Universidad de San Carlos de Guatemala
Escuela de Ciencias Físicas y Matemáticas
Departamento de Matemática

# ELEMENTOS DE TEORÍA DE MUESTREO EN ENCUESTAS DE HOGARES

**Sergio Alexander Alay Arellano**
Asesorado por Cristian José Alvarez Bran

Guatemala, septiembre de 2024

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA

ESCUELA DE CIENCIAS FÍSICAS Y MATEMÁTICAS

# ELEMENTOS DE TEORÍA DE MUESTREO EN ENCUESTAS DE HOGARES

TRABAJO DE GRADUACIÓN
PRESENTADO A LA JEFATURA DEL
DEPARTAMENTO DE MATEMÁTICA
POR

**SERGIO ALEXANDER ALAY ARELLANO**
ASESORADO POR CRISTIAN JOSÉ ALVAREZ BRAN

AL CONFERÍRSELE EL TÍTULO DE
**LICENCIADO EN MATEMÁTICA APLICADA**

GUATEMALA, SEPTIEMBRE DE 2024

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA

ESCUELA DE CIENCIAS FÍSICAS Y MATEMÁTICAS

## CONSEJO DIRECTIVO INTERINO

| | |
|---|---|
| Director | M.Sc. Jorge Marcelo Ixquiac Cabrera |
| Representante Docente | Arqta. Ana Verónica Carrera Vela |
| Representante Docente | M.A. Pedro Peláez Reyes |
| Representante de Egresados | Lic. Urías Amitaí Guzmán García |
| Representante de Estudiantes | Elvis Enrique Ramírez Mérida |
| Representante de Estudiantes | Oscar Eduardo García Orantes |
| Secretario | M.Sc. Freddy Estuardo Rodríguez Quezada |

## TRIBUNAL QUE PRACTICÓ EL EXAMEN GENERAL PRIVADO

| | |
|---|---|
| Director | Director |
| Examinador | Examinador 1 |
| Examinador | Examinador 2 |
| Examinador | Examinador 2 |
| Secretario | Secretario |

# 28 AGRADECIMIENTOS

# DEDICATORIA

# ÍNDICE GENERAL

# ÍNDICE DE FIGURAS

76

# ÍNDICE DE TABLAS

# 77 LISTA DE SÍMBOLOS

| Símbolo | Significado |
|---|---|
| $:=$ | es definido por |
| $\cong$ | es isomorfo a |
| $\Leftrightarrow$ | si y sólo si |
| $\varnothing$ | conjunto vacío |
| $E^c$ | complemento de $E$ |
| $\subsetneq$ | estrictamente contenido |
| $E \setminus F$ | diferencia entre $E$ y $F$ |
| $E\Delta F$ | diferencia simétrica entre $E$ y $F$ |
| $\mathcal{P}(X)$ | conjunto potencia de $X$ |
| $\chi_E$ | función característica de $E$ |
| $E_n\uparrow$ | $E_n$ es una sucesión creciente |
| $\mathfrak{L}$ | $\sigma$-álgebra de los conjuntos Lebesgue-medibles |
| $\mathscr{S}$ | espacio muestral |
| $\mathfrak{A}$ | $\sigma$-álgebra de eventos |
| $(\mathscr{S}, \mathfrak{A}, P)$ | espacio de probabilidad |
| $\mathscr{D}$ | espacio de las funciones de prueba |
| $\mathscr{D}'$ | espacio de las distribuciones |
| $\delta_0$ | medida de Dirac, función $\delta$ de Dirac o $\delta$-función |
| $\Phi^\times$ | espacio antidual de $\Phi$ |
| $\Phi \subset \mathcal{H} \subset \Phi^\times$ | espacio de Hilbert equipado o tripleta de Gel'fand |
| $\vert\psi\rangle$ | vector *ket* |
| $\langle\psi\vert$ | funcional *bra* |
| $\langle\varphi\vert\psi\rangle$ | *braket* |

# OBJETIVOS

## General

Escriba el objetivo general.

## Específicos

Enumere los objetivos específicos.

1.
2.

# INTRODUCCIÓN

# 1. PRELIMINARES

## 1.1. Teoría elemental de probabilidad

### 1.1.1. Espacio muestral

El espacio muestra, también llamado espacio muestra, de un experimento aleaotorio es el conjunto dede todos los posibles resultados del experimento y se le denota, generalmente por la letra maýuscula $S$.

Por otro lado, llamaremos evento o suceso a cualquier subconjunto del espacio muestral $S$. A través de algunos ejemplos ilustraremos a continuacipon

### 1.1.2. Espacio de probabilidad

En el marco de la estadística y la teoría de muestreo, el concepto de espacio de probabilidad constituye la base matemática sobre la cual se construyen los modelos de inferencia. En esencia, un espacio de probabilidad es una estructura formal que permite modelar sutuaciones aleatorias y cuantificar la incertidumbre asociada a los posibles resultados de un experimento.

Un espacio de probabiliadad se define como una terna $(S, A, P)$, donde:

- 

### 1.1.3. Variables aleatorias

### 1.1.4. Distribuciones de probabilidad

#### 1.1.4.1. Binomial

#### 1.1.4.2. Normal

#### 1.1.4.3. Teorema del Límite Central

## 1.2. Inferencia estadística

### 1.2.1. Estimadores

- Definición formal, función del estimador, ejemplos.
- Propiedades: Insesgadez, consistencia y eficiencia.

### 1.2.2. Intervalos de Confianza

¿Cómo se definen? y en la práctica, ¿cómo se interpretan?

- Definición, interpretación y conexión con la precisión.
- Fórmulas básicas para media y proporciones.
- Ejemplos conceptuales: "el valor de $\theta$ se encuentra entre...".

# 2. ¿Qué son las encuestas?

Consideraremos una encuesta como una herramienta de investigación que se utiliza para recolectar datos e información de un grupo específico de personas, conocido como muestra, con el objetivo de inferir características de la población general de donde se tomó la muestra. Una encuesta se caracteriza por utilizar instrumentos tales como cuestionarios o entrevistas, en los que se plantean preguntas estandarizadas que permiten obtener respuestas comparables y cuantificables.

A lo largo de este capítulo se discutirán las generalidades, terminología e importancia de las encuestas.

