


Departamento Administrativo Nacional de Estadística



Dirección de Regulación, Planeación,
Estandarización y Normalización
DIRPEN

**Metodología Diseño Estadístico
Encuesta de Consumo Cultural
-ECC-**

Julio 2008

	METODOLOGÍA DISEÑO ESTADÍSTICO ENCUESTA CONSUMO CULTURAL -ECC-	CÓDIGO: ME-ECC-MET-01 VERSIÓN: 02 PÁGINA: 1 FECHA: 08-07-08
ELABORÓ: METODOLOGÍA ESTADÍSTICA	REVISÓ: COORDINADOR DE ESTUDIOS ESTADÍSTICOS	APROBÓ: DIRECTOR DIRPEN

CONTENIDO

MARCO ESTADÍSTICO _____	2
1. POBLACIÓN OBJETIVO DEL DISEÑO MUESTRAL _____	2
2. VARIABLES DE INTERÉS _____	2
3. PARÁMETROS A ESTIMAR _____	2
4. COBERTURA _____	2
5. NIVELES DE DESAGREGACIÓN _____	2
6. UNIDADES ESTADÍSTICAS _____	3
7. PERIODO DE REFERENCIA _____	3
8. MARCO MUESTRAL _____	3
9. DISEÑO MUESTRAL _____	4
9.1. TAMAÑO DE LA MUESTRA _____	7
9.2. DISEÑO Y SELECCIÓN DE LA MUESTRA _____	8
9.3. METODOLOGÍA DE ESTIMACIÓN Y VARIANZA _____	9
9.3.1. Factor de expansión final _____	10
9.3.2. Estimadores de Totales y Razones _____	10
9.3.3. Ajuste del factor de expansión final _____	13
10. COEFICIENTE DE VARIACIÓN _____	14
BIBLIOGRAFÍA _____	16

MARCO ESTADÍSTICO

1. POBLACIÓN OBJETIVO DEL DISEÑO MUESTRAL

La población objetivo de estudio la conforman las personas mayores de 5 años residentes habituales de los hogares pertenecientes a las cabeceras municipales del país. En particular, se busca llegar a las personas de cinco años y más de los hogares.

2. VARIABLES DE INTERÉS

- De estudio: perfiles sobre el uso de bienes, servicios y espacios culturales así como del tiempo invertido y el tipo de actividades realizadas en el tiempo libre.
- De clasificación: por región, por municipio, sexo, edad, nivel socioeconómico.

Se asume la existencia de estratos de tamaño poblacional en el país para las cabeceras municipales. Es decir, existen grupos de clasificación para homogenizar las cabeceras por cantidad de personas residentes. Los municipios del mismo estrato, tienen aproximadamente la misma cantidad de población.

Se requiere forzosamente obtener información de ciertos municipios del país por razones argumentadas en la prueba piloto y en las definiciones temáticas. Por lo tanto, se debe llegar a ciertos municipios en los que se debe obtener obligatoriamente la información. A estos casos se les denomina de “inclusión forzosa”. Tales municipios se agrupan en un mismo estrato para hacerles un tratamiento de selección de muestra independiente.

3. PARÁMETROS A ESTIMAR

La encuesta de Consumo Cultural mide variables de tipo categórico, continuo y nominal para las cuales se debe estimar:

- Totales de población dentro de los dominios de las categorías de las preguntas para observar perfiles.
- Totales de densidad para las variables nominal.
- Totales de variables continuas.
- Proporciones de población para cuadros de salida de categorías cruzadas.
- Promedios de variables continua para establecer nivel medio de alguna característica.
- Tasas entre totales y periodos de referencia para establecer desempeño de una variable en términos del tiempo.

4. COBERTURA

La encuesta está dirigida a personas mayores de 5 años residentes habituales de un hogar, pertenecientes a las cabeceras municipales del territorio nacional en las diferentes regiones del país

5. NIVELES DE DESAGREGACIÓN

La encuesta de Consumo Cultural está diseñada para proveer estimaciones a nivel nacional con nivel de confianza de 95% y errores muestrales no superiores al 15% en el cálculo de totales, y no superiores al 10% para el cálculo de proporciones y

promedios. Esto se asegura con pruebas y simulaciones realizadas con base en la información de la encuesta de Consumo Cultural 2007, en donde para diferentes diseños muestrales de distintos tamaños en las etapas tercera y cuarta se estima una convergencia alrededor de 3% del error muestral en el cálculo del total poblacional a nivel nacional.

