

Instituto Nacional de Estadística y Geografía

***Diseño de la muestra
en proyectos de encuesta***



**INSTITUTO NACIONAL
DE ESTADÍSTICA Y GEOGRAFÍA**

Obras complementarias publicadas por el INEGI sobre el tema:

Presentación de datos estadísticos en cuadros y gráficas (edición 2011), Captación en encuestas por muestreo (edición 2010), Captación en registros administrativos (edición 2010), Diseño Conceptual para la generación de estadística básica (edición 2010), Proceso estándar para el aprovechamiento estadístico de registros administrativos (edición 2011), Proceso estándar para encuestas por muestreo (edición 2010).

Catalogación en la fuente INEGI:

310.01 Instituto Nacional de Estadísticas y Geografía (México).
Diseño de la muestra en proyectos de encuesta / Instituto Nacional de Estadística y Geografía. -- México : INEGI, c2011.

vii, 22 p. : il.

1. Estadística - Metodología.

Si requiere más información sobre esta obra, favor de contactarnos a través de:
Centros de consulta y comercialización (consulte el domicilio en Internet)
Larga distancia sin costo: 01 800 111 4634
www.inegi.org.mx
atencion.usuarios@inegi.org.mx

Presentación

De acuerdo con la Ley del Sistema Nacional de Información Estadística y Geográfica (LSNIEG), el **Instituto Nacional de Estadística y Geografía (INEGI)**, dentro de sus funciones “(...) regulará, mediante la expedición de disposiciones de carácter general, la captación, procesamiento y publicación de la Información, para el debido funcionamiento del Sistema o autorizará las que utilicen las Unidades para tales efectos” (Art. 58).

Con el fin de dar cumplimiento al mandato de Ley, el **INEGI** impulsa un programa de desarrollo normativo para la Coordinación del Sistema Nacional de Información Estadística y Geográfica (SNIEG), el Servicio Público de Información y la regulación técnica sobre los procesos de generación, integración, difusión y conservación de la información estadística y geográfica.

La determinación de la normatividad técnica para el Sistema, constituye una estrategia esencial para regular la operación de proyectos, estandarizar los procesos y homologar los conceptos, lo que permitirá avanzar en el mejoramiento de la calidad de los datos y la eficiencia del servicio público que da respuesta a la demanda social de información.

Por ello, con fundamento en el Artículo 30 de la Norma Técnica para la Generación de Estadística Básica publicada en el *Diario Oficial de la Federación* el 12 de noviembre de 2010, el **INEGI** pone a disposición de las Unidades del Estado del SNIEG el documento ***Diseño de la muestra en proyectos de encuesta***, como parte de la serie de documentación técnica y metodológica que elaborará y proporcionará con el fin de facilitar el cumplimiento de la norma.

Este documento tiene el propósito de facilitar las decisiones relacionadas con la determinación del esquema de muestreo, la determinación de la muestra, los factores de expansión, así como el cálculo de precisiones y estimaciones.

Con el fin de mejorar su contenido, se ha previsto la actualización periódica del documento, por lo cual el Instituto agradecerá los comentarios y sugerencias que contribuyan a ese propósito.

Índice

| | |
|--|-----|
| Introducción | VII |
| 1. Generalidades sobre el diseño de la muestra | 1 |
| 1.1 El diseño de la muestra como fase de un proceso | 1 |
| 1.2 Conceptos básicos | 2 |
| 1.3 Condicionantes del diseño de la muestra | 3 |
| 2. Opciones para la determinación del esquema de muestreo | 5 |
| 2.1 El muestreo probabilístico | 5 |
| 2.2 El muestreo no probabilístico o determinístico | 6 |
| 2.3 Muestreo multietápico | 7 |
| 3. Macroactividades del diseño de la muestra mediante muestreo probabilístico | 9 |
| 3.1 Elección del tipo de muestreo | 9 |
| 3.2 Preparación del marco de muestreo | 9 |
| 3.3 Determinación de la muestra | 10 |
| 3.4 Definición de estimaciones | 13 |
| 4. Documentación del diseño estadístico de la muestra | 15 |
| Glosario | 17 |
| Bibliografía | 22 |

Introducción

Entre la diversidad de actividades que comprende un proyecto de generación de información estadística, se encuentran las de diseño de la muestra, orientadas a la captura, codificación, validación, integración de la base de datos, así como la explotación de ésta para la presentación de resultados. Este conjunto de actividades impacta tanto en la calidad como en la oportunidad de los resultados.

En un proyecto de encuesta, el diseño de la muestra es crucial debido a su efecto en la confiabilidad de los resultados. Ésta se relaciona con otras fases previas como son: la planeación y el diseño conceptual, así como con las subsecuentes: la captación, el procesamiento y la presentación de resultados. En el presente documento se describen los elementos técnicos a considerar durante el proceso de decisiones que involucra esa fase, incluyendo los condicionantes en el marco del proyecto y las opciones a tener en cuenta.

Para ello se desarrollan cuatro capítulos, donde se plantean los aspectos generales; las opciones para determinar el esquema de muestreo, las macroactividades que cubre y un apartado sobre la documentación de las actividades de la fase. Se incluye también un glosario y la bibliografía consultada.

1. Generalidades sobre el diseño de la muestra

1.1 El diseño de la muestra como fase de un proceso

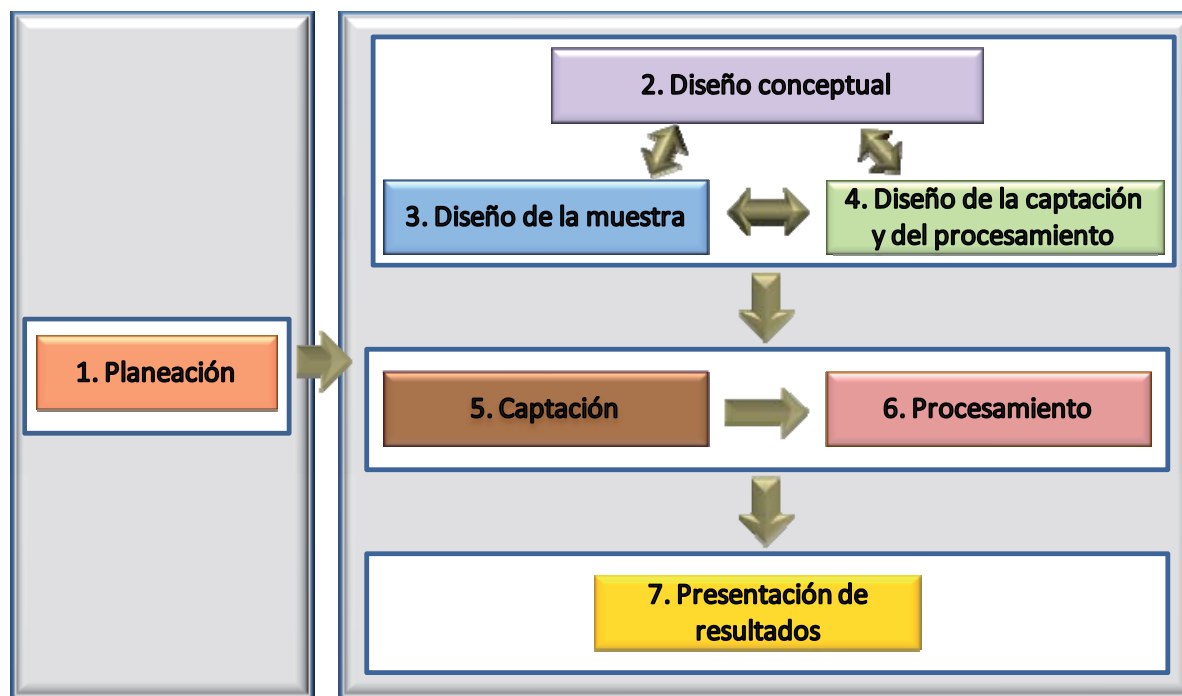
El diseño de la muestra es la fase de un proyecto de generación de estadística básica donde se define el esquema de muestreo a utilizar, se determina el tamaño y procedimiento de selección de la muestra y, en el caso del muestreo probabilístico, se calculan los factores de expansión y los estimadores que se requieren para la generación de resultados.

