{1}{{}_4n}\sum_j^{{}_4n}\frac{1}{{}_4m_j}\sum_k^{{}_4m_{lj}} {}_4y_{(abc)jk} = \frac{1}{{}_4n}\sum_j^{{}_4n} {}_4\hat{Y}^T_{(abc)(j)} $$

y el estimador de la variancia del mismo resulta

$$ \hat{s}^2_{{}_2\hat{Y}^T_{(abc)}} = \frac{1}{{}_4n({}_4n-1)}\sum_j^{{}_4n}\left({}_4\hat{Y}^T_{(abc)(j)} - \frac{\sum_j^{{}_2n} {}_4\hat{Y}^T_{(abc)(j)}}{{}_4n}\right)^2 $$

II-2- Estimación del TOTAL (${}_rY^T$) para una variable Y, en el dominio de análisis de viviendas Particulares Generales, de un CGP

Este total se compone de las viviendas particulares generales de los tres marcos. Por lo tanto la cantidad a estimar es

$$ {}_rY^T = {}_{1r}Y^T + {}_{2r}Y^T + {}_{3r}Y^T $$

tal como se definieron las categorías, las viviendas particulares generales son todas aquellas que pertenecen a la categoría a por lo tanto la cantidad a estimar es

$$ {}_rY^T = {}_1Y_a^T + {}_2Y_a^T + {}_3Y_a^T = Y_a^T $$

cada uno de los cuales pertenece a un marco cuya muestra se seleccionó en forma independiente, por lo tanto el estimador y su variancia muestral serán

$$ {}_r\hat{Y}^T = {}_1\hat{Y}_a^T + {}_2\hat{Y}_a^T + {}_3\hat{Y}_a^T = \hat{Y}_a^T \qquad \hat{s}^2_{{}_r\hat{Y}^T} = \hat{s}^2_{{}_1\hat{Y}_a^T} + \hat{s}^2_{{}_2\hat{Y}_a^T} + \hat{s}^2_{{}_3\hat{Y}_a^T} $$

de II-1, apartados 1.1, 1.2 ,1.3 se extraen los estimadores para las subpoblaciones a con lo cual el estimador resulta

$$ {}_r\hat{Y}^T = {}_a\hat{Y}^T = \frac{1}{6}\sum_l^6\left(\frac{1}{{}_1n}\sum_j^{{}_1n} {}_1M_o \frac{1}{{}_1m_j}\left(\frac{{}_1VO_{laj}}{{}_1DR_{laj}}\right)\sum_k^{{}_1DR_{laj}} {}_1y_{lajk}\right) + \frac{1}{{}_2n}\sum_j^{{}_2n}\frac{{}_2M_j^*}{{}_2P_j}\frac{{}_2m_{aj}}{{}_2m_j}\,{}_1\bar{y}_a + \sum_j^{{}_3N_a} {}_3M^*_{aj}\,{}_3\bar{y}_{ajk} $$

notar que en la expresión de arriba ${}_1\bar{y}_a$ es el promedio general de la variable en el marco 1, para la misma categoría y CGP. Más específicamente, es el promedio de todas las viviendas del CGP, del marco 1, pertenecientes a la categoría a, al cual se lo calculará:

$$ {}_1\bar{y}_a = \frac{1}{6}\sum_{l}^{6}\left(\frac{1}{{}_1n_l}\sum_{j}^{{}_1n_l}\frac{1}{{}_1DR_{laj}}\sum_{j}^{{}_1DR_{laj}}{}_1y_{lajk}\right) = \frac{1}{6}\sum_{l}^{6}\frac{1}{{}_1n_l}\sum_{j}^{{}_1n_l}{}_1\bar{y}_{laj} $$

por lo tanto el segundo sumando de la expresión de arriba puede escribírselo

$$ \frac{1}{{}_2n}\sum_{j}^{{}_2n}\frac{{}_2M_j^*}{{}_2P_j}\frac{{}_2m_{aj}}{{}_2m_j}\,{}_1\bar{y}_a = \frac{1}{{}_2n}\sum_{j}^{{}_2n}{}_2\hat{Y}_{a(j)}^T $$

con lo cual resulta

$$ {}_r\hat{Y}^T = {}_a\hat{Y}^T = \frac{1}{6}\sum_{l}^{6}\left(\frac{1}{{}_1n_l}\sum_{j}^{{}_1n_l}{}_1\hat{Y}_{la(j)}^T\right) + \frac{1}{{}_2n}\sum_{j}^{{}_2n}{}_2\hat{Y}_{a(j)}^T + \sum_{j}^{{}_3N_a}{}_3M_{aj}^*\,{}_3\bar{y}_{ajk} \tag{1} $$

siendo la variancia muestral del estimador, dado que se trata de la suma de tres variables independientes

$$ \hat{s}^2_{{}_r\hat{Y}^T} = \frac{1}{6^2}\sum_{l}^{6}\left(\frac{1}{{}_1n_l({}_1n_l-1)}\sum_{j}^{{}_1n_l}\left({}_1\hat{Y}_{la(j)}^T - \frac{\sum_{j}^{{}_1n_l}{}_1\hat{Y}_{la(j)}^T}{{}_1n_l}\right)^2\right) + \frac{1}{{}_2n({}_2n-1)}\sum_{l}^{{}_2n}\left({}_2\hat{Y}_{a(j)}^T - \frac{\sum_{j}^{{}_2n}{}_2\hat{Y}_{a(j)}^T}{{}_2n}\right)^2 + $$

$$ + \sum_{j}^{{}_3N_a}{}_3M_{aj}^{*2}\,\frac{({}_3M_{aj}^* - {}_3m_{aj})}{{}_3M_{aj}^*\,{}_3m_{aj}}\,\hat{S}^2_{{}_3y_{ajk}} \tag{2} $$

ver en 1.3 la variancia $\hat{S}^2_{{}_3y_{ajk}}$

II-3- Estimación del TOTAL (${}_qY^T$) para una variable Y, en el dominio de análisis de viviendas de: Inquilinato, Hoteles familiares, Pensiones y Casas Tomadas (usurpadas), de un CGP

Igual que en el punto anterior, el total a estimar se forma por la suma de las viviendas de I-H-P-CT de los tres marcos. Por lo tanto la cantidad a estimar es

$$ {}_qY^T = {}_{1q}Y^T + {}_{2q}Y^T + {}_{3q}Y^T $$

tal como se definieron las categorías, las viviendas en I-HF-P-CT son todas aquellas que pertenecen a las categorías b y c (b+c), por lo tanto la cantidad a estimar es

$$ {}_qY^T = {}_1Y_{bc}^T + {}_2Y_{bc}^T + {}_3Y_{bc}^T = Y_{bc}^T $$

cada uno de los cuales pertenece a un marco cuya muestra se seleccionó en forma independiente, por lo tanto el estimador y su variancia muestral serán

$$ {}_q\hat{Y}^T = {}_1\hat{Y}_{bc}^{\ T} + {}_2\hat{Y}_{bc}^{\ T} + {}_3\hat{Y}_{bc}^{\ T} = \hat{Y}_{bc}^{\ T} \qquad \hat{s}^{\,2}_{{}_q\hat{Y}^T} = \hat{s}^{\,2}_{{}_1\hat{Y}^T_{bc}} + \hat{s}^{\,2}_{{}_2\hat{Y}^T_{bc}} + \hat{s}^{\,2}_{{}_3\hat{Y}^T_{bc}} $$

de 1.1, 1.2 ,1.3 se extraen los estimadores para las subpoblaciones <u>bc</u> con lo cual el estimador de interés resulta

$$ {}_q\hat{Y}^T = \hat{Y}^T_{(bc)} = \frac{1}{6}\sum_l^6 \Big(\frac{1}{{}_1n_l}\sum_j^{{}_1n_l} {}_1M_o \frac{1}{{}_1m_j}\big({}_1m_{lbj}\,{}_1\bar{x}_{lbj}\,{}_2\bar{y}_b + \big(\frac{{}_1VO_{lcj}}{{}_1DR_{lcj}}\big)\sum_k^{{}_1DR_{lcj}} {}_1y_{lcjk}\big)\Big) + $$

$$ + \Big(\frac{1}{{}_2n}\frac{{}_2n_{bc}}{{}_2n^*_{bc}}\sum_j^{{}_2n} z_j \frac{{}_2M^*_j}{{}_2P_j}\frac{1}{{}_2m_j}\sum_k^{{}_2m_{(bc)j}} {}_2y_{(bc)jk}\Big) + \Big(\sum_j^{{}_3N_b} {}_3M^*_{bj}\,{}_3\bar{y}_{bjk} + \sum_j^{{}_3N_c} {}_3M^*_{cj}\,{}_3\bar{y}_{cjk}\Big) $$