Cualquier otro tipo de desagregación, como a nivel de municipio o de región puede superar las expectativas del diseño con valores inapropiados de precisión y confianza; pues los tamaños de muestra no alcanzan para satisfacer las expectativas en esos niveles

6. UNIDADES ESTADÍSTICAS

- **Unidad de muestreo:** son los hogares pertenecientes a las cabeceras municipales del país.
- **Unidad de observación:** es la persona mayor de 5 años residente habitual del hogar. Para las personas entre 5 y 11 años es el responsable idóneo que contesta por ellos.
- **Unidad de análisis:** Las personas mayores de 5 años.

7. PERIODO DE REFERENCIA

Últimos doce meses: de septiembre de 2007 a septiembre de 2008.

8. MARCO MUESTRAL

El marco muestral es el dispositivo que permite identificar y ubicar a cada uno de los elementos de la población objetivo, en este caso a cada una de las personas mayores de 5 años que son residentes habituales de los hogares de las cabeceras municipales del país

Para construirlo se recurre a la obtención de información por dos fuentes. La primera es el Censo de Población 2005 con las densidades totales de personas y hogares a nivel de manzana, en donde se hace una asignación de tales densidades a un número de conglomerado para homogenizar las cantidades de hogares y asegurar que cada unidad de muestreo tenga por lo menos, un número de personas satisfactorio de acuerdo con la densidad de población de la cabecera municipal. El conglomerado es una unidad que se construye con la restricción de contener cierto número de hogares y pertenecer a la misma sección cartográfica.

Se hace una partición del marco del censo por densidades de municipios en 6 estratos y se realiza una selección aleatoria simple en cada estrato de municipios. Luego, en cada municipio seleccionado, se hace una selección aleatoria simple de conglomerados y en ellos se realiza enlistamiento para construir la segunda fuente del marco de muestreo, que permite ubicar e identificar a la totalidad de hogares en esos conglomerados.

9. DISEÑO MUESTRAL

La encuesta de Consumo Cultural cuenta con una serie de procedimientos muestrales que le indican al grupo temático, la forma como se va a llegar a campo a realizar mediciones. Además, que los puntos seleccionados en la muestra sean suficientes para proveer resultados a nivel nacional. El diseño muestral cuenta con una metodología de selección por etapas y con una clasificación por estratos (estratificación de unidades).

El objetivo principal del diseño muestral es proveer la estrategia con la cual se hace el acercamiento en campo al fenómeno de interés. Por esta razón, el diseño propuesto tiene muy en cuenta la estructura de universo asumida por el estadístico encargado. Tal estructura es un supuesto que se quiere sea un reflejo más o menos fiel de la forma del universo en la realidad.

En particular, para el desarrollo de la encuesta de Consumo Cultural se hacen las siguientes suposiciones:

- El universo de estudio (tal cual como lo define la ficha metodológica) lo conforman los hogares pertenecientes a las cabeceras municipales del país. En particular, se busca llegar a las personas de cinco años y más de tales hogares.
- Se asume la existencia de estratos de tamaño poblacional en el país para las cabeceras municipales. Es decir, existen grupos de clasificación para homogenizar las cabeceras por cantidad de personas residentes. Los municipios del mismo estrato tienen aproximadamente la misma cantidad de población.
- Se requiere forzosamente obtener información de ciertos municipios del país por razones argumentadas en la prueba piloto y en las definiciones temáticas. Por lo tanto, se debe llegar a ciertos municipios en los que se debe obtener obligatoriamente la información. A estos casos se les denomina de “inclusión forzosa”. Tales municipios se agrupan en un mismo estrato para hacerles un tratamiento de selección de muestra independiente.
- Dentro de cada cabecera municipal existen unidades muestrales denominadas conglomerados. Un conglomerado es una agrupación de manzanas de una misma sección cartográfica (no necesariamente contiguas), que tiene la característica de asegurar que allí se encuentra un número establecido de hogares. Es decir, se sabe de antemano que al llegar a un conglomerado se encuentra un número aproximado de hogares, previamente calculado por metodologías muestrales.
- Evidentemente, cada conglomerado tiene un total de hogares establecido a través del Censo de Población. Dentro de cada hogar, a su vez, hay un número de personas (mayores y/o menores de cinco años). El interés se centra en aquellos individuos que tienen cinco o más años de edad.

