



Anexo 1

Lineamientos generales para elaborar un diseño estadístico aplicable a una investigación cuantitativa en Ciencias Sociales

Lic. Jorge Cunillera Canalías
Profesor/Investigador

CONTENIDOS

I. INTRODUCCIÓN.....	3	Tipos de muestreo	20
Fuentes de información	3	Selección y distribución de la muestra	22
Pasos o etapas de una investigación	5	Unidad de muestreo y unidad informante	24
II. EL DISEÑO ESTADÍSTICO	7	COROLARIO	25
Componentes del diseño estadístico	7	BIBLIOGRAFÍA	26
Determinación de la(s) población(es) de interés	8		
Elaboración del marco muestral	10		
Censo o muestreo.....	11		
Aplicación de las técnicas de muestreo.....	14		
Caso de poblaciones infinitas y finitas	14		
Criterio de las proporciones	15		
Tamaño de la muestra	19		
Tamaño de muestra para poblaciones infinitas .	19		
Tamaño de muestra para poblaciones finitas	19		

Figuras, cuadros y tablas

Figura 1. Fuentes de información al desarrollar una investigación	4
Figura 2. Pasos o etapas de una investigación	7
Cuadro 1. Distribución de los clientes actuales de la empresa XYZ, según la categoría de cliente	22
Cuadro 2. Distribución de la muestra según la categoría de cliente	23
TABLA 1. Determinación del tamaño de la muestra	27



I. INTRODUCCIÓN

Este documento ofrece al(a) lector(a) una presentación esquematizada de los principales componentes de un diseño estadístico para una investigación cuantitativa, con el fin de recabar información pertinente, oportuna y apropiada de diferentes fuentes primarias. Además, será posible determinar el número de consumidores que componen o conforman un mercado meta; así como su dispersión y otros aspectos relacionados. Se pretende con este aporte ilustrar a la persona emprendedora, que pretende subsanar este importante paso o etapa de previo al inicio formal de una empresa.

No se pretende que este instrumento sea extensivo ni exhaustivo en el plano estadístico o en el área de la investigación cuantitativa a nivel de fuentes primarias, ya que en la literatura disponible se recrean estos conceptos teóricos con propiedad. Su propósito más bien es ofrecerle al emprendedor(a) una serie de criterios para delimitar los alcances y contenidos de la investigación cuantitativa, relacionada con su proyecto para iniciar o consolidar un negocio. Tampoco se quiere presentar o incorporar definiciones o conceptos teóricos extraídos de la literatura, sino más bien aspectos prácticos que sean de utilidad inmediata en el desarrollo de las investigaciones, con el fin de implementar un futuro negocio.

Previo al desarrollo del tema central, que consiste en el diseño estadístico para una investigación a nivel de fuentes primarias, es necesario ubicar el mismo dentro del contexto de una investigación ya que es parte de esta última.

El diseño estadístico se ubica en la etapa previa a la recolección de la información, por lo que forma parte de la etapa relacionada con la preparación de la pesquisa o el plan para la investigación.

Fuentes de información

Antes de presentar las etapas de una investigación como la que se podría desarrollar, es necesario definir lo que se conoce como **fuentes primarias**, en el entendido de que una **fuentes de información** en general es o se refiere a:

una persona u organización pública o privada, que puede suministrar un dato o información específica y pertinente a su ámbito de acción o de conocimiento

Las fuentes de información para recabar los datos pertinentes a una investigación en particular pueden ser de dos tipos, a saber: primarias y secundarias. **Las fuentes secundarias** se refieren a:



datos o información ya elaborada por una persona física o jurídica y que está o no a disposición del público que tenga un interés válido en esa información. Ejemplos de este tipo de información serían: datos internos de una empresa (ventas, costos y perfil del personal), censos, bases de datos, resultados de otras investigaciones, memorias de una empresa o gremio, etc.

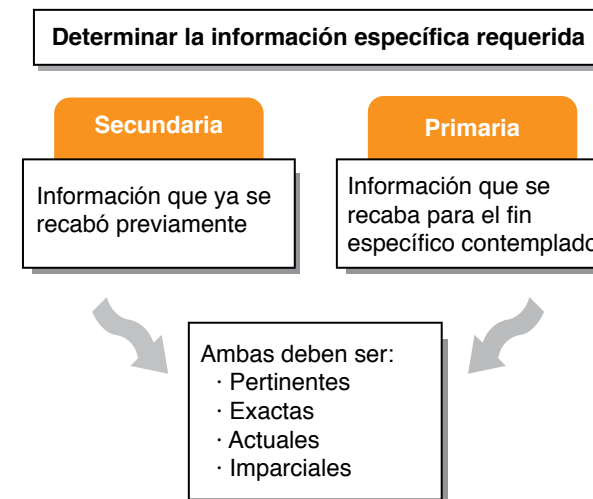
En algunos casos la información incluso se publica, como sucede con los censos nacionales, en el Informe *Estado de la Nación* y las memorias anuales de instituciones públicas y privadas. En la mayoría de los casos estos datos secundarios se mantienen en resguardo, entregándolos a la persona que los solicite si cumple con los requisitos para que se le suministren.

Las fuentes primarias se refieren a:

persona física o jurídica que está en capacidad de suministrar información, que se busca para un fin específico y relacionado con una investigación en particular.

La siguiente figura ejemplifica cómo las diferentes fuentes de información pueden suministrarla al desarrollar una investigación.

Figura 1. Fuentes de información al desarrollar una investigación



Fuente: Kotler, P. y Armstrong, G. (1999). *Marketing*. Octava edición. Editorial Prentice Hall.

Como regla general para el(la) investigador(a):

se debe recurrir inicialmente a las fuentes secundarias para obtener la información necesaria y pertinente para la investigación. En caso de que no esté disponible a este nivel, debe ir a las fuentes primarias. La razón radica en lo costoso que resulta consultar fuentes primarias así como el tiempo que se debe invertir.



Una vez que se han repasado estos conceptos, relacionados con las fuentes de información, se revisan las etapas que componen una investigación, para ubicar el nivel en el que se desarrolla el diseño estadístico, objeto de esta nota técnica.

Pasos o etapas de una investigación

Los componentes de una investigación se pueden agrupar en 4 grandes pasos o etapas que se explican a continuación.

a. Definir el problema central o la razón de ser de la investigación y sus objetivos

Esta etapa contempla la definición del problema a resolver o el por qué de la misma. En este caso debe centrarse en determinar cuál es el mercado meta para el futuro negocio. Una vez definido el problema central se establecen los objetivos para la investigación. Esto es clave, ya que uno o más objetivos mal planteados (imprecisos, no cuantificados, demasiado amplios) derivan en un plan de investigación errático.

