

TÉCNICAS DE INVESTIGACIÓN DE MERCADOS

Felicidad Marqués Asensio



Técnicas de investigación de mercados
Felicidad Marqués Asensio

ISBN: 978-84-943055-4-2
EAN: 9788494305542
IBIC: KJSM, UNC

Copyright © 2015 RC Libros
© RC Libros es un sello y marca comercial registrados

Técnicas de investigación de mercados.

Reservados todos los derechos. Ninguna parte de este libro incluida la cubierta puede ser reproducida, su contenido está protegido por la Ley vigente que establece penas de prisión y/o multas a quienes intencionadamente reprodujeren o plagiaren, en todo o en parte, una obra literaria, artística o científica, o su transformación, interpretación o ejecución en cualquier tipo de soporte existente o de próxima invención, sin autorización previa y por escrito de los titulares de los derechos de la propiedad intelectual. La infracción de los derechos citados puede constituir delito contra la propiedad intelectual. (Art. 270 y siguientes del Código Penal). Diríjase a CEDRO (Centro Español de Derechos Reprográficos) si necesita fotocopiar o escanear algún fragmento de esta obra a través de la web www.conlicencia.com; o por teléfono a: 91 702 19 70 / 93 272 04 47.

RC Libros, el Autor, y cualquier persona o empresa participante en la redacción, edición o producción de este libro, en ningún caso serán responsables de los resultados del uso de su contenido, ni de cualquier violación de patentes o derechos de terceras partes. El objetivo de la obra es proporcionar al lector conocimientos precisos y acreditados sobre el tema tratado pero su venta no supone ninguna forma de asistencia legal, administrativa ni de ningún otro tipo, si se precisase ayuda adicional o experta deberán buscarse los servicios de profesionales competentes. Productos y marcas citados en su contenido estén o no registrados, pertenecen a sus respectivos propietarios.

RC Libros
Calle Mar Mediterráneo, 2. Nave 6
28830 SAN FERNANDO DE HENARES, Madrid
Teléfono: +34 91 677 57 22
Fax: +34 91 677 57 22
Correo electrónico: info@rclibros.es
Internet: www.rclibros.es

Diseño de colección, cubierta y pre-impresión: Grupo RC
Impresión y encuadernación: Arvato
Depósito Legal: M-34247-2014
Impreso en España

19 18 17 16 15 (1 2 3 4 5 6 7 8 9 10 11 12)

INVESTIGACIÓN DE MERCADOS Y MUESTREO. MUESTREO ALEATORIO SIMPLE Y SISTEMÁTICO

LA TÉCNICAS DE MUESTREO Y LA INVESTIGACIÓN DE MERCADOS

La investigación de mercados recopila, procesa y analiza la información relativa a la mercadotecnia. Como los mercados son amplios y los clientes son numerosos, no será posible analizar a toda la población. Esta circunstancia nos obliga a hacer uso de la teoría del muestreo de modo ineludible.

En toda investigación de mercados existe un conjunto de elementos sobre los que se requiere información. Este conjunto de elementos es lo que se denota con el nombre de población o universo estadístico. Cuando el investigador toma información de todos y cada uno de los elementos de la población se dice que está realizando un *censo*. Sin embargo, esto no es muchas veces posible, ya sea por el coste que resulta de la toma de información, o bien porque la toma de información lleve consigo la destrucción de los elementos en cuestión, o que la población tenga infinitos elementos, o por otras causas.

Este problema lleva al investigador a tomar la información solo de una parte de los elementos de la población estadística, proceso que recibe el nombre de *muestreo*.

El conjunto de elementos de los que se toma información en el proceso de muestreo se llama *muestra* y el número de elementos que la componen *tamaño muestral*. Existen varios tipos de muestreo, dependiendo de que la población estadística sea finita o infinita, materia sobre la que existe amplia literatura estadística, pero en la investigación de mercados consideraremos solamente el *muestreo en poblaciones finitas*. El investigador utiliza la muestra para la toma de información, pero lo importante es que dicha muestra sea representativa.

