

## **Anexo A. Descripción técnica de la muestra**

### **Diseño de la muestra<sup>1</sup>**

En Colombia, la muestra fue diseñada para que incluyera a todos los adultos no institucionalizados; es decir que excluye a las personas que habitan en cárceles, colegios, hospitales y bases militares. Es una muestra aleatoria estratificada. La estratificación asegura la inclusión de las áreas geográficas más importantes del país: Pacífica, Atlántica, Central, Oriental, los Antiguos Territorios Nacionales (ATR) y Bogotá. La muestra fue subestratificada para incluir a las ciudades con más y con menos de 300.000 habitantes. Finalmente, la muestra fue subestratificada además en áreas urbanas y rurales.

Usamos proyecciones al 2006 del Censo de Población de 1993, el más reciente en Colombia<sup>2</sup>. De acuerdo con el Censo, 21.58% de la población habita en la región Atlántica, 17.57% en la Pacífica, 23.63% en la región Central, 18.31% en la región Oriental, 3.28% en los ATR y 15% habita en Bogotá.

La selección de la muestra fue multietápica. La primera etapa fue el municipio, luego el sector censal, luego la sección y finalmente la manzana, la vivienda y el hogar. Utilizamos un sistema de cuotas por sexo y edad para seleccionar al encuestado al interior de cada hogar.

Se entrevistaron 1491 informantes. Técnicamente, nuestro margen de error fue  $\pm 2.54\%$ . Esto quiere decir que si hiciéramos múltiples muestras en Colombia, 95% de ellas reflejarían las opiniones de la población con una precisión no inferior a  $\pm 2.54\%$ . Sin embargo, nuestra muestra fue estratificada y por conglomerados. Esto significa que, si bien la estratificación aumentó la precisión de la muestra, de alguna manera los conglomerados, que se hicieron para controlar costos del trabajo de campo, contribuyeron a reducir la precisión. Por supuesto, otros factores distintos al muestreo pueden reducir la precisión de la muestra, tales como la tasa de no respuesta, los errores al seleccionar al encuestado, la comprensión errónea de las preguntas, entre otros. Pero en términos de la técnica de muestreo, un intervalo de confianza de  $\pm 2.54\%$  es muy bueno.

La Tabla A-1 es un resumen de los errores estándar y de los efectos de diseño para algunas variables e índices de la encuesta. Los efectos de diseño (def) indican la eficiencia del diseño por conglomerados en comparación con el diseño aleatorio simple. Un def de 1 indica que las varianzas obtenidas en ambos diseños son las mismas, es decir que el diseño por conglomerados fue tan eficiente como un diseño aleatorio simple. Si el def es mayor que 1, significa que el diseño por conglomerados tuvo una varianza mayor que la producida por un diseño aleatorio simple. Y si el def es menor que 1, significa que la varianza del diseño por conglomerados es aun más pequeña que la producida por el diseño aleatorio.

---

<sup>1</sup> Esta sección y la siguiente fueron adaptadas del informe Costa Rica 2004, "Cultura democrática, seguridad ciudadana y capital social en Costa Rica", cuyos autores son Luis Rosero-Bixby y Jorge Vargas-Cullell.

<sup>2</sup> Los resultados del Censo adelantado por el gobierno desde 2005 no estaban disponibles al momento de diseñar la muestra utilizada en este estudio.

**Tabla A-1. Errores estándar y efectos de diseño para alguna variables**

<b>Variable</b>	<b>Media</b>	<b>Error est.</b>	<b>Efecto</b>
q2 (edad)	35.934	.315	.560
ed (nivel educativo)	9.367	.191	2.336
Wealth (índice de riqueza)	3.835	.097	3.767
PSA5 (respaldo al sistema)	57.211	.755	1.532
tol (tolerancia política)	53.226	1.024	1.614
soct1r (evaluación de la economía del país)	39.266	.628	1.242
l1 (autoubicación ideológica)	6.244	.082	1.246
m1r (evaluación al gobierno)	63.998	.799	1.478
exc7r (percepción de corrupción)	73.423	1.008	1.359

estrato=región (estratopri)

upm=sector (colsect)

De acuerdo con la tabla anterior, el diseño por conglomerados para esta encuesta fue muy eficiente. En efecto, con la excepción de edad, educación y riqueza, todos los def estuvieron cerca de 1. Los errores estándar para la mayoría de variables también fueron muy moderados.