## 2.1. ¿Qué elementos tiene una encuesta?

Con el propósito de asegurar la calidad y la efectividad de una encuesta, resulta fundamental considerar una serie de elementos clave que nos ayudarán para realizar su diseño, implementación y análisis. Estos elementos, además de garantizar la precisión de los datos recolectados, facilitan la interpretación y la aplicación práctica de los resultados y, principalmente, permiten generalizar a la población. A continuación se presentan los componentes que conforman una encuesta.

- **Unidad de observación.** Es el objeto sobre el que se realiza una medición. En el estudio de las poblaciones humanas, las unidades de observación suelen ser personas.

- **Población objetivo.** Es la colección completa de las unidades de observación que deseamos estudiar. Usualmente, definir la población objetivo es una parte difícil del estudio[1] y la elección de esta afectará profundamente a las estadísticas resultantes.

[1] Por ejemplo, en un estudio de eficacia de un nuevo programa educativo ¿la población objetivo debe incluir solo a los estudiantes de un cierto grado escolar?, ¿deberíamos limitar la población objetivo a estudiantes de una región específica o, incluir estudiantes de todo el país? intentá resumir esto.Ya.

- **Muestra.** Subconjunto de la población que se utiliza para inferir o sacar conclusiones sobre toda la población.

- **Unidad de muestreo.** Es una unidad que puede ser seleccionada para formar parte de la muestra. Se puede dar la situación de que queramos estudiar personas, pero no dispongamos de una lista de todas las personas de la población objetivo. En su lugar, los hogares sirven como unidades de muestreo y las unidades de observación son las personas que viven en los hogares.

- **Marco de muestreo.** Es una lista, un mapa u otra especificación de las unidades de muestreo de la población a partir de las cuales puede seleccionarse una muestra. Es importante que este marco cuente con información que permita ubicar de manera pertinente[2] a cada una de las unidades que lo conforman.

  Esto me gusta, pero no sé si aquí, dejemoslo por el momento y evaluemos si de repente hay un mejor lugar donde ponerlo Todos los marcos de muestreo presentan algún nivel de desactualización con respecto a la población objetivo. Por ejemplo, un marco de muestreo de áreas basado en cartografía, puede estar desactualizado debido a que es posible:

  - entrevistar a la misma persona en varias ocasiones,
  - nunca realizar la entrevista a una persona que no tiene un hogar fijo de residencia.

Esto también me gusta pero, de nuevo, no sé si aquí Es esencial saber que en las encuestas de personas, la población total suele ser menor que la población objetivo, es decir, no todas las peronas de la población objetivo están incluidas en el marco de muestreo, y varias personas no responderán a la encuesta.

Veamos cómo identificar algunos de estos elementos en dos de las encuestas que se han realizado en Guatemala.

**2.1 Ejemplo.** El Ministerio de Eduación -MINEDUC- realizó el estudio “Situación nutricional de los escolares inscritos en los centros educativos públicos que son beneficiarios del Programa de Alimentación Escolar, 2023” mediante la medición de peso y longitud/talla en una muestra estadísticamente significativa. En este estudio, las unidades de observación son los estudiantes matriculados en escuelas públicas,

[2]Para una encuesta de hogares con entrevistas *in situ*, las direcciones de los hogares de las personas a entrevistar. En el caso de una encuesta telefónica, los teléfonos de las personas a entrevistar.

ya que son los “objetos” sobre los cuales se realizaron las mediciones de peso y longitud. La población objetivo del estudio son todos los estudiantes matriculados en escuelas públicas durante el año 2023. Esta es la colección completa de unidades de observación que se desea estudiar.

**2.2 Ejemplo.** La Encuesta de Evaluación de la Calidad de los Servicios Públicos Básicos -ENCASBA- 2019, brinda información acerca de los servicios públicos que proporciona el Organismo Ejecutivo, la Municipalidad de Guatemala y otras instituciones públicas, la misma se sustenta en la evaluación que hizo la población informante del hogar, la cual podría ser el jefe o jefa del hogar o una persona de 18 años o más de edad. La población objetivo está compuesta por las viviendas y los hogares del municipio de Guatemala. Dentro de este contexto, las unidades de muestreo son los hogares, ya que cada hogar puede ser seleccionado para formar parte de la muestra. La muestra son aquellas unidades de muestreo que son seleccionadas para el recabamiento de su información. El marco de muestreo consiste en un listado detallado de todas las áreas geográficas y también, del listado de los hogares dentro de cada área geográfica del municipio, permitiendo identificar de manera precisa las unidades de muestreo elegibles para la encuesta.

Dependiendo de la forma en la que se recopila la información de una encuesta en el tiempo, definimos dos enfoques principales:

- **Encuestas transversales.** Recopilan datos en un solo punto en el tiempo. Se utilizan para describir las características de una población en un momento específico y son útiles para realizar análisis de prevalencia o describir condiciones actuales.
- **Encuestas longitudinales.** Siguen las mismas unidades de observación (como hogares o individuos) a lo largo del tiempo. Esto permite analizar cambios y desarrollar comprensiones causales sobre cómo y por qué ocurren los cambios.

## 2.2. Relevancia de las encuestas de hogares

Las encuestas de hogares son utensilios esenciales en la investigación social, económica y demográfica. Proveen datos en diferentes tópicos:

1. *Salud y educación.* Proveen datos importantes sobre el acceso y calidad de servicios de salud y educación, lo que permite identificar desigualdades y áreas

donde se requiere intervención. Esto es crucial para mejorar la cobertura y calidad de estos servicios (Incluir referencia: UNICEF - MICS).

2. *Investigación económica.* Datos sobre ingresos, gastos, empleo y otras variables económicas que son fundamentales para el análisis macro y microeconómico (Incluir referencia: OECD - Household Wealth).

3. *Calidad de Vida y Bienestar.* Recogen información sobre la calidad de vida y el bienestar subjetivo de las personas, incluyendo aspectos como la satisfacción con la vida, condiciones de vivienda, y acceso a servicios básicos. Ayudando así a construir un panorama integral del bienestar de la población (Incluir referencia: OECD - How's Life?).

4. *Desarrollo Sostenible.* En el contexto de los Objetivos de Desarrollo Sostenible -ODS-, las encuestas de hogares son una fuente de datos para monitorear el progreso hacia dichos objetivos (Incluir referencia: UN - SDG Indicators).

El diseño de la encuesta dependerá del objetivo de la medición, lo que se pretende al diseñar una encuesta de hogares es que esta sea un instrumento confiable, que brinde estimaciones exactas y precisas, ya que, de lo contrario, no se podrían monitorear de manera consistente las Políticas Públicas y los indicadores de interés.

