Diseño Muestral en 2 Etapas para la Caracterización del Sistema Educativo Brasileño

Juan A. Camacho
Camilo A. Raba

Departamento de Estadística
Universidad Nacional de Colombia
Bogotá - Colombia
Febrero, 2025

UNIVERSIDAD
NACIONAL
DE COLOMBIA

Tabla de Contenido

Introducción

El presente proyecto implementa un diseño muestral basado en Muestreo Aleatorio Simple por Conglomerados en la primera etapa y Muestreo Aleatorio Simple en la segunda etapa (MASC-MAS), con el fin de garantizar una adecuada representatividad de las unidades educativas en Brasil. Este enfoque metodológico permite capturar la heterogeneidad del sistema educativo y optimizar la eficiencia en la estimación de los parámetros de interés.

La estrategia de muestreo parte de la selección aleatoria de unidades federativas (UFs), las cuales funcionan como conglomerados primarios. Posteriormente, dentro de cada UF seleccionada, se eligen a las instituciones de educación básica mediante una segunda etapa de Muestreo Aleatorio Simple (MAS).

El diseño muestral también contempla la aplicación de pesos de expansión, necesarios para extrapolar los resultados muestrales a la población total. Asimismo, se evaluará la precisión de las estimaciones obtenidas mediante el cálculo de errores estándar e intervalos de confianza.

Con este enfoque, se busca garantizar que el análisis resultante sea estadísticamente sólido y represente con fidelidad la estructura del sistema educativo en Brasil.

Marco de Antecedentes

- El Censo de Educación de Brasil es la principal fuente de información sobre la educación básica y superior en el país, realizado anualmente por el Instituto Nacional de Estudios e Investigaciones Educativas Anísio Teixeira (INEP).
- Proporciona datos detallados sobre infraestructura, matrícula, docentes, recursos tecnológicos y condiciones de enseñanza en las instituciones educativas.
- Permite conocer la situación actual de la educación en Brasil, identificando desigualdades regionales y sectoriales.
- Facilita la asignación de recursos financieros a programas educativos y proyectos de infraestructura escolar.
- Proporciona insumos para la evaluación de políticas gubernamentales y la planificación de nuevas estrategias para mejorar la educación básica y superior.

Objetivos

Objetivo General

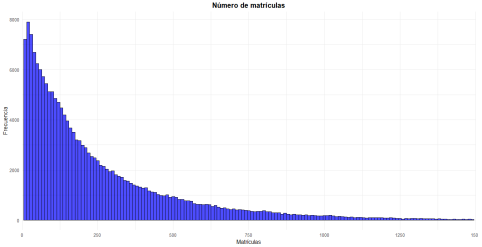
Estimar parámetros poblacionales de las instituciones de educación básica en Brasil, con el fin de caracterizar sus condiciones en términos de recursos disponibles, infraestructura y docentes.

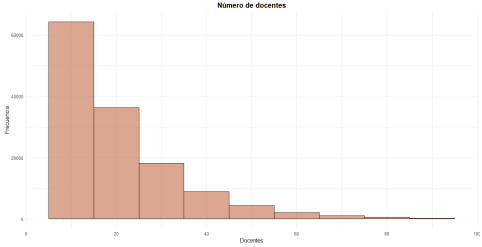
Objetivos Específicos

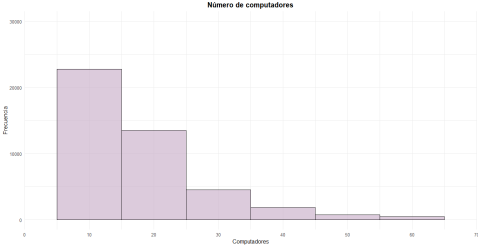
- Estimar la distribución de estudiantes, docentes, numero de clases y la infraestructura en el sistema de educación básica a nivel nacional mediante la estimacion de las medias y totales poblacionales.
- Estimar la disponibilidad estatal de infraestructura y recursos tecnológicos en las instituciones, considerando la proporción de acceso a internet y a bibliotecas.
- Estimar la distribución de las instituciones educativas según su tipo de dependencia administrativa (Federal, Estadual, Municipal o Privada) y su ubicación geográfica (urbana o rural), considerando la proporción de cada categoria.

