

Proyecto 1: Estadística Descriptiva

Torres G. Lesmes
Tecnológico de Costa Rica
Correo: lesmestorres@estudiantec.cr

Barrios S. Xavier
Tecnológico de Costa Rica
Correo: @xbs20estudiantec.cr

Calderon G. Jose Andres
Tecnológico de Costa Rica
Correo: jandres2019@estudiantec.cr

Resumen—En este documento se presenta un análisis estadístico de la temperatura en la sala de una casa mediante el uso de sensores de temperatura. Se encontró que la temperatura está definida dentro de un rango específico y posee un valor frecuente que oscila entre los 19°C y 20 °C.

Index Terms—Análisis, Dispersión, Histograma, Temperatura, Tendencia central.

I. INTRODUCCIÓN

En el presente informe se muestra un análisis estadístico de la temperatura en la sala de una casa. Para lo cual se utilizaron sensores inalámbricos zigbee, los cuales recopilan datos cada 3.3 minutos y luego son promediados en intervalos de 10 minutos. Los datos fueron recopilados en un archivo de extensión .csv a los cuales se les aplicó un análisis utilizando el lenguaje de programación python. Esto con el objetivo de determinar cual fue el rango de temperatura que osciló la sala, determinar el tipo de clima asociado a este y exponer la temperatura mas recurrente en la sala.

II. MARCO TEÓRICO

II-A. Medidas de tendencia central

Las medidas de tendencia central son: el promedio, la moda, la mediana y los cuartiles.

El promedio es el punto central o punto medio del conjunto de datos, este puede ser muestral o poblacional [1].

El promedio muestral se calcula con la ecuación 1 donde n representa el número de individuos de la muestra.

$$x = \frac{x_1 + x_2 + x_n}{n} \quad (1)$$

Por otro lado, se tiene que la moda es el dato que se repite la mayor cantidad de veces en el espacio muestral [1]. Luego tenemos la mediana la cual es el valor central en un espacio muestral ordenado de menor a mayor [1]. Esta se calcula utilizando la ecuación 2 en caso de que la cantidad de datos n sea impar y si n es par se utiliza la ecuación 3.

$$x = X\left(\frac{n+1}{2}\right) \quad (2)$$

si n es impar

$$x = \frac{1}{2} \left[X\left(\frac{n}{2}\right) + X\left(\frac{n}{2} + 1\right) \right] \quad (3)$$

Los Cuartiles dividen el conjunto en 4 partes con aproximadamente igual numero de individuos. El cálculo de los cuartiles por método simple es:

Si n es divisible por 4 se utilizan las siguientes ecuaciones:

$$Q1 = x\left(\frac{n}{4}\right) \quad (4)$$

$$Q2 = \text{mediana} \quad (5)$$

$$Q3 = x\left(\frac{3n}{4}\right) \quad (6)$$

Si n no es divisible por 4 se procede con las siguientes ecuaciones y se redondea al entero mayor [1]:

$$Q1 = x\left(\frac{n}{4}\right) \quad (7)$$

$$Q2 = \text{mediana} \quad (8)$$

$$Q3 = x\left(\frac{3n}{4}\right) \quad (9)$$

II-B. Medidas de dispersión

Las medidas de dispersión proporcionan información describiendo el comportamiento de los datos a nivel de su variabilidad. Entre sus componentes están:

El Rango muestral (r), describe el rango de la muestra. Entre más grande sea este habrá mayor variabilidad de los datos [1]. Su fórmula es:

$$r = \text{Máx}(x_i) - \text{Min}(x_i) \quad (10)$$

El Rango intercuartílico (RIQ), describe el rango entre los cuartiles pero es menos sensible a los puntos extremos de la muestra [1] con respecto al Rango muestral. Su formula es:

$$RIQ = Q3 - Q1 \quad (11)$$

La Varianza (s^2) describe la variabilidad de los datos respecto al promedio [1]. Su formula es:

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1} \quad (12)$$

La Desviación estándar muestral (s), es la raíz cuadrada de la varianza e igual que esta describe la variabilidad de los datos [1].

$$\sqrt{s^2} = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} \quad (13)$$

El coeficiente de variación (cv) permite comparar la variabilidad de dos o más conjuntos de datos que difieren de manera considerable en la magnitud de las observaciones [1]. Su formula es:

$$cv = \frac{s}{\bar{x}} \quad (14)$$

II-C. Histograma

La distribución de frecuencias ofrece un resumen más compacto de los datos que el diagrama. También es útil presentar la distribución de frecuencias en forma gráfica, esta recibe el nombre de histograma. Para dibujar un histograma el eje horizontal se utiliza para presentar la escala de medición y para dibujar las fronteras de las clases. El eje vertical representa la escala de frecuencia. Los histogramas son más estables si el conjunto de datos es grande.

