Versión del estudiante

Eureka Math

Álgebra I

Módulo 2

Se les extiende un agradecimiento especial al Centro Gordon A. Cain y al Departamento de Matemáticas de la Universidad Estatal de Luisiana por su colaboración en el desarrollo de *Eureka Math*.

Para obtener un paquete gratis de recursos de *Eureka Math* para maestros, los Consejos para padres y otros recursos, por favor, visite www.Eureka.tools.

Great Minds PBC es el creador de *Eureka Math*®,

Wit & Wisdom®, *Alexandria Plan*™ y *PhD Science*™.

Publicado por Great Minds PBC. greatminds.org

ISBN 978-1-68386-234-5

10 9 8 7 6 5 4 3 2 1

Impreso en los EE. UU.

Lección 1: Las distribuciones y sus formas

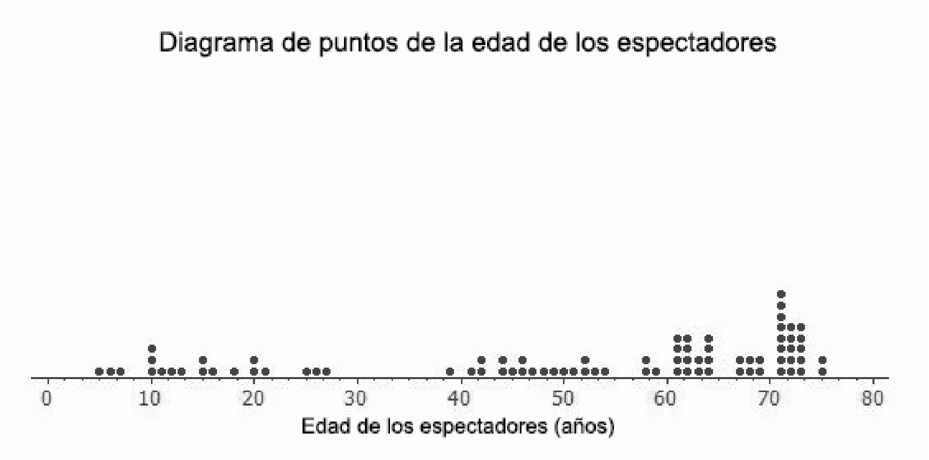
Trabajo en clase

Las estadísticas tratan sobre los datos. Sin datos que discutir, analizar o cuestionar, las estadísticas no existirían. Detrás de todos los datos hay una historia por descubrir, una historia que tiene personajes, argumentos y problemas. Las preguntas o los problemas que abordan esos datos y su historia pueden ser decepcionantes, emocionantes o simplemente corrientes. Este módulo trata sobre historias que comienzan con datos.

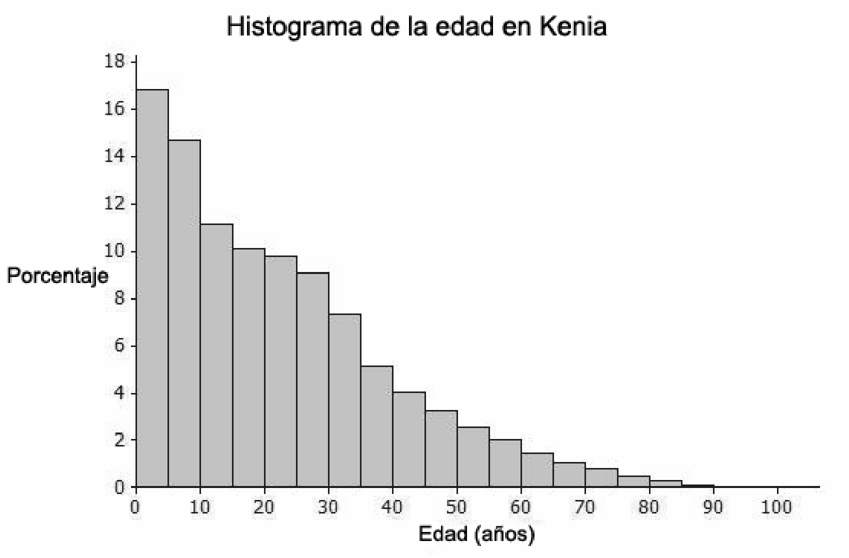
Ejemplo: Gráficas

Los datos suelen resumirse mediante gráficas; las gráficas constituyen el primer indicador de la variabilidad de los datos.

