

**Departamento Administrativo
Nacional de Estadística**



**Encuesta Nacional de Calidad de Vida
Metodología Muestral**

EQUIPO DE DISEÑOS MUESTRALES

Septiembre 2003

1. DISEÑO ESTADÍSTICO

1.1 GENERALIDADES.

Para la Encuesta de Calidad de Vida se tomo una submuestra de la muestra maestra de población de propósitos múltiples del DANE, la cual fue ampliada en términos de cobertura en el año 2000.

La muestra maestra fue utilizada por primera vez para la Evaluación de la Cobertura del Censo de población y vivienda de 1993, se incorporó al Sistema de Encuestas de Hogares para medir la fuerza de trabajo a partir de 1996 y también en la Encuesta de Calidad de Vida del año 1997.

1.2 PREMISAS DEL DISEÑO

1.2.1 Universo

El universo, que coincide con la población objetivo, está constituido por la población civil en hogares no institucionales, residentes en las viviendas particulares.

1.2.2 Cobertura

La cobertura de la ECV es nacional y corresponde a las zonas rural y urbana de los 1117 municipios del país.

1.2.3 Unidades Estadísticas

1.2.3.1 Unidad de muestreo

Estrato 1.

Primera etapa: áreas metropolitanas.

Segunda etapa: Como unidades secundarias de muestreo (USM's) se seleccionan manzanas o grupos de manzanas las cuales contienen por lo menos un segmento.

Tercera etapa: Como unidades terciarias de muestreo UTM's, segmentos.

Estrato 2.

Primera etapa: municipios o grupo de municipios correspondientes a UPMs que conforman la muestra maestra, de la cual se toma una submuestra.

Segunda etapa: manzanas o grupos de manzanas a manera de unidades secundarias de muestreo (USM's).

Tercera etapa: segmentos a modo de Unidades Terciarias de Muestreo (UTM).

1.2.3.2 Unidad de observación

Es el elemento básico al cual se le aplica el cuestionario, en este caso, el hogar.

1.2.3.3 Unidad de análisis

Hogares y personas.

1.2.4 Grado de Precisión

La precisión de las estimaciones de la ECV se mide en términos del error estándar relativo, que se espera sea del 5% para cabecera y del 7% para el resto.

1.2.5 Parámetros a ser Estimados.

Los parámetros que se estiman en la ECV básicamente toman la forma de totales y razones. Las variables de interés indagan a nivel de personas, sobre condiciones de pobreza, alfabetismo, asistencia escolar, mercado laboral, tenencia de bienes, tipo de afiliación al sistema general de seguridad social en salud. A nivel de hogares, cantidad de cuartos por vivienda, forma de eliminación de basuras, tenencia de la vivienda, materiales predominantes de las paredes y pisos, tenencia de servicios públicos, tipo de sanitario y fuente de aprovisionamiento de agua para la preparación de alimentos, entorno, seguridad y convivencia.

1.3 MARCO ESTADÍSTICO

El marco estadístico, está constituido por el inventario cartográfico y el listado agregado de viviendas y hogares a nivel de manzana, obtenidos de la información del Censo Nacional de Población y Vivienda de 1993, con actualizaciones cartográficas.

1.4 DISEÑO MUESTRAL

1.4.1 Método de muestreo

Muestreo probabilístico, estratificado, multietápico de conglomerados.

- **Probabilístico**

Es probabilístico porque cada unidad de muestreo tiene una probabilidad de selección conocida y mayor que cero. Esta información permite determinar *a priori* la precisión deseada en las estimaciones y posteriormente calcular la precisión de los resultados obtenidos a partir de la información recolectada.

- **Estratificado**

La población objetivo está clasificada en función de variables altamente correlacionadas con los parámetros a ser estimados, que comprenden dos estratos:

Estrato 1: conformado por el área urbana de los 33 municipios de las trece áreas metropolitanas.

Estrato 2: corresponde a las áreas rural y urbana de los restantes 1084 municipios del país, controlando las variables: población en cabecera, población en resto y tasa de NBI por municipio.

Se construyen 228, subestratos teniendo en cuenta que en cada unidad primaria de muestreo disponga de una población mínima de 3000 habitantes, con 2000 en cabecera. Esta subestratificación se hace agrupando UPM's municipios o grupo de municipios que tengan una cantidad de habitantes parecido tanto en cabecera como en resto, al igual que proporciones similares de NBI.

- **Multietápico**

Es Multietápico porque la selección de las unidades de muestreo se lleva a cabo en varias etapas.

- **Conglomerados**

En cada estrato, finalmente se seleccionan en cada manzana o grupos de manzanas segmentos de 10 hogares, en los cuales todos son encuestados.

El estudio de la muestra por conglomerados, aunque aumenta el error de muestreo, reduce considerablemente los costos de la operatividad de la encuesta.

Para compensar su desventaja es necesario aumentar razonablemente el tamaño de la muestra.

