

CAPITULO I

INTRODUCCION

1.1 Presentación.

¿Qué es la estadística?

Se suele pensar en una relación de datos numéricos presentada de forma ordenada y sistemática. Esta idea es la consecuencia del concepto popular que existe sobre el término y que cada vez está más extendido debido a la influencia de nuestro entorno, ya que hoy día es casi imposible que cualquier medio de difusión, periódico, radio, televisión, etc, no nos aborde diariamente con cualquier tipo de información estadística sobre accidentes de tráfico, índices de crecimiento de población, turismo, tendencias políticas, etc.

"ESTADISTICA" se derivó de la palabra "ESTADO". La función de los gobiernos entre otras cosas es llevar los registros de población, nacimientos, cosechas, impuestos y toda la información que engloba el estado, es así que, tradicionalmente se definió a la estadística como un instrumento de compilación, organización, presentación y análisis de datos numéricos.

Sólo cuando nos adentramos en un mundo más específico como es el campo de la investigación de las Ciencias Sociales: Medicina, Biología, Psicología, ... empezamos a percibir que la Estadística no sólo es algo más, sino **que se convierte en la única herramienta que, hoy por hoy, permite dar luz y obtener resultados, y por tanto beneficios, en cualquier tipo de estudio**, cuyos movimientos y relaciones, por su variabilidad intrínseca, no puedan ser abordadas desde la perspectiva de las leyes determinísticas. Podríamos, desde un punto de vista más amplio, definir la estadística como la ciencia que estudia cómo debe emplearse la información y cómo dar una guía de acción en situaciones prácticas que entrañan incertidumbre.

La estadística es una ciencia aplicada de las matemáticas y es una valiosa herramienta para la toma de decisiones. Permite el estudio de fenómenos mediante la descripción del mismo a través de inferencias mediante distribuciones probabilísticas.

La Estadística se ocupa de los métodos y procedimientos para recoger, clasificar, resumir, hallar regularidades y analizar los datos, siempre y cuando la variabilidad e incertidumbre sea una causa intrínseca de los mismos; así como de realizar inferencias a partir de ellos, con la finalidad de ayudar a la toma de decisiones y en su caso formular predicciones.

Podríamos por tanto clasificar la Estadística en descriptiva, cuando los resultados del análisis no pretenden ir más allá del conjunto de datos, e inferencias cuando el objetivo del estudio es derivar las conclusiones obtenidas a un conjunto de datos más amplio.

Estadística descriptiva: Describe, analiza y representa un grupo de datos utilizando métodos numéricos y gráficos que resumen y presentan la información contenida en ellos.

Estadística inferencial: Apoyándose en el cálculo de probabilidades y a partir de datos de muestras, efectúa estimaciones, decisiones, predicciones u otras generalizaciones sobre un conjunto mayor de datos.

1.2 CONCEPTOS BASICOS

POBLACION .- Agregado de unidades elementales, que poseen alguna característica o propiedades comunes. El estudio de toda la población constituye un CENSO.

Una población puede ser finita o infinita.

En relación al tamaño de la población, ésta puede ser:

Finita, como es el caso del número de personas que llegan al servicio de urgencia de un hospital en un día; y se conoce el tamaño N de la población.

Infinita, si por ejemplo estudiamos el mecanismo aleatorio que describe la secuencia de caras y cruces obtenida en el lanzamiento repetido de una moneda al aire.

También se considera infinita, a pesar que las poblaciones son pequeña, no se puede saber con exactitud el tamaño de la población.

Ejemplos

- Todos los agricultores que cultivan papa para comercializar en el PERU.
- Todas los arboles de pijuayo en Pucallpa.

UNIDAD ELEMENTAL .- Son los "entes" que constituyen la población y de las que se va a obtener información inicial. También conocido como elementos o individuos que contienen cierta información que se desea estudiar.

Ejemplos

- Una agricultor que cultiva papa.
- Una arbol de pijuayo.

MUESTRA .- Es una parte de la población. Se espera que la muestra sea representativa de la población, es decir reproduzca las características más importantes. El proceso de obtener la muestra de denomina MUESTREO.

MUESTRA ALEATORIA .- cuando la muestra a sido obtenida empleando algún procedimiento del azar: sorteo, extracción al azar, números aleatorios, etc.

