

Introducción al Diseño de Encuestas de Hogares

Curso de Muestreo Probabilístico en Encuestas de Hogares

Andrés Gutiérrez, Ph.D.

CEPAL - Unidad de Estadísticas Sociales

- 1 La planeación de una encuesta
- 2 El ciclo de vida de una encuesta

Motivación

Durante todo el siglo pasado, ha surgido una serie de teorías y principios que ofrecen un marco de referencia unificado en el diseño, implementación y evaluación de encuestas. Este marco de referencia se conoce comúnmente como el paradigma del “error total de muestreo” y ha encaminado la investigación moderna hacia una mejor calidad de las encuestas.

Robert M. Groves (2004)

Bibliografía y referencias

- Bradburn et. al. (2004) *Asking Questions: The Definitive Guide to Questionnaire Design*. Wiley.
- Groves, et. al. (2009) *Survey Methodology*. Wiley.
- Gutiérrez, H. A. (2016) *Estrategias de muestreo: diseño de encuestas y estimación de parámetros*. Ediciones de la U.

La planeación de una encuesta

¿Qué es una encuesta?



Figura1: Cuestionario de una encuesta

Definición de Wikipedia

Una encuesta es un estudio observacional en el cual el investigador busca recaudar datos por medio de un cuestionario pre-diseñado, y no modificar el entorno ni controlar el proceso que está en observación (como sí lo hace en un experimento).

Definición de Wikipedia

Los datos se obtienen a partir de realizar un conjunto de preguntas normalizadas dirigidas a una muestra representativa o al conjunto total de la población estadística en estudio, formada a menudo por personas, empresas o entes institucionales, con el fin de conocer estados de opinión, características o hechos específicos.

Preguntas

- ¿Es posible experimentar en una encuesta?
- ¿Todas las muestras son representativas en una encuesta?

Definición de Groves et. al. (2004):

Una encuesta es un método sistemático para obtener información de (una muestra de) entes, con el fin de construir descriptores cuantitativos de los atributos de una población más grande, de la cual los entes son miembros.

Preguntas

- ¿Qué quiere decir sistemático?
- ¿Por qué la frase “una muestra” está entre paréntesis?
- ¿Qué es un descriptor cuantitativo? ¿Puede dar algún ejemplo?

Indicadores

Ejemplos de un indicador descriptivo

- Media: promedio de años en educación.
- Proporciones: porcentaje de personas que votarán por un candidato.
- Totales: Total de personas víctimas del desplazamiento forzado.

Indicadores

Ejemplos de un indicador analítico

- Correlación: relación entre la cantidad de libros leídos y los años de escolaridad.
- Regresión: razón de incremento entre ingreso y años de experiencia

Características

Muchos procesos de recolección de información se hacen llamar encuestas. Para efectos de este curso, una encuesta tendrá las siguientes características:

- ① Los datos son recopilados mediante preguntas a personas.
- ② Las respuestas son compiladas cuando: a) un encuestador pregunta y graba las respuestas del entrevistado o b) el encuestado lee y graba sus propias respuestas.
- ③ Los datos son recolectados de un subgrupo de personas pertenecientes a la población de interés

Sondeo y encuesta

- **Sondeo:** es la traducción de *poll*, que a su vez viene del antiguo alemán referente a cabeza y se utilizaba para contar: “contar cabezas”.
- **Encuesta:** es la traducción de *survey*, que a su vez viene del latín *super* (sobre) y *videre* (observar).

Sondeo y encuesta

- En general, la primera expresión aparece muchas más veces en el sector privado, en estudios de opinión y de consumo. Un sondeo no será jamás utilizado para obtener estadísticas oficiales en estudios gubernamentales o en dominios científicos.
- Sin embargo, los sondeos muchas veces opacan la perspectiva científica de las cifras y pueden llevar a conclusiones inexactas acerca de la realidad de una problemática.
- ¡Es muy difícil no extrapolar los resultados de un sondeo!

Elecciones de 1948 en EE.UU.



Figura2: *El Chicago Daily Tribune* da como ganador a Dewey.

Elecciones de 1948 en EE.UU.



