

CAPÍTULO 1

INTRODUCCIÓN

En este capítulo, primero se introducirán algunos conceptos estadísticos básicos, luego se dará una definición y división de la estadística. Finalmente se hará una clasificación de los distintos tipos de datos que aparecen en un estudio estadístico y de que forma pueden ser recolectados.

1.1 Conceptos Estadísticos Básicos

a) Población: De una manera bastante general, una población es un conjunto de individuos u objetos que poseen la característica que se desea estudiar. En un sentido más estadístico, una población es el conjunto de mediciones de una cierta característica en todos los individuos u objetos que poseen dicha característica. Ejemplos:
Las edades de todos los estudiantes de una universidad.
Las contestaciones, SI o No, de todos los estudiantes de universidad a la pregunta: Fuma?
Los niveles de colesterol de todas las personas con problemas cardiacos.

b) Muestra: Es una parte de la población, y es el conjunto de mediciones que han sido realmente recolectados. La extracción de la muestra es un paso bien importante porque es a partir de ella que se sacan conclusiones acerca de la población. El tamaño de una muestra depende del esquema que se usa para su selección. Si el diseño es sencillo la muestra tiene que ser relativamente grande, alrededor de un 10% del tamaño de la población. Si el diseño es bien sofisticado el tamaño de la muestra puede ser bien pequeño y aún dar conclusiones confiables. Por ejemplo, en las encuestas políticas de un país donde hay millones de electores una muestra de tamaño entre 1,000 y 2,000 puede dar excelentes conclusiones.

c) Muestra Aleatoria: Es una muestra bien representativa de la población. Se considera que cada elemento de la población ha tenido la misma oportunidad de formar parte de la muestra. Las conclusiones basadas en una muestra aleatoria son confiables. Ejemplo: Si se hace una encuesta de carácter político a nivel nacional, NO sería una muestra aleatoria si solo se entrevistan personas en los pueblos donde históricamente cierto partido es siempre ganador.

d) Variable: Es la característica que se desea estudiar. Ejemplos: edad, peso, opinión, raza, tipo de sangre.

e) Dato: Es un valor particular de la variable. Por ejemplo si la variable es el número de hermanos entonces los datos pueden ser: 0, 1, 2, 3, 4, 5, 6, 7, ...

f) Parámetro: Es un valor que caracteriza a una población. El valor del parámetro es constante y por lo general es desconocido. Ejemplos:

La edad promedio de todos los estudiantes de una universidad.

El nivel de colesterol promedio de todos los pacientes con problemas cardiacos.

El porcentaje de todos los electores de un país que favorecen a cierto candidato.

g) Estadístico: Es un valor que se calcula en base a los datos que se toman en la muestra y el cual es usado para estimar el valor del parámetro. El valor del estadístico es conocido y varía con la muestra tomada. Ejemplos:

La edad promedio de una muestra de 30 estudiantes de una universidad.

El nivel de colesterol promedio de los pacientes con problemas cardiacos que están hospitalizados.

El porcentaje de votantes que favorecen a cierto candidato en una muestra de tamaño 200.

h) Censo: Es un listado de una o más características de todos los elementos de una población. Los censos poblacionales se hacen cada 10 años a nivel mundial, el próximo será en el año 2010.

Hay también otros tipos de censos como: censos de vivienda donde se trata de determinar la cantidad y el tipo de vivienda que hay en un país. Censos agrícolas donde se trata de determinar las cantidades totales de la producción agrícola del país, incluyendo cultivos y animales.

i) Encuesta: Es un listado de una o más características de todos los elementos de una muestra.

Un ejemplo son las encuestas de investigación de mercados donde se desea establecer las preferencias del consumidor con respecto a cierto tipo de producto comercial.

1.2 Definición de la Estadística.

La Estadística es la ciencia donde se aprende acerca de la población a partir de la información recolectada de una muestra extraída de dicha población. La Estadística comprende los métodos usados para recolectar la muestra, la organización y presentación de los datos recolectados y la extracción de conclusiones mediante la aplicación de técnicas adecuadas a los datos de la muestra.

Las diversas técnicas usadas para extraer la muestra constituyen el Muestreo Estadístico. Por otro lado, al extraer conclusiones acerca de la población es inevitable que se cometa un error, el cual es medido usando probabilidades. La figura 1.1 muestra un esquema del análisis estadístico.

1.3 División de la Estadística

Estadística Descriptiva: Conjunto de técnicas y métodos que son usados para recolectar, organizar, y presentar en forma de tablas y gráficas información numérica. También se incluyen aquí el cálculo de medidas estadísticas de centralidad y de variabilidad.

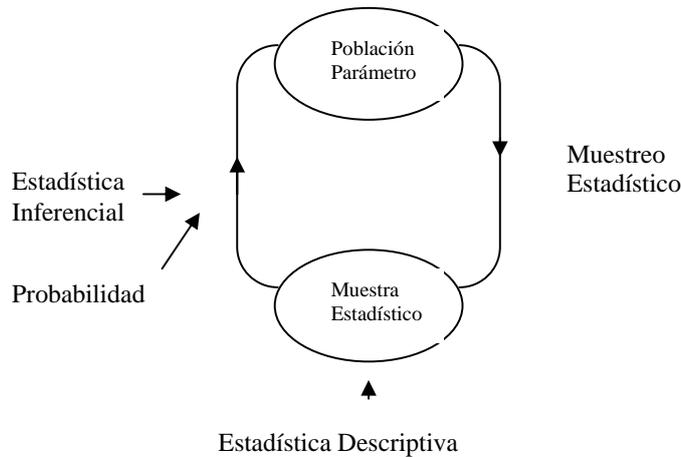


Figura 1.1. Gráfica del Análisis Estadístico

Estadística Inferencial: Conjunto de técnicas y métodos que son usados para sacar conclusiones generales acerca de una población usando datos de una muestra tomada de ella.

