



# Muestreo y recolección de datos



# Población estudiada

En una investigación cuantitativa se deben recolectar datos. ¿De qué o quienes?

Con el planteamiento y los alcances de la investigación, se puede definir los participantes, objetos, sucesos o colectividades de estudio.

A estos se les llama **unidades de muestreo** (también casos o elementos), que en general coincide con las **unidades de análisis**.

Con esta definición, se procede a delimitar el **universo** o **población**: el conjunto de todos los casos que concuerdan con una serie de especificaciones.

Es importante que la población se describa explícitamente, indicando claramente sus características de contenido, lugar y tiempo.

--- Hernández, Fernández y Baptista (2014)



# Necesidad de muestreo

Normalmente es imposible acceder a la población completa en un estudio. Los censos son extremadamente caros.

A veces, la población estudiada es infinita (e.g. todas las instancias del problema del vendedor viajero) o difícil de identificar (e.g. personas con diabetes sin diagnosticar que sufren de cambios anormales de humor).

Por esto, la gran mayoría de las investigaciones cuantitativas utilizan (una o más) **muestras**.

Una muestra es un subgrupo de la población del cual se recolectan los datos.



# Muestras

Una muestra debe ser **estadísticamente representativa** de la población, con la intención de que los resultados encontrados en la muestra se **generalicen** o extrapolen a la población.

Las muestras se categorizan en dos tipos: las **muestras probabilísticas** (también llamadas muestras dirigidas) y las **muestras no probabilísticas**.

En el primer tipo, todos los elementos de la población tienen **la misma posibilidad** de ser elegidos.

En el segundo tipo, la elección de los elementos no depende de la probabilidad, sino de las **características** de la investigación.

--- Hernández, Fernández y Baptista (2014)



# Muestras probabilísticas

Para obtener una muestra probabilística son necesarios dos pasos:

1. **calcular un tamaño de muestra** que sea representativo de la población
2. **seleccionar los casos** de manera que al inicio todos tengan la misma posibilidad de ser elegidos.

Para el primero hay fórmulas y apoyo con software. Y ya conocemos algo:

- $n > 30$  suele ser suficiente para que el teorema del límite central aplique
- conocemos funciones de R que permiten calcular el tamaño de una muestra para una prueba t de Student o una prueba de proporciones

También existen para otras inferencias (ANOVA, regresión lineal, etc.)

--- Hernández, Fernández y Baptista (2014)



# Muestras probabilísticas

Pero este principio y estas funciones **solamente** son válidas si se usa **muestreo aleatorio simple** (MAS) desde **poblaciones infinitas**.

MAS es un procedimiento en que se seleccionan los casos **al azar** y todos los elementos de la población tienen la **misma probabilidad** de ser seleccionados.

Por ejemplo, para inferencia con proporciones, las estimaciones del tamaño de la muestra se basan en la siguiente fórmula:

$$n = \frac{s^2}{e^2 / z_{\alpha/2}^2}$$

--- Hernández, Fernández y Baptista (2014)



# Muestras probabilísticas

Pero si la población es finita, la fórmula **se corrige**:

$$n = \frac{s^2 N / (N - 1)}{e^2 / z_{\alpha/2}^2 + s^2 / (N - 1)}$$

“Finita” en este contexto significa de tamaño conocido, usualmente menor al millón, con un muestreo superior al 5% de la población.

Otros tipos de muestreo, tienen asociadas **otras fórmulas** para calcular el tamaño de la muestra requerida para hacer inferencias seguras.



# Muestras probabilísticas

Para el segundo paso se requiere un **marco muestral** y un **procedimiento de selección aleatoria**.

El marco muestral permite identificar y enumerar los elementos de la población. Son listados, archivos, mapas, entre otros.

El procedimiento de selección aleatoria debe asegurar que todos los elementos de la población tengan la **misma probabilidad de ser elegidos** (inicialmente).

Se puede usar un **patrón sistemático** o tómbolas (o simulaciones de estas con tablas de números aleatorios o generando números pseudoaleatorios)





# Muestras probabilísticas

Cuando toda la población está en un listado único, el procedimiento es simple y se obtiene una **muestra aleatoria simple**.

Pero con frecuencia la población y su marco muestral son más complejos. Si bien hay muchas combinaciones de muestreo de varias etapas, muy resumidamente se puede decir que existen dos grandes alternativas.

Por un lado están las **muestras probabilísticas estratificadas**: la población se divide en segmentos y se selecciona una submuestra de cada segmento.

Por el otro están las **muestras probabilísticas por racimos**: las unidades ya se encuentran encapsuladas en *clusters* distintos, que se seleccionan al azar.

--- Hernández, Fernández y Baptista (2014)



# Muestras probabilísticas

Ejemplo de muestras probabilísticas estratificadas: por comuna, por género, por Facultad, etc. Se espera observar **efectos distintos** en cada estrato.

Ejemplo de muestra probabilística por racimos (*clusters* o **conglomerados**): diferentes asignaturas de 1er año, hospitales de la macro zona centro-sur, escuelas municipales del sector poniente de Santiago, etc. Se espera que los racimos tengan **similar composición**.

En ocasiones se combinan tipos de muestreo aleatorios.



# Muestras no probabilísticas

La selección de una muestra no probabilística (o dirigida) no sigue un criterio estadístico de generalización, sino que uno orientado por las características de la investigación.

Así, seleccionan **casos típicos**, sin intentar que sean estadísticamente representativos de la población.