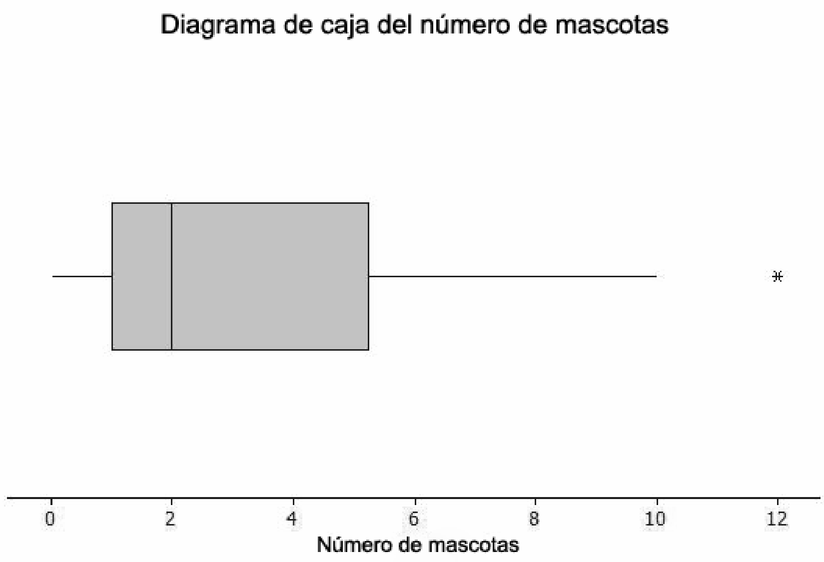
- **Diagrama de puntos:** es un diagrama que representa el valor de cada dato sobre una escala o sobre una recta numérica.



- **HISTOGRAMA:** es una gráfica que agrupa los datos en intervalos y representa los datos de cada intervalo mediante una barra.



- **DIAGRAMA DE CAJA:** es una gráfica que proporciona una imagen de los datos ordenados y divididos en cuatro intervalos, cada uno de los cuales contiene aproximadamente el 25% de los datos.

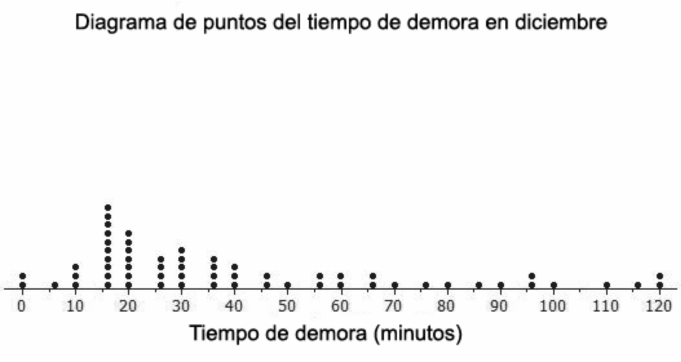


Ejercicios

Responde las preguntas que acompañan a cada gráfica para comenzar a comprender la historia detrás de los datos.

Los funcionarios de transporte recopilan datos sobre las demoras en los vuelos (el número de minutos que transcurren entre la hora programada para el despegue y el momento en que realmente parte el vuelo).

Considera el diagrama de puntos sobre el tiempo de demora para sesenta vuelos de BigAir durante diciembre de 2012.

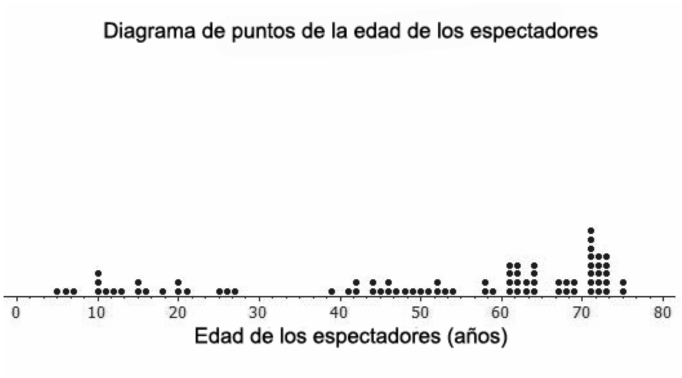


1. ¿Qué crees que nos dice esta gráfica acerca de las demoras de estos sesenta vuelos?

