



Cómo leer una ficha técnica

En esta sección discutiremos acerca de la información que contiene la ficha técnica de un estudio.

Todo estudio estadístico, ya sea generado por una empresa de investigación de mercado, organismo público u otro debe tener una ficha técnica. Esta contiene los detalles metodológicos del estudio.

Entender la información contenida en ella es de vital importancia pues nos ayuda a comprender los alcances y limitaciones del estudio evitando así sacar conclusiones erróneas.

Al revisar diversas fichas técnicas provenientes de distintos estudios y/o organismos nos encontramos con que la información más recurrente contenida en una ficha técnica es la siguiente:

- Población objetivo o universo.
- Cobertura geográfica
- Tamaño de muestra
- Diseño muestral
- Margen de error
- Técnica de recolección de los datos
- Fecha de terreno
- Ponderación

Es importante señalar que no existe una uniformidad con respecto a los nombres de estos conceptos, y estos pueden variar de acuerdo a la empresa u organismo que realizó el estudio.

A continuación, analizaremos en detalle cada uno de estos puntos.

Población objetivo o universo.

Hace referencia a la población que es interés en el estudio detallando sus características demográficas tales como género y edad. Además, puede señalar los límites geográficos del estudio.

Por ejemplo, en la encuesta Plaza Pública de Cadem (<https://plazapublica.cl/>), la ficha técnica señala:

“Universo. Hombres y mujeres mayores de 18 años, habitantes de las 16 regiones del país.”

Esto significa que el estudio considera solo personas mayores de 18 años y que habitan en nuestro país.

Cobertura geográfica

La cobertura geográfica puede hacer referencia a los límites geográficos del estudio cuando este no se ha especificado en la definición de la población objetivo, pero también puede hacer referencia al porcentaje de la población que cubre el estudio. Por ejemplo, la ficha técnica del estudio 3D de GFK señala lo siguiente (<https://www.gfk.com/es-cl/insights/press-release/chile3d-2019-marcas-de-excelencia/>);

“Abarca a las 39 comunas más pobladas de Chile, de norte a sur, **las que representan el 80%** de la población nacional. La población objetivo del estudio son hombres y mujeres, mayores de 15 años, de niveles socioeconómicos ABC1a, C1b, C2, C3 y DE.”

Tamaño de muestra

El tamaño de muestra hace referencia a la cantidad de elementos contenidos en la muestra.

Margen de error

El margen de error se refiere a la precisión de estudio. Este siempre está asociado a un nivel de confianza el cual hace referencia a la seguridad que tenemos de que ese sea el margen de error.

Por ejemplo, en el estudio de Plaza Pública de Cadem de la cuarta semana de mayo del 2019, el margen de error es del 3,7% con un 95% de confianza. Lo anterior quiere decir, que, si el porcentaje de personas que declaran aprobar la forma como Sebastián Piñera está conduciendo el gobierno es del 33%, con un 95% de confianza (de seguridad) podemos decir que el porcentaje de personas que declaran aprobar la forma como Sebastián Piñera está conduciendo el gobierno está entre $33\% \pm 3,7\%$. En otras palabras, está entre, aproximadamente, el 29% y el 37%.

El margen de error cobra aún más relevancia cuando se analiza el comportamiento de un indicador que se mide a lo largo del tiempo. Por ejemplo, en el estudio de Plaza Pública de Cadem de la tercera semana de mayo del 2019, el porcentaje de personas que declaran aprobar la forma como Sebastián Piñera está conduciendo el gobierno es del 34%. Si comparamos los resultados de la cuarta semana de mayo con los de la tercera semana de mayo el primer impulso es concluir que la aprobación de cómo Sebastián Piñera está conduciendo el gobierno cayó un punto, sin embargo, como el margen de error es casi del 4% esta variación pierde significancia, es decir, podemos concluir que no ha habido variación estadísticamente significativa del porcentaje de aprobación de cómo Sebastián Piñera está conduciendo el gobierno. En otras palabras, solo aquellas variaciones que en valor absoluto sean superiores al margen de error serán significativas.

95% suele ser el nivel de confianza estándar utilizado en diversos estudios de mercado. Es importante tener esto en cuenta ya que muchas veces se hace referencia al margen de error, pero no el nivel de confianza y es muy probable que este sea del 95%.

Diseño muestral

El diseño muestral o técnica de muestreo hace referencia a cómo se selecciona la muestra a partir de una población.

Una muestra es un subconjunto de la población que se selecciona para realizar un determinado estudio. Esto, porque en la práctica, no es posible realizar el estudio sobre el total de la población.

