

**Departamento Administrativo
Nacional de Estadística**



**Encuesta de Calidad de Vida para Bogotá
Metodología Muestral**

EQUIPO DE DISEÑOS MUESTRALES

Septiembre 2003

1. DISEÑO ESTADÍSTICO

1.2 PREMISAS DEL DISEÑO

1.2.1 Universo

El universo está constituido por la población civil en hogares no institucionales, residentes en las viviendas particulares de Bogotá.

1.2.2 Cobertura

La ECV cubre 19 localidades de Bogotá D.C., las cuales son: Usaquén, Chapinero, Santafé, San Cristóbal, Usme, Tunjuelito, Bosa, Kennedy, Fontibón, Engativa, Suba, Barrios Unidos, Teusaquillo, Mártires, Antonio Nariño, Puente Aranda, Candelaria, Rafael Uribe y Ciudad Bolívar

1.2.3 Unidades Estadísticas

1.2.3.1 Unidad de muestreo

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

Segunda etapa: Como unidades secundarias de muestreo (USM's) se seleccionan segmentos

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 6.5%

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 a nivel de personas, indagan 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, bietápico 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 las localidades

- **Bietápico**

Es bietápico porque la selección de las unidades de muestreo se lleva a cabo en dos 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 Nº 1. Tamaño de muestra

Localidad	Proyección de población al año 2002	Tamaño de muestra en segmentos
Usaquén	451.606	70
Chapinero	166.883	69
Santafé	136.097	69
San Cristobal	386.313	70
Úsme	227.663	70
Tunjuelito	196.781	70
Bosa	259.454	70
Kennedy	649.816	70
Fontibón	457.613	70
Engativá	850.529	70
Suba	574.760	70
Barrios Unidos	228.065	70

Localidad	Proyección de población al año 2002	Tamaño de muestra en segmentos
Teusaquillo	153.883	69
Mártires	119.667	69
Antonio Nariño	147.638	69
Puente Aranda	360.908	70
Candelaria	27.011	65
Rafael Uribe	304.112	70
Ciudad Bolívar	474.297	70
TOTAL		13.200

1.4.3 Método de selección

Primera etapa: Para cada una de las localidades 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.

Segunda etapa: se selecciona con muestreo aleatorio simple de las UPMs, un segmento de 10 hogares, a modo de Unidades Secundarias de Muestreo (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 sería:

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

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.

Cada una de las dos etapas las unidades de la muestra se seleccionaron aleatoriamente con una determinada probabilidad. Inicialmente es necesario calcular las probabilidades de selección de cada unidad de muestreo:

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

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

$P(Seg)$: Probabilidad de selección del segmento:

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

Probabilidad final

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

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 de la siguiente forma:

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

El total estimado de las variables y es:

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

$h = 1, \dots, H$ 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

$FEXP$ es el factor final de expansión en cada uno de los segmentos α del estrato h

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 :

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

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

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

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

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

El error estándar en cada estrato es:

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

Entonces el error estándar de una razón es:

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

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 :

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

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

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

$$A_{c_1} = \frac{\text{No.de segmentos seleccionados en la localidad}}{\text{No.de segmentos encuestados en la misma localidad}}$$

b) Factor de ajuste de la no respuesta de hogares

$$A_{c_2} = \frac{\text{No.de hogares seleccionados en un determinado seg}}{\text{No.de hogares encuestados en el mismo seg}}$$

c) Factor de expansión ajustado por no respuesta

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

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

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

Este factor aplicado a nivel de localidad 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

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

b) Factor final de expansión

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

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.