2. ¿Se te ocurre alguna razón por la cual los datos que presenta esta gráfica brindan información importante? ¿A quién podría interesarle la distribución de estos datos?

3. De acuerdo con el trabajo que hiciste anteriormente con los diagramas de puntos, ¿dirías que este diagrama de puntos representa una distribución de datos sesgada o simétrica? (Recuerda que una distribución de datos sesgada no tiene forma de campana). Explica tu respuesta.

Se seleccionó una muestra aleatoria de ochenta espectadores de un programa de televisión. El siguiente diagrama de puntos muestra la distribución de la edad (en años) de estos ochenta espectadores.

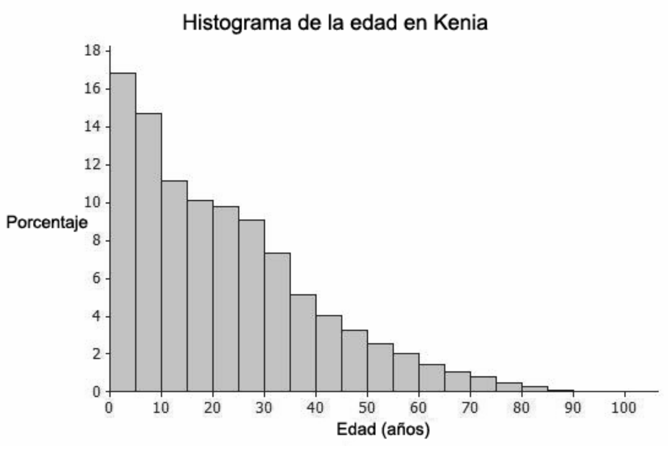


4. ¿Qué crees que nos dice esta gráfica acerca de la edad de los ochenta espectadores de esta muestra?

5. ¿Se te ocurre alguna razón por la cual los datos que presenta esta gráfica brindan información importante? ¿A quién podría interesarle la distribución de estos datos?

6. De acuerdo con el trabajo que hiciste anteriormente con los diagramas de puntos, ¿dirías que este diagrama de puntos representa una distribución de datos sesgada o simétrica? Explica tu respuesta.

El siguiente histograma representa la distribución de edad de la población de Kenia en 2010.

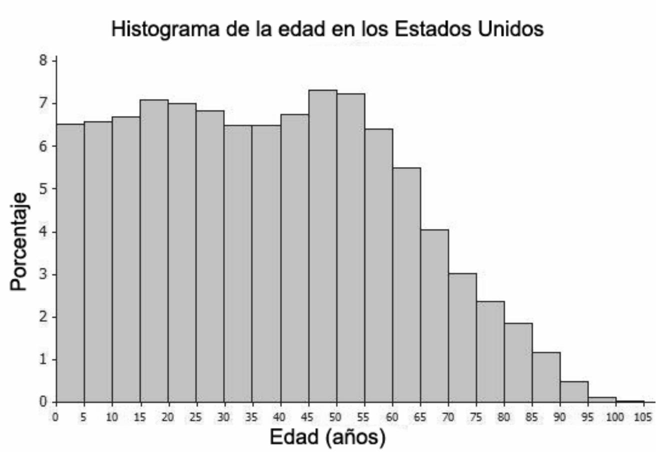


7. ¿Qué crees que nos dice esta gráfica acerca de la población de Kenia?

8. ¿Por qué podríamos querer estudiar los datos que representa esta gráfica?

9. De acuerdo con el trabajo que hiciste anteriormente con los histogramas, ¿dirías que este histograma representa una distribución de datos sesgada o simétrica? Explica tu respuesta.