Aunque se usan en investigaciones cuantitativas, son más adecuadas y utilizadas en investigaciones cualitativas.

Algunas alternativas: de **voluntarios**, de **expertos**, de casos-tipo, por **cuotas**, en cadena (**bola de nieve**), de casos **extremos**, por **oportunidad** o **conveniencia**.

--- Hernández, Fernández y Baptista (2014)



# Muestras probabilísticas versus no probabilísticas

Las muestras probabilísticas una gran ventaja: **puede medirse el tamaño del error** en predicciones. El principal objetivo de una muestra probabilística es reducir al mínimo este error.

Esto no es posible con muestras no probabilísticas. A cambio, en estudios que no requieren tanto una “representatividad” de la población, permiten una **elección controlada de casos** con ciertas características especificadas relevantes para el planteamiento del problema.

Determinada la muestra, corresponde aplicar los experimentos y/o mediciones.



# Instrumentos de medición

La **recolección de datos cuantitativos** se realiza mediante **instrumentos de medición** que deben representar fielmente las variables de interés (hipótesis).

Aquí, **medición** se refiere al “proceso de vincular conceptos abstractos con indicadores empíricos”.

Un instrumento de medición debe cumplir tres requisitos:

1. **validez** (mide lo que corresponde)
2. **confiabilidad** (resultados consistentes)
3. **objetividad**

--- Hernández, Fernández y Baptista (2014)

Formación en Investigación



# Instrumentos de medición

La **validez** se refiere al grado en que un instrumento de medición mide realmente las variables que pretende medir.

Se puede obtener **tres tipos de evidencia** para la validez cuantitativa:

1. evidencia sobre la **validez de contenido**, al contrastar el universo de ítems (dominio) frente a los ítems presentes en el instrumento
2. evidencia sobre la **validez de criterio**, al comparar los resultados de la aplicación del instrumento frente a los resultados de un criterio externo
3. evidencia sobre la **validez de constructo**, mediante el análisis de factores y al verificar la teoría subyacente

--- Hernández, Fernández y Baptista (2014)



# Instrumentos de medición

La **confiabilidad** se refiere al grado en que la aplicación repetida de un instrumento de medición, a los mismos casos, produce resultados iguales.

La confiabilidad cuantitativa se determina al calcular el **coeficiente de fiabilidad**.

Los coeficientes de fiabilidad cuantitativa varían entre 0 (nula confiabilidad) y 1 (total confiabilidad).

Los métodos más conocidos para calcular la confiabilidad son: medida de estabilidad (**test-retest**), formas alternativas o **paralelas**, mitades **partidas** y **consistencia interna**.

--- Hernández, Fernández y Baptista (2014)

Formación en Investigación



# Instrumentos de medición

Los factores que pueden afectar la confiabilidad y la validez son:

1. la **improvisación**, pensar que elegir un instrumento o desarrollarlo es algo que puede tomarse a la ligera (e.g. encuestas de usabilidad)
2. utilizar instrumentos desarrollados en el extranjero y que no han sido validados en el **contexto local**
3. poca o nula **empatía** con los participantes
4. **condiciones** en las que se aplica el instrumento de medición

**No hay medición perfecta**, pero el error de medición debe reducirse a límites tolerables.





# Instrumentos de medición

La **objetividad** se refiere al grado en que el instrumento es o no permeable a la influencia de los sesgos y tendencias del investigador o investigadores que lo administran, califican e interpretan.

La objetividad es **difícil de lograr**, particularmente en las ciencias sociales. Se refuerza:

1. **estandarizando** la aplicación del instrumento (mismas instrucciones y condiciones para todos los participantes) y la evaluación de los resultados
2. empleando **personal capacitado** y experimentado en el instrumento

--- Hernández, Fernández y Baptista (2014)

Formación en Investigación



# Instrumentos de medición

Los estudios cuantitativos buscan que la influencia de las características y las tendencias del investigador se reduzca al mínimo posible.

Pero esto es un ideal, pues la investigación siempre es realizada por seres humanos.

La validez, la confiabilidad y la objetividad **no** deben tratarse de forma separada. Sin alguna de las tres, el instrumento **no es útil** para llevar a cabo un estudio.

--- Hernández, Fernández y Baptista (2014)

Formación en Investigación



# Instrumentos de medición

Los primeros pasos para elaborar un instrumento de medición son:

1. Redefiniciones fundamentales sobre propósitos, definiciones operacionales y participantes
2. Revisar la literatura, particularmente la enfocada en los instrumentos utilizados para medir las variables de interés
3. Identificar el conjunto o dominio de conceptos o variables a medir e indicadores de cada variable
4. Tomar decisiones en cuanto a: tipo y formato; utilizar uno existente, adaptarlo o construir uno nuevo, así como el contexto de aplicación
5. Construir la primera versión del instrumento

--- Hernández, Fernández y Baptista (2014)



# Instrumentos de medición

Los siguientes pasos para elaborar un instrumento de medición son:

6. Aplicar la prueba piloto (para calcular la confiabilidad y validez iniciales)
7. Desarrollar la versión definitiva del instrumento
8. Entrenar al personal que va a aplicarlo
9. Obtener autorizaciones para aplicarlo
10. Aplicar el instrumento

Luego se debe preparar los datos para el análisis estadístico: codificarlos, limpiarlos, insertarlos en una matriz o base de datos.

# ¿Preguntas?