

Estadística es una ciencia utilizada para la toma de decisiones en situaciones de incertidumbre. Se encarga de diseñar, recolectar, describir, analizar e interpretar la información y, por ello, constituye la metodología científica principal que permite transformar datos en información. Esto la hace especialmente atractiva, ya que en cualquier sociedad desarrollada existe una inmensa variedad de problemas cuya solución sólo es posible mediante la utilización de técnicas estadísticas. La palabra estadística se deriva del vocablo "estado".

La estadística nos permite realizar inferencias y sacar conclusiones a partir de los datos. Extrayendo la información contenida en los datos podremos comprender mejor las situaciones que ellos representan. Los métodos estadísticos abarcan todas las etapas de la investigación, desde el diseño hasta el análisis final de los datos.

Podemos distinguir tres grandes etapas:

Diseño: Planeamiento y desarrollo de las investigaciones.

Descripción: Resumen y exploración de los datos.

Inferencia: Predicciones y toma de decisiones sobre las características de una población en base a la información recogida en una muestra de la población.

1. Diseño

En esta etapa se define cómo se desarrollará la investigación con el fin de responder las preguntas que le dieron origen. Un diseño bien realizado puede ahorrar esfuerzos en etapas posteriores y redundar en un análisis más sencillo. Esta etapa es crucial, pues un estudio pobremente diseñado o con datos incorrectamente recolectados o registrados puede ser incapaz de responder las preguntas que originaron el estudio.

Una vez formulado el problema, en la etapa de Diseño se definirá, entre otras cosas, la población objetivo, los tamaños de muestra, los mecanismos de selección de individuos, los criterios de inclusión y exclusión de sujetos, los métodos de asignación de tratamientos, las variables que se medirán y cómo se entrenará al equipo de trabajo para el cumplimiento del protocolo.

2. Descriptiva

Los métodos de Análisis Exploratorio o Estadística Descriptiva ayudan a comprender la estructura de los datos, a manera de detectar tanto un patrón de comportamiento general como apartamientos del mismo. Una forma de realizar esto es mediante gráficos de sencilla realización e interpretación. Otra forma de describir los datos es resumiéndolos en uno, dos o más números que caractericen al conjunto de datos con fidelidad. Explorar los datos permitirá detectar los erróneos o inesperados y nos ayudará a decidir qué métodos estadísticos pueden ser empleados en etapas posteriores del análisis para obtener conclusiones válidas.

3. Inferencia

Finalmente, la Inferencia Estadística nos permite hacer predicciones y estimaciones como decidir entre dos hipótesis opuestas relativas a la población de la cual provienen los datos (test de hipótesis). La calidad de las estimaciones puede ser muy variada y está afectada por errores. La ventaja de los métodos estadísticos es que, aplicados sobre datos obtenidos a partir de muestras aleatorias, permiten cuantificar el error que podemos cometer en una estimación o calcular la probabilidad de cometer un error al tomar una decisión en un test de hipótesis.

En este capítulo solo se abordará el diseño; la descripción e inferencia se desarrollarán en los capítulos posteriores.

DISEÑO

Consiste en definir y calcular los valores de la población, muestra, unidad de estudio (individuo), las variables, tamaño de muestra y tipo de muestreo del problema o experimento.

Comúnmente, los experimentos científicos producen observaciones o hallazgos que se expresan como datos numéricos. Las mediciones que se obtienen de los sujetos bajo estudio constituyen los datos. Generalmente, los datos son medidas de la variable. Los datos crudos u originales son aquellos datos obtenidos directamente del experimento, es decir, no han sido sometidos a tratamiento estadístico.

POBLACIÓN

Conjunto de todos los individuos (personas, objetos, animales, etc.) que porten información sobre el fenómeno que se estudia. Representa una colección completa de elementos (sujetos, objetos, fenómenos o datos) que poseen algunas características comunes. Es el conjunto de elementos más grande del cual se puede tomar una muestra representativa para el experimento científico.

La población constituye el conjunto de elementos que forma parte del grupo de estudio, por tanto, se refiere a todos los elementos que en forma individual podrían ser cobijados en la investigación. La población la define el objetivo o propósito central del estudio y no estrictamente su ubicación o límites geográficos, u otras características particulares al interior de ella.

MUESTRA

Constituye una selección al azar de una porción de la población, es decir, un subconjunto que seleccionamos de la población.

La muestra, por otro lado, consiste también en un grupo reducido de elementos de dicha población, al cual se le evalúan características particulares, generalmente, con el propósito de inferir tales características a toda la población.

INDIVIDUO (Observación, Caso, Sujeto)

Cualquier elemento que aporte información sobre el fenómeno que se estudia. El elemento o individuo muestral se refiere a la unidad más pequeña en que se puede descomponer una muestra.

VARIABLE

Es una característica de la población que le interesa al investigador y que puede tomar diferentes valores. La variable es toda aquella propiedad de algún objeto, persona o evento que posee diferentes valores. El término *variable* implica una característica que cambia.

Variable independiente: es la que habrá de ser manipulada en el experimento, es decir, aquella que cambia y posee el potencial de afectar la variable dependiente. Por lo tanto, este tipo de variable es controlada sistemáticamente por el investigador. Durante el experimento científico se busca hallar el efecto que tienen una o más variables (independientes) sobre otras variables (dependientes).