Ejemplo

- 20 agricultores de papa tomadas al azar de la población de productores de papa.

En una muestra aleatoria, cada unidad elemental tiene una determinada posibilidad (probabilidad) de pertenecer a la muestra.

OBSERVACION .- Es el registro que se obtiene al evaluar una característica en una unidad elemental.

Ejemplo

- 10 ton por hectarea (si se evaluó la producción de un agricultor).

Establecemos a continuación algunas definiciones de conceptos básicos y fundamentales básicas como son: elemento, población, muestra, caracteres, variables, etc., a las cuales haremos referencia continuamente.

VARIABLES ESTADÍSTICAS

Cuando hablemos de variable haremos referencia a un símbolo (X,Y,A,B,...) que puede tomar cualquier modalidad (valor) de un conjunto determinado, que llamaremos dominio de la variable o rango. En función del tipo de dominio, las variables las clasificamos del siguiente modo:

VARIABLES CUALITATIVAS .- Generan observaciones de carácter no numérico y son del tipo:

CUALITATIVAS JERARQUICAS .- Cuando se puede establecer una relación de orden entre las posibles observaciones. Llamadas cuasicuantitativas

CUALITATIVAS NOMINALES .- No existe un orden entre las posibles observaciones.

Ejemplos de variables cualitativas:

- Calidad de papa (V.C.J.)
- Color de la pulpa de la papa (V.C.N.)

cuando las modalidades posibles son de tipo nominal. Por ejemplo, una variable de color

VARIABLES CUANTITATIVAS .- son las que tienen por modalidades cantidades numéricas con las que podemos hacer operaciones aritméticas. Dentro de este tipo de variables podemos distinguir dos grupos:

CUANTITATIVAS DISCRETAS .- Cuando el conjunto de todas las posibles observaciones que se generan constituyen a lo más un conjunto infinito numerable. cuando no admiten siempre una modalidad intermedia entre dos cualesquiera de sus modalidades. Un ejemplo es el número de tuberculos por planta. Es obvio que cada valor de la variable es un número natural

CUANTITATIVAS CONTINUAS .- Cuando el conjunto de todas las posibles observaciones que se generan constituyen un conjunto infinito no numerable. admiten una modalidad intermedia entre dos cualesquiera de sus modalidades, v.g. el peso tuberculos de una planta cosechada. En este caso los valores de las variables son números reales, es decir ocurre a veces que una variable cuantitativa continua por naturaleza, aparece como discreta. Este es el caso en que hay limitaciones en lo que concierne a la precisión del aparato de medida de esa variable, v.g. si medimos la altura en metros de los arboles con una regla que ofrece dos decimales de precisión, podemos obtener 15.25, 18.24

En realidad lo que ocurre es que con cada mediciones se expresa que el verdadero valor de la misma se encuentra en un intervalo. Por tanto cada una de las observaciones representa más bien un intervalo que un valor concreto.

Tal como se cito anteriormente, las modalidades son las diferentes situaciones posibles que puede presentar la variable. A veces éstas son muy numerosas (v.g. cuando una variable es continua) y conviene reducir su número, agrupándolas en una cantidad inferior de clases.

Ejemplos de variables numéricas:

- Número de plantas enfermas en el campo de una hectarea (V.C.D.)
- peso de tubérculo a la cosecha. (V.C.C.)

PARAMETRO .- Es una constante que describe una característica de una población. Para poder calcular el valor de un parámetro, se requiere conocer a ciencia cierta el estado de naturaleza de la población o realizar un censo.

Principales tipos de parámetros son:

Parámetros de tendencia central o de resumen, siendo los más importantes :

- La media o promedio (μ)
- La mediana (Me)
- La moda (Mo)

Parametros de variabilidad, siendo los más importantes:

- La variancia o varianza (σ^2)
- La desviación estándar (σ)
- El coeficiente de variabilidad (C.V.)

En el caso de que la variable sea cuantitativa se usa el porcentaje o proporción.

VALOR ESTADISTICO o ESTADISTICO.- Son valores análogos a los parámetros, pero que son calculados con la información obtenida de la muestra.

Los valores estadísticos son variables porque pueden tomar diferentes valores al cambiar de muestra.

