

Anexo B. Diseño muestral

I. Diseño de la muestra¹

La muestra diseñada para Colombia incluye 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 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 2012 del Censo de Población de 2005, el más reciente en Colombia. De acuerdo con el Censo, 20% de la población habita en la región Atlántica, 17% en la Pacífica, 25% en la región Central, 18% en la región Oriental, 3% en los Antiguos Territorios Nacionales y 17% 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 1.512 personas. Técnicamente, nuestro margen de error fue $\pm 2.5\%$. 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 $\pm 2.5\%$. 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 $\pm 2.5\%$ es muy bueno.

La Tabla 10 es un resumen de los errores estándar y de los efectos de diseño para las variables edad y la calificación de la situación económica del país. 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 10. Errores estándar y efectos de diseño para algunas variables

Variable	Estimación	Error típico	Intervalo de confianza al 95%		Coeficiente de variación	Efecto de diseño	Raíz cuadrada del efecto de diseño	Acumulado
			Inferior	Superior				
Edad (q2)	36.86	.181	36.51	37.22	.005	.221	.470	
Evaluación de la economía (soct1)								
Muy buena	1.3%	.3%	.8%	2.1%	.241	1.172	1.083	1.3%
Buena	14.1%	.9%	12.4%	16.1%	.065	1.049	1.024	15.5%
Ni buena ni mala	53.3%	1.3%	50.7%	55.9%	.024	1.022	1.011	68.8%
Mala	24.0%	1.1%	21.8%	26.4%	.048	1.083	1.041	92.8%
Muy mala	7.2%	.7%	5.9%	8.8%	.100	1.164	1.079	100.0%

De acuerdo con la tabla anterior, el diseño por conglomerados para esta encuesta fue eficiente el DEF fue inferior a 1, para el caso de cómo calificaría le situación económica del país se presentaron DEF más altos. Los errores estándar también fueron muy moderados. La Tabla 11 muestra los errores estándar y DEF para la variable q2 (edad) por conglomerado (región). 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 11. Errores estándar y efectos de diseño para edad, por región

Región (estratopri)	Estimación	Error típico	Intervalo de confianza al 95%		Coeficiente de variación	Efecto de diseño	Raíz cuadrada del efecto de diseño
			Inferior	Superior			
Caribe	36.69	.353	35.99	37.38	.010	.183	.428
Bogotá	36.98	.417	36.16	37.80	.011	.216	.465
Central	37.62	.515	36.60	38.63	.014	.351	.593
Oriental	36.91	.382	36.16	37.67	.010	.179	.423
Pacífica	36.03	.376	35.29	36.77	.010	.180	.424
ATN	36.79	.565	35.68	37.90	.015	.072	.269

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 12 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 2005.

Tabla 12. Muestra vs. Censo poblacional de 2005 (mayores de 18 años)

Características	Censo de 2005	Encuesta en Colombia
N	27,184,228	1,512
% de hombres	48	50
% > 30 años	68	55
% soltero	30	35
% casado o en unión libre	56	56
% con primaria	38	28
% con secundaria	37	55
% con educación post-secundaria	25	25
% en Región Atlántica	21	20
% en Bogotá	16	17
% en Región Central	25	25
% en Región Oriental	18	18
% en Región Pacífica	17	17
% en Antiguos Territorios Nacionales	3	3

Se observa que hay congruencia entre la muestra de esta encuesta y el Censo de 2005. 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 dos variables de educación, donde la más amplia es la del porcentaje de personas con educación secundaria, que corresponde a 37% en el Censo de 2005 y pasa a 55% en la Encuesta de 2012, en primaria en el censo 2005 38% y en el 2012 28%.

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 13 compara las características de la muestra entre hombre y mujeres.

Tabla 13. Características de la muestra por sexo

	Género del entrevistado establecido en la cuota			
	Hombre		Mujer	
	Media	%	Media	%
N - %		50		50
Años	36,97		36,74	
¿Cuál es su estado civil?	Casado o Unión libre (acompañado)			52

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. Hay una pequeña diferencia con respecto al porcentaje de casados o en unión libre, donde las mujeres tienen un porcentaje un poco más alto (52%) que los hombres en este estado.

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

Puesto 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:
 - o Área urbana
 - o Área rural
- Dominios:
 - o Ciudades con más de 300.000 habitantes (de inclusión forzosa)
 - o Ciudades con menos de 300.000 habitantes

Se satisficieron, además, los siguientes criterios:

- Para cada etapa, se calcularon márgenes de error que correspondieran a los estándares mínimos de calidad.
- Se buscó facilitar la operatividad del trabajo de campo
- 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 2005. El Centro Nacional de Consultoría obtuvo las versiones 2005 del Departamento Nacional de Estadística DANE.

Cálculos por estratos

La muestra está constituida por 252 puntos muestrales: 198 urbanos y 54 rurales, distribuidos en 47 municipios de 22 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 2.5%, asumiendo una proporción 50/50 en variables dicotómicas.

