

# Plan de muestreo para el estudio de obesidad, sobrepeso y variables biopsicosociales en niños y adolescentes escolarizados de Cúcuta, Colombia

*Sampling plan for the study of obesity, overweight and biopsychosocial variables in children and adolescents in school age from Cúcuta, Colombia*

Juan Hernández-Lalinde<sup>1</sup>, Mgtr., <https://orcid.org/0000-0001-6768-1873>; [j.hernandezl@unisimonbolivar.edu.co](mailto:j.hernandezl@unisimonbolivar.edu.co), Jhon-Franklin Espinosa-Castro<sup>1</sup>, Mgtr., <https://orcid.org/0000-0003-2186-3000>; [j.espinosa@unisimonbolivar.edu.co](mailto:j.espinosa@unisimonbolivar.edu.co), Maricarmen Chacín<sup>2</sup>, Mgtr., <https://orcid.org/0000-0002-5208-9401>; [m.chacin@unisimonbolivar.edu.co](mailto:m.chacin@unisimonbolivar.edu.co), Sandra-Milena Carrillo-Sierra<sup>3</sup>, Mgtr., <https://orcid.org/0000-0001-9848-2367>; [scarrillo@unisimonbolivar.edu.co](mailto:scarrillo@unisimonbolivar.edu.co), Diego García Álvarez<sup>4</sup>, Mgtr., <https://orcid.org/0000-0002-9350-785X>; [diegogarcia\\_123@hotmail.com](mailto:diegogarcia_123@hotmail.com)

<sup>1</sup>Universidad Simón Bolívar, Departamento de Cs. Básicas, Sociales y Humanas, Cúcuta, Colombia.

<sup>2</sup>Universidad Simón Bolívar, Departamento de Investigación, Cúcuta, Colombia.

<sup>3</sup>Universidad Simón Bolívar, Facultad de Cs. Jurídicas y Sociales, Cúcuta, Colombia.

<sup>4</sup>Universidad Rafael Urdaneta, Facultad de Cs. Sociales y Administrativas, Maracaibo, Venezuela.

Autor para correspondencia: Juan Hernández-Lalinde. Universidad Simón Bolívar, Departamento de Ciencias Sociales y Humanas. Calle 14 entre avenidas 4 y 5, Barrio La Playa. C. P.: 540006. Cúcuta, Colombia. Correo electrónico: [j.hernandezl@unisimonbolivar.edu.co](mailto:j.hernandezl@unisimonbolivar.edu.co).

## Resumen

La investigación en las ciencias sociales, de la salud o del comportamiento, cuyo objeto de estudio es el hombre, obliga a enfrentarse a escenarios en los que, muchas veces, no está definida la población objetivo. Es bastante común en estas áreas del quehacer científico, que investigadores se topen con situaciones en las que deban recurrir a esquemas complejos de muestreo, incorporando elementos del muestreo aleatorio simple, estratificado y por conglomerados. La presente revisión tiene por objetivo servir como guía en este sentido, construyéndose sobre la base de un estudio real y en desarrollo, enfocado en analizar la posible relación entre el sobrepeso, la obesidad y variables biopsicosociales en niños y adolescentes escolarizados de Cúcuta, Colombia. Se ha intentado profundizar en el detalle del plan, examinando cuestiones fundamentales como las etapas utilizadas, los criterios de estratificación y conglomeración, el tamaño de la muestra de estudiantes, el número de unidades primarias y secundarias escogidas y la validación de la muestra mediante una comparación descriptiva. El elemento matemático se menciona solo de manera referencial, haciendo hincapié en el uso de programas como SPSS, para el diseño del plan, extracción de la muestra y obtención de estimaciones básicas; y G-Power, para el cálculo del tamaño de la muestra considerando aspectos como el nivel de significación, potencia, tamaño de efecto a detectar, grados de libertad, entre otros.

**Palabras clave:** muestras complejas, plan de muestreo, obesidad, sobrepeso, variables biopsicosociales, niños y adolescentes.

## Abstract

Research in social, health or behavioral sciences, whose object of study is man, requires facing circumstances in which the target population is often not defined. It is quite common in these areas of scientific work, that researchers bumping into situations in which they must resort to complex sampling schemes, incorporating elements of simple, stratified and cluster random sampling. This review aims to serve as a guide in this regard, building on the basis of a real and developing study, focused on analyzing the possible relationship between overweight, obesity and biopsychosocial variables in children and adolescents in school age from Cúcuta, Colombia. The plan has been explained in detail, examining fundamental issues such as the specific stages, the stratification and clustering criteria, the sample size of students, the number of primary and secondary sampling units needed, and the validation of the sample through a descriptive comparison. The mathematical element is mentioned only in a referential manner, with emphasis on the use of programs such as SPSS (for the design of the plan, sample extraction and basic estimation) and G-Power (for the calculation of the sample size according to the significance level, power, effect size, degrees of freedom, among others).