Variable dependiente: se determina antes o después de haber tratado la variable independiente. Representa la medida que se emplea para determinar el efecto de la variable independiente.

Variables cualitativas: son aquellas que se refieren a cualidades o atributos no medibles en números. Por ejemplo, la variable 'estado civil' puede ser: casado, viudo, soltero. Se subdivide a su vez en:

- **Cualitativa Nominal:** Surge cuando se definen categorías y se cuenta el número de observaciones sin considerar el orden. Por ejemplo, la variable 'color de ojos' puede ser: castaños, marrón, azul, etc.
- **Cualitativa Ordinal:** se tiene en cuenta el orden de acuerdo al grado que posee una determinada característica. Por ejemplo, la variable 'estudio' puede ser: 1º grado, 2º grado, 3º grado, etc.

Variables cuantitativas: son las susceptibles de medirse en términos numéricos. Se subdividen a su vez en:

- **Cuantitativas continuas.** Pueden asumir cualquier valor (números reales). Por ejemplo: estatura, peso, cantidad de colesterol, etc.
- **Cuantitativas discretas.** Asumen solo valores enteros (números enteros). Por ejemplo: número de hijos, número de trabajadores, etc.

EJERCICIOS DE POBLACIÓN Y MUESTRA

EJERCICIO 1

Objetivo: Evaluar el precio de la vivienda en la ciudad de Lima.

Población: La población será el total de las viviendas de ciudad de Lima. Así, si se estudia el precio de la vivienda de una ciudad, lo normal será no recoger información sobre todas las viviendas de la ciudad (sería una labor muy compleja).

Muestra: subgrupo o subconjunto de las viviendas de la ciudad de Lima seleccionadas aleatoriamente, que se entienda que es suficientemente representativo.

Individuo (Observación, Caso, Sujeto): una vivienda de dicha ciudad de Lima.

Variable: precio de la vivienda, tamaño, cantidad de habitaciones, ubicación, antigüedad, etc.

Variable independiente: tamaño de la vivienda, cantidad de habitaciones, ubicación, antigüedad.

Variable dependiente: precio de la vivienda (porque el precio depende de la ubicación, antigüedad, tamaño, etc.)

Variables cualitativas: tamaño (grande, mediano, pequeño), ubicación (norte, centro, sur).

- **Cualitativa Nominal:** ubicación (norte, centro, sur).
- **Cualitativa Ordinal:** tamaño (grande, mediano, pequeño).

Variables cuantitativas:

- **Cuantitativas continuas:** precio (soles).
- **Cuantitativas discretas:** antigüedad (años), cantidad de habitaciones (N° entero).

Una variable independiente se puede convertir en dependiente o viceversa; del mismo modo una variable cualitativa se puede convertir en cuantitativa o viceversa.

EJERCICIO 2

Objetivo: evaluar la percepción de los estudiantes de un colegio en torno a los niveles de ruido que existen en el plantel.

Población: está conformada por todos los estudiantes del colegio, incluidos aquellos de programas diurnos y nocturnos.

Muestra: La muestra puede tomarse eligiendo al azar un puñado de alumnos con base en un archivo o listado que contenga todos los códigos o nombres de los estudiantes.

Individuo (Observación, Caso, Sujeto): un alumno del colegio con base en un archivo o listado que contenga todos los códigos o nombres de los estudiantes.

Variable: edad del alumno, género (masculino, femenino), peso del alumno, niveles de ruido, tipo de ambiente, distancia del ambiente etc.

Variable independiente: edad del alumno, género (masculino, femenino), peso del alumno, tipo de ambiente, distancia del ambiente.

Variable dependiente: niveles de ruido.

Variables cualitativas: género (masculino, femenino), tipo de ambiente (cerrado, despegado), niveles de ruido.

Variables cuantitativas: edad y peso del alumno, distancia del ambiente (metros).

- **Cuantitativas continuas:** edad y peso del alumno, distancia del ambiente.
- **Cuantitativas discretas:** ninguna variable se ajusta a este tipo.

TAMAÑO DE MUESTRA (MUESTREO ALEATORIO SIMPLE)

En una investigación el tamaño de muestra es muy importante. Teniendo en cuenta que la calidad y validez de los resultados de una investigación dependen del tamaño de muestra. Una demasiado grande implica un desperdicio de recursos y una muestra demasiado pequeña disminuye la utilidad de los resultados.

En nuestra investigación utilizaremos el **Muestreo Aleatorio Simple** (Raj, D): Teoría del Muestreo El tamaño de muestra en el **Muestreo Aleatorio Simple** se calcula con la fórmula siguiente (Spigel, 1978: 161):

$$n = \frac{n_0}{1 + \frac{n_0}{N}} \quad \dots \dots \dots (1)$$

$$n_0 = \frac{z_{\alpha}^2 \sigma^2}{E^2} \quad \dots \dots \dots (2)$$

Donde:

- n : Tamaño de muestra.
- n_0 : Tamaño de muestra aproximado.
- N : Tamaño de la población bajo Estudio.
- Z_{α} : Valores correspondiente al nivel de Significancia.
- E : Error de tolerancia de la estimación.
- α : Nivel de significancia.
- σ^2 : varianza de la variable.