## 2.3. Errores muestrales y no muestrales

En este apartado se trata de describir el paradigma de los errores que se cometen en una encuesta y cómo medirlos de manera acertada y acotarlos sobre la base del *principio de representatividad.* Para esto empezamos definiendo dicho principio.

*Principio de representatividad en encuestas.* Afirma que cada elemento incluido en una muestra se representa a sí mismo y a un grupo de unidades que no pertenecen a la muestra seleccionada, cuyas características son cercanas a las del elemento incluido en la muestra (Incluir referencia).

Este principio es crucial para garantizar que los resultados de la encuesta sean generalizables y reflejen con precisión las opiniones, comportamientos o características de toda la población objetivo. Existen dos fuentes principales de error cuando se realiza una encuesta:

1. **Error muestral.** Este tipo de error ocurre porque no se incluyó a todas las personas de la población y se seleccionó una muestra. La magnitud del

error muestral depende del tamaño de la muestra y del método de muestreo utilizado. En general, un tamaño de muestra mayor tiende a reducir el error muestral.

2. **Error no muestral.** Este tipo de error abarca todos los demás errores que pueden ocurrir en una encuesta que no están relacionados con el hecho de que se ha utilizado una muestra en lugar de un censo. Estos errores pueden surgir en cualquier etapa del proceso de la encuesta, desde el diseño del cuestionario hasta la recolección y el análisis de los datos. El error no muestral puede ser sistemático o aleatorio. Entre los diferentes tipos de errores no muestrales están:

    - *Error de cobertura.* Ocurre cuando algunos miembros de la población objetivo no tienen la oportunidad de ser seleccionados en la muestra.
    - *Error de respuesta.* Surge cuando los encuestados no proporcionan respuestas precisas, ya sea intencionalmente o por confusión; malentendidos del cuestionario o respuestas socialmente deseables.
    - *Error de no respuesta.* Se produce cuando ciertas personas seleccionadas para la muestra no responden a la encuesta, y estas personas pueden tener características u opiniones diferentes a las de quienes sí responden.
    - *Errores de procesamiento.* Incluyen errores en la entrada de datos, codificación o análisis de los datos de la encuesta.

Dentro de los errores no muestrales existe un tipo de error denominado **sesgo** que afecta la validez de los resultados de la encuesta; es una tendencia o inclinación que distorsiona la verdad de la representatividad de los datos, llevándolos a ser no objetivos o inexactos. Los sesgos pueden ocurrir en cualquier etapa del proceso de la encuesta: desde el diseño hasta la recolección y análisis de los datos.

Según [11] las dos fuentes más importantes de sesgo se resumen de la siguiente manera:

1. **Sesgo de selección.** Ocurre cuando parte de la población objetivo no está en el marco de muestreo o cuando el marco está incompleto y presenta deficiencias. En [1] se establece que este tipo de sesgo se presenta si:

    - La selección de la muestra depende de cierta característica asociada a las propiedades de interés. Por ejemplo, la frecuencia con la que los adolescentes hablan con los padres acerca del SIDA.

- La muestra se realiza mediante elección deliberada o mediante un juicio subjetivo. Por ejemplo, si el parámetro de interés es la cantidad promedo de gastos en compras en un centro comercial y el encuestador elige a las personas que salen con muchos paquetes, entonces la información estaría sesgada puesto que no está reflejado el comportamiento verdadero de las compras.
- Existen errores en la especificación de la población objetivo. Por ejemplo, en encuestas electorales, cuando la población objetivo contiene a presonas que no están registrafas como votantes ante la organización electoral de su país.
- Existe sustitución deliberada de unidades no disponibles en la muestra. Si, por alguna razón, no fue posible obtener la medición y consecuente observación de la característica de interés para algún individuo en la población, la sustitución de este elemento debe hacerse bajo estrictos procedimientos estadísticos y no debe ser subjetiva en ningún modo.
- Existe ausencia de respuesta. Este fenómeno puede causar distorsión de los resultados cuando los que no responden a la encuesta difieren críticamente de los que sí respondieron.
- La muestra está compuesta por respondientes voluntarios. Los foros radiales, las encuestas de televisión y los estudios de portales de internet no proporcionan información confiable.

2. **Sesgo de medición.** Se da cuando el instrumento con el que se realiza la medición tiene una tendencia a diferir del valor verdadero que se desea averiguar. Éste sesgo debe ser considerado y minimizado en la etapa de diseño de la encuesta. Nótese que ningún análisis estadístico puede revelar que una pesa añadió a cada persona dos kilos de más en un estudio de salud. De igual manera, en [1] se encuentran algunas de las situaciones en las que se presenta el sesgo de medición:

   - Cuando el respondiente miente. Esta situación se presenta a menudo en encuestas que se pregunta acerca del ingreso salarial, alcoholismo y drogadicción, nivel socioeconómico e incluso edad.
   - Difícil comprensión de las preguntas. Por ejemplo: ¿No cree que no esté en buen momento para invertir? La doble negación en la pregunta es muy confusa para el respondiente.

- Las personas tienden a olvidar. Es bien sabido que las malas experiencias suelen ser olvidadas; esta sutiación debe acotarse si se está trabajando en una encuesta de criminalidad.
- Distintas respuestas a distintos entrevistadores. En algunas regiones es muy probable que la raza, edad o género del encuestador afecte directamente la respuesta del entrevistado.
- Leer mal las preguntas o polemizar con el respondiente. El encuestador puede influir notablemente en las respuestas. Por lo anterior, es muy importante que el proceso de entrenamiento del entrevistador sea riguroso y completo.

El siguiente ejemplo ilustra cómo es que los sesgos afectan a una encuesta.

**2.3 Ejemplo.** Durante las elecciones generales de 2015 en Guatemala, se llevaron a cabo varias encuestas preelectorales para predecir los resultados de la primera vuelta presidencial. Sin embargo, los resultados de las encuestas no coincidieron con los resultados finales, generando críticas y cuestionamientos sobre la metodología empleada. Entre los problemas que se identificaron, tenemos:

1. **Muestra no representativa**
   - *Subrepresentación*[3] *de votantes rurales:* Era recurrente que las encuestas no incluían adecuadamente a los votantes en áreas rurales, que representan una parte significativa del electorado en Guatemala. Esto llevó a un sesgo en los resultados hacia las preferencias de los votantes urbanos.
   - *Dificultades logísticas:* La geografía y el acceso limitado a algunas regiones hicieron que las encuestas no pudieran captar completamente las opiniones de todas las áreas del país.
2. **Tasa de respuesta baja**
   - *Desconfianza en las encuestas:* Muchos cuidadanos mostraron desconfianza hacia las encuestas, lo que resultó en tasas de respuesta bajas y potencialmente sesgadas.
   - *Acceso a tecnología:* En algunas encuestas que utilizaron métodos en línea o telefónicos, hubo subrepresentación de personas sin acceso a tecnología, como teléfonos o internet.