Figura3: *Truman sosteniendo el periódico que lo dio como perdedor.*

Tres grandes tópicos

Cuando el mundo occidental dejó de lado los grandes traumatismos del siglo XX, la investigación social tuvo un auge sobresaliente. Desde entonces, existen tres preguntas, en continua dinámica, que se deben responder para realizar una encuesta:

- ① ¿Cómo se diseñan las preguntas?
- ② ¿Cómo se selecciona la muestra?
- ③ ¿Cómo se recolectan los datos?

Inicio de los cuestionarios estandarizados

- La práctica de realizar las mismas preguntas en forma de cuestionario es reciente. En el principio cada encuestador preguntaba *lo mismo*, pero con diferentes palabras. Difícilmente, dos personas distintas eran entrevistadas con las mismas preguntas.
- Se encontró que la forma en cómo se preguntaba y cómo se recopilaba la información afectaba dramáticamente los resultados de las encuestas. Fue así como se decidió que los encuestadores debían ser entrenados (pre-operativo) formalmente.

Inicio de los métodos de muestreo

- En un principio, los investigadores trataban de recolectar datos sobre todos los elementos de la población de interés.
- Esta práctica resultaba logísticamente inadecuada cuando se trataba de poblaciones con un gran tamaño.
- Los cálculos de los indicadores sobre toda una población resultaban muy demandantes.

Inicio de los métodos de muestreo

- Aunque la teoría de la probabilidad tuvo sus orígenes en el siglo XVIII, no fue hasta la segunda década del siglo XX que se utilizó para realizar encuestas.
- La primera aplicación fue la selección sistemática de un elemento entre un conjunto de N .
- Para realizar esta selección, los registros censales se dividían en secciones y se procedía a seleccionar un elemento de la sección.

Inicio de los métodos de muestreo

- Más adelante, cuando la estadística permeó la agricultura, se definieron otros tipos de muestreo (menos demandantes) y se dio origen al muestreo de áreas.
- Es así como hoy en día es posible seleccionar muestras de bloques, manzanas, secciones y sectores cartográficos.

Inicio de los métodos de muestreo

- Se descubrió que era posible generalizar el muestreo de áreas y se creó el muestreo multietápico que permitió la selección de grandes bloques dentro de una ciudad, y áreas dentro de los bloques y el submuestreo sucesivo de unidades dentro hasta llegar a la unidad de interés.
- Todos estos submuestreos se realizan de forma probabilística.

Inicio de los métodos de muestreo

- La segunda guerra mundial y la gran depresión en EE.UU. fueron catalizadores de las encuestas a gran escala. En ese entonces, al igual que hoy, la tasa de desempleo era una cifra importante.
- Las políticas públicas empezaron a decidirse de acuerdo con las estadísticas oficiales.
- Las grandes encuestas se realizaron mensualmente.
- Hoy en día existen cientos de encuestas mensuales que dan cuenta de la realidad de las sociedades.

Inicio de la recolección de datos

- Debido a que en un principio no existía un cuestionario estandarizado, las respuestas abiertas eran la única opción de recopilar información.
- Esta práctica demandaba un gran esfuerzo en términos de resumir y sintetizar todo el corpus de palabras que los entrevistados usaban para responder.

Inicio de la recolección de datos

- En EE.UU. se dio un fenómeno en la mitad de los años 1960's: la proliferación de las entrevistas por correo.
- Los países con registros administrativos actualizados pueden contemplar este escenario puesto que induce una alta tasa de cobertura a precios más económicos (pues se prescinde del encuestador).
- Las bajas tasas de respuestas (pues el encuestado debe llenar un formulario con sus respuestas y devolverlo a la oficina postal) hicieron que paulatinamente esta forma de recolección no fuese tan apetecida.

Inicio de la recolección de datos

- Un camino intermedio entre las entrevistas cara a cara y los formularios auto-administrados por correo postal son las entrevistas telefónicas.
- Hoy en día, la mayoría de encuestas en investigación de medios y de mercado se realiza por teléfono.