Tomando como base estos supuestos acerca de la población objetivo, se procede a plantear el diseño muestral a seguir:

- 1) Como las cabeceras están agrupadas en estratos a nivel nacional, se decide hacer muestreo de cabeceras dentro de cada estrato, de forma independiente. Este proceso determina la primera etapa de selección.
- 2) Una vez se ha hecho el muestreo en los estratos de cabeceras, en cada cabecera se hace muestreo de conglomerados.
- 3) Muestrear hogares dentro de los conglomerados seleccionados en la etapa anterior. Establecer la información insumo de la encuesta para todas las personas de cinco años o más, residentes en tales hogares.

Como se deduce fácilmente, este diseño involucra tres etapas de selección de muestra y en cada etapa se seleccionan unidades muestrales diferentes.

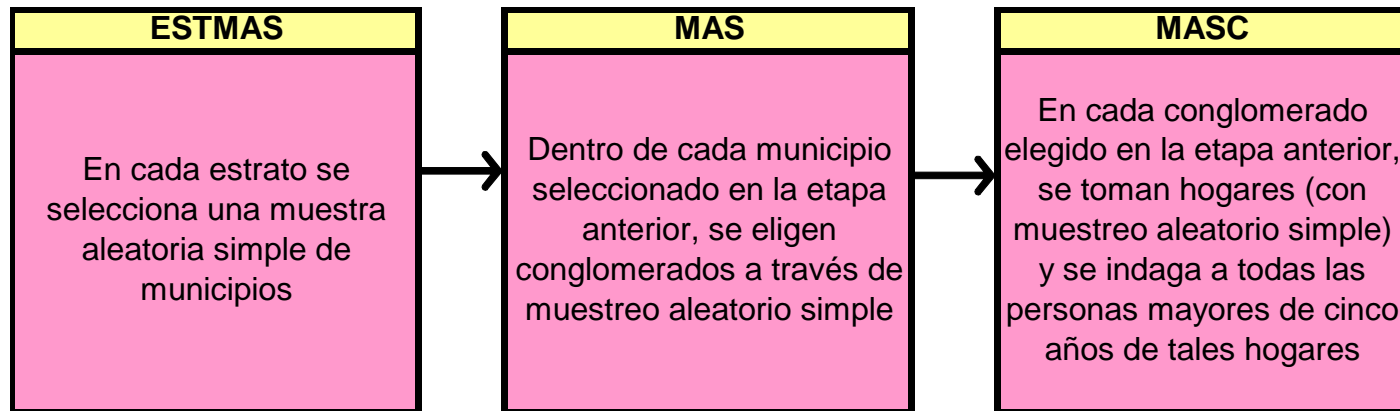
El tipo de diseño muestral a emplear en cada etapa se elige con base en observaciones hechas de una variable de estudio de interés, en estudios anteriores; o también tomando como base resultados obtenidos de una prueba piloto anterior que permita observar indicadores como el coeficiente de correlación intraclásica o las varianzas obtenidas por estrato de selección.

Finalmente, el diseño muestral que se utiliza en la Encuesta de Consumo Cultural 2008 es un ESTMAS – MAS – MASC en tres etapas:

- ESTMAS hace referencia a la primera etapa del diseño en donde se selecciona una muestra, a través de muestreo aleatorio simple (MAS), de municipios dentro de cada estrato.
- MAS se refiere a la segunda etapa del diseño, en la que se seleccionan conglomerados utilizando muestreo aleatorio simple, dentro de cada municipio escogido en la etapa anterior.
- MASC significa que en la tercera etapa se seleccionan de forma aleatoria simple hogares y se indaga a TODAS las personas de cinco años o más de tales hogares. Dentro de la teoría de muestreo esto se conoce como *diseño por conglomerados*.