Por lo tanto, entenderemos por muestra un subconjunto lo más representativo posible de una población. Naturalmente, el estudio del colectivo requerirá un cuidadoso proceso de muestreo, a los efectos de elaborar ese subconjunto en las condiciones más adecuadas de representatividad, puesto que la inferencia se caracterizará por aplicar al colectivo las conclusiones obtenidas a partir de la muestra. En este sentido, el método de selección de la muestra reviste una singular importancia, dado que, dependiendo de cómo se haya constituido esta se seguirán unos u otros resultados. Precisamente, los métodos de selección de la muestra serán el núcleo principal a tratar en los primeros temas de este libro.

Con la finalidad de medir el grado de representatividad de la muestra lo mejor posible es necesario utilizar muestreo probabilístico. Diremos que el muestreo es probabilístico cuando pueda establecerse la probabilidad de obtener cada una de las muestras que sea posible seleccionar, esto es, cuando la selección de muestras constituya un fenómeno aleatorio probabilizable. Dicha selección se verificará en condiciones de azar, siendo susceptible de medida la incertidumbre derivada de la misma. Esto permitirá medir los errores cometidos en el proceso de muestreo.

Se denomina *inferencia estadística o estadística inductiva* a la metodología consistente en inferir resultados, predicciones y generalizaciones sobre la población, basándose en la información contenida en las muestras representativas previamente elegidas por métodos de muestreo formales. La inferencia estadística está basada en la teoría de la probabilidad, pero tiene un carácter diferente. En inferencia estadística se consideran fenómenos en los que se manifiesta la regularidad estadística y se construyen modelos probabilísticos para describirlos.

Podemos definir los *métodos de muestreo* como el conjunto de técnicas estadísticas que estudian la forma de seleccionar una *muestra lo suficientemente*

representativa de una población cuya información permita inferir las propiedades o características de toda la población cometiendo un *error medible y acotable*.

A partir de una muestra, seleccionada mediante un determinado método de muestreo, se estiman las características poblacionales (media, total, proporción, etc.) con un error cuantificable y controlable. Las estimaciones se realizan a través de funciones matemáticas de la muestra denominadas *estimadores*, que se convierten en variables aleatorias al considerar la variabilidad de las muestras. Los errores se cuantifican mediante varianzas, desviaciones típicas o errores cuadráticos medios de los estimadores, que miden la precisión de los mismos.

La teoría del muestreo proporciona una técnica estadística de carácter muy práctico que sencillamente busca obtener datos de una población (clientes, empresas, hogares, etc.) en su totalidad, utilizando tan solo una parte reducida de la misma, denominada muestra, aunque como es lógico pagando algún coste (calculable) en cuanto a la precisión de las medidas poblacionales inferidas.

De forma metafórica podríamos decir que una muestra, que se supone representativa de una población, es similar a lo que representa una maqueta respecto del edificio del que ofrece una imagen. La muestra, al igual que la maqueta, será mejor o peor, según el grado de representatividad que ofrezca. La teoría del muestreo traslada la información aportada por la muestra a toda la población, dando lugar a lo que se conoce en muestreo como *elevación del dato muestral a la población* que se estudia. En la metáfora de la maqueta el factor de elevación sería la escala de la misma, que permite pasar un dato de la maqueta a su correspondiente dato para el edificio real que representa.

POBLACIÓN, MARCO Y MUESTRA

En muestreo en poblaciones finitas la población finita inicial que se desea investigar se denomina *población objetivo*, pero el muestreo de toda la población objetivo no siempre es posible debido a diferentes problemas que no permiten obtener información de algunos de sus elementos (inaccesibilidad de algunos de sus elementos, negativas a colaborar, ausencias, etc.), con lo que la población que realmente es objeto de estudio o *población investigada* no coincide con la población objetivo.