El diseño de la muestra interactúa con otras fases del proceso de generación, como se visualiza en el siguiente diagrama:

La fase de **diseño conceptual** y el **diseño de la muestra** interactúan en doble sentido, debido a que las coberturas conceptual y geográfica influyen en las decisiones sobre el esquema de muestreo, y éste a su vez puede implicar el ajuste del desglose conceptual y geográfico, dado el límite del presupuesto. Además, es en el diseño conceptual en donde se define, en su caso, el diseño de los estimadores necesarios para inferir los resultados hacia la población de estudio.

La fase de **diseño de la captación y procesamiento** también interactúa en dos sentidos con la fase de diseño de la muestra, dado que la estrategia para el

Proceso estándar para encuestas por muestreo



La interacción se da en lo general de la siguiente forma:

La fase de **planeación** interactúa como condicionante de la fase del **diseño de la muestra**, esto debido a que define el alcance de los objetivos del proyecto en cuanto a cobertura temática y geográfica, así como las restricciones impuestas por el presupuesto disponible.

diseño de captación implica considerar la distribución geográfica de la muestra, en tanto que en el sentido inverso, el diseño de la muestra requiere considerar las opciones idóneas para la captación de los datos, dadas las características del contexto donde se realiza el proyecto.

Finalmente, el diseño de la muestra se relaciona con la fase de procesamiento, mediante la entrega de los

factores de expansión, las estimaciones y precisiones correspondientes.

1.2 Conceptos básicos

Es necesario definir los conceptos más utilizados en el documento para mejor comprensión del tema.

El **esquema de muestreo** es una combinación específica de:

- Tipo de muestreo.
- Modalidad de muestreo.
- Número de etapas de selección por aplicar (según las características de la población objeto de estudio y el tipo de datos a captarse).

El **tipo de muestreo** se refiere a la característica de ser probabilístico o determinístico (llamado también no probabilístico):

- El **muestreo probabilístico** asigna a cada elemento de la población bajo estudio una probabilidad conocida y diferente de cero de ser seleccionado en la muestra.

Para ello es indispensable contar con un marco de muestreo:

- El **marco de muestreo** es el conjunto de materiales (listas, archivos, mapas, etc.) que permiten identificar a todos los elementos de la población, seleccionar una muestra y localizar sus unidades en campo una vez seleccionadas.
- El **factor de expansión** es un concepto relacionado con la probabilidad de selección y se interpreta como la cantidad de unidades en la población que representa una unidad en la muestra, llámese personas, viviendas, áreas económicas o agrícolas, etcétera, dicho factor permite dar conclusiones sobre la población total.
- El **muestreo determinístico** define el tamaño y selección de la muestra bajo aspectos subjetivos (criterio personal del investigador, conveniencia, entre otros); en este tipo de muestreo los resultados no pueden generalizarse a la población objeto de estudio, son válidos exclusivamente para el conjunto de elementos en muestra.

La **modalidad de muestreo** consiste en una subclasificación para el tipo de muestreo.

La **muestra** es un subconjunto de unidades de observación seleccionadas de una población, bajo

| En el caso probabilístico | En el caso no probabilístico o determinístico |
|---------------------------|---|
| Aleatorio simple | Convencional o accidental |
| Sistemático | Por cuotas |
| Estratificado | Bola de nieve (<i>snowball</i>) |
| Por conglomerados | intencional o por juicio |

condiciones preestablecidas que serán objeto de registro y captación de datos.

La **determinación de la muestra** contempla tanto la definición del tamaño como la selección de la muestra.

El **tamaño de muestra** es el número de unidades de observación que deben estar incluidas en la muestra.

La **selección de la muestra** se refiere a los procedimientos empleados para identificar las unidades de observación que integrarán la muestra. La selección puede realizarse con o sin reemplazo; en la primera situación se permite que una observación pueda estar en la muestra más de una vez, mientras que la segunda los elementos ya seleccionados lo hacen en forma única.

El **número de etapas de selección** significa que la determinación de la muestra se realiza en varias fases o etapas; dentro de cada una de ellas se aplica una selección individual; las etapas están vinculadas de tal forma que la muestra de una etapa cualesquiera es seleccionada sólo en aquellas unidades que fueron extraídas en la etapa inmediata anterior.

Se dice que el diseño de la muestra es **complejo** si combina el uso de al menos dos de los tres siguientes elementos:

- Varias etapas de selección.
- Conformación de estratos o estimador de razón.
- Probabilidades de selección desiguales.

1.3 Condicionantes del diseño de la muestra

Debido a que la fase del diseño de la muestra en un proyecto específico, debe responder a decisiones tomadas en fases previas. Es necesario identificar aspectos que actúan como condicionantes y que se describen a continuación:

De la fase de planeación. La población en estudio y la amplitud del área geográfica a cubrir, así como el desglose geográfico para el que se pretende generar estadísticas, influyen en las decisiones sobre el esquema de muestreo.

De la fase de diseño conceptual. La cantidad de variables a estudiar, la frecuencia con la que la característica se presenta en la población y la amplitud de los valores, son factores que inciden en la determinación del tamaño de la muestra y permiten obtener la representatividad deseada en las estimaciones correspondientes.

De la disponibilidad o estructura del marco de muestreo. Disponer o no de un marco muestral, así como de su estructura influye en la determinación del esquema de muestreo, pues de optarse por el muestreo probabilístico es necesario realizar la construcción o actualización del marco; en tanto que de no ser factible disponer de él, o bien, si las unidades muestrales son desiguales, es decir, algunas de ellas contienen la mayor parte de la información a estudiar, entonces se debe utilizar un esquema no probabilístico. Asimismo, determinadas características del marco muestral en un contexto

geográfico amplio, pueden implicar un esquema de selección de la muestra de varias etapas, partiendo de divisiones territoriales grandes hasta pequeñas áreas de selección.

De los recursos disponibles. En toda encuesta por muestreo se enfrenta el reto de obtener estadísticas con un aceptable grado de precisión en combinación con un presupuesto razonable, en la perspectiva del costo-beneficio. Existen aspectos asociados con el presupuesto y duración del evento que influyen en el cálculo del tamaño de la muestra derivado por el esquema de muestreo elegido. Por ejemplo, una muestra determinística puede ser menos costosa, pero no se puede evaluar el error; por otra parte, con una muestra probabilística se requerirá de un presupuesto mayor, pero será posible conocer el grado de precisión de las estimaciones y obtener conclusiones que se generalicen hacia toda la población.

2. Opciones para la determinación el esquema de muestreo

El esquema de muestreo se refiere a una combinación de opciones técnicas en cuanto al tipo y modalidad del muestreo, así como del número de etapas de selección, respecto al cual se toman las decisiones para un proyecto específico de encuesta.

Cabe aclarar que las alternativas metodológicas son diversas y su determinación implica tanto el análisis de los condicionantes como del conocimiento de las ventajas y desventajas de cada alternativa. En el siguiente cuadro se ilustran en general dichas opciones.

Para elegir el tipo de muestreo más adecuado en un proyecto de encuesta, la decisión debe basarse en los objetivos del estudio, el esquema de la investigación y el alcance de sus contribuciones.

medio mecánico (tablas de números aleatorios, números aleatorios generados con una calculadora u ordenador, etc.) se eligen tantos sujetos como sea necesario para completar el tamaño de muestra requerido.

- Este procedimiento, tiene poca o nula utilidad práctica cuando la población que se va a estudiar es muy grande.
- Sistemático. Se numeran todos los elementos de la población, pero en lugar de extraer n números aleatorios sólo se extrae uno. Se parte de ese número aleatorio para elegir, a intervalos constantes, todos los demás hasta completar la muestra.

| Tipo de muestreo/Modalidad de muestreo | Número de etapas de selección | | |
|--|--|--|-------------|
| | Unietápico | Bietápico | Polietápico |
| Probabilístico <ul style="list-style-type: none"> • Aleatorio simple¹ • Sistemático • Estratificado • Por conglomerados | Una sola elección del tipo y modalidad de muestreo | Combinación específica del tipo de muestreo, la modalidad y el número de etapas de selección | |
| No probabilístico o determinístico <ul style="list-style-type: none"> • Convencional o accidental • Por cuotas • Cadena o bola de nieve (<i>snow ball</i>) • Intencional o por juicio | | | |

¹ Existen dos variantes para esta modalidad diferenciadas por la manera de realizar la selección, con o sin reemplazo.

