**Guía de Diseño de la Muestra para Encuestas.**

**Introducción:**

El diseño de Muestra es un subproceso de la Fase de diseño donde se define un esquema de muestreo a usar, determinando;

* Tamaño de la muestra.
* Procedimiento de selección de unidades.
* Distribución de la muestra.
* Determinación de ponderadores y estimadores que se requieren para la generación de resultados en un Programa de Información basado en encuestas. (solo en el muestreo probabilístico).

En el diseño de muestra se interactúa con todos los subprocesos de la fase de diseño**, para estimar el tamaño de la muestra** que se empleara en el programa de información **se tienen en cuenta** los siguientes factores:

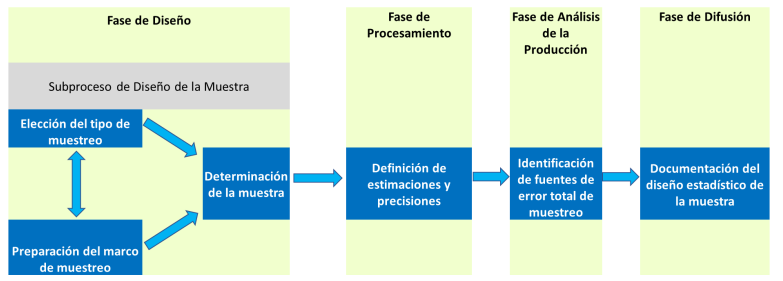
**Diseño Conceptual donde se determinan los indicadores objetivo que se requieren representar o medir.**

**Determinación de la población objeto de estudio**

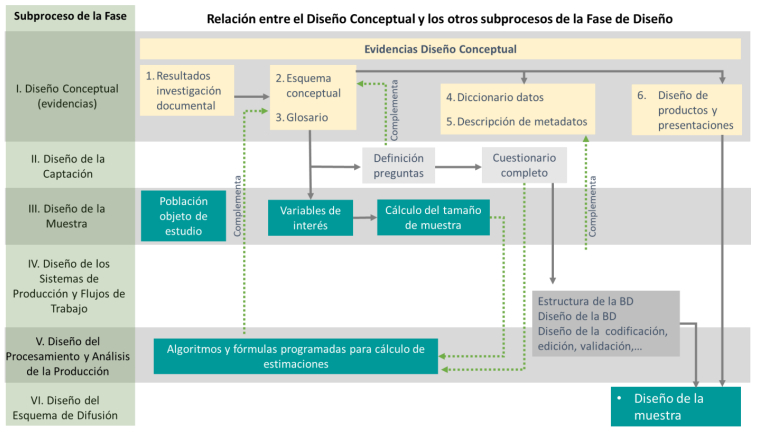
**Marco muestral.**

**Tipo de muestreo.**

**Fases involucradas en el Diseño de la Muestra.**



**Subprocesos que interactúan entre sí:**



**Diseño Conceptual y Diseño de muestra interactúan en doble sentido:** por las coberturas; conceptual y geográfica influyen en las decisiones sobre el esquema de muestreo.

La **Fase de documentación interactúa como condicionante del Diseño de la Muestra**: define el alcance de los objetivos del Programa conforme a la cobertura temática y geográfica, presupuesto.

**Fase de Captación interactúa con en doble sentido con el Diseño de Muestra**: Es la estrategia de Diseño de Captación, considera la distribución y dispersión geográfica de la muestra.

**Diseño de Muestra se relaciona con las Fases de Procesamiento y Análisis de la Producción**: Conforme a la entrega de ponderadores, estimaciones e indicadores de precisión estadística corresponden a los resultados obtenidos de la Fase de Captación.

**Un subproceso de Diseño de la Muestra de la Fase de Diseño en un Programa específico debe responder a decisiones tomadas en las otras fases**.

**Aspectos condicionantes que influyen en otras decisiones sobre la definición del esquema de muestreo:**

**Diseño de muestra**

* Unidad de muestreo

**Fase de Documentación de las Necesidades:**

* Temática de la información a generar
* Población objeto de estudio
* Dominios de estudio
* Cobertura y desglose geográfico
* Referencia temporal
* Periodicidad de captación de datos.

**Disponibilidad o estructura de marco de muestreo:**

* Disponer o no de un marco muestral y su estructura (*muestreo probabilístico* requiere de este marco)
* Si no se tiene un marco de muestreo o dispone de él, las unidades muestrales son desiguales algunas de ellas tienen la mayor información a estudiar (se usa el esquema *no probabilístico*).

**Disponibilidad de los recursos:**

* En las encuestas se enfrenta al reto de obtener estadísticas con un aceptable grado de precisión, combinándolo con un presupuesto razonable y la duración del evento que influyen en el calculo del tamaño de la muestra derivado del esquema de muestreo elegido.
* Muestra no probabilística menos costosa pero no se puede evaluar el error de muestreo ni se puede generalizar toda la población objeto de estudio
* Muestra probabilística se requiere mayor presupuesto, pero es posible conocer el grado de precisión de las estimaciones y de obtener conclusiones que se generalicen hacia toda la población objeto de estudio.

**Diseño Conceptual (Fase de Diseño):**

* La cantidad de variables a estudiar, la frecuencia con la que la característica o fenómeno se representa en la población objeto de estudio y la cobertura de los valores, determinan el tamaño de la muestra que permiten obtener representatividad deseada sobre las estimaciones correspondientes.

**Preparación del Marco de Muestreo para Muestreo Probabilístico:**

Es importante para la detección, evaluación, organización y elaboración de los listados, directorios o materiales cartográficos para identificar todas las unidades de la población objeto de estudio.

La población definida debe contener a toda la población objeto de estudio, si no los resultados de la encuesta no son confiables, el marco de muestreo debe tener estar relacionado en calidad, eficiencia y costo.

Dependiendo de la disponibilidad del marco de muestreo, se tiene diferentes implicaciones de recursos;

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

**Lo ideal sería que el marco muestral estuviera actualizado y las imperfecciones eliminadas o reducidas.**

**Componentes involucrados en el marco de muestreo y las imperfecciones que puede haber:**

Marco de muestreo: Se detona por M siendo el listado de todos los elementos de una población de estudio, permite seleccionar una muestra de esta, para hacer una estimación estadística.

Se deben identificar los elementos que se involucran en el marco de muestreo etiquetado como UM. La unión de estas dos poblaciones se define como el conjunto completo de elementos C, el cual este compuesto por tres partes:

1. Elementos de la población objeto de estudio que están listados en el marco de muestreo:



1. Los elementos de la población objeto de estudio que no están listados en el marco de muestreo:



1. Los elementos que no pertenecen a la población objeto de estudio, pero están listados en el marco de muestreo:

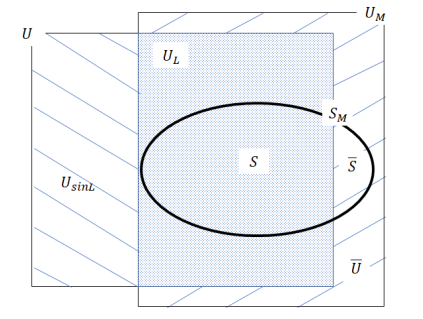


La muestra se selecciona en el marco de muestreo M con un diseño muestral específico. Las unidades de la muestra están listadas en SM (Es un subconjunto de UM). Algunos elementos en SM, son parte de la población objeto de estudio, es decir, el conjunto:



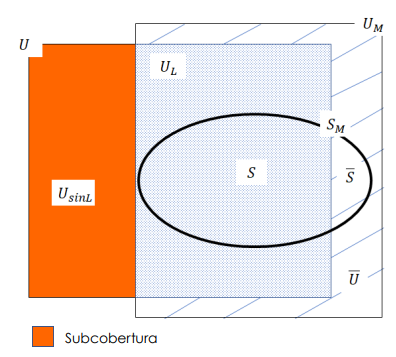
Este contiene los elementos que no pertenecen a la población objeto de estudio, donde los elementos que están dentro del resultado de este conjunto, pueden ya no existir o en caso de existir no son relevantes para la encuesta.



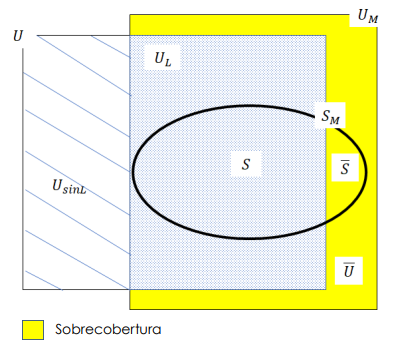


**Tipos de imperfecciones en el marco de muestreo:**

Algunos elementos de la población objeto de estudio no están listados y por lo tanto no pueden ser seleccionados en la muestra s, esto es; UsinL no está vacío.