Por otro lado, para seleccionar la muestra, necesitaremos un listado de unidades de muestreo denominado *marco* que teóricamente debiera de coincidir con la población objetivo. Un marco será más adecuado cuanto mejor cubra la población objetivo, es decir, cuanto menor sea el *error de cobertura*.

Pero en los marcos son inevitables las desactualizaciones, las omisiones de algunas unidades, las duplicaciones de otras y la presencia de unidades extrañas y otras impurezas que obligan a su depuración (*depuración de marcos imperfectos*). Idealmente podría conseguirse la población objetivo eliminando del marco las unidades erróneamente incluidas en él (unidades extrañas, duplicaciones, etc.) y añadiendo las omisiones. Asimismo, también sería una meta que al eliminar del marco las unidades de las que no se puede obtener información (inaccesibles, ausentes, no colaboradoras, etc.) se obtuviera la población investigada.

El marco puede estar constituido por unidades elementales de muestreo o por unidades compuestas. Una *unidad elemental (o simple)* es la unidad de muestreo más sencilla posible y una *unidad compuesta (o primaria)* está formada por varias unidades elementales. Como en la práctica no es fácil disponer de marcos de unidades elementales, se intentan conseguir marcos de unidades compuestas que son más accesibles.

Por ejemplo, para estudiar habitantes de una región es más fácil disponer de un listado de hogares que de un listado de individuos. Se selecciona la muestra de un marco de hogares (unidades compuestas de varios individuos) y después se estudian las propiedades de los individuos con técnicas adecuadas.

MUESTREO Y ESTIMACIÓN. ESTIMACIÓN PUNTUAL

Muestreo probabilístico. Métodos de muestreo

Consideramos la realización de un determinado experimento o fenómeno cuyos resultados se denominan *sucesos*. Entre los experimentos o fenómenos, se denominan *deterministas* aquellos en los que vamos a conocer *a priori* sus sucesos resultantes, y se denominan *aleatorios* aquellos cuyos sucesos son desconocidos *a priori*. El estudio de la probabilidad se ocupa de los fenómenos o experimentos aleatorios.

Sean S_1, S_2, \dots, S_n los sucesos elementales asociados a un fenómeno o experimento aleatorio dado, entendiéndose por *sucesos elementales* los más simples posibles, es decir, aquellos que no pueden ser descompuestos en otros sucesos. El conjunto $\{S_1, S_2, \dots, S_n\}$ se denomina *espacio muestral* asociado al fenómeno o experimento.

Si consideramos como fenómeno o experimento la extracción aleatoria de muestras dentro de una población por un procedimiento o método de muestreo dado, podemos considerar como sucesos elementales las muestras obtenidas, constituyendo el

conjunto de las mismas el espacio muestral. Si representamos el conjunto de las N unidades que constituyen la población finita objeto de estudio por $U = \{u_1, u_2, \dots, u_N\}$, una muestra de tamaño n puede considerarse como un subconjunto ordenado de n elementos de U , $S_i = \{u_{i1}, u_{i2}, \dots, u_{in}\}$, donde u_{ij} denota el elemento que ocupa el lugar j en la muestra S_i . Se considera el subconjunto S_i ordenado porque, en general, el orden de colocación de los elementos en las muestras puede ser pertinente, siendo distintas entre sí muestras con los mismos elementos colocados en distinto orden. El conjunto de las N^n muestras posibles de tamaño n que se pueden formar con los N elementos de la población U es el espacio muestral S .

Hay que especificar que, en general, el orden de colocación de los elementos en las muestras sí influye, siendo muestras distintas aquéllas que tienen los mismos elementos situados en distinto orden. Pero lo habitual es que los métodos de muestreo comunes consideren iguales muestras con los mismos elementos, aunque estén colocados en orden diferente. En este caso habitual, en el que el orden de colocación de los elementos en las muestras no se tiene en cuenta, suele expresarse una muestra de tamaño n como $s = \{u_1, u_2, \dots, u_n\}$.

