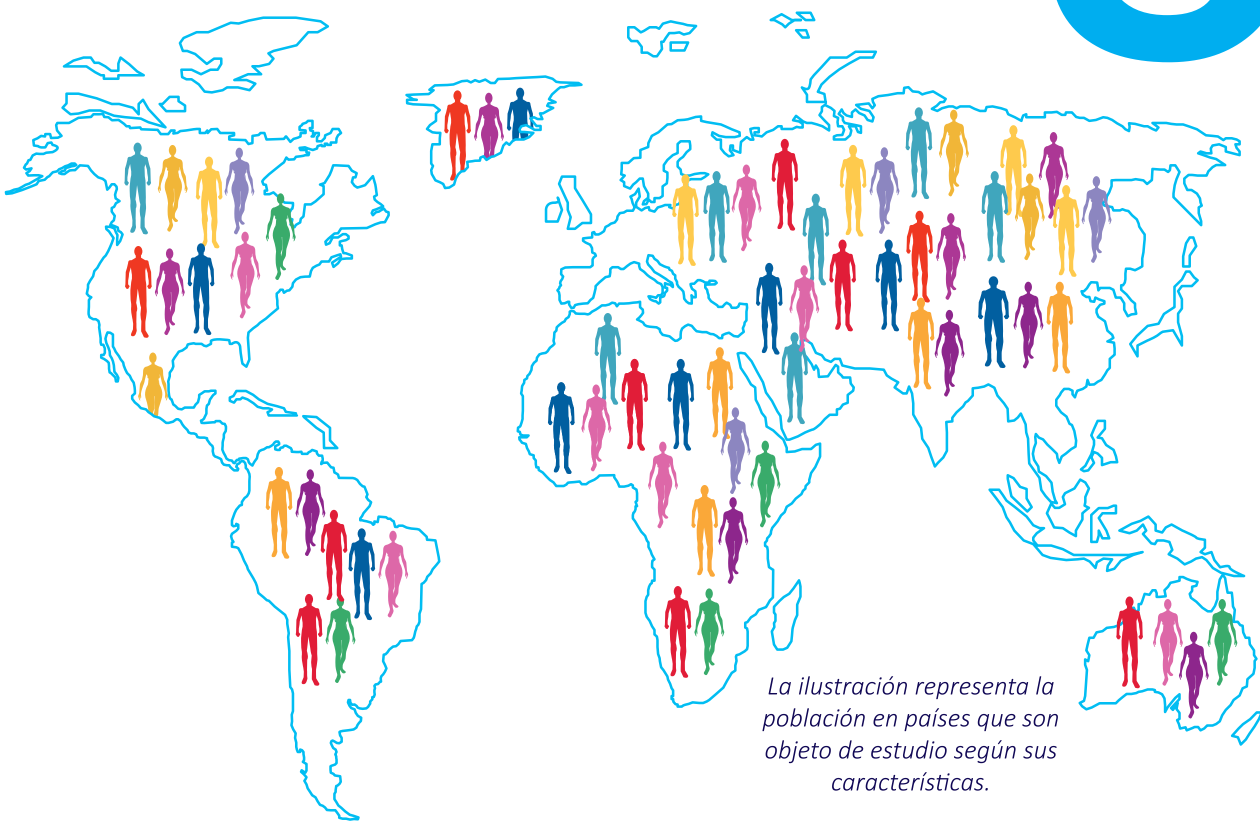
Medidas de dispersión

8 Unidad



La ilustración representa la población en países que son objeto de estudio según sus características.

Desde la antigüedad la estadística resultó ser muy útil a todas las ciencias; esta se constituyó en una herramienta importante en los procesos de investigación, ya que permite planear, recolectar y organizar información referente a individuos u observaciones de un fenómeno, al cuál se le estudian características en común en una población o muestra.

Para analizar una serie de datos no basta con conocer las medidas de tendencia central, que son las que indican donde se sitúan la mayoría de datos, también es necesario estudiar las medidas de dispersión o variabilidad, para saber qué tan próxima está entre si la información. Estas medidas tienen sus aplicaciones en las tarifas sobre servicios públicos; temperaturas por semanas; estudios sobre comportamientos de cierta población, longitudes recorridas por corredores, entre otros.

Al finalizar esta unidad sabrás cómo agrupar datos en una tabla de distribución de frecuencias, conocerás la media aritmética, la varianza y la desviación típica; para datos agrupados y no agrupados, esto se abordará con problemas de la vida cotidiana.

1.1 Rango para datos no agrupados

En la tabla se presenta la tarifa mensual (en dólares) por el servicio de agua potable en dos residenciales de San Salvador:

Residencial 1	
Casa	Tarifa mensual (en dólares)
1	12
2	11
3	12
4	13
5	12
6	18

Residencial 2	
Casa	Tarifa mensual (en dólares)
1	10
2	13
3	12
4	11
5	12
6	12
7	14

¿Cómo se calcula la media aritmética, mediana y moda para datos no agrupados?

a) Calcula la media aritmética, la mediana y la moda de las tarifas de cada una de las residenciales.

b) Para cada residencial calcula la diferencia entre la tarifa más alta y la más baja. ¿Cuál residencial tiene la mayor diferencia?

a) Para calcular la media aritmética (μ) se suman todos los datos y el resultado se divide entre el número de datos, la mediana es el dato que ocupa la posición central cuando estos se ordenan de menor a mayor y la moda es el dato con mayor frecuencia (el que más se repite).

Para la residencial 1:

$$\mu = \frac{12 + 11 + 12 + 13 + 12 + 18}{6} = \frac{78}{6} = 13$$

La media aritmética es $13.

Los datos ordenados de menor a mayor quedan de la siguiente forma:

11, 12, 12, 12, 13, 18

Por ser un número par de datos, la mediana es la media de los datos que ocupan las posiciones 3 y 4:

$$\frac{12 + 12}{2} = 12$$

La mediana es igual a $12. Por último, la moda es $12 para la residencial 1, pues es el dato que más se repite.

Para la residencial 2:

$$\mu = \frac{10 + 13 + 12 + 11 + 12 + 12 + 14}{7} = \frac{84}{7} = 12$$

La media aritmética es $12.

Los datos ordenados de menor a mayor quedan de la siguiente forma:
10, 11, 12, 12, 12, 13, 14
El dato que ocupa la posición central (cuarta posición) es 12, es decir, la mediana de la residencial 2 es igual a $12. Finalmente, la moda es igual a $12 para la residencial 2, pues es el dato que más se repite.

Los resultados anteriores se resumen en el siguiente cuadro:

	Residencial 1	Residencial 2
Media aritmética	$13	$12
Mediana	$12	$12
Moda	$12	$12

b) Para la residencial 1 la tarifa más alta es $18, la más baja es $11 y la diferencia es 18 − 11 = 7.
Para la residencial 2 la tarifa más alta es $14, la más baja es $10 y la diferencia es 14 − 10 = 4.

Por lo tanto, la diferencia de la tarifa más alta y la más baja es mayor en la residencial 1.

C

Las **medidas de dispersión** indican qué tanto se dispersan o agrupan los datos con respecto a su media aritmética.

El **rango** es una medida de dispersión que para una serie de datos no agrupados es igual a la diferencia del dato mayor y el dato menor. Al **rango** también se le llama **amplitud.** En el Problema inicial, las tarifas de la residencial 1 se encuentran más dispersas, ya que el rango es mayor.

1. Observa los datos no agrupados de las series A y B. ¿En cuál serie los datos están más dispersos?

	Serie A	Serie B
1	20.3	20.9
2	20.8	20.5
3	21.0	24.0
4	20.5	29.5
5	21.1	21.0
6	20.2	19.1
7	20.4	16.4

2. María registra la temperatura en dos semanas diferentes, obteniendo los resultados de la derecha.

a) Calcula la media aritmética, la mediana y la moda de cada semana.

b) Calcula el rango de cada semana, ¿en cuál de ellas los datos están más dispersos?

Semana 1	
Día	Temperatura
domingo	32°
lunes	31°
martes	29°
miércoles	30°
jueves	30°
viernes	29°
sábado	29°

Semana 2	
Día	Temperatura
domingo	35°
lunes	34°
martes	32°
miércoles	30°
jueves	30°
viernes	27°
sábado	25°

1.2 Desviación respecto a la media

P

Residencial 1	
Casa	Tarifa mensual (en dólares)
1	12
2	11
3	12
4	13
5	12
6	18

Residencial 2	
Casa	Tarifa mensual (en dólares)
1	10
2	13
3	12
4	11
5	12
6	12
7	14

En las series de datos de la tarifa mensual por el servicio de agua potable en dos residenciales de San Salvador, se obtuvo lo siguiente:

	Residencial 1	Residencial 2
Media aritmética	\$13	\$12
Mediana	\$12	\$12

a) ¿Cuál de las medidas; media o mediana, consideras puede ser más representativa para cada distribución?

b) En ambas series, encuentra las diferencias de cada dato y su media aritmética, ¿cómo se relacionan estas diferencias con la dispersión?

S

a) Para la residencial 1: la mayoría de los datos son menores a \$13, que es el valor de la media. Esto es debido a que la media se ve afectada por el sexto dato (\$18) que difiere considerablemente de los demás; entonces, para esta distribución la mediana puede ser un dato más representativo.

Para la residencial 2: tanto la media como la mediana tienen el mismo valor, puede tomarse cualquiera de los dos como dato más representativo de la distribución.