2.1 El muestreo probabilístico

Es útil para realizar estimaciones de variables en la población, donde todos sus elementos tienen una misma probabilidad de ser elegidos; los elementos muestrales tendrán valores muy parecidos a los de la población, de manera que las mediciones del subconjunto darán estimados precisos del conjunto mayor. Una de las principales ventajas de este tipo de muestreo es que puede medirse el tamaño de error de las predicciones y en consecuencia, reducir al mínimo el error estándar (Kish, 1965).

Las modalidades a considerar son las siguientes:

- Aleatorio simple. Se asigna un número a cada individuo de la población y a través de algún

- Estratificado. Se divide la población total en clases homogéneas, llamadas estratos (por grupos de edades, por sexo). Hecho esto la muestra se escoge aleatoriamente en número proporcional al de los componentes de cada clase o estrato. Con ello, asegurar que todos los estratos de interés estén representados adecuadamente en la muestra. Cada estrato funciona independientemente, pudiendo aplicarse dentro de ellos el muestreo aleatorio simple o el estratificado para elegir los elementos concretos que formarán parte de la muestra.

- Por conglomerados. La unidad muestral es un grupo de elementos de la población que forman una unidad, que es llamada conglomerado. Consiste en seleccionar aleatoriamente un cierto número

de conglomerados (el necesario para alcanzar el tamaño muestral establecido) y posteriormente investigar todos los elementos pertenecientes a los conglomerados elegidos. Cuando los conglomerados son áreas geográficas suele hablarse de “muestreo por áreas”.

2.2 El muestreo no probabilístico o determinístico

En el muestreo no probabilístico, también conocido como determinístico, el cálculo del tamaño y selección de la muestra se basan en juicios y criterios subjetivos, por lo tanto se desconoce la probabilidad de selección de las unidades de la población bajo estudio y no es posible establecer la precisión respecto a niveles de confianza predefinidos. No obstante esta restricción, el muestreo determinístico representa una alternativa viable, ya sea cuando la aplicación del muestreo probabilístico resulta demasiado costosa; cuando no es posible disponer de un marco de muestreo o cuando existe seguridad en que la información recabada bajo este tipo de muestreo es suficientemente útil para los fines de la investigación.

En el contexto del muestreo no probabilístico pueden identificarse las siguientes modalidades:

MUESTREO CONVENCIONAL O ACCIDENTAL

Consiste en recopilar datos acerca de los sujetos de estudio que resulten más accesibles. Es un esquema de muestreo rápido y de bajo costo, pero con deficiencias en términos de representatividad. Es útil como parte del proceso de estudios exploratorios con propósitos de orientar la definición de una investigación y no para la caracterización de estructuras o del comportamiento de una población objeto de estudio.

Muestreo por cuota

En este tipo de muestreo se utilizan los datos de subconjuntos o determinados estratos de población, tales como: sexo, edad o religión, entre otros, para seleccionar miembros que se consideren típicos según los propios fines de la investigación. El muestreo por cuotas recibe su nombre en función de la práctica que consiste en asignar ciertas proporciones de la muestra a determinados estratos de la población. En el proceso de selección, el entrevistador decide a quien aplicar el cuestionario, bajo determinados criterios generales establecidos previamente. Éstos últimos resultan insuficientes para evitar en la práctica la intervención de elementos subjetivos, por lo cual no es utilizado para fines de estadísticas que requieren buen nivel de confiabilidad.

MUESTREO EN CADENA O BOLA DE NIEVE

El muestreo en cadena, o bola de nieve, permite ubicar a sujetos o elementos de muy difícil localización o identificación. Se utiliza principalmente en investigaciones cualitativas que requieren la opinión experta, frecuentemente asociada con la toma de decisiones, pues se basa en la identificación de individuos que puedan proporcionar la mejor información sobre determinada experiencia o acontecimiento bajo estudio, para lo cual se aprovecha las relaciones entre expertos, bajo el supuesto de que se conocen mediante redes sociales. Sin embargo, para fines de generación de estadísticas esta modalidad no es apropiada.

MUESTREO INTENCIONAL O DE JUICIO

La característica principal de este tipo de muestreo es que, tanto el tamaño de muestra como la selección de los elementos que la integran están sujetos al juicio del investigador, del cual se requiere suficiente conocimiento y experiencia sobre el tema. La validez de los resultados en este caso depende del nivel de conocimiento sobre el fenómeno en estudio y de evidencias estadísticas que muestren su utilidad para conocer aspectos de comportamiento.

Las modalidades del muestreo no probabilístico son aplicadas principalmente en los medios académicos y en investigaciones privadas, pero son de baja utilidad para las oficinas nacionales de estadística por lo que, en general, se opta por no utilizarlas con la excepción del muestreo intencional o de juicio, que ha resultado útil para conocer con fidelidad, tendencias y comportamientos de determinadas variables en la generación de estadísticas económicas.

Tal es el caso cuando es posible establecer que en determinadas unidades de una población se concentre un alto porcentaje del valor de la variable objeto de estudio y se proceda a incluirlas como parte de la muestra, hasta alcanzar el nivel de cobertura pretendido o previamente acordado para la variable en cuestión.

Dicho nivel depende de los intereses y del conocimiento que se tenga, por parte del área responsable, acerca del sector de actividad económica objeto de investigación. Si bien la definición de la muestra no es representativa de la población total, se tiene la certeza de que su comportamiento en el tiempo caracteriza tanto a la variable de interés como algunas otras altamente relacionadas con ella, mientras que el tamaño de muestra resulta muy inferior al correspondiente a un diseño probabilístico.

2.3 Muestreo multietápico

Varias condiciones pueden propiciar a decidir que la selección de la muestra se realice en más de una etapa: como el no disponer de la localización exacta de cada una de las unidades de observación, o si el tamaño de la población de estudio es demasiado grande; entonces resulta conveniente conformar conglomerados mediante la agrupación de unidades de observación que deberán reunir o compartir algunas características.

Cuando la población en estudio se ha dividido en conglomerados pero hay gran variabilidad entre sus

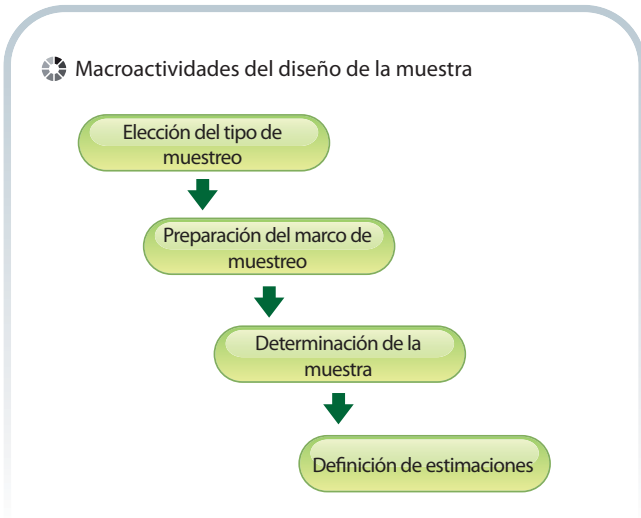
tamaños, se pueden conformar nuevos conglomerados al interior y elevar el número de etapas de selección a dos o más.

Una alternativa para la situación anterior consiste en combinar a los conglomerados con estratos. En este caso los conglomerados se estratifican, seleccionándose cierto número en cada estrato y de los grupos elegidos se toman las unidades de análisis que integrarán la muestra.

El hecho de usar muestreo por etapas conlleva a una reducción en costos pero conduce también a un incremento en las varianzas de las estimaciones.