[3] explicar qué significa esto

3. **Errores en la proyección de participación**
   *Participación inesperada:* Las encuestas no lograron capturar cambios en la participación electoral, como el aumento en el número de votantes jóvenes o de primer voto, que influyeron en los resultados finales.

4. **Formulación de preguntas**
   *Influencias en las respuestas:* Las preguntas utilizadas en algunas encuestas podrían haber sido percibidas como sesgadas o influyentes, lo que podría afectar la sinceridad de las respuestas.

# 3. Selección de la muestra

El cálculo del tamaño de muestra es un elemento fundamental en la planificación y ejecución de encuestas, ya que determina la precisión y la representatividad de los resultados obtenidos. En este capítulo se aboradará el proceso de determinar el tamaño de muestra óptimo basándose en la Teoría de Probabilidad. Este enfoque permite asegurar que la muestra seleccionada sea suficientemente grande para reflejar las características de la población, pero también lo suficientemente pequeña para ser práctica y costo-efectiva.

## 3.1. Cálculo del tamaño de muestra

Vamos a tratar de abordar el problema de *determinar el valor de alguna cantidad o indicador de la población objetivo (dichos valores pueden ser: ingreso total, cantidad total de personas, proporción de personas con cierta característica, índice total de alguna característica de las personas, etc)*. El camino seguro pero exhaustivo sería el de recabar información de toda la población para poder conocer el indicador deseado, pero, desde ya, esto tiene demasiadas desventajas: económicas, logísticas, tiempo, errores de cobertura, calidad de los datos. Por lo tanto, lo ideal es realizar la recabación de información en solamente un pequeño porcentaje de la población para lograr determinar el valor del indicador en cuestión.

Pretender usar una cantidad limitada de información para determinar con precisión un valor exacto a nivel poblacional es poco realista. Pero es razonable utilizar esa información para obtener una idea general de la situación. Por esto, en lugar de buscar un valor exacto, nos enfocamos en determinar un rango que contenga el valor del indicador, algo como "La cantidad total de personas en el país está entre 15 y 20 millones". Idealmente, quisiéramos que este rango sea lo más estrecho posible para tener una estimación más concreta del valor real del indicador.

Sin embargo, para hacerlo correctamente debemos superar obstáculos clave como asegurar que la cantidad de información disponible sea suficiente, que esta información refleje fielmente la realidad de la población, y encontrar una manera de evaluar la calidad de la estrategia empleada. Por esta razón utilizamos la Teoría la Probabilidad como marco metodológico, ya que nos proporciona las herramientas necesarias para realizar este ejercicio de manera fundamentada, considerando toda la lógica mencionada.

### 3.1.1. ¿Cómo se define el tamaño de muestra de una encuesta a partir de la Teoría de probabilidad?

Sea $s$ una muestra de $n \in \mathbb{N}$ individuos que pertenecen a una población $X = \{x_1, x_2, \ldots, x_N\}$, y sea $\mathcal{S}$ la colección de todas esas posibles muestras $s$. Cada individuo $x_i$ tiene asociado un número real $y_i$ que captura alguna cantidad ligada al parámetro poblacional $\theta$, por ejemplo, el ingreso del individuo que contribuye al ingreso promedio de la población

$$\frac{1}{N}\sum_{i=1}^{N} y_i. \tag{3.1}$$

Consideremos una función $\hat{\theta}\colon \mathcal{S} \to \mathbb{R}$ que asigna a cada posible muestra $s \in \mathcal{S}$ un valor real, que representará la información pertinente para el objetivo de estimar al parámetro poblacional y que se puede obtener de la medición de la muestra seleccionada. Por ejemplo, si el objetivo es ubicar el ingreso promedio de $X$, una función natural a considerar como punto de partida es el ingreso promedio de la muestra:

$$\hat{\theta}(s) = \frac{1}{n}\sum_{x_i \in s} y_i \tag{3.2}$$

donde $y_i$ corresponde al ingreso del individuo $x_i$.

La función es "natural" en el sentido de que emula en la muestra lo que se desea saber de la población, pero a parte de ello no tiene por qué ser necesariamente una función útil para ubicar a (3.1), puede que la muestra $s$ seleccionada tenga características muy singulares respecto al resto de la población y $\hat{\theta}(s)$ no esté ni cerca del valor poblacional. Sin embargo, si consideramos esta función en el contexto más general de todas las posibles muestras $s$ que se pueden obtener, al promediar todos los resultados de calcular $\hat{\theta}(s)$ se llega a que

$$\begin{aligned}
\frac{1}{|\mathcal{S}|}\sum_{s\in\mathcal{S}}\hat{\theta}(s) &= \frac{1}{|\mathcal{S}|}\sum_{s\in\mathcal{S}}\frac{1}{n}\sum_{x_i\in S} y_i \\
&= \frac{1}{|\mathcal{S}|}\sum_{i=1}^{N}\frac{y_i}{n}\cdot|\{s\in\mathcal{S}\colon x_i\in s\}| \\
&= \frac{1}{\binom{N}{n}}\sum_{i=1}^{N}\frac{y_i}{n}\binom{N-1}{n-1} \\
&= \frac{1}{N}\sum_{i=1}^{N} y_i.
\end{aligned} \tag{3.3}$$