${}_2\bar{y}_b$ es el promedio por vivienda de la categoría <u>b</u> (I-H-P-CT), del mismo CGP en el marco 2. En rigor será ${}_2\bar{y}_{bc}$ dado que en el marco 2 se estiman conjuntamente ambas categorías, pero donde la <u>c</u> prácticamente es inexistente. Por lo tanto resultará

$$ {}_2\bar{y}_b = \frac{1}{{}_2n^{\bullet}_{bc}}\sum_j^{{}_2n^{\bullet}_{bc}} \frac{1}{{}_2m_{(bc)j}}\sum_k^{{}_2m_{(bc)j}} {}_2y_{(bc)jk} $$

que es el promedio aritmético de la variable en todas las viviendas encuestadas. Finalmente el estimador resulta

$$ {}_r\hat{Y}^T = \hat{Y}^T_{(bc)} = \frac{1}{6}\sum_l^6\Big(\frac{1}{{}_1n}\sum_j^{{}_1n} {}_1\hat{Y}^T_{l(bc)(j)}\Big) + \frac{1}{{}_2n}\sum_j^{{}_2n} {}_2\hat{Y}^T_{(bc)(j)} + \sum_j^{{}_3N_b} {}_3M^*_{bj}\,{}_3\bar{y}_{bjk} + \sum_j^{{}_3N_c} {}_3M^*_{cj}\,{}_3\bar{y}_{cjk} $$

dado que se trata de tres variables que provienen de marcos y de muestras independientes, la variancia muestral del estimador es

$$ \hat{s}^{\,2}_{{}_1\hat{Y}^T} = \frac{1}{6^2}\sum_l^6\Big(\frac{1}{{}_1n_l({}_1n_l-1)}\sum_j^{{}_1n_l}\Big({}_1\hat{Y}^T_{l(bc)(j)} - \frac{\sum_j^{{}_1n_l} {}_1\hat{Y}^T_{l(bc)(j)}}{{}_1n_l}\Big)^2\Big) + \frac{1}{{}_2n({}_2n-1)}\sum_l^{{}_2n}\Big({}_2\hat{Y}^T_{(bc)(j)} - \frac{\sum_j^{{}_2n} {}_2\hat{Y}^T_{(bc)(j)}}{{}_2n}\Big)^2 + $$

$$ + \sum_j^{{}_3N_b} {}_3M^{*2}_{bj}\frac{({}_3M^*_{bj} - {}_3m_{bj})}{{}_3M^*_{bj}\,{}_3m_{bj}}\hat{S}^{\,2}_{{}_3y_{bjk}} + \sum_j^{{}_3N_c} {}_3M^{*2}_{cj}\frac{({}_3M^*_{cj} - {}_3m_{cj})}{{}_3M^*_{cj}\,{}_3m_{cj}}\hat{S}^{\,2}_{{}_3y_{cjk}} $$

Ver en 1.3 las variancias operativas $\hat{S}^2_{{}_3y_{bjk}}$ y $\hat{S}^2_{{}_3y_{cjk}}$

II-4-Estimador de la RAZON combinada de los totales de dos variables ($R = \dfrac{Y^T}{X^T}$ **)**

4.1 La estimación de la Razón para las categorías de viviendas (a,b,c) (a+b+c),en un CGP cualquiera, comprendiendo los cuatro marcos maestrales. Dominio de análisis D1.

El programa Encuesta Anual de Hogares (EAH) de la DGEyC esencialmente es una encuesta de viviendas de la ciudad que se efectúa a una muestra de las mismas. Las viviendas de un CGP a los efectos del muestreo fueron organizadas en cuatro marcos de muestreo tal como se describen abajo y en cada uno de ellos se aplicó un procedimiento de muestreo conveniente. Los estimadores siguientes se refieren o cubren los cuatro marcos , sin embargo es de señalar que en todos los CGP existen los marcos 1 y 2 pero no necesariamente los marcos 3 y 4.

Marco 1: viviendas particulares generales.
Marco 2: listado de Inquilinatos, Hoteles, Pensiones, Casas tomadas (I-H-P-CT).
Marco 3: listado de I-H-P de inclusión forzosa, que por simplicidad se las trata como un marco más.
Marco 4: viviendas en villas de emergencia. NOTA: en este marco no se diferencian los tipo de viviendas por lo que en (abc) se las estará incluyendo a todas.

La característica a estimar es la razón entre dos variables en el dominio 1 (D1) del CGP

$$R=\frac{Y^T}{X^T}=\frac{Y^T_{(abc)}}{X^T_{(abc)}}=\frac{{}_1Y^T_{(abc)}+{}_2Y^T_{(abc)}+{}_3Y^T_{(abc)}+{}_4Y^T_{(abc)}}{{}_1X^T_{(abc)}+{}_2X^T_{(abc)}+{}_3X^T_{(abc)}+{}_4X^T_{(abc)}}=R_{(abc)}$$ el estimador muestral de R será

$$\hat{R}=\frac{\hat{Y}^T}{\hat{X}^T}=\frac{\hat{Y}^T_{(abc)}}{\hat{X}^T_{(abc)}}=\frac{{}_1\hat{Y}^T_{(abc)}+{}_2\hat{Y}^T_{(abc)}+{}_3\hat{Y}^T_{(abc)}+{}_4\hat{Y}^T_{(abc)}}{{}_1\hat{X}^T_{(abc)}+{}_2\hat{X}^T_{(abc)}+{}_3\hat{X}^T_{(abc)}+{}_4\hat{X}^T_{(abc)}}=\hat{R}_{(abc)}$$ donde los estimadores de los

totales del numerador y del denominador se hallan en 1.1, 1.2, 1.3 y 1.4

La expresión general de la variancia de R es la siguiente

$$s^2_{\hat{R}_{(abc)}}=\frac{1}{(X^T_{(abc)})^2}\sum_{i=1}^{4}s^2_{({}_i\hat{Y}^T_{(abc)}-R_i\hat{X}^T_{(abc)})}=\sum_{i=1}^{4}\left(\frac{1}{(X^T_{(abc)})^2}s^2_{({}_i\hat{Y}^T_{(abc)}-R_i\hat{X}^T_{(abc)})}\right) \qquad (1)$$

donde i=1,2,3,4 indica "marco de muestreo", y cada sumando de la última expresión representa el aporte que le llega desde cada marco a la variancia del estimador.

El estimador muestral de dicha variancia es

$$\hat{s}^2_{\hat{R}_{(abc)}}=\frac{1}{(\hat{X}^T_{(abc)})^2}\sum_{i=1}^{4}\hat{s}^2_{({}_i\hat{Y}^T_{(abc)}-\hat{R}_i\hat{X}^T_{(abc)})} \qquad (2)$$

La variancia $\hat{s}^2_{({}_i\hat{Y}^T_{(abc)}-\hat{R}_i\hat{X}^T_{(abc)})}$ para cada marco se detalla en los puntos 4.1.1, 4.1.2, 4.1.3 y 4.1.4 siguientes.

4.1.1 Marco 1

La variancia muestral de la expresión de arriba para el marco 1, con muestreo replicado, resulta

$$\hat{s}^2_{({}_1\hat{Y}^T_{(abc)}-\hat{R}\,{}_1\hat{X}^T_{(abc)})} = \frac{1}{6^2}\sum_{l}^{6}\left(\frac{1}{{}_1n_l}\left(\hat{S}^2_{{}_1\hat{Y}^T_{l(abc)(j)}} + \hat{R}^2\hat{S}^2_{{}_1\hat{X}^T_{l(abc)(j)}} - 2\hat{R}\hat{S}_{{}_1\hat{Y}^T_{l(abc)(j)},\,{}_1\hat{X}^T_{l(abc)(j)}}\right)\right)$$

donde l indica réplica: l=1,2,...,6, y $\hat{R}$ en realidad es la razón $\hat{R}_{(abc)}$ calculada con los cuatro marcos tal como se definió arriba y que por simplicidad se lo escribió sin subíndice.

La cuasi covariancia muestral es

$$\hat{S}_{{}_1\hat{Y}^T_{l(abc)(j)},\,{}_1\hat{X}^T_{l(abc)(j)}} = \frac{1}{{}_1n_l - 1}\sum_{j}^{{}_1n_l}\left({}_1\hat{Y}^T_{l(abc)(j)} - \frac{\sum_{j}^{{}_1n_j}{}_1\hat{Y}^T_{l(abc)(j)}}{{}_1n_j}\right)\left({}_1\hat{X}^T_{l(abc)(j)} - \frac{\sum_{j}^{{}_1n_j}{}_1\hat{X}^T_{l(abc)(j)}}{{}_1n_j}\right)$$

donde ${}_1\hat{Y}^T_{l(abc)(j)}$ y ${}_1\hat{X}^T_{l(abc)(j)}$ son los totales de viviendas (abc) estimados para la réplica l, desde la sola j-ésima UP.

