

Conceptos Basicos de Estadistica

makita

10/14/2021

Conceptos básicos de estadística

Variables: género, edad, peso, talla, estrato, localidad, fecha, lugar de nacimiento..

Población (universo o colectivo)

Es el conjunto total de ELEMENTOS de la misma naturaleza cualquiera que sea, que son de interés para un problema dado

Parámetros: IMC

Clasificación de variables

N = Representación de el tamaño de la población

Cualitativas (categóricas)

Cuantitativas

muestra

Los valores de las observaciones son numéricas y en consecuencia, ordenables.

Variable aleatoria:

Son fenómenos o características de los elementos de la población.

Función de valor real que tiene como dominio el espacio muestral de un experimento aleatorio.

Variables sobre las cuales tenemos un grado de incertidumbre respecto a los valores que puede tomar

Discreta

Recorridos finitos numerables sin tomar valores intermedios e.g. conteos.

Continua

Recorridos infinitos no numerables e.g. la distribución normal

Datos

Son los resultados observados de las variables aleatorias (Cuando se hace una medición)

Escalas de medición

Parámetro

Es la medición global de cualesquier característica de los elementos de la población.

Es un valor teórico asociado a la población.

Cualitativas

Nominales Clasificación de objetos o fenómenos mediante símbolos o signos (No hay orden o dirección). e.g.

Ejemplos

Población: Los niños y niñas de 0 a 5 años de edad localizados en Bogotá

- Nombre
- Número de la cédula
- Tipo de sangre
- Color de los ojos
- Número de camiseta de los jugadores

Los números en la lista anterior no pueden ser sometidos a operaciones matemáticas

Ordinales

Categorías ordenadas (Rangos, órdenes, escalamientos)

- Sabor de un yogurt

Cuantitativas

Intervalo

Los datos medidos en una escala ordinal para los cuales pueden clasificarse las distancias entre valores pero no existe un cero absoluto o no exista ausencia total de la característica

- Temperatura: a 0°C no deja de existir la temperatura
- Notas: se corre la escala e inicia desde 3.

Razón

Tiene todas las características de un intervalo, y además tiene un cero absoluto

Resumen y descripción de datos de una variable

Datos en bruto en forma de listas (o bases no son fáciles de usar para tomar decisiones)

- Se necesita algún tipo de organización

Para esto podemos utilizar gráficos de barras, gráficos de torta, o tablas de frecuencias.

#Como agrupar los datos: Sturges

Si n no es demasiado grande, intervalos = \sqrt{n}

En caso contrario:

$$k = 1 + 3.322 \log(n)$$

k = intervalos de clase

Para la longitud de los intervalos:

$$L = \text{Dato mayor} - \text{Dato menor} / n$$

- A menudo es prueba y error

Tipos de frecuencias

- Absoluta: Conteo de observaciones que cae en cada intervalo.
- Relativa: $\frac{\text{Absoluta}}{n}$.
- Acumulada: Suma de las frecuencias absolutas
- Relativa acumulada: Suma de las frecuencias relativas.

Características a revisar de las distribuciones

- Distribucion
- Localizacion (sesgo)
- Dispersion

Medidas de localización

Media aritmética:

Si $x_1, x_2, x_3, \dots, x_n$ es una muestra de una población de tamaño N entonces la media es N

$$\mu = \frac{\sum_{i=1}^n x_i}{N} \text{ Media poblacional}$$

Estimador muestral

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \text{ Media poblacional}$$

Características:

- Es fácil de obtener
- Medida no robusta: Afectada por valores extremos o datos atípicos.

Propiedades:

Si $x_1, x_2, x_3, \dots, x_n$ es una muestra de una población de tamaño N entonces la media es N , entonces

$$\sum_{i=1}^n x_i = x_1 + x_2 + x_3 + \dots + x_n$$

Si $x_i = c$ constante entonces $\sum_{i=1}^n x_i = \sum_{i=1}^n c = c + c + c + \dots$ n veces

$$\text{Entonces } \sum_{i=1}^n x_i = nc$$

Si c es una constante que multiplica las observaciones

$$\sum_{i=1}^n cx_i = c \sum_{i=1}^n x_i$$

Si $x_1, x_2, x_3, \dots, x_n$ y $y_1, y_2, y_3, \dots, y_n$ son sucesiones de números;

$$\sum_{i=1}^n (x_i + y_i) = \sum_{i=1}^n x_i + \sum_{i=1}^n y_i$$

Si $x_1, x_2, x_3, \dots, x_n$ y $y_1, y_2, y_3, \dots, y_n$ son sucesiones de números;

$$\sum_{i=1}^n (x_i - y_i) = \sum_{i=1}^n x_i - \sum_{i=1}^n y_i$$

5.

$$\sum_{i=1}^n (x_i - \bar{x}) = 0$$

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

6.

promedio de y en función de promedio de x en regresión lineal simple

Si A

La mediana

Es el valor central (es el dato de la variable que está en el centro de la misma). Deja por encima y por debajo mitad y mitad de las observaciones.

Calculo de la mediana

Depende si el conjunto es par o impar:

Si $x_1, x_2, x_3, \dots, x_n$ Son los valores ordenados en una muestra de una población de tamaño N :

$$\hat{x} = \frac{x_{n/2} + x_{n+1/2}}{2} \text{ si } n \text{ es par}$$

$$\hat{x} = x_{n+1/2} \text{ si } n \text{ es impar}$$

Es un estimador robusto, no se ve afectado por valores extremos

Ejemplo

Edad de niños

```
x1 <- c(6,7,8,9,10)
```

n es impar, entonces $\hat{x} = x_{n+1/2} = x_{6/2} = x_3 = 8$

De la muestra analizada la mitad de los niños tienen entre 6 y 8 años, y la otra mitad entre 8 y 10 años.

Moda

- El valor que más se repite
- Usada para valores numéricos o categóricos

e.g. ¿Cuál es el color más frecuente en los ojos.

Medidas de dispersión o variación

Varianza

- Varianza poblacional:

$$\sigma^2 = \frac{\sum_{i=1}^N (x_i - \mu)^2}{N} \quad (1)$$

- Varianza muestral:

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n} \quad (2)$$