Es decir, el promedio de los $\hat{\theta}(s)$ es exactamente igual al parámetro deseado. A primera vista esto quizás pueda parecer, como mucho, un resultado bonito e interesante pero que carece de utilidad práctica. Después de todo, disponer de $\hat{\theta}(s)$ para toda posible muestra es equivalente a tener la información de toda la población, así que hemos vuelto al punto de partida. Sin embargo, intuitivamente el promedio de ciertas cantidades se interpreta como esa medida "resumen" o "punto medio" del contexto general de esas cantidades y esta noción se potencia con la Teoría de Probabilidad. Si $\mathcal{S}$ es un espacio de probabilidad, es decir, cada $s\in\mathcal{S}$ tiene asignada una probabilidad $p(s)$ de ser la muestra seleccionada, entonces $\hat{\theta}$ es una variable aleatoria. Si además, cada $s$ tiene la misma probabilidad de ser la muestra

$$p(s) = \frac{1}{|\mathcal{S}|}$$

entonces el cálculo que acabamos de hacer nos dice que $\theta$ es un estimador insesgado del parámetro poblacional, es decir

$$\begin{aligned}
\mathrm{E}[\hat{\theta}] &= \sum_{s\in\mathcal{S}}\frac{1}{|\mathcal{S}|}\theta(s) \\
&= \frac{1}{N}\sum_{i=1}^{N} y_i \\
&= \theta.
\end{aligned}$$

Luego, por el Teorema Central del Límite, es posible ubicar la esperanza del estimador utilizando un intervalo de confianza. En resumen, *por medio de un estimador insesgado $\hat{\theta}$ del parámetro poblacional $\theta$, existe una buena forma de ubicar a $\theta$ y* *todas las tareas alrededor de la muestra estarán íntimamente ligadas a esto.*

En las encuestas se considera un tamaño fijo de elementos de una población para pertenecer a una muestra $s$ sobre la cual se recopilará información. Una de las primeras tareas al diseñar una encuesta consiste en definir ese tamaño de elementos de la población, esto se hace al **ubicar el parámetro de interés** definido por el objetivo principal de la encuesta.

Ubicar el parámetro $\theta$ consiste en construir un intervalo de confianza para $\theta$ con un nivel de significancia $\alpha$ y una amplitud $A$. Tanto $\alpha$ como $A$ son cantidades que pueden depender del criterio de las personas interesadas en los resultados de la encuesta puesto que son quienes tienen, o deberían tener, una noción adecuada de las dimensiones de las cantidades recopiladas, además de las aplicaciones que se desprenden de los resultados de la encuesta, por ejemplo: la elaboración de Políticas Públicas o los análisis socioeconómicos (Manual de Encuestas Sobre Hogares).

Por el Teorema Central del Límite tenemos que para un tamaño de muestra $|s| = n$ suficientemente grande en una población grande y un estimador insesgado $\hat{\theta}$ de $\theta$ se cumple que

$$\mathrm{P}\left(\left|\hat{\theta} - \theta\right| \leq c\sqrt{\mathrm{V}(\hat{\theta})}\right) \approx \phi(c), \qquad c \in \mathbb{R}$$

donde $\phi$ es la función de distribución de una variable aleatoria normal estándar (en otras palabras, $\phi(c)$ es un valor conocido). Esto implica que el intervalo

$$I = \left[\hat{\theta} - c\sqrt{\mathrm{V}(\hat{\theta})}, \hat{\theta} + c\sqrt{\mathrm{V}(\hat{\theta})}\right] \tag{3.4}$$

contiene al parámetro $\theta$ con probabilidad $\phi(c)$. Como podemos ver en (3.4), la amplitud $A$ de $I$ depende de $c$, que está ligado al nivel de significancia que usualmente se establece en un 95 %, y de la varianza $\mathrm{V}(\hat{\theta})$ del estimador.

Supongamos que la varianza del estimador es una función decreciente del tamaño de muestra $n$, es decir $\mathrm{V}(\hat{\theta}) = h(n)$ con $n \colon \mathbb{N} \to \mathbb{R}$ tal que $a < b \implies h(a) > h(b)$. Si se busca que la amplitud de $I$ no supere un umbral $A_0$, entonces

$$\begin{aligned} 2c\sqrt{\mathrm{V}(\hat{\theta})} \leq A_0 \iff \\ h(n) \leq \left(\frac{A_0}{2c}\right)^2 \iff \\ n \geq h^{-1}\left(\left(\frac{A_0}{2c}\right)^2\right). \end{aligned}$$

Es decir, podemos hallar el tamaño "óptimo" de muestra tomando el $n$ más pequeño que satisface la desigualdad anterior. La forma que toma $\mathrm{V}(\hat{\theta})$ dependerá de la estructura probabilística de $\mathcal{S}$, así como del estimador $\hat{\theta}$.

## 3.2. Diseño muestral

Las encuestas tienen como meta principal obtener información precisa y confiable de una población objetivo a partir del estudio de una fracción relativamente pequeña de esta. La premisa básica de este enfoque radica en el Principio de Representatividad y la posibilidad de estimar parámetros de la población a través del marco teórico robusto de la Teoría de Probabilidad. Como vimos anteriormente, la precisión de la estimación depende de la estructura probabilística subyacente, conocida como el diseño muestral y durante esta sección exploraremos los principales diseños muestrales que cimentan las encuestas de hogares.

### 3.2.1. ¿Qué es un diseño muestral?

Un diseño muestral es un conjunto de procedimientos y estrategias que definen la forma en que se seleccionará una muestra a partir de una población objetivo. En términos prácticos, es el proceso que garantiza que cada unidad de la población tenga una probabilidad conocida y no nula de ser incluida en la muestra. Esto permite que las estimaciones derivadas sean represenativas y precisas, evitando sesgos y proporcionando bases sólidas para realizar inferencias estadísticas.

El diseño muestral se construye a partir de principios fundamentales como la aleatorización y la inclusión. Sí, aquí falta conectar, pero aun no se me ocurre cómo jeje...

#### 3.2.1.1. Muestra aleatoria

En el capítulo 2 ya se habló algo respecto a la definición de muestra, veremos ahora una manera de representar matemáticamente a este subconjunto. Denotaremos con la letra mayúscula $S$ a la muestra aleatoria y con una letra minúscula $s$ a una selección de la misma y será el conjunto de unidades pertenecientes a

$$s = \{1, \ldots, k, \ldots, n(S)\}.$$

El número de componenetes de $s$ es llamado **tamaño de la muestra** y en algunos casos $n(S)$ es una cantidad aleatoria.

