



Recopilación y Organización de los datos

Contenidos

1. Estadística: definición. Estadística descriptiva e inferencial. Población y muestra. Importancia y aplicaciones.
2. Variables y escalas de medición. Definición operativa
3. Recopilación de datos estadísticos. Teoría elemental del muestreo. Muestreos probabilísticos y no probabilísticos: características y procedimientos técnicos de selección.

Introducción

La estadística es el lenguaje universal de la ciencia. Como usuarios potenciales de la estadística, es necesario dominar la "ciencia" y el "arte" de utilizar correctamente su metodología. El empleo cuidadoso de los métodos estadísticos permite obtener información precisa de los datos. Estos métodos incluyen: 1) definir cuidadosamente la situación, 2) recolectar datos, 3) resumir con precisión los datos y 4) obtener y comunicar las conclusiones importantes.

La estadística implica información, números y gráficas para resumir esta información, y su interpretación. El término **estadística** posee varios significados para persona de diversos entornos e intereses. Para algunos, se trata de un medio para recolectar y representar grandes cantidades de información. Para otros, se trata de un medio para "tomar decisiones de frente a la incertidumbre". En la perspectiva idónea, cada uno de estos puntos de vista es correcto. El terreno de la estadística puede dividirse a grandes rasgos en dos áreas: estadística descriptiva y estadística inferencial. La estadística descriptiva es en lo que piensa la mayoría de las personas al escuchar la palabra *estadística* ya que incluye la recolección, presentación y descripción de datos muestrales. El término estadística inferencial se refiere a la técnica de interpretación de los valores resultantes de las técnicas descriptivas y a la toma de decisiones y obtención de conclusiones sobre la población.

La estadística es más que sólo números: son los datos, lo que se hace con los datos, lo que se aprende de los datos y las conclusiones resultantes.



1. *Estadística:*

Cuerpo de conceptos y métodos utilizados para recolectar e interpretar datos referentes a un área particular de investigación y extraer conclusiones en situaciones en que la incertidumbre y la variación están presentes. También se puede entender como la ciencia que transforma datos en información

Clasificación:

- **Estadística descriptiva:** se encarga de recoger y resumir las características de una población o muestra, deduciendo de esta descripción conclusiones sobre su estructura, además de las relaciones existentes entre otras poblaciones o muestras con las que se compara.

- **Estadística inferencial:** comprende aquellas técnicas por medio de las cuales, basándose en los resultados del análisis de una muestra de la población, induce (infiere) o estima las leyes generales de comportamiento de la población.

Componentes de una investigación estadística:

- **Población:** Conjunto de todos los elementos que cumplen ciertas propiedades y entre los cuales se desea estudiar un determinado fenómeno
Las poblaciones pueden ser:
 - **finita:** es aquella población que se puede listar
 - **infinita :** es aquella población que en la práctica no puede ser listada
- **Individuo, unidad de observación o unidad experimental:** Es cada uno de los elementos de la población o de la muestra. No tienen que ser necesariamente personas, pueden ser hospitales, empresas, viviendas, etc.
- **Censo** es un conjunto de datos de cada unidad de observación de una población.
- **Parámetro:** característica de una población resumida para su estudio. Se considera como un valor verdadero de la característica estudiada. (Esta característica se puede determinar cuando se hace un censo).)
- **Muestra:** Es un subconjunto de la población que es estudiado y a partir de la cual se sacan conclusiones acerca de las características de la población. Es deseable que la muestra sea *representativa*, es decir, que su composición sea lo suficientemente parecida a la población de la que se extrae, de manera que las observaciones de cierta característica se puedan generalizar.
- **Estadística** es una medición numérica que describe alguna característica de una muestra.



Ejemplos:

- a) Una encuesta realizada a los 19 intendentes de Mendoza, indicó que 7 de ellos (el 37%) consideraron que el personal afectado a la recolección de residuos era insuficiente, la cifra de 37 % es un parámetro porque se basa en toda la población de intendentes de Mendoza del corriente año.
- b) En una encuesta realizada a 1.215 personas seleccionadas al azar, 269 (o el 22 %) están desempleadas o no tienen un trabajo fijo. Dado que la cifra de 22% se basa en una muestra (no en toda la población), es una estadística.

2. Variable: es una característica de los individuos o unidades experimentales que son objeto del estudio estadístico. Así, por ejemplo, si consideramos a una persona, podemos distinguir en ella las siguientes **variables:** Sexo, Edad, Nivel de estudios, Profesión, Peso, Altura, Ingresos, etc.