**Keywords:** complex samples, sampling plan, obesity, overweight, biopsychosocial variables, children and adolescents.

Las técnicas elementales de muestreo, tales como el aleatorio simple o estratificado, se sustentan en el hecho de que las unidades de análisis están dadas, contenidas explícitamente en un marco muestral que bien pudiera ser un listado, mapa, croquis o cualquier otro elemento que cumpla con el propósito de reunir detalladamente a los sujetos de una población de interés<sup>[1,2]</sup>. Sin embargo, gran parte de los estudios asociados con poblaciones humanas se llevan a cabo en condiciones en las que no es posible tener acceso a documentos de esta naturaleza, por lo que la selección de una muestra de este tipo es inviable. En otras situaciones, a pesar de que existe el marco muestral, aspectos como la dispersión geográfica, los traslados a zonas lejanas, el incremento de la logística necesaria o el aumento del número de encuestadores, hacen que el uso de estas técnicas sea prohibitivo por cuestiones presupuestarias<sup>[3-5]</sup>.

Como ejemplo, supóngase un escenario en el que se desea investigar la relación entre el estrés laboral de enfermeras de una ciudad colombiana y el sector de la salud en las que trabajan. Probablemente, el equipo de investigación especula que las enfermeras que atienden en hospitales públicos experimentan niveles más elevados de estrés que aquellas que laboran en instituciones privadas. Los investigadores pudieran tener acceso al marco muestral y obtener una lista con los nombres de las enfermeras si algún organismo, como ministerios o secretarías municipales, se lo suministrasen. Aun en este contexto, que es improbable por las dificultades que entraña, implicaría un costo elevado por el traslado a distintos centros de salud, tal vez para encuestar a una sola de estas trabajadoras. Sin duda resultaría más eficiente muestrear algunas instituciones, para luego encuestar a todas o algunas de las enfermeras de dichos sitios.

Estrategias de muestreo como la del ejemplo anterior son comunes en áreas de la salud. Un ejemplo de ello se halla en la encuesta norteamericana conocida como «National Health and Nutrition Examination Survey», o simplemente NHANES<sup>[2,6-10]</sup>. Otro caso en el que se ha aplicado un diseño complejo de muestreo es el estudio de obesidad y síndrome metabólico emprendido en la ciudad de Maracaibo, Venezuela<sup>[11-13]</sup>. En ambas situaciones fue necesario combinar diferentes técnicas para poder alcanzar al individuo del que se obtuvo la información, mezclando aspectos como la estratificación y conglomeración en varias etapas. Sobre la base de esta necesidad es que se propone este artículo de revisión. El objetivo del manuscrito es ofrecer un material de consulta basado en una investigación real y en desarrollo, estudio que se enfoca en analizar la posible relación entre el sobrepeso, la obesidad y factores biopsicosociales en niños y adolescentes escolarizados de Cúcuta, Colombia. El documento ahondará en los detalles del plan, utilizando para ello los programas SPSS y G-Power, abordando someramente el aspecto matemático.

### Plan de muestreo

El diseño del plan implicó las siguientes etapas: (1), definición de la población objetivo; (2), construcción del marco

muestral; (3), diseño del esquema de muestreo; (4), cálculo del tamaño muestral; (5), definición del tamaño y cantidad de clústeres; y (6) extracción y exploración de la muestra.

### 1. Definición de la población objetivo

Como se ha mencionado, la investigación se centra en analizar la posible relación entre la obesidad, el sobrepeso y diversas variables biopsicosociales en niños y adolescentes escolarizados del municipio de Cúcuta, Colombia. En consecuencia, la población objetivo se obtuvo a partir de delimitar el universo compuesto por todos los alumnos inscritos en 2019 en el sistema educativo cucuteño. Esta base de datos fue suministrada por la Secretaría de Educación Municipal, únicamente para fines investigativos y conservando la confidencialidad de la información considerada sensible, de modo que los nombres y los documentos de identidad de los estudiantes fueron reemplazados por códigos alfanuméricos.

En primera instancia, la base de datos estaba conformada por un total de 171 971 registros. Este universo fue demarcado considerando los siguientes criterios:

- Zonas urbanas: por cuestiones logísticas y económicas, la investigación se delimitó únicamente a colegios localizados en zonas urbanas. En ocasiones, aquellos institutos ubicados en áreas rurales obligan a largos traslados, muchos de los cuales implican adentrarse en zonas selváticas, utilizar transportación animal para llegar a locaciones de difícil acceso y atravesar ríos, lagos o lagunas.
- Sectores del sistema educativo: para la estimación del sobrepeso y de la obesidad, así como para la evaluación de la relación entre estos trastornos y variables biopsicosociales, se tomaron en cuenta instituciones oficiales o públicas, pero también no oficiales o privadas.
- Colegios grandes: con el propósito de usar eficientemente los recursos del proyecto, solo se incluyeron en el marco muestral sedes cuya cantidad estudiantil fuese igual o mayor a 400 alumnos.
- Educación básica y media: por razones científicas, la estimación del sobrepeso y de la obesidad debía realizarse en niños y adolescentes. En Colombia, el sistema educativo que incluye estas etapas evolutivas es el de educación básica primaria (desde 1.º hasta 5.º), educación básica secundaria (desde 6.º hasta 9.º), y educación media (desde 10.º hasta 11.º).
- Niños y adolescentes: por el motivo esgrimido anteriormente, el rango de edad de los alumnos analizados debía oscilar desde 6 hasta 17 años.
- Jornadas académicas: en el sistema educativo del municipio hay seis jornadas; a saber: matutina, vespertina, nocturna, única, completa y fines de semana. Por cuestiones prácticas, se descartaron todas menos la matutina y vespertina.