Con la finalidad de medir el grado de representatividad de la muestra lo mejor posible, es necesario utilizar muestreo probabilístico. Diremos que *el muestreo es probabilístico* cuando pueda establecerse la probabilidad de obtener cada una de las muestras que sea posible seleccionar (elementos del espacio muestral, S) mediante un procedimiento de muestreo dado, esto es, cuando la selección de muestras constituya un fenómeno aleatorio probabilizable. Dicha selección se verificará en condiciones de azar, siendo susceptible de medida la incertidumbre derivada de la misma. Esto permitirá medir los errores cometidos en el proceso de muestreo. Evidentemente, para establecer la probabilidad de todas las muestras posibles derivadas de un procedimiento de muestreo dado, será necesario conocer ese conjunto de muestras, es decir, será necesario delimitar tanto el método de muestreo como el espacio muestral derivado del mismo.

Un *procedimiento o método de muestreo* es sencillamente un proceso o mecanismo mediante el que se seleccionan las muestras de modo que cada una tenga una determinada probabilidad de ser elegida.

Por lo tanto, el método aleatorio empleado para seleccionar la muestra define en el espacio muestral S una función de probabilidad P tal que:

- $P(S_i) \geq 0 \forall i$

- $\sum_s P(S_i) = 1$

En general, puede ocurrir que no todas las muestras del espacio muestral pueden ser elegidas. No obstante, consideraremos métodos de muestreo en los que todas las muestras puedan ser seleccionadas, es decir, $P(S_i) > 0 \forall i$; se trata entonces de *métodos de muestreo no restringidos*. En ocasiones suele expresarse un procedimiento de muestreo mediante la terna $\{U, S, P\}$, que indica que el procedimiento de muestreo definido en la población U establece en el espacio muestral S asociado la ley de probabilidad P .

Las formas básicas de selección de la muestra se clasifican atendiendo a los siguientes criterios:

1. Atendiendo a las probabilidades de selección:

- 1.1. *Con probabilidades iguales*: todas las unidades de la población tienen la misma probabilidad de ser seleccionadas en cada extracción $\pi_i = n/N$ y $P_i = 1/N$.
- 1.2. *Con probabilidades desiguales*: al menos dos unidades tienen distintas probabilidades de selección en cierta extracción. Tiene especial interés la selección con probabilidades proporcionales a los tamaños (Brewer, Murthy, Sampford, etc.).

2. Atendiendo a la mecánica de selección:

- 2.1. *Muestreo con reposición*: cada unidad que es extraída para formar parte de la muestra en una extracción se repone a la población antes de realizar la siguiente extracción; es decir, la estructura poblacional permanece invariante.
- 2.2. *Muestreo sin reposición*: cada unidad que es extraída para formar parte de la muestra en una extracción no se repone a la población antes de realizar la siguiente extracción, por lo que una unidad podrá aparecer en la muestra a lo sumo una vez y la estructura poblacional cambia de una extracción a otra.

Combinando estos cuatro tipos de muestreo tenemos: muestreo con reposición y probabilidades iguales, muestreo sin reposición y probabilidades iguales, muestreo con reposición y probabilidades desiguales y muestreo sin reposición y probabilidades desiguales.

Los *tipos de muestreo más habituales* son los siguientes:

Aleatorio simple: en una muestra aleatoria simple, las unidades de muestreo individuales se seleccionan aleatoriamente con la misma probabilidad con o sin reposición. Las unidades de muestreo son unidades elementales que se seleccionan directamente a partir de la totalidad de la población.