1. **El estudio de proporciones:** la varianza es igual a $PQ(\sigma^2)$ donde P denota la proporción estimada o esperada de la variable; si no se conoce tal valor, se reemplaza por 0.5 ($P=1/2$ y $Q=1-P$); la fórmula quedaría de la siguiente manera:

$$n_0 = \frac{z_{\alpha}^2 \sigma^2}{E^2} = \frac{z_{\alpha}^2 PQ}{E^2} \quad \dots \dots \dots (3)$$

2. **Error muestral (E) o Error de Estimación:** es el error a causa de observar una muestra en lugar de la población completa. Este valor depende del investigador y se encuentra entre 0% y 10%, en algunas ocasiones es mayor a 10%.

3. El nivel de confianza: es la probabilidad a priori de que el intervalo de confianza a calcular contenga al verdadero valor del parámetro. Se indica por $1-\alpha$ y habitualmente se da en porcentaje $(1-\alpha)\%$. Los valores que se suelen utilizar para el nivel de confianza son el 95%, 99% y 99,9%

4. Ejemplos de nivel de confianza (Ver tabla en apéndice)

Para un nivel de confianza del 88%:

$1-\alpha=0.88$, $\alpha=0.12$ y $\alpha/2=0.06$. Ver tabla de distribución normal: el valor 0.94 ($1-\alpha/2$). Se observa que para 0.9406 el valor en la fila es 1.5 y el valor en la columna es 0.06 que representa el segundo decimal, entonces $Z(1\%)=1.56$

Para un nivel de confianza del 98%:

$1-\alpha=0.98$, $\alpha=0.02$ y $\alpha/2=0.01$. Ver tabla de distribución normal: el valor 0.99 ($1-\alpha/2$). Se observa que para 0.9906 el valor en la fila es 2.3 y el valor en la columna es 0.05 que representa el segundo decimal, entonces $Z(1\%)=2.35$

Para un nivel de confianza del 95%:

$1-\alpha=0.95$, $\alpha=0.05$ y $\alpha/2=0.025$. Ver tabla de distribución normal: el valor 0.975 ($1-\alpha/2$). Se observa que para 0.975 el valor en la fila es 1.9 y el valor en la columna es 0.06 que representa el segundo decimal, entonces $Z(2.5\%)=1.96$

EJERCICIOS DE TAMAÑO DE MUESTRA

EJERCICIO 1

Problema: Se intenta estudiar el precio de la vivienda de la ciudad de Lima.

Solución:

Recoger información sobre todas las viviendas de la ciudad sería una labor muy compleja. Por tanto se halla un tamaño muestra:

Datos:

Población (N): Desconocida entonces, $N=\infty$.

P: Desconocido entonces, $P=0.5$ y $Q=0.5$

E: Este valor depende del investigador (recomendable=5%).

α : Desconocido entonces, $\alpha = 5\%$; $1-\alpha = 95\%$; $\alpha/2 = 2.5\%$, por tanto

$Z_{\alpha/2} = 1.96$ (Tabla Normal).

$n = ?$

Remplazando valores en la fórmula se tiene:

$$n_0 = [(1.96)^2 (0.5)^2] / [(0.05)^2] = 384.16$$

Luego:

$$n = (384.16) / (1 + 384.16 / \infty) = (384.16) / (1 + 0) = 384.16$$

Como nuestra unidad (sujeto, elemento) es una vivienda, se tomarán 384 viviendas de la ciudad de Lima para realizar el estudio.

Observaciones:

- Cuando $N=\infty$ se demuestra que $n=n_0$
- Muchos autores consideran el valor de $Z_{\alpha/2} = 2$, donde $\alpha = 5\%$:

$$\text{Entonces } n_0 = [(2)^2 (0.5)^2] / [(0.05)^2] = 400$$

Como $N=\infty$ se tiene que $n_0 = n = 400$,

Este motivo origina que muchos estudios sencillos se realicen con tamaño de muestra igual 400.

EJERCICIO 2

Problema: Evaluar la percepción de los estudiantes de un colegio en torno a los niveles de ruido que existen en el plantel conformado por 1,000 estudiantes, incluidos aquellos de programas diurnos y nocturnos.

Solución:

Hallando el tamaño muestra:

Datos:

Población (N): $N = 1000$.

P: Desconocido entonces, $P=0.5$ y $Q=0.5$

E: Este valor depende del investigador (recomendable=5%)

α : Desconocido entonces, $\alpha = 5\%$; $1-\alpha = 95\%$; $\alpha/2 = 2.5\%$, por tanto

$Z_{\alpha/2} = 1.96$ (Tabla Normal).

$n = ?$

Remplazando valores en la fórmula se tiene:

$$n_0 = [(1.96)^2 (0.5)^2] / [(0.05)^2] = 384.16$$

Luego:

$$n = (384.16) / (1 + 384.16 / 1000) = (384.16) / (1.38416) = 277.54$$

Como nuestra unidad (sujeto, elemento) es un alumno, se tomarán 288 alumnos del colegio para realizar el estudio.