3. Macroactividades del diseño de la muestra mediante muestreo probabilístico

El diseño de la muestra se desarrolla en una secuencia de actividades agrupadas en macroactividades, en las cuales es necesario optar por alternativas idóneas conforme al análisis de condicionantes correspondientes.



3.1 Elección del tipo de muestreo

Decidir sobre el tipo de muestreo depende, en buena medida, de la existencia o posibilidad de integrar un marco actualizado, del cual se debe seleccionar la muestra, ya que si ello es factible, es preferible aplicar el muestreo probabilístico. La imposibilidad de disponer, actualizar o integrar un marco, puede obligar a un muestreo determinístico, u optar por el muestreo sistemático.

CRITERIOS PARA ELEGIR LA MODALIDAD DE MUESTREO

- Si se dispone de un marco de muestreo actualizado y las unidades del marco cuentan con la misma probabilidad de selección, es factible utilizar un muestreo aleatorio simple sin reemplazo (MAS). Aunque es el más sencillo no suele emplearse directamente para llevar a cabo selecciones en poblaciones grandes. Su mayor relevancia la debe al hecho de ser un procedimiento básico como integrante del muestreo complejo.

- Si adicionalmente el marco de muestreo ya está estratificado o se dispone de información auxiliar para hacer una estratificación, el muestreo estratificado mejora las estimaciones del MAS (estimadores del tipo promedio o total); bajo esta situación, es preferible aplicar un muestreo sistemático con arranque aleatorio de manera independiente en cada estrato, el cual simplifica la extracción de la muestra.
- Si es necesario evitar el problema de dispersión geográfica de las unidades de observación, es útil el muestreo de conglomerados, a partir de la delimitación de áreas de características similares. Los conglomerados facilitan la utilización del muestreo con probabilidad proporcional al tamaño (PPT), ya que el número de elementos en un conglomerado representa una medida natural del tamaño del conglomerado.

3.2 Preparación del marco de muestreo

Esta actividad es esencial y se refiere a la detección, evaluación, organización y elaboración de los listados, directorios o materiales cartográficos, donde se identifican todas las unidades de la población en estudio.

La población definida a través del marco de muestreo debe contener a toda la población objeto de estudio, de lo contrario la encuesta aportará resultados poco confiables. Un marco de muestreo debe, entonces, tener propiedades relacionadas con la calidad, con la eficiencia y el costo, como se observa en la siguiente tabla.

| Lista de propiedades aconsejables de los marcos | |
|---|--|
| Propiedades relacionadas con la calidad | Unidades bien definidas |
| | Identificadores idóneos |
| | Marcos completos |
| | Marcos actualizados |
| | Unidades estables* |
| Propiedades relacionadas con la eficiencia | Inclusión de información complementaria exacta y actualizada |
| | Elección de las unidades de muestreo disponibles* |

| Lista de propiedades aconsejables de los marcos | |
|---|--|
| | Posesión de buenos mapas de las unidades |
| | Facilidad de manipulación/ procesamiento |
| Propiedades relacionadas con el costo | Bajo costo de adquisición/ preparación |
| | Bajo costo de uso |
| | Bajo costo de mantenimiento* |

* Denota las propiedades pertinentes únicas o primordialmente para los marcos de muestreo maestros u otros marcos que se van a utilizar en más de una encuesta o ronda de encuestas.

Fuente: Programa para desarrollar la capacidad nacional de efectuar encuestas de hogares. Versión preliminar, pág. 77.

En la práctica, se dan distintas situaciones en relación con la disponibilidad de un marco de muestreo, cada una con implicaciones en cuanto a recursos a utilizarse. Se destacan los siguientes casos:

- Se dispone de todo el marco al inicio del estudio
- El marco es parcial y debe actualizarse.
- No existe el marco y debe construirse.

La depuración y actualización del marco de muestreo puede realizarse con apoyo en diversos materiales disponibles y en procesos de digitalización de listados y croquis, sistemas para conformar unidades de muestreo y sistemas para la selección. Si con los materiales existentes no se puede integrar el marco, puede ser necesario construir uno específico, sin embargo, los costos de este pueden hacer inviable el proyecto si el presupuesto no alcanza a cubrirlos. En este caso se recomienda formar unidades de muestreo intermedias entre las distintas etapas de selección.

Para mejorar la eficiencia del diseño es necesario que el marco contenga información que permita separar las unidades de muestreo en estratos, formados preferentemente en función de variables correlacionadas con la(s) variable(s) en estudio. La estratificación óptima logra que los elementos que se incluyan dentro de cada estrato sean homogéneos y que haya heterogeneidad entre los estratos, así mismo, deben incluir a toda la población de modo que cada unidad de muestreo pertenezca exactamente a un solo estrato.

Por otra parte, hay que considerar la existencia y posibilidad de uso de marcos de muestreo maestros sobre viviendas y establecimientos, integrados y actualizados por el INEGI.

Este tipo de marco permite:

- La planeación operativa de las encuestas.

- Reducir los costos operativos de las encuestas por concepto de marcos.
- Tener un mejor control de los errores de muestreo y de los ajenos al muestreo.
- Mejorar la congruencia entre las estimaciones de las encuestas. (ver cuadro pág. siguiente).

3.3 Determinación de la muestra

Elegido el esquema de muestreo, ahora es necesario determinar aspectos referentes a la muestra, en particular su tamaño y su procedimiento de selección.

3.3.1 Cálculo del tamaño de la muestra

En este tipo de muestreo, y contrario a la creencia de que el tamaño de muestra debe guardar cierta proporción con el tamaño de la población, los aspectos que se involucran en el cálculo están relacionados con la característica a estudiar, el nivel de precisión y confianza que se desea lograr, el esquema de muestreo que se sigue para obtener la información, así como el dominio de estudio o área a la que se desea brindar la información.

- Cuando una característica se presenta con frecuencia en la población, el tamaño de muestra es menor que el requerido para una característica extraña o poco común; pues en éste último caso, se necesita entrevistar a una gran parte de la población para obtener algunos casos que presenten la característica de interés.
- La variabilidad de la característica a estudiar también se involucra, pues se requiere un tamaño de muestra mayor para indicadores que toman un número infinito de valores como ingreso por trabajo (que va desde 0 por trabajador familiar sin pago, hasta lo que gana el director de alguna compañía) a lo que se requiere para indicadores que toman valores más acotados como estatura al nacer.
- El nivel de precisión está relacionado tanto con el error permitido (la distancia entre la estimación y el valor "real") como con la confianza con que se va a ofrecer este resultado; por ejemplo, si se desea obtener resultados con 95% de confianza y un error máximo de 8%, significa que de 100 muestras sólo 5 tendrían un error mayor al 8 por ciento.
- En el caso de las distintas subdivisiones territoriales (o dominios de estudio) indicadas por el desglose geográfico, se debe hacer un cálculo por separado para cada una de ellas; el tamaño de muestra final será la suma de los tamaños de muestra de cada subdivisión.