ESQUEMA DE DISEÑO MUESTRAL ENCUESTA DE CONSUMO CULTURAL 2008



Nótese que este diseño indica la forma como se llega finalmente a la unidad de observación objetivo (persona de cinco años o más), a partir de una estructura sugerida para el universo de estudio. Se aclara que el diseño deja indicada la muestra que se escoge de acuerdo con los supuestos hechos. La muestra será efectiva en la medida que los supuestos se acerquen a la realidad del universo de estudio, en cada etapa de muestreo.

9.1. TAMAÑO DE LA MUESTRA

Una vez se define el diseño muestral a utilizar, se procede a fijar los tamaños de muestra pertinentes en cada etapa. La discusión respecto al tamaño de muestra es fundamental en la elaboración de cualquier investigación, porque de este aspecto se derivan los costos de una encuesta; pero también es una elección delicada cuando se trata de precisión en los resultados finales, confiabilidad de la muestra seleccionada y peso asignado a cada unidad muestral.

Un mayor tamaño de muestra mejora la precisión y confiabilidad de las estimaciones realizadas, por lo que el peso de cada individuo es menor (tanto, que podría tener una representación igual a la unidad: CENSO). Aumentar el tamaño de muestra también ayuda a que la pérdida de información de un individuo no sea tan grave, dado que en muestras pequeñas, como cada individuo se representa a sí mismo y muchos más del universo, que no están en la muestra, perder la información de uno de ellos implica una pérdida grave al generalizar hacia el universo.

Sin embargo, mayor tamaño de muestra implica más encuestas, más trabajo de campo, labores operativas, trabajo de sistemas, informático y de personal, es decir, costos muy elevados con relación al beneficio obtenido y se incurre en una pérdida de esfuerzo y recursos. Además, aunque es cierto que los errores muestrales disminuyen a mayor tamaño de muestra, es claro también que la gran cantidad de procesos que involucra el manejo de encuestas con tamaños de muestra grandes, induce un mayor error considerado como *no-muestral*, derivado de la amplia manipulación de la información recolectada. Este hecho se puede controlar ejecutando procesos de calidad, los que imputan aún más costos a la investigación.

Lo que se hace es precisar un equilibrio entre costo y beneficio. El costo muchas veces está limitado por un presupuesto fijo; y el beneficio es simulado tras intentar muchos tamaños de muestra para una estrategia muestral. Nótese que con cada elección de tamaño de muestra, también se está diciendo que el diseño muestral adopta distintas formas. Así pues, deben probarse muchos diseños muestrales hasta dar con uno óptimo en la precisión de resultados y costos.

Respecto a los parámetros de dichas simulaciones, en general estos son procesos computacionales diseñados bajo unas normas y supuestos de investigación. La simulación trata de imitar al universo de estudio de donde se extraen muestras con distintas estrategias muestrales, y se evalúan precisión y confiabilidad de diferentes estrategias muestrales, usando como base tal construcción.

En el caso específico de la Encuesta de Consumo Cultural, los tamaños de muestra por etapas están dados por:

Tamaños de muestra ETAPA 1		
ESTRATO	No MPIOs	Muestra MPIOs
1 (IF)	36	37
2	50	10
3	93	8
4	180	6
5	317	5
6	423	2
Total general	1099	68

Tamaños esperados de muestra ETAPA 2		
ESTRATO	No Conglomerados	Conglom seleccionados
1 (IF)	33646	471
2	1903	149
3	589	55
4	246	26
5	90	16
6	17	8
Total general	36491	725

Para establecer el marco muestral de hogares, se lleva a cabo una etapa en campo denominada *enlistamiento*. En este proceso se contabilizan y ubican la totalidad de hogares existentes dentro de los conglomerados que hacen parte de la muestra. El propósito de dicho enlistamiento es obtener información más confiable y actualizada de la distribución, densidad y ubicación precisa de la población o universo de estudio que se encuentra en esta instancia del diseño.