El conjunto  no está vacío, algunos elementos en el marco de muestreo no pertenecen a la población objeto de estudio.



En el marco de muestreo hay elementos que están listados en más de una ocasión (registros duplicados).

**Indicador de Tasa de Sobre cobertura a nivel unidad de observación**, es el ***porcentaje de unidades observadas contenidas en el marco de referencia, no correspondiente a unidades de la población de estudio,*** se calcula asumiendo que se puede distinguir en el marco de referencia, si hay población que no pertenece al estudio, definiéndolas sobre las unidades del marco observadas.*(en la Fase de Captación y de la Fase de Procesamiento se debe calcular la versión ponderadora)*

**Construcción del marco de muestreo y mantenimiento:**

**Etapas para la construcción y mantenimiento de un marco muestral:**

**Etapa 1 Se identifican los elementos a tomarse en cuenta:**

* Costo del establecimiento, mantenimiento y la información auxiliar.
* Disponibilidad de la información requerida para cada elemento del marco.
* Estabilidad de los elementos del marco en periodo
* Tiempo necesario para construir el marco.

**Etapa 2 Desarrollo:**

* Es la construcción de una base de datos después del proceso de investigación, recolección, estandarización y organización de la información requerida para los elementos del marco.

**Etapa 3 Validación:**

* Es la valuación de la calidad y la cobertura alcanzada en el marco obtenido después de las Etapas 1 y 2.

**Etapa 4 Administración:**

* Define los procedimientos para preservar la calidad del marco de muestreo, si será usado en futuras encuestas, debe considerarse las necesidades de otros usuarios del marco.

**Etapa 5 Mantenimiento:**

* Son las modificaciones y actualizaciones requeridas, Duplicados y “muertes” deberán removerse, los “nacimientos” incorporados y la información auxiliar actualizada. Tener acuatizado el marco de muestre es costoso.

**Información que debe estar contenida en un marco de muestreo, con los datos relacionados:**

1. **Identificación:** Se emplea para identificar de manera única a cada elemento en el marco.
2. **Contacto:** Sirve para localizar a los elementos seleccionados en la muestra durante el proceso de captación de datos.
3. **Clasificación:** esta información es útil para la inclusión en la muestra y posiblemente para hacer estimaciones para la muestra, por ejemplo, una medida de tamaño (número de empleados o cantidad de ingresos totales de una empresa).
4. **Mantenimiento:** Esta información se necesita en caso de que la encuesta se repita en el futuro, como: fechas de agregación o cambios en el marco.

**Marcos muestrales múltiples:**

Encuestas de marcos múltiples **reducen el sesgo de cobertura insuficiente**.

Los Marcos múltiples **deben adoptar un diseño de encuesta muestral combinado con el marco de área** y el marco lista como **apoyo para evitar la inestabilidad de las estimaciones**.

**Resuelven problemas como falta de completez del marco**.

**Se debe tener clasificación de las unidades muestrales en todos los marcos e identificar cuales se localizan en la intersección de ellos**, son sensibles a una mala clasificación.

**Se deben cuidar las muestras en diferentes marcos, para tener un mismo cuestionario y una misma administración**.

**Se asume que existen dos o más marcos de muestreo, cada uno tiene cierto nivel de subcobertura**, pero **su unión puede dar una cobertura más completa de la población** de estudio, asumiendo los siguiente:

* Cada elemento K en la población objeto de estudio U está en al menos uno de los dos marcos UA y UB.
* Una muestra probabilística se toma de cada marco SA de UA y SB de UB.
* La pertenencia a cada marco puede definirse para cada elemento muestreado.

Se distinguen tres dominios no traslapados de la población objeto de estudio U:

* Dominio A, lo componen los elementos solamente en UA, es decir A = U – UB
* Dominio B, lo componen los elementos solamente en UB, es decir B = U - UA
* Dominio AB, lo componen los elementos tanto en UA, como en UB, es decir AB = 𝑈𝐴 ∩ UB

Bajo estos supuestos se tiene que U = A ∪ 𝐵 ∪ 𝐴B, y el estimador de, por ejemplo, un total t se obtiene como:



Estos son los estimadores del total para la población objeto de estudio, dominio A, dominio B y dominio AB, respectivamente. Cuando los marcos muestrales son ajenos, el Dominio AB es vacío y por tanto total

**Marcos de áreas:** **Se basa en unidades de áreas geográficas**, son **usados cuando una encuesta es de naturaleza geográfica o cuando no se dispone de un marco muestral adecuado**.

Deben cubrir toda la población objeto de estudio y dividirlo en unidades geográficas mutuamente excluyentes, **se usa en encuestas que hacen estimaciones basadas en delimitaciones políticas, como municipios, estados o límites geográficos invisibles**.

**Tipos de marcos de áreas:**

* **Cuadricula:** tienen unidades terrestres que serán divididos y muestreados en más etapas, se usa en conglomerados geográficos de unidades muestrales.
* **Terreno**: tienen unidades de muestreo finales que se observan en su totalidad, se usan en encuestas agropecuarias y de medio ambiente.

**Marcos de muestreo maestros (MMM):** **Lista organizada en forma de base de datos**, **contiene a las unidades de observación registradas en el censo** que participan en cada una de las fases de diseño, distribución y selección de la muestra de una encuesta.

**Sirve para identificar y seleccionar** **unidades de muestreo y como base para realizar estimaciones basadas en los datos de la muestra**, **contiene información de la división territorial política y administrativa, del país, además de volumen de viviendas y de la población de estudio**, se usan variables significativas para clasificar los hogares, según el objetivo de la encuesta, **este método mejora con el uso de estratos de variables relacionadas al objetivo** (permite homogeneidad en las variables de cada estrato y heterogeneidad entre estratos).

Estos marcos permiten:

* 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.

**Marcos “ab initio”:** **Al construir un marco de esta forma se debe describir el método de enmienda o construcción**, **una vez definido el universo se recaba información, lo más exacta posible, de sus dimensiones, distribución espacial y temporal** para poder construir un marco muestral.

Es un **muestreo probabilístico**, donde puede reflejarse en encuestas como la de Viajeros Internacionales, donde se marcan los puntos y viajeros fronterizos, **donde su tamaño de la muestra varia por mes o temporada de levantamiento, está basado en el conteo intensivo de viajeros que cruzan de un punto a otro**.

**Muestras maestras**: **Busca hacer selecciones múltiples de unidades que permiten atender las demandas de información de encuestas**, sus ventajas son:

* Se dispone de una muestra maestra seleccionada a partir de criterios uniformes.
* Se utiliza el mismo número de etapas.
* Se utilizan las mismas unidades del marco.
* Se aplica el mismo procedimiento de cálculo para determinar las probabilidades de selección.

***El marco de muestreo debe estar actualizado***, tiene ***costo alto operativo***, ***asume que el crecimiento de las viviendas dentro de la muestra maestra será el mismo que las viviendas no incluidas en esta,*** esprobabilístico, estratificado, unietapico y por conglomerados, permite seleccionar submuestras para todas las encuestas en viviendas de INEGI.

**Desventajas:** **agotamiento de las unidades primarias de muestreo, (quedarse sin viviendas si la muestra maestra se utiliza en varias ocasiones en un corto periodo).**  Ayuda prevenir con la planeación de submuestras que podrían usarse para cada encuesta, pero depende de que el total de viviendas de cada unidad primaria de muestreo, **otra desventaja es la acumulación de sesgo cuando la muestra maestra no se lleva a cabo.**

**Entidades propietarias y usuarias distintas de un marco muestral.**

***La coordinación de ambos, ayuda a disminuir el error y mejorar las necesidades requeridas.***

**Propietario:** *organiza y coordina las actividades relacionadas con el marco, además de crear mantener y actualizar el marco.*

**Usuario:** *envía solicitudes y sugerencias, de las nuevas necesidades de información, hace los requerimientos de la información contenida en el marco muestra*

**Consideraciones adicionales en el uso de marcos muestrales.**

**El esquema de muestreo se refiere a una combinación de opciones técnicas en cuanto al tipo y modalidad del muestreo**, **además del número de etapas de selección**, con el que serán tomadas decisiones para un Programa de información especifico.

**Insumos de la determinación de la población objeto de estudio, marco muestral y tipo de muestreo:**

* Esquema conceptual incluye asociación y jerarquización de temas, universo de análisis variables y clasificaciones.
* Documentación de marco muestral.
* Metodologías anteriores si están disponibles.

