Lección

Debes saber...

- El promedio o media aritmética de un conjunto de datos $\{x_1, x_2, x_3, \ldots, x_n\}$ corresponde a

$$\overline{x} = \frac{x_1 + x_2 + x_3 + \ldots + x_n}{n}$$

- Puede expresarse también en notación de sumatoria:

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Medidas de dispersión de datos

Marcela observa sus notas semestrales en algunas asignaturas, y el promedio entre ellas, para hacer una evaluación respecto a su rendimiento en el semestre. Lo primero que le interesa saber es qué tan parecidas son sus notas entre sí. Para ello:

Matemática:	5,2
Lenguaje:	5,7
Ciencias sociales:	6,2
Ciencias naturales:	5,4
Inglés:	4,5
Promedio:	**5,4**

Paso 1 **Identifica la mayor y la menor de sus notas.**

Ciencias sociales: 6,2

Inglés: 4,5

Paso 2 **Resta estos valores:**

$$R = X_{max} - X_{min} = 6{,}2 - 4{,}5 = 1{,}7$$

Ayuda

Podemos llamar $X = \{x_1, x_2, \ldots, x_n\}$ al conjunto de notas de Marcela. Con ello, su mejor nota se simboliza como X_{max}, y la peor, X_{min}.

Ya que la escala de notas es de 1 a 7, $7 - 1 = 6$ es la mayor diferencia que podría haber en un conjunto de notas, se puede concluir que esta diferencia no es tan grande, es decir, que las notas de Marcela son relativamente parecidas entre sí. En general, a la diferencia entre el mayor valor de una muestra y el menor se le llama **rango**.

Ahora Marcela quiere averiguar si su rendimiento semestral es cercano al promedio. Para ello compara cada una de sus notas con el promedio obtenido.

Calcula el promedio de la diferencia entre las notas y el promedio.

$$\frac{(5{,}2-5{,}4)+(5{,}7-5{,}4)+(6{,}2-5{,}4)+(5{,}4-5{,}4)+(4{,}5-5{,}4)}{5}$$

$$= \frac{-0{,}2+0{,}3+0{,}8+0+-0{,}9}{5}$$

$$= \frac{0}{5} = 0$$

Se puede demostrar que cualquiera sea la cantidad de datos y el promedio este resultado será cero, por lo que es preciso tomar otras medidas. Una opción es la **desviación media** (D_m), que toma los valores absolutos de estas diferencias:

$$D_m = \frac{|5{,}2-5{,}4| + |5{,}7-5{,}4| + |6{,}2-5{,}4| + |5{,}4-5{,}4| + |4{,}5-5{,}4|}{5}$$

$$= \frac{0{,}2+0{,}3+0{,}8+0+0{,}9}{5}$$

$$= \frac{2{,}2}{5} = 0{,}44$$

En general la desviación media,

$$D_m = \frac{|x_1 - \overline{x}| + |x_2 - \overline{x}| + \ldots + |x_n - \overline{x}|}{n} \longrightarrow D_{\overline{x}} = \frac{\sum_{i=1}^{n} |(x_i - \overline{x})|}{n}$$

Observa que...

Para datos agrupados,

$$D_m = \frac{\sum_{i=1}^{N} |(x_{mci} - \overline{x})| \cdot f_i}{n}$$

x_{mci} es la marca de clase del intervalo i.

$\overline{x}$ es la media aritmética de la variable.

f_i es la frecuencia absoluta del intervalo i.

n es el número total de datos.

N es la cantidad de intervalos.

Marcela calcula ahora la desviación estándar que mide cuánto se separan los datos. Para esto debe calcular la raíz cuadrada de la varianza, que es la media de las diferencias con la media elevadas al cuadrado.

Paso 1 **Calcula el promedio de los cuadrados de las diferencias entre cada nota y el promedio, obteniendo así la varianza (σ^2) que es el cuadrado de la desviación estándar.**

$$\sigma^2 = \frac{(5{,}2-5{,}4)^2 + (5{,}7-5{,}4)^2 + (6{,}2-5{,}4)^2 + (5{,}4-5{,}4)^2 + (4{,}5-4{,}5)^2}{5}$$

$$\sigma^2 = \frac{(-0{,}2)^2 + 0{,}3^2 + 0{,}8^2 + 0^2 + (-0{,}9)^2}{5}$$

$$\sigma^2 = \frac{0{,}04 + 0{,}09 + 0{,}64 + 0 + 0{,}81}{5} = 0{,}316$$

Así la varianza es 0,316.

Paso 2 **Calcula la raíz cuadrada del valor anterior, y obtiene la desviación estandar (σ).**

$$\sigma = \sqrt{0{,}316} \approx 0{,}562$$

Mientras más parecidas sean las notas al promedio, menores serán sus diferencias con este, haciendo que la varianza y la desviación estándar sean menores.