Un valor estadístico estima al parámetro correspondiente. La notación difiere respecto a los parámetros:

- Promedio \bar{X}
- Mediana me
- Moda mo
- variancia s^2
- Desviación estándar s

s^2 estima a σ^2

A los valores estadísticos también se le conoce con el nombre de estadígrafos.

El término de "medida", suele usarse para referirse a parámetros o valores estadísticos.

1.3 ETAPAS DE UNA INVESTIGACION ESTADISTICA.

Comprende las siguientes etapas:

FORMULACION O DEFINICION DEL PROBLEMA

Delimitar con exactitud que debe ser investigado:

- Población sobre lo que se va a concluir.
- Variables que van a ser de interés.

DISEÑO DEL EXPERIMENTO

Definir si se va a realizar un censo o un muestreo. En el caso de realizar un muestreo, elegir el procedimiento que permita obtener una MUESTRA REPRESENTATIVA, que implica un adecuado tamaño de muestra, técnica de muestreo.

COMPILACION DE DATOS

Establecer los procedimientos e instrumentos a usarse para obtener información pertinente de las unidades elementales comprendidas en el estudio; puede emplearse la observación directa o indirecta.

ORGANIZACION Y DESCRIPCION DE DATOS

- Ordenación en forma legible de los datos compilados
- Clasificación en forma sistemática y presentación en cuadros y/o gráficos.
- Cálculo de medidas: parámetros o valores estadísticos para describir el comportamiento de los datos.

DECISION O INFERENCIA FINAL

Cuando se ha realizado un muestreo y es necesario concluir acerca de la población se debe realizar la estimación de parámetros y la comprobación de hipótesis.

El método científico está estrechamente ligado a las etapas mencionadas.

1.4 DEFINICION DE ESTADISTICA

Es la ciencia que brinda los métodos y procedimientos que permiten realizar la caracterización, análisis e interpretación de una serie de datos para la toma de decisiones frente a situaciones de certidumbre o incertidumbre.

RAMAS DE LA ESTADISTICA : Estadística Descriptiva, Inferencia Estadística.

Ejemplo.- En el siguiente enunciado identifique: Población, Muestra, Unidad elemental, Variable(s), Tipo(s), Observación. Además defina los parámetros y valores estadísticos correspondientes.

Las tiendas dedicadas a la comercialización de artefactos eléctricos nacionales han encargado realizar un estudio con la finalidad de obtener algunos indicadores. El estudio se realizó con una muestra de 60 tiendas obteniéndose los siguientes resultados:

- El 60% de los artículos suministrados trimestralmente por los proveedores son vendidos en ese periodo.
- La utilidad neta promedio por artefacto vendido es de 15% del costo.
- El 80% de los artículos vendidos corresponden a la denominación "línea blanca".
- Promedio de devolución: 2 artículos trimestralmente debido a fallas técnicas.

Solución:

Población: Tiendas dedicadas a la comercialización de artículos eléctricos nacionales.

Muestra: 60 tiendas elegidas en el estudio.

Unidad: una tienda de la población definida.

Variables:

1. % de artículos vendidos respecto al total suministrado.
Tipo: cuantitativa discreta.
Observación: 50%
Parámetro: % promedio de ventas trimestrales de artículos vendidos respecto al total suministrado.
Valor del parámetro: desconocido.
Valor estadístico: % promedio de ventas trimestrales de artículos vendidos respecto de la muestra. Valor: 60%.
2. Utilidad neta.
Tipo: cuantitativa continua.
Observación: 10%
Parámetro: % promedio correspondiente a la utilidad neta por artefacto respecto al costo de compra del total de artículos vendidos.
Valor del parámetro: desconocido.
Valor estadístico: % promedio correspondiente a la utilidad neta por artefacto respecto al costo de compra del total de artículos vendidos seleccionados en la muestra. valor: 15%.

Completar con las otras variables del estudio.

CAPITULO II

SUMATORIAS

La letra griega sigma mayúscula, Σ , es una notación abreviada de una suma; así,

$$x_1 + x_2 + x_3 + x_4 + x_5 = \sum_{i=1}^5 x_i$$

que se lee "sumatoria de X_i , de $i=1$ a $i=5$. A "i" se le denomina índice de adición de la sumatoria.

2.1 SUMATORIA SIMPLE

Cuando el proceso de sumarización se realiza sobre un sólo índice.

