| | **INSTITUCIÓN EDUCATIVA "EL RECUERDO"**<br>Resolución de Aprobación de Carácter Oficial No. 0143 de 2017 en los niveles de Preescolar, Básica y Media Académica<br>DANE. 123001800064 NIT. 901048820-9 | Fecha |
|---|---|---|

| | |
|---|---|
| **Guía de trabajo del área:** Estadística – Guía 4 | **Grado:** 9A - 9B |
| **Nombre del docente 9A:** Ureliano Peñata **email:** upenataieelrecuerdo@gmail.com **Celular:** 3135276620 | |
| **Nombre del docente 9B:** Rosa Cano **email:** rcanoieelrecuerdo@gmail.com **Celular:** 3105679770 | |

| TEMAS Y/O SABER | DBA (APRENDIZAJES) |
|---|---|
| ✓ **Medidas de dispersión**<br>✓ **Diagramas de caja y bigotes** | **DBA 10: Propone un diseño estadístico adecuado para resolver una pregunta que indaga por la comparación sobre las distribuciones de dos grupos de datos, para lo cual usa comprensivamente diagramas de caja, medidas de tendencia central, de variación y de localización.** |

**********Nota: Transcribe todo el contenido de la guía en tu cuaderno**********

## SABERES PREVIOS

Lee y comprende la relación entre las medidas de tendencia central y las medidas de dispersión: Cuando solo se utilizan las medidas de tendencia central para el análisis de la distribución de un conjunto de datos, se está ignorando gran parte de estos, por esta razón es recomendable que se incluya el estudio de las medidas de dispersión, para determinar cómo están distribuidos los datos entorno a los valores centrales. Qué tan alejados o que tan cerca están de la media.
Las medidas de dispersión toman gran importancia en el análisis de datos ya que dan solución a algunas limitantes que presentan las medidas de tendencia central, pues estás por sí solas no manifiestan si las diferencias entre los datos varían o no regularmente y si son grandes o pequeñas, entendiéndose la variación, como el grado en que los datos numéricos tienden a distribuirse alrededor de un valor central.
Se puede mostrar que las medidas de dispersión sirven para identificar si una medida central es adecuada para representar la población de datos e indicar la relación de un dato con los otros, y si son útiles al comparar distribuciones.

## APRENDAMOS

### TEMA 1: MEDIDAS DE DISPERSION

Las calificaciones de un grupo de diez estudiantes en un examen de estadística son las siguientes:

| | | | | |
|---|---|---|---|---|
| 56 | 58 | 67 | 69 | 75 |
| 77 | 77 | 82 | 84 | 95 |

¿Entre qué valores varían los datos?
Los datos varían entre 56 y 95, la menor y la mayor de las calificaciones, respectivamente, obtenidas por los estduiantes. A la diferencia entre estos valores se le llama **rango.**
Para comprender el comportamiento de los datos de un conjunto, se puede determinar su **dispersión** o **variabilidad** a través del rango, la varianza y la deviación típica.

Las medidas de dispersión más comunes son el rango, la varianza y la desviación estándar.

**RANGO:** El **rango** de una distribución es la diferencia entre el mayor valor y el menor valor de la variable estadística. También se llama **recorrido.**

**VARIANZA:** Antes de estudiar el concepto de varianza, es necesario definir la **desviación respecto a la media.**

Se conoce como **desviación respecto a la media,** $d_i$, a la diferencia entre cada valor de la variable estadística, $x_i$, y la media aritmética, $\overline{x}$. Es decir:

$$d_i = x_i - \overline{x}$$

**Ejemplo 1**

La media aritmética de los datos de la Tabla 4.25 es $\overline{x} = 25{,}1$ y las desviaciones respecto a la media se muestran en la Tabla 4.26.

| Número de libros $(x_i)$ | 12 | 17 | 21 | 27 | 35 | 37 | 49 |
|---|---|---|---|---|---|---|---|
| $d_i = x_i - \overline{x}$ | −13,1 | −8,1 | −4,1 | 1,9 | 9,9 | 11,9 | 23,9 |

Tabla 4.26

| Número de libros $(x_i)$ | $f_i$ |
|---|---|
| 12 | 5 |
| 17 | 3 |
| 21 | 6 |
| 27 | 8 |
| 35 | 4 |
| 37 | 3 |
| 49 | 1 |

Tabla 4.25

La **varianza** $s^2$ de una variable estadística $x$ es la media aritmética de los cuadrados de las desviaciones respecto a la media. Para datos agrupados es:

$$s^2 = \frac{f_1(x_1 - \overline{x})^2 + f_2(x_2 - \overline{x})^2 + \ldots + f_n(x_n - \overline{x})^2}{f_1 + f_2 + \ldots + f_n} = \frac{\sum f_i(x_i - \overline{x})^2}{N}$$

## DESVIACIÓN TÍPICA

La **desviación típica s** es la raíz cuadrada positiva de la varianza.

**Ejemplo 2**

La varianza y la desviación típica de los datos de la Tabla 4.25, son:

$$= \frac{(-13,1)^2 \cdot 5 + (-8,1)^2 \cdot 3 + (-4,1)^2 \cdot 6 + (1,9)^2 \cdot 8 + (9,9)^2 \cdot 4 + (11,9)^2 \cdot 3 + (23,9)^2 \cdot 1}{5+3+6+8+4+3+1}$$

$$s^2 = \frac{2\,572}{30} = 85,76 \longrightarrow s = 9,26$$

## COEFICIENTE DE VARIACIÓN

El **coeficiente de variación *CV*** sirve para comparar **la** dispersión de distribuciones que tienen diferentes medias y distintas desviaciones típicas.

$$CV = \frac{s}{\bar{x}}$$

|  | $\bar{x}$ | $s$ |
|---|---|---|
| **Sara** | 8,5 | 1,3 |
| **Lucía** | 7,5 | 1,2 |

Tabla 4.27

**Ejemplo 3**

En la Tabla 4.27, se muestran la media y la desviación típica de las notas de Sara y Lucía. El coeficiente de variación de cada una es:

$$CV_{Sara} = \frac{s}{\bar{x}} = \frac{1,3}{8,5} = 0,15 \qquad CV_{Lucía} = \frac{s}{\bar{x}} = \frac{1,2}{7,5} = 0,16$$

Aunque la desviación típica de Sara es mayor, las calificaciones de Lucía son más dispersas porque es mayor el coeficiente de variación.

Ejemplo: Analiza el siguiente ejercicio resuelto: Realiza las operaciones en tu cuaderno para comprobar los resultados obtenidos.

Luego de agrupar en intervalos de edades a los usuarios de un café internet, se completó la Tabla 6.19.

|  | $x_i$ | $f_i$ | $x_i \cdot f_i$ | $\lvert x_i - \bar{x}\rvert$ | $(x_i - \bar{x})^2$ | $\lvert x_i - \bar{x}\rvert f_i$ | $(x_i - \bar{x})^2 f_i$ |
|---|---|---|---|---|---|---|---|
| [20, 25) | 22,5 | 5 | 112,5 | 6,5 | 42,25 | 32,5 | 211,25 |
| [25, 30) | 27,5 | 3 | 82,5 | 1,5 | 2,25 | 4,5 | 6,75 |
| [30, 35) | 32,5 | 2 | 65 | 3,5 | 12,25 | 7 | 24,5 |
| [35, 40) | 37,5 | 2 | 75 | 8,5 | 72,25 | 17 | 144,5 |
| [40, 45) | 42,5 | 1 | 42,5 | 13,5 | 182,25 | 13,5 | 182,25 |
|  |  | 13 | 377,5 |  |  | 74,5 | 1 569,25 |

$$\bar{x} = \frac{377,5}{13} \approx 29 \qquad D_{\bar{x}} = \frac{\sum_{i=1}^{n} \lvert x_i - \bar{x}\rvert f_i}{N} = \frac{74,5}{13} \approx 5,7$$

$$s^2 = \frac{\sum_{i=1}^{n} f_i (x_i - \bar{x})^2}{N} = \frac{1569,25}{13} \approx 120,7 \qquad s = \sqrt{120,7} \approx 11$$

$$CV = \frac{s}{\bar{x}} = \frac{11}{29} \approx 0,38$$

En esta situación el promedio se ve afectado por datos extremos. Así que, es necesario recurrir al cálculo de las medidas de dispersión.

La desviación típica indica que en promedio los valores de la variable se desvían o alejan aproximadamente 11 años de la media. Como 29 − 11 = 18 y 29 + 11 = 40, se puede concluir que las edades inferiores a 18 y superiores a 40 son los datos más lejanos de la media.

Si se expresa el coeficiente de variación como un porcentaje, se puede decir que las edades presentan una dispersión del 38%.

**ACTIVIDAD DE APRENDIZAJE 1**

### Razonamiento

1. Los porcentajes de uso del cinturón de seguridad en dos ciudades *A* y *B* durante cuatro días se muestran en la Tabla 4.28.

| | | | | |
|---|---|---|---|---|
| **A** | 87 | 78 | 67 | 82 |
| **B** | 60 | 95 | 92 | 47 |

Tabla 4.28

Calcula el coeficiente de variación en cada ciudad e interpreta el resultado.