**Evidencias de la determinación de la población objeto de estudio, marco muestral y tipo de muestreo (Documentación del diseño de la muestra).**

* Determinación del marco muestral y tipo de muestreo.
* Justificación del marco muestral.
* Diseño de muestreo cuando corresponda.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Tipos de muestreo y etapas de selección.** | | | | |
| **Tipo de muestreo** | **Modalidad de muestreo** | **Numero de etapas de selección** | | |
| **Unietapico** | | **Bietápico o mas etapas** |
| **Probabilístico** | * **Aleatorio Simple (con remplazo y sin remplazo).** * **Sistemático.** * **Estratificado.** * **Por conglomerados.** * **Mediciones repetidas (longitudinal).** | **Una sola elección del tipo y modalidad de muestreo** | **Combinación específica del tipo de muestreo, la modalidad y el número de etapas de selección.** | |
| **Determinístico** | * **Convencional o accidental.** * **Por cuotas.** * **Cadena o bola de nieve.** * **Intencional o por juicio.** |
| **Mixto** | * **Muestreo cutoff** |

**Técnicas de muestreo Probabilístico.**

**Aleatorio simple (MAS):** Se enumeran a todos los individuos de la población de estudio de tamaño N, eligiéndose sujetos para completar el tamaño de la muestra n, n tiene que ser menor igual a N, por medio de la selección aleatoria, donde se extra el conjunto n, además de tener dos variantes.

MAS (con remplazo): Los elementos seleccionados vuelven a la población y pueden ser elegidos de nuevo, los eventos independientes.

MAS (sin remplazo): Los elementos seleccionados no vuelven a la población y no pueden ser elegidos de nuevo, cada miembro de la población solo puede seleccionarse una vez.

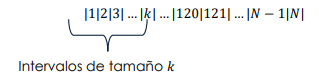
**Ventajas: fácil y sencillo de comprender, calculo rápido de medias y varianzas, se basa en teoría estadística paqueterías informáticas para analizar datos.**

**Desventajas: Requiere de antemano un listado completo de toda la población objeto de estudio, no es conveniente en pequeñas muestras no representa la población adecuadamente, tiene costo alto por lo que es poco práctico, es poco útil si la población es muy grande, tiene alta entropía.**

**Sistemático:** Se numeran los elementos de la población de estudio de forma ascendente, pero en lugar de extraer n números aleatorios solo se extrae uno, entre 1 y k (k = N/n), se parte de ese número aleatorio para elegir a intervalos constantes, todos los demás hasta completar la muestra.

***Cuando la población objeto de estudio tiene orden aleatorio con respecto al indicador Objetivo que estima a la población de estudio, este método es equivalente al muestreo aleatorio simple.***

Si el orden de la población presenta cambios paulatinos de dicho indicador objetivo, esta modalidad produce varianzas estimadas menores que el muestreo aleatorio simple, la muestra queda más dispersa sobre la población.



**Ventajas: Fácil de aplicar, No es necesario tener listada toda la población de estudio, asegura una cobertura de unidades de todos los tipos, cuando la población esta ordenada siguiendo una tendencia conocida, es el método con menor entropía.**

**Desventajas: Si la constante de muestreo está asociada a la variable de interés, las estimaciones obtenidas pueden tener sesgos, produce varianzas mayores a las estimadas por muestreo aleatorio simple, si se tiene un orden reflejado en cambios periódicos.**

**Estratificado: Se divide la población en clases homogéneas y mutualmente excluyentes, llamadas estratos (por un rasgo distintivo), la muestra se distribuye de acuerdo a los métodos de afijación o distribución:**

* Puede ser igual en todos los estratos.
* Puede ser proporcional al tamaño de cada estrato.
* Puede tener distribución de Neyman.
* Distribución optima.

Con ello se asegura que todos los estratos de interés estén representados adecuadamente en la muestra, cada estrato funciona independiente, y se selecciona una muestra probabilística, en este método con frecuencia los resultados se requieren para ciertos estratos de la población y los límites de error se establecen para cada uno de ellos, se debe calcular por separado el tamaño de cada grupo y el tamaño de muestra final será la suma de las establecidas para cada estrato.

**Ventajas: Tiende a asegurar que la muestra represente adecuadamente a la población objeto de estudio en función de las variables seleccionadas, se obtienen estimaciones mas precisas, su objetivo es conseguir una muestra lo más semejante.**

**Desventajas: Se debe conocer la distribución de la población de estudio de las variables utilizadas para la estratificación.**

**Por conglomerados:** **La población de estudio está dividida en subpoblaciones llamadas conglomerados**, en los cuales **los elementos que los componen poseen cierta característica que les hace ser propios de cierta cualidad o atributo**, conveniente si no hay marco de muestreo completo de la población.

El muestreo por conglomerados de una etapa *consiste en seleccionar aleatoriamente un cierto número de conglomerados (el necesario para alcanzar el tamaño muestral establecido) y posteriormente todos los elementos de cada conglomerado elegido formaran parte de la muestra*, es **menos preciso que el muestreo aleatorio simple o estratificado,** por la homogeneidad de las unidades de muestreo dentro de los conglomerados seleccionados, **puede generar información redundante que incrementa los errores muestrales** de los indicadores objetivo.

**Ventajas: Es muy eficiente cuando la población es muy grande y dispersa, No es preciso tener listado a toda la población de estudio, solo las unidades primarias de muestreo.**

**Desventajas: El error estándar es mayor que el muestreo aleatorio simple o estratificado, el cálculo del error estándar es complejo, objeto de estudio es caro de producirlo y cuando se requiere reducir el costo de captación de datos.**

**Muestreo multietapico (muestreo por etapas):** **Es cuando el muestreo por conglomerados se prosigue en el análisis y dentro de cada conglomerado se vuelven a seleccionar, también de toma forma aleatoria, nuevos subconglomerados, y así sucesivamente hasta seleccionar unidades ultimas al muestreo, las numero etapas dependen del tamaño de la población**.

En el muestreo Bietápico (siendo este el más frecuente del multietapico).

***En la primera etapa*:** se seleccionan el primer término y de forma aleatoria, los conglomerados o áreas.

***En la segunda etapa***: se seleccionan las unidades ultimas o más elementales del conjunto poblacional, sin necesidad de seleccionar ningún otro tipo de unidad intermedia.

**Al dividirse en conglomerados, pero hay variabilidad entre sus tamaños, se puede conformar nuevos conglomerados al interior y elevar el número de etapas, a dos o más**.

**Ventajas: Es útil si no se dispone de localización exacta de cada una de las unidades observadas o si el tamaño de la población objeto es demasiado grande (en este caso es conveniente conformar conglomerados, con agrupaciones de unidades de observación con características en común), se puede combinar conglomerados con estratos en este método (los conglomerados se estratifican).**

**Desventajas: En la segunda etapa el cálculo del estimador insesgado de la varianza puede complicarse y puede incrementar las varianzas de las estimaciones.**

**Muestreo multifase:** **Se toman dos o más muestras, estas se extraen del mismo marco y las unidades seleccionadas tienen la misma estructura en cada fase**, en esta se **recopila información de una muestra grande de unidades** y **luego para una submuestra, recopila información más detallada.**

Muestreo doble, es el más común, aunque puede tener más fases, el diseño y las estimaciones de la muestra se vuelve más complicado entre más fases tenga.

Muestreo doble fases:

1. ***Primera fase*:** Se selecciona una muestra Sa grande de unidades usando el diseño muestral simple (para estas unidades se recolecta información auxiliar cuya obtención es sencilla).
2. ***Segunda fase:*** Se selecciona una muestra S proveniente de Sa con un diseño muestral condicional a la primera fase. Los indicadores objetivo se registran para las unidades de la segunda muestra.

**Ventajas: Puede mejorar la precisión de las estimaciones, se puede utilizar para obtener la información, se puede utilizar para obtener información auxiliar que no se encuentra en el marco muestral, se pueden usar cuando el costo de recopilación de algunas variables es costoso.**

**Desventaja: Es más lenta la obtención de resultados que una encuesta de una fase, puede ser mas costoso que una encuesta de una fase, requiere revisar mas de una unidad muestreada, si la población objeto es móvil o si las características de interés cambian con frecuencia, las demoras entre fases pueden causar pueden causar problemas, si la población objeto de estudio es móvil o si las características de interés cambian con frecuencia las demoras entre fases pueden causar problemas, sus formulas para el calculo de estimaciones y varianza muestral es bastante complejo.**

***NOTA: Se debe crear un marco de información útil y confiable para la muestra Sa para que tenga éxito de un muestreo doble, el muestreo doble se ha aplicado para la estimación en encuestas con presencia de no respuesta.***

**Muestreo con Mediciones Repetidas.**

**Muestreo longitudinal:** **Mide el comportamiento de una muestra probabilística a través de periodos**, se **busca analizar datos del crecimiento o tendencia a lo largo del tiempo de los objetivos (mide con precisión los cambios de una característica de la población).**

**Se selecciona la muestra probabilística y las mismas unidades muestrales se incluyen en varios periodos**, cuando se usa un **panel rotante, una proporción de unidades se excluyen de la muestra y se remplazan por otras unidades**.