PROPIEDADES

1.- El desarrollo de una sumatoria cuyo límite inferior es "a" y el límite superior es "b".

$$\sum_{j=a}^b x_j = x_{a+1} + x_{a+2} + \dots + x_b \text{ tiene } (b - a + 1) \text{ términos.}$$

2.- La sumatoria de una constante es igual al Nro. De términos de la sumatoria multiplicada por la constante.

$$\sum_{i=a}^b k = (b - a)k$$

Se define constante para una sumatoria, a todo término no afectado por el índice de dicha sumatoria.

3.- La sumatoria de una constante por una variable es igual a la constante multiplicada por la sumatoria de la variable

$$\sum_{i=a}^b k x_i = k \sum_{i=a}^b x_i$$

4.- La sumatoria es distributiva respecto a la adición y/o sustracción.

$$\sum_{i=1}^b (x_i + y_i - z_i) = \sum_{i=1}^b x_i + \sum_{i=1}^b y_i - \sum_{i=1}^b z_i$$

5.- El índice de la sumatoria es una variable ficticia es decir se puede cambiar en toda la expresión de la sumatoria sin alterar el resultado que ésta representa.

$$\sum_{i=a}^b x_i = \sum_{j=a}^b x_j$$

6.- Se puede descomponer en la suma de dos o más sumatorias

$$\sum_{i=1}^n x_i = \sum_{i=1}^m x_i + \sum_{i=m+1}^n x_i; \quad m < n$$

EJERCICIO

Si $Q = \sum_{i=1}^n (x_i - h)^2$, demostrar que esta función dependiente de "h", alcanza el valor más pequeño cuando este es igual a:

$$h = \bar{x} = \sum_{i=1}^n \frac{x_i}{n}$$

Solución:

Del enunciado debe cumplirse que:

$$\sum_{i=1}^n (x_i - h)^2 \geq \sum_{i=1}^n (x_i - \bar{x})^2$$

$$\sum_{i=1}^n (x_i^2 - 2hx_i + h^2) \geq \sum_{i=1}^n (x_i^2 - 2\bar{x}x_i + \bar{x}^2)$$

$$\sum_{i=1}^n x_i^2 - 2h \sum_{i=1}^n x_i + nh^2 \geq \sum_{i=1}^n x_i^2 - 2\bar{x} \sum_{i=1}^n x_i + n\bar{x}^2$$

Simplificando

$$h^2 - 2h\bar{x} + \bar{x}^2 \geq 0$$

$$(h - \bar{x})^2 \geq 0$$

Lo cual es siempre cierto.

2.2 SUMATORIA DOBLE

Cuando la variable observada se categoriza por dos criterios, se tiene dos subíndices para la variable, y un elemento en particular se define como X_{ij} .

Por ejemplo:

$$\sum_{i=1}^3 \sum_{j=1}^2 x_{ij} = x_{11} + x_{12} + x_{21} + x_{22} + x_{31} + x_{32}$$

PROPIEDADES

1.- El desarrollo de una sumatoria $\sum_{i=a}^b \sum_{j=c}^d x_{ij}$ tiene $(b-a+1)(d-c+1)$ términos

2.- $\sum_{i=1}^n \sum_{j=1}^r k = nrk$

3.- $\sum_{i=1}^n \sum_{j=1}^r x_i = \sum_{i=1}^n r x_i = r \sum_{i=1}^n x_i$

4.- $\sum_{i=1}^n \sum_{j=1}^r y_j = \sum_{j=1}^r n y_j = n \sum_{j=1}^r y_j$

5.- $\sum_{i=1}^n \sum_{j=1}^r x_i y_j = \sum_{i=1}^n x_i \sum_{j=1}^r y_j$

6.- $\sum_{i=1}^n \sum_{j=1}^r (x_i \pm y_j) = \sum_{i=1}^n x_i \sum_{j=1}^r y_j \pm \sum_{j=1}^r y_j \sum_{i=1}^n x_i$