EJERCICIO 3

Problema: Se pretende determinar la prevalencia de una enfermedad en una población de 250 ovejas: se supone una prevalencia del 30% y se desea una precisión del 10% para un nivel de confianza del 95%.

Solución:

Datos:

Población: $N = 250$ ovejas

Unidad de estudio: una oveja

$P=30\%$ entonces $Q=70\%$

$E = 10\%$

$\alpha = 95\%; \alpha = 5\% ; \alpha/2 = 2.5\%$ entonces;

$Z_{\alpha/2} = 1.96$ (Tabla Normal)

$n = ?$

Remplazando valores en la fórmula:

$$n_0 = (1.96)^2(0.30)(0.70)/(0.10)^2 = 80.67$$

Luego:

$$n = (80.67)/(1+80.67/250) = 60.99$$

El tamaño de muestra será 61 ovejas.

EJERCICIO 4

Problema: Se pretende determinar el tiempo que demora un proceso penal en la Corte Superior de Lima Norte, sabiendo que existen 7470 expedientes con proceso penal en la sede de la Corte Superior.

Solución:

Datos:

Población: $N = 7470$ expedientes

$P=50\%$ entonces $Q=50\%$

$E = 5\%$

$\alpha = 95\%; \alpha = 5\% ; \alpha/2 = 2.5\%$ entonces;

$Z_{\alpha/2} = 1.96$ (Tabla Normal)

$n = ?$

Remplazando valores en la fórmula:

$$n_0 = [(1.96)^2(0.5)^2]/[(0.05)^2] = 384.16$$

Luego:

$$n = (384.16)/(1+384.16/7470) = 365.37$$

El tamaño de muestra será 365 expedientes con proceso penal, para realizar el estudio.

n : cantidad de expedientes penales donde se realizará el estudio.

N : cantidad total de expedientes penales existentes en la sede de la Corte Superior.

TIPOS DE MUESTREO

Existen dos tipos: el muestreo por selección intencionada o muestreo de conveniencia y el muestreo aleatorio (probabilístico).

EL MUESTREO DE CONVENIENCIA

Consiste en la elección por métodos no aleatorios de una muestra cuyas características sean similares a las de la población objetivo. En este tipo de muestreos la "representatividad" la determina el investigador de modo subjetivo, siendo este el mayor inconveniente del método ya que no podemos cuantificar la representatividad de la muestra.

Casi siempre presenta sesgos y, por tanto, debe aplicarse únicamente cuando no existe alternativa. En algunos casos, especialmente cuando se requiere una estrecha colaboración por parte de los ganaderos o veterinarios de campo, es la única opción para que el estudio sea viable. Supongamos que queremos realizar un estudio longitudinal consistente en tomar muestras de los animales de la explotación cada mes o llevar diariamente registros determinados de la granja, para lo cual, la mejor opción será realizar el estudio en granjas de confianza que permitan las manipulaciones y tengamos garantías de que el trabajo se llevará a cabo correctamente.

También puede ser útil cuando se pretende realizar una primera prospección de la población o cuando no existe un marco de la encuesta definido. Este tipo de muestreos puede incluir individuos próximos a la media o no, pero casi nunca representará la variabilidad de la población, que normalmente quedará subestimada.

MUESTREO ALEATORIO

En el muestreo aleatorio todos los elementos tienen la misma probabilidad de ser elegidos. Los individuos que formarán parte de la muestra se elegirán al azar mediante números aleatorios. Existen varios métodos para obtenerlos, siendo los más frecuentes la utilización de tablas de números aleatorios o generarlos por ordenador.

El muestreo aleatorio puede realizarse de distintas maneras, las más frecuentes son el muestreo simple, el sistemático, el estratificado y el muestreo por conglomerados.

1. Muestreo aleatorio simple

Es el método conceptualmente más simple. Consiste en extraer todos los individuos al azar de una lista (marco de la encuesta). En la práctica, a menos que se trate de poblaciones pequeñas o de estructura muy simple, es difícil de llevarlo a cabo de forma eficaz.

2. Muestreo estratificado

Consiste en la división previa de la población de estudio en grupos o clases que se suponen homogéneos respecto a característica a estudiar. A cada uno de estos estratos se le asignaría una cuota que determinaría el número de miembros del mismo que compondrán la muestra. Dentro de cada estrato se suele usar la técnica de muestreo sistemático, que es una de las más usadas en la práctica. Según la cantidad de elementos de la muestra que se han de elegir de cada uno de los estratos, existen dos técnicas de muestreo estratificado:

2.1 Asignación proporcional: el tamaño de cada estrato en la muestra es proporcional a su tamaño en la población.

$$n_i = n \frac{N_i}{N}$$

Donde: N_i : Tamaño de Población del estrato 'i'.
 n_i : Tamaño de muestra del estrato 'i'.

2.2 Asignación óptima: la muestra recogerá más individuos de aquellos estratos que tengan más variabilidad. Para ello es necesario un conocimiento previo de la población.

3. Muestreo sistemático

Se utiliza cuando el universo o población es de gran tamaño o ha de extenderse en el tiempo. En este caso se elige el primer individuo al azar y el resto viene condicionado por aquél. Luego hay que calcular una constante, que se denomina coeficiente de elevación $K = N/n$, donde N es el tamaño del universo y n el tamaño de la muestra.

