|  |  |  |
| --- | --- | --- |
| LogoIN | DEPARTAMENTO DE MATEMÁTICA  **GUÍA N°2**  **MUESTREO**  **v. 1.1** | 1er Semestre 2025 |
| Probabilidades y Estadística Descriptiva e Inferencial |
| **Coordinador:**  Philippe Vandeperre Venegas |

**Objetivo de aprendizaje:** Identificar el mecanismo de muestreo empleado en diversas situaciones de recolección de datos.

Uno de los objetivos principales de la Estadística es extraer conclusiones e información sobre una determinada **población**. Recordemos que por población se denomina al conjunto completo de elementos, con alguna característica común, que es objeto de nuestro estudio (personas, objetos, mediciones, etc.). Evidentemente, la forma más directa de cumplir dicho objetivo sería estudiar todos y cada uno de los elementos de la población (**censar**). Sin embargo, en numerosas ocasiones esto no es posible ya que, por ejemplo, el tamaño de la población puede ser demasiado grande (ej. estrellas en la Vía Láctea) e incluso infinito. En ocasiones estudiar los elementos supone la destrucción de estos (ej. ensayos destructivos de control de calidad). En el caso más habitual, el coste económico de medir/observar a la población es prohibitivo. En estos casos, es necesario trabajar con un subconjunto de elementos de la población, es decir, una **muestra**. Al proceso de obtener muestras se le denomina **muestreo**.

En Estadística se suele designar por **N** al tamaño de la población y por **n** al de la muestra. La decisión de qué tan grande debe ser la muestra pasa la mayoría de los casos por criterios probabilísticos que abordaremos más adelante en este curso (el segundo semestre calcularemos qué tan grande debe ser una muestra para obtener cierto grado de certeza respecto a las inferencias que queremos realizar sobre la población).



Para poder estudiar correctamente una población mediante la inferencia estadística es fundamental que la muestra esté bien escogida. Si la muestra no es representativa de la población, **todas las conclusiones que obtengamos de ella no tendrán utilidad.**

*"Gargabe in, garbage out"*

La clave de un proceso de muestreo es que **la muestra sea representativa de la población**. Una forma de conseguir esto es introducir un grado de **aleatoriedad** en la elección de los elementos que compondrán la muestra. La forma más elemental de muestreo aleatorio es el **muestreo aleatorio simple**.

**Muestreo aleatorio**

Una **muestra aleatoria simple (m.a.s.)** de *n* sujetos se selecciona de modo que cada muestra posible del mismo tamaño *n* tiene la misma posibilidad de ser elegida. Con frecuencia, una muestra aleatoria simple se denomina muestra aleatoria, pero en sentido estricto una *muestra aleatoria* tiene el requisito más débil de que todos los miembros de la población tengan la misma posibilidad de ser seleccionados. Esta distinción no será importante en este curso.

**Actividad (recuerdo):**

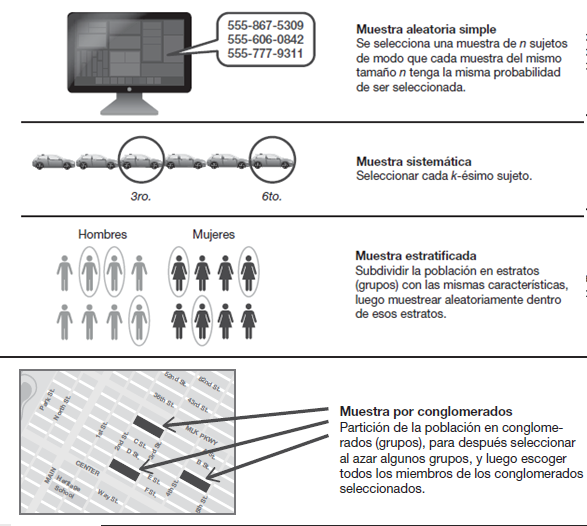
*1. Si se tiene una población finita de tamaño N, ¿cuántas muestras de tamaño n pueden obtenerse, si...*