Sistemático 1 en n: se ordenan los elementos y se elige uno de cada n

Estratificado: el muestreo estratificado implica seleccionar muestras aleatorias simples independientemente dentro de subgrupos de la población homogéneos dentro de sí y heterogéneos entre sí y que no se solapen (estratos). Por ejemplo, los estratos pueden ser grupos socioeconómicos, categorías laborales, grupos de edad o grupos étnicos. Con la estratificación, puede asegurarse que los tamaños muestrales de los subgrupos de interés son adecuados, mejorar la precisión de las estimaciones globales e incluso utilizar distintos métodos de muestreo entre los diferentes estratos.

Conglomerados: el muestreo por conglomerados implica la selección en la población de grupos de unidades muestrales o conglomerados heterogéneos dentro de sí y homogéneos entre sí y que no se solapen. Por ejemplo, los conglomerados pueden ser escuelas, hospitales, hogares o zonas geográficas y las unidades muestrales son ahora unidades compuestas que contienen cada una varias unidades elementales en estudio (alumnos, pacientes, personas o ciudadanos). Subdivisiones sucesivas de las unidades de muestreo compuestas pueden llevar a muestreo monoetápico, bietápico o polietápico en general.

Múltiples etapas: en el muestreo polietápico, se selecciona una muestra de primera etapa basada en conglomerados. A continuación, se crea una muestra de segunda etapa extrayendo submuestras a partir de los conglomerados seleccionados. Si la muestra de segunda etapa está basada en subconglomerados, entonces puede añadir una tercera etapa a la muestra. Por ejemplo, en la primera etapa de una encuesta, se podría extraer una muestra de ciudades. A continuación, y a partir de las ciudades seleccionadas, se podrían muestrear unidades familiares. Finalmente, a partir de las unidades familiares seleccionadas, se podría encuestar a individuos. Los Asistentes de muestreo y preparación del análisis permiten especificar tres etapas en un diseño.

Muestreo no aleatorio: cuando es difícil obtener la muestra aleatoriamente, las unidades se pueden muestrear sistemáticamente (con un intervalo fijo) o secuencialmente.

Probabilidades de selección desiguales: cuando se muestrean conglomerados que contienen números de unidades desiguales, puede utilizarse el muestreo probabilístico proporcional al tamaño (PPS) para que la probabilidad de selección del conglomerado

sea igual a la proporción de unidades que contiene. El muestreo PPS también puede utilizar esquemas de ponderación más generales para seleccionar unidades.

Muestreo no restringido: el muestreo no restringido selecciona las unidades con reposición (CR), por lo que se puede seleccionar más de una vez una unidad individual para la muestra.

Ponderaciones muestrales, pesos o factores de elevación: las ponderaciones muestrales se calculan automáticamente al extraer una muestra compleja y de forma ideal se corresponden con la frecuencia que cada unidad muestral representa en la población objetivo.

Por lo tanto, la suma de las ponderaciones muestrales debe estimar el tamaño de la población. Los procedimientos de análisis de muestras complejas requieren las ponderaciones muestrales para poder analizar correctamente una muestra compleja.

Estimadores

A partir de una muestra, seleccionada mediante un determinado método de muestreo, se estiman las características poblacionales (media, total, proporción, etc.), con un error cuantificable y controlable. Las estimaciones se realizan a través de funciones matemáticas de la muestra denominadas *estimadores*, que se convierten en variables aleatorias al considerar la variabilidad de selección de las muestras, y que por tanto cumplen las condiciones de una función de medida. Los errores se cuantifican mediante varianzas, desviaciones típicas o errores cuadráticos medios de los estimadores, que miden la precisión de los mismos.