Los márgenes de error para un intervalo de 95% de confianza aparecen en la Tabla 14:

Tabla 14. Tamaño de muestra y márgenes de error - Nivel de confianza del 95%

Estratos Regiones	Tamaño de muestra	Margen de error M.A.S.	% M.P.C.
Atlántica	288	5,77	6,33
Bogotá	264	6,03	6,61
Central	360	5,17	5,66
Oriental	288	5,77	6,33
Pacífica	264	6,03	6,61
Territorios nacionales	48	14,15	15,50
Áreas			
Urbana	1.188	2,84	3,11
Rural	324	5,44	5,96
Total país	1.512	2,52	2,76

II. Otros aspectos relacionados con la muestra y 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 113 miembros en el equipo, distribuidos como se ve en la Tabla 15:

Tabla 15. Personal involucrado en el proyecto

Actividad	Total de personal
Coordinadores de campo	5
Supervisores	15
Encuestadores	60
Supervisores de calidad en campo	10
Codificadores	3
Digitadores	3
Subtotal campo y digitación	96
Personal directivo y profesional	4
Personal administrativo	3
Total personal del equipo	103

Información adicional sobre la muestra

Tabla 16. Universo, población total por región y áreas urbanas y rurales

Colombia: población total, proyección año 2011

Región	Urbano	Rural	Total
ATLANTICA	7.373.200	2.649.872	10.023.072
BOGOTA	7.555.165	16.180	7.571.345
CENTRAL	8.626.045	3.037.225	11.663.270
ORIENTAL	6.038.624	2.639.940	8.678.564
PACIFICA	5.481.003	2.502.354	7.983.357
TERRITORIOS NACIONALES	303.101	359.114	662.215
Total general	35.377.138	11.204.685	46.581.823

Distribución Porcentual (%)

Región	Urbano	Rural	Total
ATLANTICA	74%	26%	100%
BOGOTA	100%	0%	100%
CENTRAL	74%	26%	100%
ORIENTAL	70%	30%	100%
PACIFICA	69%	31%	100%

TERRITORIOS NACIONALES	46%	54%	100%
Total general	76%	24%	100%

Tabla 17. Tamaño y distribución de la muestra por estrato

	Urbano	Rural	Total
Atlántica	222	66	288
- de 300,000 habitantes	84	66	150
+ de 300,000 habitantes	138	0	138
Bogotá	264	0	264
+ de 300,000 habitantes	264	0	264
Central	288	72	360
- de 300,000 habitantes	120	72	192
+ de 300,000 habitantes	168	0	168
Oriental	198	90	288
- de 300,000 habitantes	126	90	216
+ de 300,000 habitantes	72		72
Pacífica	180	84	264
- de 300,000 habitantes	66	84	150
+ de 300,000 habitantes	114	0	114
Territorios nacionales	36	12	48
- de 300,000 habitantes	36	12	48
Total	1188	324	1512

Ubicación geográfica de la muestra

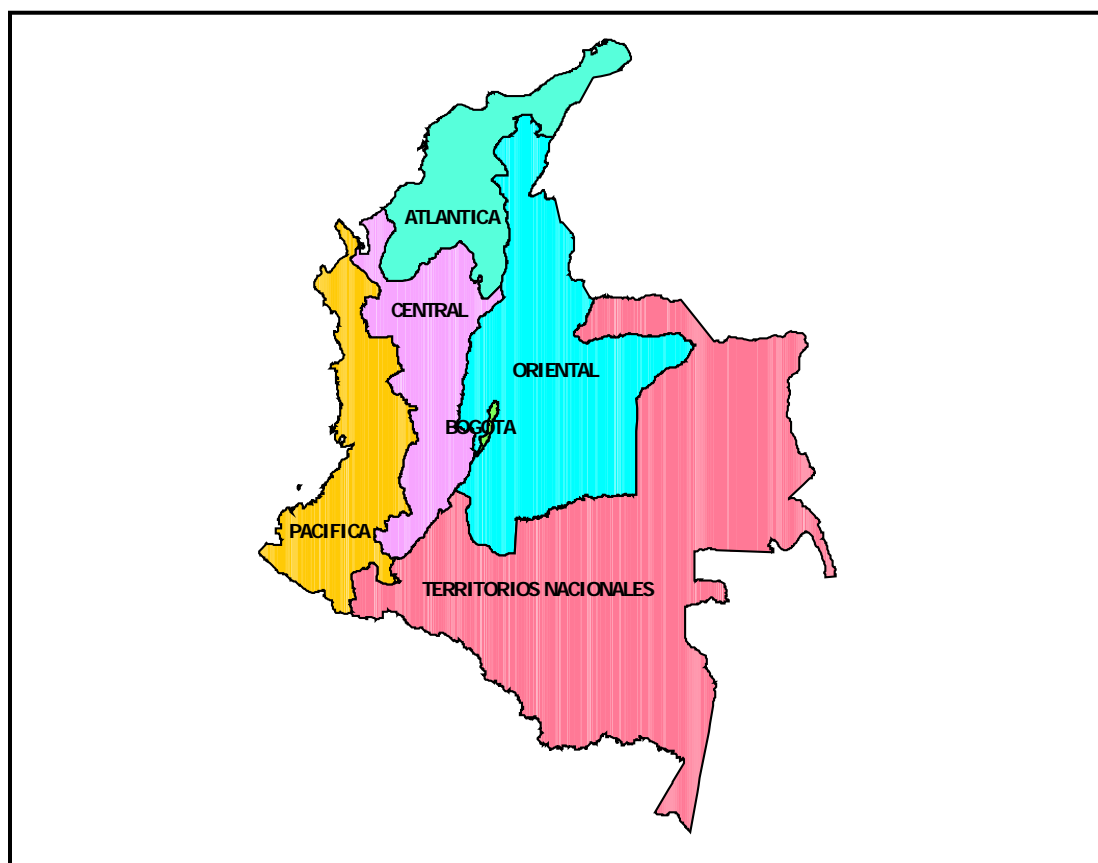


Gráfico 186. Distribución del país por estratos (regiones)

Comentarios finales sobre el trabajo de campo

Sobre el cuestionario

El cuestionario fue largo, 60 minutos en promedio, pero en general los informantes estuvieron dispuestos a contestar las encuestas y tuvimos muy pocas entrevistas interrumpidas.

Acerca del trabajo de campo

Durante esta medición no se presentaron problemas de orden público, por lo que pudimos cumplir con la recolección de todas las encuestas en la totalidad de los municipios contemplados en la muestra.

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.