Por los resultados obtenidos en la prueba piloto con el indicador de correlación intraclase, se mantiene el diseño de conglomerado dentro de los hogares, es decir, se indaga a todas las personas mayores de cinco años de los hogares. El presupuesto fijado permite encuestar aproximadamente 20.000 personas repartidas en alrededor de 5.000 hogares.

9.2. DISEÑO Y SELECCIÓN DE LA MUESTRA

De acuerdo con la teoría de muestreo el factor de expansión es la capacidad que tiene cada individuo seleccionado en una muestra probabilística para representar el universo en el cual está contenido. Es decir, es la magnitud de representación que cada selección posee para describir una parte del universo de estudio. Cuando el diseño es MAS se asume que individuos dentro de una misma unidad de muestreo tienen la misma capacidad de representar al universo en consideración, en tanto que diferentes unidades

de muestreo deben reflejar lo mejor posible la densidad y distribución del universo estudiado.

Todas las pautas anteriores se logran con la construcción u obtención de marcos muestrales que deben proveer información fiel sobre las características demográficas principales del universo que se pretende estudiar. Además debe permitir de alguna forma, ubicar a todos y cada uno de los individuos pertenecientes a dicho universo.

Un diseño muestral MAS se traduce como un diseño de muestreo aleatorio simple y se dice que dicho diseño es probabilístico porque asigna una probabilidad de selección a todas y cada una de las unidades del universo o la unidad muestral (cuando el diseño es en etapas). Es decir, todos los individuos en consideración de los que se pretende inferir tienen una probabilidad de ser seleccionados; en particular esta probabilidad es la misma con el diseño MAS.

El factor de expansión por teoría para un diseño MAS sobre k unidades de muestreo está definido por:

$$\pi_{1k} = \frac{N_{1k}}{n_{1k}}$$

Donde

N_{1k} Denota el tamaño total de elementos en la unidad de muestro.

n_{1k} Denota el número de elementos a ser seleccionados dentro de la unidad de muestreo.

9.3. METODOLOGÍA DE ESTIMACIÓN Y VARIANZA

Es decir, para el diseño en consideración de la encuesta de Consumo Cultural *ESTMAS – MAS – MASC* y de acuerdo con los diagramas presentados anteriormente, se tienen los factores de expansión por etapas de la siguiente manera:

ETAPA 1: Selección de municipios dentro de los estratos

$$f \exp_{I(i)} = \frac{N_{I(i)}}{n_{I(i)}} \text{ para } i = 1, 2, 3, 4$$

con

$N_{I(i)}$: Cantidad de municipios en el estrato i

$n_{I(i)}$: Número de municipios seleccionados en el estrato i

ETAPA 2: Selección de conglomerados dentro de municipios

$$f \exp_{II(j)} = \frac{N_{II(j)}}{n_{II(j)}} \text{ para } j = 1, \dots, n_{I(i)}$$

con

$N_{II(j)}$: Cantidad de conglomerados en el municipio j

$n_{II(j)}$: Cantidad de conglomerados seleccionados en el municipio j

ETAPA 3: Selección de hogares dentro de conglomerado

$$f \exp_{III(k)} = \frac{N_{III(k)}}{n_{III(k)}} \text{ para } k = 1, \dots, n_{II(j)}$$

con

$N_{III(k)}$: Cantidad de hogares dentro del conglomerado k

$n_{III(k)}$: Cantidad de hogares seleccionados dentro del conglomerado k

Nota: El factor de expansión correspondiente a cada individuo dentro de un hogar seleccionado es igual a la unidad ($f \exp_{individuo} = 1$), puesto que se eligen todas las personas de cinco años y más, es decir, esta parte del diseño es la que se conoce como muestreo por conglomerados.

9.3.1. Factor de expansión final

El factor de expansión final se define como el producto de los factores de expansión de las tres etapas de diseño:

$$F \exp = (f \exp_{I(i)}) (f \exp_{II(j)}) (f \exp_{III(k)})$$

Este factor se aplica a cada uno de los individuos pertenecientes a los hogares seleccionados en la última etapa del diseño. La suma de los factores de expansión de todos los individuos de la muestra provee una estimación del tamaño del universo de estudio.

