

**Departamento Administrativo
Nacional de Estadística**



Dirección de Censos y Demografía

**ESTIMACIÓN E INTERPRETACIÓN DEL
COEFICIENTE DE VARIACIÓN DE LA
ENCUESTA COCENSAL**

CENSO GENERAL 2005 - CGRAL

Junio de 2008

INTRODUCCIÓN

El Censo General 2005 que se llevó a cabo en Colombia entre los meses de mayo de 2005 y mayo de 2006, busca llenar los vacíos de información estadística de interés para el país que se presentan luego de 12 años del último Censo de Población y Vivienda.

Además de suplir una buena cantidad de las actuales necesidades de información estadística del país, el Censo General 2005 incorpora innovaciones metodológicas con las que espera contribuir a la generación y adopción de metodologías modernizadas para efectuar Censos en los demás países del mundo.

En lo referente a la información estadística del país, se llevo a cabo una metodología censal estándar, y una Encuesta Cocensal aplicada a una muestra de hogares, viviendas y personas para cada uno de los municipios en Colombia.

Por tal motivo se realizó dos tipos de cuestionarios: Uno corto, con toda la información demográfica básica (Censo básico) realizado a toda la población y uno largo que permite recolectar profundizaciones temáticas en lo concerniente a Población, Hogar y Vivienda (Censo ampliado).

En este documento, se describe brevemente algunos aspectos generales sobre la metodología de la Encuesta Censal y la interpretación de sus resultados, a la luz del diseño muestral.

DISEÑO MUESTRAL

La muestra cocensal, es una muestra probabilística de hogares seleccionada en tiempo real durante el operativo del Censo que entrega estimaciones en las grandes ciudades, a niveles de comuna o localidad y, a nivel de cabecera y resto para los demás municipios.

Para la selección de muestra se empleó un diseño *Estratificado Bernoulli* donde la unidad de selección es el hogar y se estratifica por municipio, cabecera y resto, y para las grandes ciudades (Bogotá, Medellín, Cali y Barranquilla) con estructura de localidades o comunas, según corresponda.

Es así como, de los 1179 municipios, comunas y corregimientos departamentales que existían en la época, se tienen:

- 625 municipios de inclusión forzosa, es decir, se aplicó el cuestionario ampliado en todo el municipio.

- 57 municipios con muestra en el resto y aplicación de ampliado en toda la cabecera.
- 119 municipios con muestra en la cabecera y aplicación de ampliado en todo el resto.
- 378 municipios restantes, tuvieron selección de muestra en la cabecera y en el resto.

Dado que en el Censo existen tres tipos de unidades de observación: La vivienda, el hogar y la persona y la unidad de selección de la muestra es el hogar, entonces, una vez que un hogar ha sido seleccionado, todas las personas miembros del hogar lo son también, mientras que una vivienda se considera como seleccionada cuando al menos uno de los hogares que la habita ha sido seleccionado. En este orden de ideas:

- Los hogares son seleccionados bajo un diseño *Estratificado Bernoulli de elementos*.
- Las personas se seleccionan bajo un *diseño Estratificado Bernoulli de conglomerados*.
- Las viviendas son seleccionadas bajo un diseño *Poisson de elementos*.

Para calcular los tamaños de muestra, se utilizaron las proyecciones de población a nivel municipal para el año 2005, tamaños promedio de hogar a nivel de cabecera y resto municipal y las fórmulas de tamaño de muestra del muestreo aleatorio simple de elementos, suponiendo la selección de una muestra de hogares y la estimación de parámetros del tipo proporción (con denominador conocido) para categorías de variables con una presencia del fenómeno de al menos el 10% en la población de estudio y un coeficiente de variación (error de muestreo) de 7% o menos a nivel municipal.

PROCEDIMIENTO DE ESTIMACIÓN

La estimación de las diferentes características poblacionales o parámetros se realizó a partir de la metodología de “Estimadores de calibración” propuesta por Särndal, y Deville en 1992, la cual proporciona estimaciones validas y precisas. Lo que significa obtener estimaciones con errores de muestreo minimizados.

El software recomendado para llevar a cabo estos procedimientos es la **macro CLAN**, desarrollada por *Statistics Sweden*, la cual es utilizada para ejecutar la calibración de los factores de expansión básicos y con ello, el ajuste por estructura de la muestra cocensal, también permite la estimación de parámetros y sus correspondientes errores de muestreo.

En el documento: “*Metodología estadística de diseño y estimación de la muestra cocensal*”, realizado por la Dirección de Censos y Demografía del Dane, se explica en detalle:

- La Encuesta Cocensal.
- Diseño muestral.
- Cálculo del tamaño de la muestra.
- Estimación del estimador y estimación de varianza de los principales tipos de parámetros, para los diferentes niveles de desagregación utilizados.

COEFICIENTE DE VARIACIÓN

Las estimaciones hechas a través de muestra tienen dos tipos de error, que son:

- Errores ajenos al muestreo: Se presentan por procedimientos de observación imperfectos, estos no dependen directamente del desarrollo matemático, esta clase de errores es inherente a toda investigación estadística. En el Censo General 2005 se minimizó cualquier error de este tipo.
- Errores de muestreo: Se presentan debido a que solo se estudia una fracción de la población total. Los errores de este tipo, se pueden estimar a partir de los resultados obtenidos y permite evaluar la calidad de la estimación, siempre y cuando el diseño muestral sea probabilístico.