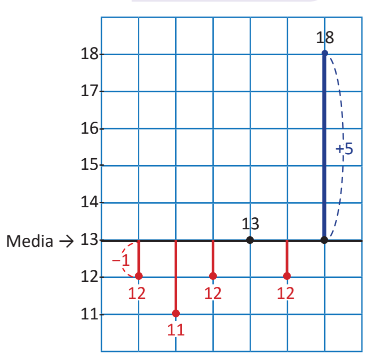
b) En la tabla se presentan las diferencias de cada uno de los datos y la media:

Residencial 1	
x	$x-\mu$
12	12−13 = −1
11	11−13 = −2
12	12−13 = −1
13	13−13 = 0
12	12−13 = −1
18	18−13 = 5

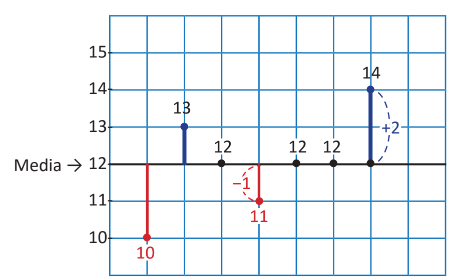
Residencial 2	
x	$x-\mu$
10	10−12 = −2
13	13−12 = 1
12	12−12 = 0
11	11−12 = −1
12	12−12 = 0
12	12−12 = 0
14	14−12 = 2

Sin tomar en cuenta los signos negativos, las diferencias reflejan la distancia de cada uno de los datos a su media aritmética. Lo anterior puede observarse mejor en los siguientes esquemas:

Residencial 1



Residencial 2



En ellos se observa que los datos de la residencial 2 se encuentran a menor distancia con respecto a su media aritmética ($12); mientras que en los datos de la residencial 1, el último de ellos está relativamente lejos de su media aritmética ($13).

En una distribución, a la diferencia de cada uno de los datos (x) y su media aritmética (μ) se le llama **desviación** respecto a la media (o simplemente desviación), se simboliza por $x - \mu$ e indica la diferencia de cada uno de los datos a la media aritmética. La suma de todas las desviaciones se simboliza por $\Sigma(x - \mu)$ y siempre es igual a cero:

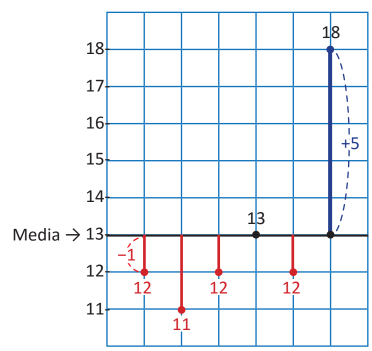
Suma de todas las desviaciones = 0

es decir,

$$\Sigma(x - \mu) = 0$$

Utilizando el Problema inicial, verifica que la suma de todas las desviaciones con respecto a la media es cero.

Residencial 1



En el esquema puede notarse que el valor absoluto de la suma de las distancias negativas es igual al de la positiva, haciendo que el resultado sea cero. También puede hacerse el cálculo:

$$\begin{aligned}\Sigma(x - \mu) &= (-1) + (-2) + (-1) + 0 + (-1) + 5\\ &= -1 - 2 - 1 - 1 + 5\\ &= -5 + 5\\ &= 0\end{aligned}$$

En la siguiente tabla se presentan tres series de datos no agrupados.

Completa cada una de las tablas y con base en las desviaciones, respecto a la media responde:

¿En cuál distribución los datos se encuentran más dispersos con respecto a la media?

Serie A	
x	$x - \mu$
15	
4	
6	
3	
2	
Media	
Mediana	
Rango	

Serie B	
x	$x - \mu$
8	
9	
6	
7	
5	
Media	
Mediana	
Rango	

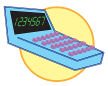
Serie C	
x	$x - \mu$
15	
5	
8	
10	
7	
Media	
Mediana	
Rango	

1.3 Varianza para datos no agrupados

P

Las desviaciones con respecto a la media pueden resultar complicadas de interpretar debido al signo negativo en alguna de ellas y cuando se tienen muchos datos.

En las tablas aparecen las desviaciones respecto a la media de los datos de la clase anterior:



Residencial 1	
x	$x - \mu$
12	12 − 13 = − 1
11	11 − 13 = − 2
12	12 − 13 = − 1
13	13 − 13 = 0
12	12 − 13 = − 1
18	18 − 13 = 5

Residencial 2	
x	$x - \mu$
10	10 − 12 = − 2
13	13 − 12 = 1
12	12 − 12 = 0
11	11 − 12 = − 1
12	12 − 12 = 0
12	12 − 12 = 0
14	14− 12 = 2

a) Calcula el cuadrado de cada una de las desviaciones con respecto a su media.

b) Calcula la media aritmética de los cuadrados de las desviaciones del literal anterior. Esta media aritmética se simboliza con σ^2 (σ es la letra griega sigma).

S

a) En las tablas se presentan los cuadrados de cada una de las desviaciones, en la columna $(x - \mu)^2$.

Residencial 1		
x	$x - \mu$	$(x - \mu)^2$
12	12 − 13 = − 1	$(-1)^2 = 1$
11	11 − 13 = − 2	$(-2)^2 = 4$
12	12 − 13 = − 1	$(-1)^2 = 1$
13	13 − 13 = 0	$0^2 = 0$
12	12 − 13 = − 1	$(-1)^2 = 1$
18	18 − 13 = 5	$5^2 = 25$

Residencial 2		
x	$x - \mu$	$(x - \mu)^2$
10	10−12 = −2	$(-2)^2 = 4$
13	13−12 = 1	$1^2 = 1$
12	12−12 = 0	$0^2 = 0$
11	11−12 = −1	$(-1)^2 = 1$
12	12−12 = 0	$0^2 = 0$
12	12−12 = 0	$0^2 = 0$
14	14−12 = 2	$2^2 = 4$

b) La media aritmética de los cuadrados de las desviaciones del literal anterior (que se simboliza por σ^2) se calcula sumando todos los resultados, de la última columna, de cada tabla y dividiéndolos entre el total de datos.

Para la residencial 1:

$$\sigma^2 = \frac{1 + 4 + 1 + 0 + 1 + 25}{6} = \frac{32}{6} \approx 5.33$$

Para la residencial 2:

$$\sigma^2 = \frac{4 + 1 + 0 + 1 + 0 + 0 + 4}{7} = \frac{10}{7} \approx 1.43$$

La medida (σ^2) sirve también para calcular la dispersión de los datos con respecto a su media. Se puede observar que cuanto mayor sean las desviaciones respecto a la media, mayor es σ^2 y por consiguiente más dispersos se encontrarán los datos.

σ^2 en la residencial 1 se ve afectada por la desviación del último dato cuyo cuadrado es 25, dando como resultado que sea mayor a σ^2 de la residencial 2. Por lo tanto, las tarifas de la residencial 1 se encuentran más dispersas.

C

A la media aritmética de los cuadrados de las desviaciones se le llama **varianza**, se denota por σ^2 y se calcula:

$$Varianza = \frac{\text{Suma de los cuadrados de las desviaciones}}{\text{Número de datos}}$$

es decir,

$$\sigma^2 = \frac{\Sigma(x - \mu)^2}{n}$$

Donde n es el número total de datos y μ es la media aritmética de la serie de datos. En el problema inicial, la varianza de la serie de datos de la residencial 1 es $\sigma^2 \approx 5.33$; mientras que la varianza de la serie de datos de la residencial 2 es $\sigma^2 \approx 1.43$.

Como esta medida es sensible a cada uno de los datos de la serie, la varianza revela aspectos en la dispersión que no refleja el rango. Cuanto mayor sea la varianza, más dispersos se encontrarán los datos con respecto a su media aritmética y puede recurrirse a la mediana como dato representativo de la distribución.

En las tablas se presentan las tres series de datos no agrupados de la clase anterior.

Serie A		
x	$x - \mu$	$(x - \mu)^2$
15		
4		
6		
3		
2		

Varianza (σ^2)	

Serie B		
x	$x - \mu$	$(x - \mu)^2$
8		
9		
6		
7		
5		

Varianza (σ^2)	

Serie C		
x	$x - \mu$	$(x - \mu)^2$
15		
5		
8		
10		
7		

Varianza (σ^2)	

Completa cada una de las tablas y calcula la varianza de cada serie. Con base en ella, justifica en cuál serie los datos se encuentran más dispersos. Compáralo con el resultado obtenido en la clase anterior.

1.4 Desviación típica para datos no agrupados

P

Con la tarifa mensual por el servicio de agua potable, en dos residenciales de San Salvador, realiza lo siguiente:

a) Calcula la raíz cuadrada de la varianza de ambas series y simbolízala por σ (sin el cuadrado). ¿Sigue siendo mayor el resultado de la residencial 1 que cuenta con datos más dispersos?
b) Coloca los datos de cada residencial como puntos sobre la recta numérica.
c) Resta y suma el respectivo valor de σ a cada media aritmética. Coloca estos números sobre la recta.
d) Según lo observado en la recta, ¿cuáles datos están más dispersos?