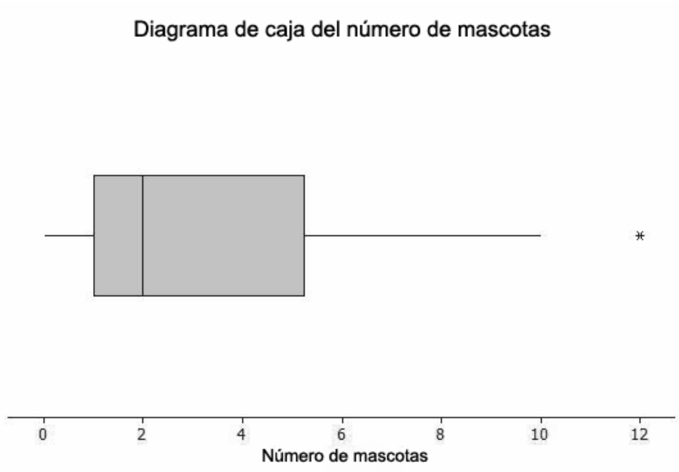
El siguiente histograma representa la distribución de edad de la población de los Estados Unidos en 2010.



10. ¿Qué crees que nos dice esta gráfica acerca de la población de los Estados Unidos?

11. ¿Por qué podríamos querer estudiar los datos que representa esta gráfica?

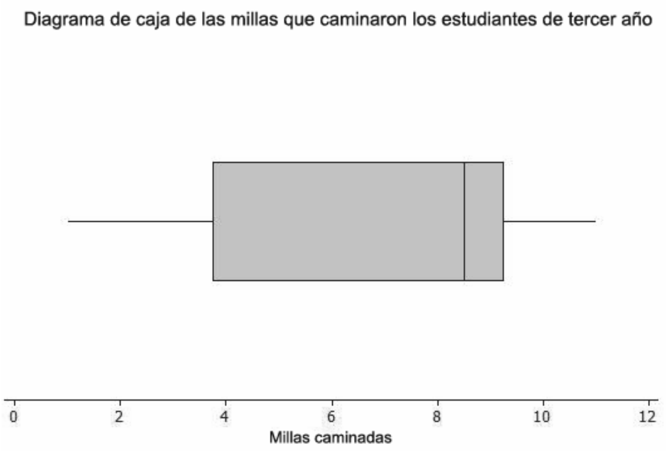
Se les preguntó a treinta estudiantes de la Escuela Secundaria River City cuántas mascotas tienen. A partir de sus respuestas, se creó el siguiente diagrama de caja.



12. ¿Qué nos dice el diagrama de caja acerca del número de mascotas que poseen los treinta estudiantes de la Escuela Secundaria River City?

13. ¿Por qué podría ser importante comprender los datos en los cuales se basa esta gráfica?

Veintidós estudiantes de tercer año de la Escuela Secundaria River City participaron en una caminata destinada a recaudar fondos para la banda de la escuela. A partir del número de millas que caminó cada uno de los veintidós estudiantes, se creó el siguiente diagrama de caja.



14. ¿Qué crees que nos dice el diagrama de caja acerca del número de millas que caminaron los veintidós estudiantes?

15. ¿Por qué podría ser importante comprender los datos en los cuales se basa esta gráfica?

Resumen de la lección

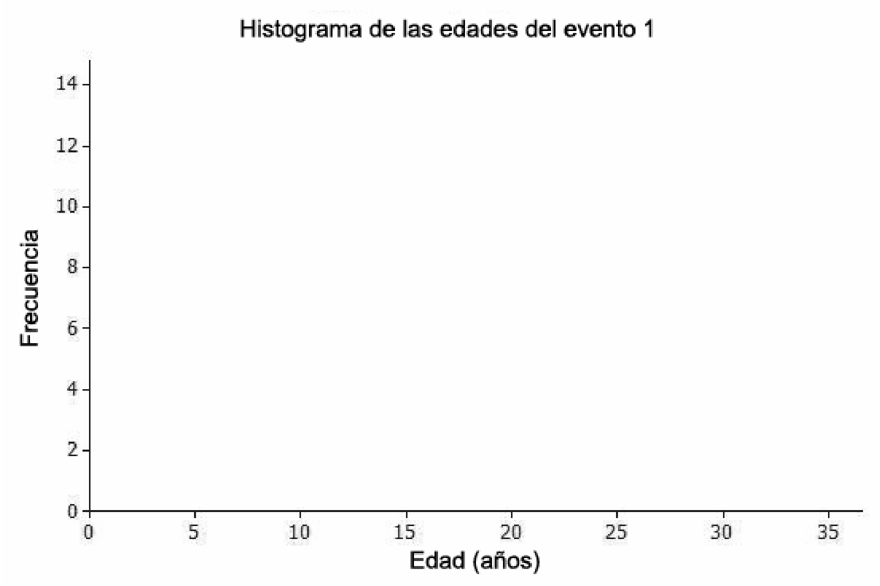
Las estadísticas tratan sobre los datos. Las gráficas brindan una representación de una distribución de datos; se utilizan para comprender dichos datos y para responder preguntas acerca de la distribución.