Ventajas e inconvenientes de los distintos tipos de muestreo probabilístico

| | Características | Ventajas | Inconvenientes |
|-------------------------|--|--|---|
| Aleatorio simple | Se selecciona una muestra de tamaño n de una población de N unidades, cada elemento tiene una probabilidad de inclusión igual y conocida de n/N | <ul style="list-style-type: none"> • Sencillo y de fácil comprensión • Cálculo rápido de medias y varianzas. • Se basa en la teoría estadística y, por tanto, existen paquetes informáticos para analizar los datos. | Requiere de antemano un listado completo de toda la población. Cuando se trabaja con muestras pequeñas es posible que no represente a la población adecuadamente. |
| Sistemático | <ul style="list-style-type: none"> • Conseguir un listado de los N elementos de la población. • Determinar tamaño muestral n. • Definir un intervalo $K=N/n$. • Elegir un número aleatorio, r, entre 1 y k (r=arranque aleatorio). • Seleccionar los elementos de la lista. | <ul style="list-style-type: none"> • Fácil de aplicar. • No siempre es necesario tener un listado de toda la población. • Cuando la población está ordenada siguiendo una tendencia conocida, asegura una cobertura de unidades de todos los tipos. | Si la constante de muestreo está asociada con el fenómeno de interés, las estimaciones obtenidas a partir de la muestra pueden contener sesgo de selección. |
| Estratificado | En ocasiones será conveniente estratificar la muestra según ciertas variables de interés. Para ello se debe conocer la composición estratificada de la población objetivo al hacer un muestreo. Ya calculado el tamaño muestral apropiado, éste se reparte de manera proporcional entre los distintos estratos definidos en la población usando una simple regla de tres. | <ul style="list-style-type: none"> • Tiende a asegurar que la muestra represente adecuadamente a la población en función de unas variables seleccionadas. • Se obtienen estimaciones más precisas. • Su objetivo es conseguir una muestra lo más semejante posible a la población en lo que a la o las variables estratificadoras se refiere. | Se ha de conocer la distribución en la población de las variables utilizadas para la estratificación. |
| Conglomerados | Se realizan varias fases de muestreo sucesivas (polietápico). La necesidad de listados de las unidades de una etapa se limita a aquellas unidades de muestreo seleccionadas en la etapa anterior. | <ul style="list-style-type: none"> • Es muy eficiente cuando la población es muy grande y dispersa. • No es preciso tener un listado de toda la población, sólo de las unidades primarias de muestreo. | <ul style="list-style-type: none"> • El error estándar es mayor que en el muestreo aleatorio simple o estratificado. • El cálculo del error estándar es complejo. |

Es común que en encuestas de propósitos múltiples se desee brindar información de distintos indicadores, de esta manera es necesario calcular un tamaño de muestra para cada uno y elegir el tamaño de muestra mayor, ya que éste cubrirá las especificaciones de precisión de todos los indicadores.

Sin embargo, este tamaño de muestra puede resultar muy costoso, entonces se debe tomar la decisión de no incluir algunas variables en la encuesta, o bien, admitir para ellas un error de estimación más alto que para el resto.

Considerando todo lo anterior se debe elegir la expresión matemática que permita calcular el tamaño de muestra.

Es necesario destacar que la expresión definida considerando los aspectos anteriores, se ve afectada por la modalidad de muestreo elegida y el número de etapas realizadas para obtener la información, es decir por el efecto del diseño (DEFF), la tasa de no respuesta (TNR), y el coeficiente de variación (CV), por lo cual la expresión matemática final debe incluir cada uno de estos términos.

Por otra parte, el grado de complejidad de la ecuación también depende de los valores particulares que adquieran algunos de los aspectos arriba mencionados, así por ejemplo, se puede tener una ecuación muy simple cuando: el error es absoluto, el tipo de parámetro es una proporción, la distribución de la variable es normal y el esquema por aplicar es MAS.

Adicionalmente, se debe mencionar que en el muestreo estratificado, con frecuencia los resultados se requieren para ciertos estratos de la población y las cotas de error deseados se establecen para cada uno de ellos; en este caso, se debe calcular por separado el tamaño en cada grupo y el tamaño de muestra final será la suma de las establecidas para cada estrato.

Por otra parte, aun cuando los resultados no se requieran a ese nivel en particular, sí se debe tener cuidado de asignar muestra a todos los estratos (L) contemplados en el marco de muestreo, en caso de que el tamaño de muestra (n) no lo permita, se deben hacer los ajustes (uniones de estratos) necesarios para no dejar a ninguno sin representación.

Al respecto, existen varias formas de distribuir (o “afijar”) la muestra dentro de los diferentes estratos (nh); la elección de una de estas opciones está en función del conocimiento y comportamiento de las varianzas (S_h) y de los costos (C_h) por captar un cuestionario de cada uno de los estratos.

DISTRIBUCIÓN DE IGUAL NÚMERO DE CADA ESTRATO

Esta opción es la más sencilla de aplicar y asume que los estratos presentan las varianzas, los costos y sus tamaños iguales (Nh); si esto no es así, redundará en estimaciones más pobres que las tres siguientes alternativas; sin embargo, puede ser útil cuando se pretende obtener resultados con precisiones semejantes en los diferentes estratos. Para obtener la distribución se aplica la expresión:

$$n_h = \frac{n}{L}$$

DISTRIBUCIÓN PROPORCIONAL

Se usa si los estratos presentan varianzas iguales, costos también iguales y sus tamaños son distintos. La distribución se obtiene con:

$$n_h = \frac{N_h}{N} n$$

DISTRIBUCIÓN DE NEYMAN

Se aplica cuando los costos son iguales en todos los estratos, las varianzas son distintas y los tamaños de los estratos también son distintos. En este caso, la expresión a emplear es:

$$n_h = \frac{N_h S_h}{\sum_{h=1}^L N_h S_h} n$$

DISTRIBUCIÓN ÓPTIMA

Se emplea cuando se tienen costos muy diferentes por estrato, las varianzas son distintas y los tamaños de los estratos también son diferentes. La distribución se genera con:

$$n_h = \frac{N_h S_h / \sqrt{C_h}}{\sum_{h=1}^L N_h S_h / \sqrt{C_h}} n$$

3.3.2 Selección de la muestra

En este caso, el algoritmo empleado debe garantizar que la selección sea aleatoria y que la probabilidad de selección de las unidades de muestreo sea la que establece el esquema de muestreo elegido, para lo cual debe tenerse información clara y precisa del número de elementos que integran cada grupo de unidades del marco de muestreo sujeto de selección independiente, o en su defecto, la magnitud de la variable que se utiliza como referente para la selección. Es recomendable que la selección se haga de manera automatizada a fin de evitar al máximo los errores de conteo.

Es prudente mencionar que pueden existir varios procedimientos de selección para un mismo esquema de muestreo; por ejemplo, en el muestreo aleatorio sin reemplazo (MAS) se puede hacer la selección considerando:

- Una tabla de números aleatorios.
- Generar números en el intervalo de cero a uno con distribución uniforme.
- El uso de la probabilidad uniforme.

Cualquiera de estas opciones produce la muestra con la probabilidad de selección que el MAS requiere (n/N)

Si se emplea un muestreo estratificado o por conglomerados, la selección debe realizarse de manera independiente en cada estrato o conglomerado, se debe verificar que el número de unidades de muestreo seleccionadas por grupo coincida con el tamaño de muestra total.

3.4 Definición de estimaciones

Dado el esquema de muestreo y las características de la muestra, se definen las bases técnicas para el cálculo de las estimaciones, lo cual implica considerar los factores de expansión y las medidas de precisión.

3.4.1 Cálculo de factores de expansión

El hecho de que a partir de una muestra de la población de estudio se infieran estimaciones sobre variables específicas para la población total, implica la expansión de la muestra de acuerdo con los llamados factores de expansión, que técnicamente se definen a la inversa de la probabilidad de selección. Esto se aplica en dos actividades básicas: el cálculo de los factores de expansión y el análisis ajuste de los mismos.

Dada su definición, los factores de expansión se calculan sin información de campo, solo se requiere conocer la probabilidad de selección de cada unidad de observación acorde con el esquema de muestreo elegido.

AJUSTE DE LOS FACTORES DE EXPANSIÓN

Esta actividad se realiza hasta que culmina la fase de captación e implica evaluar el nivel de la no respuesta, procediéndose a los ajustes necesarios considerando lo siguiente:

- Que las unidades con respuesta absorban el peso de aquellas sin respuesta, y expandan por sí mismas a la población en estudio (ajuste por no respuesta).
- Asegurar que en cada dominio de estudio se obtenga la misma población total a la determinada por una proyección confiable y referida a la misma fecha del levantamiento de la encuesta (ajuste por proyección).

Posterior a estos posibles ajustes y antes de declarar como liberados los factores de expansión “definitivos”, debe verificarse que en cada dominio para el que se pretende obtener estimaciones, la expansión obtenida para la(s) variable(s) en estudio sea congruente con el total obtenido con el marco.

Los factores de expansión definitivos son aplicados durante la fase del procesamiento para la explotación de resultados.

