

# Estadística Descriptiva

José Miguel González Arias, *ITCR, 201271801*, Jose Andrés Sandoval Díaz, *ITCR, 2018319279*, Alonso Azofeifa Jiménez, *ITCR, 2019199079*

**Index Terms**—Estadística descriptiva, Media, Moda, Mediana, Histograma.

## I. ABSTRACT

EL siguiente documento se realizó para documentar el trabajo realizado en la obtención de métricas como la media, la mediana, la moda, desviación estándar y media, varianza, rangos máximos y mínimos, así como los diferentes Cuartiles en los que se divide, una serie de datos obtenidos de un estudio a lo largo del año 2016.

Para este caso en específico se tomó en consideración el valor de la humedad en una sala, en valores tomados en múltiples días. Estos datos se pueden encontrar en [1], donde se puede descargar el archivo csv con todos los datos del estudio. Además como se especificó anteriormente, para el presente estudio se tomó como referencia la fecha y hora en que se tomó el dato y adicionalmente el valor del porcentaje de humedad en la sala, dato que se encuentra en la columna RH\_2.

## II. OBJETIVOS

Calcular los valores de las medidas de variabilidad o dispersión con base a los datos suministrados (Humedad en una habitación).

Confeccionar un gráfico que represente el comportamiento de los datos de una forma estadística por medio de un histograma y diagrama de cajas. A la vez determinar cual de los 2 representa de mejor forma los datos.

Analizar estadísticamente los resultados obtenidos.

## III. CONCEPTOS

### III-A. Desviación estandar

La desviación estandar nos indica que tan dispersos están los datos unos de otros; entre mayor sea el valor de esta, más dispersos estarán los datos entre sí. [2]

### III-B. Desviación media

Consiste en hallar la media de los valores absolutos (esto es, sin tener en cuenta si son positivos o negativos) de las desviaciones de una serie de observaciones con respecto a su media o a su mediana. [3]

### III-C. Varianza

Esta mide que tan dispersos están los datos al rededor de la media, es de suma importancia porque nos dice lo confiables los que estamos calculando comparado a la realidad de los datos a como están agrupados. La varianza es igual a la desviación estándar elevada al cuadrado. [4]

### III-D. Media

Es la medida más usada para encontrar el promedio. Comúnmente, las personas utilizan la palabra "promedio" para referirse a la "media". [5]

### III-E. Moda

Consiste en encontrar el dato que más se repite en un conjunto de datos, como lo puede ser en un conjunto de mediciones de temperatura. [5]

### III-F. Mediana

Es el número que se encuentra en el medio de un conjunto ordenado de números. [6]

Parámetro	Valor
Cantidad de datos	19735
Desviación Estándar	4,069813
Desviación Media	3,178364
Varianza	16,562535
Rango Mínimo	20,463333
Rango Máximo	56,026666
Media	40,420420
Moda	40
Mediana	40,5
$Q_1$	37,9
$Q_2$	40,5
$Q_3$	43,26

Cuadro I

DATOS MÁS RELEVANTES DEL ESTUDIO

## IV. REPRESENTACIÓN DE LOS DATOS

En primera instancia para ubicar la frecuencia en que se repiten los datos, se escoge utilizar un histograma, para representar las frecuencias absolutas de los datos, respecto a la humedad que corresponde.

Para realizar este histograma de la figura 1, fue necesario encontrar el número de clases en que se podían dividir las diferentes humedades. Para obtener este valor se utilizó como base la ecuación 1, donde  $n$  representa la cantidad de datos total que se observa en el cuadro I.

Sustituyendo el valor de  $n$  como se observa en la ecuación 2, recordando siempre redondear al número entero mayor, se obtiene que el total de clases a utilizado fue 141.

$$\text{Número de clases} = \sqrt{n} \quad (1)$$

$$\text{Número de clases} = \sqrt{19735}$$

$$\text{Número de clases} = 140,48 \approx 141 \quad (2)$$

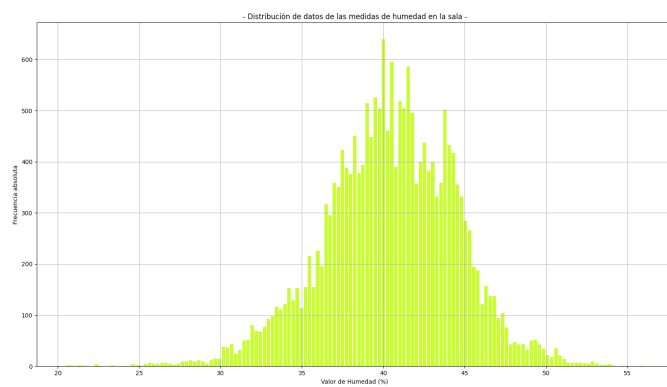


Figura 1. Gráfica Porcentaje de Humedad vs Frecuencia absoluta

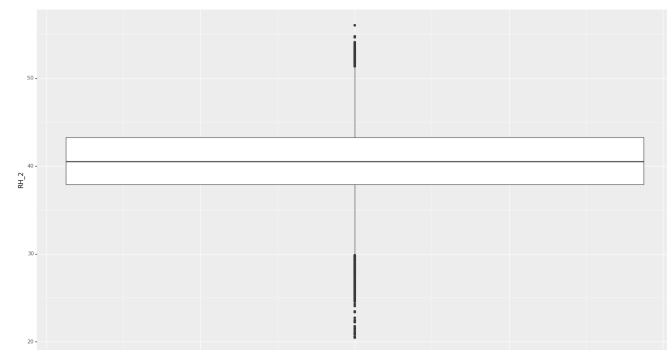


Figura 2. Diagrama de cajas de la humedad en la sala

Adicionalmente se grafica el promedio de todos los datos generados por día, para crear la figura 3. Para esto se graficó todo el año en estudio.