En general la varianza,

$$\sigma^2 = \frac{(x_1-\bar{x})^2 + (x_2-\bar{x})^2 + \ldots (x_n-\bar{x})^2}{n} \longrightarrow \sigma^2 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n}$$

En general la desviación estándar,

$$\sigma = \sqrt{\frac{(x_1-\bar{x}) + (x_2-\bar{x}) + \ldots (x_n-\bar{x})}{n}} \longrightarrow \sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n}}$$

El porqué usar la **desviación estándar** se debe a otras consideraciones que verás en cursos posteriores, pero en la lección siguiente podrás apreciar su uso para comparar conjuntos de datos y analizar en cuál los datos son más parecidos entre sí (es decir, el conjunto es más **homogéneo**) o si se diferencian más (conjunto **heterogéneo**).

En resumen

Se llama **dispersión** de un conjunto $X = \{x_1, x_2, x_3, \ldots x_n\}$ a la variabilidad que existe entre los datos y las medidas de tendencia central. Generalmente estas medidas tienen que ver con el grado de dispersión que tiene el conjunto de datos con **respecto a su media**. Mientras más dispersos sean, más **heterogéneo** es el conjunto, y si es menos disperso es más **homogéneo**. La dispersión se puede cuantificar utilizando el **rango (R)**, la **desviación media** (D_m), la **desviación estándar ($\sigma(x)$)** y la **varianza (var(x)** o $\sigma^2(x)$).

Observa que...

La varianza se encuentra en unidades "al cuadrado", mientras que la desviación estándar está en la misma unidad que los datos (notas), por lo que esta última nos puede dar una idea más cercana de lo disperso que es el conjunto.

Observa que...

Para datos agrupados,

$$\sigma^2 = \frac{\sum_{i=1}^{N}(x_{mci}-\bar{x})^2 \cdot f_i}{n}$$

Razona y comenta

- Si las notas de Marcela hubieran sido 2; 2; 4; 6; 6, ¿cuál habría sido el rango de sus notas? ¿Y su varianza y desviación estándar? Estos valores, ¿habrían sido mayores o menores que los anteriormente obtenidos?
- En el caso anterior, las notas de Marcela, ¿son más o menos dispersas? El promedio que obtuvo, ¿es más o menos parecido a sus notas que en el ejemplo anterior?

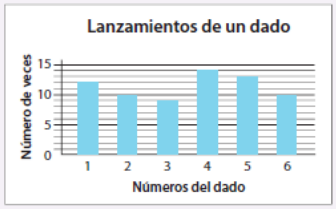
Practiquemos lo aprendido

Repaso

1. El siguiente conjunto de datos refleja el número de páginas de los 24 libros que integran una colección. Haz una tabla de frecuencia absoluta y **calcula** el promedio de páginas.

90	96	97	96	97	91	96	98
99	98	91	96	99	98	92	96
94	98	94	96	97	95	97	95

2. El siguiente gráfico de barras expresa las puntuaciones obtenidas al lanzar 68 veces un dado:



Calcula el promedio obtenido.

Práctica guiada

3. **Calcula** el rango, la varianza y la desviación estándar de los siguientes conjuntos de datos, siguiendo los pasos utilizados en la página anterior.

Ejemplo: 20; 20; 5; 8; 11

Rango = 20 – 5 = 15

Varianza

$$\overline{X}=\frac{20+20+5+8+11}{5}=\frac{64}{5}=12,8$$

$(20 - 12,8)^2 = 7,2^2 = 51,84$

$(20 - 12,8)^2 = 7,2^2 = 51,84$

$(5 - 12,8)^2 = (-7,8)^2 = 60,84$

$(8 - 12,8)^2 = (4,8)^2 = 23,04$

$(11 - 12,8)^2 = (-1,8)^2 = 8,64$

$$\sigma^2 = \frac{51,84+51,84+60,84+23,04+8,64}{5}$$

$$= 39,24$$

Desviación estándar

$$\sigma = \sqrt{39,24}$$

$$\approx 6,26$$

a) 5; 14; 15; 13; 1

b) 6; 2; 13; 1; 12

c) 11; 6; 14; 2; 7; 11; 18; 19; 17; 6

d) 5; 14; 18; 19; 14; 19; 13; 5; 20; 12

4. **Analiza** la siguiente situación. Luego resuelve.

El largo en pulgadas de 10 clavos de una bolsa es el siguiente:

3,2 - 3 - 3,5 - 3,5 - 3,5 - X - 3,2 - 3,1 - 3,2 - 3

Si se sabe que el rango de las medidas es 0,6; ¿qué valores puede tomar X?

Para determinarlo:

- Identificamos los valores mínimo y máximo, y calculamos el rango, sin considerar X.

Mín = 3 Max = 3,5

Máx – Mín = 0,5

- Para que el rango sea 0,6, necesitamos aumentar el valor máximo en 0,1, o disminuir el mínimo de igual manera. Por lo tanto, los posibles valores de X son:

X = 2,9 X = 3,6

Determina valores posibles de **x** para que se cumpla la condición dada.

a) 19,4; 15,5; x; 2,7; 8,7, si el rango es 17,6.

b) 10,6; 6,8; 13,4; 4,2; x; 6,3, si el rango es 9,3.

c) 18,9; x; 0,9; 5,7; 5,1; 5,8, si el rango es 19.