4.1.2 Marco 2

En este marco la variancia muestral es

$$\hat{s}^2_{({}_2\hat{Y}^T_{(abc)}-\hat{R}\,{}_2\hat{X}^T_{(abc)})} = \frac{1}{{}_2n}\left(\hat{S}^2_{{}_2\hat{Y}^T_{(abc)(j)}} + \hat{R}^2\hat{S}^2_{{}_2\hat{X}^T_{(abc)(j)}} - 2\hat{R}\hat{S}_{{}_2\hat{Y}^T_{(abc)(j)},\,{}_2\hat{X}^T_{(abc)(j)}}\right)$$ donde es

$$\hat{S}_{{}_2\hat{Y}^T_{(abc)(j)},\,{}_2\hat{X}^T_{(abc)(j)}} = \frac{1}{{}_2n-1}\sum_{j}^{{}_1n}\left({}_2\hat{Y}^T_{abc)(j)} - \frac{\sum_{j}^{{}_2n}{}_2\hat{Y}^T_{(abc)(j)}}{{}_2n-1}\right)\left({}_2\hat{X}^T_{(abc)(j)} - \frac{\sum_{j}^{{}_2n}{}_2\hat{X}^T_{(abc)(j)}}{{}_2n-1}\right)$$

donde a su vez ${}_2\hat{Y}^T_{(abc)(j)}$ e ${}_2\hat{X}^T_{(abc)(j)}$ representan los totales de ambas variables, estimados desde la sola j-ésima UP

4.1.3 Marco 3

En este marco las UP fueron de inclusión forzosa con submuestreo en cada una. Submuestreo que, del modo usual, es independiente entre UP. La expresión puede escribirse

$$\hat{s}^2_{({}_3\hat{Y}^T_{(abc)}-\hat{R}\,{}_3\hat{X}^T_{(abc)})} = \hat{s}^2_{(\sum_{j}^{{}_3N}{}_3\hat{Y}^T_{(abc)j} - \hat{R}\sum_{j}^{{}_3N}{}_3\hat{X}^T_{(abc)j})}$$

donde ${}_3N$ representa ahora el total de UP en el marco, en tanto que con j se indica a cada una de ellas. Por tratarse de muestras independientes la fórmula anterior puede escribirse del siguiente modo

$$\hat{s}^2_{({}_3\hat{Y}^T_{(abc)}-\hat{R}_3\hat{X}^T_{(abc)})} = \hat{s}^2_{(\sum_j^{{}_3N} {}_3\hat{Y}^T_{(abc)j} - \hat{R}\sum_j^{{}_3N} {}_3\hat{X}^T_{(abc)j})} = \hat{s}^2_{(\sum_j^{{}_3N}({}_3\hat{Y}^T_{(abc)j}-\hat{R}_3\hat{X}^T_{(abc)j}))} = \sum_j^{{}_3N} \hat{s}^2_{({}_3\hat{Y}^T_{(abc)j}-\hat{R}_3\hat{X}^T_{(abc)j}))}$$

y por tratarse de submuestreo simple al azar sin reposición en cada UP la variancia será

$$\sum_j^{{}_3N} \hat{s}^2_{({}_3\hat{Y}^T_{(abc)j}-\hat{R}_3\hat{X}^T_{(abc)j}))} = \sum_j^{{}_3N} {}_3M_j^{*2} \frac{{}_3M_j^{*} - {}_3m_j}{{}_3M_j^{*}} \frac{1}{{}_3m_j} (\hat{S}^2_{{}_3y_{(abc)jk}} + \hat{R}^2 \hat{S}^2_{{}_3x_{(abc)jk}} - 2\hat{R}\hat{S}_{{}_3y_{(abc)jk},{}_3x_{(abc)jk}})$$

siendo la estimación muestral de la covariancia operativa o cuasi covariancia

$$\hat{S}_{{}_3y_{(abc)jk},{}_3x_{(abc)jk}} = \frac{1}{({}_3m_j - 1)} \sum_k^{{}_3m_j} \left({}_3y_{(abc)jk} - \frac{\sum_k^{{}_3m_j} {}_3y_{(abc)jk}}{{}_3m_j}\right)\left({}_3x_{(abc)jk} - \frac{\sum_k^{{}_3m_j} {}_3x_{(abc)jk}}{{}_3m_j}\right)$$

mientras que $\hat{R}$ es la razón definida al principio de este apartado 4.

4.1.4 Marco 4

Para el marco 4 el procedimiento de muestreo, y consecuentemente las estimaciones, se realizan de forma similar a las del marco 2, y por lo tanto las expresiones de cálculo pueden tomarse de dicho punto 4.1.2.

4.1.5 Uso de las expresiones anteriores para estimar una razón combinada en el ámbito de los dominios D2 y D5

Para estimar en el dominio D2: se aplican las mismas fórmulas de los puntos anteriores con la sola eliminación del marco 4 de muestreo. Es decir, se quitan todos los términos que se refieren a este marco.

Para estimar en el dominio D5: también se aplican las mismas fórmulas de los puntos anteriores con la eliminación de los marcos de muestreo 1 ,2 y 3. Es decir, se trabaja sólo con el marco 4 que corresponde a viviendas en villas de emergencia.

4.2. Estimador de la RAZON de los totales de dos variables en el conjunto de viviendas de categoría (a), en un CGP cualquiera, en los marcos 1, 2 y 3. Se excluye villas de emergencia.

Se excluye el marco 4 dado que en el mismo no se diferencian los tipos de viviendas.
Notar que se trata de la estimación de una razón en el dominio 3 de análisis (D3) del CGP.
La característica a estimar es

$$ {}_rR = \frac{{}_rY^T}{{}_rX^T} = \frac{Y^T_{(a)}}{X^T_{(a)}} = \frac{{}_1Y^T_{(a)} + {}_2Y^T_{(a)} + {}_3Y^T_{(a)}}{{}_1X^T_{(a)} + {}_2X^T_{(a)} + {}_3X^T_{(a)}} = R_{(a)} $$ el estimador muestral de ${}_rR$ será

$$ {}_r\hat{R} = \frac{{}_r\hat{Y}^T}{{}_r\hat{X}^T} = \frac{\hat{Y}^T_{(a)}}{\hat{X}^T_{(a)}} = \frac{{}_1\hat{Y}^T_{(a)} + {}_2\hat{Y}^T_{(a)} + {}_3\hat{Y}^T_{(a)}}{{}_1\hat{X}^T_{(a)} + {}_2\hat{X}^T_{(a)} + {}_3\hat{X}^T_{(a)}} = \hat{R}_{(a)} $$ donde los estimadores de los totales del numerador y del denominador se hallan en el punto 2.

La expresión general de la variancia de $\hat{R}_a$ es similar a la precedente, aplicada ahora a la categoría a

$$ s^2_{\hat{R}_{(a)}} = \frac{1}{(X^T_{(a)})^2} \sum_{i=1}^{3} s^2_{({}_i\hat{Y}^T_{(a)} - R_i\hat{X}^T_{(a)})} = \sum_{i=1}^{3} \left(\frac{1}{(X^T_{(a)})^2} s^2_{({}_i\hat{Y}^T_{(a)} - R_i\hat{X}^T_{(a)})}\right) $$ el estimador muestral sería

$$ \hat{s}^2_{\hat{R}_{(a)}} = \frac{1}{(\hat{X}^T_{(a)})^2} \sum_{i=1}^{3} \hat{s}^2_{({}_i\hat{Y}^T_{(a)} - \hat{R}_i\hat{X}^T_{(a)})} $$ donde la razón $\hat{R}$ ahora en rigor es $\hat{R}_a$ definida arriba.

4.2.1 Marco 1

Las expresiones de cálculo que corresponde usar en este marco son las mismas dadas en 4.1.1 cambiando (abc) por la sola categoría (a). Esto es, donde ya no se estima una razón para el conjunto de categorías (abc) sino sólo para las viviendas particulares generales (a), por lo tanto las fórmulas de cálculo directamente pueden tomarse del punto 4.1.1, efectuando los reemplazos indicados.
Lo cual conduciría al siguiente variancia muestral para este marco

$$ \hat{s}^2_{({}_1\hat{Y}^T_{(a)} - \hat{R}_1\hat{X}^T_{(a)})} = \frac{1}{6^2} \sum_{l}^{6} \left(\frac{1}{{}_1n_l}\left(\hat{S}^2_{{}_1\hat{Y}^T_{l(a)(j)}} + \hat{R}^2 \hat{S}^2_{{}_1\hat{X}^T_{l(a)(j)}} - 2\hat{R}\hat{S}_{{}_1\hat{Y}^T_{l(a)(j)},\, {}_1\hat{X}^T_{l(a)(j)}}\right)\right) $$

donde l indica réplica: l=1,2,...,6, y $\hat{R}$ es la razón $\hat{R}_{(a)}$ calculada con los tres marcos tal como se señaló arriba y que por simplicidad se lo escribió sin subíndice.

La cuasi covariancia muestral es

$$ \hat{S}_{{}_1\hat{Y}^T_{l(a)(j)},\, {}_1\hat{X}^T_{l(a)(j)}} = \frac{1}{{}_1n_l - 1} \sum_{j}^{{}_1n_l} \left({}_1\hat{Y}^T_{l(a)(j)} - \frac{\sum_j^{{}_1n_j} {}_1\hat{Y}^T_{l(a)(j)}}{{}_1n_j - 1}\right)\left({}_1\hat{X}^T_{l(a)(j)} - \frac{\sum_j^{{}_1n_j} {}_1\hat{X}^T_{l(a)(j)}}{{}_1n_j - 1}\right) $$

donde ${}_1\hat{Y}^T_{l(a)(j)}$ y ${}_1\hat{X}^T_{l(a)(j)}$ son los totales de viviendas (a) estimados para la réplica l, desde la sola j-ésima UP.

4.2.2 Marco 2

Para el marco 2 se procede realizando los mismos cambios indicados en 4.2.1

4.2.3 Marco 3
Para el marco 3 se procede del mismo modo realizando los cambios indicados en 4.2.1

4.3. Estimador de la RAZON de los totales de dos variables en el conjunto de viviendas categoría (bc) (Inquilinatos, Hoteles, Pensiones y Casas tomadas), de un CGP cualquiera, en los marcos 1, 2 y 3. Se excluyen villas de emergencia.

