

Estadística

Unidad I

- Estadística. Concepto. Importancia. Aplicaciones.

¿Qué es la Estadística?

Problema

Identifica

... .. Población

.

.. .. .

.

.

mide variables

Estadística Descriptiva describe muestras gráficas

efectúa pruebas construye intervalos

Estadística Inferencial hipótesis de confianza

toma decisiones estima parámetros

s/ parámetros

Resuelve el

problema

Objetivo Estadística

Identifica parámetros variables,

Selecciona tamaños de la muestra,

Significancia, confiabilidad.

Diseña un experimento, plan de

muestreo.

1.2 Conceptos Básicos

Población: Conjunto total de elementos objeto de estudios, también se denomina universo colectivo. La población puede estar formada por personas, hogares, instituciones u objetos.

Marco: Conjunto de elementos de la población total disponibles para la elección de la muestra.

En ocasiones todos los elementos de la población están disponibles y por lo tanto marco = población.

Ej. Población total alumnos, marco alumnos presentes.

Unidad muestral: Unidad seleccionada de la población para la aplicación de la técnica de investigación. Puede ser un elemento poblacional único o un conjunto de elementos.

Ej. personas, familias, etc.

Muestra: Conjunto de unidades muestrales o porción de la población que se selecciona para su análisis.

Parámetro: Es una medida de resumen que se calcula para describir una característica de toda la población.

Ej. censo, promedio de notas.

Estadística: Es una medida de resumen que se calcula para describir una característica de una sola muestra de la población.

Ej. media de notas de 10 alumnos seleccionados.

Cambia un elemento de la muestra ya es otra muestra.

1.3 Estadística descriptiva e inferencia estadística

Estadística Descriptiva: Abarca los métodos que incluyen la recolección, presentación y caracterización de un conjunto de datos con el fin de describir apropiadamente las diversas características de ese conjunto de datos.

Estadística Inferencial: Abarca aquellos métodos que hacen posible la estimación de una característica de una población o la toma de una decisión referente a una población, basándose sólo en los resultados de una muestra.

Por lo tanto un aspecto importante de la estadística inferencial es el proceso de usar las estadísticas de muestras y descriptivas para sacar conclusiones respecto a los parámetros de una población.

Tipos de estudios estadísticos.

Enumerativos: Involucran las tomas de decisiones, respecto a las características de una población. Ej.: sobre el curso, conteo de quienes no tienen auto, y tomo la decisión de implementar un transporte para los alumnos.

Analítico: Involucran realizar algún cambio sobre un proceso para mejorar el desempeño en el futuro. Ej.: decidir poner la caja registradora al final y no al inicio para que el cliente vea todo primero. Estudio de preferencia de un detergente, se estudió que el envase no era el correcto, se analizó y se lo cambió.

El análisis de ambos es importante para seleccionar el método de muestreo a utilizar.

1.4 Datos o Variables: Información suministrada u obtenida, necesaria para la toma de decisiones.

Objetivos o propósitos importantes de la recolección de datos

- Proporcionar el inicio a un estudio de investigación.
- Permiten evaluar el desempeño en un servicio o proceso de producción en curso.
- Ayudan a establecer cuales serán las medidas de acción en el proceso de toma de decisiones.
- Satisfacen las curiosidades.

Obtención de datos

Básicamente son 4 métodos:

- Datos publicados por fuentes gubernamentales, industriales o individuales (INDEC).

Fuentes de datos:

- ◆ Primario: recolector original de los datos. Ej. p/ censos Ministerio Educación.
- ◆ Secundario: Órganos o individuos que presentan estos datos en tablas o diagramas. Ej. INDEC.
 - ◇ Datos obtenidos a través del diseño de un experimento. Ej. Automotriz que busca un nuevo sistema de encastre de bulones con experimentos a cargo de especialistas.
 - ◇ Datos recopilados al conducir una encuesta.
 - ◇ Datos recogidos de hacer observaciones del comportamiento, aptitudes u opiniones de los individuos en los que estamos interesados, generalmente en su entorno natural. Ej. rata laboratorio, astrología, geología.

GIGO (Garbage In Garbage Out)

Basura entra Basura sale

Si se ponen datos inciertos o erróneos, se obtendrán resultados inciertos o erróneos.

Obtención de datos mediante investigación de encuestas.