Las variables se clasifican en:

➤ Variables **Cualitativas, Categóricas o Atributos.**

Los atributos, cualidades o categorías son aquellas variables que para su definición precisan de palabras, es decir, no le podemos asignar un número. Por ejemplo: Sexo, Profesión, Estado Civil, Nacionalidad, etc.

A su vez las podemos clasificar en:

- **Ordinales:** Aquellas que sugieren una ordenación, por ejemplo graduación militar, nivel de estudios, etc.
- **Nominales:** Aquellas que sólo admiten una mera ordenación alfabética, pero no establece orden por su naturaleza, por ejemplo el color de pelo, sexo, estado civil, etc.

➤ Variables **Cuantitativas.**

Las variables cuantitativas son las que se describen por medio de números, como por ejemplo el Peso, Altura, Edad, Número de alumnos por curso , etc

A su vez, este tipo de variables se puede dividir en dos subclases:

- **Cuantitativas discretas.** Aquellas a las que se les puede asociar un número entero, es decir, aquellas que por su naturaleza no admiten un fraccionamiento de la unidad, por ejemplo número de hermanos, páginas de un libro, etc.



- **Cuantitativas continuas:** Aquellas que no se pueden expresar mediante un número entero, es decir, aquellas que por su naturaleza admiten que entre dos valores cualesquiera la variable pueda tomar cualquier valor intermedio, por ejemplo peso, tiempo. etc.

En muchos casos el tratamiento estadístico hace que a variables discretas las trabajemos como si fuesen continuas y viceversa.

- **Dato:** es el valor correspondiente a cierta variable medido en una unidad de observación.

Ejemplo:

<i>Población:</i>	VIVIENDAS DE LA PROVINCIA DE MENDOZA	
<i>Unidad de observación</i>	VIVIENDAS	
<i>Variables</i>	Tipo	Dato
Material de construcción	Cualitativa nominal	Ladrillo
Metros cuadrados cubiertos	Cuantitativa continua	150 m ²
Servicios (gas natural, cloacas, etc).	Cualitativa nominal	gas natural
Habitantes por vivienda	Cuantitativa discreta	5 personas
Antigüedad (en años)	Cuantitativa discreta	10 años
Avalúo	Cuantitativa continua	\$ 75.000
Estilo arquitectónico	Cualitativa nominal	clásico
Calefacción	Cualitativa nominal	no tiene
Impuestos anuales (en pesos)	Cuantitativa continua	\$ 3.540
Cantidad de habitaciones	Cuantitativa discreta	3 habitaciones
Cantidad de baños	Cuantitativa discreta	2 baños

Escalas de Medida

Para llevar a cabo una medición necesitamos:

a) *Un instrumento de medición:*

Un instrumento de medición es un artificio usado para medir . Puede ser una balanza para medir el peso de un individuo, un test para medir personalidad, inteligencia o aptitud, un cuestionario para medir conocimientos, etc. Para que el instrumento sea eficaz debe reunir 2 requisitos: *confiabilidad* y *validez*.

La *confiabilidad* significa estabilidad o constancia de los resultados obtenidos al aplicar el instrumento, o sea que al aplicarlo más de una vez dé resultados muy parecidos (que presente variaciones poco significativas).

La *validez* se refiere a que el instrumento mida lo que efectivamente pretenda medir. Esto parece una redundancia, pero no es tal si nos damos cuenta de los casos en que se quiere medir una



característica y se mide otra. Por ejemplo queremos medir aptitud para un empleo y lo que realmente medimos es conocimiento.

b) Un sistema de medición (o escala)

Vamos a considerar la definición de medición usada por Stevens que dice “*medición es la asignación de números a objetos o hechos de acuerdo a ciertas reglas*” . De acuerdo con esto podemos dividir los niveles de medición en cuatro escalas fundamentales: **nominal, ordinal, de intervalos y de razón**. Cada uno de estos niveles requiere la utilización de un determinado conjunto de instrumentos estadísticos.

Cualquiera de estos procedimientos de medición debe ser **exhaustivo**, es decir que debe tener la cantidad suficiente de categorías en las que puedan clasificarse cada uno de los casos considerados. Estas clases o categorías deben ser **mutuamente excluyentes**, es decir que debe ser posible clasificar cada caso individual tan solo en una categoría.