Supongamos que tenemos definida una característica X en la población $U = \{u_1, u_2, \dots, u_N\}$ que toma el valor numérico X_i sobre la unidad u_i $i = 1, 2, \dots, N$, dando lugar al conjunto de valores $\{X_1, X_2, \dots, X_N\}$. Consideramos ahora una cierta función θ de los N valores X_i , que suele denominarse parámetro poblacional. Seleccionamos una muestra $s = \{u_1, u_2, \dots, u_n\}$ de U mediante un procedimiento de muestreo dado, y consideramos los valores $s(X) = \{X_1, X_2, \dots, X_n\}$ que toma la característica X en estudio sobre los elementos de la muestra. A partir de estos valores estimamos puntualmente el parámetro poblacional θ mediante la expresión $\hat{\theta} = \hat{\theta}(s(X)) = \hat{\theta}(X_1, \dots, X_n)$, basada en los valores X_i $i = 1, 2, \dots, n$, que toma la característica X sobre las unidades de la muestra s .

La función $\hat{\theta}$ que asocia a cada muestra s el valor numérico $\hat{\theta}(s(X)) = \hat{\theta}(X_1, \dots, X_n)$, se denomina *estimador* del parámetro poblacional θ . A los valores $\hat{\theta}(s(X))$ para cada s del espacio muestral se les denomina *estimaciones puntuales*.

Entre los parámetros poblacionales θ (función de los N valores X_i) más comunes a estimar, tenemos el total poblacional y la media poblacional para la característica X , definidos de la forma siguiente:

- *Total poblacional:* $X = \theta(X_1, \dots, X_N) = \sum_{i=1}^N X_i$
- *Media poblacional:* $\bar{X} = \theta(X_1, \dots, X_N) = \frac{X}{N} = \frac{1}{N} \sum_{i=1}^N X_i = \sum_{i=1}^N \frac{X_i}{N}$

Hasta ahora hemos supuesto que la característica X definida sobre los elementos de la población es cuantitativa, es decir, cuantificable numéricamente. Sin embargo, también se pueden definir características cualitativas sobre los elementos de la población, como por ejemplo su pertenencia o no a una determinada clase A . Si, para cada unidad u_i , $i = 1, 2, \dots, N$ de la población definimos la característica A_i , que toma valor 1 si la unidad u_i pertenece a la clase A , y que toma valor 0 si la unidad u_i no pertenece a la clase A , podemos definir el total de elementos de la población que pertenecen a la clase A (total de clase) y la proporción de elementos de la población que pertenecen a la clase A (proporción de clase) de la forma siguiente:

- *Total de clase:* $A = \theta(A_1, \dots, A_N) = \sum_{i=1}^N A_i$
- *Proporción de clase:* $P = \theta(A_1, \dots, A_N) = \frac{A}{N} = \frac{1}{N} \sum_{i=1}^N A_i = \sum_{i=1}^N \frac{A_i}{N}$

Analizados ya los cuatro parámetros poblacionales más típicos a estimar, vemos que, en general, un parámetro poblacional θ puede expresarse como una suma de elementos Y_i función de los valores que la característica cuantitativa X o cualitativa A considerada toma sobre los elementos de la población. De esta forma, podemos escribir:

$$\theta = \sum_{i=1}^N Y_i$$

en cuyo caso tenemos:

$$\begin{cases} Y_i = X_i & \text{para el total poblacional } X \\ Y_i = \frac{X_i}{N} & \text{para la media poblacional } \bar{X} \\ Y_i = A_i & \text{para el total de clase } A \\ Y_i = \frac{A_i}{N} & \text{para la proporción de clase } P \end{cases}$$

Ahora surge el problema de analizar la forma de los estimadores puntuales óptimos $\hat{\theta} = \hat{\theta}(X_1, \dots, X_n)$ para estos parámetros poblacionales típicos. Resulta que las mejores propiedades suelen presentarlas los estimadores lineales insesgados de la forma $\hat{\theta} = \sum_{i=1}^n w_i Y_i$.