Pasos:

- **Determinación de los objetivos:** Hablar con la persona que nos encargó el trabajo para saber que es lo que necesita. A través de preguntas.
- **Determinación del tipo de encuesta:**

Tipos de encuesta:

- Personal

Ventajas

- ◇ Contacto directo
- ◇ Como se expresa
- ◇ Sinceridad
- ◇ Lo podemos guiar para ayudarlo a responder las preguntas

Desventajas

- ◆ Tiempo
- ◆ Podemos inducir a una respuesta
 - ◇ Telefónica

Ventajas

- Costos
- Tiempo
- Volumen de encuestados mayor

Desventajas

- No se conoce a la persona, se puede mentir al encuestador
- No todos tienen teléfono
- Horario indeseado
- ◇ Postal (fax, correo, mail)

Desventajas

- No te contestan
- Me contestarían únicamente mis clientes

La elección del tipo de encuesta se basa principalmente en 3 aspectos:

- Tema de investigación
- Características del entrevistado
- Presupuesto disponible
- **Diseño del cuestionario:** Es un instrumento que contiene varias preguntas y que trata de una diversidad de fenómenos o características. Éstos fenómenos o características se denominan *variables aleatorias* –cambia mucho bajo las mismas condiciones– (los datos generados por las variables aleatorias cambian de una respuesta a otra).

Las variables aleatorias pueden generar 2 tipos de datos:

- ◇ Datos categóricos o respuestas categóricas cualitativas
 - Por si o por no, o explica, selecciona entre varias opciones
- ◆ Datos numéricos o datos cuantitativos.
 - ◇ Discretos: Sólo pueden tomar determinados valores (surgen de un proceso de conteo) que cantidad de materia tenemos aprobadas nos dan números enteros. Ej. cantidad de años, edad.
 - ◇ Continuos: cuando pueden tomar valores tan próximos como se quiere (surgen de un proceso de medición) Ej. longitud, talla de una persona, peso, etc.

Niveles de escala de medición.

Éstos datos generados por las variables aleatorias (preguntas) pueden describirse de acuerdo con el nivel de medición obtenido, o bien, de acuerdo a la información brindada.

Datos categóricos:

Escala nominal: permite determinar solamente las categorías u opciones de respuesta con las que se identifica el entrevistado (sólo la nombra).

Variable: Ej.: ¿cuál es la marca de gaseosa preferida?

() Marca A

(X) Marca B

() Marca C

Sólo preferencia, clasificación.

Escala ordinal: Define la relación ordenada entre las categorías o alternativas de respuesta.

Variable: Ej.: Ordene según su preferencia las siguientes marcas de gaseosa.

(3) Marca A

(1) Marca B

(2) Marca C

Preferencia y orden de preferencia, clasificación y orden.

Datos numéricos:

Escala de Intervalo: Proporciona información de clasificación, ordenación de alternativas y adicionalmente sobre sus diferencias. La asignación del origen o punto cero es arbitraria y las distancias entre los valores de la escala son iguales pero no existe proporcionalidad entre ellos.

Variable: Ej.: Valore según su preferencia en una escala del 1 al 10 las siguientes marcas de gaseosa:

(01) Marca A

(10) Marca B

(02) Marca C

Escala de cociente o de raíz a razón: Administra la misma información que la escala de intervalo pero además se pueden interpretar los valores de la escala. El origen o cero es absoluto.

Variables: Ej.: Aproximadamente ¿cuántas gaseosas ha comprado durante la última semana de las siguientes marcas?

(00) Marca A

(15) Marca B

(05) Marca C

Orden, preferencia y cantidad de consumo (proporcionalidad)

- **Codificación del cuestionario**: Una vez diseñado y testeado el cuestionario, conviene codificarlo antes de realizar el trabajo de campo. Con el objeto de evitar posibles errores de diseño. La codificación del cuestionario facilita la tabulación de los datos resultantes y su análisis posterior a través de un programa informático de análisis estadístico.
- **Muestreo**: En vez de hacer un censo completo, los procedimientos de muestro estadístico se han convertido en la herramienta preferida en la mayoría de las situaciones de investigación.

Existen 3 razones principales para extraer una muestra:

- ◊ lleva menos tiempo que un censo completo
- ◊ es menos costoso
- ◊ es más ágil y eficiente que obtener un conteo completo de la población objeto.

Métodos de muestreo: como voy a seleccionar la muestra.

Existen 2 métodos de muestreo a los cuáles se denomina:

- Muestras no Probabilísticas.
- Muestras Probabilísticas
- Muestras no Probabilísticas: En éste tipo de procedimientos, los miembros de la población no tienen una probabilidad conocida de pertenecer a la muestra. Éstos métodos no permiten establecer las desviaciones sufridas en los resultados de la investigación, y por lo tanto, las estimaciones obtenidas no pueden proyectarse estadísticamente a la totalidad de la población.

Normalmente los métodos no probabilísticos se utilizan en estudios exploratorios o intencionales, en los cuales no es necesario proyectar los resultados. Por lo tanto, los resultados de un estudio con muestreo no probabilístico pueden ser totalmente válidos siempre que se utilicen adecuadamente y se asuman sus limitaciones.