➤ **Escalas nominales o de clasificación:** Se realiza cuando la propiedad estudiada solo puede agruparse en categorías lógicamente exhaustivas y mutuamente excluyentes, de modo que puedan establecerse equivalencias o diferencias. Con esta escala de medición se realiza la operación más simple y básica de toda la ciencia que es la *clasificación*.

A cada categoría se le asignan nombres o números. Si se asignan números, sólo se pueden interpretar como que son categorías diferentes entre sí, sin que se pueda afirmar que uno es superior a otro y, por lo tanto, sin que se puedan ordenar.

Ejemplos:

- a) Partidos políticos
- b) País de origen
- c) N° de las líneas de ómnibus

➤ **Escalas ordinales o por orden jerárquico:** Además de incluir las propiedades de las medidas nominales, se incluye la propiedad de que las categorías pueden ser ordenadas en el sentido de "menor que" o "mayor que". Este tipo de medición no ofrece ningún tipo de información acerca de la magnitud de las diferencias entre las categorías.

Ejemplos:

- a) Nivel salarial (alto, medio, bajo)
- b) Rango docente (profesor titular, asociado, adjunto, jefe de trabajos prácticos, ayudante)
- c) Valores de los naipes en una partida de truco



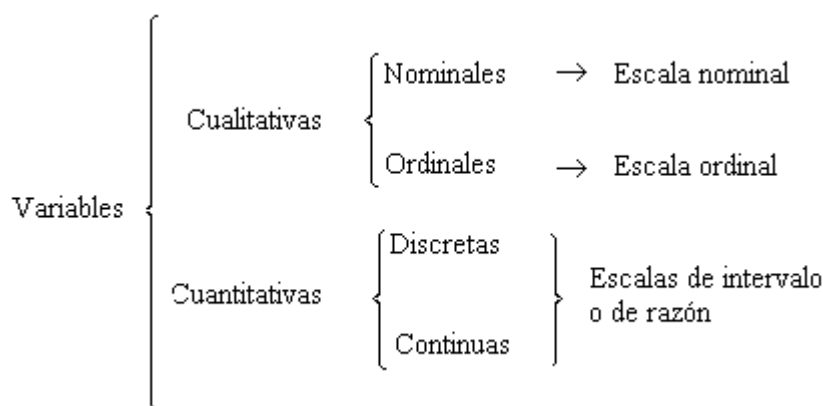
➤ **Escalas de intervalo o de unidades iguales:** Incluye las propiedades lógicas de las escalas nominales y ordinales y además sus categorías se definen en términos de una unidad de medición estándar (años de antigüedad, pesos de ingresos, grados de temperatura, etc). Distancias numéricamente iguales representan distancias iguales empíricas en las variables que miden. En estas escalas el origen y la unidad de medida son arbitrarios

➤ **Escalas de razones o cocientes:** Cuando una medición tiene todas las características de una medida de intervalo y, además, se le puede asignar un punto de origen de verdadero valor cero (escala con cero absoluto), se tiene una medida de razón. Esto quiere decir que el 0 de la escala debe coincidir con la “nada” de la variable

Ejemplos:

- a) Temperatura máxima de una ciudad en grados centígrados (Intervalo)
- b) Cantidad de alumnos por curso (intervalo y razón)
- c) Rendimiento académico medido en una escala de 0 a 7 (Intervalo)
- d) Tiempo de reacción ante un estímulo (intervalo y razón)

Resumen:



3. Recopilación de datos estadísticos

La teoría del muestreo es el estudio de las relaciones existentes entre una población y las muestras extraídas de ellas.

Las muestras tienen que ser representativas porque los datos que obtengamos de ellas, los generalizamos a la población. La falta de representatividad anula por completo la generalización.

El **diseño de la muestra** comprende el *establecimiento de un plan de muestreo* (o sea los pasos a seguir para obtener la muestra) y la *elección del estimador* (que depende del objetivo del muestreo y da la precisión de la estimación)

Pasos a seguir para diseñar una muestra:



1- Definir la población:

La población debe estar perfectamente definida en cuanto a *las unidades que la componen* (unidad de muestreo: personas, empresas, cultivos, etc.) y a su posible *distribución*. También se debe establecer la *extensión*, o sea el lugar físico donde se va a llevar a cabo el muestreo, y el *tiempo* asignada a la recolección de la información.

2- Identificar el marco de muestreo

El marco de muestreo es un listado de los elementos que son de interés en nuestra investigación. Debe tenerse mucho cuidado de evitar repeticiones y de no excluir a ningún grupo de elemento que puedan aportar información.