**Ventajas: Reduce la varianza muestral de las estimaciones muestrales, se puede usar para obtener información del comportamiento de las unidades de observación a través del tiempo, los costos se pueden reducir.**

**Desventajas: Las estimaciones y el tratamiento de la no respuesta son complejos, es más complicado mantener representativa la muestra de la población porque los cambios ocurren a través del tiempo, se requiere que el presupuesto para el mantenimiento de la muestra para que dure toda la encuesta.**

***El muestreo con panel rotante es la mezcla de muestras independientes seleccionadas sucesivamente y una muestra longitudinal,*** donde ***una fracción de muestra se remplaza en cada ocasión que se lleva a cabo la muestra, se busca tener estimaciones a niveles agregados***, es una muestra controlada en un tiempo para garantizar estimaciones transversales insesgadas.

**Muestreo con probabilidades proporcionales al tamaño:**  Es un método que usa información auxiliar y genera probabilidades de selección distintas, es parte de los diseños de probabilidades de selección desiguales, las varianzas de las estimaciones pueden reducirse, tiene más precisión en las estimaciones si las medidas se correlacionan al indicador objetivo.

**Con Variación de tamaños y se conocen en la población de estudio,** **si se usa en la selección de la mejorar la eficiencia estadística**, tiene sesgo de selección, tiene variantes como; con remplazo, sin remplazo, de Poisson y sistemático.

**Ventajas: Puede mejorar la eficiencia estadística del diseño muestral al usar información auxiliar, en comparación con el muestreo aleatorio simple y el muestreo estratificado.**

**Desventajas: Se requiere que el marco muestral se mantenga actualizado para todas las unidades para que pueda usar la medida de tamaño, la estimación de la varianza muestral es más complicada, si la medida de tamaño no tiene correlación con los indicadores de interés, el diseño muestral puede ser menos estadísticamente eficiente.**

**Muestreo con Marcos de flujo de entradas y salidas.**

**Muestreo con marcos “ab intio”:** **Útil para medir flujos de entrada y salida, esta encuesta sirve para las mercancías que cruzan un lado a otro o poblaciones que cruzan de un lado a otro, se usan en áreas de entrada y salida**.

**Se usan “ab initio” y un muestreo sistemático o uno de conglomerados, en varias etapas de muestreo sistemático dentro de los conglomerados muestreados**, el **marco muestral puede ser una lista de las unidades de la población enumeradas dentro de un intervalo de tiempo** en ubicaciones particulares, para que tenga cobertura completa debe tener las ubicaciones de las áreas donde se concentra la población objetivo.

**Ventajas: El marco para la etapa final se puede crear mientras está en el campo, útil si no se tiene marco de referencia completo de la población de estudio o no hay posición fija, se usa para encuestas de flujos de entrada y salida.**

Desventajas: Puede resultar difícil relacionar la población objeto de estudio con una población conocida, esto se debe a que las encuestas de flujo miden a los visitantes en lugar de las personas, es difícil administrar las operaciones de campo debido a los flujos variables en la población, se recomienda que las entrevistas sean breves, se producen tasas de respuestas bajas.

**Diseños muestrales complejos:**

* Se usa este en vez del muestreo aleatorios simple al ser más fácil y barato.
* Usan diseños complejos, que usan estratificación, conglomeración, varias etapas o fases de muestreo de probabilidades de selección distintas y uso de varios marcos de muestreo.
* El objetivo es la minimización de la varianza y disminuir la dificultad de calcular estimaciones muestrales.
* Se busca obtener estimaciones puntuales y de varianza para los parámetros poblaciones de interés.

**Técnicas de muestreo Determinístico (no probabilístico).**

**Muestreo convencional o accidental:** Se recopilan datos acerca de los sujetos de estudio que resulten más accesibles, en el esquema de muestreo rápido de bajo costo, tiene deficiencias en representatividad, sirve para fines exploratorios o para orientar la investigación.

**Muestreo por cuotas:** Se usan datos de subconjuntos o determinados estratos de población objeto de estudio (sexo, edad, religión, etc.), para seleccionar miembros considerados típicos, según la investigación.

Se asignan proporciones de la muestra en determinados estratos de la población, donde en el proceso de selección, el entrevistador decide quien aplicara el cuestionario, bajo determinados criterios generales establecidos previamente, se puede combinar con el muestreo multietapico para tener buenos resultados, mejorando la representatividad de los grupos.

**Muestreo en cadena o bola de nieve:** Se basa en un supuesto donde los miembros de una población rara se conocen, se identifican unas cuantas personas con la característica rara, se pide a cada una que identifique personas con la misma característica rara para la muestra, donde estas nuevas personas de la muestra también identifican a más personas con la característica rara y así sucesivamente, hasta obtener el tamaño deseado de la muestra.

Se puede crear una muestra bastante grande de una población rara, además de que las personas que participan deben tener muchos conocidos, se requieren suposiciones fuertes para generalizar los resultados de una muestra de bola de nieve en la población objeto de estudio.

**Muestreo intencional o de juicio:** La característica de este es que la muestra y la selección de los elementos que la integran son sujetos a juicio del investigador, este requiere mucha experiencia y conocimiento del tema, la eficiencia y validez del resultado depende del nivel de conocimiento del investigador sobre el estudio además de las evidencias estadísticas que muestren su utilidad para conocer su comportamiento.

Es el único método útil de los determinísticos para las oficinas nacionales de estadística, relacionado a temas de tendencias, comportamientos de determinados indicadores para generar estadísticas económicas.

**Técnicas de muestreo Mixtas (combina probabilística con determinísticas).**

**Muestreo cutoff:** En este se combina la selección probabilística y determinística de elementos de la muestra, una parte de la población se incluye de forma determinística. Produce estimaciones sesgadas (se utiliza como error de medición, el error cuadrático medio, sumando la varianza y el sesgo al cuadrado), se pueden justificar con algún argumento de los siguientes:

Es muy cara la construcción y mantenimiento de un marco de muestreo confiable y la ganancia de precisión no es tan significativa.

El sesgo causado por la exclusión de algunas unidades es insignificante, se usa si la distribución de los valores es muy asimétrica y no tiene marco de muestreo para unidades pequeñas, puede haber dos casos:

* Una parte de la población tiene unidades económicas de pequeño tamaño que contribuyen muy poco al total del indicador Objetivo.
* Otro caso es cuando las poblaciones tienen alguna variable económica gigante cuya inclusión en la muestra es obligatoria para no tener un error grande de estimación total.

**Ventajas: Pueden usarse cuando el indicador objetivo tiene una distribución muy asimétrica, no es necesario el mantenimiento del marco de muestreo para las unidades con medidas de tamaño pequeñas.**

**Desventajas: No produce estimadores insesgados, la varianza muestral es cero por definición en el estrato take all, se tiene que estimar la varianza y el sesgo como medidas del error muestral.**

**Versiones de Cutoff de mayor uso.**

*Cutoff 1:* Cada unidad económica con medida de tamaño superior al límite definido tiene probabilidad de inclusión igual a 1 y las unidades debajo del límite tienen probabilidad igual a cero, las unidades no incluidas no se estiman, funciona cuando el indicador objetivo es muy asimétrico, no tiene sesgo significativo en los estimadores en considerar los que tienen probabilidad cero.

*Cutoff 2:* Se hace lo mismo que en Cutoff 1, pero la estimación se hace para las unidades no incluidas en la muestra, se requiere información auxiliar de todas las unidades.

*Cutoff 3:* Se establece un límite para incluir a las unidades en la muestra, también se incluyen algunas unidades debajo del límite, esta estratificación se conoce como toma todo (take all) y toma algunos (take some).

*Cutoff 4:* Ahora dos limites definen 3 estratos de la muestra “toma todo”, “toma algunos” y “toma ninguna”.

*Cutoff 5:* Las unidades se ordenan de forma descendente de acuerdo con una medida de tamaño (ingresos, número de empleados, entre otros). Los se captan a partir de la unidad económica de mayor tamaño hasta completarlos, este método puede ser arbitrario, ante esto se recomienda usar algoritmos existentes para determinar los mencionados puntos de corte y asignación de tamaños de muestra en cada estrato (se debe usar información auxiliar con relación alta con el indicador objetivo), además se parece al Cutoff 2.

**Criterios para elegir la modalidad de muestreo.**

**Muestreo probabilístico:** Si se tiene existencia o posibilidad de integrar un marco de muestreo actualizado, donde se debe seleccionar la muestra, si es factible se puede aplicar esta técnica.

**Muestreo Determinístico:** Si no se dispone, actualiza o integra un marco, se puede optar por método, u optar por el muestreo sistemático.