Con base en lo anterior, la población objetivo quedó conformada por 82 259 alumnos, 50.30 % (n = 41 379) de los cuales eran mujeres, y 49.70 % (n = 40 880) eran hombres. La edad

promedio de la población fue de 11.16, con una desviación típica de 2.95 años. El 90.20 % (n = 74 194) de los estudiantes pertenecían al sector oficial de la educación, en tanto que el 9.80 % (n = 8065) restante estaba inscrito en colegios privados. La distribución según jornada fue la siguiente: 60.10 % (n = 49 438) de los niños o adolescentes asistían a clases en la mañana, mientras que el 39.90 % (n = 32 821) lo hacía en la tarde. Los porcentajes según grado, estrato y otros detalles pueden visualizarse en las **tablas 1 y 2**.

**Tabla 1. Descripción de la población objetivo según variables sociodemográficas**

| Variable               | Categorías | Recuento | Porcentaje |
|------------------------|------------|----------|------------|
| Sector educativo       | Oficial    | 74 194   | 90.20      |
|                        | No oficial | 8065     | 9.80       |
| Jornada académica      | Matutina   | 49 438   | 60.10      |
|                        | Vespertina | 32 821   | 39.90      |
| Grado                  | Primero    | 7976     | 9.70       |
|                        | Segundo    | 7604     | 9.24       |
|                        | Tercero    | 7741     | 9.41       |
|                        | Cuarto     | 8036     | 9.77       |
|                        | Quinto     | 8169     | 9.93       |
|                        | Sexto      | 12 252   | 14.89      |
|                        | Séptimo    | 11 057   | 13.44      |
|                        | Octavo     | 9042     | 10.99      |
|                        | Noveno     | 7491     | 9.11       |
|                        | Décimo     | 1683     | 2.05       |
| Estrato socioeconómico | Undécimo   | 1208     | 1.47       |
|                        | Estrato 0  | 932      | 1.13       |
|                        | Estrato 1  | 45 690   | 55.54      |
|                        | Estrato 2  | 25 761   | 31.32      |
|                        | Estrato 3  | 7015     | 8.53       |
|                        | Estrato 4  | 2417     | 2.94       |
|                        | Estrato 5  | 406      | 0.49       |
| Sexo del estudiante    | Estrato 6  | 38       | 0.05       |
|                        | Femenino   | 41 379   | 50.30      |
|                        | Masculino  | 40 880   | 49.70      |

**Tabla 2. Distribución de la población objetivo según sector, jornada y grado académico**

| Grado    | Sector oficial |            |            |            | Sector no oficial |            |
|----------|----------------|------------|------------|------------|-------------------|------------|
|          | Matutina       |            | Vespertina |            | Matutina          |            |
|          | Recuento       | Porcentaje | Recuento   | Porcentaje | Recuento          | Porcentaje |
| Primero  | 2322           | 2.82       | 4872       | 5.92       | 782               | 0.95       |
| Segundo  | 2301           | 2.80       | 4514       | 5.49       | 789               | 0.96       |
| Tercero  | 2325           | 2.83       | 4630       | 5.63       | 786               | 0.96       |
| Cuarto   | 2594           | 3.15       | 4608       | 5.60       | 834               | 1.01       |
| Quinto   | 3238           | 3.94       | 4109       | 5.00       | 822               | 1.00       |
| Sexto    | 7732           | 9.40       | 3651       | 4.44       | 869               | 1.06       |
| Séptimo  | 7388           | 8.98       | 2751       | 3.34       | 918               | 1.12       |
| Octavo   | 6369           | 7.74       | 1832       | 2.23       | 841               | 1.02       |
| Noveno   | 5269           | 6.41       | 1454       | 1.77       | 768               | 0.93       |
| Décimo   | 1132           | 1.38       | 240        | 0.29       | 311               | 0.38       |
| Undécimo | 703            | 0.85       | 160        | 0.19       | 345               | 0.42       |

## 2. Construcción del marco muestral

El marco muestral se elaboró a partir del listado suministrado por la Secretaría de Educación Municipal, manejando esta información en archivos con valores delimitados por coma (.csv). A esta base de datos se le aplicaron los filtros respectivos de acuerdo a los criterios enumerados en el apartado anterior y se utilizaron las siguientes características: (1), número correlativo; (2), sede o colegio; (3), sector educativo; (4), jornada académica; (5), grado; (6), grupo o sección; (7), código del estudiante; (8), código de identificación; (9), estrato socioeconómico; (10), fecha de nacimiento; (11), edad; y (12), peso del colegio según cantidad estudiantil. Como se verá más adelante, esta última variable fue empleada para ponderar la probabilidad de escogencia de los conglomerados en la primera etapa de muestreo. En definitiva, el marco muestral se constituyó en una matriz de 82259 filas por 12 columnas.