El ciclo de vida de una encuesta

De ideas abstractas a acciones concretas

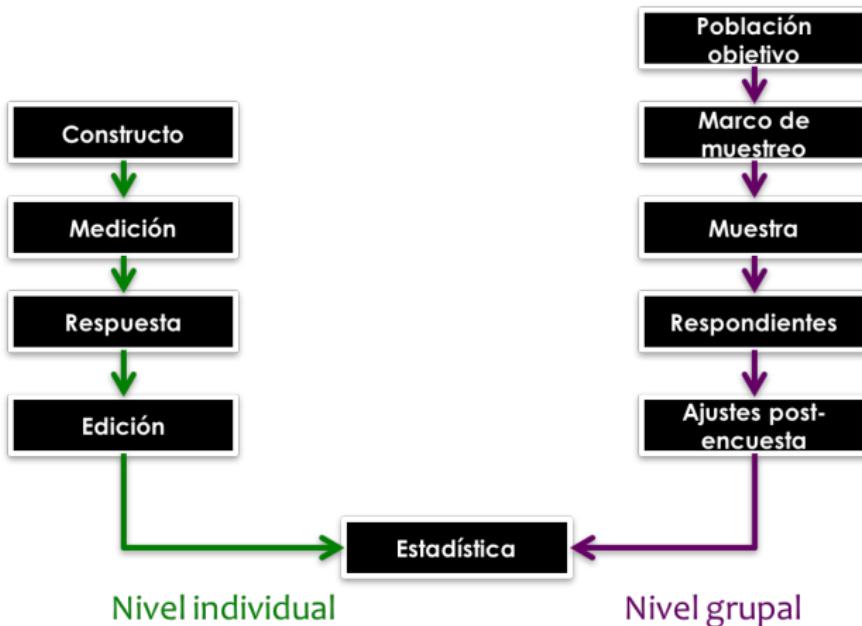


Figura4: *Modelo de Groves et. al. (2004)*

Constructos

- Son las ideas abstractas (ambiguas) sobre las cuales el investigador desea inferir.
- Los constructos dan origen a la investigación y son la simiente de la encuesta.
- Las palabras con que se describen los constructos son siempre simples, pero la redacción elaborada de los constructos no siempre es precisa.

Constructos

Una **encuesta de victimización** mide la cantidad de incidentes relacionados con crímenes en un año determinado. Pero:

- ¿Qué es un crimen?
- ¿Cómo se define a una víctima?
- Cuando se realiza un graffiti en un espacio público, ¿quién es la víctima?

Constructos

Una **encuesta de goce efectivo de derechos** sobre menores de edad mide la efectividad del estado al garantizar los derechos básicos a la primera infancia. Pero:

- ¿Qué es un derecho?
- ¿Cómo se define primera infancia?

Constructos

Algunos constructos son más abstractos que otros:

- Optimismo en la economía.
- Confianza inversionista.
- Percepción del Plan Nacional de Desarrollo.

Algunos constructos son observables más fácilmente:

- Consumo de alcohol y otras drogas.
- Nutrición en la primera infancia.
- Productividad de una intervención en el sector agrícola.
- Factores de riesgo asociados a una enfermedad.

Mediciones

- La medición es una caracterización mucho más concreta que el constructo. Es una forma de obtener información de los constructos de interés.

Mediciones

La cuestión clave para realizar una buena medición es realizar preguntas que induzcan respuestas que reflejen claramente los constructos que se desean medir.

Mediciones

Estas preguntas pueden ser:

- Comunicadas en forma oral (encuestas cara a cara o telefónicas).
- Comunicadas en forma visual (atributos de un producto - marketing).
- Observaciones directas del encuestador (condiciones de la vivienda).
- Electrónicas o físicas (precios de productos en supermercados, muestra de suelo en una hacienda).

Respuesta y edición

- El resultado de la medición es la respuesta.
- La naturaleza de las respuesta está determinada por la naturaleza de las preguntas.

Mirando hacia el futuro, ¿usted cree que de aquí a un año usted y su familia estarán financieramente mejor, peor o igual que ahora?

Respuesta y edición

- Después de que los entrevistados han respondido, los datos deben pasar por un proceso de edición y validación de inconsistencias.
- En este proceso de edición se debe examinar la distribución completa de las respuestas y buscar *datos atípicos* para que sean revisados con detenimiento.
- Los datos editados constituyen el insumo para realizar todo el proceso de inferencia estadística pertinente para que las cifras resultantes sean confiables y precisas.