*a) ... se extrae la muestra sin reemplazo (reposición)?*

*b) ... se extrae la muestra con reemplazo?*

*2. Si la población es infinita, ¿es realmente relevante si la extracción se hace con reemplazo o no? Discutir e investigar.*

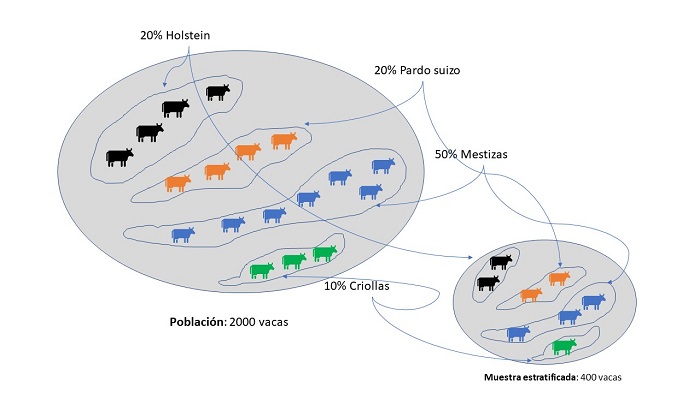
Además del muestreo aleatorio simple, a continuación se presentan algunos otros métodos de muestreo comúnmente utilizados en Estadística. La siguiente figura ilustra algunos de estos diferentes métodos de muestreo.



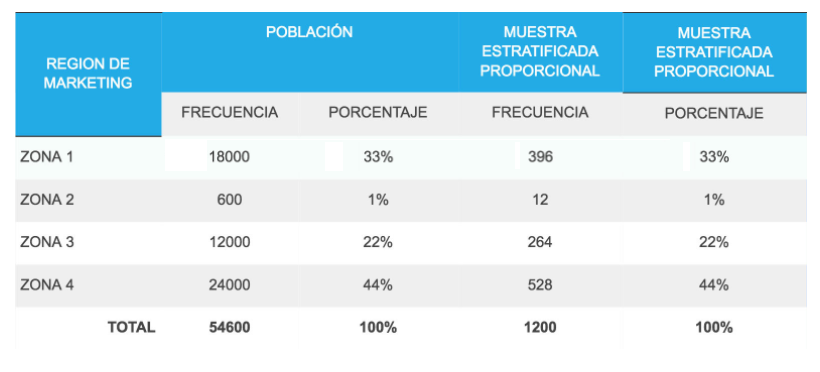
**Muestreo estratificado**

El **muestreo estratificado** es un procedimiento de muestreo en el que se considera que la población no es homogénea en relación a los objetivos de la investigación. Entonces, la población se separa en segmentos exclusivos (disjuntos), cada uno de ellos homogéneo respecto a la característica que se va a estudiar, y luego una muestra aleatoria simple se selecciona de cada segmento (estrato). Las muestras seleccionadas de los diversos estratos se combinan en una sola muestra para análisis.

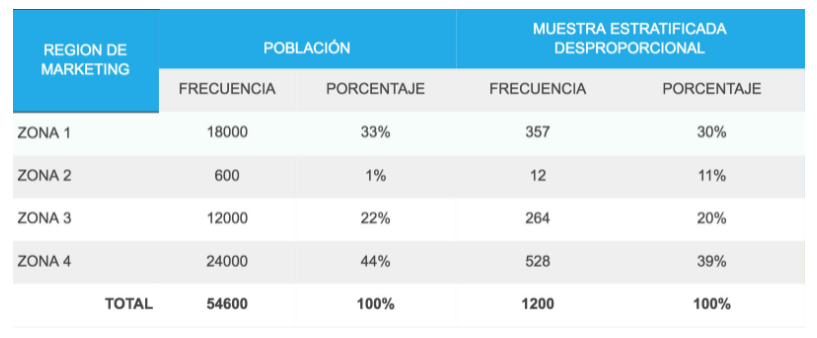
Es relativamente habitual definir estratos de acuerdo a algunas variables sociodemográficas de la población como la edad, sexo, clase social o región geográfica. Estas variables permiten dividir fácilmente la muestra en grupos mutuamente excluyentes y, con bastante frecuencia, permiten discriminar comportamientos diferentes dentro de la población.