El PROBLEMA CENTRAL y los OBJETIVOS para la investigación se centran, en este caso, en determinar cuál es el MERCADO META y cuál su dispersión.

b. El plan para la investigación

Consiste precisamente en definir el diseño estadístico que orientará la recopilación de la información a nivel de fuentes primarias. En esta etapa también se define el tipo o técnica de investigación a utilizar. A continuación se mencionan tres de los tipos de investigación que se pueden emplear.

- **ENCUESTA:** Se refiere a consultar a un grupo seleccionado de personas físicas o jurídicas sobre un tema en particular, y que es de interés para el(la) investigador(a) que se vale de esta técnica para recabar los datos. La herramienta que más se utiliza para recopilar la información de las personas entrevistadas es el cuestionario. La forma de abordar a las personas seleccionadas es por la vía presencial, telefónica o por correo electrónico.
- **OBSERVACIÓN:** Esta técnica se basa en recabar la información de las personas seleccionadas para este fin, mediante la observación del comportamiento y acciones previamente establecidas según los fines de la investigación. Tiene la particularidad de que esta observación se realiza a las personas durante su vida cotidiana. Se utilizan cámaras fotográficas o de filmación o simplemente una persona entrenada registra los movimientos y acciones de interés en el lugar



seleccionado, por medio de un formulario diseñado con ese fin.

- **EXPERIMENTACIÓN:** Consiste en someter a un grupo de personas a alguna prueba con características de prueba de laboratorio, en donde a los participantes se les somete a una rutina preestablecida y luego se les consulta sobre aspectos específicos de la experiencia vivida. Se acostumbra filmar y/o grabar el evento. Este es el caso de los Grupos focales o *Focus groups*.

En nuestro caso el tipo de investigación a utilizar con mayor frecuencia será la ENCUESTA y como instrumento, el cuestionario.

c. Implementación del plan para la investigación o trabajo de campo

Una vez definido el plan de la investigación se procede a llevarlo a la práctica. Esto es lo que se conoce como **trabajo de campo** y consiste en:

recorrir a las fuentes de información previamente definidas y ubicadas y, empleando la técnica de investigación seleccionada, proceder a recabar los datos de interés plasmándolos en el instrumento diseñado para este fin.

Recuerde que en la mayoría de los casos la técnica de investigación sería la encuesta y el instrumento para recabar los datos el cuestionario.

d. Procesamiento de la información recabada y elaboración del informe de la investigación

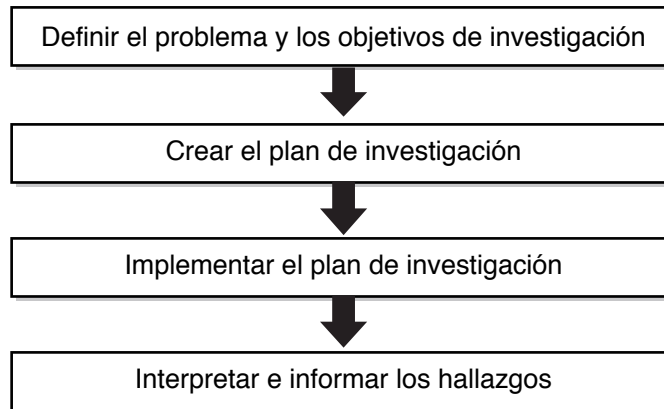
Este es el último paso de la investigación y consiste en procesar los datos recabados y presentarlos de una forma ordenada, lógica y entendible, como un informe de investigación. El procesamiento de los datos se puede hacer de forma manual o utilizar paquetes de *software* estadístico diseñados con el fin de facilitar la captura y el procesamiento de la información.

En la **figura 2** se destacan las etapas que componen una investigación.

Una vez explicada la ubicación del diseño estadístico en una investigación, se procede al desarrollo del tema central de esta nota técnica.



Figura 2. Pasos o etapas de una investigación



Basada en la figura *Pasos o etapas de una investigación* presentada en Kotler, P. y Armstrong, G. (1999). *Marketing*. Octava edición. Editorial Prentice Hall.

II. EL DISEÑO ESTADÍSTICO

Sin pretender esbozar una definición, sino más bien, establecer un concepto práctico sobre lo que es o representa un diseño estadístico, se dirá que se trata de **una guía para desarrollar una investigación en la que se detallan las diferentes partes y componentes de la misma, ordenados en una forma secuencial y lógica.**

No se debe perder de vista que la consulta a fuentes primarias como parte de la investigación, no es un fin en sí. Es una herramienta útil para recopilar información pertinente para resolver los objetivos relacionados o involucrados con la misma.

Componentes del diseño estadístico

La base fundamental de un diseño estadístico es el **planteamiento de objetivos** que se desprenden del **objeto o razón de ser de la investigación**. Por tanto, el paso o etapa previa al establecimiento del diseño estadístico, es tener muy claros los **objetivos de la investigación**. Estas premisas son las que delimitarán los alcances de la misma (diseño muestral y trabajo de campo, entre otros) y su razón de ser. Sin objetivos o sin claridad en la formulación de los mismos, la investigación puede caer en el riesgo de que se pierda el sentido del para qué de la información, o bien, recabar datos innecesarios o poco útiles para la investigación.

Una vez hecha esta advertencia se procede a presentar las partes o etapas componentes de un diseño estadístico.

- 1.1 Definición de la **población de interés**
- 1.2 Diseño del **marco muestral**
- 1.3 Determinación del **tamaño de la muestra**
- 1.4 Escogencia del **tipo de muestreo**
- 1.5 **Selección y distribución de la muestra**
- 1.6 Delimitación de la **unidad de muestreo**



1.7 Definición de la **unidad informante**

Estas son las partes componentes. De previo al desarrollo de las mismas es necesario tener presente tres aspectos relacionados con las consultas o pesquisas a nivel de fuentes primarias. Aún cuando éstas se derivan de los objetivos de la investigación, también tienen implícita su relación con la o las poblaciones de interés que serán objeto de estudio. Los aspectos son:

¿A quiénes se debe consultar? Se refiere a determinar la o las poblaciones de interés para la investigación. Pueden ser estudiantes, funcionarios de empresas, amas de casa, jefes de familia, informantes clave u otros.

a- ¿Qué se les debe preguntar? Consiste en la información que interesa recopilar de cada población de interés. Se puede consultar sobre los materiales didácticos, el mercado meta de la empresa, criterios para decidir sobre uno u otro producto de consumo masivo y otras áreas según sea la población de interés.

b- ¿Cómo y cuándo se les debe preguntar? Se refiere a la **técnica de investigación, el instrumento** a utilizar para recopilar la información y la **logística** para acceder al informante. Se puede emplear

como **técnica** de investigación la entrevista por medio de correo electrónico, por teléfono o en forma presencial. También es posible utilizar **las técnicas de observación y experimentación. El instrumento** puede ser un cuestionario o una guía de conversación. El **componente logístico** sería determinar dónde y cómo aplicar el instrumento. Puede ser en la casa de habitación, el lugar de trabajo, un lugar público o por medios electrónicos. También es posible utilizar otras técnicas de investigación más bien **cualitativas** como lo son desarrollar talleres y grupos focales.