3.4.2 Cálculo de las estimaciones y precisiones

Liberada la base de datos definitiva y disponiéndose de los factores de expansión, una actividad más que se considera en la generación de resultados, en el **diseño de la muestra** se refiere al cálculo de estimaciones y las precisiones correspondientes.

CÁLCULO DE LAS ESTIMACIONES

Disponiéndose de la información captada, los factores de expansión y la expresión matemática para cada

estimador, el cálculo de las estimaciones se realiza por sustitución. Así, para estimar el total de una variable de estudio, se suma los productos generados a partir de multiplicar el valor de cada unidad de observación por el factor de expansión correspondiente. De manera similar y respetando la expresión del estimador, se construyen las estimaciones para promedios, proporciones, y razones.

El estimador elegido debe satisfacer las siguientes características:

- Ser insesgado.
- Ser consistente.
- Tener varianza mínima.
- Ser fácil de obtener y calcular.

Habiendo definido las fórmulas para obtener la varianza de los estimadores elegidos, el cálculo de la varianza se realiza de igual forma por sustitución numérica.

CÁLCULO DE LAS PRECISIONES

Como parte de la evaluación de la calidad de la información captada por la encuesta, en particular para conocer si se cumplieron las expectativas de confiabilidad de los estimadores, se calculan las precisiones de los mismos, esta actividad consiste en obtener, para todas aquella(s) variable(s) en estudio:

- El error de muestreo, previo cálculo de la estimación de la varianza del esquema de muestreo empleado.
- Los intervalos de confianza de acuerdo al nivel de confianza establecido con el cálculo del tamaño de muestra.
- El efecto de diseño.
- El coeficiente de variación.

En la determinación del intervalo de confianza, en casos de un diseño muestral complejo, debe considerarse el efecto del diseño, multiplicando los límites del mismo por la desviación estándar del DEFF calculado (DEFT); por ejemplo, al aplicar un paquete de cómputo, que supone independencia y distribución idéntica en las unidades de observación, se debe corregir el intervalo reportado por el *software*, multiplicando por el DEFT.

4. Documentación del diseño estadístico de la muestra

Finalmente la actividad de cierre de la fase de diseño estadístico de la muestra, implica la elaboración de una nota metodológica en la cual se documenten todos los aspectos importantes relacionados con el desarrollo de la propia fase.

Entre ellos cabe destacar, en función del caso de que se trate, los siguientes:

La población objeto de estudio.

- El marco de muestreo.
- La unidad de observación.
- La estratificación.
- El esquema de muestreo empleado.
- La determinación de la muestra.
- Cálculo y ajuste de los factores de expansión.
- Cálculo de los estimadores empleados.
- Cálculo de las precisiones.

En el desarrollo de estos puntos, se deben plasmar tanto las características originalmente planteadas para el diseño de la muestra como los ajustes realizados a las mismas durante el proceso de aplicación. Ambas situaciones deben quedar establecidas con la mayor claridad posible.

La documentación de los aspectos mencionados tiene como propósito presentar la metodología utilizada en la encuesta de una forma organizada, clara y lo más sencilla posible para facilitar su comprensión. Se pueden utilizar cuadros, gráficos e ilustraciones a fin de hacerla más atractiva.

La descripción detallada debe presentarse en el orden lógico en que se desarrolló el trabajo, considerando los problemas enfrentados durante el diseño de la muestra y las soluciones correspondientes que se adoptaron.

Para incrementar la utilidad de esta actividad, los expertos que participaron en este trabajo, pueden aportar conclusiones y recomendaciones que se pueden incluir en un capítulo complementario.

Con ello se ofrecen elementos para realizar mejoras en el proyecto actual o para encuestas futuras.

Glosario

Categoría. Conjunto objeto de cuantificación y caracterización. **INEGI.** 2010. *Diseño conceptual para la generación de estadística básica.*

Cobertura geográfica. Territorio o ámbito espacial al que se refiere la captación de datos en un proyecto estadístico. **INEGI.** 2010. *Diseño conceptual para la generación de estadística básica.*

Coefficiente de variación. Es el cociente entre la desviación estándar y la media aritmética.

Conglomerado. Es una subpoblación en la cual los elementos que la componen poseen cierta característica que les hace ser propios de cierta cualidad o atributo.

Por ejemplo, lugar geográfico, grupo étnico, ideología y organización social.

Cruce de variables. Combinación de cada una de las clases de una clasificación con cada una de las clases de otra, respecto a las variables involucradas. **INEGI.** 2010. *Presentación de datos estadísticos en cuadros y gráficas.*

Cuestionario. Tipo de instrumento de captación que presenta, bajo un orden determinado, las preguntas e indicaciones necesarias para el registro de los datos correspondientes a las unidades de observación, en un proyecto de generación de estadística básica. **INEGI.** 2010. *Norma técnica para la generación de estadística básica.*

Desagregación geográfica. Nivel de detalle de una división territorial. **INEGI.** 2010. *Diseño conceptual para la generación de estadística básica.*

Determinación de la muestra. Contempla tanto la definición del tamaño como la selección de la muestra.

Diseño de la muestra. Conjunto de actividades mediante las cuales se determina el método de muestreo por aplicar, el tamaño de la muestra y los procedimientos de selección, así como los elementos técnicos para la determinación de estimadores. **INEGI.** 2010. *Norma técnica para la generación de estadística básica.*

Distribución (o “afijación”) de la muestra. Se refiere a la distribución de la muestra entre los diferentes estratos de diseño y dominios de estudio para los que se requieren realizar estimaciones.

Documentación. Texto descriptivo utilizado para definir o describir un objeto, diseño, especificación, instrucciones o procedimiento.

Dominio de estudio. Véase definición de categoría.

Encuesta por muestreo. Método para generar información estadística mediante la captación de datos para un subconjunto de unidades seleccionadas de la población objeto de estudio. **INEGI.** 2010. *Norma técnica para la generación de estadística básica.*

Encuesta transversal. Es una encuesta donde se obtiene únicamente una medición de los sujetos de estudio en un momento particular dado.

Enfoque asistido por un modelo. Es un caso particular del enfoque basado en el diseño, se introduce un modelo para construir estimadores, pero la inferencia se hace basada en el diseño (estimadores de razón o de regresión).

Entrevista. Procedimiento para obtener información mediante un cuestionario en el que se presenta una serie de preguntas realizadas a un interlocutor o entrevistado. **INEGI.** 2010. *Captación en encuestas por muestreo.*

Esquema de muestreo. Es una combinación específica del tipo de muestreo, la modalidad de muestreo y el número de etapas de selección por aplicar, según las características de la población objeto de estudio y el tipo de datos a captarse.

Error absoluto. Es el valor positivo del error de estimación.

Error de estimación. Es la diferencia entre un estadístico y su parámetro correspondiente.

Error de muestreo. Véase error de estimación.

Error de no respuesta. Error que ocurre cuando la encuesta falla en obtener la respuesta de una, o posiblemente de todas las preguntas.

Error estadístico. La diferencia no conocida entre el valor estimado y el valor verdadero.

Error no muestral. Error en las estimaciones de la muestra que no pueden atribuirse a las fluctuaciones del muestreo.

Error del marco. Error causado por limitaciones inherentes a los insumos utilizados para producir los datos, o por retrasos y errores en la obtención y procesamiento de los datos.

Error estándar. La desviación estándar de la distribución de un valor estadístico.

Error relativo. Es el cociente entre el error absoluto y el parámetro.

Estadística básica. Información generada a partir de un conjunto de datos obtenidos de un proyecto censal, de una encuesta por muestreo o del aprovechamiento de registros administrativos. **INEGI.** 2010. *Norma técnica para la generación de estadística básica.*

Estadístico. Es una medida cuantitativa, derivada de un conjunto de datos de una muestra.

Estimación. Es el valor numérico de un estimador.

Estimador. Es un estadístico usado para estimar un parámetro desconocido de la población.

Estimador consistente. Un estimador se dice ser consistente en el sentido de que al incrementar el tamaño de muestra, la estimación se acerca cada vez más al parámetro poblacional.