Concretamente, *para muestreo sin reposición, el estimador óptimo es el de Horvitz y Thompson* $\hat{\theta}_{HT} = \sum_{i=1}^n \frac{Y_i}{\pi_i}$, donde π_i es la probabilidad que tiene la unidad u_i de la población de pertenecer a la muestra; *para muestreo con reposición el estimador óptimo es el de Hansen y Hurwitz* $\hat{\theta}_{HH} = \sum_{i=1}^n \frac{Y_i}{nP_i}$, donde P_i es la probabilidad de seleccionar la unidad u_i de la población para la muestra (probabilidad unitaria de selección de la unidad u_i).

Existen justificaciones para considerar que el parámetro poblacional:

$$\theta = \sum_{i=1}^N Y_i$$

puede estimarse convenientemente mediante el estimador:

$$\hat{\theta} = \sum_{i=1}^n w_i Y_i$$

entre las que podemos citar las siguientes:

- Todas las mediciones de la variable en estudio sobre las unidades de la muestra intervienen en la formación del estimador.

- La importancia de la aportación al estimador de la unidad muestral u_i puede controlarse mediante el coeficiente de ponderación w_i .
- Cuando $w_i = 1$, todas las unidades muestrales intervienen de igual forma en la formación del estimador.
- Los coeficientes w_i pueden depender, entre otros factores, del tamaño de las unidades muestrales, del orden de colocación de las mismas en la muestra, y sobre todo de la probabilidad que tiene la unidad u_i de pertenecer a la muestra según el método de muestreo considerado.

ESTIMACIÓN POR INTERVALOS

Cuando se realiza una afirmación acerca de los parámetros de la población en estudio basándose en la información contenida en la muestra, bien sea mediante los valores puntuales de un estadístico basado en la misma, bien sea señalando un intervalo de valores dentro del cual se tiene confianza de que esté el valor del parámetro, decimos que estamos ante *estimaciones*. En el primer caso estamos ante el proceso de *estimación puntual*, en el que utilizamos directamente los valores de un estadístico, denominado *estimador puntual*, sobre la muestra dada (*estimaciones puntuales*), para estimar los valores poblacionales. En el segundo caso estamos ante la *estimación por intervalos*, donde se calcula un intervalo de confianza en el que razonablemente cae el valor estimado con un *nivel de confianza* prefijado.

Realizar una estimación por intervalos (o definir un intervalo de confianza) para un parámetro poblacional θ al nivel de confianza α es hallar un intervalo real para el que se tiene una probabilidad $1 - \alpha$ de que el verdadero valor del parámetro θ caiga dentro del citado intervalo. El valor $1 - \alpha$ suele denominarse *coeficiente de confianza*.

Intervalos de confianza cuando el estimador es insesgado

Se trata de estimar el parámetro poblacional θ mediante un intervalo de confianza basado en el estimador $\hat{\theta}$ insesgado para θ ($E(\hat{\theta}) = \theta$). Para estimadores insesgados, es necesario distinguir entre el caso en que la distribución del estimador es normal, y el caso en que dicha distribución no puede asegurarse que sea normal.

a) El estimador $\hat{\theta}$ tiene una distribución normal

El intervalo de confianza para el parámetro poblacional θ basado en $\hat{\theta}$ será:

$$\left[\hat{\theta} - \lambda_{\alpha} \sigma(\hat{\theta}), \hat{\theta} + \lambda_{\alpha} \sigma(\hat{\theta}) \right] \text{ con } \lambda_{\alpha} = F_{N(0,1)}^{-1} \left(1 - \frac{\alpha}{2} \right)$$

F es la función de distribución de la normal (0,1), y α es el nivel de confianza.