### Resolución de problemas

2. En un colegio hay matriculados la siguiente cantidad de estudiantes:

- En grado sexto hay 112 estudiantes.
- En grado séptimo hay 123 estudiantes.
- En grado octavo hay 130 estudiantes.
- En grado noveno hay 110 estudiantes.
- En grado décimo hay 150 estudiantes.
- En grado once hay 146 estudiantes.

a. Elabora una tabla que contenga los anteriores datos.

b. Halla el rango.

c. Calcula la varianza y la desviación típica.

### Evaluación del aprendizaje

i. Las figuras 4.19 y 4.20 muestran los puntos anotados por dos jugadores de baloncesto.

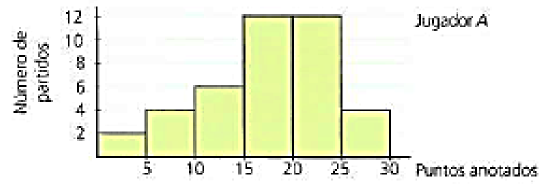



Figura 4.19

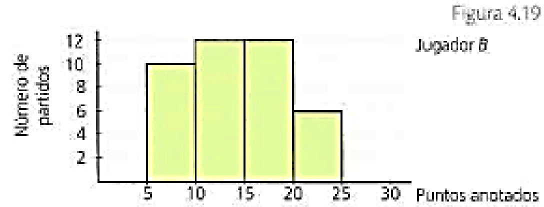



Figura 4.20

¿Cuál de ellos alcanza la mejor media anotadora? ¿Quién es más regular en su posición?

ii. En la Tabla 4.29, se registró el número de goles que hicieron dos equipos de fútbol *A* y *B*.

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **A** | 25 | 24 | 27 | 24 | 26 | 25 | 27 | 24 |
| **B** | 28 | 30 | 21 | 22 | 27 | 20 | 28 | 30 |

Tabla 4.29

Calcula el promedio de goles de cada equipo. ¿Cuál de ellos es más regular en su desempeño?

### Medidas de dispersión

#### Ejercitación

2. Calcula el rango, la varianza y la desviación típica de los datos presentados en la Tabla 4.45.

| Puntaje | Número de personas |
|---|---|
| [5, 9] | 6 |
| [9, 13] | 9 |
| [13, 17] | 7 |
| [17, 21] | 15 |
| [21, 25] | 12 |

Tabla 4.45

3. Observa los datos de la Tabla 4,46. Luego, halla el coeficiente de variación. Interpreta los resultados.

| | | | | | | |
|---|---|---|---|---|---|---|
| **X** | 41 | 29 | 35 | 24 | 25 | 19 |
| **Y** | 41 | 45 | 56 | 49 | 38 | 48 |

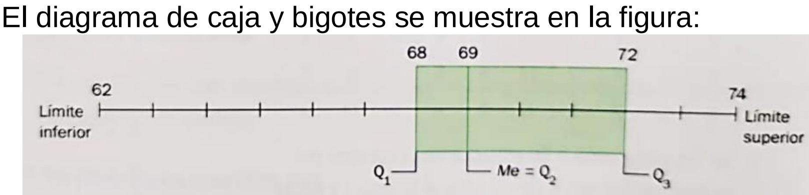
Tabla 4.46

## TEMA 2: DIAGRAMAS DE CAJAS Y BIGOTES

Un **diagrama de cajas y bigotes** es una respresentación gráfica que permite estudiar la **simetría** o **asimetría** de una distribución. En estos diagramas se reflejan cinco parámetros: los límites inferior y superior, y los cuartiles.
El número de pulsaciones por minuto de 50 estudiantes se registró en la tabla:

| Número de pulsaciones por minuto | $f_i$ | $F_i$ |
|---|---|---|
| 62 | 6 | 6 |
| 65 | 5 | 11 |
| 68 | 4 | 15 |
| 69 | 11 | 26 |
| 72 | 12 | 38 |
| 74 | 12 | 50 |

Representa los datos en un diagrama de caja, y estudia la simetría y concentración de los mismos.
Para representar los datos en un diagrama de caja, se calculan los siguientes parámetros: límite inferior ($L_i$), $Q_1$, $Me$, $Q_3$ y el límite superior ($L_s$).
$L_i = 62$ $Q_1 = 68$ $Me = Q_2 = 69$ $Q_3 = 72$ $L_s = 74$
Luego, se dibuja un segmento que tiene como extremos $L_i$ y $L_s$. Sobre él se marcan $Q_1$ hasta $Q_3$. Las líneas que sobresalen de la caja se llaman **bigotes.**
El diagrama de caja y bigotes se muestra en la figura:



Al observar el diagrama se puede concluir que:

- El bigote de la derecha ($Q_3$, $L_s$) es más corto que el de la izquierda ($L_i$, $Q1$), lo que indica que las pulsaciones se encuentran más concentradas en la cuarta parte más alta de los datos.
- Las pulsaciones que se hallan entre el 50% y el 75% de los datos ($Q_2$, $Q_3$) están más dispersas que las que están entre el 25% y el 50% ($Q_1$, $Q_2$).

**Ejemplo 1**

Para determinar si una distribución de frecuencias es simétrica o asimétrica es necesario calcular su media. La media de la distribución de la Tabla 4.22 es 69,6. Como el valor de la media no está en el centro de la caja se puede concluir que la distribución no es simétrica.

El diagrama de cajas da información sobre cómo se encuentran concentrados los datos. Para saber si hay algún valor más alejado o valor atípico que distorsione el estudio de los diferentes parámetros existe el siguiente criterio:

$$x \text{ es un valor atípico si } \begin{cases} x > Q_3 + 1{,}5[Q_3 - Q_1] \\ \text{o} \\ x < Q_1 - 1{,}5[Q_3 - Q_1] \end{cases}$$

**Ejemplo 2**

Los siguientes datos corresponden al número de fallas de asistencia al colegio de 30 estudiantes durante un mes.

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 3 | 2 | 4 | 1 | 8 | 1 | 3 | 2 | 2 | 3 | 1 | 3 | 1 |
| 5 | 2 | 6 | 2 | 1 | 2 | 1 | 2 | 4 | 3 | 6 | 1 | 4 | 1 | 3 |

$L_i = 1$ $\quad Q_1 = 1$ $\quad Me = Q_2 = 2$ $\quad Q_3 = 3$ $\quad L_s = 8$

Los valores atípicos son:

$x < Q_1 - 1{,}5[Q_3 - Q_1]$ $\qquad x > Q_3 + 1{,}5[Q_3 - Q_1]$

$x < 1 - 1{,}5(2)$ $\qquad x > 3 + 1{,}5(2)$

$x < -2$ $\qquad x > 6$

Por lo tanto, 8 es el único valor atípico. En el diagrama de caja de la Figura 4.15 se ubicó un asterisco que señala la ubicación del dato atípico.

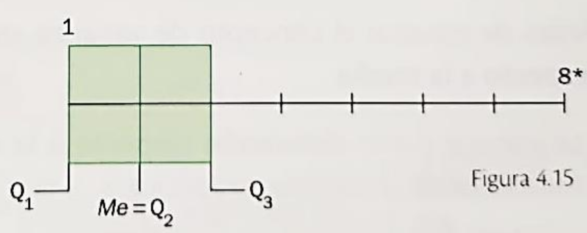



Figura 4.15

**ACTIVIDAD DE APRENDIZAJE 2**

## Resolución de problemas

1. Representa el diagrama de caja y bigotes de cada distribución. Ten en cuenta que para datos agrupados los límites inferior y superior corresponden a las marcas de clase del primer y último intervalo.

a. Edades de los estudiantes que acuden a una biblioteca en un día.

| Edad (años) | [6, 8) | [8, 10) | [10, 12) | [12, 14) |
|---|---|---|---|---|
| Número de estudiantes | 5 | 12 | 14 | 13 |

Tabla 4.23

b. Resultados de unos estudiantes en la prueba de salto de longitud.

| Salto (m) | [2; 2,5) | [2,5; 3) | [3; 3,5) | [12, 14) |
|---|---|---|---|---|
| Número de estudiantes | 10 | 9 | 11 | 7 |

Tabla 4.24

## Evaluación del aprendizaje

En las figuras 4.16 a 4.18 se observan los diagramas de caja y bigotes de tres variables.


Figura 4.16


Figura 4.17

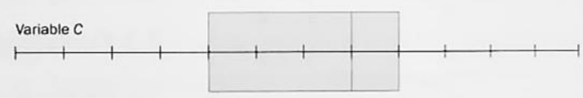



Figura 4.18

Determina qué se puede decir de cada uno de ellos en cuanto a los siguientes aspectos.

a. Concentración de los datos.

b. Dispersión de los datos.

c. Simetría.

*"ASESORIA: si tiene alguna duda o no entiende algo sobre esta guía, comuníquese con el número que aparece en la parte de arriba*