Residencial 1	
Casa	Tarifa mensual (en dólares)
1	12
2	11
3	12
4	13
5	12
6	18
σ^2	5.33 (dólares al cuadrado)

Residencial 2	
Casa	Tarifa mensual (en dólares)
1	10
2	13
3	12
4	11
5	12
6	12
7	14
σ^2	1.43 (dólares al cuadrado)

S

a) La raíz cuadrada de la varianza de los datos de la residencial 1 es:

$$\sigma = \sqrt{5.33}$$
$$\approx 2.31$$

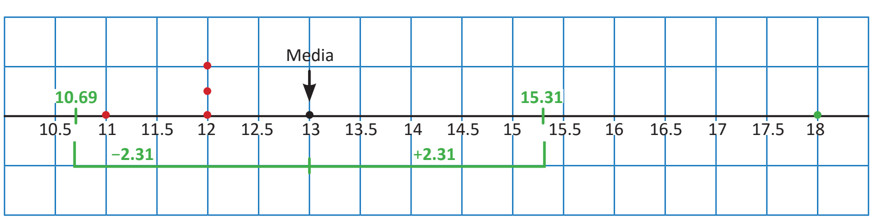
Mientras que la raíz cuadrada de la varianza de los datos de la residencial 2 es:

$$\sigma = \sqrt{1.43}$$
$$\approx 1.20$$

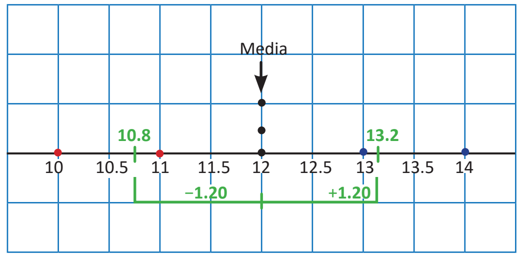
El resultado de la residencial 1 sigue siendo mayor que el de la residencial 2.

b) Cada punto representa uno de los datos; si dos o más datos tienen el mismo valor, entonces se ubican verticalmente sobre el valor correspondiente (los puntos rojos son los datos menores que la media, los azules los mayores, y los negros los que tienen igual valor que la media).

c) **En la serie de la residencial 1:** Para conocer la cantidad de datos que quedan a una distancia σ de su media (13) se le resta y suma a μ el valor de σ (que es 2.31) dando como resultado 10.69 y 15.31 respectivamente. En el esquema de abajo se observa que cinco de los seis datos de la serie quedan a una distancia de 2.31 de la media aritmética.



En la serie de la residencial 2: Al restar y sumar σ (1.20) a la media (12) se obtiene como resultado 10.8 y 13.2 respectivamente. En el esquema de abajo se observa que cinco de los siete datos de la serie quedan a una distancia de 1.20 de la media aritmética.



d) Aparentemente no hay mucha diferencia en las dos series, sin embargo, el hecho que σ sea menor para la residencial 2 indica que los datos se encuentran a una menor distancia de su media aritmética que los datos de la residencial 1, y por tanto, las tarifas mensuales de la residencial 1 se encuentran más dispersas (esto por influencia del dato cuyo valor es \$18).

A la raíz cuadrada de la varianza se le denomina **desviación típica**, se denota por σ y se calcula así:

$$\textit{Desviación Típica} = \sqrt{\textit{Varianza}}$$

$$= \sqrt{\frac{\textit{Suma de los cuadrados de las desviaciones}}{\textit{Número de datos}}}$$

Es decir,

$$\sigma = \sqrt{\frac{\Sigma(x - \mu)^2}{n}}$$

A la desviación típica también se le llama **desviación estándar.**

La desviación típica da un tipo de promedio de las desviaciones con respecto de la media μ, o sea, un promedio de las distancias de cada dato a su media aritmética, algo que no hace la varianza por expresarse en unidades cuadradas.

Cuanto mayor sea la desviación típica, más dispersos se encontrarán los datos con respecto a su media aritmética y puede recurrirse a la mediana como medida representativa de la serie de datos. La desviación típica siempre es un número mayor que cero o igual a cero (en su defecto), nunca será negativo.

Con las series de datos A, B y C del ejercicio de la clase anterior realiza los siguiente:

Serie A		
x	$x - \mu$	$(x - \mu)^2$
15		
4		
6		
3		
2		

σ^2	
σ	

Serie B		
x	$x - \mu$	$(x - \mu)^2$
8		
9		
6		
7		
5		

σ^2	
σ	

Serie C		
x	$x - \mu$	$(x - \mu)^2$
15		
5		
8		
10		
7		

σ^2	
σ	

a) Calcula la desviación típica de cada una de las series de datos.
b) Determina, en cada serie, la cantidad de datos que quedan a una distancia de una desviación típica con respecto a su media.

1.5 Agrupación de datos

P

Carlos y Antonio trabajan en la librería Maquilishuat. Durante 30 días registran la cantidad de cuadernos vendidos cada día, obteniendo el siguiente registro:

Carlos				
5	15	23	11	20
10	6	9	10	22
15	21	15	16	34
20	18	13	26	18
16	22	21	24	12
14	17	19	16	11

Antonio				
9	15	5	18	22
13	17	11	24	14
19	22	23	10	11
20	12	16	28	18
10	13	21	17	8
21	20	15	15	6

Cada casilla representa un día.

a) Clasifica el número de cuadernos vendidos en 6 grupos de 5 en 5, inicia en 5 y termina en 35.
b) Organiza los grupos en una tabla y determina el total de datos en cada grupo.

S

a) Como deben ser 6 grupos y el primero de ellos debe comenzar en 5 y el último terminar en 35, entonces los grupos serán: de 5 a 10 cuadernos, de 10 a 15 cuadernos, de 15 a 20 cuadernos, de 20 a 25 cuadernos, de 25 a 30 cuadernos y de 30 a 35 cuadernos. Según lo anterior, los cuadernos vendidos por Carlos quedan clasificados de la siguiente forma:

Carlos			16			
			19			
			17	24		
		11	16	21		
		14	18	22		
		12	18	20		
		13	16	21		
	9	10	15	22		
	6	10	15	20		
	5	11	15	23	26	34
	De 5 a 10	**De 10 a 15**	**De 15 a 20**	**De 20 a 25**	**De 25 a 30**	**De 30 a 35**

En el grupo “de 5 a 10” se colocan las cantidades 5, 6, 7, 8 y 9, si las hay; la cantidad final (10) se coloca en el grupo siguiente. De manera similar se hace para los demás grupos.

De forma similar se clasifican los cuadernos vendidos por Antonio:

En 8 días, Antonio vendió de 20 a 25

Antonio						
			15			
		13	15	20		
		10	17	21		
		12	18	21		
		11	16	20		
	6	10	19	23		
	8	14	17	22		
	5	11	18	24		
	9	13	15	22	28	
	De 5 a 10	**De 10 a 15**	**De 15 a 20**	**De 20 a 25**	**De 25 a 30**	**De 30 a 35**

b) La tabla queda de la siguiente manera:

Cantidad de cuadernos vendidos	Número de días	
	Carlos	Antonio
5 a 10	3	4
10 a 15	7	8
15 a 20	10	9
20 a 25	8	8
25 a 30	1	1
30 a 35	1	0
TOTAL	**30**	**30**

Este número representa la cantidad de días en los que Antonio vendió de 20 a 25 cuadernos.

C

La tabla en que se organizan los grupos de datos de una serie tal como en el Problema inicial se llama: **tabla de distribución de frecuencias.**

A los intervalos de datos formados se les llama **clases** y el total de datos que corresponde a cada clase se le llama **frecuencia**. Al tamaño de una clase se le llama ancho de clase y a los valores extremos **límites de clase**.

Por ejemplo, para la primera clase del Problema inicial los límites de clase son 5 y 10, el límite inferior es 5, el límite superior es 10 y el ancho de clase es 5. El número que está en el centro de cada clase se llama **punto medio**, se denota por Pm y se determina mediante la ecuación:

$$Pm = \frac{\textit{Límite superior + Límite inferior}}{2}$$

El punto medio de la primera clase es: $Pm = \frac{5 + 10}{2} = 7.5$

En dos comunidades de Morazán se hace un estudio sobre la edad de los menores de 21 años, obteniendo los siguientes resultados:

Comunidad 1					
9	14	15	14	19	16
11	18	9	12	20	12
12	11	10	19	14	13
15	12	11	18	11	16
14	16	17	12	13	17

Comunidad 2					
14	13	9	17	15	9
9	14	15	20	18	12
13	10	9	11	10	13
16	12	12	11	10	13
18	11	14	10	19	9

a) Clasifica las edades de los menores de 21 años de cada comunidad en 4 grupos de 3 en 3, inicia en 9 y termina en 21.
b) Organiza los datos en una tabla de distribución de frecuencias.
c) Con la tabla creada, agrega otra columna donde se muestre el punto medio de cada clase.

1.6 Media aritmética y rango para datos agrupados

P

Carlos y Antonio trabajan en la librería Maquilishuat. Durante un mes registran la cantidad de cuadernos vendidos por día, obteniendo el siguiente dato:

Cantidad de cuadernos vendidos	Número de días		Punto medio de cada clase (Pm)	$f_C \times Pm$	$f_A \times Pm$
	Carlos (f_C)	Antonio (f_A)			
5 a 10	3	4	7.5	22.5	30.0
10 a 15	7	8			
15 a 20	10	9			
20 a 25	8	8			
25 a 30	1	1			
30 a 35	1	0			
TOTAL	30	30			

a) Completa la tabla y calcula la media aritmética para cada una de las series de datos (de Carlos y Antonio), ¿qué ocurre?
b) Identifica en cada una el límite superior de la última clase que posee frecuencia distinta de cero y el límite inferior de la primera clase que posee frecuencia distinta de cero.
c) Realiza, para cada serie, la diferencia del límite superior y el límite inferior encontrado en el literal b. ¿En cuál serie los datos se encuentran más dispersos?

S

a) Los valores de la tabla quedan de la siguiente manera:

Cantidad de cuadernos vendidos	Número de días		Punto medio de cada clase (Pm)	$f_C \times Pm$	$f_A \times Pm$
	Carlos (f_C)	Antonio (f_A)			
5 a 10	3	4	7.5	22.5	30.0
10 a 15	7	8	12.5	87.5	100.0
15 a 20	10	9	17.5	175.0	157.5
20 a 25	8	8	22.5	180.0	180.0
25 a 30	1	1	27.5	27.5	27.5
30 a 35	1	0	32.5	32.5	0.0
TOTAL	30	30			

La media aritmética de la serie de datos de Carlos se calcula:

$$\mu = \frac{\textit{Suma de los productos } f \times Pm}{\textit{Número de datos}}$$

$$\mu = \frac{22.5 + 87.5 + 175.0 + 180.0 + 27.5 + 32.5}{30}$$

$$= \frac{525}{30}$$

$$= 17.5$$

La media aritmética de la serie de datos de Antonio, se calcula de la misma manera que la de Carlos, y el resultado es 16.5.

b) Para el caso de Carlos la última clase que posee frecuencia distinta de cero es la clase de 30 a 35, que tiene frecuencia igual a 1 cuyo límite superior es 35, y la primera clase que posee frecuencia distinta de cero es la clase de 5 a 10 (tiene frecuencia 3) cuyo límite inferior es 5.