**3.1 Definición.** Una *muestra sin reemplazo* se denota mediante un vector columna

$$s = (I_1, I_2, \ldots, I_N)' \in \{0,1\}^N$$

donde

$$I_k = \begin{cases} 1 & \text{si el } k\text{-ésimo elemento pertenece a la muestra,} \\ 0 & \text{en otro caso.} \end{cases}$$

Si la inclusión de cada uno de los elementos se hace entre los elementos que no han sido escogidos aún, estamos hablando de una muestra aleatoria sin reemplazo, así, el conjunto $s$ nunca tendrá elementos repetidos. Ya que $n(S)$ no es una cantidad fija, es posible que ocurran los siguientes escenarios:

1. Que la muestra no contenga ningún elemento, por lo que se dice que la muestra es vacía.

2. Que la muestra contenga a todos los elementos de la población. A esta muestra se le conoce como *censo*.

**3.2 Definición.** Una *muestra con reeemplazo* se denota mediante un vector columna

$$s = (n_1, n_2, \ldots, n_N)' \in \mathbb{N}^N$$

en donde $n_k$ es el número de veces que el elemento $k$ está en la muestra.

El número de elementos distintos en una muestra aleatoria $S$ con reemplazo es llamado *tamaño de muestra efectivo*.

En las secciones siguientes de este capítulo se desarrollarán las estrategias de muestreo específicas que se ajustan a ciertas situaciones y estimadores que mejoran la eficiencia de la estrategia. Según (Gutiérrez 2015) las estrategias de muestreo se definen en términos del tipo de muestreo que se utiliza para la selección de muestras y, en general, existen dos distinciones básicas:

1. Tipo de muestreo: La selección de unidades es con reemplazo o sin reemplazo.

2. Tamaño de muestra: Fijo o aleatorio.

Para continuar, es necesario tomar en cuenta las siguientes definiciones.

**3.3 Definición.** Un *soporte* $Q$ es un conjunto de muestras.

**3.4 Definición.** Decimos que un soporte es *simétrico* si para cualquier $s \in Q$, todas las permutaciones de $s$ están también en $Q$.

**3.5 Ejemplo.** A continuación tenemos algunos ejemplos de soportes simétricos.

1. El soporte simétrico sin reemplazo definido como

   $$S = \{0,1\}^N$$

   Notemos que $|S| = 2^N$. De modo que si $N = 3$, $S$ se define por las siguientes muestras:

   $$S = \{(0,0,0),(1,0,0),(0,1,0),(0,0,1),(1,1,0),(1,0,1),(0,1,1),(1,1,1)\}.$$

2. El soporte simétrico sin reemplazo de tamaño fijo definido como

   $$S_n = \left\{ s \in S : \sum_{k \in X} s_k = n \right\}.$$

   Como sabemos que el número de formas de elegir $n$ elementos de un conjunto de $N$, es $\binom{N}{n}$, entonces
   $$|S_n| = \binom{N}{n}.$$

3. No sé si colocar más ejemplos o si valdrá la pena desarrollar un poco más de teoría de soportes, lo que coloqué, me sirvió pa' entederlos...

La mayoría de las encuestas de hogares se fundamentan en el uso de una *muestra probabilística*. Este enfoque garantiza la posibilidad de realizar inferencias válidas y generalizables sobre las características de la población de hogares o de las personas que los conforman, a partir de los datos recolectados en el estudio.

**3.6 Definición.** Una muestra (con o sin reemplazo) se dice que es *probabilística* si:

1. Es posible definir teóricamente un soporte $Q$ tal que $Q = \{s_1, s_2, \ldots, s_q, s_{q+1}, \ldots, s_Q\}$, de todas las muestras posibles obtenidas por un método de selección, en donde $s_q \in Q$.

2. Las probabilidades de selección que el proceso aleatorio le otorga a cada posible muestra en el soporte, son conocidas de antemano a la selección de la muestra final.

**3.7 Definición.** Dado un soporte $Q$, un *diseño de muestreo* $p()$ es una distribución de probabilidad tal que $p\colon Q \to (0,1]$ con la que para todo $s \in Q$ se tiene que $p(s) > 0$ y

$$\sum_{s\in Q} p(s) = 1.$$

### 3.2.2. Muestreo Aleatorio Simple

El diseño muestral más sencillo de todos, sobre el cual toda la teoría parte, es el Muestreo Aleatorio Simple. La definición consiste en considerar cada una de las $s \in \mathcal{S}$ como igual de probables, es decir:

$$p(s) = \frac{1}{|\mathcal{S}|}. \tag{3.5}$$

Bajo este diseño, el estimador natural de un promedio poblacional resulta tener varianza

$$\mathrm{V}_{\mathrm{MAS}}(\hat{\theta}) = \frac{S^2}{n}\left(1 - \frac{n}{N}\right)$$

donde $S^2$ y $N$ son constantes de la población, y $n$ es el tamaño de la muestra. Se puede verificar que esta es una función decreciente en $n$, así que por lo descrito anteriormente existe una forma sistemática en la cual uno define el tamaño de muestra para una encuesta bajo este diseño:

$$n = \frac{S^2}{\left(\frac{A_0}{2}\right)^2 \cdot \frac{1}{c^2} + \frac{S^2}{N}} \tag{3.6}$$

Al hacer $A_0/2 = e$, $c = z_{\alpha/2}$ y suponiendo que $N$ es mucho más grande que $S$ se llega a la fórmula que se encuentra en los libros de texto:

$$n = \left(\frac{z_{\alpha/2} S}{e}\right)^2.$$

### 3.2.3. Muestreo estratificado

El muestreo estratificado es una técnica de muestreo que mejora la precisión de las estimaciones al dividir la población en subgrupos homogéneos llamados **estratos**. Estos estratos, que a menudo corresponden a grandes unidades geográficas, permiten garantizar una representación adecuada de cada grupo en la muestra, facilitando así inferencias más precisas sobre subpoblaciones específicas de interés.

Dividimos la población de $N$ en $H$ estratos, con $N_h$ unidades de muestreo en el estrato $h$. Para que el muestreo estratificado funcione, debemos conocer los valores de $N_1, N_2, \ldots, N_H$ y debe cumplirse que

$$N_1 + N_2 + \cdots + N_H = N$$

donde $N$ es el número total de unidades en toda la población. Cada estrato es tratado como una subpoblación independiente, y se selecciona una muestra de tamaño $n_h$ dentro de cada estrato, asegurando que

$$n_1 + n_2 + \cdots + n_H = n.$$

La asignación de la muestra a los estratos puede hacerse de distintas maneras, dependiendo de los objetivos de la encuesta y de la variabilidad dentro de cada estrato.

1. **Asignación simple o uniforme.** Cada estrato recibe el mismo tamaño de muestra sin importar su tamaño en la población:

   $$n_h = \frac{n}{H}$$

   Este método es poco eficiente si los estratos tienen tamaños mu distintos o varianzas diferentes.