Determinación de la(s) población(es) de interés

Al plantear u orientar la investigación, los objetivos de la misma encauzan, en la inmensa mayoría de los casos, hacia un grupo de elementos que a su vez se perfilan como las partes componentes de un todo. Este todo como veremos, será la población de interés, y sus componentes, los elementos en los cuales se desgrana esta totalidad que es la población. Se plantean dos conceptos básicos.

a- Elemento: Es la unidad mínima que integra o compone un todo. En ciencias sociales se refiere, por ejemplo, a cualquiera de los(as) siguientes: ama de casa, estudiante, empresa



del sector, profesor, jefe de familia, vehículo, profesional de una determinada especialidad, funcionario de una empresa, etc.

- b- Población:** Conjunto o grupo de elementos que presentan una o más características en común. Ejemplos de este segundo concepto son: los(as) estudiantes de la UNED, los(as) consumidores(as) de bebidas gaseosas, los(as) usuarios(as) del servicio remunerado de personas, los(as) funcionarios(as) de la empresa Global Aeronáutica, los(as) profesionales en ciencias médicas, las empresas maquiladoras, etc.

Se han definido una serie de poblaciones, sin embargo, se observa que las mismas son amplias y hasta confusas. **En una investigación en ciencias sociales esta definición de población no es útil.** ¿Por qué? Porque las poblaciones descritas son tan amplias que resultaría casi imposible trabajar con ellas, en cualquier sentido.

¿Qué se debe hacer para subsanar el inconveniente? Es necesario definir y delimitar claramente la población de interés. En este punto se deben recordar de nuevo los objetivos de la investigación. Los mismos son el norte y

referencia para resolver esta situación y cualquier otra que se presente en la investigación.

En el caso de la población de interés, se debe definir con claridad y respetando en todos sus extremos los objetivos planteados. Rescatando los casos utilizados como ejemplo, una definición de población de interés podría ser así (asumiendo que es congruente con los objetivos de la investigación):

- a- Estudiantes activos de la UNED en todo el país, que cursan el énfasis en Contaduría Pública y que obtuvieron el grado de bachiller en los años 2006 ó 2007.
- b- Consumidores(as) de bebidas gaseosas, que cursan la educación secundaria tanto en centros de estudio públicos como privados y que viven en las cabeceras de provincia.
- c- Usuarios(as) del servicio remunerado de personas en la modalidad de autobús, que viven en el Área Metropolitana de San José y que son mayores de 55 años (ciudadanos de oro).
- d- Funcionarios(as) de Global Aeronáutica que son graduados(as) en alguna rama de la Ingeniería.



- e- Profesionales de las ciencias médicas (independientemente del género) dentro de la modalidad de Enfermería, que laboran en hospitales privados.
- f- Empresas maquiladoras, que se dedican a la actividad textilera y que se ubican en alguna zona franca de la Gran Área Metropolitana.

Han sido presentadas las principales consideraciones relacionadas con la definición de la población de interés para la investigación. Este es el primer paso o etapa en la estructuración de un diseño estadístico. El siguiente es el marco muestral.

Elaboración del marco muestral

Una vez que se ha definido o delimitado la(s) población(es) de interés, se hace indispensable establecer lo que se conoce como marco muestral. **¿Qué es un marco muestral? Es un documento que reúne todos los elementos que componen la población de interés, diferenciándolos claramente uno de otro y permitiendo individualizarlos en un momento dado.** Esta denominación de “documento” se refiere a una lista, un mapa, un archivo digital, un plano o cualquier otro instrumento que resuma la totalidad de elementos de la población de interés.

En algunos casos, es menester que el (la) o los (las) investigadores (as) elaboren el marco muestral, en vista de que el mismo no existe como dato secundario o la información disponible está incompleta u organizada de una forma tal, que no es congruente con nuestros fines. Ha sido en casos como estos de donde surgió la expresión, ya popular entre los investigadores en ciencias sociales: *fue tanto más complicado hacer el marco muestral que el resto de la investigación*. En otros es tan sencillo como obtener un listado de un archivo digital o una fotocopia de un mapa o plano.

Ejemplos de marcos muestrales para una investigación, continuando con los ejemplos iniciados con la definición de poblaciones de interés, serían:

1. Listado emitido por la Oficina de Registro de la UNED de los estudiantes del énfasis en Contaduría Pública, que se mantienen activos y que obtuvieron su título de bachiller en los años 2006 ó 2007.
2. Con base en datos del ministerio de Educación, se enlistan los centros de estudio ubicados en el área de atracción de cada una de las siete cabeceras de provincia. Acto seguido se determina el número de estudiantes que están matriculados en cada centro de educación secundaria. Este dato se obtiene por medio del mismo Ministerio o consultando a cada uno de estos centros de forma directa. Tenemos un listado de colegios y el número de estudiantes en cada uno



de ellos. Otra forma de acceder a esta población de interés, sería a través de los hogares en la zona de atracción. Por medio de los segmentos censales¹ que elabora el Instituto Nacional de Estadística y Censos (INEC), se puede ubicar la zona bajo estudio y localizar en sus casas a los estudiantes.

3. En la oficina respectiva de la Caja Costarricense del Seguro Social se ubica un listado de personas calificadas como ciudadanos de oro y que viven en la zona de interés.
4. Este es un caso sencillo. En una empresa, se recurre a una lista de funcionarios que ostentan el título de ingeniero. Si se tratara del ICE serían cientos, pero en una firma independiente pueden ser unos cuantos. Este es el caso de una investigación en la que se deben contactar informantes clave. Estos son pocos y fáciles de ubicar.
5. Por medio del colegio de enfermeros(as) se puede obtener un listado. También es posible elaborarlo u obtenerlo directamente de cada hospital.
6. Por medio de cada zona franca en el área de interés, se obtiene esta lista. También se puede hacer de forma indirecta obteniendo un listado con entidades gubernamentales que rigen el sector, tales como COMEX y PROCOMER. Otra fuente para elaborar este marco muestral sería a través de cámaras y otras organizaciones que aglutinen empresas de esta actividad.