Hay dos subtipos principales de muestreo estratificado: el **proporcional** y el muestreo **desproporcionado**. En el estratificado proporcional, el número de elementos extraídos de los diversos estratos es proporcional a la representación de los estratos de la población objetivo. Es decir, el tamaño de la muestra extraída de cada estrato es proporcional con el tamaño relativo de ese estrato de la población. Este procedimiento de muestreo se utiliza cuando el propósito de la investigación es estimar los parámetros poblacionales.



El **muestreo estratificado desproporcionado** es un procedimiento en que el número de elementos incluidos en la muestra de cada estrato no es proporcional a su representación en la población total. Los elementos de la población no tienen la misma oportunidad de ser incluidos en la muestra. La misma fracción de muestreo no se aplica a cada estrato. Por otra parte, los estratos tienen diferentes fracciones de muestreo, y como tal, este procedimiento de muestreo no garantiza la equiprobabilidad de los elementos de ser escogidos en la muestra. Con el fin de estimar los parámetros de población, la composición poblacional debe utilizarse para compensar la desproporción en la muestra (esto quiere decir que el método requiere un conocimiento previo de la población). Sin embargo, para algunos proyectos de investigación el muestreo estratificado desproporcionado puede ser más apropiado que el proporcional.



***Actividad:***

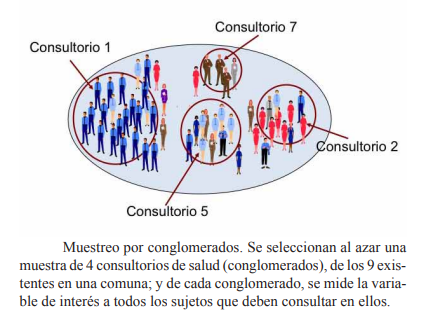
*La probabilidad de que una oveja esté enferma está directamente relacionada con la edad. Supongamos que se tiene una población de 250 ovejas, donde el 44% de las ovejas son de menos de 2 años, el 28% son de 3 a 4 años, el 18% son de 5 a 6 años y el 10% son de más de 6 años. Se establece (luego veremos cómo) que la muestra debe ser de 61 ovejas. Queremos realizar un muestreo aleatorio estratificado proporcional. Determine cuántas ovejas debemos extraer de cada estrato para conformar la muestra.*

**Muestreo por conglomerados**

El **muestreo por conglomerados**, también conocido como muestreo por racimos, es un procedimiento de muestreo probabilístico en que los elementos de la población son seleccionados al azar en forma natural por agrupaciones (*clusters*). Los elementos del muestreo se seleccionan de la población de manera individual, uno a la vez.

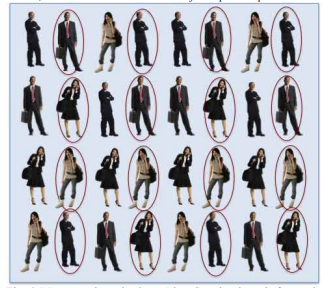
Las unidades de muestreo o grupos pueden ser espaciados, tal como ocurre naturalmente en las unidades geográficas o físicas (por ejemplo: estados, delegaciones o distritos); en base a una organización como escuelas, grado escolar; o servicio telefónico tales como códigos de área.

La heterogeneidad del grupo es fundamental para un buen diseño del muestreo por conglomerados. Es decir, los elementos dentro de cada grupo debe representar la variabilidad de la población objetivo. Muchas veces este principio no se cumple, lo que aumenta el error de muestreo, pero de todas formas se procede por un asunto de costos. Los conglomerados son parecidos entre sí.