2. **Asignación proporcional.** Se asigna el tamaño de muestra a cada estrato de acuerdo con su peso en la población

   $$n_h = n \cdot \frac{N_h}{N}.$$

   Este método es útil cuando los estratos tienen tamaños muy diferentes pero similar variabilidad.

3. **Asignación óptima (Neyman).** Si la variabilidad dentro de los estratos es diferente, se usa la asignación óptima de Neyman, que minimiza la varianza del estimador global

   $$n_h = n \cdot \frac{N_h S_h}{\sum_{h=1}^{H} N_h S_h},$$

   donde $S_h$ es la desviación estándar de la variable de interés dentro del estrato $h$. Esta asignación es más eficiente porque asigna más muestra a los estratos con mayor variabilidad, mejorando la precisión del estimador.

La estimación del promedio poblacional $\bar{Y}$ en un diseño estratificado se calcula como

$$\hat{Y}_s = \sum_{h=1}^{H} W_h \hat{Y}_h,$$

donde:

- $W_h = N_h/N$ es el peso del estrato en la población,
- $\bar{Y}_h$ es la media muestral del estrato h.

La varianza del estimador es:

$$\mathrm{V}(\hat{Y}_s) = \sum_{h=1}^{H} W_h^2 \frac{S_h^2}{n_h}\left(1 - \frac{n_h}{N_h}\right).$$

Si la asignación de la muestra se hace con la regla de Neyman, se minimiza esta varianza, logrando una estimación más eficiente.

El tamaño de muestra total $n$ para un muestreo estratificado se puede determinar a partir de la varianza deseada y del margen de error permitido en la estimación. Si el objetivo es estimad una media con error estándar máximo $\epsilon$, el tamaño de muestra se obtiene resolviendo la ecuación

$$n = \frac{\sum_{h=1}^{H} W_h^2 S_h^2}{\epsilon^2/z^2}. \tag{3.7}$$

Si se estima una proporción poblacional P, el tamaño de muestra se puede aproximar por

$$n = \frac{\sum_{h=1}^{H} W_h^2 P_h(1 - P_h)}{\epsilon^2/z^2}. \tag{3.8}$$

Donde:

- $z$ es el valor crítico de la distribución normal estándar según el nivel de confianza deseado (por ejemplo, 1.96 para 95 %),
- $\epsilon$ es el margen de error máximo permitido.

### 3.2.4. Encuesta Multi-etápica

Las encuestas multi-etápicas son un tipo de diseño muestral utilizado en la investigación estadística y de encuestas. En este enfoque, la selección de la muestra se realiza en varias etapas, en lugar de seleccionar directamente a los individuos de la población objetivo. Este método es particularmente útil cuando se trata de poblaciones grandes y dispersas, donde sería costoso o logísticamente difícil realizar un muestreo directo. En general. en América Latina son muy comunes los diseños de selección en dos etapas:

- **Primera etapa.** Se seleccionan unidades grandes, llamadas Unidades Primarias de Muestreo (UPM), que pueden ser regiones geográficas, bloques o conglomerados.

- **Segunda etapa.** Dentro de cada UPM se seleccionan unidades más pequeñas como viviendas u hogares.

También es posible encontrar en algunos países diseños divididos en más de dos etapas. Por ejemplo, en una primera etapa se seleccionan municipios, en una segunda etapa se seleccionan UPM dentro de los municipios seleccionados y, en la tercera, se selecciona una muestra de hogares en aquellas UPM seleccionadas en la segunda etapa pertenecientes a los municipios seleccionados en la primera etapa de muestreo.

### 3.2.5. Características de las Encuestas Multi-etápicas

Según [10], el muestreo multi-etápico presenta dos características fundamentales que lo hacen estadísticamente robusto y eficiente en la planificación logística del proceso de recopilación de datos

- **La independencia,** que implica que no hay ninguna correlación en el diseño de muestreo de las UPM. Esto quiere decir que en cada UPM se puede ejecutar con independencia cualquier estrategia de muestreo que se considere apropiada para seleccionar la submuestra de hogares.

- **La varianza,** que implica que sin importar qué diseño de muestreo se haya ejecutado en la primera etapa para seleccionar las UPM, la segunda etapa de selección podrá ejecutarse de manera independiente a la primera. Es decir, el submuestreo de los hogares es independiente del muestreo de las UPM.

## 3.3. ¿Qué es el DEFF?

El efecto de diseño (DEFF, por sus siglas en inglés “Design Effect”) es una medida que cuantifica cuánto más grande debe ser una muestra obtenida con un diseño muestral complejo (como el estratificado, por conglomerados, o una combinación de ambos) en comparación con una muestra obtenida mediante muestreo aleatorio simple, para lograr la misma precisión en las estimaciones.

Se define como

$$\mathrm{DEFF} = \frac{\mathrm{V}(\hat{\theta}_{\text{complejo}})}{\mathrm{V}(\hat{\theta}_{\text{MAS}})}, \tag{3.9}$$

donde:

- $\mathrm{V}(\hat{\theta}_{\text{complejo}})$ es la varianza del estimador en el diseño muestral utilizado (estratificado, por conglomerados, etc.).
- $\mathrm{V}(\hat{\theta}_{\text{MAS}})$ es la varianza del estimador si se hubiera usado un muestreo aleatorio simple.

Un DEFF $> 1$ indica que el diseño muestral utilizado requiere una muestr mayor para alcanzar la misma precisión que el MAS, muestras que un DEFF $< 1$ indica que el diseño muestral es más eficiente que el MAS.

### 3.3.1. Cálculo del DEFF en muestreo por conglomerados

En el caso del muestreo por conglomerados, el DEFF tiende a ser mayor que 1 porque los individuos dentro de un conglomerado tienden a ser más similares entre sí que si fueran seleccionados aleatoriamente. En este caso, el DEFF se aproxima como:

$$\mathrm{DEFF} \approx 1 + (b-1)\rho \tag{3.10}$$

donde:

- $\rho$ es el coeficiente de correlación intraclase, que mide la similitud dentro de los conglomerados.
- $b$ es el tamaño promedio de los conglomerados.

Si la correlación intraclase $\rho$ es alta, significa que los individuos dentro de cada conglomerado son muy similares, lo que aumenta la varianza y hace que el DEFF sea mayor que 1, reduciendo la eficiencia del diseño. Cuando se usa un diseño muestral complejo, el tmaño de muestra necesario se ajusta multiplicándolo por el DEFF,

$$n_{\text{ajustado}} = n_{\text{MAS}\times\text{DEFF}}.$$

Donde:

- $n_{MAS}$ es el tamño de muestra que se habría usado con muestreo aleatorio simple.