1.4 Tipos de Datos

Es importante clasificar los datos estadísticos porque hay algunas gráficas y métodos estadísticos que son solamente adecuados para cierto tipo de datos. La siguiente es la clasificación de datos más frecuentemente usada.

A. Datos Cuantitativos. Son aquellos que resultan de hacer mediciones o conteos. Se clasifican a su vez en dos subtipos:

A1. Datos Discretos. Son los que resultan de hacer conteos y por lo general son números enteros. Ejemplos:
 Número de llamadas que llegan a un cuadro telefónico en un período de 10 minutos.
 Número de personas que habitan una casa.
 Número de accidentes que ocurren semanalmente en una fábrica.

A2. Datos Continuos. Son los que resultan de hacer mediciones y pueden asumir cualquier valor de la recta real.
 Ejemplos: Tiempo de espera para ser atendido en un banco. Temperatura, peso, altura, salario, etc.

B. Datos Cualitativos o Categóricos. Son aquellos que expresan atributos o categorías. Para facilitar el análisis estadístico de este tipo de datos frecuentemente se codifican a números, esta codificación da lugar a dos subtipos de datos categóricos:

B1. Datos Nominales. Son aquellos datos categóricos que pueden ser codificados numéricamente pero donde hay una relación arbitraria entre los números asignados y el valor de la variable.

Ejemplo: Sexo, se puede asignar indistintamente la codificación 0 ó 1 a femenino y masculino. O inclusive usar otros números.

Otros ejemplos son: Raza de la persona, Programa de estudio, Profesión, Tipo de Sangre.

B2. Datos Ordinales. Son aquellos que al ser codificados numéricamente deben guardar una correspondencia entre los números asignados y el verdadero valor de la variable.

Ejemplo. La respuesta a una pregunta de un cuestionario de evaluación puede ser: Deficiente, regular, promedio, bueno y excelente. Al momento de codificar estos datos debería usarse 0, 1, 2, 3, 4, o una secuencia ordenada de números similar a ella, pero no se puede asignar algo como 3, 5, 4, 1, 2 porque se perdería el significado de los verdaderos datos.

Otros ejemplos son:

Categoría de Profesor: Instructor, Catedrático Auxiliar, Catedrático Asociado, Catedrático Principal.

Nivel de empleo: Desempleado, Subempleado, Empleado.

1.5 Técnicas de Muestreo

Las siguientes son las maneras más usadas de extraer muestras de una población.

a) **Muestreo Aleatorio.** Se usa cuando a cada elemento de la población se le quiere dar la misma oportunidad de ser elegido en la muestra.

b) **Muestreo Estratificado.** Se usa cuando se conoce de antemano que la población está dividida en estratos, que son equivalentes a categorías y los cuales por lo general no son de igual tamaño. Luego, de cada estrato se saca una muestra aleatoria, usualmente proporcional al tamaño del estrato. La muestra combinada formará la muestra estratificada. Este método de muestreo es muy usado en las encuestas de opinión pública.

c) **Muestreo por conglomerados (“Clusters”).** En este caso la población se divide en grupos llamados conglomerados. Luego, se elige al azar un cierto número de ellos y todos los elementos de los conglomerados elegidos forman la muestra. Un ejemplo típico es cuando se hacen encuestas en una ciudad y cada uno de los bloques de la misma, forma un estrato.

d) **Muestreo Sistemático.** Se usa cuando los datos de la población están ordenados en forma numérica. La primera observación es elegida al azar de entre los primeros elementos de la población y las siguientes observaciones son elegidas guardando la misma distancia entre si. Por ejemplo, supongamos que la población es de tamaño 1000 y deseamos sacar una muestra de tamaño 20, en este caso se divide a la población en $1000/20 = 50$ partes. Luego de entre las observaciones 1 al 50 se elige una de ellas al

azar, supongamos que salió la observación 37, ese sería el primer elemento de la muestra, los siguientes serían elegidos de 50 en 50. La muestra consistiría de las observaciones 37, 87, 137, 187, 237, 287, 337, 387, 437, 487, 537, 587, 637, 687, 737, 787, 837, 887, 937 y 987.

1.6 Maneras de Recolectar Datos

- a) Haciendo entrevistas personales. Puede ser el método más efectivo en muchas ocasiones pero es costoso y requiere bastante tiempo para ser ejecutado.
- b) Haciendo entrevistas por teléfono. Tiene la desventaja de que el entrevistado puede no ser sincero en sus contestaciones o puede negarse a contestar. También, muchas veces nadie contesta el teléfono, lo que implica que hay que hacer un gran número de llamadas para obtener información. En consecuencia, el método se vuelve costoso y demanda tiempo. Solo es conveniente usarlo cuando el cuestionario es corto.
- c) Mediante cuestionarios emitidos por correo. Es costoso y por lo general no más del 30% de los entrevistados retornan el cuestionario. Es más adecuado usarlo cuando el cuestionario es extenso.
- d) Por observación directa. Por ejemplo, si uno quiere hacer un estudio de tiempo de espera en la fila y el tiempo de servicio para ser atendidos en un restaurante. Lo que se hace es observar el proceso y tomar los tiempos de espera y de servicio a los clientes en un periodo de tiempo dado.
- e) A través de la Internet. Existen muchos lugares de internet donde hay datos disponibles de estudios pasados que se han hecho acerca de experimentos en distintas disciplinas.
- f) Usando simulación por computadoras. Por ejemplo, con la ayuda de una computadora uno puede simular el proceso de atención de los clientes de un restaurante. Para ello se simulan datos que siguen una distribución probabilística pre-establecida.