****

**Muestreo aleatorio sistemático**

Se emplea cuando el criterio de distribución de los sujetos a estudio en una serie es tal, que los más similares tienden a estar más cercanos. Este tipo de muestreo suele ser más preciso que el aleatorio simple, debido a que recorre la población de forma más uniforme. De este modo, se seleccionará cada "ésimo" caso. Por ejemplo: ante la pregunta ¿cuál es la muestra necesaria para establecer la prevalencia de cambios inflamatorios en biopsias hepáticas de pacientes con cáncer al hígado? Un muestreo aleatorio sistemático aplicaría de la siguiente forma: entre todos los sujetos con cáncer hepático, seleccionar aquellos que ingresan los días impares del mes, o aquellos cuyo primer dígito del RUT sea par, hasta completar la muestra estimada. El método más común es escoger un primer elemento de la lista al azar, y luego sacar uno cada “k”, hasta completar el tamaño de la muestra al recorrer la lista completa.

****

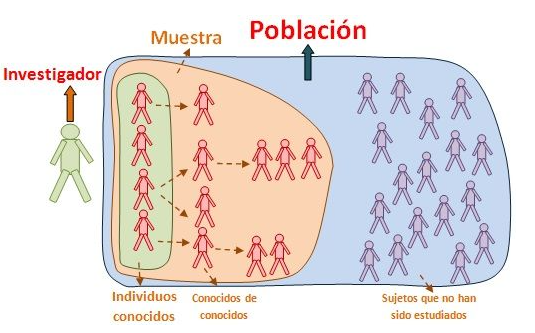
**Muestreo no aleatorio**

En ocasiones, y por diversos motivos, no es posible o fácil realizar un muestreo aleatorio, por lo que se recurre a un proceso de selección con una base distinta a las consideraciones de probabilidad. En Estadística existen diversos mecanismos de muestreo no aleatorio, basados, por ejemplo, en el juicio de un experto, la conveniencia en el acceso a los datos u otros.

**Muestreo por cuotas**: es la técnica más difundida sobre todo en estudios de mercado y sondeos de opinión, y es semejante al muestreo aleatorio estratificado en el sentido de que la población se encuentra dividida en grupos (por ejemplo, género o edad) de los cuales se obtendrá, en este caso por medio de un procedimiento no aleatorio, la cota o cuota de elementos que constituirán la muestra para estudiar u observar una característica de interés.

Ejemplo: En la ciudad de México se desea conocer la preferencia de compra de un producto para el arreglo personal, luego se elegirán personas que pasen por varios puntos fijos establecidos en el norte, sur, el este y el oeste de la metrópoli, descartando niños y adultos mayores, para encuestar solo personas con poder adquisitivo, escogidas por el recolector o investigador, hasta llegar a la cota de elementos deseados de la muestra.

**Muestreo bola de nieve**: Indicado para estudio de poblaciones clandestinas, minoritarias o muy dispersas pero en contacto entre sí (consumidores de droga, coleccionistas, pacientes con una enfermedad rara, etc.). Consiste en identificar sujetos que se incluirán en la muestra a partir de los propios entrevistados, partiendo con una pequeña cantidad de individuos que cumplen los requisitos necesarios y que sirven como localizadores (o reclutadores) de otros con características análogas.



**Muestreo por juicio**: se establece a partir del juicio (criterio) de la persona que hace el muestreo, el cual se basa en su experiencia, su práctica o formación profesional.

**Muestreo según el número de muestras extraídas**

En muchas ocasiones, y dependiendo de la naturaleza del estudio o de la población, no solo se extrae una única muestra, pueden ser varias.

**Muestreo simple**: es cuando se toma una muestra de una población dada para el propósito de inferencia estadística. Es importante en este tipo de muestreo que el tamaño de la muestra sea lo suficientemente grande para extraer una conclusión válida.

**Muestreo doble:** es utilizado cuando el resultado del estudio de la primera muestra no es decisivo, por lo que entonces es necesario extraer una segunda muestra de una misma población. Las dos muestras son combinadas para analizar los datos y tomar una decisión.

Por ejemplo, supóngase una máquina industrial que elabora ciertas piezas. Por más que se ajuste la máquina, ninguna pieza es idéntica a otra, ya que en sus dimensiones y peso pueden presentarse pequeñas variaciones. Se trata de determinar si un lote de piezas fabricadas por dicha máquina cumple el criterio de tolerancia para que el mismo sea aceptado o rechazado.