De ideas abstractas a acciones concretas

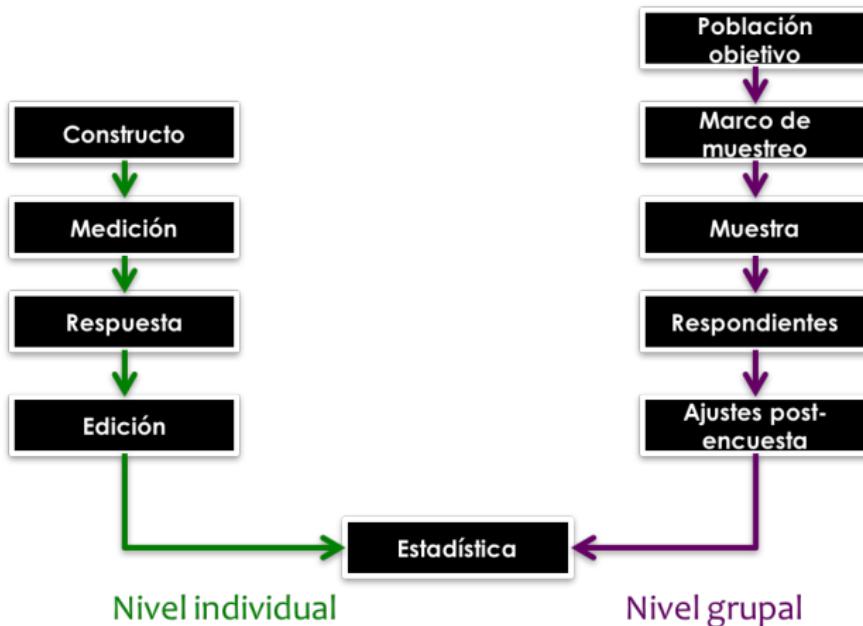


Figura5: *Modelo de Groves et. al. (2004)*

La población objetivo

- De las definiciones concernientes a agregados, esta es la más abstracta.
- En general **la población objetivo representa el conjunto de unidades que serán estudiadas.**

La población objetivo

En una encuesta es posible definir la población objetivo como, por ejemplo, *los adultos salvadoreños*. Sin embargo, esta definición de población:

- No contempla el periodo de referencia de la medición.
- No aclara si se incluyen los adultos salvadoreños residentes en el exterior.
- No precisa cómo se verificará la nacionalidad de un entrevistado.

La población enmarcada

No es posible realizar una encuesta probabilística sin un **marco de muestreo**, definido como: un dispositivo que permite *ubicar* e *identificar* (ambas acciones al mismo tiempo) las unidades pertenecientes a la población de interés.

La población enmarcada

Todos los marcos de muestreo presentan algún nivel de desactualización con respecto a la población de interés.

- Un marco de muestreo de líneas telefónicas puede no contener a todos los residentes de una ciudad.
- Un marco de muestreo de áreas, basado en cartografía, puede estar desactualizado

La población enmarcada

Nótese que con un marco de áreas es posible:

- Entrevistar a la misma persona en varias ocasiones (múltiples residencias).
- Nunca realizar la entrevista a una persona que no tienen un lugar fijo de residencia.

La población enmarcada

La población enmarcada está definida por el conjunto de miembros de la población de interés que efectivamente tienen chance no nula de ser seleccionados en una muestra probabilística.

La muestra

- El tamaño de muestra define directamente la precisión y confiabilidad de las estimaciones.
- El tamaño de muestra se incrementará a medida que lo hagan los niveles de desagregación (grupos etarios, regiones geográficas, niveles de escolaridad, etc.).
- La muestra es sólo una y no es posible replicarla varias veces en un mismo estudio.

La muestra

- Pueden existir escenarios en donde una encuesta con un tamaño de muestra menor induzca menores errores de muestreo que una encuesta con un mayor tamaño de muestra.
- En un censo el error de muestreo es nulo.

Los respondientes

- No siempre los esfuerzos realizados para que los individuos seleccionados en la muestra respondan son fructíferos.
- Los individuos que son **efectivamente entrevistados** se denominan respondientes.
- Al complemento de este conjunto se les denomina **no respondientes**.

Los respondientes

- Pueden existir casos de no respondientes parciales (no respondientes de ítems), para los cuales debe existir un proceso de **decisión** en términos de su reemplazo.
- No todas las ausencias parciales son reemplazadas.