Grupo de problemas

1. Veinticinco personas asistieron a un evento. Sus edades son las siguientes:

$$3, 3, 4, 4, 4, 4, 5, 6, 6, 6, 6, 6, 6, 6, 7, 7, 7, 7, 7, 7, 16, 17, 22, 22, 25.$$

   a. Utiliza los ejes que se proporcionan para crear un histograma de las edades.



   b. ¿Describirías tu gráfica como simétrica o sesgada? Explica tu elección.

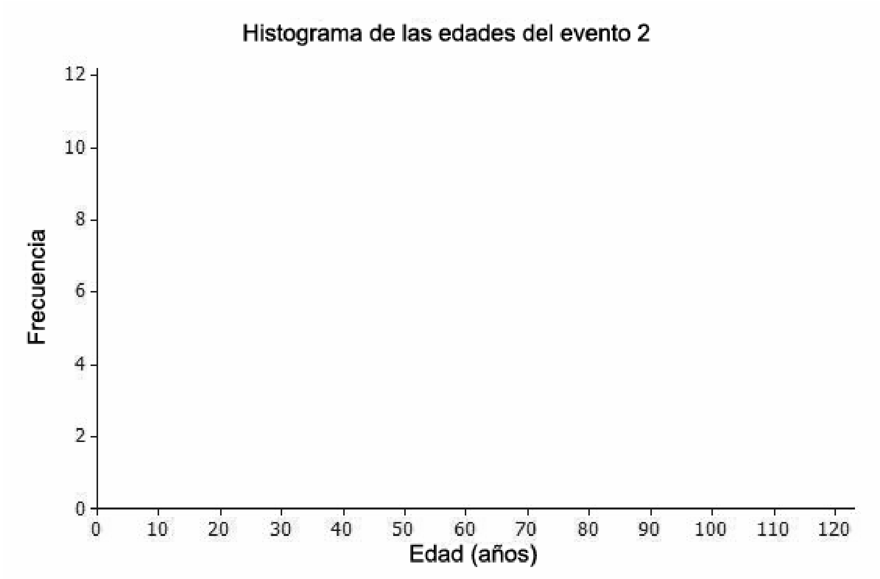
   c. Identifica una edad típica de las veinticinco personas.

   d. ¿A qué evento crees que asistieron estas veinticinco personas? Utiliza tu histograma para justificar tu conjetura.

2. Otras cuarenta personas también asistieron a un evento. Sus edades son las siguientes:

$$6, 13, 24, 27, 28, 32, 32, 34, 38, 42, 42, 43, 48, 49, 49, 49, 51, 52, 52, 53,$$
$$53, 53, 54, 55, 56, 57, 57, 60, 61, 61, 62, 66, 66, 66, 68, 70, 72, 78, 83, 97.$$

a. Utiliza los ejes que se proporcionan para crear un histograma de las edades.



b. ¿Describirías tu gráfica de edades como simétrica o sesgada? Explica tu elección.

c. Identifica una edad típica de las cuarenta personas.

d. ¿A qué evento crees que asistieron estas cuarenta personas? Utiliza tu histograma para justificar tu conjetura.

e. ¿Cómo explicarías las diferencias entre los dos histogramas?

Esta página queda en blanco intencionalmente.

Lección 2: Describir la tendencia central de una distribución

Trabajo en clase

En el trabajo que hiciste anteriormente con las distribuciones de datos, aprendiste cómo derivar la media y la mediana de una distribución de datos. En esta lección, se profundiza tu trabajo previo sobre la tendencia central.

Desafío exploratorio/Ejercicios 1 a 9

Considera los siguientes tres conjuntos de datos.

