

Análisis de datos con Python, módulo III - Ejercicio integrador final. Estadística

Cálculo de la media

Para calcular la media, simplemente se calcula la sumatoria de todos los datos registrados y se la divide por la cantidad de esos datos.

En este caso:

$$\bar{X} = \frac{\sum_{i=1}^{90} x_i}{90}; \bar{X} = \frac{65011}{90}, \text{ cálculo que arroja un valor para la media de } \bar{X} = 722.34$$

Cálculo de la mediana, M_a

Se define a la mediana (Ma) como el valor que ocupa la posición central de un conjunto de observaciones, debiendo estar los datos previamente ordenados en sentido ascendente (o descendente) de magnitud.

Procedimiento:

En este caso los datos se presentan de manera atómica, es decir, individualmente y no agrupados en rangos, por lo que el procedimiento para establecer la mediana será el siguiente:

1. Se ordenan los datos de modo creciente
2. Se ubica el dato que se encuentra en el punto medio de la colección de datos
3. Dado que en este caso la población tiene número par de componentes ($n = 90$), se calcula el promedio entre los dos valores centrales (x_{45} y x_{46}) y se establece ese valor como la mediana, M_a

$$M_a = \frac{740 + 741}{2} \Rightarrow M_a = 740.5$$

Cálculo de la moda, M_o

La moda o modo (Mo) es el dato individual que más se repite en la serie. M_o será el valor más típico, más recurrente, o bien, el que reúne la mayor frecuencia absoluta entre todos los valores (categorías) individuales observados en el conjunto de datos que se analiza.

Para este caso, el valor 846 es el más frecuente (se repite tres veces), por lo que se establece el valor de la moda,

$$M_o = 846$$

Análisis de los cinco números

El análisis de los cinco números consiste en identificar el **mínimo** (X_{\min}) el **cuartil 1** (Q_1), la **mediana** (M_a), el **cuartil 3** (Q_3) y el **máximo** (X_{\max})

Para el caso analizado, el valor *mínimo* corresponde al dato número 34 y es

$$X_{\min} = 456$$

El cuartil 1:

$$Q_1 = \frac{n}{4} ; Q_1 = \frac{90}{4} \Rightarrow Q_1 = 22.5$$

Como no se obtiene un número entero para la posición del cuartil, se calcula su valor como la suma de la parte entera del valor obtenido más la parte decimal multiplicada por la diferencia entre los valores de las posiciones entre las que se encuentra esa posición obtenida para el cuartil

En este caso :

$$Q_1 = 588 + 0.5 (589 - 588) \Rightarrow Q_1 = 588.5$$

La mediana ya fue calculada, $M_a = 740.5$

El cuartil 3:

$$Q_3 = \frac{3n}{4} ; Q_3 = \frac{270}{4} \Rightarrow Q_3 = 67.5$$

Nuevamente, al no existir la posición 67.5, se calcula el valor del cuartil como el valor en la posición 67 más el producto de la parte decimal (0.5) por la diferencia entre el valor en la posición 68 y el valor en la posición 67:

$$Q_3 = 841 + 0.5 (845 - 841) \Rightarrow Q_3 = 843$$

El valor máximo se obtiene directamente de la tabla, y es

$$X_{\max} = 982$$