Para determinar la primera unidad muestral se debe producir la primera extracción, para ello hay que elegir al azar un número entre 1 y K ; de ahí en adelante tomar uno de cada K a intervalos regulares. Ocasionalmente, es conveniente tener en cuenta la periodicidad del fenómeno.

Esto quiere decir que si tenemos un determinado número de personas que es la población y queremos escoger de esa población un número más pequeño, el cual es la muestra, dividimos el número de la población por el número de la muestra que queremos tomar y el resultado de esta operación será el intervalo, entonces escogemos un número al azar desde uno hasta el número del intervalo, y a partir de este número escogemos los demás siguiendo el orden del intervalo.

4. Muestreo por conglomerados

Se divide la población en varios grupos de características parecidas entre ellos y luego se analizan completamente algunos de los grupos, descartando los demás. Dentro de cada conglomerado existe una variación importante, pero los distintos conglomerados son parecidos. Requiere una muestra más grande, pero suele simplificar la recogida de muestras. Frecuentemente los conglomerados se aplican a zonas geográficas.

Técnica similar al muestreo por estadios múltiples. Se utiliza cuando la población se encuentra dividida, de manera natural, en grupos que se supone contienen toda la variabilidad de la población, es decir, la representan fielmente respecto a la característica a elegir. Pueden seleccionarse sólo algunos de estos grupos o conglomerados para la realización del estudio. Cuando, dentro de cada conglomerado, se extraen los individuos que formarán parte de la muestra por muestreo aleatorio simple, el muestreo se llama bietápico.

Las ideas de estratificación y conglomerados son opuestas. El primer método funciona mejor cuanto más homogénea es la población respecto del estrato, aunque más diferentes son estos entre sí. En el segundo, ocurre lo contrario. Los conglomerados deben presentar toda la variabilidad, aunque deben ser muy parecidos entre sí. Recolección y organización de datos: una vez identificada la población se procede a recoger los datos; en muchas ocasiones la población es muy grande y no sería posible realizar la investigación totalmente con el fin de obtener todos los datos asignados a cada uno.

5. Muestreo mixto

Cuando la población es compleja, cualquiera de los métodos descritos pueden ser difíciles de aplicar; en estos casos se aplica un muestreo mixto que combina dos o más de los anteriores sobre distintas unidades de la encuesta.

6. Muestreo por estadios múltiples

Esta técnica es la única opción cuando no se dispone de lista completa de la población de referencia o bien cuando por medio de la técnica de muestreo simple o estratificado se obtiene una muestra con unidades distribuidas de tal forma que resultan de difícil acceso.

En el muestreo a estadios múltiples se subdivide la población en varios niveles ordenados que se extraen sucesivamente por medio de un procedimiento de embudo. El muestreo se desarrolla en varias fases o extracciones sucesivas para cada nivel.

Por ejemplo, si tenemos que construir una muestra de profesores de primaria en un país determinado, éstos pueden subdividirse en unidades primarias representadas por circunscripciones didácticas y unidades secundarias que serían los propios profesores. En primer lugar extraemos una muestra de las unidades primarias (para lo cual debemos tener la lista completa de estas unidades) y en segundo lugar extraemos aleatoriamente una muestra de unidades secundarias de cada una de las primarias seleccionadas en la primera extracción.

7. Muestreo por cuotas

Es la técnica más difundida sobre todo en estudios de mercado y sondeos de opinión. En primer lugar es necesario dividir la población de referencia en varios estratos definidos por algunas variables de distribución conocida (como el género o la edad). Posteriormente, se calcula el peso proporcional de cada estrato, es decir, la parte proporcional de población que representan. Finalmente se multiplica cada peso por el tamaño de n de la muestra para determinar la cuota precisa en cada estrato. Se diferencia del muestreo estratificado en que una vez determinada la cuota, el investigador es libre de elegir a los sujetos de la muestra dentro de cada estrato.

8. Muestreo de "bola de nieve"

Indicado para estudios de poblaciones clandestinas, minoritarias o muy dispersas pero en contacto entre sí. Consiste en identificar sujetos que se incluirán en la muestra a partir de los propios entrevistados. Partiendo de una pequeña cantidad de individuos que cumplen los requisitos necesarios estos sirven como localizadores de otros con características análogas.

- Es diferente hallar el tamaño de muestra con elegir las muestras.
- Existen diferentes técnicas de muestreo, por tanto el tamaño de muestra se puede hallar de diferentes maneras teniendo presente el tipo de muestreo.

La muestra se puede elegir de diferentes maneras teniendo presente el tipo de muestreo. El muestreo aleatorio puede realizarse de diversas formas, las más frecuentes son el muestreo simple, el sistemático, el estratificado y el muestreo por conglomerados; de la combinación de las mencionadas anteriormente nacen un sin número de muestreos.

EJERCICIOS DE TIPOS DE MUESTREO (ELEGIR LA MUESTRA)

EJERCICIO 1

Problema: Se pretende determinar la prevalencia de una enfermedad en una población de 250 ovejas; se supone una prevalencia del 30% y se desea una precisión del 10% para un nivel de confianza del 95%.