**Criterios a tener en cuenta para elegir una modalidad de muestreo:**

-Se puede usar el muestreo aleatorio simple si se dispone de un muestreo actualizado y las unidades del marco cuentan con la misma probabilidad de selección, no se usa directamente en poblaciones grandes (el muestreo estratificado mejora las estimaciones de este método).

-Si se cuenta un marco de muestreo, estratificado o de información auxiliar para hacer una estratificación, mejor utilizar el muestreo sistemático con arranque aleatorio de manera independiente en cada estrato, donde simplifica la extracción de la muestra.

|  |  |  |
| --- | --- | --- |
| **Tipos de Muestreo.** | | |
| **Probabilístico** | **Determinístico** | **Mixto** |
| * Misma posibilidad por parte de todos los individuos de ser seleccionados. * Cada unidad es seleccionada Aleatoriamente. * No hay sesgo de selección. * Probabilidades de inclusión, se pueden calcular los estimadores confiables y estimar el error estándar. * El muestreo es objetivo, los resultados pueden generalizarse a la población total, puede ser representativa de la población. * Si se dispone de un marco muestral y su estructura se usa este método. * Puede que se requiera mayor presupuesto, pero es posible conocer el grado de precisión de las estimaciones y de obtener conclusiones que se generalicen hacia toda la población objeto de estudio. | * No todos tienen la misma posibilidad de ser seleccionados. * Tamaño y selección de muestra es a juicio, conforme a la experiencia, conocimiento y criterio del que desarrolla el estudio. * No es posible tener la precisión con niveles de confianza definidos. * Se usa cuando no se dispone de un marco de muestreo. * El muestreo es más subjetivo y no es una muestra representativa de la población. * Puede ser menos costosa pero no se puede evaluar el error de muestreo ni se puede generalizar toda la población objeto de estudio | * Sus probabilidades de inclusión deben ser mayores a cero para todas las unidades de la población. * Hay métodos donde una parte de la población sus probabilidades de selección son mayores a cero mientras que para la otra son iguales a cero. * Son una versión intermedia del muestreo probabilístico y el determinístico, con probabilidades de inclusión desconocidas. |

**Determinación de la muestra para muestreo probabilístico.**

1. Principio de aleatorización: consiste en seleccionar un diseño de muestreo de entropía máxima o cercana a esta.
2. Principio de sobrerrepresentación: es sobremuestrear las unidades con mayor incertidumbre debido a que la muestra debe recolectar tanta información de la población objeto de estudio como sea posible, lo que implica diseños muestrales con probabilidades de selección distintas.
3. Principio de restricción: es seleccionar solo muestras con un conjunto de características deseables, es decir, utilizar un proceso de verificación (muestreo balanceado) para evitar muestras con dominios, estratos o categorías vacías.

***Ya elegido el esquema de muestreo, se determinan aspectos de la muestra en cuanto a tamaño y procedimiento de selección.***

**Cálculo del tamaño de la muestra.**

***El tamaño de la muestra es el numero de unidades de observación que van a ser incluidas en la muestra.***

En el probabilístico, los aspectos que se involucran en el cálculo están relacionados con la característica o fenómeno a estudiar, 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 busca brindar información.

* Cuando una característica se presenta con frecuencia en la población objeto de estudio, el tamaño de muestra es menor que el requerido para una característica extraña o poco común, porque se necesita entrevistar a gran parte de la población, para tener casos que tengan un rasgo 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 numero finito de valores, diferente a lo que se requiere para indicadores que toman valores más acotados.
* El nivel de precisión esta relacionado tanto con el error permitido, como con la confianza con la que se ofrece el resultado (intervalo de confianza y error).
* En el caso de distintas subdivisiones territoriales indicadas por el desglose geográfico, se debe hacer un calculo por separado para cada una de ellas, el tamaño de muestra final será la suma de tamaños de muestra de cada subdivisión.

**Ejemplos de cálculos de los tamaños de las muestras.**

Se debe elegir la expresión matemática que permita calcular el tamaño de la muestra, esta se ve afectada por el muestreo elegido, numero de etapas, la tasa de no respuesta, y el coeficiente de variación, por lo que deberá incluir algunos de estos términos. A su vez el grado de complejidad de la ecuación depende de los valores particulares adquiridos de estos efectos, ejemplo:



n = muestra / z = valor del cuantil (normal dist.) / p = estimación proporción interés muestra piloto

r = error relativo fijo / q = 1 – p / tnr = tasa de no respuesta / DEFF = efecto diseño.

Se tiene un intervalo de confianza de 90%, precisión de 15%, tasa de no respuesta de 15%, valores de proporción al interés p varían de 0.01 a 1, se consideran 3 valores DEFF (1.00, 2.16, 3.65)

**Distribución de la muestra en los estratos.**

Se debe tener cuidado de asignar muestra a todos los estratos, contemplados en el marco de muestreo, donde el tamaño de muestra (n), no lo permita, se debe hacer los ajustes (uniones de los estratos), necesarios para no dejar ninguno sin representación.

Hay varias formas de distribuir o afijar la muestra dentro de los diferentes estratos (nh), la elección de alguna de estas opciones está en función del conocimiento y comportamiento de las varianzas (Sh) y de los costos (Ch) de captar un cuestionario en cada uno de los estratos.

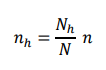
***Distribución de la muestra en los estratos:***

Esta opción es la más sencilla de aplicar y asume que los estratos presentan varianzas, costos y tamaños iguales (Nh). Si esto no es así, redunda en estimaciones mas pobres que a las tres 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 siguiente formula:



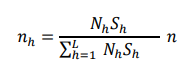
***Distribución proporcional:***

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

******

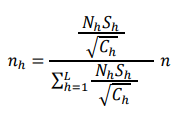
***Distribución de Neyman:***

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

******

***Distribución optima:***

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:



En esta distribución también, se puede minimizar el costo total de la muestra sujeto a una varianza fija o un coeficiente de variación fijo.

*Valores definidos de las anteriores ecuaciones:*

N = tamaño de la población objeto de estudio / L = número de estratos / n = tamaño de muestra

Nh = tamaño de la población objeto de estudio en el estrato / nh = tamaño de muestra de estrato

Sh = varianza del estrato / Ch = costo en el estrato.

***Selección de la muestra:***

La selección de la muestra se refiere a los procedimientos empleados para identificar las unidades de observación que integran la muestra.

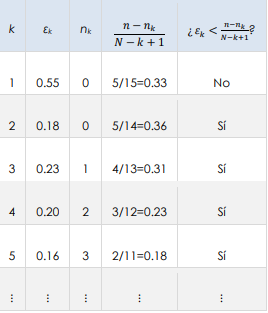
Selección de muestra con remplazo: Se permite una observación que pueda estar en la muestra mas de una vez, mientras la segunda los elementos ya seleccionados lo hacen de forma única, el algoritmo empleado debe garantizar selección aleatoria, la probabilidad de elección debe ser la que establece el esquema elegido (debe tener información clara, precisa de los elementos de cada grupo, sujeto a selección independiente).

Hay muchos procedimientos de selección en un mismo esquema de muestreo: ejemplos de selección de muestreo aleatorio, teniendo disponibilidad de un listado de los elementos de la muestra.

**Esquema A:** Se generan ε1, ε2, …, números pseudo aleatorios independientes de una distribución uniforme en el intervalo (0,1). Solamente ocurre que ε >n/N, entonces el elemento k = 1 se selecciona, para los subsecuentes números k =2, 3..., nkes el número de elementos seleccionados de los primeros k -1 elementos en la lista. Solamente si ocurre que:



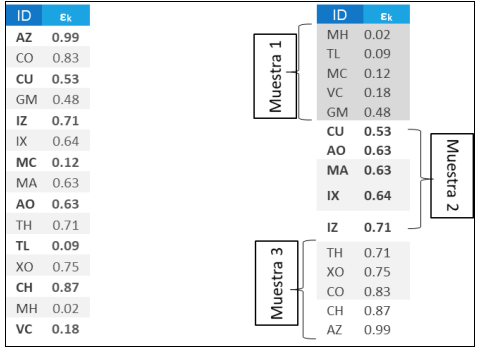
Entonces el elemento k es seleccionado. El procedimiento se detiene en nk = n. (N =15 / n = 5)



**Esquema B:** Este procedimiento es recomendable para poblaciones de viviendas o unidades económicas que se requiere que no sean seleccionadas frecuentemente, este tipo de muestras sin traslape se les denomina negativamente coordinadas; las muestras con traslape máximo se conocen como positivamente coordinadas. Se generan en N números pesudo aleatorios con distribución uniforme en el intervalo (0,1) ε1, …,εk,…,εN, donde εK esta asignado al elemento k. Estos números se ordenan de manera ascendente.