- $n_{\text{ajustado}}$ es el tamaño de muestra corregido por el diseño muestral.

Si el DEFF $> 1$, se necesita un tamaño de muestra mayor para mantener la precisión deseada en las estimaciones.

## 3.4. ¿Cómo se calcula el tamaño de muestra en encuestas de hogares?

El cálculo del tamaño de muestra para la mayoría de encuestas de hogares inicia con el planteamiento de un diseño muestral adecuado y este comienza con la definición de los objetivos de la encuesta. Esto permite delimitar aspectos como: la población objetivo, los dominios de estudio, la variable de diseño y los estratos. Una vez fijados estos elementos se procede al cálculo del tamaño de la muestra necesario a tener, es decir, unidades informantes sobre las que hay que recolectar información. En encuestas de gran alcance, en las que se predende obtener información en varias etapas: primero se seleccionan sectores cartográficos (Unidades Primarias de Muestreo) y luego hogares dentro de estos sectores seleccionados (Unidades Secundarias de Muestreo).

El tamaño de muestra está intimamente asociado a la precisión que se desee obtener en la estimación del parámetro de interés (variable de diseño). Esta precisión se define tomando en consideración el contexto de la información a recopilar y las capacidades logísticas y presupuestarias disponibles. El proceso del cálculo de este tamaño debe ser realizado en cada dominio de estudio a considerar. Finalmente, una vez calculado el tamaño de muestra, se define una estrategia de afijación de la muestra, es decir, en cada dominio de estudio se distribuye el tamaño de muestra calculado en los estratos que conforman el dominio.

En particular, cuando se trabaja en R, existe un paquete muy util llamado *samplesize4surveys* el cual contiene una serie de funciones que consideran el tipo de la variable de diseño y las etapas del diseño muestral. Por lo tanto, resulta muy sencillo realizar una rutina en R para determinar el tamaño muestral en cada dominio de estudio para la encuesta en cuestión. Entonces, lo principal es establecer

## 3.5. Algoritmos de selección

Los algoritmos de selección son la base operativa de los diseños muestrales, permitiendo la implementación sistemática de distintos métodos para elegir unidades de análisis dentro de una población. Su función principal es garantizar que la selección de la muestra se realice de manera precisa, objetiva y conforme a los principios probabilísticos que rigen la inferencia estadística.

Existen diversos algoritmos de selección, cada uno asociado a un diseño muestral específico. Los más comunes son: el de *muestreo aleatorio simple*, el *muestreo estratificado*, el *muestreo por conglomerados*, y el *muestreo sistempático*. En cada caso, los algoritmos determinan la forma en que se eligen las unidades, asegurando que la muestra cumpla con los requisitos de aleatorización y representatividad.

El diseño de un algoritmo de selcción depende de varios factores, como la estructura de la población, la disponibilidad de un marco muestral y los requerimientos del estudio. Mientras que un MAS puede implmentarse mediante un procedimiento sencillo de selección uniforme, un muestreo estratificado requiere dividir la población en subgrupos y aplicar la selección dentro de cada uno de ellos. Por otro lado, en el muestreo por conglomerados, las unidades se seleccionan en grupos predefinidos, en el muestreo sistemático, la selección sigue un patrón regular basado en intervalo fijo.

A continuación, presentariemos los principales algoritmos de selección, detallando sus procedimientos:

### 3.5.1. Muestreo aleatorio simple

Es la forma más sencilla de muestreo probabilístico. Se obtiene un MAS de tamaño $n$ cuando cada posible subconjunto de $n$ unidades de la población tiene la misma probabilidad de ser seleccionado como muestra. Al seleccionar una muestra aleatoria, el investigador está mezclando la población ante de elegir $n$ unidades. No es necesario examinar cada miembro de la población por la misma razón que un ténico de laboratorio no necesita extraer toda la sanbre de una persona para medir su conteo de glóbulos rojos, pues, la sangre está suficientemente bien mezclada, por lo que cualquier muestra debe ser rerpresentativa.

### 3.5.2. Muestreo estratificado

### 3.5.3. Muestreo por conglomerados

### 3.5.4. Muestreo sistemático

# 4. Factores de expansión

# CONCLUSIONES

1. Conclusión 1.
2. Conclusión 2.
3. Conclusión 3.

# RECOMENDACIONES

1. Recomendación 1.
2. Recomendación 2.
3. Recomendación 3.

# BIBLIOGRAFÍA

[1] S. Lohr. *Sampling: Design And Analysis.* 2.ª ed. CRC press/Chapman & Hall/Taylor & Francis Group, Nueva York, 2019.

[2] R. De la Madrid. The rigged Hilbert space of the free hamiltonian. Consultado en marzo de 2005 en `http://arxiv.org/abs/quant-ph/0210167`.

[3] J. Escamilla-Castillo. *Topología.* 2.ª ed. s.e., Guatemala, 1992.

[4] N. Haaser y J. Sullivan. *Análisis real.* Tr. Ricardo Vinós. Trillas, México, 1978.

[5] P. Halmos. *Teoría intuitiva de los conjuntos.* 8.ª ed. Tr. Antonio Martín. Compañía Editorial Continental, S.A., México, 1973.

[6] F. Kronz. Quantum theory: von Neumann versus Dirac. Consultado en marzo de 2005 en `http://plato.stanford.edu/entries/qt-nvd/`.

[7] K. Liu, X. Sun, and S.-T. Yau. Goodness of canonical metrics on the moduli space of Riemann surfaces. *Pure Appl. Math. Q.*, **10**(2):223–243, 2014

[8] E. Leader and C. Lorcé, The angular momentum controversy: What's it all about and does it matter?, *Phys. Rept.* **541**, 163 (2014).

[9] S. Sternberg. Theory of functions of a real variable. Consultado en abril de 2005 en `http://www.math.harvard.edu/~shlomo`.

[10] Comisión Económica para América Latina y el Caribe (CEPAL), *Diseño y análisis estadístico de las encuestas de hogares de América Latina*, Metodologías de la CEPAL, N° 5 (LC/PUB.2023/14-P), Santiago, 2023.

[11] A. Gutiérrez, *Estrategias de muestreo: diseño de encuestas y estimación de parámetros*, Ediciónes de la U, Bogotá, 2016.