Primera clase con frecuencia distinta de cero. → 5 a 10

Última clase con frecuencia distinta de cero. → 30 a 35

Cantidad de cuadernos vendidos	Número de días Carlos (f_C)
5 a 10	3
10 a 15	7
15 a 20	10
20 a 25	8
25 a 30	1
30 a 35	1
TOTAL	30

De igual forma se identifican las dos clases para el caso de Antonio:

Primera clase con frecuencia distinta de cero. → 5 a 10

Última clase con frecuencia distinta de cero. → 25 a 30

Cantidad de cuadernos vendidos	Número de días Antonio (f_A)
5 a 10	4
10 a 15	8
15 a 20	9
20 a 25	8
25 a 30	1
30 a 35	0
TOTAL	30

c) Para la serie de Carlos la diferencia es: 35 − 5 = 30.
Y para la serie de Antonio la diferencia es: 30 − 5 = 25.

Por lo tanto, los datos de la serie de Carlos se encuentran más dispersos.

El **rango** para una serie de datos agrupados es la diferencia del límite superior de la última clase con frecuencia distinta de cero y el límite inferior de la primera clase con frecuencia distinta de cero. La **media aritmética** para series de datos agrupados se calcula así:

$$\mu = \frac{\textit{Suma de los productos } f\times Pm}{\textit{Número de datos}}$$

En un centro escolar se registra el tiempo, en minutos, que los estudiantes de octavo y noveno grado miran televisión al día, los datos se muestran en la siguiente tabla de distribución de frecuencia:

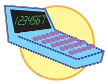
Minutos	8° grado (f_1)	9° grado (f_2)	Pm	$f_1 \times Pm$	$f_2 \times Pm$
30 a 40	0	3	35	0	105
40 a 50	10	8			
50 a 60	11	9			
60 a 70	12	12			
70 a 80	11	10			
80 a 90	6	8			
TOTAL	50	50			

a) Completa la tabla, encuentra la media aritmética para cada una de las series de datos y compáralas.
b) Comparando los rangos, determina en cuál serie los datos se encuentran más dispersos.

1.7 Varianza para datos agrupados

P

En la tabla aparecen los datos correspondientes a la cantidad de cuadernos vendidos por Carlos durante 30 días:



Cantidad de cuadernos vendidos	Número de días Carlos (f_C)	Punto medio (Pm)	$f \times Pm$	$Pm - \mu$	$(Pm - \mu)^2$	$f\,(Pm - \mu)^2$
5 a 10	3	7.5	22.5	−10	100	300
10 a 15	7	12.5	87.5			
15 a 20	10	17.5	175.0			
20 a 25	8	22.5	180.0			
25 a 30	1	27.5	27.5			
30 a 35	1	32.5	32.5			
TOTAL	30					
Media aritmética (μ)	**17.5**					

a) Completa la tabla y calcula la suma de los datos de la última columna.
b) ¿Cómo podrías calcular la varianza para esta serie de datos agrupados?

S

a) La tabla completa se presenta a continuación:

Cantidad de cuadernos vendidos	Número de días Carlos (f_C)	Punto medio (Pm)	$f \times Pm$	$Pm - \mu$	$(Pm - \mu)^2$	$f(Pm - \mu)^2$
5 a 10	3	7.5	22.5	−10	100	300
10 a 15	7	12.5	87.5	−5	25	175
15 a 20	10	17.5	175.0	0	0	0
20 a 25	8	22.5	180.0	5	25	200
25 a 30	1	27.5	27.5	10	100	100
30 a 35	1	32.5	32.5	15	225	225
TOTAL	**30**					
Media aritmética (μ)	**17.5**					

La sumatoria de los datos de la última columna, $f(Pm - \mu)^2$, es:

$$300 + 175 + 0 + 200 + 100 + 225 = 1\,000$$

b) Para calcular la varianza, basta dividir entre el número total de datos el resultado de la suma calculada en el literal a), es decir:

$$\sigma^2 = \frac{1\,000}{30}$$

$$\approx 33.33$$

Por lo tanto, $\sigma^2 \approx 33.33$.

Igual que en las series de datos no agrupados, la varianza se encuentra expresada en unidades cuadradas. Para este caso serían "días al cuadrado".

C La varianza de una serie de datos agrupados se calcula de la siguiente forma:

$$\text{Varianza} = \frac{\text{Suma de los productos } f\,(Pm - \mu)^2}{\text{Número de datos}}$$

es decir,

$$\sigma^2 = \frac{\Sigma\, f(Pm - \mu)^2}{n}$$

Donde n es el número total de datos, Σ es el símbolo de sumatoria, f es la frecuencia de cada clase, Pm es el punto medio de cada clase y μ es la media aritmética de la serie de datos.

Cuanto mayor sea la varianza, más dispersos se encontrarán los datos con respecto a su media aritmética.

1. Completa la tabla y calcula la varianza para la cantidad de cuadernos vendidos por Antonio. Luego determina en cuál serie los datos se encuentran más dispersos, comparando las varianzas.

Cantidad de cuadernos vendidos	Número de días Antonio (f_c)	Punto medio (Pm)	$f \times Pm$	$Pm - \mu$	$(Pm - \mu)^2$	$f(Pm - \mu)^2$
5 a 10	4	7.5	30.0	−9	81	324
10 a 15	8	12.5	100.0			
15 a 20	9	17.5	157.5			
20 a 25	8	22.5	180.0			
25 a 30	1	27.5	27.5			
30 a 35	0	32.5	0.0			
TOTAL	**30**					
Media aritmética (μ)	**16.5**					

2. Con las series de datos agrupados de las dos comunidades de Morazán de la clase 5 realiza lo siguiente:

a) Completa la siguiente tabla y calcula la varianza de los datos de la comunidad 1:

Edad en años	Cantidad de personas	Punto medio (Pm)	$f_1 \times Pm$	$Pm - \mu$	$(Pm - \mu)^2$	$f_1(Pm - \mu)^2$
	Comunidad 1 (f_1)					
De 9 a 12	7	10.5	73.5	−4	16	112
De 12 a 15	11	13.5	148.5			
De 15 a 18	7	16.5	115.5			
De 18 a 21	5	19.5	97.5			
TOTAL	**30**					
Media aritmética (μ)	**14.5**					

b) Elabora una tabla como la anterior para la comunidad 2 y calcula la varianza de sus datos.
c) Con base a lo anterior responde, ¿en cuál comunidad los datos se encuentran más dispersos?

1.8 Desviación típica para datos agrupados

Calcula la desviación típica de la serie de datos correspondientes a la cantidad de cuadernos vendidos por Carlos y Antonio durante 30 días y justifica en cuál de ellas los datos se encuentran más dispersos.

Cantidad de cuadernos vendidos	Número de días	
	Carlos (f_C)	Antonio (f_A)
5 a 10	3	4
10 a 15	7	8
15 a 20	10	9
20 a 25	8	8
25 a 30	1	1
30 a 35	1	0
TOTAL	**30**	**30**
Media aritmética (μ)	**17.5**	**16.5**
Varianza (σ^2)	33.33	29

Para datos agrupados en clases, la desviación típica sigue siendo igual a la raíz cuadrada de la varianza.
Para la serie de datos de Carlos:

$$\sigma = \sqrt{33.33}$$
$$\approx \ 5.77$$

Y para la serie de datos de Antonio:

$$\sigma = \sqrt{29}$$
$$\approx \ 5.39$$

Como la desviación típica de la distribución de Carlos es mayor a la de Antonio, se concluye que los datos de la distribución de Carlos se encuentran más dispersos, con respecto a su media aritmética 17.5.

La desviación típica de una serie de datos agrupados se calcula:

$$\textit{Desviación típica} = \sqrt{\textit{varianza}}$$

Es decir,

$$\sigma = \sqrt{\frac{\sum f(Pm - \mu)^2}{n}}$$

Donde n es el número total de datos, Σ es el símbolo de sumatoria, f es la frecuencia de cada clase, Pm es el punto medio de cada clase y μ es la media aritmética de la serie de datos. Tanto para datos agrupados como no agrupados, la desviación típica siempre es un número mayor que cero o igual a cero (en su defecto), nunca será un número negativo.

Con base al ejercicio 2 de la clase anterior. Calcula la desviación típica (σ) de las series de datos agrupados de las dos comunidades de Morazán y responde con base a esta medida, ¿en cuál comunidad los datos se encuentran más dispersos?

1.9 Practica lo aprendido

1. Los siguientes datos representan las estaturas de 8 estudiantes en centímetros:
163, 162, 164, 163, 164, 162, 161, 185.

a) Calcula la media aritmética, la mediana y el rango de la serie de datos.
b) ¿Cuál de las medidas, media o mediana, escogerías para representar la distribución? Justifica tu respuesta.

2. Con los datos presentados en la siguiente tabla, determina cuáles de las series de datos B, C y D tienen igual desviación típica que la serie de datos de A. Justifica tu respuesta.