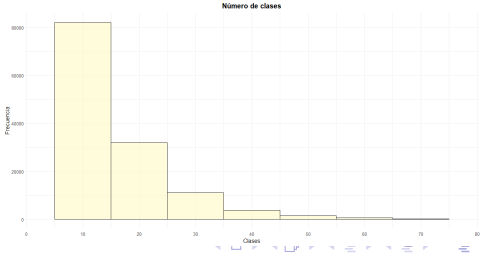
Tabla de Contenido

Variables cuantitativas



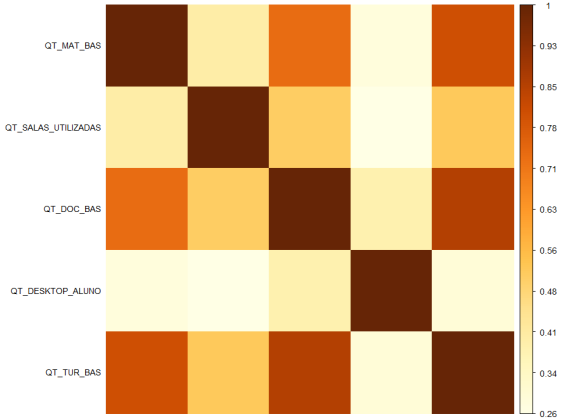




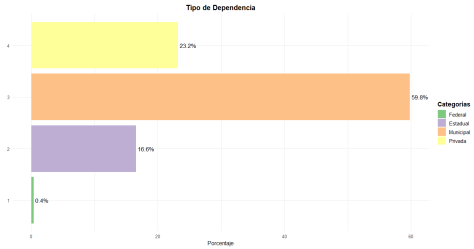


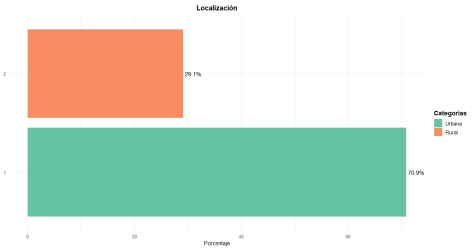
Variables cuantitativas

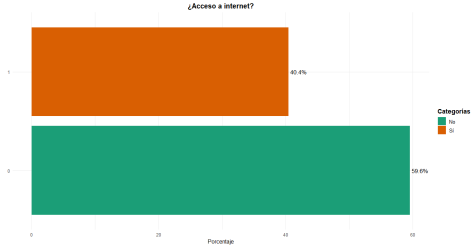
Matriz de correlaciones



Variables cualitativas







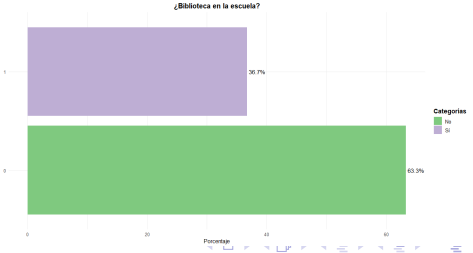


Tabla de Contenido

Diseño muestral

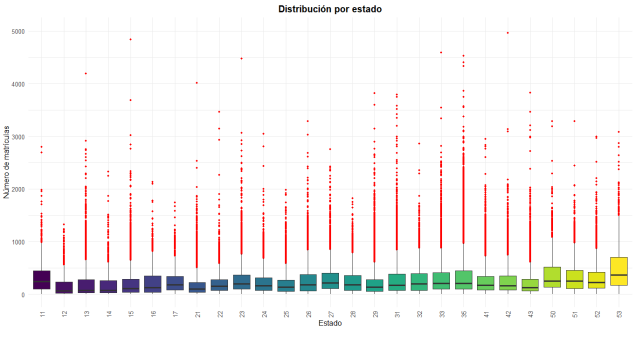
Se escogió un diseño MASC-MAS de acuerdo a los siguientes criterios:

- Brasil es un país que se caracteriza por su gran extensión territorial, sin embargo su organización política en UFs y municipios facilita la identificación de conglomerados naturales, lo que simplifica la selección y el acceso a datos.
- Las regiones del país (Norte, Nordeste, Sudeste, Sur y Centro-Oeste) presentan disparidades significativas en ingresos, educación y acceso a servicios, un muestreo por conglomerados nos da garantía que estas diferencias se vean representadas en la muestra.
- Se evaluo el comportamiento de la variable de interés segmentada por UFs, en donde se obtuvo que el valor del Coeficiente de Correlación Intraclase (ICC) es 0.039, lo que indica que solo el 3.9 % de la variabilidad en la variable matrículas en educación básica se explica por diferencias entre los estados. Esto sugiere que hay homogeneidad relativa en la distribución de matrículas entre los estados brasileños analizados.