La característica a estimar es una razón entre dos variables en el dominio 4 (D4)

$$ {}_qR = \frac{{}_qY^T}{{}_qX^T} = \frac{Y^T_{(bc)}}{X^T_{(bc)}} = \frac{{}_1Y^T_{(bc)} + {}_2Y^T_{(bc)} + {}_3Y^T_{(bc)}}{{}_1X^T_{(bc)} + {}_2X^T_{(bc)} + {}_3X^T_{(bc)}} = R_{(bc)} $$ el estimador muestral de R será

$$ {}_q\hat{R} = \frac{{}_q\hat{Y}^T}{{}_q\hat{X}^T} = \frac{\hat{Y}^T_{(bc)}}{\hat{X}^T_{(bc)}} = \frac{{}_1\hat{Y}^T_{(bc)} + {}_2\hat{Y}^T_{(bc)} + {}_3\hat{Y}^T_{(bc)}}{{}_1\hat{X}^T_{(bc)} + {}_2\hat{X}^T_{(bc)} + {}_3\hat{X}^T_{(bc)}} = \hat{R}_{(bc)} $$ donde los estimadores de los

La variancia del estimador $\hat{R}_{(bc)}$ será de forma similar a la indicada en 4.1.1 y en 4.2.1, aplicada a las categorías (bc), esto es:

$$ s^2_{\hat{R}_{(bc)}} = \frac{1}{(X^T_{(bc)})^2} \sum_{i=1}^{3} s^2_{({}_iY^T_{(bc)} - R_i X^T_{(bc)})} = \sum_{i=1}^{3} \left(\frac{1}{(X^T_{(bc)})^2} s^2_{({}_i\hat{Y}^T_{(bc)} - R_i \hat{X}^T_{(bc)})} \right) $$ el estimador muestral sería

$$ s^2_{\hat{R}_{(bc)}} = \frac{1}{(\hat{X}^T_{(bc)})^2} \sum_{i=1}^{3} \hat{s}^2_{({}_i\hat{Y}^T_{(bc)} - \hat{R}_i \hat{X}^T_{(bc)})} $$ donde la razón $\hat{R}$ ahora es $\hat{R}_{(bc)}$

Para los tres marcos 1, 2 y 3 se procede del modo indicado arriba para 4.2.

II-5- Estimador del TOTAL de una variable en el muestreo estratificado

5.1 Estimación del total $Y^T_{(abc)}$ para las categorías de viviendas (abc) en el muestreo estratificado. Dominio 1 (D1)

El total a estimar es el agregado de la variable Y a lo largo de todas las viviendas (unidades de listado) de la ciudad, que simbolizamos $Y^T_{(abc)}$. Esto es, el agregado de Y en el dominio de análisis D1. Se trata de estimar un total que surge de agregar los estimadores correspondientes a cada marco, a lo largo de los 16 Centros de Gestión y Participación de la ciudad, esto es:

$$ Y^T_{(abc)} = \sum_h^H \sum_i^4 {}_iY^T_{(abc)h} = \sum_i^4 \sum_h^H {}_iY^T_{(abc)h} $$ donde i =1,...,4 indica los marcos de muestreo de

cada CGP, y h =1,...,H señala los 16 CGP de la ciudad. Un estimador del total, acorde a los procedimientos de muestreo empleados, sería:

$$\hat{Y}^T_{(abc)} = \sum_h^H \sum_i^4 {}_i\hat{Y}^T_{(abc)h} = \sum_i^4 \sum_h^H {}_i\hat{Y}^T_{(abc)h}$$

donde los estimadores de total son independientes entre marcos de un mismo CGP y además son independientes entre CGP dado que se tratan de selecciones mutuamente independientes, por lo que la variancia del total será puede expresarse

$$s^2_{\hat{Y}^T_{(abc)}} = \sum_h^H \sum_i^4 {}_i s^2_{\hat{Y}^T_{(abc)h}} = \sum_i^4 \sum_h^H {}_i s^2_{\hat{Y}^T_{(abc)h}}$$

donde ${}_i s^2_{\hat{Y}^T_{(abc)h}}$ constituye la variancia del estimador ${}_i\hat{Y}^T_{(abc)h}$ en el marco i-ésimo, del CGP h-ésimo.

El estimador de la variancia será

$$\hat{s}^2_{\hat{Y}^T_{(abc)}} = \sum_h^H \sum_i^4 {}_{ih}\hat{s}^2_{\hat{Y}^T_{(abc)}} = \sum_i^4 \sum_h^H {}_{ih}\hat{s}^2_{\hat{Y}^T_{(abc)}}$$

tanto los estimadores ${}_i\hat{Y}^T_{(abc)h}$ como de ${}_i\hat{s}^2_{\hat{Y}^T_{(abc)h}}$ se encuentran explicitados en los puntos 1.1, 1.2, 1.3 y 1.4.

5.2 La estimación de un TOTAL en los dominios D2, D3, D4 y D5.

Las expresiones de cálculo anteriores son válidas para estimar “totales” (agregados) en todos los dominios suprimiendo: aquellos marcos que no hacen al dominio de estimación y/o las categorías de viviendas que no están involucradas en la definición del mismo. Por lo señalado, las expresiones generales de 3.1 junto con aquellas más específicas contenidas en 1.1, 1.2 ,1.3 ,1.4 ,2 y 3 explican todos los procedimientos de estimación necesarios para abordar la estimación de totales para todos los dominios de análisis de la EAH2004.

II-6- La estimación de la RAZON de los totales de dos variables en el muestreo estratificado

Para extender la estimación de una razón a todos los CGP de la ciudad, o para un subconjunto de ellos, corresponde considerar a cada CGP como un estrato de muestreo, cada uno de los cuales también se encuentra dividido en subestratos, que son los marcos de muestreo. Como ya se señaló con anterioridad, cada marco dentro de un CGP fue muestreado en forma independiente y también entre CGP.

En las expresiones siguientes se utilizará el subíndice h para indicar CGP, con h=1,..,H, donde H representa los 16 CGPs de la ciudad.

6.1 Estimación de la RAZON para las categorías de viviendas (abc) en el muestreo estratificado

La característica a estimar ahora es la razón combinada

$$R = \frac{Y^T}{X^T} = \frac{\sum_h^H Y_h^T}{\sum_h^H X_h^T} = \frac{\sum_h^H Y_{(abc)h}^T}{\sum_h^H X_{(abc)h}^T} = \frac{\sum_h^H (\sum_i^4 {}_iY_{(abc)h}^T)}{\sum_h^H (\sum_i^4 {}_iX_{(abc)h}^T)} = \frac{\sum_i^4 (\sum_h^H {}_iY_{(abc)h}^T)}{\sum_h^H (\sum_i^4 {}_iX_{(abc)h}^T)} = R_{(abc)}$$

donde i es el indicador de marco y h el de CGP.

El estimador muestral de R puede escribirse

$$\hat{R} = \frac{\hat{Y}^T}{\hat{X}^T} = \frac{\sum_h^H \hat{Y}_h^T}{\sum_h^H \hat{X}_h^T} = \frac{\sum_h^H (\sum_i^4 {}_i\hat{Y}_{(abc)h}^T)}{\sum_h^H (\sum_i^4 {}_i\hat{X}_{(abc)h}^T)} = \frac{\sum_i^4 (\sum_h^H {}_i\hat{Y}_{(abc)h}^T)}{\sum_h^H (\sum_i^4 {}_i\hat{X}_{(abc)h}^T)} = \hat{R}_{(abc)} \qquad (1)$$

Notar que $R_{(abc)}$ representa la razón calculada sobre toda la ciudad en el dominio 1 (D1). Esto es, una razón que involucra a todas las viviendas de la ciudad o, dicho de otro modo, a todos los marcos de todos los CGP de Buenos Aires.

La variancia del estimador es la simple extensión de la expresión (1) del punto 4.1 dado la independencia entre marcos y entre CGP, por lo cual resulta

$$s^2_{\hat{R}(abc)} = \frac{1}{(X^T_{(abc)})^2} \sum_h^H \sum_{i=1}^4 s^2_{({}_i\hat{Y}^T_{(abc)h} - R_i\hat{X}^T_{(abc)h})} = \frac{1}{(X^T_{(abc)})^2} \sum_i^4 \sum_h^H s^2_{({}_i\hat{Y}^T_{(abc)h} - R_i\hat{X}^T_{(abc)h})} \qquad (2)$$

El estimador muestral de dicha variancia es

$$\hat{s}^2_{\hat{R}(abc)} = \frac{1}{(\hat{X}^T_{(abc)})^2} \sum_h^H \sum_i^4 \hat{s}^2_{({}_i\hat{Y}^T_{(abc)h} - \hat{R}_i\hat{X}^T_{(abc)h})} = \frac{1}{(\hat{X}^T_{(abc)})^2} \sum_i^4 \sum_h^H \hat{s}^2_{({}_i\hat{Y}^T_{(abc)h} - \hat{R}_i\hat{X}^T_{(abc)h})} \qquad (3)$$

El cálculo de la variancia $\hat{s}^2_{({}_i\hat{Y}^T_{(abc)h} - \hat{R}_i\hat{X}^T_{(abc)h})}$ para cada marco se detalla en los puntos 4.1.1, 4.1.2, 4.1.3 y 4.1.4 anteriores. Observar que en cada sumando de la variancia $\hat{R}$ es la razón estimada con todas las vivienda de la ciudad.