9.3.2. Estimadores de Totales y Razones

Con los factores de expansión calculados se define el estimador del total de una variable de estudio X (t_x) en el universo, en función de los valores observados con la muestra, de la siguiente manera:

$$\hat{t}_x = \hat{t}_\pi(x) = \sum_{k \in s_\Omega} (F \exp)^* x_k$$

Donde s_Ω es una muestra del universo de estudio con valores particulares medidos x_k .

Y el estimador de la varianza para la anterior expresión es:

$$\hat{Var}(\hat{t}_x) = \frac{N^2(1-f)}{n} S^2_{xk} \quad \text{Con} \quad S^2_{xk} = \frac{\sum (x_k - \bar{x}_s)^2}{(n-1)} ; f = \frac{n}{N}$$

con N y n tamaños de universo y muestra respectivamente; siempre que el diseño de muestreo sea MAS.

Si Y es otra variable de estudio, en general el estimador de la razón $Z = \frac{t_y}{t_x}$ se define por

$$\hat{Z} = \frac{\hat{t}_\pi(y)}{\hat{t}_\pi(x)}$$

el cual no es un estimador lineal, luego el cálculo de la estimación de su varianza difiere de los métodos convencionales. Aplicando el método de linearización de Taylor y construyendo una nueva variable que recoge dicho término de varianza, se obtiene un estimador de la varianza del estimador.

El estimador de la varianza para el estimador de razón es

$$\hat{Var}(\hat{z}) = \frac{1}{\hat{t}_\pi^2} \left[\hat{V}(\hat{t}_y) + \hat{z}^2 \hat{V}(\hat{t}_x) - 2\hat{z} \text{Cov}(\hat{t}_y; \hat{t}_x) \right]$$

o calculándose directamente con la fórmula común de varianza en el diseño MAS, a través de la construcción de una nueva variable U para cada uno de los elementos de la muestra

$$u_k = \frac{1}{\hat{t}_x} y_k - \frac{\hat{t}_y}{\hat{t}_x^2} x_k$$

Los promedios y las proporciones se trabajan como casos particulares del estimador de razón cuando los numeradores y los denominadores cambian en concepto, y estos pueden pasar de ser variables continuas a categóricas.

Una forma de generar estimaciones para diferentes proporciones o promedios (dependiendo del requerimiento de investigación) es mediante la formación de variables indicadoras de dominio que toman el valor de 1 si un individuo de la muestra pertenece a un dominio de interés y 0 en caso contrario.

Entiéndase al dominio de un universo como un conjunto de individuos contenido en éste, que presentan una característica discriminatoria Z que los ubica dentro de una categoría particular observada.

La variable indicadora se establece como

$$I^z_k = \begin{cases} 1 & \text{si el individuo k tiene la característica Z} \\ 0 & \text{en caso contrario} \end{cases}$$

Es así que el estimador del total para la variable I^z_k se convierte en el estimador del total de individuos del universo perteneciente a un dominio Z; y en particular, si I^z_k siempre es 1 para cualquier individuo de la muestra, es simplemente el estimador del total de individuos en el universo.

Las proporciones y los promedios solicitados en cuadros de salida, se construyen formando cocientes de diversos estimadores de dominios y de totales. Las formulas no varían.

Para el diseño multi-etápico *ESTMAS – MAS – MASC* las formulas de estimación por etapas están definidas como sigue

- *Etapa 1:* $\hat{t}_{I(i)x} = \sum_{j \in sI} \frac{N_{I(i)}}{n_{I(i)}} (\hat{t}_{II(i)(j)x})$ con estimador de varianza

$$\hat{V}_I(\hat{t}_{I(i)x}) = \left(\frac{N_{I(i)}^2}{n_{I(i)}} \right) \left(1 - \frac{n_{I(i)}}{N_{I(i)}} \right) S_{sI}^2(\hat{t}_{II(i)(j)x}) + \sum_{j \in sI} \frac{N_{I(i)}}{n_{I(i)}} \hat{V}_{II}(\hat{t}_{II(i)(j)x})$$

$$\text{donde } S_{sI}^2(\hat{t}_{II(i)(j)x}) = \frac{\sum (\hat{t}_{II(i)(j)x} - \bar{\hat{t}}_{II(i)x})^2}{(n_{I(i)} - 1)}$$