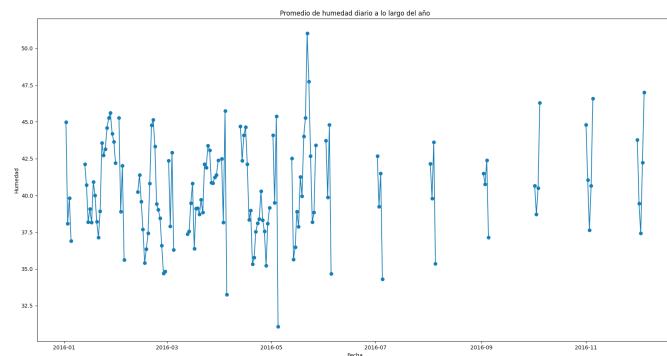


Figura 3. Distribución del promedio de la humedad por día a través del año

## V. ANÁLISIS DE LOS DATOS

En la figura 1, se presentan los datos agrupados por rangos donde se puede notar una gran frecuencia de datos cercano al punto de la gráfica marcado como 40 % de humedad. Esto era de esperar pues la moda de los datos ronda el valor de 40, entonces el pico de la gráfica se observa en este punto de humedad. Cabe destacar como la figura que representa este histograma es simétrica, simbolizando que la mayor cantidad

de repeticiones de un determinado porcentaje de humedad, se encuentra cercano al centro de la gráfica, generando así una disminución considerable de repeticiones en datos que se acercan a los rangos máximos y mínimos que se mostraron en el cuadro I.

Como se puede observar en la figura 3, los datos se agruparon por día, mostrando el promedio generado en cada uno.

Lo primero que cabe destacar de esta gráfica de la figura 3 es la cantidad de días en que no se tomaron datos. Se observa que posterior a Junio 2016, la cantidad de días en que se tomaron datos fueron muy pocos. Esta es una visualización muy importante debido a que se puede empezar un estudio de la razón por la cuál no se generaron estos datos y determinar si fue premeditado o bien si fue algún error en la toma de datos.

Adicionalmente se ve como la gran mayoría de puntos se concentran alrededor del valor de humedad 40. Esto se puede comparar con el cuadro I, donde se tiene que la moda y la media ronda el valor de 40. Y comparando también con el histograma de la figura 1, se ve donde la frecuencia absoluta genera el pico más alto de la gráfica justo en el valor 40.

Se realizó un digrama de cajas sólo para comparar las ventajas que tiene un histograma en el manejo de esta cantidad de datos. Se observa como hay gran cantidad de puntos atípicos, tanto en altos valores de humedad como en valores bajos de humedad. Para esta cantidad de datos

## VI. CONCLUSIÓN

Los resultados obtenidos en los gráficos con respecto a las medidas de variabilidad o dispersión se pudieron observar que la moda y la media no difirieron con respecto a lo visto en la figura 1, porque la distribución de los datos es la figura 1 formaron una gráfica simétrica. Esto se comprobó por medio de los valores obtenidos en el cuadro 1, como fue la desviación estándar, varianza y desviación media las cuales nos indicaron desde un principio la posible forma del histograma que íbamos a obtener, siendo este con una forma simétrica.

Por lo tanto, se logró comprobar la importancia que tienen las medidas de variabilidad o dispersión hacia el comportamiento de un conjunto de datos como lo fue las mediciones de humedad que se obtuvieron en esa sala en un tiempo definido.

Se concluye que el diagrama de cajas es más funcional en casos en donde se comparan datos de muestras distintas, como lo puede haber sido la comparación de las mediciones de humedad en distintas habitaciones de la casa.

## REFERENCIAS

- [1] Luis M. Candanedo, Veronique Feldheim y Dominique Deramaix. «Data driven prediction models of energy use of appliances in a low-energy house, Energy and Buildings». En: *ISSN 0378-7788* 140 (abr. de 2017), págs. 81-97.
- [2] LLC Minitab. *¿Qué es la desviación estándar?* Available at <https://support.minitab.com/es-mx/minitab/18/help-and-how-to/statistics/basic-statistics/supporting-topics/data-concepts/what-is-the-standard-deviation/>.

- [3] Palacios Martínez Ignacio (dir.) y col. *Diccionario electrónico de enseñanza y aprendizaje de lenguas*. Available at <https://www.dicenlen.eu/es/diccionario/entradas/desviacion-media>. 2019.
- [4] LLC Minitab. *¿Qué es la varianza?* Available at <https://support.minitab.com/es-mx/minitab/18/help-and-how-to/statistics/basic-statistics/supporting-topics/data-concepts/what-is-the-variance/>.
- [5] Shmoop Editorial Team. *Media, mediana, modo y rango*. Available at <https://www.shmoop.com/estadistica-basica-probabilidades/media-mediana-modo-rango.html>. Nov. de 2018.
- [6] Khan Academy. *Introducción a la estadística: media, mediana y moda*. Available at <https://es.khanacademy.org/math/cc-sixth-grade-math/cc-6th-data-statistics/mean-and-median/v/statistics-intro-mean-median-and-mode>.