Existen distintas técnicas de muestreo. Estas se dividen en dos grupos:

- Técnicas probabilísticas
- Técnicas no probabilísticas

Las técnicas probabilísticas tienen las siguientes características:

- La elección de los elementos es al azar
- Todos los elementos de la población tienen probabilidad conocida y no nula de ser elegidos
- Es posible estimar el error muestral
- Sus resultados pueden inferirse al universo o población

Las técnicas probabilísticas requieren un marco muestral. En palabras simples un marco muestral es un "listado" de todos los elementos de la población. Hoy en día y en este contexto, la palabra listado es un poco antigua y podría decirse que el marco muestral es una base de datos que contiene todos los elementos de la población.

Los métodos no probabilísticos se caracterizan por no cumplir con las características mencionadas anteriormente. Es decir:

- La elección de los elementos no es al azar, sino de forma cómoda a juicio del propio investigador
- No es posible estimar el error muestral
- Sus resultados no pueden extrapolarse al universo o población
- Su única ventaja es que su costo es inferior al de los métodos probabilísticos

Los principales métodos de muestreo probabilístico son:

- Muestreo aleatorio simple
- Muestreo sistemático
- Muestreo estratificado
- Muestro por conglomerados o por área

Muestreo aleatorio simple

El principio básico tras el muestreo aleatorio simple es que cada elemento de la población tiene igual probabilidad de ser seleccionado.

El procedimiento del muestreo aleatorio simple puede describirse de la siguiente manera:

- Se enumeran todos los elementos del marco muestral
- En una urna se colocan papeles numerados de 1 a N, donde N corresponde al total de elementos de la población.
- Se seleccionan uno a uno n papeles de la urna, donde n corresponde al tamaño de muestra.
- Sí, por ejemplo, se extrajo de la urna el papel con el número 1, entonces eso implica que el primer elemento del marco muestral fue seleccionado.

Muestreo sistemático

El muestreo sistemático Es similar al muestreo aleatorio simple salvo que en vez de elegir los elementos de la muestra como extrayendo números al azar desde una muestra, se seleccionan desde el marco muestral mediante un proceso que se llama salto sistemático.

Consideremos el siguiente ejemplo. Suponga que se tiene una población de tamaño $N = 100$ y que se desea extraer una muestra de tamaño $n = 30$. Si hacemos $100/30 = 3,33 \approx 3$. Esto implica que, del marco muestral, se seleccionará cada tercer elemento del marco muestral. El primer elemento será cualquiera entre los tres primeros, pero este debe ser seleccionado al azar. Una vez seleccionado el primer elemento, se selecciona cada tercer elemento. De esta manera, si el primer elemento seleccionado es el 2, el siguiente será el 5, el subsiguiente el 8 y así consecutivamente.

Muestreo estratificado

En el muestreo estratificado, la población se subdivide en estratos. Los estratos son subgrupos de la población que se caracterizan por que dentro de ellos los elementos son homogéneos, mientras que los elementos de los diferentes estratos son heterogéneos. Además, los estratos son excluyentes y exhaustivos. Lo primero significa que cada elemento de la población puede pertenecer a un solo estrato y lo segundo que todo elemento de la población debe pertenecer a algún estrato.

Algunas de las estratificaciones más comunes son: género, tramo etario, nivel socioeconómico, región y comuna entre otras.

Una vez definido los estratos, se procede a extraer una muestra aleatoria simple de cada uno de ellos.

El muestreo estratificado busca mejorar la precisión de las estimaciones asegurando que cada estrato de la población este representado en la muestra.

Muestro por conglomerados o de área

En el muestreo por conglomerados, la población se subdivide en conglomerados. En contraposición a los estratos, los conglomerados son subgrupos de la población que se caracterizan por que dentro de ellos los elementos son heterogéneos, mientras que los elementos de los diferentes estratos son homogéneos. Los conglomerados también son excluyentes y exhaustivos.

Los conglomerados por lo general son bloques de manzanas u otra subdivisión geográfica y es por esto que también recibe el nombre de muestreo por áreas.

En el muestreo por conglomerados no necesariamente se toma una muestra de cada conglomerado. En muchos casos, en principio se toma una muestra de conglomerados y luego o bien se levanta la

información de todo el conglomerado o se selecciona, mediante algún mecanismo, una muestra en cada conglomerado seleccionado.

El muestreo por conglomerados busca principalmente reducir los costos haciendo el proceso de toma de muestra más eficiente ya que disminuye los tiempos de desplazamiento del encuestador.

Es bastante usual que en la descripción de un diseño muestral se hable de diseños bietápico o trietápico. Esto se refiere a que los elementos que conformarán la muestra no se seleccionan de una sola vez, si no que para llegar a ellos se debe hacer dos (diseño bietápico) o tres (diseño trietápico) muestreos de manera sucesiva.