El tamaño de muestra es 61 ovejas: Población(N)=250 ovejas, P=30%, E=10%, $1-\alpha=95%$, $Z_{\alpha/2}=1.96$
 $n_0 = (1.96)^2(0.3)(0.7)/(0.10)^2 = 80.67$. $n = (80.67)/(1+80.67/250) = 60.99$.

Calculado el tamaño de muestra, esta se puede elegir utilizando los siguientes tipos de muestreo:

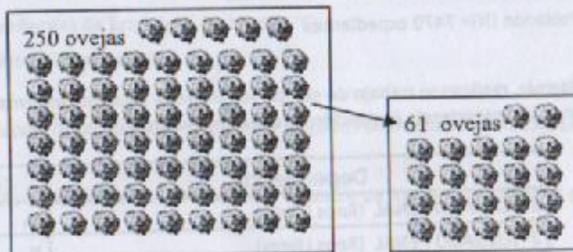
1. Muestreo Aleatorio Simple

Tamaño de muestra (n) = 61 ovejas

Población (N)=250 ovejas

La elección de la muestra se realiza de la siguiente manera:

A cada oveja se le asigna un número (del 1 al 250), luego se seleccionan al azar 61 ovejas entre la oveja número 1 y la oveja 250 utilizando una tabla de números aleatorios o una computadora.



Representación gráfica del muestreo aleatorio simple: de las 250 ovejas se seleccionan al azar (aleatorio) las 61 ovejas para realizar el estudio.

2. Muestreo Aleatorio Sistemático

Tamaño de muestra (n) = 61 ovejas

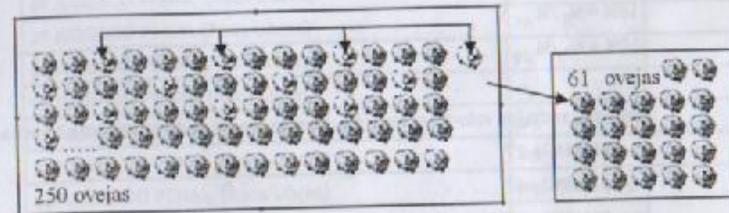
Población (N)=250 ovejas

Coefficiente de elevación $K = N/n = 250/61 = 4.09$

A cada oveja se le asigna un número (del 1 al 250).

El valor de $K=4$.

Entonces elegimos en forma aleatoria un número entre 1 y 4; supongamos que resultó el número 3 (1,2,3,4) de modo que tomaremos la oveja número 3, y a continuación cada cuarto oveja de la siguiente manera: 1,2,3,4,5,6,7,8,9,10,11,12,13,14,15... y así sucesivamente hasta completa la muestra de 61 ovejas (oveja número ovejas).



Representación gráfica del muestreo sistemático: de las 250 ovejas se seleccionan sistemáticamente las 61 ovejas (a las ovejas seleccionadas se les ha cambiado de color).

3. Muestreo Aleatorio Estratificado

La probabilidad de que una oveja esté enferma está directamente relacionada con la edad. En el ejemplo anterior, se tiene que el 44% de las ovejas son de menos de 2 años, el 28% de las ovejas son de 3 a 4 años, el 18% de las ovejas son de 5 a 6 años y el 10% son ovejas de más de seis años:

Solución:

Población (N)=250 ovejas.

Tamaño de muestra (n) = 61 ovejas (Muestra es igual a Tamaño de Muestra).

Existen 4 estratos y son los siguientes:

E_1 : Ovejas de menos de 2 años, con Población = N_1 y Muestra = n_1

E_2 : Ovejas de 3 a 4 años, con Población = N_2 y Muestra = n_2

E_3 : Ovejas de 5 a 6 años, con Población = N_3 y Muestra = n_3

E_4 : Ovejas de más de seis años, con Población = N_4 y Muestra = n_4

Aplicando asignación proporcional: El tamaño de la muestra en cada estrato es proporcional a su tamaño en la población y la fórmula es la siguiente:

$$n_i = n \left[\frac{N_i}{N} \right] \dots (1)$$

Donde:

N_i : Población del estrato i ,

n_i : Tamaño de muestra del estrato i

De la información se tiene que:

$$44\% = N_1 / N ;$$

$$28\% = N_2 / N ;$$

$$18\% = N_3 / N ;$$

$$10\% = N_4 / N$$

Remplazando los valores al tamaño de muestra en cada estrato según ecuación (1)

$$n_1 = 61(44\%) = 27$$

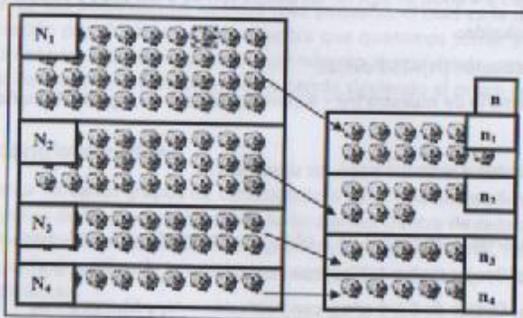
$$n_2 = 61(28\%) = 17$$

$$n_3 = 61(18\%) = 11$$

$$n_4 = 61(10\%) = 6$$

Esto es, 27 ovejas del estrato 1, 17 ovejas de estrato 2, 11 ovejas de estrato 3 y 6 ovejas del estrato 4.