## 3. Diseño del esquema de muestreo

Debido a cuestiones económicas, se optó por un esquema de muestras complejas. Vale la pena aclarar que, al contar con el marco muestral hasta la unidad de información (código del estudiante), pudo haberse empleado una técnica más precisa como el muestreo aleatorio estratificado. Se descartó este método debido a la posibilidad de extraer una muestra considerablemente dispersa en términos geográficos, lo que incrementaría enormemente el gasto en recursos humanos, logística y transporte. Como se verá posteriormente, la pérdida de precisión fue compensada al incluir el efecto de diseño en el plan.

Así pues, se utilizó un muestreo aleatorio polietápico, estratificado y por conglomerados. Aleatorio debido a que las unidades primarias, secundarias y terciarias de muestreo fueron escogidas al azar; polietápico debido a que se emplearon tres fases para poder llegar al estudiante; estratificado en vista de que se usaron criterios de segmentación que garantizaran la presencia de ciertos elementos en la muestra; y por conglomerados ya que se necesitó de la formación de clústeres para alcanzar la unidad de análisis. A continuación, se describe en detalle cada aspecto del diseño:

- Primera etapa: en la primera fase se utilizó el sector educativo y la jornada académica como variables de estratificación, conformando así tres subpoblaciones: (1), sector oficial, jornada matutina; (2), sector oficial, jornada vespertina; y (3), sector no oficial, jornada matutina. Es importante aclarar que en la población delimitada no había colegios privados que funcionasen en horario vespertino, por tal razón este estrato no figura en el esquema anterior. En lo que respecta a los conglomerados, se formaron de manera natural según las sedes educativas; por lo tanto, con el diseño de la primera etapa se aseguró la presencia de institutos públicos y privados, de ambas jornadas y se abarataron los costos del proyecto al emplear los colegios como unidades primarias de muestreo.
- Segunda etapa: una vez seleccionadas las instituciones educativas se procedió a desarrollar la segunda etapa del plan. En este caso, la segmentación se hizo según grado y la conglomeración se realizó por grupos o secciones. Con esto, se garantizaría que en la muestra habría representa-

ción de cada uno de los grados y se optimizaría el uso de los recursos al elegir al azar los salones como unidades secundarias de muestreo.

- Tercera etapa: en este nivel del diseño se pudo alcanzar la unidad de información, de modo que en esta etapa solo se seleccionó al azar el número de estudiantes correspondiente a cada uno de los salones previamente escogidos. Nótese que es equivalente hablar en este momento de unidades terciarias o unidades de análisis.
- Métodos de muestreo en cada etapa: las unidades secundarias (salones) y las de información (alumnos) fueron elegidas mediante muestreo aleatorio simple sin reposición (MASSR). Por su parte, las unidades primarias (sedes) fueron seleccionadas mediante muestreo proporcional al tamaño (MPT), empleando el número de alumnos matriculados como factor de ponderación.
- Programa utilizado: el diseño del plan y la extracción de la muestra fue realizado con el módulo de muestras complejas de SPSS, versión 25 para Windows de 64 bits.

#### 4. Cálculo del tamaño de muestra

En este apartado se exponen los criterios necesarios para calcular el tamaño de muestra. Para ello, se asumieron ciertas condiciones y se tomaron en cuenta, tanto las variables que serían analizadas en la investigación, como los procedimientos estadísticos empleados.

- Tipo de variables: los instrumentos aplicados a los estudiantes permitieron medir características cuantitativas y cualitativas. De estas, las categóricas tuvieron un papel preponderante en el cálculo del tamaño muestral debido a que fueron utilizadas como factores, y la cantidad de niveles impactó en dicha estimación determinando los grados de libertad de algunos procedimientos. Para simplificar la explicación, se presentan en el **cuadro 1** las variables cualitativas más relevantes para el diseño del plan de muestreo y sus respectivas categorías.

**Cuadro 1. Descripción de las variables cualitativas más relevantes para el cálculo del tamaño muestral**

| Variables                                    | Niveles | Categorías   |
|--|---------|--|
| Sector educativo                             | 2       | Oficial y no oficial   |
| Jornada académica                            | 2       | Matutina y vespertina  |
| Grado  | 11      | Desde 1.º hasta 11.º   |
| Estrato socioeconómico                       | 7       | Desde 0 hasta 6  |
| Sexo   | 2       | Femenino y masculino   |
| Grupos etarios                               | 4       | 6-8, 9-11, 12-14 y 15-17   |
| Tipo de familia                              | 4       | Nuclear, recompuesta, monoparental y extensa   |
| Clasificación del IMC según OMS              | 5       | Bajo peso, zona riesgo para bajo peso, peso saludable, sobrepeso y obesidad  |
| Desempeño académico                          | 4       | Deficiente, aceptable, sobresaliente y excelente   |
| Desempeño disciplinario                      | 4       | Deficiente, aceptable, sobresaliente y excelente   |
| Antecedentes médicos personales o familiares | 2       | Sí y no  |
| Práctica deportiva diversa                   | 5       | Nunca, 1 a 2 veces en los últimos 7 días, 3 a 4 veces en los últimos 7 días, 5 a 6 veces en los últimos 7 días y 7 o más veces en los últimos 7 días |
| Nivel de actividad física                    | 4       | Inactivo, muy poco activo, moderadamente activo y muy activo   |
| Nivel de autoconcepto                        | 3       | Bajo, medio y alto   |
| Nivel de acoso escolar                       | 3       | Bajo, medio y alto   |