Conjunto de datos 1: Dueños de mascotas

Se seleccionó aleatoriamente a un grupo de estudiantes de la Escuela Secundaria River City y se les preguntó: “¿Cuántas mascotas posees actualmente?”. A continuación, se indican los resultados registrados.

0	0	0	0	1	1	1	1	1	1	1	1	1	1	2
2	2	2	3	3	4	5	5	6	6	7	8	9	10	12

Conjunto de datos 2: Longitud del pasillo del ala este de la Escuela Secundaria River City

Se seleccionó a veinte estudiantes para que midieran la longitud del pasillo del ala este. Se realizaron dos marcas en el piso del pasillo: una al comienzo y otra al final del pasillo. Se entregó a cada estudiante una regla de un metro y se les pidió que la utilicen para determinar la longitud entre las marcas, redondeando a la décima de metro más cercana. A continuación, se indican los resultados registrados.

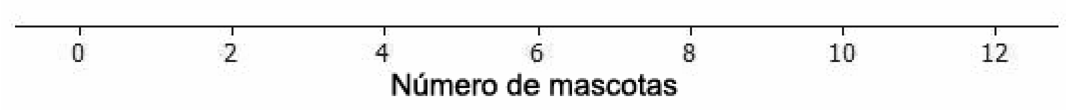
8.2	8.3	8.3	8.4	8.4	8.5	8.5	8.5	8.5	8.5
8.6	8.6	8.6	8.6	8.7	8.7	8.8	8.8	8.9	8.9

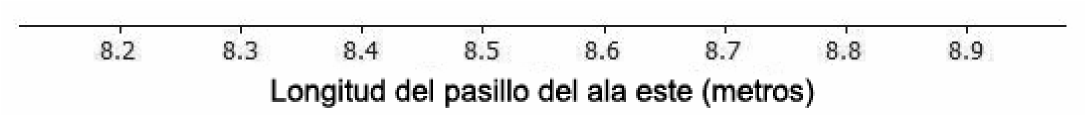
Conjunto de datos 3: Antigüedad de los carros

Se preguntó a veinticinco dueños la antigüedad de sus carros en años. A continuación, se indican los resultados registrados.

0	1	2	2	3	4	5	5	6	6	6	7	7
7	7	7	7	8	8	8	8	8	8	8	8	

1. Crea un diagrama de puntos con cada conjunto de datos. Utiliza las siguientes escalas.





2. Calcula la media del número de mascotas que poseen los treinta estudiantes de la Escuela Secundaria River City. Calcula la mediana del número de mascotas que poseen los treinta estudiantes.

3. ¿Cuál crees que es un número típico de mascotas para los estudiantes de la Escuela Secundaria River City? Explica cómo hiciste tu estimación.

4. ¿Por qué crees que distintos estudiantes obtuvieron diferentes resultados cuando midieron la misma distancia del pasillo del ala este?

5. ¿Cuál es la media de la longitud del conjunto de datos del pasillo del ala este? ¿Cuál es la mediana?

6. Una empresa de construcción va a instalar un pasamanos a lo largo de una pared desde el punto donde comienza hasta el punto donde termina el pasillo del ala este. La empresa te pregunta cuál debería ser la longitud del pasamanos. ¿Qué responderías? Explica tu respuesta.

7. Describe la distribución de la antigüedad de los carros.

8. ¿Cuál es la media de la antigüedad de los veinticinco carros? ¿Cuál es la mediana? ¿Por qué son diferentes la media y la mediana?

9. ¿Qué número utilizarías como estimación de la antigüedad típica de un carro para los veinticinco dueños? Explica tu respuesta.

Resumen de la lección

- Un diagrama de puntos brinda una representación gráfica de una distribución de datos y nos ayuda a visualizar dicha distribución.
- La media y la mediana de la distribución son resúmenes numéricos de la tendencia central de una distribución de datos.
- Cuando la distribución es casi simétrica, la media y la mediana de la distribución son aproximadamente iguales. Cuando la distribución no es simétrica (en cuyo caso suele describirse como “sesgada”), la media y la mediana no son iguales.
- En el caso de las distribuciones simétricas, la media es una elección apropiada para describir un valor típico de la distribución. En el caso de las distribuciones de datos sesgadas, la mediana constituye una mejor descripción de un valor típico.