- *Etapa 2:* $\hat{t}_{II(i)(j)x} = \sum_{k \in sII} \frac{N_{II(j)}}{n_{II(j)}} (\hat{t}_{III(i)(j)(k)x})$ con estimador de varianza

$$\hat{V}_{II}(\hat{t}_{II(i)(j)x}) = \left(\frac{N_{II(j)}^2}{n_{II(j)}} \right) \left(1 - \frac{n_{II(j)}}{N_{II(j)}} \right) S_{sII}^2(\hat{t}_{III(i)(j)(k)x}) + \sum_{k \in sII} \frac{N_{II(j)}}{n_{II(j)}} \hat{V}_{III}(\hat{t}_{III(i)(j)(k)x})$$

$$\text{donde } S_{sII}^2(\hat{t}_{III(i)(j)(k)x}) = \frac{\sum (\hat{t}_{III(i)(j)(k)x} - \bar{\hat{t}}_{III(i)(j)x})^2}{(n_{II(j)} - 1)}$$

- *Etapa 3:* $\hat{t}_{III(i)(j)(k)x} = \sum_{l \in sIII} \frac{N_{III(k)}}{n_{III(k)}} (\hat{t}_{IV(i)(j)(k)(l)x})$ con estimador de varianza

$$\hat{V}_{III}(\hat{t}_{II(i)(j)(k)x}) = \left(\frac{N_{III(k)}^2}{n_{III(k)}} \right) \left(1 - \frac{n_{III(k)}}{N_{III(k)}} \right) S_{sIII}^2(\hat{t}_{IV(i)(j)(k)(l)x}) + \sum_{k \in sIII} \frac{N_{III(k)}}{n_{III(k)}} \hat{V}_{IV}(\hat{t}_{IV(i)(j)(k)(l)x})$$

$$\text{donde } S_{sIII}^2(\hat{t}_{IV(i)(j)(k)(l)x}) = \frac{\sum_{l=1}^s (\hat{t}_{IV(i)(j)(k)(l)x} - \bar{\hat{t}}_{IV(i)(j)(k)x})^2}{(n_{III(k)} - 1)}$$

Finalmente, para obtener la estimación general a la que se quiere llegar, se tiene:

$$\hat{t}_x = \sum_{i=1}^3 (\hat{t}_{I(i)x})$$

Con estimador de varianza

$$\hat{V}_I(\hat{t}_{I(i)x}) = \sum_{l=1}^3 \hat{V}_I(\hat{t}_{I(i)x})$$

Si el estimador resulta ser de razones, las formulas para el calculo de la varianza no se alteran si se emplean en lugar de x_l los valores de u_l .

9.3.3. Ajuste del factor de expansión final

Como ya se mencionó antes, el factor de expansión es un término multiplicativo que le corresponde a cada individuo de la muestra para expandir su observación en extrapolación a la población de interés. Se habla del factor de expansión como la magnitud que le corresponde a cada elemento de la muestra para representar una porción del universo.

Por lo general, cuando el diseño muestral consta de la metodología MAS como ya se vio antes, y se dispone de información de los tamaños poblacionales se dice que la capacidad de representación de cada individuo o su factor de expansión es de la forma:

$$f \text{ exp} = \frac{N}{n}$$

El cual es un término que esta en función de las probabilidades de inclusión de cada individuo de una población objetivo a pertenecer a la muestra, pues el muestreo es probabilístico

Por otro lado, los factores de ajuste son aquellos que se proponen a partir de la validación de los resultados con datos externos para hacer coherente la descripción de la población. Por lo general, un factor de ajuste lo que hace es comparar el valor estimado de una variable de interés, con este mismo valor estimado de otra encuesta o con un dato más confiable del resultado del mismo. Es decir:

$$f_{ajust} = \frac{\hat{\theta}}{\theta_{ext}}$$

Entonces, este valor asegura que el estimativo de una variable θ coincida con el resultado del valor externo, pero implica que todas las demás estimaciones cambiaran:

$$\hat{\theta}_{ajust} = \theta_{ext}$$

Finalmente, se habla del factor de corrección como el término que ajusta por pérdida de individuos al factor de expansión. Nótese que el factor de expansión depende de la cantidad de individuos incluidos en la muestra, pero cuando no todos los individuos de la muestra fueron obtenidos, el factor sobredimensiona la pérdida de dichos individuos ampliando la representación de una no respuesta por cientos o miles.