Clasificación:

- Muestreo por conveniencia: Éste procedimiento consiste en seleccionar las unidades muestrales más convenientes para el estudio, o en permitir que la participación de la muestra sea totalmente voluntaria.
- Muestreo por criterio: Éste procedimiento se basa en el criterio o juicio del investigador para seleccionar unidades muestrales representativas.
- Muestreo por cuotas: En éste método, primero se realiza una estratificación de la muestra que garantice la variedad de criterios y características de la población objeto en estudio, y luego se aplica un muestreo por criterio para seleccionar las unidades muestrales de cada estrato. Ej.: dividido en sexo y edad.
- Muestreo de bola de nieve: Éste método consiste en solicitar de las propias unidades muestrales captadas, la identificación de posibles nuevos elementos de la muestra. Va incrementando la cantidad de gente. Ej.: cuando nos preguntan 3 nombres que pudieran estar interesados, y así sucesivamente.

2. Muestreos Probabilísticos: Son los más usados.

Ventajas: Se pueden evaluar las desviaciones cometidas en las estimaciones de las características, y por lo tanto realizar una proyección hacia la población.

Desventajas: Hay que tener censado todo el marco de la población objeto de estudio

Estas muestras son aquellas en las que los sujetos se eligen sobre la base de probabilidades conocidas. Éstas son:

- Muestreo aleatorio simple: una vez censado el marco de la población, se asigna un número a cada miembro y se eligen aleatoriamente las unidades muestrales, es decir que cada individuo o elemento tiene la misma oportunidad de selección que cualquier otro, y la selección de un individuo o elemento en particular no afecta la probabilidad de que se elija cualquier otro. Ej.: bolillero de Sabado Bus (luego se incrementa la posibilidad), bolilla o pecera, números aleatorios.

Con reemplazo: Se pone de nuevo la bolilla.

Sin reemplazo: No se pone de nuevo la bolilla.

En ambos tipos de procedimientos (bolillero o nros. Aleatorios) la muestra podría obtenerse con reemplazo o sin reemplazo.

Como se usa una tabla de números aleatorios: Ej.: Si tenemos una población $N=96$, extraiga una muestra de tamaño $n=16$ sin reemplazo, comenzando de la fila 13 de la tabla (55-70-02-45-86-93-24-73-25-96-11-56-33-44-14-31)

- **Muestro Sistemático:** Es un muestreo aleatorio en el cual se eligen los elementos de la población a intervalos uniformes a partir de un listado ordenado (dentro del marco, tomar un número y repetirlo varias veces hasta terminar la lista)

Se utilizan en tipos de estudio enumerativos.

- **Muestreo estratificado:** Lo primero que hace el investigador es clasificar los elementos de la población en subgrupos separados de acuerdo a una o más características. Después se obtiene por separado una muestra aleatoria simple o sistemática de cada estrato. Ej.: 2 estratos- varones, mujeres- dentro de c/u edades.

Ventajas: mayor representatividad de la muestra seleccionada. Que la opinión que se toma sea acorde a todo el grupo.

- **Muestreo por conglomerado:** Es un tipo de muestreo aleatorio en el que los elementos de la población se dividen en forma natural en subgrupos. En éste muestreo, la unidad muestral no son los individuos sino un conjunto de individuos que, bajo determinados aspectos, se puede considerar que forman una unidad o conglomerado. En vez de censar toda la población, es necesario censar sólo los conglomerados. El método consiste en seleccionar aleatoriamente un número de conglomerados cuya suma total de elementos proporcione el tamaño muestral buscado. Ej.: se hace una campaña de un medicamento que se quiere saber si lo acepta la farmacia, son elegidas en forma aleatoria, se censa sólo a la farmacia no a los empleados.

Ventajas: menos costoso, pues sólo es necesario censar un conglomerado.

Desventajas: Bajo nivel de representatividad. Es muy probable que las personas que trabajen en la farmacia piensen relativamente lo mismo.

- **Trabajo de campo:** Una vez desarrolladas las fases anteriores, se realiza el trabajo de campo entrevistando a todos los integrantes de la muestra seleccionada. El trabajo de campo es controlado y revisado por el investigador para evitar distorsiones de la información resultante.
- **Tabulación de Datos:** Construir tablas. Al finalizar el trabajo de campo, y una vez revisados y depurados los cuestionarios, se tabulan estadísticamente los datos obtenidos.
- **Análisis de Resultados y Elaboración del Informe:** Por último, se analizan los datos tabulados, y en base a los resultados y las conclusiones se redacta el informe final de la investigación.

Luis Gonzalo Omar Molina