¹ División administrativa del país hecha por el INEC con fines censales. Todo el país se dividió para el censo de 1984 en unidades geográficas que abarcaban de 60 a 80 unidades habitacionales. Esta división facilita localizar barrios y viviendas dentro de estos, incluyendo en algunos casos la división en clases sociales de estos sectores.

Censo o muestreo

Una vez definida la población de interés y elaborado el marco muestral, el siguiente paso consiste en definir la amplitud de la consulta. En este punto se debe decidir sobre el número de elementos poblacionales a consultar.

¿Qué hacer, un censo² o abordar una muestra³? Los criterios básicos o primarios son dos: **tiempo para realizar la investigación y presupuesto disponible**. Si se cuenta con poco tiempo y/o los recursos son exiguos, es mejor inclinarse por investigar una muestra de un tamaño y distribución que se pueda estudiar dentro de las limitaciones que plantean estos dos criterios. Pero, ¿qué hacer cuando tenemos recursos suficientes y el tiempo necesario?; más aún, ¿qué hacer aún cuando la población es “pequeña” y está focalizada? En general, los parámetros para decidir, adicionales a los dos anteriores, son los siguientes:

- a- **Tamaño de la población:** Se debe sopesar qué tan “grande” o “pequeña” es la población. Por ejemplo, 60 empresas de uno u otro sector puede ser una población “grande”, mientras que 500 estudiantes de una determinada institución, una más bien “pequeña”.

² Un censo es una investigación de fuentes primarias en la cual se consulta a todos los elementos de la población de interés sobre un tema prefijado.

³ Una muestra es una parte de la población de interés que se obtiene por algún mecanismo estadístico probabilístico o no probabilístico.



b- Distribución de la población: Paralelo al tamaño, la distribución es el complemento de este criterio para decidir. ¿Está la población concentrada en el Valle Central o en el Área Metropolitana de San José? ¿Se encuentra dispersa a lo largo y ancho del país? ¿Está concentrada en una ciudad o un sector de la misma?

c- Homogeneidad de la población: ¿Qué tan parecidos son sus elementos? Se trata de una población con pocas diferencias o, por el contrario, aún dentro de una misma actividad existen diferencias claras (entre regiones del país, por ejemplo).

Los criterios de tiempo y presupuesto sumados a los tres anteriores, relacionados propiamente con las características peculiares de cada población, **son la base para decidir entre realizar un censo o aplicar un muestreo.** Si una población está concentrada en el Valle Central, es pequeña, accesible y muy homogénea, el investigador se podría inclinar por hacer un censo. Sin embargo, si la población es tan homogénea, pequeña y accesible, ¿no es mejor trabajar con una muestra?

En el caso de las ciencias exactas la decisión es más sencilla y focalizada. Si las poblaciones son muy homogéneas, una pequeña muestra es suficiente para inferir a

toda la población. En el caso típico de la sangre, se sabe que de una simple gota se puede, incluso, extrapolar a los 5 litros que componen el torrente sanguíneo de una persona adulta. ¿Cuántos bombillos incandescentes deben someterse a prueba para evaluar todo un lote de producción? La respuesta es unos cuantos, podrían ser de 5 a 10 unidades. En estos ejemplos el uso de muestras muy pequeñas también se justifica porque la muestra, al someterla a examen, se destruye.

En ciencias sociales se trabaja con seres humanos, con las connotaciones que esto implica. Esta es una razón para utilizar muestras incluso en el caso de poblaciones pequeñas, ya que entre los seres humanos “cada cabeza es un mundo”.

Se anotan a continuación algunas de las **situaciones en las que aplicar un censo es viable**, dadas las diferentes poblaciones de interés. Estos ejemplos ayudan a ubicar cuándo proceder con un censo y en cuáles casos recurrir a alguna técnica de muestreo.

- 1) Se va a evaluar el control interno en una empresa comercial, destacando los puntos de control, los procedimientos, los formularios utilizados y los funcionarios involucrados. En este caso se justifica entrevistar o consultar a **todos** los informantes



clave y, de ser necesario, a todos los funcionarios involucrados en los procesos. ¿Quiénes son funcionarios clave? El gerente general, el encargado del área de Contabilidad, los jefes de área o de operaciones y por último, los restantes funcionarios que participan en el proceso. La diferencia en este caso estribaría en que, a los informantes clave se les aplicaría una guía de conversación, mientras que a los restantes funcionarios se les consultaría con un cuestionario estructurado.

- 2) Una empresa fabricante de un producto de consumo masivo cuenta con 3 empresas que se dedican a la distribución del 100% de su producto. Se desea conocer las razones por las cuales una de ellas ha desmejorado su ejecutoria, mientras que las otras dos no han logrado cubrir lo que la tercera ha perdido. Dentro de las fuentes primarias a consultar se tendría de nuevo a informantes clave de las 3 empresas. Estos serían los gerentes, encargados de ventas y los vendedores. Se consultaría a **todos los involucrados**. Podría establecerse alguna salvedad en el caso de los vendedores, ya que, supongamos, podría ser una cantidad difícil de manejar; 80 ó 100 personas. Se recurriría en este caso a censar a los que consideramos informantes clave, recurriendo a una

muestra para consultar a los vendedores. En este caso se estaría atendiendo dos poblaciones diferentes.

- 3) Se necesita conocer las diferencias y similitudes en las políticas para el reclutamiento y selección de personal entre dos entidades bancarias. En este caso es pertinente entrevistar a los encargados del área de Recursos Humanos de cada banco, así como a las personas responsables de estos procesos de reclutamiento y selección. En total serían unas 3 a 5 personas en cada institución. **Se haría un censo de la población de interés.**

En estos 3 casos se trata de poblaciones de interés pequeñas y focalizadas. Esto permite inclinarse por censar a los elementos de la o las poblaciones.

Aplicación de las técnicas de muestreo

Lo común en ciencias sociales es trabajar con muestras, pero, **¿cómo calcular el tamaño de esa muestra de manera que sea representativa?** Si la muestra no lo es, cualquier dato obtenido de ella queda como un simple sondeo. Esto es, una cifra para la cual tenemos un margen de error "alto"⁴, o peor aún, no es posible siquiera calcular

⁴ En ciencias sociales se considera que un margen de error superior al 5% es alto, y por lo tanto los resultados se consideran como un sondeo.



el error asociado a la inferencia estadística. Es por estas razones que el investigador debe velar para que el tamaño de la muestra sea representativo.

