

Apéndice I. Descripción técnica del diseño muestral

Diseño de la muestra¹

La Muestra Especial diseñada para para los 116 municipios Colombianos que hacen parte del Plan Nacional de Consolidación, 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 y los Antiguos Territorios Nacionales. 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.

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 1.505 personas. Técnicamente, nuestro margen de error fue $\pm 2.5\%$. Esto quiere decir que si hiciéramos múltiples muestras de este conjunto de municipios, 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.

Tabla AI.1 - Errores estándar y efectos de diseño para algunas variables

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.

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	37,00	,173	36,66	37,34	,005	,204	,452

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	,6%	,2%	,3%	1,1%	,360	1,080	1,039
Buena	7,2%	,8%	5,9%	8,9%	,106	1,307	1,143
Ni buena, ni mala (regular)	55,1%	1,5%	52,1%	58,1%	,028	1,428	1,195
Mala	30,6%	1,4%	27,9%	33,5%	,046	1,377	1,174
Muy mala (pésima)	6,4%	,7%	5,2%	8,0%	,112	1,279	1,131
Total	100,0%	,0%	100,0%	100,0%	,000		

De acuerdo a las anteriores tablas el diseño por conglomerados para esta encuesta fue eficiente, la pregunta edad obtuvo un def inferior a 1, y la pregunta ¿Cómo calificaría la situación económica del País? Obtuvo def un poco por encima de 1. Los errores estándar 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	37,50	,391	36,73	38,28	,010	,175	,419
Central Media Años	37,09	,357	36,38	37,79	,010	,150	,388
Oriental Media Años	37,02	,421	36,19	37,84	,011	,149	,385
Pacífica Media Años	36,80	,288	36,24	37,37	,008	,266	,516
Antiguos Territorios nacionales Media Años	36,74	,535	35,69	37,79	,015	,130	,360

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.

A nivel operativo se toman diferentes medidas para evitar sesgos: la selección de cada una de las unidades muestrales se hace mediante mecanismos completamente aleatorios; la muestra incluye población rural y urbana, 65 de los 116 municipios que hacen parte del Plan Nacional de Consolidación y se fijan cuotas para controlar el número de encuestas por sexo y por grupos de edad para que la muestra se asemeje en estas características a la población representada.

Usando la proyección de población 2013 del DANE, los municipios incluidos en la estrategia presentan un 59,7 de población en área urbana 40,3% en área rural, como lo muestra la siguiente tabla, se puede afirmar que la estructura de la muestra se asemeja bastante a la estructura poblacional.

**Tabla AI.3 – Distribución de la muestra por área rural y área urbana
[ESTRATER] Urbano / rural [Usar definición censal del país]**

	Estimación	Error típico	Intervalo de confianza al 95%		Coeficiente de variación
			Inferior	Superior	
% del total Urbano	61,1%	3,3%	54,4%	67,5%	,055
Rural	38,9%	3,3%	32,5%	45,6%	,086
Total	100,0%	,0%	100,0%	100,0%	,000

La proyección poblacional del DANE (habitantes de 18 años o más) indica que el 49% son hombres y 51% mujeres, teniendo en cuenta que el plan operativo controlaba cuotas la estructura poblacional de la muestra según esta característica también es muy similar a la estructura poblacional.

La ¡Error! No se encuentra el origen de la referencia. Distribución de la muestra por género

Género del entrevistado establecido en la cuota

	Estimación	Error típico	Intervalo de confianza al 95%		Coeficiente de variación	Efecto del diseño	Raíz cuadrada del efecto del diseño
			Inferior	Superior			
% del total Hombre	50,0%	,0%	49,9%	50,1%	,001	,001	,036
Mujer	50,0%	,0%	49,9%	50,1%	,001	,001	,036
Total	100,0%	,0%	100,0%	100,0%	,000	.	.

La tabla AI.5 expone algunas características de los encuestados comparando hombre y mujer, se evidencia que por grupos etareos no se presentan diferencias significativas, en cuanto a nivel educativo se observa que en nivel primaria hay una mayor proporción de mujeres que de

hombres, en nivel secundaria es mayor la proporción de hombres que de mujeres. En cuanto estado civil se observa que la proporción de solteros es mayor para hombres.

La ¡Error! No se encuentra el origen de la referencia.5 compara las características de la muestra entre hombre y mujeres.

		Género del entrevistado	
		Hombre	Mujer
Rango de edad del entrevistado establecido en la cuota	18-27	33,2%	33,4%
	28-40	33,5%	33,2%
	41 o más	33,3%	33,4%
Nivel	Ninguno	4,6%	3,2%
	Primaria	29,7%	35,2%
	Secundaria	51,8%	46,5%
	Universitaria	8,9%	7,0%
	Superior no universitaria	4,9%	7,8%
	NS	,0%	,1%
	NR	,0%	,0%
¿Cuál es su estado civil? [Leer alternativas]	Soltero	31,9%	23,5%
	Casado	19,4%	20,9%
	Unión libre (acompañado)	44,0%	43,9%
	Divorciado	,4%	1,5%
	Separado	3,1%	5,2%
	Viudo	1,3%	5,1%

Descripción técnica del diseño muestral

Universo

El universo de la encuesta tiene una cobertura nacional de todos los adultos que habitan en los 116 municipios pertenecientes al Plan Nacional de Consolidación, los cuales se encuentran divididos en cinco regiones del país: 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

Se diseñó un 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 (Municipios pertenecientes al Plan Nacional de Consolidación)
- 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 Región Atlántica
 - o 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

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 63 municipios de los 116 que componen el Universo.

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.

Información adicional sobre la muestra

Tabla AI.3 - Universo, población total por región y áreas urbanas y rurales-
Municipios PNAC: población, personas de 18 años o más, proyección año 2013

Región	Número municipios	Población total	Cabecera	Resto
ATLANTICA	14	561.968	315.869	246.099
CENTRAL	27	577.350	316.927	260.423
ORIENTAL	25	413.267	233.038	180.229
PACIFICA	39	1.442.292	940.286	502.006
TERRITORIOS NACIONALES	11	209.676	107.622	102.054
Total general	116	3.204.553	1.913.743	1.290.810

Región	Número municipios	Población total	Cabecera	Resto
ATLANTICA	14	18%	17%	19%
CENTRAL	27	18%	17%	20%
ORIENTAL	25	13%	12%	14%
PACIFICA	39	45%	49%	39%
TERRITORIOS NACIONALES	11	7%	6%	8%
Total general	116	100%	100%	100%