- Procedimientos estadísticos: el cálculo del tamaño muestral se sustentó en la necesidad de realizar los siguientes análisis estadísticos: (1), análisis de asociación mediante prueba chi-cuadrado; (2), comparación de proporciones por medio de prueba z; (3), contraste de medias a través de prueba t, ANOVA<sup>1</sup> o similares; y (4), estimación y comparación de OR<sup>2</sup> obtenidos por regresión logística.
- Aspectos generales: aunque las ecuaciones para el cálculo de n varían según el propósito, muchos de sus términos son recurrentes y permanecen invariables. Por consiguiente, se han asumido los siguientes criterios habituales: (1), nivel de significación para contrastes de hipótesis bilaterales de 0.05; y (2), potencia de prueba de 0.80.
- Tamaño del efecto: la magnitud de las diferencias observadas en un contraste se mide a partir del tamaño del efecto, lo que a su vez influye significativamente en la cantidad de sujetos necesarios para llevar a cabo tal comparación. En esta revisión se utilizaron las sugerencias planteadas por Cohen<sup>14</sup> para identificar efectos de tamaño moderado, información que puede consultarse en el **cuadro 2**. Nótese que las condiciones varían según el tipo de análisis estadístico.

**Cuadro 2. Tamaño del efecto según Cohen para varios procedimientos estadísticos**

| Procedimiento  | Nulo   | Pequeño     | Moderado    | Alto   |
|--|--------|-------------|-------------|--------|
| Prueba chi-cuadrado de independencia                   | < 0.10 | 0.10 – 0.29 | 0.30 – 0.49 | ≥ 0.50 |
| Prueba z para comparar dos proporciones independientes | < 0.20 | 0.20 – 0.49 | 0.50 – 0.79 | ≥ 0.80 |
| Prueba t-Student para muestras independientes          | < 0.20 | 0.20 – 0.49 | 0.50 – 0.79 | ≥ 0.80 |
| Análisis de varianza unifactorial de efectos fijos     | < 0.10 | 0.10 – 0.24 | 0.25 – 0.39 | ≥ 0.40 |

- Programa utilizado: las operaciones de cálculo se han efectuado con el programa G-Power, versión 3.1.9.2 para Windows de 64 bits. En consecuencia, el aspecto matemático se ha relajado y se ha omitido la presentación de ecuaciones. Para una consulta detallada, remitirse a Cochran<sup>[15]</sup>, Lohr<sup>[16]</sup>, Lemeshow et al.<sup>[17]</sup>, Levy y Lemeshow<sup>[18]</sup>, entre otros.
- Tamaños de muestra no ajustados: los procedimientos anteriores se han realizado asumiendo que el muestreo es equivalente a seleccionar elementos de un conjunto infinito, o que tras la extracción de cada individuo se repone al lote el número que lo identifica dentro de la población. Puesto que este no es el caso, se ajustaron los cálculos anteriores al total de unidades presentes en este universo. Tal corrección se obtiene al aplicar la fórmula  $n_{\text{final}} = n_0 / (1 + n_0 / N)$ , donde  $n_0$  corresponde a los tamaños muestrales no ajustados.
- Tasa de no respuesta y efecto de diseño: también fue necesario considerar fenómenos como el de la no respuesta y el efecto de diseño, en especial porque el plan involucró estrategias de muestras complejas. En lo concerniente al primer aspecto, se asumió un porcentaje de no respuesta de

1 Analysis of variance (análisis de varianza).

2 Odds ratio (razón de oportunidades).

30 %, valor que se obtuvo de consultar investigaciones similares. En cuanto al efecto de diseño, se empleó un ajuste de 3, cifra que suele ser usada en estudios vinculados a las ciencias de la salud<sup>19</sup>. Es importante aclarar que, aunque se pudiera haber realizado una estimación más precisa de este factor, utilizando para ello la correlación intraclase de los clústeres, se optó por una aproximación menos formal debido a dos limitaciones: por un lado, la falta de datos que permitiera estimar el coeficiente de correlación dentro de las sedes académicas y dentro de los grupos; y por el otro, el hecho de que al formarse de manera natural los conglomerados, presentaron tamaños desiguales<sup>16,19</sup>.

Una vez definidos todos estos puntos se emplearon las herramientas del programa G-Power para obtener los tamaños muestrales correspondientes. En la **tabla 3** puede visualizarse el  $n_0$  de cada análisis, además del  $n_{final}$  luego del ajuste por población finita, no respuesta y efecto de diseño. Compruébese que, para cubrir las exigencias estadísticas de todos los análisis a la vez, se seleccionó el  $n_{final}$  más grande, en este caso, el asociado a la prueba z de proporciones cuyo tamaño exige la inclusión de al menos 1344 sujetos.