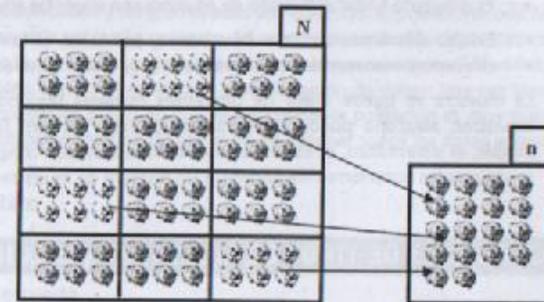
Este método evita que se tomen más individuos de un grupo que de los demás y esto pueda condicionar el resultado.



Representación gráfica del muestreo estratificado con asignación proporcional al tamaño, es decir, a población grande le corresponde tamaño de MUESTRA y a población pequeña le corresponde tamaño de muestra pequeña.

Muestreo aleatorio por conglomerados

Se divide la población en varios grupos de características parecidas entre ellos y luego se analizan completamente algunos de los grupos, descartando los demás. Dentro de cada conglomerado existe una variación importante, pero los distintos conglomerados son parecidos. Requiere una muestra más grande, pero suele simplificar la recogida de muestras.



Representación gráfica del muestreo por conglomerados.

Muestreo mixto

Cuando se desea realizar un estudio más preciso, se pueden combinar los muestreos anteriormente descritos. También puede suceder que la población bajo estudio sea compleja, entonces se puede aplicar un muestreo mixto que combine dos o más muestreos anteriores.

EJERCICIO 2

Problema: se pretende determinar el tiempo que demora un proceso penal en la Corte Superior de Lima Norte, sabiendo que existen 7470 expedientes con proceso penal en la sede de la Corte Superior.

Por el tema anterior se sabe que: el tamaño de muestras es: $n=365$ ($N=7470$; $\alpha=5\%$; $E=5\%$; $p=0.5$; $q=0.5$, $n_0=384$)

Solución:

Tamaño de muestra (n) = 365 expedientes

Población (N) = 7470 expedientes

Además, realizando trabajo de campo se obtuvo la siguiente información de la cantidad de expedientes en cada dependencia (Población de cada dependencia).

Dependencia	Estratos	Población
1º JUZGADO PENAL (Reos Libres)	N_1	914
2º JUZGADO PENAL (Reos Libres)	N_2	912
3º JUZGADO PENAL (Reos Libres)	N_3	696
4º JUZGADO PENAL (Reos Libres)	N_4	789
5º JUZGADO PENAL (Reos Libres)	N_5	814
6º JUZGADO PENAL (Reos Libres)	N_6	1395
7º JUZGADO PENAL (Reos Libres)	N_7	380
8º JUZGADO PENAL (Reos Libres)	N_8	127
9º JUZGADO PENAL (Reos Libres)	N_9	601

10º JUZGADO PENAL (Reos Libres)	N_{10}	275
11º JUZGADO PENAL (Reos Libres)	N_{11}	71
12º JUZGADO PENAL (Reos Libres)	N_{12}	195
13º JUZGADO PENAL (Reos Libres)	N_{13}	301
Total	N	7470

La selección de la muestra se realizará utilizando el muestreo estratificado porque se requiere una muestra representativa para poder determinar el objetivo de estudio (tiempo de duración de un proceso).

- No se utiliza el muestreo aleatorio simple porque no es representativa, ya que en la muestra podrían seleccionarse más expedientes de determinado juzgado.
- No se utiliza el muestreo por conglomerado porque no es representativa, ya que al seleccionarse solo algunos juzgados perderíamos información relevante de los otros no seleccionados.

Muestreo estratificado

Existen 13 estratos desde el primer Juzgado al 13 Juzgado.

La elección de la muestra en cada estrato se realizará aplicando un reparto proporcional, porque se puede observar algunos juzgados con elevado número de expedientes y otros con cantidades pequeñas de expedientes. Para la proporcionalidad se aplica la fórmula siguiente:

$$n_i = n \left[\frac{N_i}{N} \right]$$

Donde

N_i : Cantidad de expedientes de juzgado 'i' (Población 'i').

n_i : Tamaño de muestra del juzgado 'i'

Aplicando la fórmula para hallar la muestra del primer Juzgado Penal se tiene: $n_1 = 365(914/7470) = 44.65 = 45$

Aplicando la fórmula sucesivamente para hallar la muestra de los demás juzgados se tiene el cuadro siguiente:

Dependencia	Estrato	Muestra
1º JUZGADO PENAL (Reos Libres)	n_1	45
2º JUZGADO PENAL (Reos Libres)	n_2	45
3º JUZGADO PENAL (Reos Libres)	n_3	34
4º JUZGADO PENAL (Reos Libres)	n_4	39
5º JUZGADO PENAL (Reos Libres)	n_5	40
6º JUZGADO PENAL (Reos Libres)	n_6	68
7º JUZGADO PENAL (Reos Libres)	n_7	19
8º JUZGADO PENAL (Reos Libres)	n_8	6
9º JUZGADO PENAL (Reos Libres)	n_9	29
10º JUZGADO PENAL (Reos Libres)	n_{10}	13
11º JUZGADO PENAL (Reos Libres)	n_{11}	3
12º JUZGADO PENAL (Reos Libres)	n_{12}	9
13º JUZGADO PENAL (Reos Libres)	n_{13}	15
Total	N	365

Encontrado el tamaño de muestra para cada estrato, el siguiente paso es elegir las muestras que participarán en el estudio.