Esta indica el 𝑖 −ésimo valor más pequeño de los N valores ε ligado al elemento 𝑘𝑖 ; 𝑖 = 1, … , N. La primera muestra esta conformada por los primeros n valores ε correspondientes a {𝑘1, …, 𝑘𝑛} la segunda muestra corresponde a los elementos ligados con los siguientes valores ε asociados con el conjunto {𝑘𝑛+1, … , 𝑘2𝑛} que no se traslapa con la primera muestra.



**Esquema de muestreo proporcional al tamaño:** Muestra probabilidades proporcionales al tamaño, siendo un método de selección de muestra cuando la probabilidad de selección es proporcional a la medida de tamaño positiva, que es disponible para todas las unidades en el marco muestral y además aproximadamente proporcional al indicador objetivo.

Las probabilidades de inclusión son: π𝑘 = 𝑛 × 𝑥𝑘/𝑡x.

Si se usa tx como el total de la población medida de tamaño x, se generan ε1, ε2, . . ., εN números pseudo aleatorios independientes de la distribución uniforme del intervalo (0,1). Solamente si ocurre que εk < πk, entonces el elemento k se selecciona. Este procedimiento se repite para todos los elementos en la población objeto de estudio, tamaño de muestra no es fijo.

Si se usa el esquema estratificado o por conglomerados, la selección debe hacerse de forma independiente en cada estrato o conglomerado y se debe verificar que el numero de unidades de muestreo de una distribución ~ N (0,1).

***Cálculo de ponderadores:***

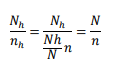
Es el valor asociado a la probabilidad de selección e interpreta la cantidad de unidades de la población, que representa a una unidad de la muestra, este permite dar conclusiones, son el inverso de la probabilidad de selección, requiere de dos tareas:

Calculo de ponderadores y el análisis de los ajustes que se deban considerar en estos, ajuste por no respuesta, por estimación, o tamaño de la población, etc.

El diseño muestral sistemático y el de muestreo aleatorio sin remplazo producen muestras auto ponderadas.



Sabiendo que la distribución es proporcional se verifica que el ponderador es igual para cada elemento en la población objeto de estudio.



Si el diseño de muestra no usa distribución proporcional de la muestra, por ajustes de no respuesta o calibraciones diferentes por cada estrato a los ponderadores, el diseño no es auto ponderado.

***Ajustes de ponderadores:***

Esta actividad se realiza hasta que culmina la Fase de Captación, en el Diseño de muestra se definen los ajustes a los ponderadores, como evaluar el nivel de no respuesta:

* Una es que las unidades de respuestas tendrán pesos mayores a los planeados en a la estimación.
* Que en cada dominio de estudio se obtenga la misma población de determinada estimación siendo confiable y referida a la fecha de levantamiento de la encuesta.

Ajuste por no respuesta y estimación, pueden incluirse en el marco teórico más general, suponiendo el diseño muestral aleatorio simple con ponderadores de diseño dk = n /N, con información auxiliar x (relacionada a la población de estudio).



Los ponderadores calibrados remplazan el estimador por:



El ponderador calibrado Wk se define con la ecuación:



gk = peso de corrección / se deben cumplir dos condiciones para calibrar el ponderador:

* Los ponderadores calibrados gk tienen que ser cercanos a 1 (garantiza que los estimadores obtenidos serán insesgados o aproximadamente insesgados).
* La distribución muestral calibrada auxiliar debe coincidir con la distribución de la población (esta garantiza que la muestra ponderada por los ponderadores calibrados Wk , es representativa de la información auxiliar usada).



La función de la distancia entre gk y 1 denota D (gk, 1) se utiliza para minimizar.



Sujeta la condición de calibración 2, este se puede resolver utilizando el método de multiplicadores de Lagrange, la función de distancia no es única:

* 𝐷 (𝑔𝑘, 1) = (𝑔𝑘 − 1) 2 define una calibración lineal (regresión lineal)
* 𝐷 (𝑔𝑘, 1) = 𝑔𝑘 𝑙𝑜𝑔 (𝑔𝑘) − 𝑔𝑘 + 1 define una calibración multiplicativa (raking)

Los ponderadores calibrados Wk obtenidos después de aplicar cualquier método de calibración pueden tener las siguientes desventajas:

* Ponderadores extremadamente grandes pueden generar estimaciones altamente inestables.
* La calibración lineal puede producir ponderadores negativos.

***La no respuesta en encuestas probabilísticas:***

La no respuesta ocurre cuando un informante no proporciona la información solicitada, genera sesgo por no respuesta, se ajusta con ponderadores muestrales y métodos de imputación, hay dos tipos de no respuesta:

* Por unidad o total: un elemento seleccionado no proporciona ninguna información.
* Por variable: el informante seleccionado responde algunas preguntas, pero no todas.

La tasa de no respuesta mide la calidad de la información, se calcula a nivel muestral conforme a los códigos de no respuesta obtenidos, de acuerdo al inverso de la probabilidad de selección de las unidades de muestreo.

*La tasa de respuesta sin ponderar*, usa conteos simples de la muestra, da la descripción útil del texto operativo de la encuesta, se puede monitorear el éxito para obtener respuestas en la encuesta.

La tasa de no respuesta ponderada se calcula como el cociente de la suma de los pesos de las unidades que responden y la suma de los pesos de las unidades de la muestra, estas tienen una mejor descripción del éxito de la encuesta conforme a la población, son útiles en encuestas con probabilidades distintas de inclusión.

***Cálculo de estimaciones:***

Disponiendo de la información captada, los ponderadores, la expresión matemática de cada estimador y las estimaciones se hacen por sustitución, estimando el total del indicador objetivo, sumando productos generados a partir del multiplicador del valor obtenido de la variable.

**El estimador debe cumplir las condiciones.**

• Ser insesgado; en promedio las estimaciones deben ser iguales al valor poblacional.

• Consistente; cuando se incrementa el tamaño de la muestra la estimación se acerca al valor poblacional.

• Eficiente; las variaciones de los resultados de las posibles muestras deben ser pequeñas (precisión).

• Ser fácil de obtener y calcular (una combinación lineal de valores observados).

El sesgo de un estimador y el valor verdadero de un estadístico poblacional, si θ es el estadístico poblacional y θ̂ es el estimador de dicho estadístico, el sesgo se define como:



Si el 𝑆𝑒𝑠𝑔𝑜(θ̂) = 0 entonces θ̂ es un estimador insesgado del valor poblacional del estadístico θ. Si el estimador es asintóticamente insesgado, se puede considerar cuando n es suficiente grande, es consistente y el estimador se cumple y la distribución del estimador se considera que este alrededor del valor poblacional.

El calculo de la varianza de los estimadores requiere estimar otras cantidades que no son funciones totales, con base a datos de una encuesta, para los cuales no se tienen formulas exactas de la varianza.

𝑅̂ = θ̂𝑦 θ̂𝑥 la varianza del estimador 𝑅̂ no es igual al cociente de las varianzas de los estimadores θ̂𝑦 y θ̂𝑥 .

***Cálculo de las precisiones estadísticas:***

Para tener evaluación de la calidad de la información captada por la encuesta, saber si se cumplen las expectativas de la confianza de los estimadores, se calcula las precisiones.

Error estándar: medida de dispersión esperada de las estimaciones alrededor del parámetro poblacional, se obtiene del calculo de estimación de la varianza Var(θ̂) para el estimador θ, un error estándar grande sugiere menos confianza del estimador muestral, se define como:



El factor de corrección por población finita (fpc) se usa para ajustar la varianza estimada del estimador θ̂, debido a que se obtiene con datos de la muestra, no es completamente observada la población finita, la variabilidad es por datos que no están en la muestra, si se ignora el fpc, se sobreestima el error estándar de θ.

La varianza estimada debe ajustarse en la muestra, son conocidos y la distribución de la varianza θ, proviene de los casos N-n que no se incluyen en la muestra.

Para el muestreo aleatorio sin reemplazo fpc= 𝑁−𝑛 𝑁 = 1 − 𝑓 donde f = n/N.

Se denomina la fracción de muestreo, la cantidad 1 − 𝑓 se aplica a la varianza estimada de θ̂.

La varianza se reduce a cero cuando n = N se acerca al valor completo cuando el tamaño de muestra n se reduce:



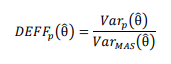
Una muestra con estratificación y se aplican los fpc en cada estrato, uno o más estratos tendrán fracciones de muestreo f altas, reduciendo los errores estándar, en muestreos multiétapicos habrá fracciones de muestreo f distintas en cada etapa siendo involucradas en la estimación de la varianza de θ̂. El efecto de los mismos en la estimación de la varianza depende de la variación de los indicadores objetivos en las unidades primarias de muestreo.

Si la fracción de muestreo de la primera etapa es pequeña, la contribución a la varianza θ del submuestreo en las siguientes etapas no será considerable, los intervalos de confianza se determinan según el nivel de confianza para la muestra y se emplean en la varianza del estimador:

Intervalo de confianza: e 100(1 − α) % para θ (asumiendo la normal del estimador).