En primer lugar, se toma una muestra aleatoria de piezas con la que se desea chequear si una de las variables, por ejemplo, el largo de la pieza, está dentro de la tolerancia.

En caso de que el largo promedio esté por debajo o por encima del grado de tolerancia deseada para dicha variable, en esta primera muestra, entonces se infiere que el lote es defectuoso y el mismo ha de ser desechado. En este caso no se requiere tomar nuevas muestras.

Por el contrario, si el valor promedio está dentro del rango de tolerancia, pero la desviación estándar de la muestra es lo suficientemente grande como para que la suma o la resta al valor promedio quede fuera del margen, entonces será necesario recabar una segunda muestra mayor.

Esta segunda muestra debe incluir a la muestra original para rehacer los cálculos y así poder tomar una decisión definitiva respecto a la variable investigada. De esta manera  puede saberse si el lote es defectuoso o no.

**Muestreo en etapas múltiples:** Con frecuencia, los encuestadores profesionales y los investigadores gubernamentales o científicos recopilan datos utilizando una combinación de los métodos de muestreo anteriores. En un diseño en etapas múltiples, los encuestadores seleccionan una muestra en diferentes etapas, y cada una de ellas puede utilizar distintos métodos de muestreo, como en el ejemplo siguiente.

**Ejemplo:**

Las estadísticas de desempleo del gobierno de Estados Unidos se basan en encuestas a domicilio. No es práctico examinar personalmente cada hogar en una muestra aleatoria simple, ya que estarían dispersos por todo el país. En cambio, la Oficina de Censos de Estados Unidos y la Oficina de Estadísticas Laborales colaboran para llevar a cabo un estudio llamado Encuesta de la Población Actual. Una encuesta reciente incorpora un diseño muestral en etapas múltiples, siguiendo en general los pasos descritos a continuación:

**1.** La totalidad de Estados Unidos se divide en 2,007 regiones llamadas *unidades primarias de muestreo* (PSU por sus siglas en ingles). Las unidades de muestreo primarias son áreas metropolitanas, condados grandes o combinaciones de condados más pequeños. Las 2,007 unidades primarias de muestreo se agrupan en 824 estratos.

**2.** En cada uno de los 824 estratos se selecciona una de las unidades de muestreo primarias de manera que la probabilidad de selección sea proporcional al tamaño de la población en cada unidad de muestreo primario.

**3.** En cada una de las 824 unidades de muestreo primario seleccionadas, los datos del censo se usan para identificar un *distrito de enumeración* del censo, cada uno con aproximadamente 300 hogares. Los distritos de enumeración se seleccionan al azar.

**4.** En cada uno de los distritos de enumeración seleccionados, los grupos de aproximadamente cuatro direcciones (contiguas siempre que sea posible) se seleccionan al azar.

**5.** Una persona responsable en cada uno de los 60,000 hogares seleccionados es entrevistada sobre la situación laboral de cada miembro del hogar de 16 años de edad o más.

Este diseño muestral en etapas múltiples incluye una combinación de muestreo aleatorio, estratificado y de conglomerados en diferentes etapas. El resultado final es un diseño muestral muy complicado, pero es mucho más práctico, menos costoso y más rápido que usar un diseño más sencillo, como una muestra aleatoria simple.

***Actividad:*** *Identificar los diferentes mecanismos de muestreo utilizados en los distintos pasos del ejemplo recién presentado.*

**Bibliografía**

Muestreo estratificado, un tipo de muestreo de probabilidad:

https://www.questionpro.com/blog/es/muestreo-estratificado/

Muestreo por conglomerados: Un tipo de muestreo probabilístico:

https://www.questionpro.com/blog/es/muestreo-por-conglomerados/

Otzen, T. & Manterola, C. (2017) *Técnicas de Muestreo sobre una Población a Estudio*

https://scielo.conicyt.cl/pdf/ijmorphol/v35n1/art37.pdf

Roldán, P. L. & Fachelli S. (2008) *Metodología de la investigación social cuantitativa*, Universitat Autonoma de Barcelona