Estimador de razón. La estimación se forma por un cociente basado en la relación existente entre dos variables y que se miden en el mismo conjunto de elementos.

Estimador insesgado. Es cuando el valor esperado del estimador es igual al parámetro estimado.

Estrato. Es una subpoblación en la cual los elementos que la componen reúnen características comunes que la hacen ser homogénea.

Factor de corrección por finitud. Se determina restando al tamaño del universo, el tamaño de la muestra

y dividiendo este diferencial nuevamente entre el tamaño del universo; es un ajuste que se hace a la varianza cuando el universo es finito.

- **Factor de expansión.** Es un concepto relacionado con la probabilidad de selección y se interpreta como la cantidad de unidades en la población que representa una unidad en la muestra, llámese personas, viviendas, áreas económicas o agrícolas etcétera, dicho factor permite dar conclusiones sobre la población total.

Fase de captación. Serie de actividades para obtener los datos a nivel de las unidades de observación, conforme a determinado método de generación de estadísticas. **INEGI.** 2010. *Norma técnica para la generación de estadística básica.*

Fase de diseño conceptual. Serie de actividades para identificar las necesidades de información y determinar el marco conceptual, los instrumentos de captación, los criterios de validación y la presentación de resultados. **INEGI.** 2010. *Norma técnica para la generación de estadística básica.*

Fase de diseño de la captación y el procesamiento

- **Diseño de la captación.** Serie de actividades para determinar, desarrollar y probar las estrategias para el levantamiento de los datos, así como los procedimientos y sistemas para su seguimiento y control.
- **Diseño del procesamiento.** Serie de actividades para determinar, desarrollar, y probar estrategias y procedimientos que habrán de aplicarse para la validación de los datos captados y la generación de resultados estadísticos. **INEGI.** 2010. *Norma técnica para la generación de estadística básica.*

Fase del diseño de la muestra. Conjunto de actividades mediante las cuales se determinan el método de muestreo por aplicar, el tamaño de la muestra y los procedimientos de selección, así como los elementos técnicos para la determinación de estimadores. **INEGI.** 2010. *Norma técnica para la generación de estadística básica.*

Fase de planeación en un proyecto estadístico. Proceso para determinar los objetivos y estrategia de un proyecto, así como la secuencia de actividades y su calendarización, los recursos y la organización requeridos para su realización. **INEGI.** 2010. *Norma técnica para la generación de estadística básica.*

Fase de presentación de resultados. Serie de actividades para la elaboración de productos a partir de la información estadística generada en un proyecto determinado. **INEGI.** 2010. *Norma técnica para la generación de estadística básica.*

Fase de procesamiento. Serie de actividades mediante las cuales se ordenan, almacenan y preparan los archivos con la información captada, asegurando su congruencia a fin de proceder a su explotación para la presentación de resultados estadísticos. **INEGI.** 2010. *Norma técnica para la generación de estadística básica.*

Fase del proceso de generación de estadística. Cada una de las series de actividades que se distinguen por su naturaleza técnica específica y los momentos de realización, dado un programa y calendario del proyecto estadístico. **INEGI.** 2010. *Norma técnica para la generación de estadística básica.*

Intervalo de confianza. Rango o recorrido de valores numéricos dentro del cual se espera que esté con cierto grado de confianza el parámetro de estudio (habitualmente se emplea el 95% de confianza).

Marco de muestreo. Listado en el cual se identifica a todos los elementos de una población y que permite seleccionar una muestra de la misma con fines de estimación estadística.

Muestra. Subconjunto de unidades seleccionadas de una población, bajo condiciones preestablecidas que serán objeto de registro y captación de datos.

Muestra no probabilística. Una muestra en la que la selección de las unidades está basada en factores diferentes a la oportunidad aleatoria, por ejemplo conveniencia, experiencia anterior o el juicio del investigador.

Muestra probabilística. Una muestra seleccionada mediante un método basado en la teoría de la probabilidad (proceso aleatorio), esto es, por medio de un método que incluye el conocimiento de las posibilidades de que alguna unidad sea seleccionada.

Muestreo. El proceso de seleccionar un número de casos de todos los casos en un grupo particular o universo.

Muestreo aleatorio simple. Es una modalidad del muestreo probabilístico donde cada elemento de la población tiene la misma probabilidad de ser seleccionado para integrar la muestra.

Muestreo aleatorio simple con reemplazo (irrestringido). Es un caso particular del muestreo aleatorio simple donde la selección se hace con reemplazo. Genera observaciones independientes e idénticamente distribuidas, lo cual es un supuesto básico de la teoría estadística que es necesario para confirmar relaciones entre variables (véase muestreo aleatorio simple).

Muestreo aleatorio simple sin reemplazo (MAS). Es un caso particular del muestreo aleatorio simple donde la selección se hace sin reemplazo. Esta modalidad no produce observaciones independientes ni idénticamente distribuidas, pero su demanda de uso es alta, ya que sus varianzas estimadas son menores que el caso con reemplazo. Además, para solventar el inconveniente de independencia hace uso del factor de corrección por finitud (véase muestreo aleatorio simple).

Muestreo bietápico. La muestra se genera en dos etapas de selección.

Muestreo bola de nieve o snowball. Es una modalidad del muestreo determinístico y consiste en localizar a algunos individuos que integran la población en estudio, los cuales conducen a otros y éstos a su vez a otros, así hasta conseguir una muestra suficiente. Esta modalidad de muestreo se emplea muy frecuentemente cuando se hacen estudios con poblaciones “marginales”, delincuentes, sectas o determinados tipos de enfermedades.

Muestreo casual, fortuito o incidental. Es una modalidad del muestreo determinístico y ocurre cuando el investigador selecciona directa e intencionadamente a los individuos de la población. Por ejemplo, cuando se entrevista a una de cada diez personas que pasan por una calle cualquiera.

Muestreo con probabilidad proporcional al tamaño (PPT). Es una modalidad del muestreo probabilístico y de fuerte aplicación en la actualidad; puede llevarse a cabo cuando el marco de muestreo contiene información sobre variables “auxiliares” que tienen una buena relación proporcional con la variable de estudio. Las unidades de observación se seleccionan con probabilidades desiguales y, a la vez, proporcionales a cierta variable auxiliar, logrando con esto que el cociente formado por la variable de interés y la auxiliar tenga mucho menor variabilidad que la de las lecturas individuales de la variable en estudio; este método maneja al igual que el muestreo aleatorio simple, la opción de seleccionar la muestra con o sin reemplazo, aunque el cálculo de las varianzas estimadas se obtienen suponiendo un muestreo con reemplazo, ya

que el caso sin reemplazo es un problema teórico de difícil solución; es decir, se adopta una posición conservadora.

Muestreo estratificado. Es una modalidad del muestreo probabilístico y se basa en la conformación de subpoblaciones (estratos) mutuamente excluyentes, de tal manera que cada una de éstas, sea lo más homogénea posible en su interior y, a la vez, entre ellas sean lo más diferente posible. Para elaborar los estratos adecuadamente se requiere entre otras cosas, de una selección adecuada de variables “auxiliares”, de definir el número óptimo de grupos a formar y de la existencia y manejo de un sistema de cómputo adecuado; a pesar de lo anterior, en general, esta modalidad redonda en ciertos beneficios versus la opción de no conformar estratos a saber: mejores estimaciones, permite un manejo óptimo de recursos y facilita administrativamente el control en la fase de captación.

Muestreo intencionado, por juicio, opinático o por conveniencia. Es una modalidad del muestreo determinístico y se basa en que los elementos que integran la muestra son resultado de la *expertis* del investigador. Las principales ventajas de esta opción son la facilidad de obtener la muestra y que el costo usualmente es bajo.

Muestreo polietápico. La muestra se genera en más de dos etapas de selección.