1.4.2 Tamaño de muestra.

El tamaño de la muestra a nivel de región, por segmento se distribuye de la siguiente manera:

Cuadro N° 1. Tamaño de muestra

REGIÓN	CANTIDAD DE SEGMENTOS		
	TOTAL	CABECERA	RESTO
	2.407	1.927	480
Atlántico	160	80	80
Pacífico	175	95	80
Central	160	80	80
Oriental	160	80	80
Orinoquía y Amazonía	42	42	
San Andrés y Providencia	42	42	
Bogotá	1.318	1.318	
Antioquia	175	95	80
Valle del Cauca	175	95	80

1.4.3 Método de selección

Estrato 1.

Primera etapa: inclusión forzosa de las 13 áreas metropolitanas.

Segunda etapa: Como unidades secundarias de muestreo (USM's) se seleccionan manzanas o grupos de manzanas las cuales contienen por lo menos un segmento; la selección se realiza proporcional a la cantidad de viviendas.

Tercera etapa: se selecciona con muestreo aleatorio simple de las USMs, un segmento conformado por 10 hogares, a modo de Unidades Terciarias de Muestreo (UTM).

Estrato 2.

Primera etapa: se selecciona una UPM por cada estrato con probabilidad proporcional al tamaño de la población y se obtienen 228 UPM's, cada una representativa de su correspondiente estrato. De la muestra maestra, mediante la técnica de selección controlada, se conforman 4 submuestras, cada una con 57 UPM's, de la cual se selecciona una submuestra con muestreo aleatorio simple.

Segunda etapa: se seleccionan manzanas o grupos de manzanas a manera de unidades secundarias de muestreo (USM's), las cuales contienen por lo menos un segmento; la selección se realiza proporcional al número de viviendas.

Tercera etapa: se seleccionan segmentos como Unidades Terciarias de Muestreo (UTM), empleando un muestreo aleatorio simple dentro de cada USM.

1.4.4 Método de estimación

Una vez obtenidos los datos muestrales se procede a realizar la inferencia estadística a cerca de la población objetivo utilizando los factores de expansión de la siguiente forma:

Factor básico de expansión es:

$$F_b = \frac{1}{P}$$

Donde los factores de expansión son función de las probabilidades finales de selección, que aplicados rutinariamente a los datos muestrales corrigen el sesgo que se generaría en el uso no ponderado de la información recolectada.

En cada una de las tres etapas, las unidades de la muestra se seleccionan aleatoriamente con una determinada probabilidad. Inicialmente es necesario calcular las probabilidades de selección de cada unidad de muestreo:

$P(Mun)$: Probabilidad del municipio, es igual a 1 para los municipios de las 13 áreas metropolitanas y para los demás municipios toma la siguiente forma:

$$P(Mun) = \frac{\text{Población de la UPM}}{\text{Población del estrato}}$$

$P(Maz)$: Probabilidad asociada a la manzana:

$$P(maz) = \frac{\text{Total de viviendas de la manzana}}{\text{Total de viviendas del municipio}}$$

$P(\text{Seg})$: Probabilidad de selección del segmento:

$$P(\text{Seg}) = \frac{1}{\text{Total de segmentos en la manzana}}$$

Probabilidad final

$$P_{(mun)} \times P_{(maz)} \times P_{(seg)} = P$$

1.4.4.1 Estimadores

Al remplazar esta fórmula con los valores provenientes de una muestra particular, se obtiene un valor denominado estimación o valor estimado¹.

El estimador de totales

Los estimadores se calculan tanto para cabecera como para resto de la siguiente forma:

El total estimado de las variables y es:

$$\hat{y} = \sum_{h=1}^H \sum_{\alpha=1}^{a_h} y_{h\alpha} FEXP_{h\alpha CAB}$$

donde

$h = 1, \dots, H$ son cada uno de los estratos

$\alpha = 1, \dots, a_h$ son cada uno de los segmentos del estrato h

a_h es la cantidad de segmentos seleccionados en el estrato h

$FEXP_{CAB}$ es el factor final de expansión en cada uno de los segmentos α del estrato h en cabecera

¹ Documento de la metodología muestral de la Encuesta continua de hogares, 2002.

El estimador de razones

Las tasas, razones, proporciones y promedios, generadas a partir de este diseño muestral son de la forma de estimadores de razón, en la cual el numerador y el denominador son variables aleatorias, entonces la razón r está dada por :

$$\hat{r} = \frac{\hat{y}}{\hat{x}} = \frac{\sum_h^H \sum_\alpha^{a_h} y_{h\alpha} FEXP_{h\alpha CAB}}{\sum_h^H \sum_\alpha^{a_h} x_{h\alpha} FEXP_{h\alpha CAB}}$$