La forma de medir el error muestral se hace a través del **coeficiente de variación estimado** (*cve*), el cual mide la magnitud de la variabilidad de la distribución muestral del estimador, es decir, es el indicador del grado de aproximación con que se estiman las características del universo y esta dado por:

$$cve = \frac{\sqrt{\hat{V}(\hat{\theta})}}{\hat{\theta}} * 100,$$

Donde $\hat{V}(\hat{\theta})$ es la varianza muestral de la estimación del parámetro y $\hat{\theta}$ es el parámetro estimado, o, en otras palabras, se define como la variación porcentual del error estándar con respecto a su estimación, es decir es el cociente entre el error estándar del estimador y el estimador multiplicado por 100.

INTERPRETACIÓN DEL COEFICIENTE DE VARIACIÓN

El coeficiente de variación es una calificación que permite a los usuarios evaluar la calidad estadística de las estimaciones.

Para la muestra cocensal se considera que una estimación con un coeficiente de variación:

- Hasta del 7%, es precisa;
- Entre el 8 y el 14% significa que existe una precisión aceptable;
- Entre el 15% y 20% precisión regular y por lo tanto se debe utilizar con precaución Mayor del 20% indica que la estimación es poco precisa y por lo tanto se recomienda utilizarla sólo con fines descriptivos (tendencias no niveles).

El diseño de la muestra cocensal se realizó para obtener estimaciones con alta precisión, sin embargo, cualquier otro nivel de desagregación está sujeto a que su precisión no necesariamente sea buena y por tanto la cifra no sea de gran utilidad (a mayor nivel de desagregación, menor será el nivel de precisión), por lo tanto es de vital importancia que el investigador que tome los resultados estimados a partir de la Encuesta Cocensal tenga presente su respectivo coeficiente de variación – *cve*- y evalúe el nivel de precisión de dicha estimación.

Un ejemplo de cómo se debe interpretar el *cve* se presenta en la tabla 1, donde se observa la cantidad de hogares que tienen ducha eléctrica para los departamentos de Caquetá y Cundinamarca por área. Los coeficientes de variación de Caquetá son superiores a 20%, (esto se debe a la baja frecuencia de ocurrencia del fenómeno en el departamento), luego estas estimaciones son poco precisas, indicando que esta información se debe utilizar sólo con fines descriptivos.

Por otro lado los *cve* de la variable en el departamento de Cundinamarca están dentro de los límites establecidos, luego representan alta precisión en las estimaciones obtenidas.

Tabla 1. Cantidad de hogares con ducha eléctrica por departamento y área

Nombre del Departamento	Hogares que tienen ducha electrica					
	Total		Cabecera		Resto	
	Estimación	<i>cve</i>	Estimación	<i>cve</i>	Estimación	<i>cve</i>
Cundinamarca	133705	0,99%	109240	1,19%	24465	1,02%
Caquetá	471	19,13%	409	21,63%	62	28,16%

Fuente: Censo General 2005, Encuesta Cocensal

En razón de lo anterior, se presenta una metodología alternativa que facilita la lectura e interpretación de los coeficientes de variación a través de los **intervalos de confianza**.

El intervalo de confianza para un estimador, es un rango dentro del cual los valores no pueden ser considerados estadísticamente diferentes unos de otros, y se encuentran sujetos a un nivel de confiabilidad (generalmente este nivel es del 95%).

La forma de calcular el intervalo es la siguiente:

1. Se toma el valor puntual de la estimación.
2. Se calcula el error estándar del estimador, definido como:

$$error\ std = cve * \hat{\theta}$$

3. A la estimación se le suma un valor correspondiente a:

$$\pm 1.96 * error\ std$$

Siendo el intervalo de confianza (I.C.):

$$I.C. = \hat{\theta} \pm 1.96 * error\ std$$

El valor de 1.96 corresponde al valor crítico a dos colas de una distribución normal con media 0 y varianza 1 y un nivel de confianza de 95%.

A modo de ejemplo, para los datos que se encuentran en la tabla 1, en la cabecera del departamento Caquetá, su intervalo de confianza (I.C.) se calcula así:

1. $\hat{\theta} = 409$
2. $error\ std = 0.2163 * 409$
 $= 88.46$
3. $I.C. = 409 \pm 1.96 * 88.46$

Lo anterior indica que el verdadero valor del parámetro se encuentra entre **235.61** y **582.38** con un nivel de confiabilidad del **95%**.

El usuario puede proceder de la misma forma para construir el intervalo de confianza, para cualquier tipo de parámetro y nivel de confianza (entre más amplio sea el intervalo, la estimación del parámetro es menos precisa).

Finalmente se realizó un análisis por municipio en las diferentes variables que pertenecen a la Encuesta Cocensal con base en los coeficientes de variación, donde se encontró que la mayoría de los municipios presentan fenómenos con una presencia superior al 10% y un coeficiente de variación de menor del 7%, indicando que los tamaños de muestra son suficientes para representar a la totalidad de la población municipal.

BIBLIOGRAFÍA

1. Särndal, Carl Erik, Swenson, B. ,Wretman, J. Model Assisted Survey Sampling, Springer-Verlang, New York (1992).
2. Bautista S., Leonardo, Diseños de Muestreo Estadístico, Universidad Nacional de Colombia , Departamento de Matemáticas y Estadística, (1998)
3. Luna, Angela, Censo General 2005: Un proyecto con enfoque sistemático, Revista ib No. 1, CANDANE, Bogotá, Colombia 2006
4. 2001 Census Technical Report. Sampling and Weighting. Statistics Canada. Canadá
5. Metodología estadística de diseño y estimación de la muestra cocensal, Dane, Censo General 2005, 2006.