6.2 La estimación de una RAZON en los dominios D2, D3, D4 y D5.

Las expresiones de cálculo anteriores son válidas para estimar razones en todos los dominios, suprimiendo aquellos marcos que no hacen al dominio de estimación y/o a las categorías de viviendas que no están involucradas en la definición del mismo. Por lo señalado, las expresiones generales de 5.1 junto con aquellas más específicas contenidas en 4.1, 4.2 y 4.3 explican todos los procedimientos de estimación necesarios para abordar la estimación de razones de todos los dominios de análisis de la EAH de 2004.

III -Estimaciones para AREAS ESPECIALES (o dominios especiales) de la ciudad de Buenos Aires

Por área especial se entiende un continuo de superficie cuyos limites no necesariamente son coincidentes con los límites de los CGP, y obviamente su definición responde a una necesidad específica y precisa de producción de información. Tales como fueron los casos del área "Cuenca del arroyo Maldonado" y el área "Zona de influencia del Mercado de Liniers".

III-1- Consideraciones a tener en cuenta en el momento de establecer un área especial de trabajo

La definición de una de estas áreas debería ceñirse a algunas reglas básicas que tienen que ver con posibilidad posterior de poder producir la información objetivo a partir de los datos de la EAH 2004, a saber:

1-La totalidad del área especial no puede ser menor que un CGP promedio, lo cual puede lograrse mediante la unión de partes de dos o más CGPs contiguos, de modo que las estimaciones emergentes no se alejen demasiado de la precisión que habitualmente se alcanza con un CGP, que de algún modo marca el mínimo pretendido por la DGEyC en la precisión de los datos que suministra desagregadamente.

2-La/s villas de emergencia de un CGP involucrado en el área especial se las incluye totalmente o bien se las excluye en su totalidad.

3-Se tratará que un CGP involucrado en el área especial mínimamente se encuentre incluido en un sexto de su superficie (o un sexto de sus unidades primarias de muestreo), excluyendo o sin contar el área de villas, a fin tener una elevada certidumbre (99,8%) de contar con al menos dos unidades primarias de "viviendas particulares generales" dentro esa superficie, y consecuentemente poder calcular el desvío estándar de los estimadores.

4-Si bien el área la define el demandante, será indispensable que quien lo realice lo efectúe con total ignorancia o desconocimiento de la ubicación de las UP que componen la muestra de la EAH, a fin de evitar que el área especial se vea afectada por un sesgo en los resultados proveniente de incluir o excluir intencionalmente algunas UP, para inducir resultados.

5. La superficie de un CGP que participe del área de trabajo se debe delimitar sólo por unidades primarias de muestreo completas (UP). Es decir, el área delimitada no contendrá ninguna UP incompleta. Durante esta operación se identifican las UP incluidas y su cantidad.

6. No puede existir un dominio especial, geográficamente definido, que se refiera sólo a I-H-P-CT o sólo a villas de emergencia, o a la conjunción de ambos, dado que estos son dominios de análisis sólo válidos para el total ciudad.

III-2 - Los dominios de análisis posibles dentro del área especial de trabajo

Dentro del área establecida es posible tabular los datos y producir información para los siguientes dominios de análisis:

Dominio 1 (D1): incluye todos los tipos de viviendas existentes en el área definida, sin excepción. Viviendas “particulares generales”, “viviendas (pieza/habitación de...) de inquilinatos, hoteles, pensiones y casas tomadas” y “viviendas de villas de emergencia”

Dominio 2 (D2): idem D1, pero sin incluir villas de emergencia

Dominio 3 (D3): idem D1, pero sin incluir “viviendas (pieza/habitación de...) de inquilinatos, hoteles, pensiones y casas tomadas” y “viviendas de villas de emergencia”. Es decir, solo viviendas “particulares generales”.

III-3 - Procedimientos de estimación

3.1 En los CGP totalmente incluidos dentro del área especial

Las estimaciones dentro de bs CGP que han sido completamente incluidos dentro del área de trabajo seguirán el procedimiento habitual de estimación establecidos para la EAH 04.
Los CGP incluidos en forma parcial serán tratados de un modo distinto, tal como se describe seguidamente.

3.2 – Marcos de muestreo y estimadores para CGPs parcialmente incluidos en el área especial

En estos casos la porción de un CGP incluida en el área la identificaremos con **A**. Básicamente toda esta porción se la considerará como un estrato con cuatro subestratos de muestreo o marcos independientes de muestreo compuestos del siguiente modo:

Marco 1A -Marco de viviendas particulares generales

-Todas las UP del marco 1 (de viviendas particulares generales) que se encuentran dentro de la porción incluida (A) constituirán el marco 1A de esa parte, y las UP maestrales contenidas en 1A serán consideradas como una muestra aleatoria (prescindiendo de la replicación), seleccionadas con probabilidad proporcional al tamaño de las mismas, en un muestreo a dos etapas con reposición en la primera etapa y sin reposición en la segunda etapa, conformando todas ellas el nuevo marco 1A o sub-estrato de muestreo, dentro de la porción incluida del CGP.

Nomenclatura

${}_{1A}N$: designa el total de UP que han caído en el nuevo estrato así definido.
${}_{1A}n$: total de unidades primarias seleccionadas que quedaron en el nuevo marco

${}_{1A}M_o$: total de viviendas, según el CNP01, en todo el estrato 1A. Valor que se puede determinar del marco muestral original del CGP. Deberá verificarse que $_{1A}M_o = \sum_{j}^{_{1A}N} {}_1M_j$

${}_{1A}P_j$: designa la probabilidad de selección de la j-ésima UP en el nuevo estrato 1A. Las mismas deben ajustarse o redefinirse acorde al tamaño final del mismo del modo siguiente: ${}_{1A}P_j = \frac{{}_1M_o}{{}_{1A}M_o}\,{}_1P_j$ solo si $j \in al - nuevo - estrato$; donde

${}_1M_o$: cantidad de viviendas en el original marco 1, del CGP

${}_1P_j$: probabilidad de selección correspondiente a la j-ésima UP en el marco original.

Marco 2A

El marco 2A de la parte incluida va a quedar conformado por todos los domicilios de I-H-P-CT del listado-marco 2 original del CGP, que quedaron incluidos en el área de trabajo A.

En este marco corresponde realizar los mismos ajustes planteados en el caso anterior, con lo cual se obtendrían

$${}_{2A}N\,,\quad {}_{2A}M_o\,,\quad {}_{2A}P_j\,,\quad {}_{2A}n\,,\quad {}_{2A}P_j = \frac{{}_2M_o}{{}_{2A}M_o}\,{}_2P_j$$

Marco 3A

Este marco quedará compuesto por los domicilios autorrepresentados del original marco 3, que quedaron incluidos en la porción del CGP que pasó a formar parte del área de estudio. No existen cambios en los estimadores dado que son domicilios (o UP) autorrepresentados.

Marco 4A

Este marco no cambia respecto del marco 4 original, ya que las viviendas en villas, tal como se indicó arriba, participan todas o se las excluye a todas. Por lo cual los estimadores dentro de este marco serán exactamente los mismos usados en el propio CGP de origen.

III-4-Procedimiento de estimación en los marcos 1A, 2A, 3A, 4A, para el Dominio 1. Estimación del TOTAL (${}_AY^T$ o también ${}_AY^T_{(abc)}$) de una variable Y, en un CGP parcialmente incluido en el AREA ESPECIAL (dominio especial DE 1), incluyendo las viviendas en villas de emergencia

La magnitud a estimar a lo largo de los cuatro marcos, y de todas las categorías de viviendas será

$$ {}_{A}Y^{T} = {}_{1A}Y^{T} + {}_{2A}Y^{T} + {}_{3A}Y^{T} + {}_{4A}Y^{T} = {}_{1A}Y^{T}_{(abc)} + {}_{2A}Y^{T}_{(abc)} + {}_{3A}Y^{T}_{(abc)} + {}_{4A}Y^{T}_{(abc)} = {}_{A}Y^{T}_{(abc)} $$

Siendo los cuatro marcos poblaciones independientes con muestras también independientes, básicamente el estimador del total ${}_{A}Y^{T}_{(abc)}$ será

$$ {}_{A}\hat{Y}^{T}_{(abc)} = {}_{1A}\hat{Y}^{T}_{(abc)} + {}_{2A}\hat{Y}^{T}_{(abc)} + {}_{3A}\hat{Y}^{T}_{(abc)} + {}_{A4}\hat{Y}^{4}_{(abc)} \quad ; \quad \hat{s}^{2}_{{}_{A}\hat{Y}^{T}_{(abc)}} = \hat{s}^{2}_{{}_{1A}\hat{Y}^{T}_{(abc)}} + \hat{s}^{2}_{{}_{2A}\hat{Y}^{T}_{(abc)}} + \hat{s}^{2}_{{}_{3A}\hat{Y}^{T}_{(abc)}} + \hat{s}^{2}_{{}_{4A}\hat{Y}^{T}_{(abc)}} \quad ; $$

4.1 Los estimadores de ${}_{1A}\hat{Y}^{T}_{(abc)}$ y $\hat{s}^{2}_{{}_{1A}\hat{Y}^{T}_{(abc)}}$ en el marco 1A

Por estar en un muestreo con selección *ppt* y sin réplicas los estimadores serán

$$ {}_{1A}\hat{Y}^{T}_{(abc)} = {}_{1A}M_o \frac{1}{{}_{1A}n} \sum_{j}^{{}_{1A}n} \frac{1}{{}_{1}m_j} \sum_{k}^{{}_{1}m_j} {}_{1}y_{(abc)jk} = \frac{1}{{}_{1A}n} \sum_{j}^{{}_{1A}n} {}_{1}M_o \frac{1}{{}_{1}m_j} \sum_{k}^{{}_{1}m_j} {}_{1}y_{(abc)jk} = $$

Notar que tanto ${}_{1}m_j$ como ${}_{1}y_{(abc)jk}$ en realidad pertenecen al marco 1A porque toda la UP lo está, por lo que en rigor son ${}_{1A}m_j$ como ${}_{1A}y_{(abc)jk}$, sin embargo por simplicidad la pertenencia a 1A sólo quedará indicado en las sumatorias.