Causas de la ausencia de respuesta

Algunos de los factores que inciden en el aumento de la ausencia de respuesta pueden ser:

- *Contenido*: encuestas relacionadas con el uso de drogas, finanzas. Se puede acotar la tasa de respuesta si se ordenan las preguntas de manera adecuada.
- *Tiempo de la encuesta*: algunas temporadas arrojan tasas de no respuestas más altas que otras.
- *Encuestadores*: aplicar métodos estándar de mejoramiento de la calidad para aumentar la precisión y tasa de respuesta de los entrevistadores involucrados en el estudio.

Causas de la ausencia de respuesta

Algunos de los factores que inciden en el aumento de la ausencia de respuesta pueden ser:

- *Método de recolección*: las encuestas telefónicas y por correo tienen una tasa de respuesta menor que las entrevistas personales.
- *Diseño de cuestionario*: formulación de las preguntas.
- *Agobio*: encuestas demasiado largas que indisponen al respondiente.

Los ajustes post-encuesta

- Toda encuesta cuenta con personas que no quisieron responder y/o con un marco de muestreo que no cubre a toda la población.
- Es necesario reajustar los factores de expansión para evitar la sub-estimación de los parámetros de interés.

Los ajustes post-encuesta

Se utiliza una reponderación diferencial cuando es evidente que hay un patrón de ausencia de respuesta en algunos subgrupos de la población:

- Los desempleados no responden sistemáticamente.
- Las tasas de respuestas a nivel urbano son menores que las tasas de respuesta a nivel rural.

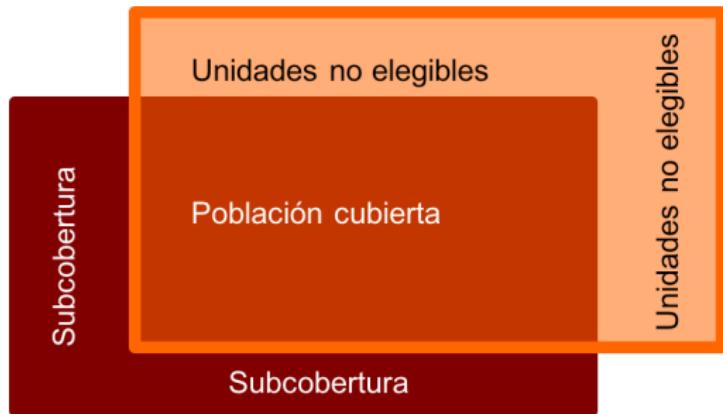
Los ajustes post-encuesta

Es necesario imputar¹ los valores perdidos cuando no se seleccionó un reemplazo.

- Es posible utilizar metodologías estocásticas complejas para imputar valores, o técnicas simples sistemáticas.
- En cualquier caso, siempre es preferible obtener la respuesta directa del entrevistado.
- ¡La imputación y la reponderación están enmarcados en los ajustes post-encuesta!

¹Imputar viene del inglés *input*.

Esquema de cobertura



Población objetivo

Población enmarcada

Figura6: *Población objetivo vs.población enmarcada.*

Caracterización de una encuesta de fuerza laboral

- **El propósito de la encuesta:** proporcionar información básica sobre el tamaño y estructura de la fuerza de trabajo (empleo, desempleo e inactividad) de la población del país.
- **Entidad que recopila la información:** Instituto/Oficina Nacional de Estadística.
- **Población objetivo:** población civil no institucional residente en el país.

Caracterización de una encuesta de fuerza laboral

- La PCNI contiene a todas las personas que no hacen parte de la fuerza pública y no pertenecen a instituciones de aislamiento como prisiones, hospitales, sanatorios, ancianatos, etc.
- La PCNI contiene a la población en edad de trabajar y a los no pertenecientes a la fuerza laboral. La edad para empezar a trabajar en un país puede variar (generalmente 15 años).
- A su vez, la PET contiene a Inactivos y Ocupados. La clasificación de ocupado es una variable derivada que está inducida por muchos filtros.

¡Gracias!

Andrés Gutiérrez

Experto Regional en Estadísticas Sociales

División de Estadísticas

Email: andres.GUTIERREZ@cepal.org