Y si $p_{h\alpha}$ es la probabilidad de selección del segmento α en el estrato h , se tiene que:

$$\hat{\text{var}}(\hat{y}_h) = \text{varianza de } \hat{y}_h = \frac{1-p_{h\alpha}}{a_h-1} [a_h \sum_\alpha^{a_h} (y_{h\alpha} FEXP_{h\alpha})^2 - (\sum_h^H \sum_\alpha^{a_h} y_{h\alpha} FEXP_{h\alpha})^2]$$

$$\hat{\text{var}}(\hat{x}_h) = \text{varianza de } \hat{x}_h = \frac{1-p_{h\alpha}}{a_h-1} [a_h \sum_\alpha^{a_h} (x_{h\alpha} FEXP_{h\alpha})^2 - (\sum_h^H \sum_\alpha^{a_h} x_{h\alpha} FEXP_{h\alpha})^2]$$

$$\hat{\text{cov}}(\hat{y}_h, \hat{x}_h) = \text{covarianza de } \hat{y}_h, \hat{x}_h = \frac{1-p_{h\alpha}}{a_h-1} \left[a_h \sum_\alpha^{a_h} (y_{h\alpha} x_{h\alpha} FEXP_{h\alpha}) - \sum_h^H \sum_\alpha^{a_h} x_{h\alpha} y_{h\alpha} FEXP_{h\alpha} \right]$$

El error estándar en cada estrato es:

$$\hat{ES}(\hat{r}_h) = \sqrt{\frac{1}{(\hat{x}_h)^2} [\hat{\text{var}}(\hat{y}_h) + \hat{r}_h^2 \hat{\text{var}}(\hat{x}_h) - 2\hat{r}_h \hat{\text{cov}}(\hat{y}_h, \hat{x}_h)]}$$

Entonces el error estándar de una razón es:

$$\hat{ES}(\hat{r}) = \sqrt{\frac{1}{(\hat{x})^2} [\sum_h^H \hat{\text{var}}(\hat{y}_h) + \hat{r}^2 \sum_h^H \hat{\text{var}}(\hat{x}_h) - 2\hat{r} \sum_h^H \hat{\text{cov}}(\hat{y}_h, \hat{x}_h)]}$$

1.4.4.2 Estimación de errores muestrales

Una manera de medir la calidad de las estimaciones es a través del error de muestreo, el cual indica la magnitud de la variabilidad de las estimaciones. Entre

menor sea la magnitud de esta variabilidad, mejor será la precisión de la estimación del indicador de interés.

La varianza del estimador esta dada en unidades generalmente de difícil manejo y por ello se utiliza una medida relativa con base en valores porcentuales, denominada coeficiente de variación o error relativo del estimador :

$$CV(\hat{y}) = \frac{\sqrt{\hat{V}(\hat{y})}}{\hat{y}} \times 100$$

Criterios para utilizar el coeficiente de variación estimado:

Tabla Nº 2 Criterios para utilizar el coeficiente de variación estimado

Coeficiente de variación	Criterio
<5%	Buena calidad
5 al 15 %	Aceptablemente práctico
15 al 25 %	Baja precisión, debe usarse a discreción
> 25%	No publicable

1.4.4.3 Ajustes por no respuesta y por proyecciones de población

Ajuste por no respuesta en la muestra²

Dado que la tasa de no respuesta de la muestra varía en las diferentes unidades muestrales, es conveniente corregir el Factor Básico de Expansión por factores que ajusten la no respuesta así:

- a) Factor de ajuste de la no respuesta de segmentos

$$Ac_1 = \frac{\text{No.de segmentos seleccionados}}{\text{No.de segmentos encuestados}}$$

- b) Factor de ajuste de la no respuesta de hogares

² Metodología de la Encuesta nacional de calidad de vida, 1997.

$$A_{c_2} = \frac{\text{No.de hogaresseleccionados en un determinado segmento}}{\text{No.de hogaresencuestados en el mismosegmento}}$$

c) Factor de expansión ajustado por no respuesta

$$F_{bc} = F_b \times A_{c_1} \times A_{c_2}$$

Ajuste por las proyecciones de población a la fecha de la encuesta

Este factor aplicado a nivel de región o subregión o por zona, corrige la estructura de la población expandida a partir de la muestra con base en el Factor básico ajustado por no respuesta, e iguala la población total expandida a la proyectada con base en los datos del último censo.

a) Factor de ajuste

$$A_p = \frac{\text{Pob. proyectada a la fecha de la encuesta en una zona}}{\text{Pob.exp. a partir de la muestra con base en los fac. ajustados en la misma zona}}$$

b) Factor final de expansión

$$FEXP = F_{bc} \times A_p$$

4. BIBLIOGRAFIA

- DANE, *Guía para Documentar la Actividad Estadística*, 1998
- Kish, Leslie, *Survey Sampling*, 1976
- Cochran, William, *Técnicas de Muestreo*, 1985
- DANE, *Metodología de la Encuesta Nacional de Calidad de Vida 1.997*.
- SÄRNDAL, Carl., Swensson, Bengt. *Model Assisted Survey Sampling*. New York: Springer-Verlag. 1992.