Aplica

Resuelve los siguientes problemas.

5. Manuel ha sacado la misma nota en todas las pruebas de Lenguaje del semestre. ¿Cuál es la desviación estándar de sus notas?

6. Sara se propuso mejorar sus notas durante el segundo semestre, y logró subir un punto todas ellas. ¿Qué ocurrió con la dispersión de sus notas? Justifica y pon un ejemplo.

7. Una central telefónica registra las siguientes cantidades de llamadas, de lunes a viernes:

Día	Llamadas
Lunes	4
Martes	9
Miércoles	7
Jueves	12
Viernes	8

a) Durante la última semana del mes, el número de llamadas se duplicó cada día. ¿Qué sucede con el rango? ¿y con la varianza?

b) Debido a una promoción, el número de llamadas a la central se quintuplicó cada día. ¿Qué sucede con el rango? ¿y con la varianza?

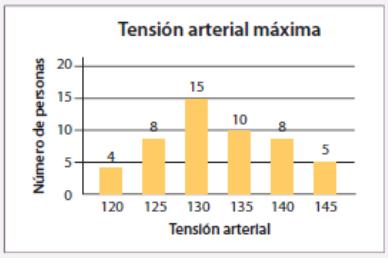
c) Considera tus respuestas anteriores para responder: ¿qué sucede con el rango y la varianza si todos los valores se multiplican por un número dado **x**?

8. **En un análisis de la sangre de unos pacientes se obtuvieron, en miles por centímetro cúbico, las siguientes cantidades de leucocitos:**

9,5 12 11,8 14,5 10 17,5 13,5

Halla el rango, la varianza y la desviación media.

9. **Se midió la tensión arterial máxima (presión que ejerce la sangre sobre la pared de los vasos sanguíneos) de 50 personas. El resultado viene reflejado por el siguiente gráfico de barras:**



Calcula el promedio y la desviación estándar de la tensión arterial de las 50 personas.

10. **Analiza** la siguiente situación.

Camilo está analizando nuevas formas de calcular la varianza de un conjunto. Para ello calcula la varianza de x = {2, 3, 5, 8} utilizando la siguiente tabla, sabiendo que el promedio es 4,5.

$(x_i - \overline{x})^2 =$	$x_i^2 - 2x_i\overline{x} + \overline{x}^2$
$(2-4{,}5)^2 =$	$2^2 - 2(2 \cdot 4{,}5) + 4{,}5^2$
$(3-4{,}5)^2 =$	$3^2 - 2(3 \cdot 4{,}5) + 4{,}5^2$
$(5-4{,}5)^2 =$	$5^2 - 2(5 \cdot 4{,}5) + 4{,}5^2$
$(8-4{,}5)^2 =$	$8^2 - 2(8 \cdot 4{,}5) + 4{,}5^2$

Luego, escribe lo siguiente:

$$\sigma^2 = \frac{2^2+3^2+5^2+8^2}{4} - \frac{2 \cdot 4{,}5(2+3+5+8)}{4} + \frac{4 \cdot 4{,}5^2}{4}$$

a) Desarrolla los términos de la expresión anotada por Camilo. ¿Qué obtienes en cada caso?

b) Explica algebraicamente cómo se calcula cada uno de los términos.

c) Utiliza la expresión obtenida por Camilo y tus resultados de las preguntas a) y b) para demostrar que, si $\overline{x^2}$ corresponde al promedio de los cuadrados de los valores y $\overline{x}$ es el promedio de ellos, entonces $\sigma^2(x) = \overline{x^2} - (\overline{x})^2$.

11. **Desafío:** **dado un conjunto $x = \{x_1, x_2, \ldots, x_n\}$, cuyo promedio es $\overline{x}$, demuestra las relaciones:**

a) Si $x + a = \{x_1 + a, x_2 + a, \ldots, x_n + a\}$, entonces $\sigma^2(x + a) = \sigma^2(x)$.

b) Si $ax = \{ax_1, ax_2, \ldots, ax_n\}$, entonces $\sigma^2(ax) = a^2\sigma^2(x)$.

c) Si $ax = \{ax_1, ax_2, \ldots, ax_n\}$, entonces $R(ax) = a \cdot R(x)$.

d) $\sigma^2(x) = \dfrac{x_1^2 + x_2^2 + \ldots + x_n^2}{n} - \overline{x}^2$

12. **Desafío:** **David tiene la siguiente información respecto de las notas de su curso en una prueba**

Nota	Cantidad de alumnos
Entre 1 y 1,9	4
Entre 2 y 2,9	8
Entre 3 y 3,9	9
Entre 4 y 4,9	11
Entre 5 y 5,9	7
Entre 6 y 7	6

Calcula la varianza de los datos.

13. **Conexiones:** **averigua de qué manera se calcula el puntaje de la PSU. ¿Cuál es la importancia de la desviación estándar?**

Reflexiona

- ¿Por qué es importante determinar la dispersión de un conjunto de datos?
- Si tus notas tienen una dispersión muy alta, ¿qué significa eso respecto a tu rendimiento académico? Explica.