Dado que las viviendas de la muestra pertenecientes a las categorías a y c fueron encuestadas, en tanto que aquellas pertenecientes a la b (I-H-P) no se encuestaron (sólo se registró la cantidad de viviendas ocupadas), es necesario introducir el siguiente ajuste al estimador del total

$$ = \frac{1}{{}_{1A}n} \sum_{j}^{{}_{1A}n} {}_{1A}M_o \frac{1}{{}_{1}m_j} \left(\sum_{k}^{{}_{1}m_{aj}} {}_{1}y_{ajk} + {}_{1}m_{bj}\, {}_{1}\bar{y}_{bj} + \sum_{k}^{{}_{1}m_{cj}} {}_{1}y_{cjk} \right) $$

${}_{1}\bar{y}_{bj}$ es el promedio aritmético simple de la variable, por unidad de listado, observado en las ${}_{1}m_{bj}$ unidades secundarias de la muestra, perteneciente a la j-ésima UP. Sin embargo dado que en el marco 1, y consecuentemente en 1A, no se han relevado los valores de la variable para la categoría b, en lo sucesivo se sustituirá a ${}_{1}\bar{y}_{bj}$ por el promedio de la misma variable y categoría (b) del marco 2 original, es decir de todo el CGG. Se empleará el promedio de todo el CGP dado que el marco 2A podría estar compuesto por muy escasas UP.

Esto es, se lo sustituirá por el promedio aritmético simple ${}_{2}\bar{y}_{b}$ (en rigor por ${}_{2}\bar{y}_{bc}$) del mismo CGP, lo cual significa que se dará a las unidades b del marco 1 el mismo comportamiento de b en el marco 2 del mismo CGP.
Como sustituto del término ${}_{1}\bar{y}_{bj}$ en rigor se utilizará el promedio por vivienda de las categorías (bc) (I-H-P-CT) del marco 2 donde prácticamente no existen casas tomadas por lo que ${}_{2}\bar{y}_{bc}$ prácticamente es igual a ${}_{2}\bar{y}_{b}$. Se lo calculará:

$${}_2\bar{y}_{bc} = \frac{1}{{}_2 n^{\bullet}_{bc}} \sum_{j}^{{}_2 n^{\bullet}_{bc}} \frac{1}{{}_2 m_{(bc)j}} \sum_{k}^{{}_2 m_{(bc)j}} {}_2 y_{(bc)jk}$$

que es el promedio aritmético puro de la variable y en todas las viviendas encuestadas que pertenecen a las categorías (bc) en el marco 1 de todo el CGP.

Además, dado que las viviendas de categoría b en algunos casos se encontraron más de una vivienda de I-H-P, se introducirá una variable auxiliar ${}_1\bar{x}_{bj}$ la que se definirá en las UP del marco 1A.

${}_1\bar{x}_{bj}$ = cantidad promedio (por unidad de listado, seleccionadas y de reserva) de viviendas de I-H-P halladas en la unidad primaria correspondiente. Es decir: ${}_1\bar{x}_{bj} = \frac{1}{{}_1DR_{bj}} \sum_{k}^{{}_1DR_{bj}} {}_1x_{bjk}$

Finalmente, para dar participación en las estimaciones a las viviendas de reserva se introducirán los conceptos ya utilizados en el punto 1.1 de viviendas ocupadas y domicilios realizados con lo cual el estimador tomaría la forma

$${}_{1A}\hat{Y}^{T}_{(abc)} = \frac{1}{{}_{1A}n} \sum_{j}^{{}_{1A}n} {}_1M_o \frac{1}{{}_1m_j} \left(\frac{{}_1VO_j}{{}_1DR_{aj}} \sum_{k}^{{}_1DR_{aj}} {}_1y_{jk} + {}_1m_{bj}\, {}_1\bar{x}_{bj}\, {}_2\bar{y}_b + \frac{{}_1VO_{cj}}{{}_1DR_{cj}} \sum_{k}^{{}_1DR_{cj}} {}_1y_{cjk} \right) =$$

$$= \frac{1}{{}_{1A}n} \sum_{j}^{{}_{1A}n} \left(\left({}_{1A}M_o \frac{1}{{}_1m_j} \frac{{}_1VO_j}{{}_1DR_{aj}} \sum_{k}^{{}_1DR_{aj}} {}_1y_{ajk} \right) + \left({}_{1A}M_o \frac{1}{{}_1m_j} {}_1m_{bj}\, {}_1\bar{x}_{lj}\, {}_2\bar{y}_b \right) + \left({}_{1A}M_o \frac{1}{{}_1m_j} \frac{{}_1VO_{lj}}{{}_1DR_{cj}} \sum_{k}^{{}_1DR_{lj}} {}_1y_{cjk} \right) \right) =$$

$$= \frac{1}{{}_{1A}n} \sum_{j}^{{}_{1A}n} \left({}_{1A}\hat{Y}^{T}_{a(j)} + {}_{1A}\hat{Y}^{T}_{b(j)} + {}_{1A}\hat{Y}^{T}_{c(j)} \right) = \frac{1}{{}_{1A}n} \sum_{j}^{{}_{1A}n} {}_{1A}\hat{Y}^{T}_{(abc)(j)} \qquad (1)$$

Una vez calculados los valores ${}_1\hat{Y}_{(abc)(j)}$ se calculará del modo usual la variancia del total

$$\hat{s}^{2}_{{}_{1A}\hat{Y}^{T}_{(abc)}} = \frac{1}{{}_{1A}n({}_{1A}n - 1)} \sum_{j}^{{}_{1A}n} \left({}_{1A}\hat{Y}^{T}_{(abc)(j)} - \frac{\sum_{j}^{{}_{1A}n} {}_{1A}\hat{Y}^{T}_{(abc)(j)}}{{}_{1A}n} \right)^{2} \qquad (2)$$

4.2 Los estimadores de ${}_{2A}\hat{Y}^{T}_{(abc)}$ y $\hat{s}^{2}_{{}_{2A}\hat{Y}^{T}_{(abc)}}$, en el marco 2A.

Como particularidad de este estrato de muestreo de I-H-P-CT es de señalarse que, dado la generalizada pequeñez de las muestras en la mayoría de los CGPs (son excepciones los CGP: 1, 3, 4, y 16), resulta verdaderamente elevada la probabilidad que cuando se particiona un CGP en dos o más porciones o sectores, alguna/s de esas partes contengan sólo una o aún ninguna UP muestral, con lo cual las estimaciones directas no son posibles, por lo que en estos casos será necesario recurrir a aproximaciones.

Por lo señalado se describe el procedimiento de estimación a seguir en los distintos casos.

4.2.1 Cuando en el marco 2A hay 2 o más UP de la muestra, efectivamente encuestadas ($_{2A}n^{\bullet} \geq 2$)

En este caso se aplicarán los mismos los mismos estimadores dados en el punto II.1.2 para un CGP cualquiera. Esto es:

$$ {}_{2A}\hat{Y}^T_{(abc)} = \frac{1}{_{2A}n}\sum_j^{_{2A}n} \frac{_{2A}\hat{Y}^T_j}{_{2A}P_j} = \frac{1}{_{2A}n}\sum_j^{_{2A}n} \frac{_{2A}M^*_j\, _2\bar{y}_j}{_{2A}P_j} = \frac{1}{_{2A}n}\sum_j^{_{2A}n} \frac{_{2A}M^*_j}{_{2A}P_j}\frac{1}{_2m_j}\left(\sum_k^{_2m_j} {}_2y_{jk}\right) = \frac{1}{_{2A}n}\sum_j^{_{2A}n} {}_{2A}\hat{Y}^T_j = $$

$$ = \frac{1}{_{2A}n}\sum_j^{_{2A}n} \frac{_{2A}M^*_j}{_{2A}P_j}\frac{1}{_2m_j}\left(\sum_k^{_2m_{aj}} {}_2y_{ajk} + \sum_k^{_2m_{bj}} {}_2y_{bjk} + \sum_k^{_2m_{cj}} {}_2y_{cjk}\right) \tag{1} $$

donde $_{2A}M^*_j$ es el total de viviendas que se han obtenido por recuento total de las mismas, en cada UP de la muestra y también de las reservas, en tanto que $_{2A}P_j$ es la probabilidad en el marco 2A, definida en 9.3.2 . En esta estimación $_{2A}n$ representa el total de UP de la muestra $_2n$ en este marco 2A, incluyendo las de reserva que hubiesen sido utilizadas y $_{2A}n^{\bullet}$ las efectivamente encuestas.