Caso de poblaciones infinitas y finitas

El primer paso para alcanzar este logro se centra en **definir si la población de interés es finita o infinita en el plano estadístico**. Se presenta antes, un concepto de cada tipo de población.

Población infinita. En ciencias sociales una población se considera o podría considerarse como infinita, más por conveniencia para una investigación que por estar constituida por un cierto número de elementos. Se debetener presente que en el plano matemático, astronómico, científico y, en general dentro de las ciencias exactas, el infinito es algo interminable. A nivel de las ciencias sociales no existe una población infinita con esta connotación. La razón básica es que lo humano no es compatible con el infinito absoluto.

El calificar una de las poblaciones de interés como infinita, es más bien un subterfugio estadístico para facilitar el cálculo de un tamaño de muestra que sea representativo. Ya que, si a lo anterior se le suma que **las poblaciones en ciencias sociales generalmente presentan**

una distribución normal, basta con agregar un nivel de confianza y definir el área bajo la curva a ambos extremos de la misma que se excluye del valor de interés (este valor es conocido como zeta alfa medios), para procesar la fórmula resultante y obtener el tamaño de la muestra.

Pero el objetivo de este documento no es teorizar sobre el fundamento estadístico y matemático de los fenómenos de interés. En la literatura disponible se pueden revisar en detalle y con toda propiedad, estos conceptos que fundamentan las fórmulas y criterios que posteriormente serán expuestos para solventar el cálculo del tamaño de una muestra.

El concepto de población infinita es más bien una referencia teórica, ya que a partir de este concepto y dada una población de interés, es posible agilizar el cálculo del tamaño de muestra; **así como decidir si la población es más bien finita, de nuevo, en el plano meramente estadístico.**

Población finita: “Si el tamaño de muestra resultante representa 10% o más de la población, debe aplicarse la corrección de la población finita” (Malhotra, 2004: 349). Este concepto extraído de la literatura ejemplifica las últimas líneas esbozadas para la población infinita. Pero, ¿qué significa, o, qué implicación tiene al calcular



el tamaño de una muestra? El impacto se centra en determinar, por este medio, si la población de interés es infinita o en su defecto finita en el plano estadístico.

Si al calcular el tamaño de una muestra para una población determinada, resulta que ese tamaño de muestra es igual o superior al 10% de la población, la misma se debe considerar como una población finita, y por lo tanto debe reorientarse el enfoque relacionado con la magnitud de la misma.

En un inciso posterior, se detallan las fórmulas para el cálculo del tamaño de muestra para cada tipo de población. De previo, y antes de proceder, se debe incluir un último componente involucrado en estos cálculos. El mismo se centra en los parámetros a incluir en la fórmula. Si revisamos la literatura, las fórmulas para el cálculo del tamaño de muestra incluyen, como dato disponible, **la desviación estándar y la varianza de la población o de un dato muestral obtenido de previo**. La disponibilidad de estos datos en las investigaciones, es casi imposible. En la inmensa y abrumadora mayoría de los casos no se cuenta con este antecedente. Es por esta razón, que en el proceso de cálculo del tamaño de muestra requerimos de un último componente, **el cual se denomina criterio de las proporciones**.

Criterio de las proporciones

Entonces, en las ciencias sociales en general y en las investigaciones en particular, es raro y escaso que se cuente con antecedentes útiles (promedio, varianza, desviación estándar) de investigaciones desarrolladas con anterioridad, sobre el tema de interés. Debido a lo anterior, para el cálculo de tamaños de muestra se debe recurrir al **criterio de las proporciones, también conocido como la probabilidad de éxito o fracaso según sea que presente o no la(s) característica(s) de interés en la investigación**. Sumado a lo anterior, se debe apuntar que la experiencia comprueba que las poblaciones, en vista de que se trabaja con personas, presentan una distribución normal. Esto facilita el cálculo de tamaños de muestra. Lo que varía en estos casos es el nivel de confianza, reflejándose su impacto en la fórmula a través del componente “z” (zeta alfa medios).

Amparados en este criterio, se determinan de previo estas probabilidades de éxito y fracaso para luego proceder al cálculo del tamaño de muestra. Hay un pero. Este es **¿cómo determinamos esta probabilidad de éxito y fracaso? ¿en qué nos basamos?** Tenemos tres formas de hacerlo.



Antes de este detalle se debe tener presente que estas probabilidades de éxito o fracaso son complementarias, por tanto, al determinar una se obtiene la otra por diferencia.

a- Investigación exploratoria: Antes de cualquier cálculo, se recurre a un grupo reducido de elementos de la población de interés, empleando un criterio de conveniencia o juicio de experto y se les consulta sobre la característica de interés. De estas respuestas y su forma, resulta la probabilidad de que esta(s) característica(s) se presente(n) en la muestra (probabilidad de éxito). Esta opción se debe dejar como último recurso, y una vez que no ha sido posible resolver el asunto por medio de las otras dos vías que a continuación se presentan. La práctica común es consultar al menos 10 elementos de la población.

b- Opinión de un experto(a): Es perfectamente válido que el investigador recurra a un experto en la materia para decidir sobre estas probabilidades. De hecho, si el investigador lo es, procede a la fijación de la probabilidad de éxito y acto seguido calcula el tamaño de la muestra. Este experto en la mayoría de los casos surge dentro de la orga-

nización para la cual se está desarrollando la investigación.

c- Se da por un hecho: En algunos casos esta probabilidad es obvia y no se requiere de asesoría o consejo de experto, ni de consulta previa con este fin, a la población de interés. Se procede a definirla de forma directa.

La probabilidad de éxito debe ser alta. Esto es 90% o más de los casos. **¿Por qué debe ser alta?** Porque si tuviéramos una probabilidad intermedia es un indicador de que la definición inicial de objetivos no fue acorde. También podría darse que la definición de la población de interés no se hizo de forma concreta. En ambos casos el resultado es una situación de incertidumbre o de indefinición que provoca dificultades para el cálculo o fijación de estas probabilidades y el dato en última instancia, va a ser intermedio, lo que provoca la necesidad de una muestra “muy grande” para investigar la población. Esto implica un costo elevado y/o un período de tiempo prolongado para recabar la información.

A continuación, un ejemplo. Intento de que el estudiante evalúe la unidad didáctica utilizada en un determinado curso de la carrera de Contaduría. Si se define como la población de interés a los estudiantes de la UNED que



estudian la carrera de Contaduría Pública, que se encuentran como estudiantes activos y que obtuvieron su título de bachillerato en los años 2006 ó 2007; queda claro quiénes son y dónde están. Las características de interés serán diáfanos. En el otro extremo, si la población la definimos así: estudiantes de la UNED que cursan el nivel de licenciatura de cualquier énfasis de la Escuela de Ciencias de la Administración, ¿qué puede pasar? Varias situaciones. Que un porcentaje no haya llevado el curso porque no forma parte de su plan de estudios, que los estudiantes de otros énfasis, que no tienen el concepto integral de la carrera de Contaduría, no opinen o que sus opiniones sean vagas. Estas situaciones van a desembocar en una muestra grande para que el dato final sea representativo.

En este segundo caso se refleja cómo la probabilidad de éxito se torna ambigua y más bien cercana al 50%, cifra que desacredita la calidad de la investigación, o como se indicó, obliga a trabajar con una muestra grande y sus implicaciones.

Ejemplos de poblaciones finitas e infinitas

Para un investigador una población puede ser infinita mientras que otro, perfectamente, puede darle un trato de población finita. Hay casos en los que la delimitación de la población de interés resulta hasta obvia. La base científica

que usted puede utilizar como argumento para diferenciar la extensión de una población, es la que se expuso al inicio de este inciso y que se resume en la cita del autor Narres K. Malhotra: “Si el tamaño de muestra resultante representa 10% o más de la población, debe aplicarse la corrección de la población finita” (Malhotra, 2004: 349). Esto significa, en otras palabras, que si se presenta la situación descrita por Malhotra, la población debe considerarse como finita.

La diferenciación sobre el tamaño de la población en muchos casos, como ya se indicó, es un subterfugio estadístico para reducir el tamaño de muestra. Ya que al asumir que una población es infinita, el tamaño de la muestra en la inmensa mayoría de los casos, resulta inferior que si se asume a ese mismo conglomerado como población finita. Este fenómeno tiene una explicación matemática relacionada con el área bajo la curva de la distribución normal de los valores muestrales y poblacionales, que no se van a abordar en este escrito.

Algunos ejemplos se muestran a continuación.

Poblaciones infinitas típicas

- a- Amas de casa con o sin hijos, de entre 25 y 45 años, que viven en el Área Metropolitana de San José.



- b- Total de estudiantes matriculados en la UNED durante el primer cuatrimestre del 2007.
- c- Funcionarios del Instituto Costarricense de Electricidad (ICE) que laboran en cualquier lugar del país.
- d- Clientes del Banco Nacional que mantienen cuentas de ahorro activas.

Poblaciones finitas típicas

Funcionarios que desempeñan puestos de jefatura en una empresa.

- a- Pacientes de un determinado dentista.
- b- Número de patentes comerciales extendidas por la municipalidad de Buenos Aires y que se mantienen vigentes.
- c- Número de vendedores de una distribuidora de productos de consumo masivo.

Tal y como se advirtió, un investigador puede revertir algunos de estos ejemplos y pasar de población finita a infinita o viceversa. Lo crucial es la justificación que se presente para sustentar la postura.

Tamaño de la muestra

Una vez perfilados los componentes que se presentaron en incisos anteriores, lo que resta es repasar las fórmulas y procedimientos para el cálculo del tamaño de muestra. Pero, ¿qué es una muestra y a qué nos referimos con tamaño de muestra?

Para el proceso cuantitativo la muestra es un subgrupo de la población de interés (sobre el cual se recolectarán datos, y que tiene que definirse o delimitarse de antemano con precisión), éste – se refiere al subgrupo- deberá ser representativo de la población. (Hernández Sampieri, 2006: 236)

Queda claro que la muestra debe ser una parte representativa de la población de interés, cuyo tamaño se calcula por medio de fórmulas válidas y diseñadas según se trate de una población finita o infinita.

Tamaño de muestra para poblaciones infinitas

Si la población de interés es o se considera **infinita**, se utiliza la siguiente fórmula para calcular el tamaño de la muestra. Recuerde que la misma se centra o contempla el criterio de las proporciones.



$$n = \frac{n^2 \times p(1 - p)}{e^2}$$

Donde: n : tamaño de la muestra
z : zeta alfa medios
p: probabilidad de éxito
e: error de muestreo

En caso de que al calcular el tamaño de la muestra utilizando el criterio de que la población es infinita, el mismo resulte igual o mayor al 10% de esa población, como se indicó en un inciso anterior; se le debe considerar como una población más bien finita; por tanto, se debe recurrir a la fórmula respectiva.

En la Tabla 1, al final de esta nota, se resume el cálculo de diversos tamaños de muestra, con un nivel de confianza del 95% y según varios márgenes de error y de probabilidades de éxito o fracaso.

Tamaño de muestra para poblaciones finitas

Si la población se considera **finita**, debe usarse la fórmula que a continuación se presenta.

$$n = \frac{p(1 - p)}{\frac{e^2}{z^2} + \frac{p(1 - p)}{N}}$$

En esta fórmula el único componente nuevo es la “N” que representa el tamaño de la población.

Cuando el nivel de confianza es del 95% (este es un nivel de confianza de uso común entre los investigadores), el valor de “z” es igual a 1,96. Este dato aplica para cualquiera de las dos fórmulas detalladas.

Tipos de muestreo

Una vez que se tiene el tamaño de la muestra, el siguiente paso consiste en definir el tipo de muestreo que utilizaremos para seleccionar y distribuir la muestra entre los elementos de la población. **Observe usted como queda implícito que el tipo de muestreo y la selección y distribución de la muestra son pasos secuenciales y ligados entre sí.** Al definir uno queda delimitado el otro.

Para estos fines se cuenta con **dos tipos de muestreo, el probabilístico y el no probabilístico.** La diferencia, y ésta es fundamental, estriba en la representatividad de los resultados. Si se utiliza alguna de las variantes del **muestreo probabilístico**, cada elemento de la población tiene una **probabilidad conocida** de ser seleccionado en la muestra. Esto implica que cualquiera de los elementos poblacionales podría ser incluido en la muestra.



En el caso de los tipos de **muestro no probabilístico** esta máxima no se cumple, ya que los elementos de la población no tienen esa probabilidad conocida; más aún, **un porcentaje de ellos ni siquiera es considerado entre los elegibles para integrar la muestra.**

Las variantes de cada tipo de muestreo son las siguientes.

Muestreo probabilístico	Muestreo no probabilístico
Simple al azar	Conveniencia
Estratificado	Juicio de experto
Sistemático	Por cuotas
Por conglomerados	

Manteniendo la orientación de este documento, no se pasa a definir o argumentar sobre estas variantes de los tipos de muestreo ya que no es el objetivo del mismo.