## **Resultados de la muestra y descripción de los encuestados**

El diseño probabilístico de la muestra, así como la disponibilidad de un buen marco muestral, son condiciones suficientes para esperar que el grupo de entrevistados sea representativo de la población colombiana. Sin embargo, debido al efecto de errores aleatorios y por las distorsiones inevitables del diseño muestral, la muestra puede desviarse en algunas características de la población que representa. Puede incluir sesgos que deben ser especificados. La Tabla A-2 permite responder a la pregunta: ¿qué tan representativa es la muestra de la población? En ella comparamos algunas características de la muestra con las del Censo de 1993<sup>3</sup>.

---

<sup>3</sup> No hay proyecciones recientes para la mayoría de indicadores que aquí se analizan.

**Tabla A-2. Muestra vs. Censo poblacional de 1993 (mayores de 18 años)**

<b>Características</b>	<b>Censo de 1993</b>	<b>Encuesta en Colombia</b>
(N)	26,735,000	1,491
% de hombres	49	50
% > 30 años	62	60
% soltero	40	32
% casado o en unión libre	50	59
% con primaria	44	34
% con secundaria	34	46
% con educación post-secundaria	11	20
% en Región Atlántica	22	22
% en Bogotá	15	15
% en Región Central	24	24
% en Región Oriental	18	18
% en Región Pacífica	18	17
% en Antiguos Territorios Nacionales	3	4

Se observa que hay congruencia entre la muestra de esta encuesta y el Censo de 1993. Algunas características como edad, sexo y habitantes de las diferentes regiones son prácticamente idénticas. Hay una pequeña desviación en los porcentajes de casados y solteros. Y, finalmente, hay una brecha en las tres variables de educación, donde la más amplia es la del porcentaje de personas con educación secundaria, que corresponde a 34% en el Censo de 1993 y pasa a 46% en la Encuesta de 2006. La explicación para las brechas de casados y educación puede ser que el Censo tiene 13 años de haberse realizado. Éste es un período lo suficientemente largo para que se hayan producido grandes cambios en estas variables. Desafortunadamente, a la fecha no existen proyecciones confiables del Censo 1993 en educación y por rangos de edad ni se dispone de las cifras del censo 2005. Existen algunas proyecciones que incluyen estudiantes actuales, lo cual hace que no sea una estadística comparable con las variables de esta encuesta.

Pero dado que, en general, la muestra es representativa de la población, no se hace necesario el uso de ponderadores. Por lo tanto, la muestra de este estudio es auto-ponderada.

La Tabla A-3 compara las características de la muestra entre hombre y mujeres.

**Tabla A-3. Características de la muestra, por sexo**

<b>Característica</b>	<b>Total</b>	<b>Hombres</b>	<b>Mujeres</b>
(N)	1,491 (100%)	50%	50%
Edad promedio	37.33	37.55	37.11
% casado o en unión libre	58.82	55.90	61.74
Ingresos familiares (Moda) % entre 361,000 y 420,000 pesos	26.95	26.48	27.43

Usamos una cuota por sexo y edad para seleccionar al encuestado. Por lo tanto, nuestros porcentajes de hombres y mujeres son muy similares entre sí. Sus edades también prácticamente iguales. Hay una pequeña diferencia con respecto al porcentaje de casados o en unión libre, donde las mujeres muestran un porcentaje un poco más alto (61.74%) que los hombres en este estado. Con respecto al ingreso familiar, hay un mayor porcentaje de mujeres en el rango entre 361.000 y 420.000 pesos, que corresponde a la moda de ingresos familiares, aunque la brecha es muy reducida.

## **Descripción técnica del diseño muestral**

### ***Universo***

El universo de la encuesta tiene una cobertura nacional de todos los adultos que habitan en las seis regiones del país: Bogotá, las regiones Atlántica, Pacífica, Central, Oriental y los Antiguos Territorios Nacionales. El universo también comprendió a todos los adultos que habitan en las áreas urbanas y rurales.

El universo se dividió en dos sectores: uno con las ciudades con más de 300.000 habitantes y otro con las ciudades con menos de 300.000 habitantes.

### ***Población***

La muestra se circunscribió a todos los adultos no institucionalizados; es decir que excluye a las personas que habitan en cárceles, colegios, hospitales y bases militares. Los hogares privados en estas áreas sí fueron contemplados.