**Tabla 3. Tamaños de muestra para cada uno de los procedimientos realizados**

| Procedimiento   | $n_0$ | $n_{final}$ |
|---|-------|-------------|
| Prueba chi-cuadrado de independencia <sup>(1)</sup>                         | 307   | 1193        |
| Prueba z para comparar dos proporciones independientes <sup>(2)</sup>       | 346   | 1344        |
| Prueba t para muestras independientes <sup>(3)</sup>                        | 128   | 498         |
| Análisis de varianza unifactorial de efectos fijos <sup>(4)</sup>           | 275   | 1069        |
| Regresión logística binaria simple con regresor dicotómico <sup>(5)</sup>   | 285   | 1108        |
| Regresión logística binaria múltiple con regresores variados <sup>(6)</sup> | 334   | 1297        |

(1) Se asumieron 40 grados de libertad como máximo: tabla de contingencia de  $(11 - 1) \times (5 - 1) = 40$ . (2) Se asumió  $p_1 = 0.60$ ,  $p_2 = 0.45$  (efecto moderado) y asignación  $n_2 = n_1$ . La muestra para cada grupo es de 672 sujetos. (3) Se utilizó una asignación  $n_2 = n_1$ . La muestra para cada grupo es de 64 sujetos. (4) Se asumió una cantidad máxima de 11 grupos. (5) Se asumió un OR = 2, una frecuencia de verdaderos positivos de 0.30, un  $R^2 = 0$  y una distribución binomial con  $p = 0.50$ . (6) Se asumió un modelo con 10 regresores y una probabilidad de verdaderos positivos de 0.30.

### 5. Definición del tamaño y cantidad de conglomerados a seleccionar

Ya se ha mencionado que las unidades primarias (UPM) y secundarias de muestreo (USM) fueron las instituciones educativas y los grupos académicos, respectivamente. Como es lógico, la formación natural de estos conglomerados hace que exhiban tamaños desiguales, situación que se presenta casi siempre en aquellas encuestas relacionadas con las ciencias sociales, de la salud o del comportamiento, en las que establecimientos como colegios, hospitales, universidades o centros de atención psicológica constituyen clústeres habituales<sup>17,20</sup>. Tomando en cuenta este escenario, se descartaron del universo aquellos colegios cuyo número de estudiantes fuese menor que 400, esto con la intención de maximizar la eficiencia en el uso de los recursos, pero también de

incrementar la homogeneidad en el tamaño de las UPM. En lo que respecta a las USM, el trabajo de establecer su tamaño fue más sencillo debido a que la cantidad de alumnos por sección fue bastante similar, salvo ciertas excepciones. Ahora bien, para definir cuántas unidades de muestreo serían elegidas en cada etapa, se realizó una aproximación a la inversa, empleando el tamaño muestral calculado en numeral anterior y los tamaños representativos de las sedes y grupos. El detalle de este procedimiento se explica a continuación:

- Paso 1: el total de sujetos se dividió en partes iguales para cada sector académico; esto es, 672 alumnos pertenecientes a colegios públicos y 672 a escuelas privadas. Conviene aclarar por qué no se utilizó una asignación proporcional al tamaño de los sectores educativos registrados en la población (recuérdese que la cantidad de estudiantes que asisten a instituciones oficiales supera en aproximadamente nueve a uno a la de los que toman clases en colegios privados). La razón de emplear una afijación simple reside en el interés de la investigación y en el tipo de análisis estadístico que se va a desarrollar. Nótese que la intención principal no es la estimación de parámetros, sino el contraste y los test de asociación. En estos casos, conviene que los n de los grupos sean relativamente similares.
- Paso 2: para definir cuántas secciones debían escogerse por cada sede, se usaron medidas de tendencia central robustas para identificar el número de alumnos por grupo. Se registraron aproximadamente 40 estudiantes por sección en los colegios públicos y 30 en los institutos privados, de modo que se optó por nivelar estas cifras y usar un tamaño promedio de 35 estudiantes. Al dividir 672 entre esta cantidad, se determinó que serían necesarios 20 grupos por cada sector educativo, y puesto que también se requería la presencia de todos los grados, se optó por escoger dos salones por cada nivel.
- Paso 3: la decisión de cuántas UPM seleccionar no se basó en criterios exclusivamente estadísticos, sino también en razonamientos logísticos y económicos. Así pues, y con la intención de lograr un equilibrio entre la potencia deseada y la disponibilidad presupuestaria, se decidió muestrear dos sedes públicas en horario matutino y dos en horario vespertino, más cuatro privadas en jornada matutina.
- Paso 4: el número de alumnos a escoger en cada salón se determinó al dividir el total de elementos necesarios en la muestra entre el número de grupos. Puesto que se consideraron dos aulas por cada grado en cada una de las ocho sedes, la cantidad de estudiantes quedó definida por  $1344 / (2 \times 11 \times 8) = 7.63 = 8$  (por redondeo). En el **cuadro 3** se puede visualizar esto en forma resumida y notar que el tamaño muestral proyectado es de 1408 individuos.