Siguiendo con el tema, no debe usted concluir que el muestreo probabilístico es mejor o superior que el no probabilístico. De hecho, son, una vez más, la naturaleza y los objetivos de la investigación quienes dictan el derrotero a seguir al decidir sobre el tipo de muestreo.

A continuación algunos ejemplos:

a- Una empresa desea obtener información de entre sus miles de clientes, dando preferencia a los más representativos o importantes. Asumiendo que el 100% de los clientes está clasificado con base en variables de interés, como podrían ser volumen de compras, exactitud en los pagos y/o ubicación geográfica del cliente; ¿cómo se determina a quién consultar si lo que se desea es muy puntual y enfocado a un grupo de estos clientes? ¿Se apega el investigador al rigor estadístico o se inclina por una selección orientada de los elementos a consultar? Remítase a la página donde se definen los clientes por categoría.

Este es un caso en el que la empresa interesada debe pensar más en ubicar los clientes clave (categoría A) a través de la opinión de un experto, que en aplicar procesos probabilísticos. El experto va a indicar cuántas y/o cuáles son las empresas clave para obtener información pertinente. Pero, ¿quién es el experto, y por qué no hacerlo por conveniencia que es aún más directo y expedito? El experto puede surgir de diferentes orígenes, a saber:

- El más común es un funcionario de la misma empresa que conoce el devenir de la misma.
- Otro sería un conocedor de la industria en la cual se desenvuelve la empresa u organización.
- Una tercera fuente sería un investigador ocasional que, con base en su experiencia como tal, organiza la



investigación y define él mismo sobre la muestra y a quién consultar.

Aplicar el muestreo por conveniencia sería válido, sin embargo, si se apega a los clientes más cercanos en el plano geográfico y/o fáciles de acceder y/o que cuentan entre sus funcionarios con amigos del interesado; se puede dejar de lado clientes representativos. Esto deriva en una muestra poco fructífera. El caso en que el muestreo por conveniencia es muy utilizado, se presenta en aquellas situaciones en que se desea delimitar o determinar las probabilidades de éxito o fracaso, previo al cálculo del tamaño de la muestra.

b- La misma empresa del ejemplo anterior **desea ahora consultar entre todos sus clientes** sobre la atención que recibe de la fuerza de ventas y el servicio al cliente. Esta situación es muy diferente y seguimos involucrando a los miles de clientes de la firma. En este caso predominaría el criterio de calcular un tamaño de muestra representativo y de utilizar un tipo de muestreo probabilístico, presumiblemente el estratificado, para garantizar que la muestra se distribuya con un criterio de proporcionalidad, entre los estratos o grupos de clientes.

Usted debe tener presente que optar por un tipo de muestreo probabilístico no representa, por definición, una decisión correcta. Tampoco por usar un muestreo no

probabilístico se va a desacreditar la investigación y sus resultados. Puede ser tan válido uno u otro tipo de muestreo. La base que sustenta la decisión es la clave y los objetivos de la investigación el norte que se debe tener presente en cada acción que se emprenda.

Selección y distribución de la muestra

Tal y como se mencionó en el inciso anterior, la selección y distribución de la muestra está aparejada con la definición del tipo de muestreo. Una acción lleva a la otra y, de hecho, la define.

¿En qué consiste esta acción? Recuerde que, a este punto, sobre la investigación ya se definió lo siguiente: La población de interés, el marco muestral, si se va a realizar un censo o se va a trabajar con una muestra y si fuera este segundo el caso, se define el tamaño de la muestra y por último se ubica el tipo de muestreo a utilizar. Como efecto inmediato del tipo de muestreo, resulta el procedimiento para seleccionar la muestra, acción con la cual queda, a su vez, definida la distribución de la misma.

Se repasa a continuación, en qué consiste este procedimiento. Se recurre al caso mencionado en el inciso anterior, el de la empresa que cuenta con miles de clientes. Se considera el segundo escenario de investigación,



relacionado con la evaluación de la atención que recibe de la fuerza de ventas y el servicio al cliente. En el Cuadro 1 se detallan datos de interés, los cuales serán la base para nuestro desarrollo hipotético del caso práctico.

Cuadro 1. Distribución de los clientes actuales de la empresa XYZ, según la categoría de cliente.
Al 30 de junio de 2007

Categoría de cliente	Número de clientes	Participación porcentual
A	108	1,28
B	723	8,58
C	1 398	16,58
D	2 119	25,14
E	4 082	48,42
Total	8 430	100,00

Fuente: Registros internos de la empresa XYZ (esto es hipotético, se incluye para ejemplificar este componente)

¿Esta población es finita o infinita? Al partir de los criterios planteados en su oportunidad, si se asume que la población es infinita, se tienen diferentes tamaños de muestra representativos, calculados con los siguientes parámetros:

Parámetro	Tamaño de muestra 1	Tamaño de muestra 2	Tamaño de muestra 3
Nivel de confianza (%)	95	95	95
Probabilidad de éxito (%)	95	95	95
Error de muestreo (%)	5	4	3
Tamaño de la muestra	73	114	203

Se observa que en cada uno de los tres casos, el tamaño de muestra es inferior al 10% de la población, incluso ninguno llega siquiera al 2,5 % de la misma.

El criterio más adecuado para distribuir el tamaño de la muestra, apegándonos al criterio del muestreo probabilístico, es utilizar la modalidad del estratificado. Tenemos



claramente diferenciada la población según categoría de cliente. Cada una de estas categorías corresponde a un estrato, ya que, y aún cuando todos los elementos pertenecen a una misma población, los elementos de cada estrato tienen características en común que no presentan los miembros de los otros estratos o subgrupos. El procedimiento para distribuir la muestra, que se basa en el peso porcentual de cada estrato para asignarle ese mismo porcentaje del tamaño de muestra, es el siguiente:

Cuadro 2. Distribución de la muestra según la categoría de cliente. Al 30 de junio de 2007

Categoría de cliente	Número de clientes	Particip. porcen.	Distribución tamaño muestra 1	Distribución tamaño muestra 2	Distribución tamaño muestra 3
A	108	1,28	1	2	3
B	723	8,58	6	10	17
C	1 398	16,58	12	18	34
D	2 119	25,14	18	29	51
E	4 082	48,42	36	55	98
Total	8 430	100,00	73	114	203

Fuente: Registros internos de la empresa XYZ y elaboración propia.