Diseño Muestral

- El hecho de que la variabilidad dentro de los conglomerados sea mayor que la variabilidad entre conglomerados valida el uso de MAS en la segunda etapa, ya que garantiza que todas las unidades dentro de un conglomerado tengan igual probabilidad de ser seleccionadas, evitando sesgos.
- La implementación del MAS en la segunda etapa permite minimizar errores de selección y facilita el cálculo de estimadores y varianzas.

Justificación del diseño



- La variabilidad interna en los estados es elevada (rangos intercuantilicos grandes).
- Entre estados, si bien hay variabilidad, la distribución de las matrículas se observa relativamente homogénea.

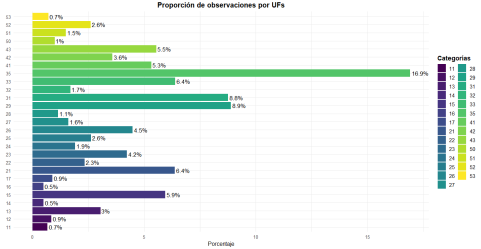
Notación

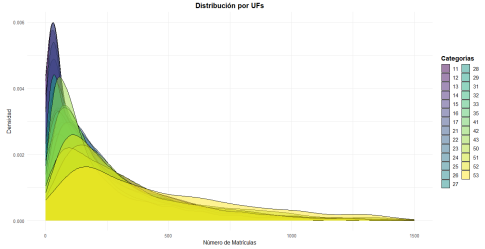
Usaremos la siguiente notación

- N : número de UFs en la población
- n : número de UFs seleccionados en una muestra aleatoria simple
- M_i : número de elementos de la UFs i
- m_i : número de elementos seleccionados en una muestra aleatoria de la UFs i
- $M = \sum_{i=1}^{N} M_i$: número de elementos de la población
- $\bar{M} = \frac{M}{N}$: tamaño de UFs promedio para la población
- y_{ij} : j-ésima observación en la muestra de la i-ésima UFs
- $\bar{y}_i$: $\frac{1}{m_i} \sum_{j=1}^{m_i} y_{ij}$ = media muestral para la i-ésima UFs

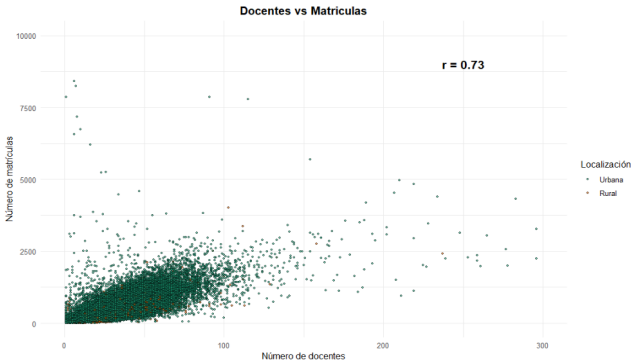
UPMs

Una UF (Unidad Federativa) en Brasil es una división político-administrativa que conforma la estructura federal del país. Brasil se divide políticamente en 27 Unidades Federativas (UFs), compuestas por 26 estados y 1 Distrito Federal (Brasilia, capital del país). Cada UF cuenta con autonomía para gestionar políticas locales en educación, seguridad y salud.





Variable auxiliar



La variable número de docentes tiene una importante correlación con la variable principal de estudio, número de matrículas, por tanto nos servirá como variable auxiliar para el proceso de muestreo.

Cálculo del tamaño de muestra total

El calculo del tamaño de muestra lo realizamos usando:

$$n = \text{Deff} \times n_{\text{MAS}} = [1 + (d-1)\rho_c] \times n_{\text{MAS}}$$

Donde:

$$n_{\text{MAS}} = \frac{z^2_{1-\frac{\alpha}{2}} S^2_{Y_U}}{EM^2} \quad \text{y} \quad \rho_c = 1 - \frac{S^2_{Y_W}}{S^2_{Y_U}}$$

- $\alpha = 0{,}05$
- $EM = 10$
- $d = \bar{M}$

Cálculo del número de conglomerados

Se utilizó la fórmula de Kish para optimizar el número de conglomerados:

$$m_{opt} = \sqrt{\frac{C_1(1-\rho_c)}{C_2\rho_c}}$$

Donde:

- C_1: costo por conglomerado.
- C_2: costo por unidad del conglomerado

Para asignar un valor apropiado para los costos, teniendo en cuenta los tamaños altos de los conclomerados, se supuso que $C1 = 30 \times C2$.