### ***Unidad de selección final***

Dado que el cuestionario incluye preguntas no sólo referentes al encuestado sino también a otros miembros del hogar, la unidad de observación estadística seleccionada fue el hogar. El encuestado sólo podía habitar en un hogar.

Dado que cada hogar pertenece a una vivienda, a veces compartida con otros hogares relativamente estables en el tiempo, cada vivienda fue seleccionada como la unidad de selección final.

## ***Método de muestreo***

Escogimos el muestreo probabilístico, estratificado, multietápico, con selección aleatoria de las unidades en cada etapa. Primero, la muestra fue estratificada por tamaño de municipio (ciudades con más y menos de 300.000 habitantes), luego por regiones y luego por áreas (rural y urbana).

Es un muestreo multietápico puesto que al interior de las áreas urbanas se empezó con unidades primarias de muestreo (sectores), luego con unidades secundarias (secciones), luego con unidades terciarias (manzanas) y con unidades finales de muestreo (conglomerados de viviendas) de 6 a 8 en áreas urbanas y de 10 a 12 en áreas rurales. En cada vivienda el encuestador seleccionó un solo hogar como unidad de observación.

El informante fue seleccionado de acuerdo con cuotas de edad y sexo. En efecto, en cada manzana el encuestador tenía que incluir al menos un hombre y una mujer de los siguientes rangos de edad:

18 a 27 años  
28 a 40 años  
Más de 40 años

A cada encuestador se le asignó una manzana específica. Una vez en la manzana, el encuestador hacía un recuento de las primeras 20 viviendas que encontraba. Tenía la instrucción de aplicar un mínimo de 8 encuestas de las 20 viviendas listadas, procurando un balance de las cuotas de sexo y edad.

El método de selección se eligió de acuerdo con las siguientes consideraciones:

Se requerían muestras representativas en los siguientes niveles:

- Nacional
- Estratos de primera etapa:
  - o Ciudades con más de 300.000 habitantes
  - o Ciudades con menos de 300.000 habitantes
- Estratos de segunda etapa:
  - o Bogotá
  - o Región Atlántica
  - o Región Pacífica
  - o Región Oriental
  - o Región Central
  - o Antiguos Territorios Nacionales
- Estratos de tercera etapa:

- Área urbana
  - Área rural
- Dominios:
- Ciudades con más de 300.000 habitantes (de inclusión forzosa)
  - Ciudades con menos de 300.000 habitantes
- a) Para cada etapa, se calcularon márgenes de error que correspondieran a los estándares mínimos de calidad.
- b) Se buscó facilitar la operatividad del trabajo de campo
- c) Se trabajó con el mejor y más actualizado marco muestral disponible para cada municipio (censo poblacional, cartografía, listados de viviendas actualizados, entre otros)

### ***Marco muestral***

El marco muestral está constituido por el inventario actualizado de cartografía y viviendas obtenidos a partir del Censo de 1993. El Centro Nacional de Consultoría obtuvo las versiones 2003 del Departamento Nacional de Estadística DANE.

### ***Cálculos por estratos***

La muestra está constituida por 193 puntos muestrales: 160 urbanos y 33 rurales, distribuidos en 53 municipios de 26 de los 32 departamentos de Colombia.

### ***Tamaños de muestra, intervalos de confianza y márgenes de error***

El intervalo de confianza anticipado para la muestra nacional fue de 95% con un margen de error de 2.6%, asumiendo una proporción 50/50 en variables dicotómicas.

Los márgenes de error para un intervalo de 95% de confianza se indican en la Tabla A-4:

**Tabla A-4. Tamaño de la muestra y márgenes de error (nivel de confianza del 95%)**

<b>Estratos Regiones</b>	<b>Tamaño de muestra</b>	<b>Margen de error M.A.S.</b>	<b>% M.P.C.</b>
Atlántica	323	5.45	5.58
Bogotá	231	6.45	6.59
Central	358	5.18	5.29
Oriental	274	5.92	6.05
Pacífica	251	6.19	6.32
Antiguos Territorios Nacionales	54	13.34	13.63
<b>Áreas</b>			
Urbana	1107	2.96	3.01
Rural	384	5.00	5.11
<b>Total País</b>	<b>1491</b>	<b>2.54</b>	<b>2.6</b>

***Personal que trabajó en esta encuesta***

El CNC involucró a sus cinco sedes (Bogotá, Cali, Medellín, Barranquilla y Bucaramanga) para asegurar una calidad óptima en el mínimo tiempo posible. Debido a la actual situación de orden público del país, se nos aconsejó permanecer el mínimo tiempo posible en la mayoría de áreas visitadas, lo cual complicó las operaciones.

Debido a la complejidad del formulario, utilizamos nuestros encuestadores más experimentados, muchos de ellos con más de 15 años de experiencia.

El CNC involucró a un total de 101 miembros en el equipo, distribuidos así:

**Tabla A-5. Personal del CNC que participó en el estudio**

<b>Actividad</b>	<b>Total de personal</b>
Coordinadores de campo	5
Supervisores	15
Encuestadores	40
Supervisores de calidad en campo	10
Codificadores	7
Digitadores	7
Verificadores de datos	7
Subtotal campo y digitación	91
Personal directivo y profesional	5
Personal administrativo	3
<b>Total personal del equipo</b>	<b>99</b>

**Tabla A-6. Población por región: áreas urbana y rural (proyección a 2006)**

Región	Urbana	Rural	Total
Atlántica	6,912,796 (70.9%)	2,843,894 (29.1%)	9,756,690
Bogotá	7,014,111 (100.0%)	1,517 (0.0%)	7,015,628
Central	7,659,319 (71.2%)	3,093,365 (28.8%)	10,752,684
Oriental	5,207,407 (62.6%)	3,106,566 (37.4%)	8,313,973
Pacífica	5,322,949 (66.8%)	2,644,004 (33.2%)	7,966,953
Antiguos Territorios Nacionales	583,895 (39.6%)	890,830 (60.4%)	1,474,725
<b>Total</b>	<b>32,700,477 (72.2%)</b>	<b>12,580,176 (27.8%)</b>	<b>45,280,653</b>

**Tabla A-7. Tamaño y distribución de la muestra por estrato (región)**

Región	Urbana	Rural	Total
<b>Atlántica</b>			
> 300,000 habitantes	127		127
< 300,000 habitantes	112	84	196
Total Atlantica	239	84	323
<b>Bogotá</b>			
> 300,000 habitantes	231		231
Total Bogota	231		231
<b>Central</b>			
> 300,000 habitantes	150		150
< 300,000 habitantes	112	96	208
Total Central	262	96	358
<b>Oriental</b>			
> 300,000 habitantes	66		66
< 300,000 habitantes	112	96	208
Total Oriental	178	96	274
<b>Pacífica</b>			
> 300,000 habitantes	91		91
< 300,000 habitantes	88	84	160
Total Pacífica	179	72	251
<b>Antiguos Territorios Nacionales</b>			
< 300,000 habitantes	18	36	54
Total Antiguos Territorios Nacionales	18	36	54
<b>Total</b>	<b>1,107</b>	<b>384</b>	<b>1,491</b>

**Tabla A-8. Encuestados y unidades primarias de muestreo (UPM) por estrato (región)**

Región	Muestra	UPM
Atlántica	323	32
Bogotá	231	34
Central	358	39
Oriental	274	26
Pacífica	251	26
Antiguos Territorios Nacionales	54	6
<b>Total</b>	<b>1,491</b>	<b>163</b>



## **Comentarios finales sobre el trabajo de campo**

**Sobre el cuestionario:** el cuestionario fue más largo que las anteriores versiones de LAPOP, 55 minutos en promedio, de acuerdo con la opinión de los entrevistadores, pero en general los informantes estuvieron dispuestos a contestar las encuestas y tuvimos muy pocas entrevistas interrumpidas.

**Acerca del trabajo de campo:** Por razones de seguridad no fue posible la realización de las encuestas rurales en uno de los municipios . Además, dado que tanto en 2004 como en 2005 no fue posible hacer encuestas en Chalan (Sucre), este municipio fue reemplazado en la muestra de 2006.

Aparte de estos incidentes, a pesar de que algunos informantes estaban ubicados en áreas con fuerte presencia de grupos armados ilegales, no hubo reportes de ningún tipo de presión sobre los informantes para inducir las respuestas. Por el contrario, los encuestadores enfatizaron la libertad de opinión de quienes aceptaron ser parte de este estudio.

Como todos los años, el CNC quisiera extender su gratitud a cada uno de los miembros del equipo de este estudio, especialmente a los valientes hombres y mujeres que desafiaron las normas de seguridad y asumieron muchos riesgos para lograr un trabajo muy bien hecho.