El efecto del 𝐷𝐸𝐹𝐹𝑝(θ̂) para el estimador θ bajo el muestral p con muestra n.



Conceptos:

𝑉𝑎𝑟𝑝(θ̂) Varianza del estimador bajo el diseño muestral p.

𝑉𝑎𝑟𝑀𝐴𝑆(θ̂) Varianza para el estimador con el diseño aleatorio simple sin remplazo.

𝐷𝐸𝐹𝐹𝑝(θ̂) > 1; la varianza del estimador θ es mayor usando el diseño p.

El intervalo de confianza debe considerarse el efecto del diseño, multiplicando los limites por la raíz cuadrada del DEFF.



***DEFT es el factor de ajuste de error estándar, en un diseño muestral complejo, se interpreta:***

* DEFT=1: no tiene efecto en las estimaciones de los errores estándares del diseño muestral complejo.
* DEFT>1: el diseño muestral complejo incrementa los errores estándares de los estimadores.
* DEFT<1: incrementa la eficiencia usando el diseño muestral complejo, reduce el error estándar

Coeficiente de variación 𝐶𝑉(θ̂), es una medida relativa de precisión, si es más próximo a cero la estimación es más precisa, no tiene medidas de unidad, la precisión de dos o mas indicadores puede compararse, el estimador θ debe ser positivo.

***Indicadores de calidad de precisión para indicadores objetivo:***

Se adoptaron 4 indicadores para medir la precisión de las encuestas, se calculan para las encuestas probabilísticas, siendo estos; coeficiente de variación, el error estándar y el intervalo de confianza, la cobertura de la variable de diseño, mide la calidad de las encuestas de muestreo no probabilístico, representa el porcentaje alcanzado de la variable de diseño, pudiendo ser la variable que se basa la selección no probabilística de la muestra u otra considerada el diseño conforme a la muestra.



𝑇̃ 𝐷, 𝑚= total de la variable de diseño en la muestra para el dominio D.

𝑇𝐷, 𝑀= total de la variable de diseño en el marco muestral para el dominio D

***Identificadores de las fuentes de error de muestreo:***

El error total de muestreo se refiere a las fuentes de sesgo y varianza de la estimación que pueden afectar la precisión de datos muestrales y surgen del diseño de la muestra, captación, procesamiento, análisis de los datos de la encuesta, estas se agrupan en dos clases:

*Representación:* Relacionado con la población objeto de estudio de la encuesta, variables contempladas en la muestra.

*Medición:* Datos que se obtienen por unidad de observación, como la temática de la encuesta.

**La representatividad de una muestra está afectada por las siguientes fuentes de error:**

*Error de cobertura*. Deficiencias de un marco de muestreo, no representa completamente la muestra de la población objeto de estudio, genera poca precisión de estimaciones muestrales y sesgo de cobertura.

***Definición de población objetivo, procedimientos para seleccionar al informante, disponibilidad y calidad del marco de muestreo, poblaciones difíciles de alcanzar.***

*Error de muestreo*. Error ocurrido cuando una muestra es medida en lugar de la población objeto de estudio completa, el sesgo muestral es la falla sistemática en observar los elementos de la muestra con valores distintos del indicador objetivo, es la variabilidad de las estimaciones, afectado por el tamaño de muestra, además de la conglomeración, estratificación y ponderación, se compone por error fijo (sesgo) y un error variable (varianza).

***Tamaño de muestra, método de captación, conglomeración, estratificación, ponderación en el diseño muestral, procedimientos de campo, definiciones (hogar, residente, habitual).***

*Error por no respuesta*. La no respuesta por unidad o total, pasa cuando un informante seleccionado no se encuentra, no participa en la entrevista, es cuando la no respuesta por unidad está correlacionada con una o más indicadores objetivo.

***Condiciones para la entrevista (acceso a viviendas, seguridad), reglas respecto al numero de contactos, tratamiento de rechazos, definiciones, poblaciones difíciles de alcanzar.***

*Error por ajustes posteriores*. Al concluir la captación de datos de la encuesta probabilística, se hacen ajustes para sesgo de selección, errores de cobertura y errores por no respuesta.

***Disponibilidad de datos, errores de edición, reglas y prácticas para divulgación de la información.***

**Componentes del error de medición:**

*Validez.* Hay una medición correcta del constructo planeado por le temática de la encuesta.

***Existe el concepto, adaptación o operacionalización del constructo.***

*Error de medición y en el proceso de respuesta.* Aparecen cuando los valores verdaderos son diferentes a los del valor medido en la unidad muestreada.

***Proceso de respuesta, comprensión, adaptación, contexto (orden de las preguntas), determinantes sociales, juicio, tendencia a estimar, respuestas socialmente aceptables, desgaste.***

*Error de procesamiento*. Cuando se procesa la información de la encuesta incluye la captura, codificación y edición. El personal del procesamiento de datos puede tener errores fijos y aleatorios.

***Capacidades distintas, edición de respuestas.***

**Diseño muestral de la de la Encuesta de Comercio Internacional de Servicios 2023.**

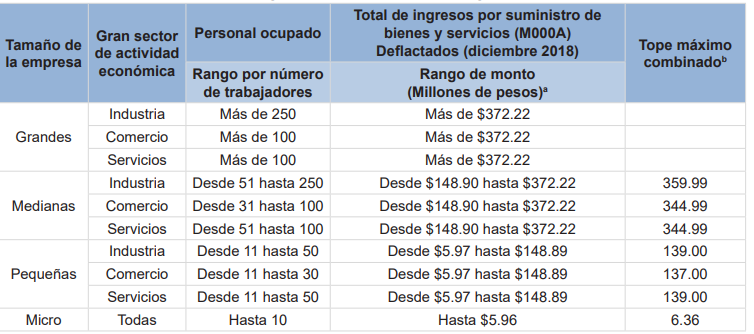
Población objetivo: Empresas del país, pertenecientes a grandes sectores de actividad económica, según el SCIAN (Sistema de Clasificación Industrial de América del Norte), que tuvieron alguna actividad de servicios con el extranjero, o alguna propiedad u operación en el exterior, cobertura nacional.

Sectorial: Permite tener estimaciones para los grandes sectores de actividad económica según SCIAN, siendo los siguientes:

* Sector industria, conformado por las actividades de minería, energía eléctrica, suministro de agua y gas, construcción y manufacturas.
* Sector comercio, integrado por las actividades del comercio al por mayor y comercio al por menor.
* Sector servicios, constituido por todos los servicios privados no financieros, incluyendo transportes.

Dominio de estudio: Denotan subconjuntos que fueron planeados para el diseño de la muestra:

* Nacional-gran sector de actividad económica (según el SCIAN).
* Nacional-tamaño de la empresa (denominado por el Diario Oficial de la Federación)

******

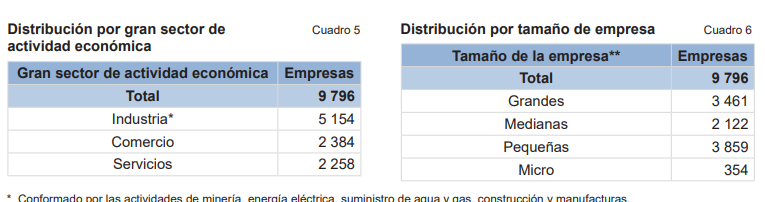
Unidad de Muestreo: Empresas del país de instalaciones fijas que reportan una actividad con el extranjero, o alguna relación de propiedad u operación con el exterior, según actividades de Industria, Comercio, Servicios privados no financieros.

Marco de muestreo: El marco se integró con 9,796 empresas, distribuidas por dominio, gran sector de actividad económica y tamaño de la empresa, que cumplen con los siguientes rubros:

Grupo I: Empresas que tienen información de al menos una de las 27 variables de gastos, en 13 variables de ingresos.

Grupo II: Pertenece algún holding, grupo empresarial, grupo financiero, empresa controladora, con participación de capital extranjero en el capital social en proceso integrados con empresas ubicadas en otros países.

1. Presenta semejanza estadística con el grupo I
2. Cuentan con personal ocupado mayor a 1000.



Esquema de muestreo:

Probabilístico: Las unidades de muestreo tienen una probabilidad conocida y distinta de cero de ser seleccionadas.

Estratificado: Las unidades de muestreo con características similares se agrupan para formar estratos.

**Estratificación:**

Se usa el método de Dalnius-Hodges para la primera estratificación, usando este método a 5 estratos, con el objetivo de definir las empresas en el estrato 1 de certeza para los grupos I y II.a

Grupo I: Se define una variable auxiliar nombrada bienes y servicios en el extranjero, usando la suma de las 40 variables auxiliares cuantitativas de gastos e ingresos en el extranjero.