A	B	C	D
25	30	35.5	28
24	29	34.5	27
25	30	35.5	28
26	31	36.5	29
23	28	33.5	26
21	26	31.5	21
25	30	35.5	28
24	29	34.5	27
23	28	33.5	23
22	27	32.5	22

3. Observa las siguientes series de datos no agrupados:

	Serie A
1	30
2	25
3	11
4	20
5	14
6	26

	Serie B
1	18
2	20
3	19
4	21
5	22

a) Calcula las desviaciones respecto a su media aritmética de cada serie y la desviación típica.
b) Comparando las desviaciones típicas, determina en cuál serie los datos se encuentran más dispersos.

4. Los siguientes datos representan el peso en libras de 9 personas que trabajan en una oficina.
160 l, 200 l, 164 l, 130 l, 140 l, 162 l, 161 l, 185 l, 154 l.

a) Calcula la media aritmética, la mediana y el rango de la serie de datos.
b) ¿Cuál de las medidas, media o mediana, escogerías para representar la distribución? Justifica tu respuesta.

1.10 Practica lo aprendido

1. Una tienda de ropa tiene dos sucursales A y B. En 100 días registran la cantidad de clientes atendidos en cada sucursal, los datos se presentan en la siguiente tabla:

Cantidad de clientes	Cantidad de días	
	Sucursal A	Sucursal B
50 a 60	15	17
60 a 70	20	21
70 a 80	24	27
80 a 90	22	20
90 a 100	19	15

a) Calcula la varianza para cada una de las sucursales.
b) Con base a la varianza, ¿en cuál sucursal los datos se encuentran más dispersos?

2. Se realizó un estudio sobre el peso, en libras, de los estudiantes de noveno grado de un centro escolar. Los resultados se presentan en la siguiente tabla:

Peso en libras	Sección A	Sección B
120 a 130	7	5
130 a 140	12	9
140 a 150	13	12
150 a 160	10	14
160 a 170	8	10

Utiliza la desviación típica para determinar en cuál de las secciones los pesos se encuentran más dispersos, con respecto a su media aritmética.

3. La estatura en pulgadas de cierto grupo de personas se muestra en la siguiente tabla. Sabiendo que $\mu = 67.45$

Estatura	f	Pm
60 - 62	1	61
62 - 64	4	63
64 - 66	8	65
66 - 68	30	67
68 - 70	37	69

a) Calcula la varianza.
b) Calcula la desviación típica.

2.1 Desviación típica de una variable, más una constante

En una empresa se aumenta \$50 al salario de 10 trabajadores; en la tabla de la derecha se muestran los salarios anteriores y el salario actual.

a) ¿Cuál es la media aritmética de ambas series de datos?
b) Calcula la desviación típica para ambas series de datos y compáralas, ¿qué ocurre?
c) ¿Qué pasaría con la desviación típica de los datos del salario actual si el aumento fuera de \$60?

Trabajador	Salario anterior (en dólares)	Salario actual (en dólares)
1	485	535
2	488	538
3	486	536
4	489	539
5	486	536
6	485	535
7	488	538
8	487	537
9	500	550
10	486	536

a) La media aritmética de los salarios anteriores se calcula:

$$\mu = \frac{485 + 488 + 486 + 489 + 486 + 485 + 488 + 487 + 500 + 486}{10}$$
$$= 488$$

Si a los datos de una serie A se les suma una constante dando como resultado otra serie B, entonces la media de B es igual a la media de A más la constante.

De manera similar se calcula la media aritmética de los salarios actuales, cuyo resultado es 538. Por lo tanto, la media aritmética de los salarios anteriores es \$488 y la de los salarios actuales es \$538.

b) En la tabla se muestran las desviaciones, de los salarios anteriores, con respecto a su media aritmética \$488 y sus respectivos cuadrados:

Trabajador	Salario anterior (en dólares)	$x - \mu$	$(x - \mu)^2$
1	485	−3	9
2	488	0	0
3	486	−2	4
4	489	1	1
5	486	−2	4
6	485	−3	9
7	488	0	0
8	487	−1	1
9	500	12	144
10	486	−2	4
Media (μ)	**488**		

Luego, la desviación típica se calcula:

$$\sigma = \sqrt{\frac{9+0+4+1+4+9+0+1+144+4}{10}}$$

$$= \sqrt{\frac{176}{10}}$$

$$= 4.2$$

De manera similar se calcula la desviación típica de los salarios actuales, cuyo resultado también es 4.2; es decir, la desviación típica a diferencia de la media aritmética, no se vio afectada al sumar 50 a cada uno de los datos.

c) Si el aumento fuera de $60, entonces la desviación típica sería igual a la calculada para el salario anterior, o sea 4.2.

C

Si a cada uno de los datos de una distribución A se les suma la misma constante c (c es un número cualquiera) dando como resultado otra distribución B, entonces la desviación típica de la distribución B es igual a la desviación típica de la distribución A.

1. Observa la tabla con dos series de datos A y B, ¿tienen ambas distribuciones la misma desviación típica? Justifica tu respuesta y calcula el valor de la misma.

	Serie A	Serie B
1	25.1	37.1
2	26.4	38.4
3	27.5	39.5
4	20.7	32.7
5	21.2	33.2

2. En la residencial Centroamérica aumentarán $5 a la tarifa mensual por el servicio de agua potable. ¿Cuál será el valor de la desviación típica de la distribución teniendo en cuenta este cambio?

Casa	Tarifa a cancelar (en dólares)
1	10.50
2	10.60
3	12.20
4	11.50
5	12.90
6	11.40
7	12.60
8	12.50
9	11.30
10	35.50

2.2 Desviación típica de una variable multiplicada por una constante

P

Cinco corredores deciden que para el mes de febrero aumentarán al doble las longitudes que recorren cada semana para entrenar. En la tabla se presenta la longitud recorrida en enero y la longitud que se recorrerá en febrero.

a) Calcula la desviación típica de ambas series de datos.

b) Efectúa el cociente entre la desviación típica de febrero y la desviación típica de enero, ¿cuál es la relación entre ambos datos?

Corredores	Longitud recorrida en metros	
	Enero	Febrero
1	150	300
2	160	320
3	145	290
4	165	330
5	150	300

S

a) Para calcular la desviación típica es necesario tener la media aritmética de cada serie de datos. Para enero la media es:

$$\mu = \frac{150 + 160 + 145 + 165 + 150}{5} = 154.$$

Se calculan los cuadrados de las desviaciones con respecto a la media aritmética 154 m, los resultados se presentan en la siguiente tabla:

Corredores	Enero	$x - \mu$	$(x - \mu)^2$
1	150	–4	16
2	160	6	36
3	145	–9	81
4	165	11	121
5	150	–4	16
Media (μ)	**154**		

$$\sigma = \sqrt{\frac{16 + 36 + 81 + 121 + 16}{5}}$$

$$= \sqrt{\frac{270}{5}}$$

$$= 7.35$$

La desviación típica de enero es 7.35.

Para los datos de febrero: como las longitudes se han aumentado al doble, (se han multiplicado por 2) entonces la media aritmética de febrero también aumenta al doble, o sea 308 m. La desviación típica se calcula de manera similar a la de enero, dando como resultado 14.7.

b) El cociente es:

$$\frac{\textit{Desviación típica de febrero}}{\textit{Desviación típica de enero}} = \frac{14.7}{7.35} = 2$$

Es decir, la desviación típica de febrero es el doble de la desviación típica de enero. Cuando los datos se multiplican por un número positivo, la desviación típica también se multiplica por ese número.

C

Si a cada uno de los datos de una distribución A se les multiplica por la misma constante c (c es un número positivo), dando como resultado otra distribución B, entonces la desviación típica de la distribución B es igual a multiplicar la desviación típica de la distribución A por la constante c.

1. En la tabla de abajo se presentan tres series de datos:

   a) ¿Cuál es el número por el que se tienen que multiplicar los datos de la serie A para obtener la serie B?, ¿y para obtener la serie C?

   b) Calcula la desviación típica de la serie A, con base a ella calcula la desviación típica de las series B y C.

A	B	C
12.5	18.75	5.0
11.0	16.5	4.4
11.5	17.25	4.6
12.8	19.2	5.12
12.2	18.30	4.88

2. En una serie de datos, la media aritmética de la distribución es 35 y la desviación típica es 17.07; si cada uno de los datos se reduce a la mitad, ¿cuál será el valor de la nueva desviación típica?

3. Una librería registra la cantidad de libros vendidos, de lunes a viernes, durante dos semanas.

   a) ¿Cuál es el número por el que se tienen que multiplicar los datos de la semana 1 para obtener los de la semana 2?
   b) Calcula la desviación típica para ambas semanas.