Casi siempre se eligen las muestras en forma aleatoria, por ejemplo de los 914 expedientes del primer Juzgado Penal se eligen al azar 45 expedientes. Para hacer el ejercicio más interesante utilizaremos la elección sistemática.

Elección de la muestra en forma Sistemática

La selección de la muestra en forma sistemática se realizará en cada juzgado (estrato), por ejemplo:

El 12º Juzgado Penal tiene 195 expedientes, a cada expediente se le asigna un número: 1,2,3...195. Luego se calcula el coeficiente $K = N/n = 195/10 = 19.5$ ($K=20$), entonces elegimos en forma aleatoria un número del 1 al 20, supongamos que resultó 9 (primera muestra elegida; 1,2,3,...8,9,10...20).

Ahora para las demás muestras, a la primera muestra seleccionada (9) se le suma $K=20$, y así hasta completar el tamaño de muestra de 10 expedientes. Del modo siguiente: 1,2,...9,...,27,28,29,30,...,48, 49,50,...,188,189,190,...195.

La muestra son los expedientes representados por los siguientes números 9, 29, 49, 69, 89, 109, 129, 149, 169, 189.

De esta forma se hace la elección de la muestra para los demás juzgados.

EJERCICIOS PROPUESTOS

1).- Se quiere estimar la incidencia de la hipertensión arterial en el embarazo. Cuántas embarazadas tenemos que observar para una confianza del 95%; estimar dicha incidencia con un error del 2% en los siguientes casos:

- Sabiendo que un sondeo previo se ha observado un 9% de hipertensas.
- Sin ninguna información previa.

2).- Un productor de semillas desea saber, con un error de estimación del 5%, el porcentaje de semillas que germinan en la granja de su competidor. ¿Qué tamaño de muestra debe tomarse para obtener un nivel de confianza del 95%? Con la muestra hallada realice una selección sistemática si se sabe que M1, M2, M3 y M4, son los estratos de población (M1 es el doble de M4, M3 es el triple de M4, mientras que M1 es la mitad de M2). Suponga usted el número ideal de los conglomerados en cada estrato.

PILOTO

En lo posible debe realizar un piloto en campo de estudio, el mismo que consiste en realizar un muestreo que le permita familiarizarse con el lugar, las condiciones, el equipo, las personas, etc. La función del muestreo no es tomar datos útiles para la investigación, sino validar las condiciones de muestreo frente a los objetivos propuestos: ¿qué variables incidentes hay? ¿qué población se debe evaluar? ¿cuántas muestras deberían tomarse? ¿cuántos sujetos se deben entrevistar? ¿a qué condición social o cultural pertenecen? etc.

Estadística Descriptiva

La Estadística Descriptiva ayuda a comprender la estructura de los datos, pues detectan tanto un patrón de comportamiento general como apartamientos del mismo. Una forma de realizar esto es mediante gráficos de sencilla realización e interpretación. Otra forma de describir los datos es resumiéndolos en uno, dos o más números que caractericen al conjunto de datos con fidelidad. Explorarlos permitirá detectar datos erróneos o inesperados y nos ayudará a decidir qué métodos estadísticos pueden ser empleados en etapas posteriores del análisis para obtener conclusiones válidas.

DISTRIBUCIÓN DE FRECUENCIA

Distribución de frecuencia es la representación estructurada, en forma de tabla, de toda la información que se ha recogido sobre la variable que se estudia.

Variable (Valor)	Frecuencias absolutas		Frecuencias relativas	
	Simple (fi)	Acumulada (Fi)	Simple(hi)	Acumulada (Hi)
X1	f1	F1=f1	h1 = f1/n	H1=h1
X2	f2	F2=f1+ f2	h2 = f2/n	H2=h1+ h2
.
X-1	fk-1	Fk-1=f1+ f2 +..+ fk-1	hk-1=fk-1/n	Hk-1=h1+ h2 +..+ hk-1
Xk	fk	Fk = $\sum fi=f1+f2+..+fk$	hk =fk/n	Hk= $\sum hi=h1+h2+..+hk$

Donde:

- X_i : los distintos valores que puede tomar la variable. (i=1,2,3...k).
 f_i : el número de veces que se repite cada valor.
 h_i : el porcentaje que la repetición de cada valor supone sobre el total.
 n : el número de observaciones realizadas.

Propiedades

$$i) n = \sum_{i=1}^k f_i = f_1 + f_2 + \dots + f_k$$

$$ii) n = F_k$$

$$iii) h_i = \frac{f_i}{n}$$

$$iv) H_i = F_i / n$$

$$v) H_k = \sum h_i = h_1 + h_2 + \dots + h_k = 1 = 100\%$$