3- Determinar el tamaño de la muestra

Tener una muestra representativa de la población no es suficiente. Necesitamos que esa muestra tenga un cierto tamaño para que quede libre de esos errores que pueden ocurrir por azar y anularían la representatividad de la muestra.

La determinación del tamaño de la muestra se realiza por procedimientos estadístico, alguno de los cuales veremos más adelante.

Básicamente, el tamaño de la muestra depende de:

- a) el tamaño de la población
- b) el estadístico con el que estemos midiendo
- c) la variabilidad de la medida
- d) los márgenes de nuestras predicciones
- e) el nivel de seguridad que deseamos que tenga el procedimiento.

4- Seleccionar una técnica de muestreo

Los procedimientos que nos aseguran que los individuos seleccionados representan a su población se denominan ***técnicas de muestreo***.

Cuando cada muestra tiene una probabilidad conocida de ser elegida, los muestreos se denominan *probabilísticos*. Este tipo de muestreo permite medir la precisión de las estimaciones de las características de la población realizadas y nos permite describir matemáticamente qué tan objetivas son esas estimaciones.

Algunos de las técnicas más importantes de ***muestreo probabilístico*** son:

A) ***Muestreo aleatorio simple***



El muestreo aleatorio simple selecciona muestras mediante métodos que permiten que *cada posible muestra tenga igual oportunidad de ser elegida* y que *cada elemento de la población total tenga una oportunidad igual de ser incluido en la muestra*.

Un ejemplo sencillo sería el siguiente:

Suponga que una agencia de viajes regala un pasaje para 4 alumnos de 1º año (muestra) para que asistan a un Congreso de Administración Pública. La población son todos los alumnos de 1º año, que aparecen en el listado de la Facultad y a quienes le asignamos un número. Luego por medio de una tabla de números aleatorios (puede ser la que trae incorporada las calculadoras), o utilizando un bolillero, o cualquier otro procedimiento similar, que asegure la igualdad de oportunidades, se seleccionan los 4 ganadores.

B) Muestreo sistemático

En el muestreo sistemático, los elementos son seleccionados de la población dentro de un intervalo uniforme que se mide respecto al tiempo, al orden o al espacio. Si quisiéramos entrevistar a cada vigésimo estudiante de la Facultad, elegiríamos un punto de inicio aleatorio entre los primeros veinte nombres del listado de todos los estudiantes de la Facultad y luego seleccionaríamos cada veinteavo nombre de allí en adelante.

En el muestreo sistemático cada *elemento tiene igual oportunidad de ser seleccionado*, pero cada muestra *no* tiene una posibilidad igual de ser seleccionada. En el ejemplo que vimos, los estudiantes numerados 4, 5, 6 y 7 no hubieran tenido ninguna oportunidad de haber sido seleccionados juntos.

En este tipo de muestreo existe el problema de introducir errores ya que puede haber factores periódicos que nos afecten los resultados (al incluirlos o excluirlos siempre). La ventaja es que requiere menos tiempo y a veces menos costo que el muestreo aleatorio simple.

C) Muestreo aleatorio estratificado

Para utilizar el muestreo estratificado dividimos la población en grupos relativamente homogéneos, llamados *estratos*. Suponemos que cada grupo tiene una pequeña variación dentro de sí mismo, pero hay una amplia variación entre los grupos.

Después utilizamos uno de los siguientes procedimientos:

- a) seleccionamos aleatoriamente de cada estrato un número específico de elementos correspondientes a la fracción de ese estrato en la población
- b) seleccionamos el mismo número de elementos en todos los estratos y damos peso a los resultados de acuerdo con la posición del estrato con respecto a la población total.

Con cualquiera de los planteamientos se garantiza que todos los elementos de la población tenga oportunidad de ser seleccionados



Por ejemplo se desea seleccionar una muestra de 50 alumnos para recabar información acerca de un nuevo Plan de Estudios, y queremos que estén representados todos los cursos, de 1° a 5°. Cada curso corresponde a un estrato (los alumnos pueden estar incluidos en un solo curso, por lo que al dar la definición operativa de alumno se debe prever en qué categoría entra un alumno que recurre alguna materia).