### Cuadro 3. Esquema resumido del plan de muestreo y cantidad de UPM, USM y alumnos escogidos

**Etapas 1:** dos colegios públicos en horario matutino, dos colegios públicos en horario vespertino y cuatro colegios privados en horario matutino.  
**Total UPM = 8.**

**Etapas 2:** dos grupos o secciones por cada grado. Once grados considerados (desde 1.º hasta 11.º). Ocho sedes seleccionadas en la primera etapa.  
**Total USM = 2 × 11 × 8 = 176.**

**Etapas 3:** 176 grupos seleccionados en la segunda etapa. Cantidad total de estudiantes mínima de 1344 (tamaño muestral). Número promedio de estudiantes por grupo a escoger de  $1344 / 176 = 7.63 = 8$  (por redondeo).  
**Total alumnos = 176 × 8 = 1408 (proyectado).**

## 6. Extracción y exploración de la muestra

La extracción de la muestra fue realizada con el módulo de muestras complejas de SPSS, razón por la cual no se profundizará en este punto acerca del procedimiento. Para realizar una consulta que sirva como guía en este sentido,

puede remitirse a la página de soporte de IBM<sup>[21]</sup>. Por otro lado, siempre resulta conveniente realizar una exploración de la muestra extraída a objeto de comprobar que se ha alcanzado la representatividad deseada. La **tabla 4** exhibe una comparación descriptiva entre los porcentajes de la población objetivo y los hallados en la muestra, además de señalar las estimaciones obtenidas al incorporar las ponderaciones generadas de forma automática por el programa. Compruébese que el n final no fue de 1408 sujetos, sino de 1395. Esto sucede porque en el diseño se emplearon tamaños aproximados para los conglomerados. Por otra parte, nótese que los porcentajes presentes en la población objetivo tienden a ser bastante similares a los encontrados en la muestra, salvo en el caso del sector educativo por las razones ya explicadas. Por último, se observa una ligera sobreestimación de la población ya que la proyección obtenida es de 86 536 sujetos, cuando la cantidad real es de 82 259, generando un error absoluto porcentual de 5.20 %.

Tabla 4. Comparación descriptiva entre la población objetivo y la muestra extraída

| Variable            | Categorías    | Porcentajes        |                        | Recuentos                               |                   |
|---------------------|---------------|--------------------|------------------------|---|-------------------|
|                     |               | Muestra (n = 1395) | Población (N = 82 259) | Estimado <sup>(1)</sup><br>(N = 86 536) | Real (N = 82 259) |
| Sector educativo    | Oficial       | 47.11              | 90.20                  | 78 528                                  | 74 194            |
|                     | No oficial    | 52.89              | 9.80                   | 8008                                    | 8065              |
| Jornada académica   | Matutina      | 66.24              | 60.10                  | 48 851                                  | 49 438            |
|                     | Vespertina    | 33.76              | 39.90                  | 37 686                                  | 32 821            |
| Grado               | Primer grado  | 10.32              | 9.70                   | 8935                                    | 7976              |
|                     | Segundo grado | 9.82               | 9.24                   | 8246                                    | 7604              |
|                     | Tercer grado  | 10.32              | 9.41                   | 8775                                    | 7741              |
|                     | Cuarto grado  | 10.32              | 9.77                   | 8526                                    | 8036              |
|                     | Quinto grado  | 11.47              | 9.93                   | 11 674                                  | 8169              |
|                     | Sexto grado   | 9.82               | 14.89                  | 15 552                                  | 12 252            |
|                     | Séptimo grado | 9.25               | 13.44                  | 11 651                                  | 11 057            |
|                     | Octavo grado  | 9.18               | 10.99                  | 5309                                    | 9042              |
|                     | Noveno grado  | 9.18               | 9.11                   | 4451                                    | 7491              |
|                     | Décimo grado  | 5.09               | 2.05                   | 1813                                    | 1683              |
| Estrato socioecon.  | Undéc. grado  | 5.23               | 1.47                   | 1603                                    | 1208              |
|                     | Estrato 0     | 0.43               | 1.13                   | 52                                      | 932               |
|                     | Estrato 1     | 33.55              | 55.54                  | 50 919                                  | 45 690            |
|                     | Estrato 2     | 24.52              | 31.32                  | 26 916                                  | 25 761            |
|                     | Estrato 3     | 19.57              | 8.53                   | 6001                                    | 7015              |
|                     | Estrato 4     | 19.28              | 2.94                   | 2335                                    | 2417              |
|                     | Estrato 5     | 2.37               | 0.49                   | 281                                     | 406               |
| Sexo del estudiante | Estrato 6     | 0.29               | 0.05                   | 32                                      | 38                |
|                     | Femenino      | 54.05              | 50.30                  | 46 330                                  | 41 379            |
|                     | Masculino     | 45.95              | 49.70                  | 40 206                                  | 40 880            |

(1): las estimaciones se han calculado con base en las ponderaciones que se generan automáticamente en el programa, al diseñar el plan de muestreo y extraer la muestra.

## Referencias

- Lehtonen R, Pahkinen E. Practical Methods for Design and Analysis of Complex Surveys. 2.ª ed. John Wiley & Sons, Ltd; 2004. 362 p. (Statistics in practice).
- Korn EL, Graubard BI. Epidemiologic studies utilizing surveys: accounting for the sampling design. Am J Public Health. septiembre de 1991;81(9):1166-73.
- Mukhopadhyay P. Topics in survey sampling. New York: Springer; 2001. 292 p. (Lecture notes in statistics).
- Chaudhuri A, Chaudhuri A, Stenger H, Stenger H. Survey Sampling: Theory and Methods, Second Edition [Internet]. 0 ed. CRC Press; 2005 [citado 14 de agosto de 2019]. Disponible en: <https://www.taylorfrancis.com/books/9781420028638>
- Thompson SK. Sampling. 3rd ed. Hoboken, N.J.: Wiley; 2012. 436 p. (Wiley series in probability and statistics).
- Alexander CM, Landsman PB, Teutsch SM, Haffner SM. NCEP-Defined Metabolic Syndrome, Diabetes, and Prevalence of Coronary