7.- $\sum_{i=1}^n \sum_{j=1}^r K x_{ij} = K \sum_{i=1}^n \sum_{j=1}^r x_{ij}$

8.- $\sum_{i=1}^n \sum_{j=1}^r x_{ij} = \sum_{j=1}^r \sum_{i=1}^n x_{ij}$

EJERCICIOS

1.- Expresar mediante una sumatoria doble la siguiente expresión
 $(X_{11} + X_{12} - 4) \frac{1}{3} + (X_{21} + X_{22} - 4) \frac{1}{3} + (X_{31} + X_{32} - 4) \frac{1}{3} + (X_{41} + X_{42} - 4) \frac{1}{3} + 8b$

SOLUCION

$$\sum_{i=1}^4 \left[\left[\begin{array}{c} 2 \\ \sum_{j=1} X_{ij} - 4 \end{array} \right]^{1/3} \right] + 8b$$

2.- Probar si A y B son iguales, siendo:

$$A = \sum_{i=1}^n \sum_{j=1}^m \left[(X_{ij} + 2\bar{X})^2 - 2\bar{X} \sum_{k=1}^m X_{ik}/m \right]$$

$$B = \sum_{i=1}^n \sum_{j=1}^m X_{ij}^2 + 6nm \left[\sum_{i=1}^n \sum_{j=1}^m X_{ij} \right]^2$$

Donde:

$$\bar{X} = \frac{\sum_{i=1}^n \sum_{j=1}^m X_{ij}}{nm}$$

SOLUCION

$$A = \sum_{i=1}^n \sum_{j=1}^m (X_{ij} + 2\bar{X})^2 - \sum_{i=1}^n \sum_{j=1}^m 2\bar{X} \sum_{k=1}^m X_{ik}/m$$

$$A = \sum_{i=1}^n \sum_{j=1}^m (X_{ij} + 2\bar{X})^2 - 2m\bar{X} \sum_{i=1}^n \sum_{k=1}^m X_{ik}/m$$

$$A = \sum_{i=1}^n \sum_{j=1}^m (X_{ij} + 2\bar{X})^2 - 2mn\bar{X}^2$$

$$A = \sum_{i=1}^n \sum_{j=1}^m X_{ij}^2 + 4\bar{X} \sum_{i=1}^n \sum_{j=1}^m X_{ij} + 4mn\bar{X}^2 - 2mn\bar{X}^2$$

$$A = \sum_{i=1}^n \sum_{j=1}^m X_{ij}^2 + 6nm\bar{X}^2$$

en la expresión de B

$$B = \sum_{i=1}^n \sum_{j=1}^m X_{ij}^2 + 6nm \left[\sum_{i=1}^n \sum_{j=1}^m X_{ij} \right]^2$$

$$B = \sum_{i=1}^n \sum_{j=1}^m X_{ij}^2 + 6nm \left[nm \bar{X} \right]^2$$

$$B = \sum_{i=1}^n \sum_{j=1}^m X_{ij}^2 + 6 n^2 m^2 \bar{X}^2$$

comparando A y B, se concluye que son diferentes.

2.3 SUMATORIA DOBLE CON UNO DE LOS INDICES CON LIMITE SUPERIOR VARIABLE.

La sumatoria se expresa como:

$$\sum_{i=1}^n \sum_{j=1}^{r_i} X_{ij}$$

el límite superior de "j" depende del valor de "i"

PROPIEDADES

A las propiedades mencionadas en la sumatoria doble, se agregan:

1.- No se puede permutar el operador sumatoria.

$$\sum_{i=1}^n \sum_{j=1}^{r_i} X_{ij} \quad \text{por} \quad \sum_{j=1}^{r_i} \sum_{i=1}^n X_{ij}$$

2.- El desarrollo de $\sum_{i=1}^n \sum_{j=1}^{r_i} X_{ij}$ tiene $\sum_{i=1}^n r_i$ términos.

2.4 NOTACION PUNTUAL.- Se usa para simplificar la escritura y se definen como:

1.- $\sum_{i=1}^n r_i = r.$

2.- $\sum_{i=1}^n \sum_{j=1}^m X_{ij} = \sum_{i=1}^n X_{i.} = X_{..}$

3.- $\sum_{i=1}^n X_{ij} = X_{.j}; \quad \sum_{j=1}^m X_{ij} = X_{i.}; \quad \sum_{j=1}^{r_i} X_{ij} = X_{i.}$

Ejercicio : simplificar

$\sum_{i=1}^t \left[\sum_{j=1}^b n_{ij} \bar{X}_{.j} - r \bar{X}_{..} \right]$, siendo