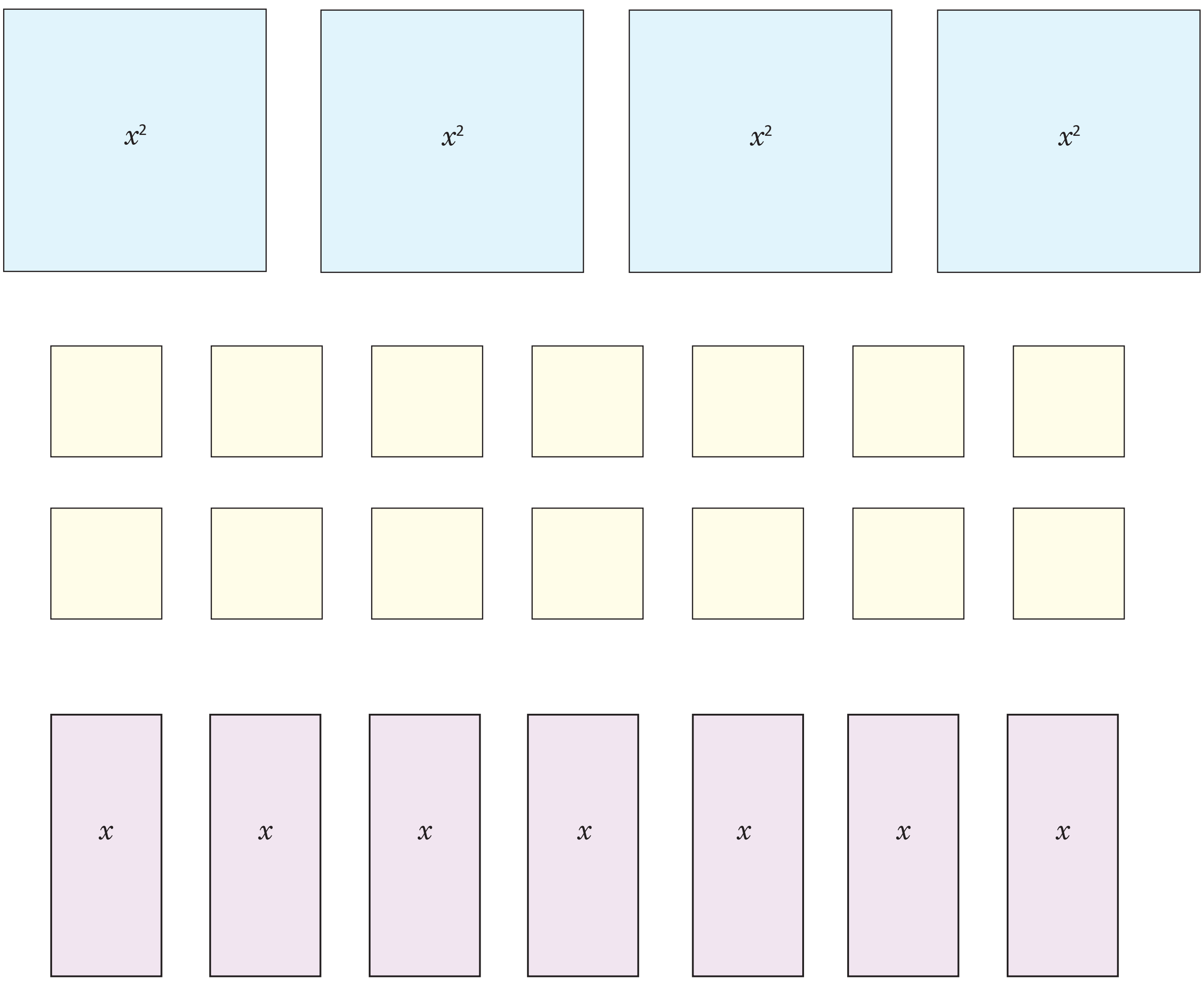
Días	Cantidad de libros vendidos	
	Semana 1	Semana 2
lunes	8	24
martes	9	27
miércoles	5	15
jueves	7	21
viernes	11	33

Material complementario

Se presenta el siguiente recurso para que pueda servir como apoyo en algunas clases de este libro. Este material no es para recortar, sino para sacar fotocopias en caso de que sea necesario.

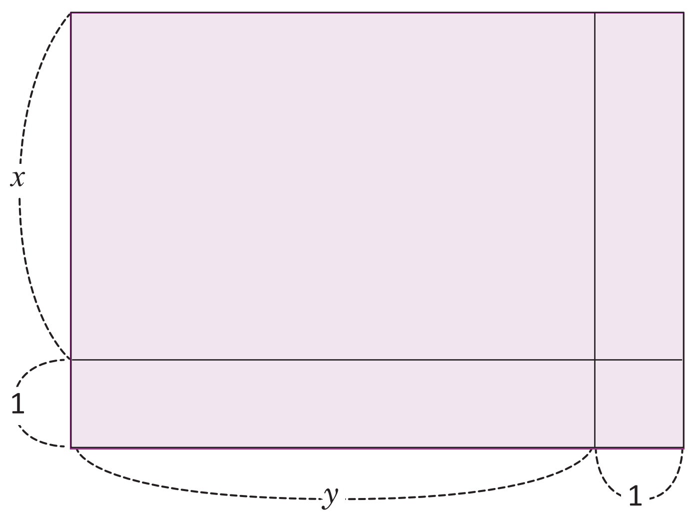
UNIDAD 1: Multiplicación de polinomios

Figuras que se pueden utilizar para las siguientes clases: 1.1, 2.1, 3.1, 3.3, 3.5

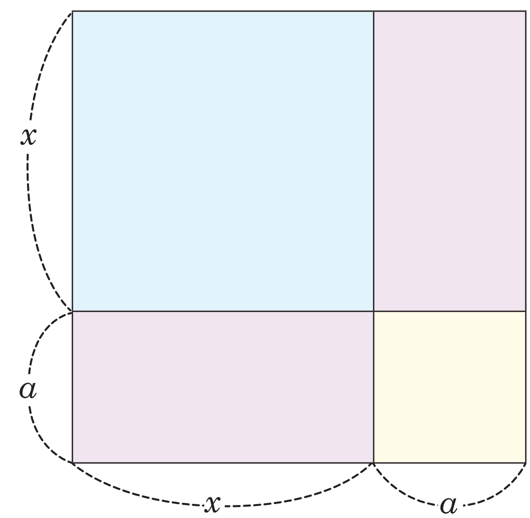


Figuras que se pueden utilizar para las siguientes clases: 1.1, 2.1, 3.1, 3.3, 3.5