Uniendo las categorías b y c y considerando la no encuesta, en 2A, de las viviendas de la categoría a, el estimador queda

$$ {}_{2A}\hat{Y}^T_{(abc)} = \frac{1}{_{2A}n}\sum_j^{_{2A}n} \frac{_{2A}M^*_j}{_{2A}P_j}\frac{1}{_2m_j}\left(_2m_{aj}\, _{2A}\bar{y}_{aj} + \sum_k^{_2m_{(bc)j}} {}_2y_{(bc)jk}\right) \tag{2} $$

notar que una UP de este marco generalmente tendrá viviendas que pertenecerán a las categorías bc o, de no ser así, toda la UP será una sola vivienda de la categoría a, en muy raras ocasiones la UP contendrá viviendas de las tres categorías. Cuando toda la UP sea una sola vivienda de la categoría a entonces, en ese caso, la vivienda se la incluirá en el proceso de estimación como si la misma hubiese sido realizada (se la considerará en $_{2A}n$) dándole el valor de $_{2A}M^{\bullet}_j = 1$, el valor de $_2m_{aj} = 1$, y asignándole (imputándole) como aproximación del valor faltante $_{2A}\bar{y}_{aj}$ el promedio por unidad de listado de la misma variable en el marco 1A de la misma partición del CGP, calculado como se indica seguidamente.

$$ {}_{1A}\bar{y}_a = \frac{1}{_{1A}n}\sum_j^{_{1A}n}\left(\frac{1}{_{1A}DR_{aj}}\sum_k^{_{1A}DR_{aj}} {}_1y_{ajk}\right) = \frac{1}{_{1A}n}\sum_j^{_{1A}n} {}_{1A}\bar{y}_{aj} $$

con lo cual finalmente, y de forma similar al punto II.1.2, la ecuación (2) queda

$$ {}_{2A}\hat{Y}^T_{(abc)} = \frac{1}{_{2A}n}\sum_j^{_{2A}n} \frac{_{2A}M^*_j}{_{2A}P_j}\frac{1}{_2m_j}\left(_2m_{aj}\, _{1A}\bar{y}_a + \frac{_2n_{bc}}{_2n^{\bullet}_{bc}} z_j \sum_k^{_2m_{(bc)j}} {}_2y_{(bc)jk}\right) = \frac{1}{_{2A}n}\sum_j^{_{2A}n}\left(\frac{_{2A}M^*_j}{_{2A}P_j}\frac{1}{_2m_j}\, _2m_{aj}\, _{1A}\bar{y}_a + \frac{_{2A}M^*_j}{_{2A}P_j}\frac{_2n_{bc}}{_2n^{\bullet}_{bc}} z_j \frac{1}{_2m_j}\sum_k^{_2m_{(bc)j}} {}_2y_{(bc)jk}\right) = $$

$$ = \frac{1}{_{2A}n}\sum_j^{_{2A}n}\left({}_{2A}\hat{Y}^T_{a(j)} + {}_{2A}\hat{Y}^T_{bc(j)}\right) = \frac{1}{_{2A}n}\sum_j^{_{2A}n} {}_{2A}\hat{Y}^T_{(abc)(j)} $$

Procediendo de este modo una vez calculados los valores $_{2}\hat{Y}_{(j)} = {}_{2}\hat{Y}^{T}_{(abc)(j)}$ se estima la variancia empleando la fórmula usual

$$\hat{s}^{2}_{_{2A}\hat{Y}^{T}_{abc}} = \frac{1}{_{2A}n(_{2A}n-1)} \sum_{j}^{_{2A}n} (_{2A}\hat{Y}^{T}_{(abc)(j)} - \frac{\sum_{j}^{_{2A}n} {}_{2A}\hat{Y}^{T}_{(abc)(j)}}{_{2A}n})^{2} \qquad (3)$$

4.2.2 Cuando en el <u>marco 2A</u> ha caído sólo una (1) UP efectivamente encuestada ($_{2A}n^{\bullet} = 1$)

Cuando ha caído una sola UP muestral de I-H-P-CT en la partición A, el estimador del total $_{2A}\hat{Y}^{T}_{(abc)}$, dado en la ecuación (2) de arriba, se reduce para una sola UP encuestada, el cálculo se realizará del modo habitual expresado por la ecuación (2) del punto precedente. Ver también II.1.2.

$$_{2A}\hat{Y}^{T}_{(abc)} = \frac{1}{_{2A}n} \sum_{j}^{_{2A}n} \frac{_{2A}M^{*}_{j}}{_{2A}P_{j}} \frac{1}{_{2}m_{j}} (_{2}m_{aj1A}\bar{y}_{a} + \frac{_{2}n_{bc}}{_{2}n^{\bullet}_{bc}} z_{j} \sum_{k}^{_{2}m_{(bc)j}} {}_{2}y_{(bc)jk}) = \frac{1}{_{2A}n} \sum_{j}^{_{2A}n} (\frac{_{2A}M^{*}_{j}}{_{2A}P_{j}} \frac{1}{_{2}m_{j}} {}_{2}m_{aj1A}\bar{y}_{a} + \frac{_{2A}M^{*}_{j}}{_{2A}P_{j}} \frac{_{2}n_{bc}}{_{2}n^{\bullet}_{bc}} z_{j} \frac{1}{_{2}m_{j}} \sum_{k}^{_{2}m_{(bc)j}} {}_{2}y_{(bc)jk}) =$$

$$= \frac{1}{_{2A}n} \sum_{j}^{_{2A}n} (_{2A}\hat{Y}^{T}_{a(j)} + {}_{2A}\hat{Y}^{T}_{bc(j)}) = \frac{1}{_{2A}n} \sum_{j}^{_{2A}n} {}_{2A}\hat{Y}^{T}_{(abc)(j)}$$

Como variancia aproximada se tomará la expresión siguiente, que es el producto de la variancia del promedio $_{2}\bar{y}_{(abc)}$ de la variable por unidad de listado, para las categorías de viviendas (abc) en todo el <u>marco 2</u> del CGP, multiplicada por el cuadrado de la cantidad de unidades de listado del <u>marco 2A</u>. Esto es:

$$\hat{s}^{2}_{_{2A}\hat{Y}^{T}_{(abc)}} = \frac{_{2A}M^{2}_{o}}{(_{2}n-1)} \sum_{j}^{_{2}n} (_{2}\bar{y}_{(abc)j} - \frac{\sum_{j}^{_{2}n} {}_{2}\bar{y}_{(abc)j}}{_{2A}n})^{2} = {}_{2A}M^{2}_{o} \; (\frac{1}{(_{2}n-1)} \sum_{j}^{_{2}n} (_{2}\bar{y}_{(abc)j} - \frac{\sum_{j}^{_{2}n} {}_{2}\bar{y}_{(abc)j}}{_{2A}n})^{2})$$

Notar que la variancia de la media se calcula con todas las unidades primarias del marco 2 del mismo CGP.

La expresión anterior para un cálculo más simple también puede escribirse

$$\hat{s}^{2}_{_{2A}\hat{Y}^{T}} = \frac{_{2A}M^{2}_{o}}{_{2}M^{2}_{o}} (\frac{1}{(_{2}n-1)} \sum_{j}^{_{2}n} (_{2}\hat{Y}^{T}_{(abc)(j)} - \frac{\sum_{j}^{_{2}n} {}_{2}\hat{Y}^{T}_{(abc)(j)}}{_{2}n})^{2})$$

4.2.3 Cuando en el <u>marco 2A</u> no ha caído ninguna UP efectivamente encuestada ($_{2A}n^{\bullet} = 0$), pero es $_{2A}M_{o}$>0.

En este caso también se aplicará la expresión del punto anterior, ecuación (2), pero remplazando los promedios del modo siguiente:

$$ {}_{2A}\hat{Y}^{T}_{(abc)} = {}_{2A}M_{o}\,{}_{2}\bar{y}_{(abc)} = \frac{{}_{2A}M_{o}}{{}_{2}M_{o}}\,{}_{2}\hat{Y}^{T}_{(abc)} = \frac{{}_{2A}M_{o}}{{}_{2}M_{o}}\left({}_{2}\hat{Y}^{T}_{(a)} + {}_{2}\hat{Y}^{T}_{(bc)}\right) $$

Notar que para estimar el total de la variable Y se imputan valores promedio de todo el CGP ante la falta información propia del marco 2A

La variancia en estos casos se aproximará exactamente del mismo modo que el indicado en el punto precedente 9.4.2.2, cuando ${}_{2A}n = 1$.

4.2.4 Cuando en el marco 2A no ha caído ninguna UP de la muestra del CGP (${}_{2A}n = 0$), y también es ${}_{2A}M_{o} = 0$

Cuando se da la condición que ${}_{2A}M_{o} = 0$ significa que en el marco 2A no existe ninguna unidad primaria del tipo I-H-P-CT. Es un marco vacío, por lo que no hay nada que estimar en el mismo.

4.3 Los estimadores ${}_{3A}\hat{Y}^{T}_{(abc)}$ y $\hat{s}^{2}_{{}_{3A}\hat{Y}^{T}_{(abc)}}$, en el marco 3A

Estos estimadores son exactamente los mismos dados en el punto 1.1.3 de este mismo informe, aplicados ahora a aquellos domicilios autorrepresentados que se encuentran contenidos en el nuevo estrato. Por lo cual se remite directamente al punto II.1.3

4.4 Los estimadores ${}_{4A}\hat{Y}^{T}_{(abc)}$ y $\hat{s}^{2}_{{}_{4A}\hat{Y}^{T}_{(abc)}}$, en el marco 4A

Lo mismo ocurre en este caso, los estimadores son exactamente los mismos dados en el punto II.1.4 de este mimo informe, aplicados ahora a aquellas viviendas en villas contenidas en el nuevo estrato. En este punto resulta necesario recordar que: en cualquier partición de un CGP las viviendas en villas del mismo participan en su totalidad de la partición, o bien se las excluye en su totalidad de la misma.