Por ejemplo, en la ficha técnica de la nueva encuesta nacional de empleo (http://historico.ine.cl/canales/chile_estadistico/mercado_del_trabajo/nene/pdf/ficha_tecnica_NENE.pdf) en Fuente de Información se señala lo siguiente: *“El esquema de muestreo es probabilístico y bietápico. Tiene como unidad de selección de primera etapa (UPE) las manzanas (o secciones en la parte RAU-Rural del marco), como unidad de segunda etapa (USE) a las viviendas particulares...”* Aquí se seleccionan secuencialmente dos muestras, en una primera etapa se seleccionan manzanas y posteriormente dentro de las manzanas seleccionadas se seleccionan viviendas.

Los principales métodos de muestreo no probabilístico son:

- Muestro por conveniencia
- Muestreo por juicio
- Muestro por cuotas
- Muestreo de bola de nieve

Muestro por conveniencia

Como su nombre lo indica, el muestreo por conveniencia selecciona los elementos de la muestra de manera conveniente. Por ejemplo, si no interesa hacer un estudio acerca de la experiencia de compra de una persona en el Costanera Center, lo que más no conviene es pararnos afuera del Costanera Center y entrevistara a las personas que salen de él. De esta manera garantizamos que las personas a encuestar cumplan con el requisito.

Muestro por juicio

En este caso se confía en la experiencia del encuestador para seleccionar a los elementos de la muestra. Por ejemplo, si para un estudio se necesita encuestar a dueñas de casa de los niveles socioeconómicos C2 y C3, queda a juicio del encuestador determinar cuáles son los casos que cumplen con los requisitos.

Muestro por cuotas

Podría decirse que el muestreo por cuotas es una especie de muestreo estratificado pero no probabilístico. Esto en el sentido de que se extraen muestras de distintos segmentos presentes en la población y además se establece una cuota (cantidad fija) de elementos a seleccionar de cada grupo. Por lo general estas cuotas son proporcionales a la composición de la población con respecto a dichos segmentos. El encuestador selecciona, a veces haciendo un muestro por juicio, cada uno de los elementos hasta completar las cuotas.

Por ejemplo, supongamos que se quiere hacer un estudio que considera los siguientes tramos etarios: 18 a 20 años, 21 a 23 años y 24 a 25 años. Dentro del rango de los 18 a 25 años, el primer tramo pesa un 35%, el segundo un 38% y el último un 27%. Si vamos a tomar una muestra de 100 individuos mediante el muestreo por cuotas se seleccionarán 35 individuos entre 18 a 20 años, 38 de 21 a 23 años y 27 de 24 a 25 años.

Muestro bola de nieve.

El muestreo bola de nieve también es conocido como muestreo por referidos. Se utiliza cuando la población de interés es muy pequeña y/o difícil de contactar. Por ejemplo, si queremos hacer un estudio sobre los inmigrantes que se encuentran de manera ilegal en Chile probablemente será difícil localizarlos. Basta comunicarse con un inmigrante ilegal para que este te contacte con otra y así sucesivamente.

Técnica de recolección de los datos

La técnica de recolección de los datos hace referencia a como se levantó la información. Años atrás era usual que estas fueran solo presenciales o telefónicas. Hoy en día gracias a la masificación de los medios digitales hay varios estudios que se hacen por internet.

Una técnica muy utilizada hoy en día es la metodología CATI. CATI significa Computer Assisted Telephone Interview.

En esta metodología el encuestador registra las respuestas del entrevistado en un cuestionario que aparece en la pantalla de un computador. Además este sistema selecciona aleatoriamente la muestra a partir de una base de números telefónicos y marca automáticamente los números de teléfono de las personas a entrevistar.

Fecha de terreno. Corresponde a la fecha en que se realizó la encuesta.

Ponderación

En la ponderación, a cada individuo se le asigna un peso de acuerdo a su importancia en relación con otros casos o encuestados. El valor 1.0 representa el caso no ponderado.

El objetivo de la ponderación es aumentar o disminuir la importancia del número de casos de la muestra que poseen determinadas características tales como, género, edad, nivel socioeconómico entre otras, con el fin de que estos sean más representativos de la población evitando así que algunos segmentos de la población estén sub o sobre representados.

Si asumimos que en muchos casos las características de un individuo pueden estar relacionadas con su opinión acerca de un tema en particular, por ejemplo, el nivel socioeconómico y la intención de voto en las próximas elecciones municipales, el hecho de que un segmento se encuentre sub o sobre representado puede llevar a interpretaciones erróneas.