La forma de hacer el muestreo aleatorio sería la siguiente:

Nº de alumnos de cada curso		Proporción con respecto al total	Tamaño de muestra
primero	650	$(650/2400) = 0,271$	$650 \cdot 0,271 = 14$
segundo	550	0,229	12
tercero	450	0,187	9
cuarto	400	0,167	8
quinto	350	0,146	7
Total	2 400		

Luego por muestreo aleatorio simple seleccionamos ese número de alumnos por curso

D) *Muestreo por conglomerados*

Se divide la población en grupos o *conglomerados* y luego se selecciona una muestra aleatoria de esos grupos. Suponemos que esos grupos son representativos de la población como un todo. Es decir que dentro de cada grupo hay una gran variación, pero los grupos son esencialmente similares entre sí.

Por ejemplo no asignamos un número a cada uno de los habitantes de Mendoza y luego sacamos al azar los números que han de componer la muestra, sino mostramos conglomerados de individuos, como por ejemplo, los departamentos. Si queremos tener una muestra de 500 personas seleccionamos, aleatoriamente, 10 departamentos de forma que cada uno aporte 50 personas.

Un muestreo por conglomerados puede producir una muestra a un costo considerablemente menor que con los otros muestreos aleatorios.

En general estas técnicas de muestreo se utilizan combinadas.

En el desarrollo de la estadística inferencial, vamos a suponer que el muestreo es el aleatorio simple. Conociendo los principios, su extensión a otros tipos de muestreo es simple conceptualmente, pero algo elaborado matemáticamente.

Si la elección de las unidades que integran la muestra se efectúa a través del conocimiento (de un experto, por ejemplo) o de la opinión personal del investigador acerca de quiénes integrarán la muestra, se efectúa un muestreo **no probabilístico**, donde no se conoce la posibilidad de que una muestra sea seleccionada y no se puede medir el error de muestreo. Estos muestreos se utilizan en etapas tempranas de la investigación (al plantear las hipótesis, por ejemplo).



Algunas técnicas de muestreo no probabilístico son:

a) **Muestreo accidental**

En un programa de televisión (Hora Clave) se dan números de teléfono para que la gente se pronuncie a favor de ciertos eventos: "Si está a favor de la re-reelección disque el número, si está en contra disque el número" . Al final del programa se presenta el recuento de llamadas.

La muestra de las personas que llamaron es accidental, porque es un subconjunto de las personas que están viendo televisión en ese momento y no podemos asegurar que representen a todos los argentinos (o a todos los porteños), porque están excluidas las personas que no le gustan los programas de índole política, las personas que a esa hora están trabajando, las que no reciben la señal del canal (o reciben el programa en forma diferida), las que no están de acuerdo con la conducción del programa, etc.

Otro ejemplo muy común es cuando una emisora de radio, que selecciona un determinado tipo de música porque cree que es la que tiene mejor acogida entre su público, realiza una vez a la semana una encuesta para determinar las mejores canciones. Lógicamente, sólo llaman sus oyentes, porque a los que no les gusta esa selección musical, están en otro punto de la sintonía, y, si accidentalmente están escuchando, no votarían, ya que sus temas preferidos no estarían entre los que se pueden votar. El reparo surge cuando la emisora diga cosas como "éstos son los temas preferidos por la juventud", "aquí están los temas preferido por los radioyentes"

b) **Muestreo a propósito:**

- ***Muestreo por conveniencia***

Selecciona los elementos de la población de acuerdo a la conveniencia del investigador. Por ejemplo buscar voluntarios para una prueba psicológica; entrevistar a las personas que están en las colas de las cajas de un supermercado para conocer su opinión acerca de determinado producto

- ***Muestreo por juicio***

Consiste en seleccionar elementos de la población en base a lo que cree algún experto respecto a las unidades que darán respuesta a determinado problema. Por ejemplo,; si queremos construir un test que va a medir el grado de desarrollo en la escritura en niños, necesitamos una muestra donde estén representados los distintos niveles de desarrollo. Una forma de seleccionar es pedir a maestros de distintos colegios y de distintos entornos que nos proporcionen 2 de los mejores, 2 de los peores y 2 medios. Los "expertos" (maestros) nos proporcionan la muestra

- ***Muestreo por cuotas***

Se establecen pasos para obtener una muestra que sea similar a la población especificada ejerciendo ciertos "controles" sobre algunas características



de sus elementos. Se estiman los tamaños de subconjuntos de la población en base a datos y se calculan proporcionalmente "cuotas" o número de observaciones muestrales con respecto a esos subconjuntos. Por ejemplo se seleccionan 30 personas que padecen cierta enfermedad y 30 que no la padecen ("controles").