Si realmente es dudoso que $\hat{\theta}$ tenga una distribución normal, puede utilizarse la distribución t de Student con $n - 1$ grados de libertad para calcular el intervalo de confianza para θ , que en este caso será:

$$\left[\hat{\theta} - t_{\alpha} \hat{\sigma}(\hat{\theta}), \hat{\theta} + t_{\alpha} \hat{\sigma}(\hat{\theta}) \right] \text{ con } t_{\alpha} = F_{t_{n-1}}^{-1} \left(1 - \frac{\alpha}{2} \right)$$

F es la función de distribución de una t de Student con $n - 1$ grados de libertad.

b) El estimador $\hat{\theta}$ no tiene una distribución normal

El intervalo de confianza, derivado de la desigualdad de Tchevichev, para el parámetro poblacional θ basado en $\hat{\theta}$ que cubre el valor de θ con una probabilidad $1 - \alpha$ (coeficiente de confianza), será:

$$\left[\hat{\theta} - \frac{\sigma(\hat{\theta})}{\sqrt{\alpha}}, \hat{\theta} + \frac{\sigma(\hat{\theta})}{\sqrt{\alpha}} \right]$$

Este intervalo suele ser más ancho que el obtenido cuando la distribución de $\hat{\theta}$ es normal. A medida que $\hat{\theta}$ se aleja más de la normalidad, la anchura de este intervalo es mucho mayor respecto del obtenido para normalidad. Ya sabemos que una estimación por intervalos es tanto mejor cuanto más reducido sea el intervalo de confianza correspondiente, de ahí que la propiedad de normalidad sea muy deseable, pues en este caso los intervalos obtenidos son muy estrechos, lo que implica una buena estimación por intervalos.

Intervalos de confianza en estimadores sesgados

El intervalo de confianza para θ basado en el estimador $\hat{\theta}$ en presencia del sesgo no despreciable $B(\hat{\theta}) = E(\hat{\theta}) - \theta$ es el siguiente:

$$\left[\hat{\theta} - \lambda_{\alpha} \sigma(\hat{\theta}) - B(\hat{\theta}), \hat{\theta} + \lambda_{\alpha} \sigma(\hat{\theta}) - B(\hat{\theta}) \right]$$

Observamos que se trata de un intervalo no centrado en $\hat{\theta}$ y desplazado en la cantidad $B(\hat{\theta})$ respecto del intervalo sin sesgo, que debe centrarse situándonos en la peor de las circunstancias, es decir, tomando como extremo fijo del intervalo el más lejano del centro $\hat{\theta}$, y calculando el otro extremo por equidistancia al centro. Ante esta situación, la presencia del sesgo $B(\hat{\theta})$ origina que el intervalo de confianza para θ basado en el estimador $\hat{\theta}$ y centrado en $\hat{\theta}$, tenga una longitud superior al intervalo cuando no hay sesgo. Por lo tanto, la presencia de sesgo conduce a una estimación por intervalos menos precisa.

PRECISIÓN DE LOS ESTIMADORES

Como un estimador $\hat{\theta}$ de un parámetro poblacional θ es sencillamente una variable aleatoria unidimensional, nos interesarán sus características de centralización y dispersión, particularmente su esperanza, su varianza y sus momentos, así como otras medidas relativas a su precisión.

Para analizar la precisión de un estimador suelen utilizarse los conceptos de error de muestreo (o desviación típica), acuracidad (o error cuadrático medio) y sesgo. Suele llamarse precisión a la acuracidad, lo que no es del todo correcto, ya que, aunque la acuracidad sea la magnitud más general para la medición de la precisión, hay casos en los que el análisis puede realizarse en función de otras magnitudes, como el sesgo o la desviación típica. Todas estas magnitudes que influyen en la precisión de un estimador pueden relacionarse a partir de la *descomposición del error cuadrático medio en sus componentes* de la forma siguiente:

$$ECM(\hat{\theta}) = \sigma(\hat{\theta})^2 + B(\hat{\theta})^2$$

Por tanto, la acuracidad (error cuadrático medio) de un estimador se descompone en la suma del cuadrado del error de muestreo y el cuadrado del sesgo.

En la práctica, se considera que el sesgo de $\hat{\theta}$ no es influyente cuando:

$$\left| \frac{B(\hat{\theta})}{\sigma(\hat{\theta})} \right| < \frac{1}{10}$$