Existen tres tipos de sesgo [1]: el sesgo positivo que se da cuando los datos tienden a acomodarse en mayor medida del lado izquierdo de la media como se puede observar en la Fig.1, por otro lado se tiene el sesgo negativo en el cual los datos se acumulan del lado derecho de la media Fig.3 y por último se tiene el sesgo simétrico el cual la mayor concentración de datos se encuentra en la media y de forma relativamente equitativa tanto del lado derecho como el izquierdo de la media Fig2.

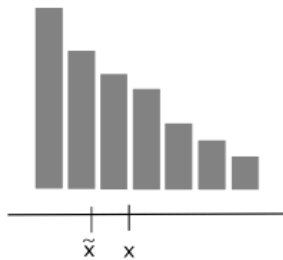


Figura 1: Sesgo positivo o a la derecha [1]

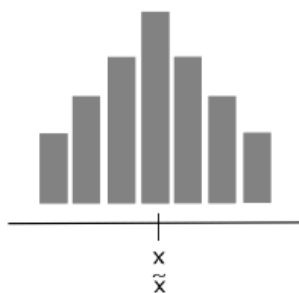


Figura 2: Sesgo simétrico [1]

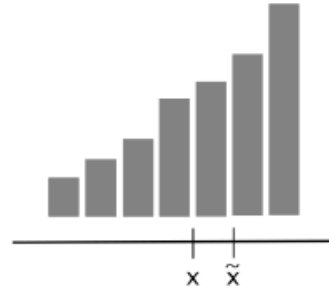


Figura 3: Sesgo negativo o a la izquierda [1]

III. ANÁLISIS DE RESULTADOS

La muestra y las medidas de localización están contenidas en la Tabla I.

Tabla I

Muestra y medidas de localización	
Parámetro	Valor
Muestra	19735
Media	20.34 °C
Cuartil 1	18.79 °C
Mediana o Cuartil 2	20.00 °C
Cuartil 3	21.50 °C
Moda	19.20 °C

Las medidas de dispersión se encuentran registradas en la Tabla II.

Tabla II

Medidas de dispersión	
Parámetro	Valor
Varianza	4.81 °C ²
Desviación estándar	2.19 °C
Coeficiente de variación	10.78 %
Rango muestral	13.76 °C
Rango intercuatílico	21.50 °C

Los valores mínimo y máximo se encuentran registrados en la Tabla III.

Tabla III

Máximo y mínimo de la muestra	
Parámetro	Valor (°C)
Máximo	16.21
Mínimo	29.85

Los datos de temperatura de la muestra, mostrada en la tabla I, son mediciones recopiladas y promediadas que se encuentran entre el rango de 16.21 °C y 29.85 °C, esto indica que la temperatura captada por los sensores en la sala durante el lapso de tiempo de medición no se encuentra en un clima de condición extrema [2]. Esta temperatura puede ser la temperatura ambiente o una adaptada por algún sistema de calefacción.

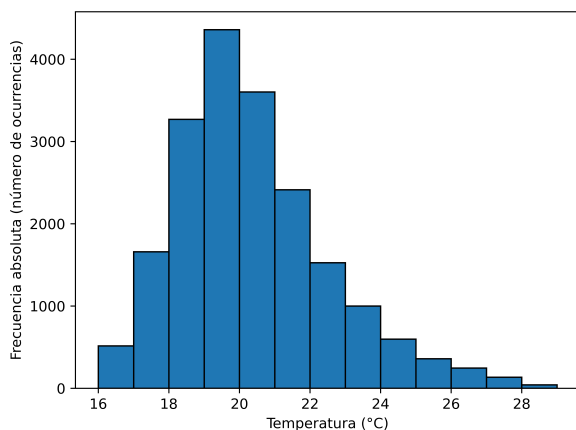


Figura 4: Resultados de monitoreo de la temperatura en la sala

En este caso se presenta un sesgo positivo, ya que se tiene una mayor frecuencia absoluta de los datos del lado izquierdo de la media como se puede apreciar en la figura 4. Según la figura 4 la temperatura de la casa osciló mayoritariamente entre los 19°C a 20°C. Esto significa que la sala de la casa estuvo mayormente fresca, ya que la temperatura ambiente oscila frecuentemente entre los 20°C y 25 °C y como dijimos anteriormente el histograma tiene un sesgo positivo lo cual significa que la mayoría de sus datos están a la izquierda.

IV. CONCLUSIONES

- Del histograma se puede apreciar que el dato de moda está dentro del rango más frecuente y se puede concluir que esta es una referente a la temperatura más probable a encontrar en dicha sala.
- La temperatura oscila entre los 16.21 °C y 29.85 °C, por lo cual, la casa no se encuentra en un clima de condición extrema.

REFERENCIAS

- [1] D. Montgomery y G. Runger. Probabilidad y estadística aplicadas a la ingeniería. Second. Limusa Wiley, 2006.
- [2] Meteorología y climatología de Navarra. [En línea]. <http://meteo.navarra.es/definiciones/koppen.cfm#:~:text=Divide%20los%20climas%20del%20mundo,la%20primera%20letra%20en%20may%C3%BAscula..> Marzo, 2023.