

Apéndice I. Descripción técnica del diseño muestral

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.

Se usaron proyecciones al 2013 del Censo de Población de 2005, el más reciente en Colombia. De acuerdo con el Censo, 20,3% de la población habita en la región Atlántica, 17,2% en la Pacífica, 25,5% en la región Central, 18,4% en la región Oriental, 1,2% en los Antiguos Territorios Nacionales y 17,3% 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 1.511 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 AI.1 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 AI.1 - Errores estándar y efectos de diseño para algunas variables

Estadísticos univariantes

		Estimación	Error típico	Intervalo de confianza al 95%		Coeficiente de variación	Efecto del diseño	Raíz cuadrada del efecto del diseño
				Inferior	Superior			
Media	Años	37,68	,208	37,27	38,09	,006	,270	,519

Ahora, hablando de la economía... ¿Cómo calificaría la situación económica DEL PAÍS? ¿Diría usted que es muy buena, buena, ni buena ni mala, mala o muy mala?

		Estimación	Error típico	Intervalo de confianza al 95%		Coeficiente de variación	Efecto del diseño	Raíz cuadrada del efecto del diseño
				Inferior	Superior			
% del total	Muy buena	,4%	,2%	,2%	1,0%	,482	1,388	1,178
	Buena	10,0%	,8%	8,5%	11,8%	,084	1,173	1,083
	Ni buena, ni mala (regular)	51,8%	1,4%	49,0%	54,5%	,027	1,199	1,095
	Mala	28,9%	1,3%	26,4%	31,4%	,044	1,158	1,076
	Muy mala (pésima)	9,0%	,8%	7,5%	10,7%	,089	1,178	1,085
	Total	100,0%	,0%	100,0%	100,0%	,000		

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 AI.2 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 AI.2 - Errores estándar y efectos de diseño para edad, por región

Estadísticos univariantes

[COESTRA] Estrato primario de la muestra		Estimación	Error típico	Intervalo de confianza al 95%		Coeficiente de variación	Efecto del diseño	Raíz cuadrada del efecto del diseño
				Inferior	Superior			
Atlántica	Media Años	36,86	,346	36,18	37,55	,009	,171	,413
Bogotá	Media Años	37,83	,505	36,83	38,82	,013	,485	,696
Central	Media Años	37,99	,342	37,32	38,67	,009	,150	,388
Oriental	Media Años	38,41	,487	37,45	39,37	,013	,207	,455
Pacífica	Media Años	37,04	,407	36,24	37,84	,011	,128	,357
Antiguos Territorios nacionales	Media Años	38,71	1,102	36,54	40,88	,028	,031	,177

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 AI.3 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 AI.3 - Muestra vs. Censo poblacional de 2005 (mayores de 18 años)

Características	Censo de 2005	Encuesta nacional 2013
N	31.590.796 ²	1.511
% de hombres	48	50
% > 30 años	68	62
% soltero	30	37
% casado o en unión libre	56	54
% con primaria	38	22
% con secundaria	37	51
% con educación post-secundaria	25	25
% en Región Atlántica	21	19
% en Bogotá	16	17
% en Región Central	25	24
% en Región Oriental	18	19
% 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 solteros además de 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 51% en la Encuesta de 2013, en primaria en el censo 2005 38% y en el 2013 22%.

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.

² Según proyección de población del DANE 2013

La Tabla AI.4 compara las características de la muestra entre hombre y mujeres.

Tabla AI.4 - Características de la muestra, por sexo

	Género del entrevistado establecido en la cuota			
	Hombre		Mujer	
	Media	% del N de la columna	Media	% del N de la columna
Años	38		37	
¿Cuál es su estado civil? [Leer alternativas]	Soltero	40,8%		33,9%
	Casado	27,0%		24,8%
	Unión libre (acompañado)	27,5%		29,1%
	Divorciado	,6%		1,1%
	Separado	3,1%		5,4%
	Viudo	,9%		5,7%

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 (50% para cada uno). Sus edades también son muy similares entre sí. Se observan pequeñas diferencias en estado civil entre hombres y mujeres.

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:
 - 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:
 - 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 distribuida 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 AI.5:

Tabla AI.5 - Tamaño de muestra y márgenes de error - Nivel de confianza del 95%

	Población	Muestra	error
ATLANTICA	6.419.395	288	5,8%
BOGOTA	5.475.880	264	6,0%
CENTRAL	8.054.813	360	5,2%
ORIENTAL	5.822.234	287	5,8%
PACIFICA	5.439.181	264	6,0%
TERRITORIOS NACIONALES	379.293	48	14,1%
Total general	31.590.796	1.511	2,5%

Información adicional sobre la muestra

Tabla AI.6 - Universo, población total por región y áreas urbanas y rurales

Colombia: población, personas de 18 años o más, proyección año 2013

Región	Urbano	Rural	Total
ATLANTICA	4.814.872	1.604.523	6.419.395
BOGOTA	5.464.260	11.620	5.475.880
CENTRAL	6.112.155	1.942.658	8.054.813
ORIENTAL	4.109.216	1.713.018	5.822.234
PACIFICA	3.826.821	1.612.360	5.439.181
TERRITORIOS NACIONALES	177.994	201.299	379.293
Total general	24.505.319	7.085.477	31.590.796

Distribución porcentual

Región	Urbano	Rural	Total
ATLANTICA	75%	25%	100%
BOGOTA	100%	0%	100%
CENTRAL	76%	24%	100%
ORIENTAL	71%	29%	100%
PACIFICA	70%	30%	100%
TERRITORIOS NACIONALES	47%	53%	100%
Total general	78%	22%	100%