Muestreo por conglomerados. Es una modalidad del muestreo probabilístico y consiste en la identificación de subpoblaciones considerando, para su delimitación, rasgos de proximidad territorial (se dice ser la contraparte de formar estratos); enseguida, se seleccionan sólo algunas de ellas y se capta información del total de unidades de las mismas. La gran ventaja de manejar esta alternativa de muestreo es que definiendo conglomerados de tamaño adecuado los entrevistadores no tienen que hacer grandes traslados para lograr más entrevistas durante el levantamiento obteniendo con esto la muestra en un corto periodo y a bajo costo.

Muestreo por cuotas o accidental. Es una modalidad del muestreo determinístico y consiste en facilitar al entrevistador tanto el número como las características de las personas que tiene que seleccionar dejando, a su criterio la elección de las mismas, siempre y cuando cumplan con las características indicadas, cuando se combina con una selección por etapas generalmente se aplica en la última de ellas.

Muestreo sistemático con arranque aleatorio. Es una modalidad particular en el muestreo probabilístico, el cual divide a la población en k subgrupos de n elementos y aleatoriamente toma como muestra a uno de ellos; por su forma de construcción el tamaño de muestra de los subgrupos no es constante (puede ser menor al requerido). Cuando la población está en un orden aleatorio con respecto a la variable que estima a la población en estudio, este método es equivalente al MAS, aunque su uso facilita la extracción de la muestra. Por otro lado, cuando el orden de la población presenta tendencia a cambios paulatinos de dicha variable, esta modalidad produce varianzas estimadas menores que el MAS, esto se debe a que la muestra queda más dispersa sobre la población, es decir, que la muestra es más representativa; aunque en este caso no hay expresiones válidas para obtener los estimadores y las varianzas respectivas, se pueden utilizar las expresiones dadas por el MAS como una aproximación conservadora; en caso contrario, es decir, cuando se tiene un orden que se refleja en cambios periódicos, este método puede producir varianzas estimadas mayores a las generadas mediante el MAS.

Muestreo unietápico. La muestra se obtiene en una sola emisión (no hay etapas de selección).

Nivel de confianza. Es la probabilidad de que el intervalo construido en torno a un estadístico capte el verdadero valor del parámetro.

Número de etapas de selección. Significa que la determinación de la muestra se realiza en varias fases o etapas; dentro de cada una de ellas se aplica una selección individual; las etapas están vinculadas de tal forma que la muestra de una etapa cualesquiera es seleccionada solo en aquellas unidades que fueron extraídas en la etapa inmediata anterior.

Parámetro. Son las medidas o datos que se obtienen sobre la población, tales como la media, la varianza, la proporción.

Población en estudio. Es el conjunto de unidades de observación, seres humanos, establecimientos, u otro tipo de elementos que se van a estudiar, en cierta ubicación de tiempo y espacio. En un esquema de tipo probabilístico y para fines de calcular el tamaño de la muestra, se requiere conocer de ese conjunto una estimación de su varianza, el coeficiente de variación, el efecto del diseño (DEFF), obtenidos de encuestas similares previas o de un estudio piloto.

Precisión. Se refiere a la correspondencia de los resultados de las mediciones obtenidas de la muestra (estimador) con respecto a los resultados que se medirían en toda la población (parámetro).

Probabilidad de selección. Oportunidad que tiene cada elemento de la población o universo de ser incluido en una muestra.

Proceso para la generación de estadística básica. Conjunto de procedimientos y actividades para producir estadísticas, a partir de la aplicación de un instrumento de captación a unidades de la población objeto de estudio. **INEGI.** 2010. *Norma técnica para la generación de estadística básica.*

Representatividad de la muestra. Una muestra es representativa si los rasgos de los elementos que la integran son similares a los de toda la población, es decir, si la muestra es capaz de reproducir las características de la población.

Selección de la muestra. Se refiere a los procedimientos empleados para identificar las unidades de observación que integrarán la muestra. La selección puede realizarse con o sin reemplazo; en la primera situación se permite que una observación pueda estar en la muestra más de una vez, mientras que la segunda los elementos ya seleccionados lo hacen en forma única.

Sistema nacional de información estadística y geográfica o sistema. Conjunto de unidades organizadas

a través de los subsistemas, coordinadas por el instituto y articuladas mediante la red nacional de información, con el propósito de producir y difundir la información de interés nacional. **INEGI.** 2008. *Ley del sistema nacional de información estadística y geográfica.*

Tamaño de muestra. Es el número de unidades de observación que deben estar incluidas en la muestra.

Tema. Enunciado genérico referente a un campo de conocimiento. **INEGI.** 2010. *Diseño conceptual para la generación de estadística básica.*

Unidad de observación. Elemento unitario del cual se obtienen datos con propósitos estadísticos sobre el conjunto al que pertenece. **INEGI.** 2010. *Norma técnica para la generación de estadística básica.*

Universo (o población) de estudio. Ver categoría.

Valor. Cada uno de los números o modalidades nominales que puede admitir una variable.

Variable. Concepto que admite distintos valores para la caracterización o clasificación de un elemento o un conjunto. **INEGI.** 2010. *Diseño conceptual para la generación de estadística básica.*

Varianza. Es un estimador de la dispersión de una variable aleatoria respecto de su media.

Bibliografía

- DANE.** *La Calidad Estadística a Través de las Normas ISO.* Colombia.
http://http://www.dane.gov.co/publicaciones/Estadisticas_al_dia_No4.pdf
- DANE.** *Metodología para Formulación de Planes Estadísticos.* Colombia.
http://www.dane.gov.co/sniet/archivos_zip/metod_planes.zip
- Elmar Wein.** *The Planning of Data Editing.* Federal Statistical Office, Germany
<http://www.unece.org/stats/documents/2000/10/sde/3.e.pdf>
- EUROSTAT-OECD.** *Glossary Of Statistical Terms (OECD)*
<http://cs3-hq.oecd.org/scripts/stats/glossary/index.htm>
- EUROSTAT.** *Harald Sonnberger and Nick Maine. Editing and Imputation in Eurostat*
<http://www.unece.org/stats/documents/2000/10/sde/21.e.pdf>
- EUROSTAT-OECD.** *Procedures and Checklists for Oecd Statistical Activities*
<http://www.oecd.org/dataoecd/26/40/21687687.pdf>
- Felligi, I.P. Holt D.** (1976). *A systematic approach to automatic edit imputation, journal of the American statistical association.* March 1976.
- Granquist, Leopold,** 1997. *The new view on editing.*
- Grossh Margaret E., Muñoz Juan.** (1998). *Manual de diseño sobre condiciones de vida (LSMS)* documento de trabajo no. 1265.
- INE.** *Nuevas Tecnologías para Difundir Datos Estadísticos. España.*
http://www.fuentesestadisticas.com/numero2/paginas/feu_nuevas.html
- INEGI.** (2006). *Desarrollo y documentación de software, Guía.*
- INEGI.** (1999). *Estrategias generales del XII Censo General de Población y Vivienda 2000.*
- INEGI.** (2005). *Sistema de codificación automático y manual de la Encuesta Nacional de Ocupación y Empleo (ENOE).*
- INEGI.** (2002). *Memoria XII Censo General de Población y Vivienda 2000.*
- Leiv Solheim,** *How to Measure the Effect of Data Editing.* Statistics Norway.
<http://www.unece.org/stats/documents/2000/10/sde/2.e.pdf>
- ONU.** *Diccionario de datos de definiciones oficiales.*
- ONU.** *Principios y recomendaciones para los censos de población y habitación.* Serie M, No. 67/rev.1.
- ONU.** *Recomendaciones internacionales para las estadísticas industriales.* Serie M No. 48.
- STATCAN.** *Characteristics of an Effective Statistical System* (Ivan, P. Fellegi)
<http://unstats.un.org/unsd/goodprac/bpform.asp?DocId=190&KeyId=25>
- STATCAN.** (2003) *Gestión de la Calidad de los Datos en un Organismo Estadístico* (Gordon Brackstone).
www.eclac.cl/deype/ceacepal/documentos/lcl1891e.pdf
- Villan, Idelfonso, Bravo, Maria Soledad.** *Procedimiento de depuración de datos estadísticos.* Eustat, 1990. Seminario internacional de estadística en EUS-KADI.