III-5 Las estimaciones para el Dominio Especial 2 (DE 2)

Para producir las estimaciones para el dominio 3, que incluye sólo las viviendas residenciales generales y viviendas de I-H-P-CT, excluidos villas de emergencia, es suficiente con seguir los procedimientos dados arriba, en punto III, apartados 4.1, 4.2 y 4.3, sin incluir en los cómputos el marco 4A.

III-6 Las estimaciones para el Dominio Especial 3 (DE 3)

Estas estimaciones se logran trabajando dentro de los marcos 1A, 2A, 3A, con las viviendas de la sola categoría a, es decir, sin dar participación a las categorías b y c.
En este caso se tendrá en cuenta de considerar que:

$y_{jk} = \;\; y_{jk}$ si jk pertenece a la categoría de viviendas tipo a

$y_{jk} =$ 0 si jk NO pertenece a la categoría de viviendas a

III-7- Estimador del TOTAL de una variable en el área especial

7.1 Estimación del total $Y^{T}_{(abc)}$ para las categorías de viviendas (abc) en toda el área especial. Dominio Especial 1 (DE1)

El total a estimar es el agregado de la variable Y en toda el área especial, en todos los tipos de vivienda, que simbolizamos, igual que en II.6.1, con $Y^{T}_{(abc)}$. Esto es, el agregado de Y en el dominio especial DE1. Se trata de agregar el total de la variable Y que aportan los CGP incluidos totalmente, más la porción incluida de aquellos CGP parcialmente tomados por el área de trabajo. Por lo tanto el total a estimar puede escribirse:

$$Y^T_{(abc)} = \sum_h^H \sum_i^4 {}_iY^T_{(abc)h} = \sum_i^4 \sum_h^H {}_iY^T_{(abc)h}$$ donde, igual que en II.5.1, i =1,...,4

indica los marcos de muestreo de cada CGP total o parcialmente tomados, y h =1,...,H ahora indica sólo y cada uno de los CGP involucrados en el área especial. Un estimador del total, acorde a los procedimientos de muestreo empleados, sería:

$$\hat{Y}^T_{(abc)} = \sum_h^H \sum_i^4 {}_i\hat{Y}^T_{(abc)h} = \sum_i^4 \sum_h^{H4} {}_i\hat{Y}^T_{(abc)h}$$ donde los estimadores de total son

independientes entre marcos de un mismo CGP (ya sea en un CGP total o parcialmente incluido) y además son independientes entre CGP, dado que se tratan de selecciones mutuamente independientes, por lo que la variancia del total puede expresarse

$$s^2_{\hat{Y}^T_{(abc)}} = \sum_h^H \sum_i^4 {}_i s^2_{\hat{Y}^T_{(abc)h}} = \sum_i^4 \sum_h^H {}_i s^2_{\hat{Y}^T_{(abc)h}}$$ donde ${}_i s^2_{\hat{Y}^T_{(abc)h}}$ constituye la variancia del

estimador ${}_i\hat{Y}^T_{(abc)h}$ en el marco i-ésimo, del CGP h-ésimo.

El estimador muestral de la misma será

$$\hat{s}^2_{\hat{Y}^T_{(abc)}} = \sum_h^H \sum_i^4 {}_i \hat{s}^2_{\hat{Y}^T_{(abc)h}} = \sum_i^4 \sum_h^H {}_i \hat{s}^2_{\hat{Y}^T_{(abc)h}}$$

Notar que son los mismos estimadores dados en II.5.1 pero donde participan tanto los CGP totalmente incluidos como los parcialmente involucrados. Por lo que los estimadores ${}_i\hat{Y}^T_{(abc)h}$ y ${}_i \hat{s}^2_{\hat{Y}^T_{(abc)h}}$ se encuentran explicitados en el punto II apartados 1.1, 1.2, 1.3 y 1.4. para los CGP completamente incluidos; en tanto que los CGPs parcialmente incluidos, es decir, que participan con una porción A, los estimadores correspondientes están dados en III.4.1 a 4.4.

7.2 La estimación de un TOTAL en los dominios DE2, DE3

Igual que en III.5.2, las expresiones de cálculo anteriores son válidas para estimar "totales" (agregados) en todos los dominios suprimiendo: aquellos marcos que no hacen al dominio de estimación (como en el DE2 que debe suprimirse del marco 4), y/o las categorías de viviendas que no están involucradas en el dominio especial (DE) a estimar. En el caso del dominio DE3 corresponde trabajar sólo con los marcos 1,2,3 y las viviendas pertenecientes a la categoría A solamente.
Por lo señalado, las expresiones generales de 3.1 junto con aquellas más específicas contenidas en 1.1, 1.2 ,1.3 ,1.4 ,2 y 3 y en particular las detalladas en los punto III.4.2 explican todos los

procedimientos de estimación necesarios para abordar la estimación de totales para todos los dominios de análisis de la EAH2004.

III-8- La estimación de la RAZON de los totales de dos variables, para las categorías de viviendas (abc) en el Dominio Especial 1 (DE1)

Para extender la estimación de una razón a todos los CGPs involucrados en el área especial corresponde considerar a cada uno como un estrato de muestreo, cuales también se encuentra dividido en subestratos, que en rigor son las distintas subpoblaciones de muestreo. Como ya se señaló con anterioridad, cada marco, o subpoblación, dentro de un CGP fue muestreado en forma independiente y también entre CGP.

8.1 Estimación de la RAZON para las categorías de viviendas (abc) en el muestreo estratificado

La característica a estimar en este caso es la razón combinada

$$R=\frac{Y^T}{X^T}=\frac{\sum_h^H Y_h^T}{\sum_h^H X_h^T}=\frac{\sum_h^H Y_{(abc)h}^T}{\sum_h^H X_{(abc)h}^T}=\frac{\sum_h^H(\sum_i^4 {}_iY_{(abc)h}^T)}{\sum_h^H(\sum_i^4 {}_iX_{(abc)h}^T)}=\frac{\sum_i^4(\sum_h^H {}_iY_{(abc)h}^T)}{\sum_h^H(\sum_i^4 {}_iX_{(abc)h}^T)}=R_{(abc)}$$

donde i es el indicador de marco y h señala los CGPs tomados por el área especial.

El estimador muestral de R puede escribirse

$$\hat{R}=\frac{\hat{Y}^T}{\hat{X}^T}=\frac{\sum_h^H \hat{Y}_h^T}{\sum_h^H \hat{X}_h^T}=\frac{\sum_h^H(\sum_i^4 {}_i\hat{Y}_{(abc)h}^T)}{\sum_h^H(\sum_i^4 {}_i\hat{X}_{(abc)h}^T)}=\frac{\sum_i^4(\sum_h^H {}_i\hat{Y}_{(abc)h}^T)}{\sum_h^H(\sum_i^4 {}_i\hat{X}_{(abc)h}^T)}=\hat{R}_{(abc)}$$

Notar que $R_{(abc)}$ representa la razón calculada sobre todo el área especial de trabajo D1. Esto es, una razón que involucra a todas las viviendas (abc).

La variancia del estimador es la extensión de la expresión (1) del punto III.4.1 dado la independencia entre marcos y entre CGP, por lo cual resulta

$$s^2_{\hat{R}_{(abc)}}=\frac{1}{(X^T_{(abc)})^2}\sum_h^H\sum_{i=1}^4 s^2_{({}_i\hat{Y}^T_{(abc)h}-R_i\hat{X}^T_{(abc)h})}=\frac{1}{(X^T_{(abc)})^2}\sum_i^4\sum_h^H s^2_{({}_i\hat{Y}^T_{(abc)h}-R_i\hat{X}^T_{(abc)h})}$$

El estimador muestral de dicha variancia es

$$\hat{s}^2_{\hat{R}_{(abc)}}=\frac{1}{(\hat{X}^T_{(abc)})^2}\sum_h^H\sum_i^4 \hat{s}^2_{({}_i\hat{Y}^T_{(abc)h}-\hat{R}_i\hat{X}^T_{(abc)h})}=\frac{1}{(\hat{X}^T_{(abc)})^2}\sum_i^4\sum_h^H \hat{s}^2_{({}_i\hat{Y}^T_{(abc)h}-\hat{R}_i\hat{X}^T_{(abc)h})}$$

El cálculo de la variancia $\hat{s}^2_{({}_i\hat{Y}^T_{(abc)h} - \hat{R}\,{}_i\hat{X}^T_{(abc)h})}$ para cada marco se detalla en el punto II, apartados 4.1.1, 4.1.2, 4.1.3 y 4.1.4 . Observar que en cada sumando de la variancia el término $\hat{R}$ es la razón estimada con todas las vivienda del área de trabajo.

8.2 La estimación de una RAZON en los dominios DE2 y DE3

Las expresiones de cálculo anteriores son válidas para estimar razones en todos los dominios, suprimiendo aquellos marcos que no hacen al dominio de estimación y/o a las categorías de viviendas que no están involucradas en la definición del mismo. Por lo señalado, las expresiones generales de II.5.1 junto con aquellas más específicas contenidas en II,apartados 4.1, 4.2 y 4.3, explican todos los procedimientos de estimación necesarios para abordar la estimación de razones en los distintos dominios de análisis.

...........oooooo..........