- Heart Disease Among NHANES III Participants Age 50 Years and Older. *Diabetes*. 1 de mayo de 2003;52(5):1210-4.
7. Seligman HK, Laraia BA, Kushel MB. Food Insecurity Is Associated with Chronic Disease among Low-Income NHANES Participants. *J Nutr*. 1 de febrero de 2010;140(2):304-10.
  8. Looker AC, Orwoll ES, Johnston CC, Lindsay RL, Wahner HW, Dunn WL, et al. Prevalence of Low Femoral Bone Density in Older U.S. Adults from NHANES III. *J Bone Miner Res*. 1997;12(11):1761-8.
  9. Merikangas KR, He J-P, Brody D, Fisher PW, Bourdon K, Koretz DS. Prevalence and Treatment of Mental Disorders Among US Children in the 2001–2004 NHANES. *Pediatrics*. 1 de enero de 2010;125(1):75-81.
  10. Mozumdar A, Liguori G. Persistent Increase of Prevalence of Metabolic Syndrome Among U.S. Adults: NHANES III to NHANES 1999–2006. *Diabetes Care*. 1 de enero de 2011;34(1):216-9.
  11. Bermúdez V, Pacheco M, Rojas J, Córdova E, Velázquez R, Carrillo D, et al. Epidemiologic Behavior of Obesity in the Maracaibo City Metabolic Syndrome Prevalence Study. Maedler K, editor. *PLoS ONE*. 18 de abril de 2012;7(4):e35392.
  12. Bermúdez V, Rojas J, Salazar J, Calvo MJ, Morillo J, Torres W, et al. The Maracaibo city metabolic syndrome prevalence study: primary results and agreement level of 3 diagnostic criteria. *Rev Latinoam Hipertens*. 2014;9(4):20-32.
  13. Bermúdez V, Marcano RP, Cano C, Arráiz N, Amell A, Cabrera M, et al. The Maracaibo City Metabolic Syndrome Prevalence Study: Design and Scope. *Am J Ther*. mayo de 2010;17(3):288-94.
  14. Cohen J. *Statistical power analysis for the behavioral sciences*. 2nd ed. Hillsdale, N.J.: L. Erlbaum Associates; 1988. 567 p.
  15. William G. Cochran. *Sampling Techniques*. 3rd ed. New York, NY: John Wiley & Sons, Inc.; 1977. 442 p. (Wiley series in probability and mathematical statistics).
  16. Sharon L. Lohr. *Sampling: Design and Analysis*. 2nd ed. Boston, MA: Brooks/Cole Cengage Learning; 2010. 609 p. (Advanced Series).
  17. Stanley Lemeshow, David W. Hosmer, Janelle Klar, Stephen K. Lwanga. *Adequacy of sample size in health studies*. New York, NY: John Wiley & Sons, Ltd; 1990. 247 p.
  18. Levy PS, Lemeshow S. *Sampling of populations: methods and applications*. 3rd ed. New York: John Wiley & Sons, Inc.; 1999. 525 p. (Wiley series in probability and statistics).
  19. Alf C, Lohr S. Sampling Assumptions in Introductory Statistics Classes. *Am Stat*. febrero de 2007;61(1):71-7.
  20. Tashakkori A, Teddlie C, Teddlie CB. *Handbook of Mixed Methods in Social & Behavioral Research*. SAGE; 2003. 792 p.
  21. IBM SPSS Complex Samples - Overview - United States [Internet]. 2019 [citado 14 de agosto de 2019]. Disponible en: <https://www.ibm.com/us-en/marketplace/spss-complex-samples>

### Indices y Bases de Datos:

AVFT está incluida en las bases de datos de publicaciones científicas en salud:

**OPEN JOURNAL SYSTEMS**

**REDALYC** (Red de Revistas Científicas de América Latina, el Caribe, España y Portugal)

**SCOPUS** de Excerpta Medica

**GOOGLE SCHOLAR**

**Scielo**

**BIREME** (Centro Latinoamericano y del Caribe de Información en Ciencias de la Salud)

**LATINDEX** (Sistema Regional de Información en Línea para Revistas Científicas de América Latina, el Caribe, España y Portugal)

**Índice de Revistas Latinoamericanas en Ciencias** (Universidad Nacional Autónoma de México)

**LIVECS** (Literatura Venezolana de Ciencias de la Salud)

**LILACS** (Literatura Latinoamericana y del Caribe en Ciencias de la Salud)

**PERIÓDICA** (Índices de Revistas Latinoamericanas en Ciencias)

**REVENCYT** (Índice y Biblioteca Electrónica de Revistas Venezolanas de Ciencias y Tecnología)

**SABER - UCV**

**EBSCO Publishing**

**PROQUEST**

[www.revhipertension.com](http://www.revhipertension.com)

[www.revdiabetes.com](http://www.revdiabetes.com)

[www.revsindrome.com](http://www.revsindrome.com)

[www.revistaavft.com](http://www.revistaavft.com)