

”Probabilidad y estadística”

Deniso Xocuis

5 de abril del 2023

Estadística

Una **población** es el conjunto de mediciones de interés para el investigador. Una **muestra** es un subconjunto de mediciones seleccionado de la población de interés. Se pueden clasificar variables en dos categorías: cualitativas y cuantitativas.

Las variables **cualitativas** miden una cualidad o característica en cada unidad experimental. Las **cuantitativas** miden una cantidad numérica en cada unidad experimental.

Una vez recolectados los datos, éstos pueden consolidarse y resumirse para mostrarla en una tabla estadística para la distribución de datos dependiendo los tipos de variables. Cuando la variable de interés es cualitativa, la tabla estadística es una lista de las categorías siendo consideradas junto con una medida de la frecuencia con que se presenta cada valor. Se puede medir “la frecuencia” en tres formas diferentes:

- La frecuencia o número de mediciones en cada categoría
- La frecuencia relativa o proporción de mediciones en cada categoría
- El porcentaje de mediciones en cada categoría

$$\text{Frecuencia relativa} = \frac{\text{frecuencia}}{n}$$

Donde “n” es la suma total de frecuencia, es decir, el total de datos

Ejemplo:

	f	x	x*f	$(x - \bar{x})^2$	$(x - \bar{x})^2 * f$
17-23	370	20	7400	12.18	4506.6
23-29	35	26	910	6.30	220.5
29-35	5	32	160	72.42	362.1
35-41	3	38	114	210.54	631.62
41-47	7	44	308	420.06	2940.42
47-53	5	50	250	702.7	3513.5
53-59	2	56	112	1056.9	2113.8
59-65	8	62	496	1483	11864
65-71	10	68	680	1981	19810
71-77	4	74	296	2551	10204
77-83	1	80	80	3193	3193
	n = 460		= 10806		= 59359

$$\text{media} = x = \frac{L_2 + L_1}{2}$$

Donde L2 es el límite superior y L1 el límite inferior.

$$\bar{x} = \frac{\Sigma(x * F)}{n} = \frac{10806}{460} = 23.49(\text{promedio})$$

$$S^2 = \frac{\Sigma(x - \bar{x})^2 * F}{n} = \frac{59359}{460} = 129.04(\text{varianza}) \longrightarrow S = \sqrt{129.04} = 11.35 (\text{desviación estándar})$$

$$CV = \frac{S}{\bar{x}} = \frac{11.35}{23.49} * 100 = 48.36\% \text{ (heterogéneo, muy dispersos)}$$

El coeficiente de varianza nos dice que es homogéneo si $CV < 25\%$ y heterogéneo si $CV > 25\%$

1 Frecuencia

Número de veces se repite un dato, la suma total de frecuencia son los datos totales.

1.1 Frecuencia relativa

$$f_{ri} = \frac{f_i}{n}$$

Para datos agrupados en intervalos:

$$\bar{x} = \frac{\sum x_i f_i}{n} \text{ (media)}$$

$$Me = Li + \frac{\frac{n}{2} - F_i - 1}{f_i} * a_i \text{ (mediana, donde Li es límite inferior)}$$

Se puede encontrar el dato de en medio de la siguiente forma:

si el número es impar: $\frac{n+1}{2}$, y si es par: $\frac{n}{2}$, el dato que nos de se busca en **la frecuencia absoluta acumulada**, si el dato es exacto a alguno de la frecuencia absoluta acumulada, su media es el límite superior.