Para evitar dicho inconveniente de algún modo, se recurre a la corrección del factor de expansión por un procedimiento de selección de muestra en dos fases, el cual consiste en asumir que la recolección fue una fase del diseño estadístico y que después hubo otra fase de selección en donde se volvió a seleccionar muestra y entonces quedaron los individuos finales. El factor de corrección entonces adopta la forma:

$$f_{corr} = \frac{\sum M}{\sum i}$$

Donde $\sum M$ hace referencia al tamaño total de la muestra que debió haberse obtenido e $\sum i$ hace referencia a la cantidad de individuos que finalmente se entrevistan

10. COEFICIENTE DE VARIACIÓN

Cuando la magnitud de la variabilidad es muy grande los parámetros estimados pierden utilidad, pues el valor verdadero del parámetro, en el universo, puede estar en un intervalo muy amplio lo cual no proporciona información útil. Por lo tanto esta medida de variabilidad es un criterio muy importante para determinar la calidad de una estimación.

Pero el valor de esta magnitud está directamente relacionada con la unidad de medida de la variable de interés, por ejemplo si se está estimando el total de los ingresos para una actividad económica dada la unidad de medida en miles de pesos, y como además la varianza está en unidades al cuadrado, se estaría hablando de miles de pesos al cuadrado. La raíz cuadrada de esta varianza es la que se denomina Desviación estándar de la distribución o error estándar. Esta medida de dispersión tiene la ventaja de que la unidad de la medida de dispersión corresponde a la unidad de la variable de interés, así por ejemplo, una variabilidad de cien en un millón en la estimación del total es pequeña lo cual indicaría que la estimación tiene muy buena precisión, pero si el parámetro a estimar es el promedio de personas ocupadas por empresas, tener una variabilidad de cinco

personas es alta, y aunque es un valor muchísimo menor puede significar que no es tan preciso.

Un criterio más uniforme para determinar la precisión de un estimador es el coeficiente de variación el cual se define como la variación porcentual del error estándar a la estimación central, es decir es el cociente entre el error estándar del estimador y el estimador multiplicado por 100.

$$CV = \frac{\sqrt{V(\hat{\theta})}}{E(\hat{\theta})} * 100$$

el coeficiente de variación estimado está dado por

$$c\hat{v} = \frac{\sqrt{\hat{V}(\hat{\theta})}}{(\hat{\theta})} * 100$$

Aunque la varianza, el error estándar y el coeficiente de variación, miden la magnitud de la variabilidad de la distribución muestral del estimador, es decir, lo que algunos denominan el error de muestreo, el coeficiente de variación tiene la ventaja de dar esta medida en términos porcentuales, por lo cual es una medida común para estimaciones.

Se suele considerar que el resultado de una estimación es bueno si su coeficiente de variación es menor del 5 %; aceptablemente práctico, entre el 5 % y el 10%; de baja precisión si es mayor del 10 %. Y menor del 15% y no útil si es mayor del 15%.

Para el mejor entendimiento del significado y los diferentes valores que toman los coeficientes de variación en los cuadros presentados se debe tener en cuenta los siguientes aspectos: el diseño de la muestra se realiza para obtener estimaciones con alta precisión a nivel entidad. Cualquier otro nivel de desagregación o parámetro estimado está sujeto a que su precisión no necesariamente sea buena y por tanto el dato no sea publicable.



BIBLIOGRAFÍA

Bautista S., Leonardo, Diseños de Muestreo Estadístico, Universidad Nacional de Colombia , Departamento de Matemáticas y Estadística, (1998).

Särndal, Carl Erik, Swenson, B. , Wretman, J. Model Assisted Survey Sampling, Springer-Verlang, New York (1992).