Valor de la variable de bienes y servicios del extranjero mayor o igual a 33,404.57 o personal ocupado mayor a 1,000.

Grupo II.a y Grupo II.b: Se define la variable auxiliar nombrada bienes y servicios totales como la suma de variables total de gastos por consumo de bienes y servicios, total de ingresos por suministros de bienes y servicios.

Grupo II.a: Valor de la variable de bienes y servicios totales mayor o igual a 1,960,393.31 o personal ocupado mayor a 1,000.

Grupo II.b: Se determina como certeza.

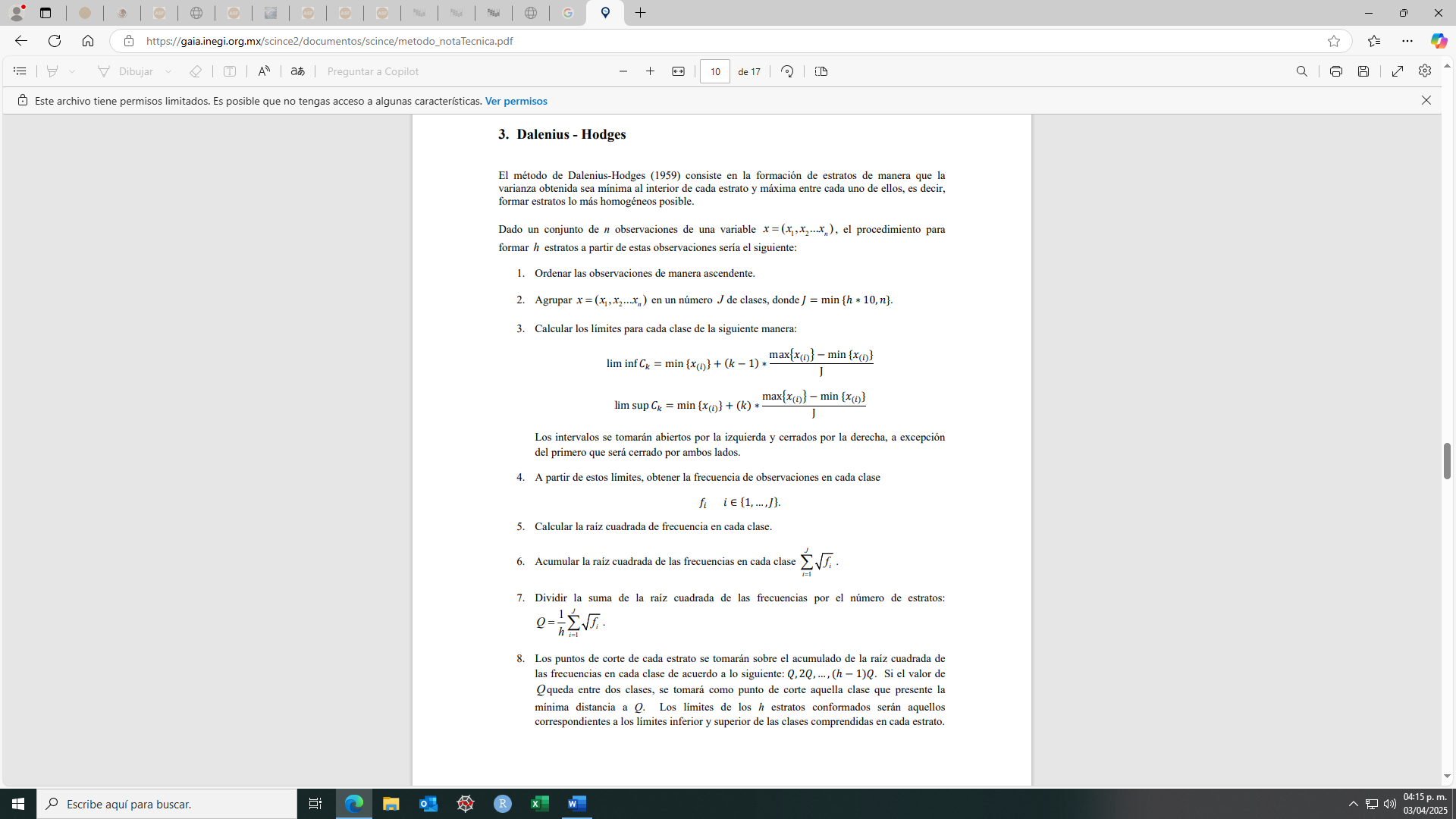
***Método Dalenius-Hodges:*** *Consiste en la formación de estratos, de forma que la varianza sea mínima al interior de cada estrato y maximiza entre cada estrato cada uno de ellos, forma estratos lo más homogéneos posible.*

Dado el conjunto de n observaciones de la variable X=(X1,X2,..,Xn), el procedimiento para formar h estratos, es el siguiente:

1-Ordenar las observaciones de manera ascendente.

2- Agrupar X= (X1, X2,..,Xn) en un numero de J clases, donde J = min {h \* 10, n}.

3- Calcular los limites para cada clase de la siguiente manera:



Los intervalos se tomaran abiertos por la izquierda y cerrados por la derecha, exceptuando el primero, que sera cerrado de ambos lados.

4- Con los limites, obtener la frecuencia de observaciones en cada clase.

5-Acumular la raiz cuadrada de frecuencia en cada Clase.

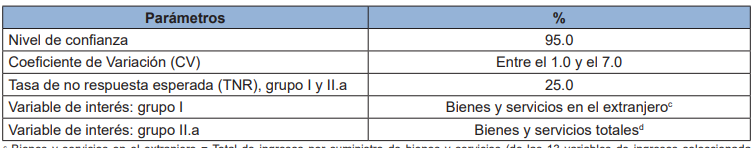
6- Acumular la raiz cuadrada de las frecuencias en cada clase.

7- Dividir la suma de la raíz cuadrada de las frecuencias por el numero de estratos.

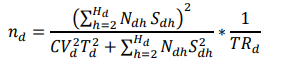
8-Los puntos de corte de cada estrato se toman sobre el acumulado de la raíz cuadrada de las frecuencias de cada clase, según lo siguiente Q, 2Q, (h-1) Q, si el valor de Q queda entre dos clases, se tomará como punto de corte aquella clase que presente la mínima distancia de Q.

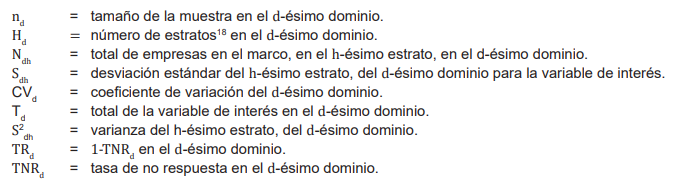
**Tamaño de muestra:**

Se calcula de forma independiente para gran sector de la actividad económica y tamaño de la empresa considerando los parámetros.



Calculándose con la siguiente ecuación:





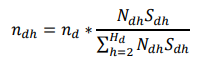
El tamaño de muestra resultante se obtiene después de la afijación correspondiente a nivel gran sector de actividad económica, tamaño de la empresa y estrato.

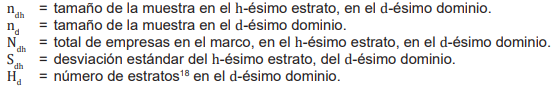
**Afijación de la muestra:**

Para cada dominio de estudio se distribuye la muestra en cada estrato con el método de Neyman, el tamaño de la muestra total afijada de dominios es de 5,539 empresas considerando para:

Grupo I: variable bienes y servicios en el extranjero.

Dominio II bienes y servicios totales para el dominio.





**Calculo de factores de expansión.**

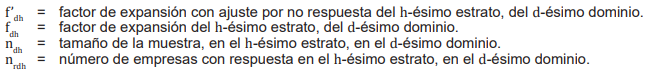
Se define como el inverso de la probabilidad de seleccionar una empresa de h-esimo estrato y de d-esimo dominio de estudio, utilizando el siguiente calculo:





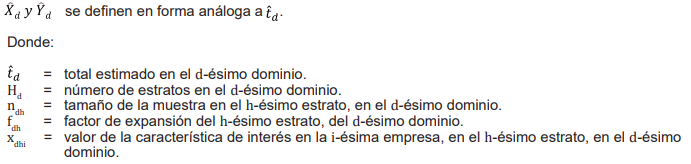
**Ajuste por no respuesta (realizado a nivel dominio - estrato).**

****

****

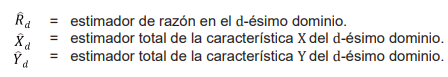
**Estimadores (estimador total nacional de característica X).**

****

****

Para la estimación de proporciones de tasas y promedios se usa el estimador de razón:

****



Estimador de la varianza del estimador total de característica X:



