

Apéndice A: Descripción Técnica de la Muestra

Diseño de la muestra¹

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 a 2004 del Censo de Población de 1993, el más reciente en Colombia. De acuerdo con el Censo, 22% de la población habita en la región Atlántica, 18% en la Pacífica, 24% en la región Central, 18% en la región Oriental, 3% 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 1,479 informantes. Técnicamente, nuestro margen de error fue $\pm 2.6\%$. 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 $\pm 2.6\%$. 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 $\pm 2.6\%$ 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.

¹ 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.

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

	Totales			
	N	X	EE	DEF
q2 (edad)	1,479	36.8	.22	.38
ed (educación, total años)	1,478	9.4	.21	3
wealth	1,476	4.6	.128	5.3
psa5	1,456	57.3	.66	1.48
tol	1,451	50.3	1.14	2.24
exctot	1,439	.15	.014	1.3
vb2r	1,478	.67	.013	1.20
It1r	1,455	64.1	1.08	1.9
vic1r	1,479	.144	.01	1.3
vicconf	1,479	16.14	.80	1.4
m1r	1,470	70.3	.61	1.27
ls3r	1,475	81.7	.74	1.36
ing4r	1,402	72.6	.83	1.21
coups	1,422	37.1	1.3	1.8

De acuerdo con la tabla anterior, el diseño por conglomerados para esta encuesta fue muy eficiente. En efecto, con la excepción de educación, riqueza y tolerancia, todos los DEF estuvieron cerca de 1. Los errores estándar para la mayoría de variables también fueron muy moderados. La Tabla A.2 muestra los errores estándar y DEF para la variable q2 (edad) por conglomerado (región).

La Tabla A.3, hace lo mismo para el índice de tolerancia. Los DEF, así como también los errores estándar indican que el diseño por conglomerados para las regiones fue más eficiente que un diseño aleatorio simple.

Tabla A.2 Errores estándar y efectos de diseño para edad, por región

Q2	N	X	EE	DEF
Atlántica	1,479	36.9	.46	.31
Bogotá	1,479	36.3	.81	.83
Central	1,479	37.3	.40	.27
Oriental	1,479	36.7	.48	.33
Pacífica	1,479	36.6	.46	.29
Antiguos Territorios Nacionales	1,479	36.5	1.09	.46

Tabla A.3 errores estándar y efectos de diseño para tolerancia, por región

Tol	N	X	EE	DEF
Atlántica	1,451	48.3	1.97	1.4
Bogotá	1,451	50.8	1.8	.98
Central	1,451	45.2	3.3	4.18
Oriental	1,451	51.7	2.4	1.9
Pacífica	1,451	57.7	2.4	1.8
Antiguos Territorios Nacionales	1,451	52.8	3.1	.55

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.4** permite responder a la pregunta: ¿qué tan representativa es la muestra de la población? Abajo, comparamos algunas características de la muestra con las del Censo de 1993².

Tabla A.4 Muestra vs. Censo poblacional de 1993 (mayores de 18 años)

Características	Censo de 1993	Encuesta en Colombia
(N)	26,735,000	1,479
% de hombres	49	49
% > 30 años	42	43
% soltero	40	34
% casado o en unión libre	50	56
% con primaria	44	30
% con secundaria	34	52
% con educación post-secundaria	11	17
% en Región Atlántica	22	21
% en Bogotá	15	16
% en Región Central	24	24
% en Región Oriental	18	19
% 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 post-secundaria, que corresponde a 34% en el Censo de 1993 y pasa a 52% en la Encuesta de 2004. La explicación para las brechas de casados y educación puede ser que el Censo tiene 11 años de haberse realizado. Éste es un período lo suficientemente largo para que se hayan producido grandes cambios en estas variables. Desafortunadamente, no existen proyecciones confiables del Censo 1993 para las variables de educación, discriminando a la población mayor de edad. Existen algunas proyecciones que incluyen estudiantes actuales, lo cual hace que no sea una estadística comparable con las variables de esta encuesta.

² No hay proyecciones recientes para la mayoría de indicadores que aquí se analizan.

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 autoponderada.

La Tabla A.5 compara las características de la muestra entre hombre y mujeres.

Tabla A.5 Características de la Muestra, por sexo

Característica	Total	Hombres	Mujeres
(N)	1,479 (100%)	50.1%	49.9%
Edad promedio	36.8	37	36
% casado o en unión libre	55.9	56.7	55.1
Ingresos familiares (Moda) % entre 181,000 y 360,000 pesos	26.3	25	27.6

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 son muy similares entre sí, con una diferencia de sólo un año. Hay una pequeña diferencia con respecto al porcentaje de casados o en unión libre, donde los hombres tienen un porcentaje un poco más alto (56.7%) que las mujeres en este estado. Con respecto al ingreso familiar, hay un mayor porcentaje de mujeres en el rango entre 181.000 y 360.000 pesos, que corresponde a la moda de ingresos familiares, aunque la brecha es muy reducida (2.6 puntos porcentuales).

El Centro Nacional de Consultoría (CNC), una empresa con más de 20 años en la investigación de mercados y en la realización de complejos estudios sociales y políticos, fue seleccionada para desarrollar el proceso de muestreo, el trabajo de campo, así como la digitación, verificación y procesamiento de los datos.

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:

Ciudades con más de 300.000 habitantes
Ciudades con menos de 300.000 habitantes

Estratos de segunda etapa:

Bogotá
Región Atlántica
Región Pacífica
Región Oriental
Región Central
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.

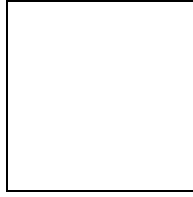
Cálculos por estratos

La muestra está constituida por 138 puntos muestrales: 103 urbanos y 35 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 fueron:



Ajuste por no cobertura

Para asegurar la eficiencia, suficiencia y cobertura de la muestra, adoptamos un sistema de ajuste por no cobertura que garantiza un mínimo nivel de error y un máximo nivel de confianza. Además, el sistema elimina sesgos que resultan de reemplazar las viviendas.

El sistema aplica al tamaño de muestra (n) de cada estrato un factor de no cobertura definido por:

$$N^* = (1+t) n$$

En donde:

T = tasa de no cobertura.

N^* = tamaño de muestra final

Escogimos un promedio t de 0.18.

En consecuencia,

$$N^* = (1+0.18) 1,500 = 1,770 \text{ mayores de edad.}$$

Al final del trabajo de campo, se logró un tamaño de muestra de 1.479 entrevistados, 21 menos de los esperados debido a problemas de orden público (ver “Comentarios finales sobre el trabajo de campo”).

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:

Actividad	Total de personal
Coordinadores de campo	15
Supervisores	9
Encuestadores	35
Supervisores de calidad en campo	10
Codificadores	8
Digitadores	8
Verificadores de datos	8
Subtotal campo y digitación	93
Personal directivo y profesional	5
Personal administrativo	3
Total personal del equipo	101

Figura A.1 Universo, población total por región y áreas urbanas y rurales

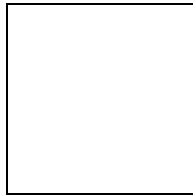


Figura A.2 Tamaño y distribución de la muestra por estrato

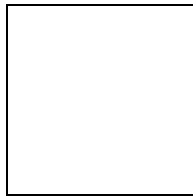
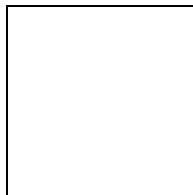
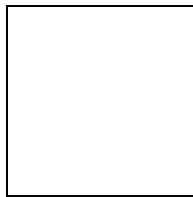
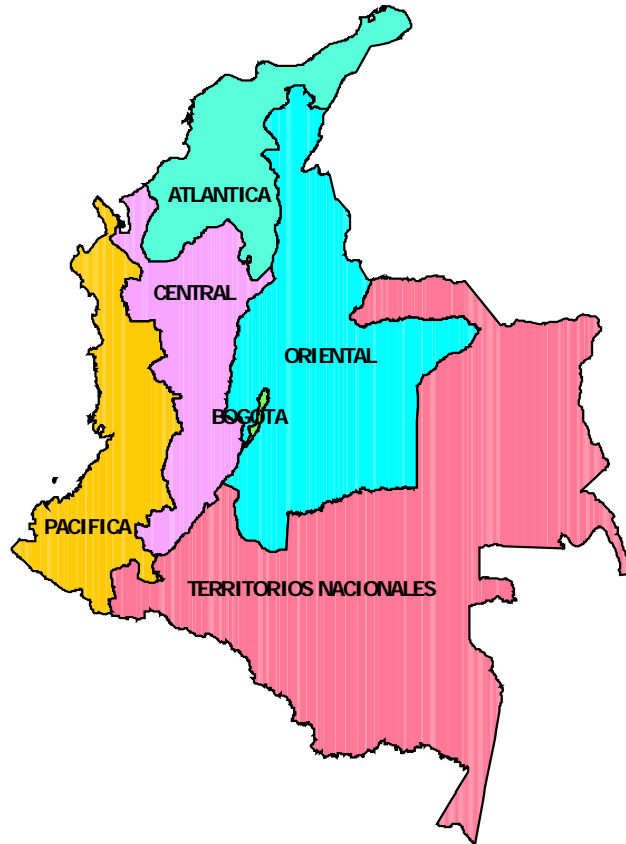


Figura A.3 Selección de la muestra por estrato, áreas y municipios



Ubicación geográfica de la muestra



Comentarios finales sobre el trabajo de campo

Sobre el cuestionario: el cuestionario fue largo, 45 minutos en promedio, pero en general los informantes estuvieron dispuestos a contestar las encuestas y tuvimos muy pocas entrevistas interrumpidas.

Debido al conflicto interno colombiano, algunas preguntas eran especialmente delicadas. Algunos encuestadores mencionaron que las series de la página 8 del cuestionario generaron cierta incomodidad entre algunos informantes. Pero para nuestra sorpresa, tuvimos muy poca resistencia a contestarlas.

Acerca del trabajo de campo: por solicitud de las autoridades locales, no se logró aplicar la encuesta en las zonas rurales de Cumbitara, Nariño y de Chalán, Sucre, debido a la frágil seguridad de estos municipios en los días del trabajo de campo.

En el resto del país, algunos de nuestros encuestadores fueron interceptados e interrogados por los grupos armados ilegales, quienes finalmente permitieron la recolección de la información.

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.

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.