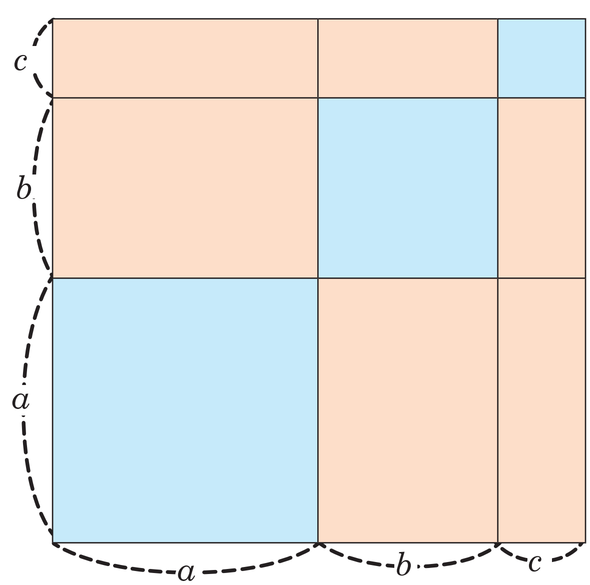
Clase: 1.2



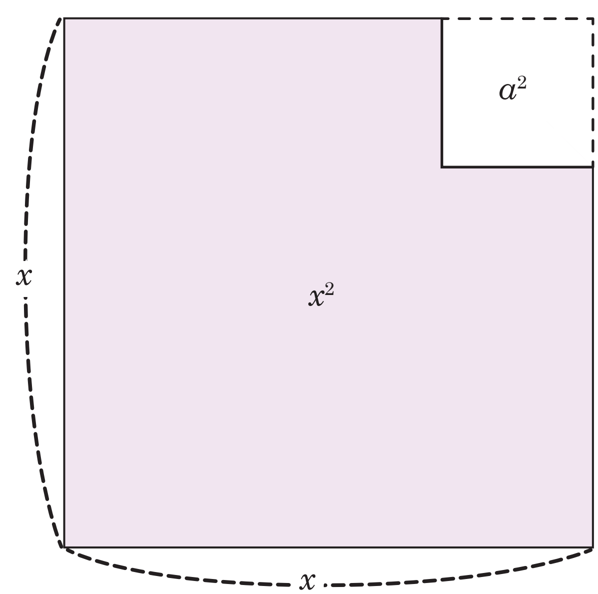
Clase: 2.2



Clase: 2.7



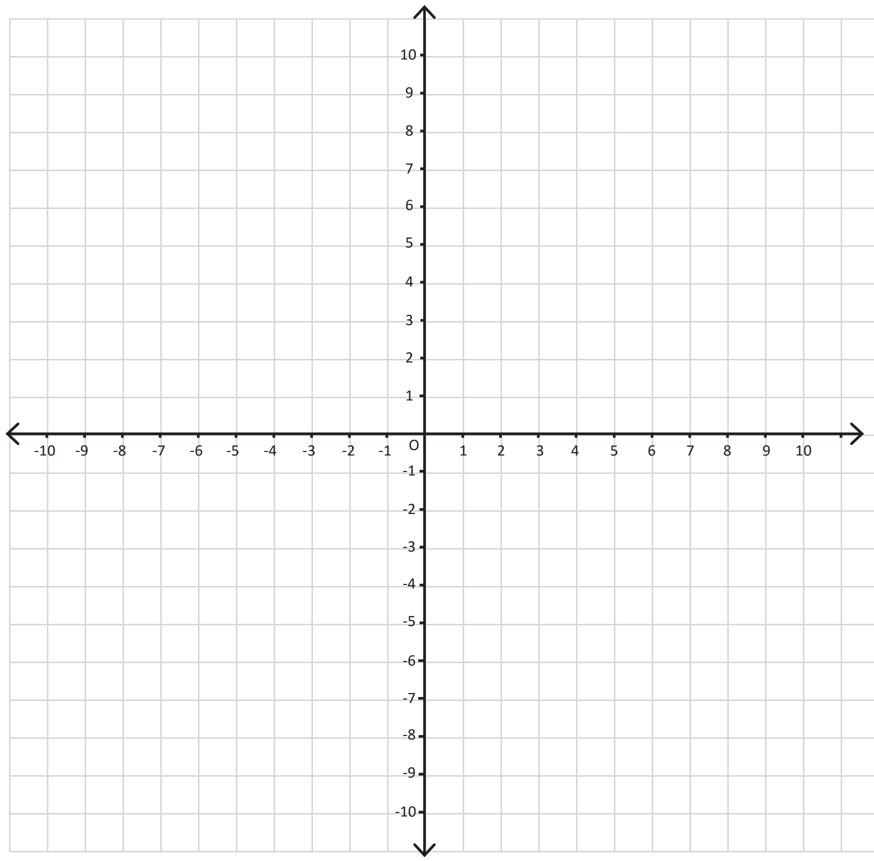
Clase: 3.6



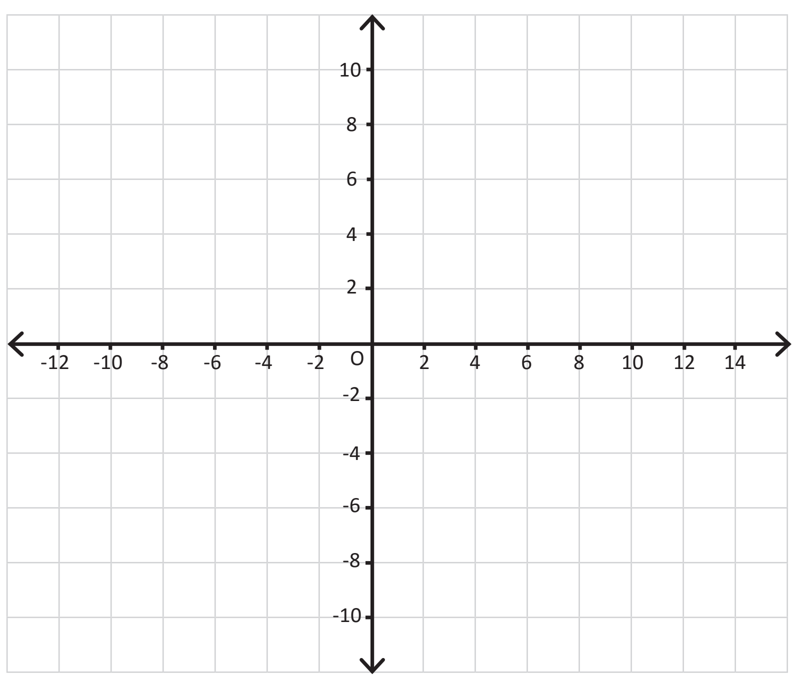
UNIDAD 4: Función cuadrática de la forma $y = ax^2 + c$

Figuras que se pueden utilizar para las clases:

1.4, 1.5, 1.8, 1.9, 2.1, 2.2, 2.3

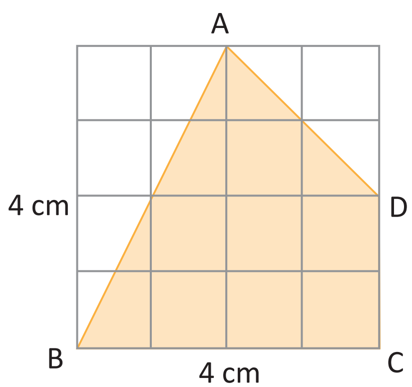


Clase: 1.6

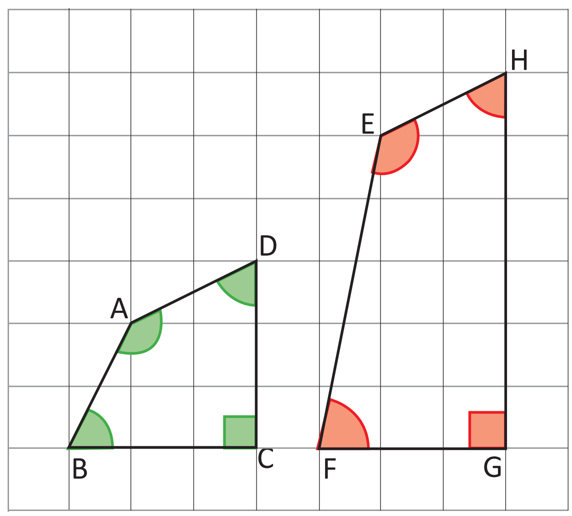


Figuras que se pueden utilizar:

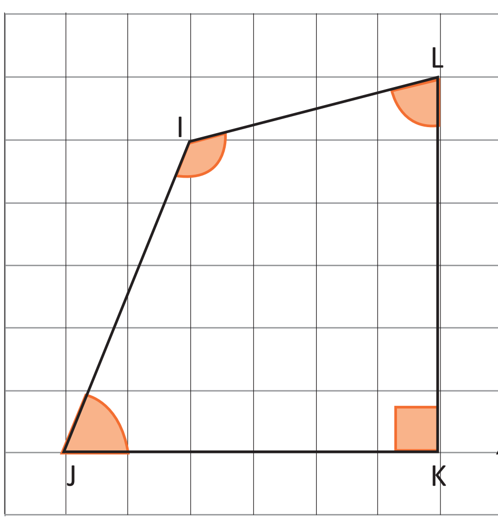
Clase 1.3



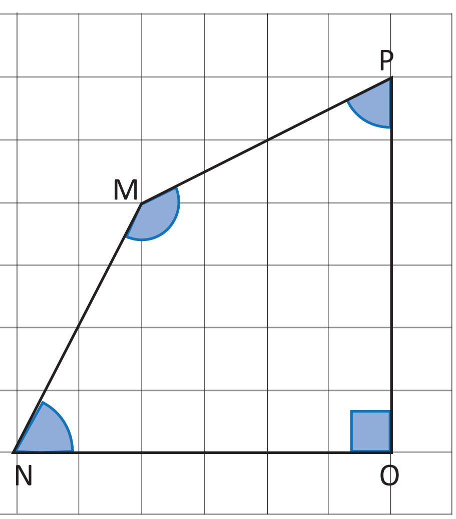
Clase 1.4



Clase 1.4



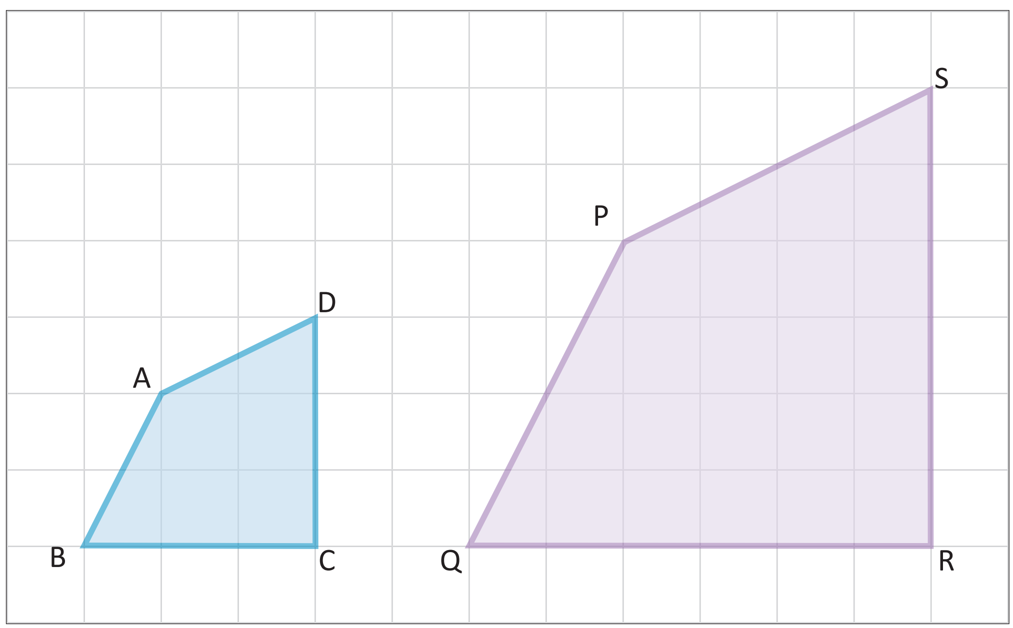
Clase 1.5



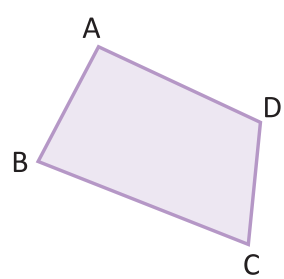
UNIDAD 5: Figuras semejantes

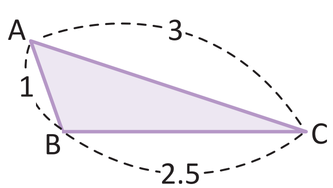
Figuras que se pueden utilizar:

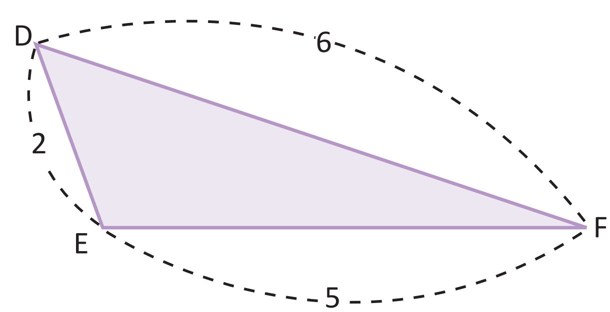
Clase 5 Clase 5

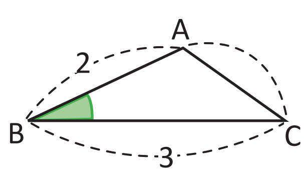


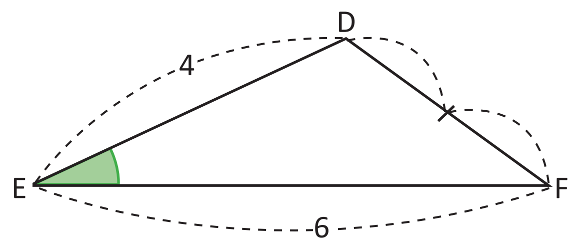
Clase 1.6







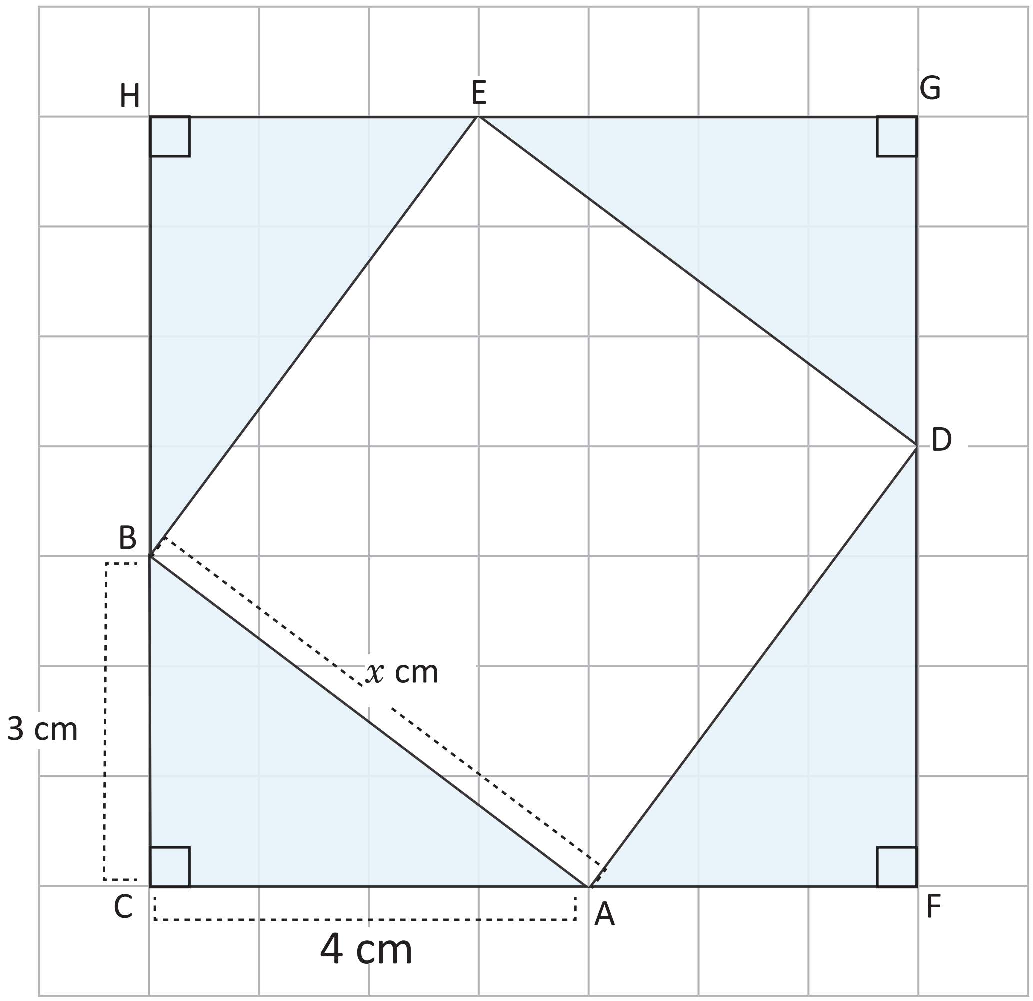




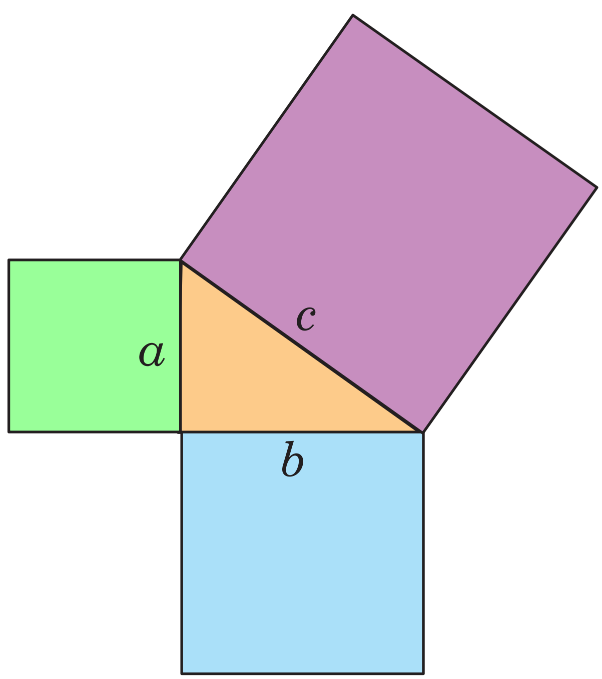
Unidad 8

Figuras que se pueden utilizar:

clase 1.1



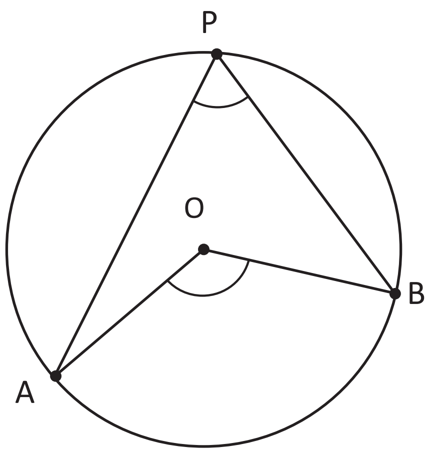
clase 1.4



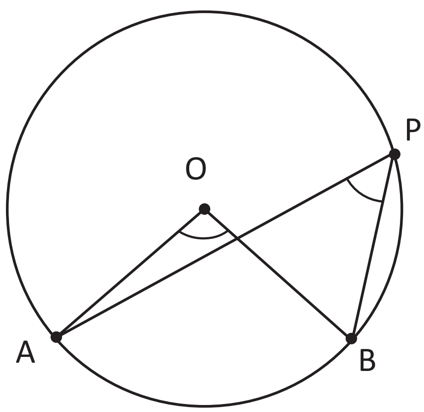
UNIDAD 7: Ángulo inscrito y central

Figuras que se pueden utilizar:

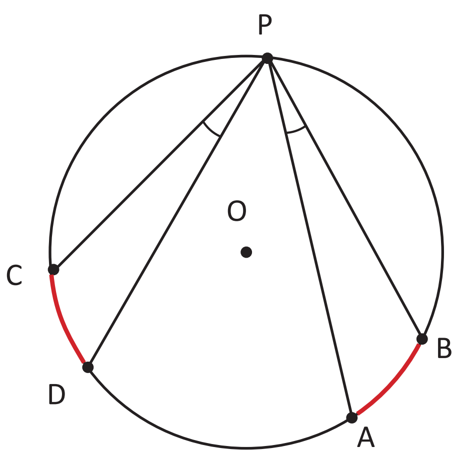
clase 1.4



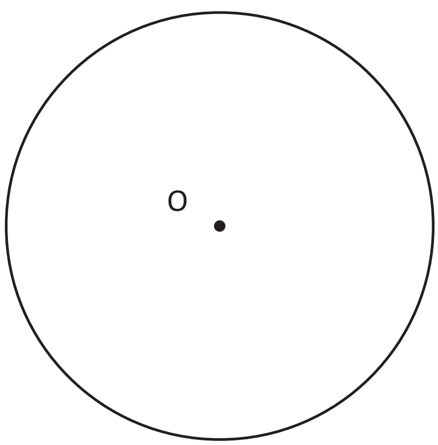
clase 1.5



clase 1.7



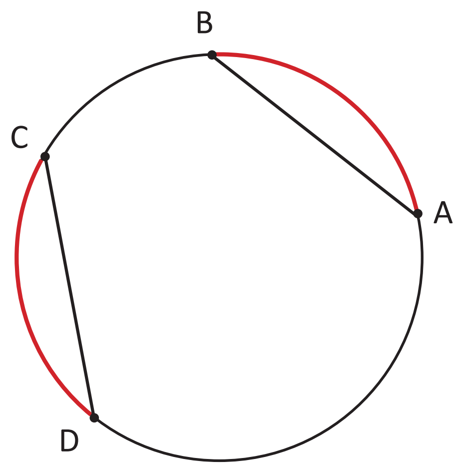
clase 2.1



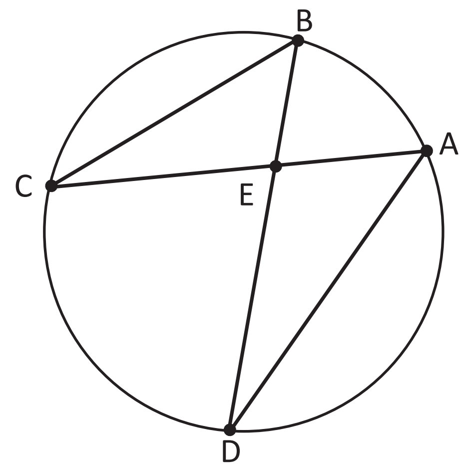
UNIDAD 7: Ángulo inscrito y central

Figuras que se pueden utilizar:

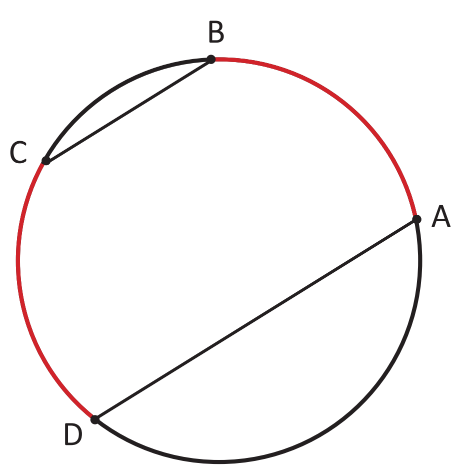
Clase 2.2



Clase 2.3



Clase 2.4



Clase 2.6