Medidas Tendencia Central

Supongamos que se tiene una muestra $\{x_i\}_{i=1}^{n}$ de cierta variable numérica. Dentro de las medidas resumen se distinguen tres en particular: de tendencia central, de dispersión, cuantiles.

Una medida de tendencia central, también conocida como promedio, permite representar a toda la población mediante un valor.

Media y mediana

La **media** de la muestra aleatoria $\{x_i\}_{i=1}^{n}$ es

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad (1.1)$$

La **mediana** es el valor en la posición intermedia (50 % de los datos) una vez que se ordenan de manera creciente.

En el caso de la mediana si la cantidad de datos es impar coincide con uno de los datos, si es par se toma el promedio entre los dos valores intermedios.

Ejemplo 1. Considere el conjunto de datos ordenados $\{3, 5, 6, 6, 7, 9, 12, 13\}$. Entonces,

$$\bar{x} = \frac{1}{8}(3 + 5 + \cdots + 13) = 7.625$$

$$m = 6.5$$

En el caso de la mediana, de los datos en la posición intermedia, corresponden a las posiciones 4 y 5, esto es, los datos 6 y 7, tomando su promedio se tiene m = 6.5.

Cuando hay una gran cantidad de datos la media es recomendable. Para pocos datos es preferible el uso de la mediana como promedio. En muestra pequeñas la diferencia puede ser grande en presencia de puntos atípicos.

Ejemplo 2. Considere el conjunto de datos ordenados
{3, 5, 6, 6, 7, 9, 12, 30}. Entonces,

$$\bar{x} = \frac{1}{}(3 + 5 + \cdots + 30) = 9.75$$

$$m = 6.5$$

Comparando con los resultados de los ejemplos 1 y 2, la mediana no cambia su valor, sin embargo, la media se incrementa de 7.625 a 9.75. Es claro que este incremento de la media se debe al valor atípico de 30 y por tanto es preferible el valor de la mediana al de la media para representar a la población.

Medidas de dispersión

Las medidas de dispersión establecen el grado de homogeneidad (heterogeneidad) de un conjunto de datos.

rango,varianza y desviación estándar

Sea $\{x_i\}_{i=1}^{n}$ una muestra aleatoria. el **rango**, denotado como rg es

$$rg = \text{máx}(\{x_i\}) - min(\{x_i\}) \tag{1.2}$$

La **varianza** se obtiene mediante la expresión

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 \tag{1.3}$$

La **desviación estándar** se es la raíz cuadrada de la varianza

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{1.4}$$

Una medida cero del rango o la desviación estándar implica un conjunto completamente homogéneo. Esto significa que todos los datos de la muestra son iguales.

Ejemplo 3. Tomando como referencia los datos de los ejemplos 1 y 2, el rango, varianza y desviación estándar para cada uno de estos son, respectivamente.

$$rg_1 = 10$$

$$s_1^2 = 11.98214$$

$$s_1 = 3.461523$$

$$rg_2 = 27$$

$$s_2^2 = 74.21429$$

$$s_2 = 8.614771$$

Como se esperaba, el punto atípico del segundo caso incrementa la variabilidad de los datos.

Cuartiles

Un cuartil permite dividir a una muestra aleatoria en cuatro grupos con la misma cantidad de datos. El procedimiento es similar al del cálculo de la mediana, con la diferencia que cada uno de los valores será parte de la muestra.

Cuartil

Sea $\{x_i\}_{i=1}^{n}$ una muestra aleatoria ordenada de manera creciente. El **cuartil** x_q, donde $q \in \{0.25, 0.5, 0.75\}$, es el valor mínimo en $\{x_i\}$ tal que la proporción de datos menores a x_q sea cuando menos q.

Ejemplo 4. Considere la muestra aleatoria ordenada de 25 datos {8, 9, 13, 16, 16, 23, 24, 27, 32, 34, 37, 37, 38, 39, 47, 48,50, 50, 50, 51, 52, 52, 53, 57, 59}. Las posiciones correspondientes para cada uno de los cuartiles son

$$(25)\ \frac{1}{4} = 6.25 \rightarrow 7$$

$$(25)\ \frac{1}{2} = 12.5 \rightarrow 13$$

$$(25)\ \frac{3}{4} = 18.75 \rightarrow 19$$

Los valores correspondientes a las posiciones determinan los *cuartiles.*

$$q_{0.25} = 24$$

$$q_{0.5} = 38$$

$$q_{0.75} = 50$$

Un cuartil es un caso particular de un concepto más general denominado *cuantil.* Otros casos son los *quintiles* (5 grupos), *deciles* (10 grupos) y *percentiles* (100 grupos).

Referencias:

Hoel, P. G. (1984). Elementary Statistics. John Wiley & Sons.

Mendenhall, W. (2015). Introducción a la Probabilidad y Estadística. Cengage Learning Editores.