Con este paso queda distribuido el tamaño de muestra, según cada uno de los tres márgenes de error evaluados. Estos tres escenarios se desarrollan con fines ilustrativos, ya que en una encuesta se calcula un solo tamaño de

muestra, que sea acorde al presupuesto disponible y a las expectativas de representatividad de la investigación.

¿Cómo se hace la selección de la muestra? Se tiene claro el total de elementos a seleccionar dentro de cada categoría de cliente, resta definir a quiénes o cuáles escoger. Tenemos dos vías, a saber:

- **Seleccionar por medio del muestreo simple al azar** tantos elementos del estrato como la asignación porcentual indica. Por ejemplo, en el caso del tamaño de muestra cuando el error de muestreo es igual a 4%, se deben seleccionar 2 clientes de la categoría A y 18 de la categoría C. El procedimiento consiste en asignarle a cada elemento del estrato un número. Luego, con base en la tabla de números aleatorios, se seleccionan los elementos a ser entrevistados. **Otra forma** de hacer la selección sería aplicando **el muestreo sistemático**, una vez que todos los elementos del estrato han sido numerados. En estos dos casos se está empleando un criterio totalmente aleatorio para distribuir y seleccionar los elementos a ser consultados.
- Otra fórmula sería seleccionar los elementos de cada estrato utilizando criterios no probabilísticos. Este mecanismo contempla utilizar el criterio de experto o la conveniencia para decidir sobre cuáles clientes serán accesados, dentro de cada estrato, respetando, eso sí, el tamaño de la muestra asignado a ese subgrupo.



En el primer ejemplo, en el que solamente nos enfocamos en los clientes tipo A, la escogencia de las empresas se hace de una forma orientada o no probabilística y entre los 108 clientes que conforman esa categoría.

Unidad de muestreo y unidad informante

Estos dos conceptos se presentan de forma conjunta y van aparejados con un fin común. Por unidad de muestreo se entiende, en la mayoría de los casos, todos y cada uno de los elementos de la población de interés, y sobre los cuales va a recaer la selección de la muestra. Es aquella referencia sobre la cual se escogen los elementos de la población a ser consultados. En un porcentaje de los casos, la unidad de muestreo es igual al elemento de la población. En algunas investigaciones cada elemento de la población constituye una unidad de muestreo.

En otras investigaciones la unidad de muestreo y el elemento de la población son diferentes. Esto significa que para fines de seleccionar y distribuir la muestra, hay que tener presente la unidad de muestreo, sin embargo ésta no es la que brindará la información que se requiera.

Los datos requeridos por medio del instrumento desarrollado para ese fin, y empleando la técnica de investigación seleccionada, son facilitados, en última instancia, por

la unidad informante. Este componente del proceso de investigación es, por tanto, quien brindará la información que se solicite.

A continuación se presentan unos ejemplos sobre casos en los que se tipifican tanto la unidad de muestreo como la unidad informante.

Unidad de muestreo	Unidad informante
La empresa	El(la) funcionario(a) clave
La casa de habitación	El jefe de familia
La casa de habitación	El ama de casa
La casa de habitación	El(la) hijo(a) adolescente
El centro de estudio	El(la) profesor(a)
El centro de estudio	El(la) alumno(a)
El punto de venta	El(la) dueño(a) o encargado(a)
La línea aérea	El(la) pasajero(a)

En estos ejemplos, cada línea se debe interpretar como una investigación independiente, aún cuando la unidad de muestreo es la misma en varios casos.

COROLARIO

Con este punto finaliza este escrito que pretende aclarar aspectos prácticos al definir e implementar un diseño



estadístico para una investigación cuantitativa en ciencias sociales. No es una receta de uso universal. Más bien se refiere al planteamiento de una serie de pasos e interrogantes que usted debe plantearse al planear una pesquisa a nivel de fuentes primarias.

En este documento no se incluyeron conceptos teóricos, ya que los mismos se pueden consultar en la literatura disponible. La intención aquí fue desarrollar aspectos prácticos que le serán de utilidad al diseñar su propia investigación.

Tenga siempre presentes los objetivos de la investigación al decidir sobre cualquier aspecto subordinado con la misma.

BIBLIOGRAFÍA

Brown, T. (1935). The use of statistical techniques in market research. Harvard University Graduate School of Business Administration. Business Research Studies, 12.

Gómez, M. (1985). *Elementos de estadística descriptiva*. Costa Rica: EUNED.

Hernández, S. (2006). *Metodología de la investigación*. Cuarta edición. México: McGraw Hill.

Kinnear, T. (1981). *Investigación de mercados: un enfoque aplicado*. México: McGraw Hill.

Kotler, P. y Armstrong, G. (1999). *Marketing*. Octava edición. México: Prentice Hall.

Malhotra, N. (2004). *Investigación de mercados un enfoque aplicado*. Cuarta edición. México: Pearson Educación.

Weirs, R. (1986). *Investigación de mercados*. México: Prentice Hall.



Reconocimiento, No comercial,
Compartir bajo la misma licencia (3.0 Unported)



TABLA 1. Determinación del tamaño de la muestra

Caso de poblaciones infinitas

(El tamaño preciso de la muestra para garantizar dentro del 95% que los resultados de la encuesta se encuentran dentro de un número permisible de puntos de porcentaje, en cualquiera de los sentidos del valor real)

Probabilidades de éxito o fracaso

Porcentaje permisible de error en ambos sentidos.	5/95%	10/90%	15/80%	20/80%	27/75%	30/70%	35/65%	40/60%	45/55%	50/50%
0.5%	7296	13824	19584	24576	28800	32256	34944	36864	38016	38400
1.0%	1824	3456	4896	6144	1200	8064	8736	9216	9504	9600
1.5%	812	1536	2176	2732	200	3584	3883	4096	4224	4266
2.0%	456	864	1224	1536	800	2016	2184	2304	2376	2400
2.5%	292	552	784	984	1152	1292	1400	1476	1520	1536
3.0%	203	384	544	683	800	896	970	1024	1056	1066
3.5%	149	282	400	501	587	658	713	752	715	783
4.0%	114	216	306	384	450	504	546	576	394	600
4.5%	90	170	242	304	355	398	431	455	469	474
5.0%	73	138	196	246	288	323	350	369	380	384
10.0%						81	88	92	95	96

Adaptado del libro de Theodore H. Brown de 1935, *The use of statistical techniques in market research*. Harvard University Graduate School of Business Administration. Business Research Studies,12.

* Basado en la fórmula $n = \frac{3.84 p \cdot q}{e^2}$ (p y q igual al 100 %).

Si un porcentaje se escoge para p, el otro u otros factores restantes serán iguales a q.