Estimador del total

Un estimador insesgado para el total está dado por

$$\hat{\tau} = M\hat{\mu} = \frac{N}{n}\sum_{i=1}^{n} M_i \bar{y}_i$$

La varianza del estimador está dada por

$$\hat{Var}[\hat{\tau}] = \left(\frac{N-n}{N}\right)\left(\frac{N^2}{n}\right)s_b^2 + \frac{N}{n}\sum_{i=1}^{n} M_i^2 \left(\frac{M_i - m_i}{M_i}\right)\left(\frac{s_i^2}{m_i}\right)$$

Donde

$$s_b^2 = \frac{\sum_{i=1}^{n}(M_i\bar{y}_i - \bar{M}\hat{\mu})^2}{n-1} \quad \text{y} \quad s_i^2 = \frac{\sum_{j=1}^{m_i}(y_{ij} - \bar{y}_i)^2}{m_i - 1}$$

Estimador de la media poblacional

Un estimador insesgado para la media poblacional está dado por

$$\hat{\mu} = \left(\frac{N}{M}\right) \frac{\sum_{i=1}^{n} M_i \bar{y}_i}{n_i}$$

La estimación de la varianza del estimador está dada por

$$\hat{Var}[\hat{\mu}] = \left(\frac{N-n}{N}\right)\left(\frac{1}{n\bar{M}^2}\right) s_b^2 + \frac{1}{nN\bar{M}^2} \sum_{i=1}^{n} M_i^2 \left(\frac{M_i - m_i}{M_i}\right)\left(\frac{s_i^2}{m_i}\right)$$

Donde s_b^2 y s_i^2 son como se definieron previamente.

Estimador de la proporción

Un estimador insesgado para la porporción está dado por

$$\hat{P} = \left(\frac{N}{M}\right) \frac{\sum_{i=1}^{n} M_i p_i}{n_i}$$

La estimación de la varianza del estimador está dada por

$$\hat{Var}[\hat{P}] = \left(\frac{N-n}{N}\right)\left(\frac{1}{n\bar{M}^2}\right) s_b^2 + \frac{1}{nN\bar{M}^2} \sum_{i=1}^{n} M_i^2 \left(\frac{M_i - m_i}{M_i}\right)\left(\frac{s_i^2}{m_i}\right)$$

Donde:

$$s_i^2 = \frac{m_i}{m_i - 1} p_i(1 - p_i)$$

Tabla de Contenido

Resultados generales

- Coeficiente de correlación intraclásica de la variable auxiliar: $\rho_c = 0{,}067$
- Tamaño promedio de la UFs: $\bar{M} = 6595$
- Tamaño de muestra total: $n_{TOT} = 3920$
- Número de UFs seleccionadas: $n = 20$

Resultados de las estimaciones

Resultados - Objetivo Especifico 1

- Estimar la distribución de estudiantes, docentes, numero de clases y la infraestructura en el sistema de educación básica a nivel nacional mediante la estimacion de las medias y totales poblacionales.
  - Estimación de Totales en el Sistema de Educación Básica en Brasil:

Descripción	Total Estimado	Total Real	SE	CV	IC_LI	IC_LS	Dentro_IC
Matrículas en Educación Básica	45558482	47279412	962945.93	2.11	43671142.7	47445821	Sí
Docentes en Educación Básica	2926286	2933396	47582.70	1.63	2833026.0	3019547	Sí
Numero de clases	2181970	2211154	32892.35	1.51	2117502.1	2246438	Sí
Cantidad de Salas Utilizadas	1695211	1560075	145255.42	8.57	1410515.1	1979906	Sí
Cantidad de Computadores para los Alumnos	1142851	1109359	82590.60	7.23	980976.9	1304726	Sí

  - Estimación de Medias en el Sistema de Educación Básica en Brasil:

Descripción	Media Estimada	Media Real	SE	CV	IC_LI	IC_LS	Dentro_IC
Promedio de Matrículas en Educación Básica	255.84	265.51	34.71	13.57	187.80	323.88	Sí
Promedio de Docentes en Educación Básica	16.43	16.47	2.31	14.04	11.91	20.96	Sí
Promedio de Clases en Educación Básica	12.25	12.42	1.68	13.74	8.95	15.55	Sí
Promedio de Salas Utilizadas	9.52	8.76	1.85	19.41	5.90	13.14	Sí
Promedio de Computadores de Escritorio por Alumno	6.42	6.23	1.21	18.89	4.04	8.79	Sí

Resultados - Objetivos Especificos 2 y 3

- Estimar la disponibilidad estatal de infraestructura y recursos tecnológicos en las instituciones, considerando la proporción de acceso a internet y a bibliotecas.
  - Estimación de la Proporción de Acceso a Internet y Bibliotecas en Instituciones Educativas:

Descripción	Proporción Estimada	Proporción Real	SE	CV	IC_LI	IC_LS	Dentro_IC
Acceso a Internet para Alumnos	0.4144	0.4044	0.0631	15.23	0.2907	0.5381	Sí
Presencia de Biblioteca en la Institución	0.3382	0.3670	0.0448	13.25	0.2504	0.4260	Sí

- Estimar la distribución de las instituciones educativas según su tipo de dependencia administrativa (Federal, Estadual, Municipal o Privada) y su ubicación geográfica (urbana o rural), considerando la proporción de cada categoria.
  - Estimación de la Distribución de Instituciones Educativas según Dependencia Administrativa y Ubicación Geográfica:

Categoría	Proporción Estimada	Proporción Real	SE	CV	IC_LI	IC_LS	Dentro_IC
Instituciones Federales	0.0038	0.0039	0.0011	28.59	0.0017	0.0059	Sí
Instituciones Estaduales	0.1622	0.1656	0.0231	14.25	0.1169	0.2076	Sí
Instituciones Municipales	0.5708	0.5984	0.0629	11.02	0.4475	0.6942	Sí
Instituciones Privadas	0.2385	0.2321	0.0520	21.81	0.1365	0.3404	Sí
Instituciones en Zona Urbana	0.7145	0.7089	0.1140	15.95	0.4911	0.9378	Sí
Instituciones en Zona Rural	0.2609	0.2911	0.0343	13.15	0.1936	0.3281	Sí

Tabla de Contenido

Conclusiones

- El diseño de muestreo en dos etapas (MASC-MAS) demostró ser adecuado para capturar la heterogeneidad del sistema educativo brasileño. La baja correlación intraclase indicó homogeneidad entre las Unidades Federativas (UFs), validando el uso de Muestreo Aleatorio Simple en la segunda etapa.
- Las estimaciones de totales y medias (matrículas, docentes, infraestructura) presentaron errores estándar y coeficientes de variación bajos (CV $< 15\,\%$ en la mayoría), confirmando la solidez metodológica. Ademas, los intervalos de confianza reflejaron alta fiabilidad en los resultados.
- Se evidenciaron desigualdades en el acceso a recursos educativos, como la disponibilidad de internet y bibliotecas, resaltando la importancia de políticas públicas enfocadas en reducir estas brechas.
- Los resultados proporcionan insumos críticos para la planificación educativa en Brasil, como la asignación de recursos tecnológicos y la expansión de infraestructura en zonas rurales. La metodología empleada puede replicarse en otros contextos con estructuras administrativas similares, garantizando estimaciones robustas para la toma de decisiones basada en evidencia.

Bibliografía I

[Bia01] Zélia Magalhães Bianchini. *Determinación del tamaño de la muestra para encuestas de hogares en dos etapas considerando el efecto de diseño*. es. 2001.

[Coc77] William G. Cochran. *Sampling Techniques (3rd Edition)*. en. John Wiley & Sons, 1977.

[Est23] Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP). *Microdatos del Censo Escolar de la Educación Básica*. pt. 2023. URL: https://www.gov.br/inep/pt-br/acesso-a-informacao/dados-abertos/microdados/censo-escolar.

[Jim21] Jerry Jiménez. *Técnicas de Muestreo C14. Muestreo por conglomerado de dos etapas*. es. YouTube Video. 2021.

[Kis65] Leslie Kish. *Survey Sampling*. en. John Wiley & Sons, 1965.

[Loh22] Sharon L. Lohr. *Sampling: Design and Analysis*. en. Chapman & Hall/CRC, 2022.

[SS92] Carl-Erik Särndal y Jan Swensson Bengt y Wretman. *Model Assisted Survey Sampling*. en. Springer-Verlag, 1992.