$\bar{X}_{.j} = X_{.j} / k \quad \bar{X}_{..} = X_{..} / (rt) \quad X_{.j} = \sum_{i=1}^t n_{ij} X_{ij}$

$X_{..} = \sum_{i=1}^t \sum_{j=1}^b n_{ij} X_{ij} \quad \sum_{i=1}^t n_{ij} = k$

SOLUCION

$$\begin{aligned} \sum_{i=1}^t \left[\sum_{j=1}^b n_{ij} \bar{X}_{.j} - r \bar{X}_{..} \right] &= \sum_{i=1}^t \sum_{j=1}^b n_{ij} \bar{X}_{.j} - \sum_{i=1}^t r \bar{X}_{..} = \\ &= \sum_{i=1}^t \sum_{j=1}^b n_{ij} X_{.j} / k - X_{..} = \left[\sum_{j=1}^b X_{.j} / k \right] \sum_{i=1}^t n_{ij} - \sum_{j=1}^b X_{.j} - X_{..} \\ &= \sum_{i=1}^t \sum_{j=1}^b n_{ij} X_{ij} - X_{..} = 0 \end{aligned}$$

APLICACION

El siguiente cuadro muestra la cantidad de toneladas de azúcar transportada desde los almacenes mayorista de San Luis a los mercados de Santa Anita.

		M E R C A D O S		
		1	2	3
A	1	4	6	7
L	2	6	7	2
M	3	5	4	9
A	4	7	0	2
C	5	3	0	0
E				
N		25	17	20

X_{ij} : cantidad en toneladas de azúcar transportada desde el almacén "i" a los mercados "j".

$$j=1,2,3; \quad i=1,2,\dots,5$$

Represente mediante la expresión más sintética en términos de sumatoria y encuentre el valor de:

1.- Cantidad de azúcar transportada a los mercados.

$$\sum_{i=1}^5 \sum_{j=1}^3 x_{ij} = 62 \text{ ton.}$$

2.- Cantidad total de azúcar transportada desde los almacenes a los mercados 1 y 3.

$$x_{11} + x_{21} + x_{31} + x_{41} + x_{51} + x_{13} + x_{23} + x_{33} + x_{43} + x_{53} = \sum_{i=1}^5 \sum_{j=1}^2 x_{i,2j-1} = 45$$

3.- Si los precios de venta por kilos de azúcar de los mercados 1, 2 y 3 son respectivamente:

$$\begin{aligned} P_1 &= S/. 2.00 \\ P_2 &= S/. 2.30 \end{aligned}$$

$$P_3 = S/. 2.10$$

Halle el ingreso bruto por las ventas en los mercados 1 y 3.

$$1000 p_1 (x_{11} + x_{21} + x_{31} + x_{41} + x_{51}) + 1000 p_3 (x_{13} + x_{23} + x_{33} + x_{43} + x_{53})$$

$$1000 p_1 \sum_{i=1}^5 x_{i1} + 1000 p_3 \sum_{i=1}^5 x_{i3} = 1000 \sum_{j=1}^2 p_{2j-1} \sum_{i=1}^5 x_{i,2j-1} = S/.92,000$$

- 4.- Conociendo la información dada en (3) y sabiendo además el costo del azúcar por sacos de 50 kilos es de S/.80 y los costos de transporte por tonelada desde los almacenes 1, 2, 3, 4 y 5 son variables:

$$C_1 = S/. 120.00$$

$$C_2 = S/. 72.00$$

$$C_3 = S/. 96.00$$

$$C_4 = S/. 120.00$$

$$C_5 = S/. 72.00$$

Halle la ganancia total de los mercados 1 y 3.

1 ton = 1000 kilos, costo por ton del azúcar = 1600,

Costo del azúcar:

$$C_1 x_{11} + C_2 x_{21} + C_3 x_{31} + C_4 x_{41} + C_5 x_{51} + C_1 x_{13} + C_2 x_{23} + C_3 x_{33} + C_4 x_{43} + C_5 x_{53}$$

$$\sum_{i=1}^5 c_i \sum_{j=1}^2 x_{i,2j-1} = 4,536; \text{ Utilidad} = 92000 - 45 \cdot 1600 - 4536 = S/.15,464.00$$