Grupo de problemas

Considera el siguiente escenario. La empresa que creó el popular videojuego “Líderes” planea lanzar una importante actualización del juego. Los usuarios ganan o pierden puntos por tomar decisiones como líderes de un país imaginario. En la mayoría de los casos, al jugar repetidas veces, el usuario mejora su capacidad para tomar decisiones. La empresa lanzará una campaña publicitaria en línea pero, por el momento, no están seguros de cómo enfocar la publicidad. Tu objetivo es ayudar a la empresa a decidir cómo enfocar la campaña publicitaria. Se han propuesto cinco videos para los siguientes públicos objetivo:

Video 1: Mujeres con puntuación de nivel inicial

Video 2: Varones con puntuación de nivel avanzado

Video 3: Todos lo usuarios con puntuación de nivel medio

Video 4: Varones con puntuación de nivel inicial

Video 5: Mujeres con puntuación de nivel avanzado

1. ¿Por qué podría tener interés la empresa en desarrollar diferentes videos en función de la puntuación de los usuarios?

2. Se seleccionó aleatoriamente a treinta usuarias mujeres y a veinticinco usuarios varones de una base de datos compuesta por personas que juegan regularmente a ese juego. Todos ellos accedieron a formar parte de un estudio de investigación y a informar de su puntuación. La puntuación se basa en las respuestas del jugador a preguntas sobre liderazgo. Una puntuación de 1 a 40 se considera nivel inicial de liderazgo; de 41 a 60, nivel medio, y más de 60, nivel avanzado.

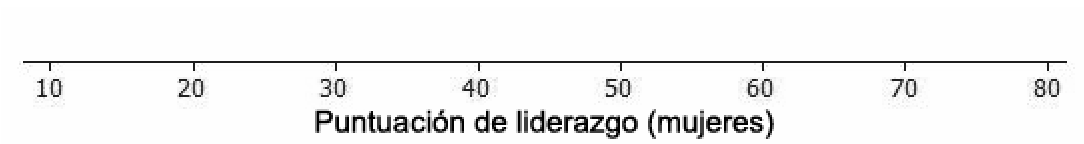
   Utiliza los siguientes datos para crear un diagrama de puntos con las puntuaciones de las mujeres, otro con las puntuaciones de los varones y otro con las puntuaciones de los dos grupos, varones y mujeres, combinados.

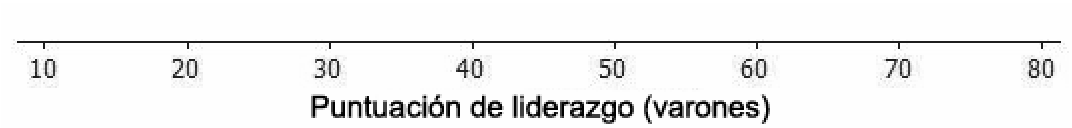
   Puntuaciones de las mujeres:

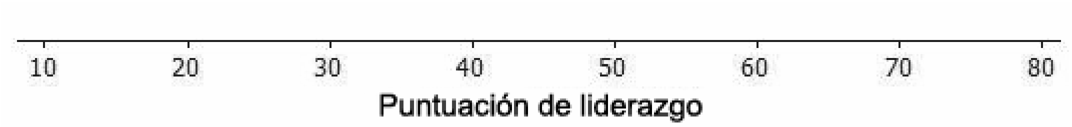
10	20	20	20	30	30	30	40	40	40
50	50	55	65	65	65	65	65	70	70
70	70	76	76	76	76	76	76	76	76

   Puntuaciones de los varones:

15	20	20	25	25	25	25	30	30	30
30	30	30	35	35	35	35	35	40	40
40	45	45	45	50					







3. ¿Cuál crees que es una puntuación típica para una usuaria mujer? ¿Cuál crees que es una puntuación típica para un usuario varón? Explica cómo determinaste estas puntuaciones típicas.

4. ¿Por qué es más difícil informar de una puntuación típica para el grupo general que incluye a varones y mujeres?

5. Los costos de producción solo permitirán desarrollar dos videos publicitarios. ¿Qué dos videos recomendarías desarrollar? Explica tus recomendaciones.

Esta página queda en blanco intencionalmente.

Lección 3: Estimar las tendencias centrales e interpretar la media como un punto de equilibrio

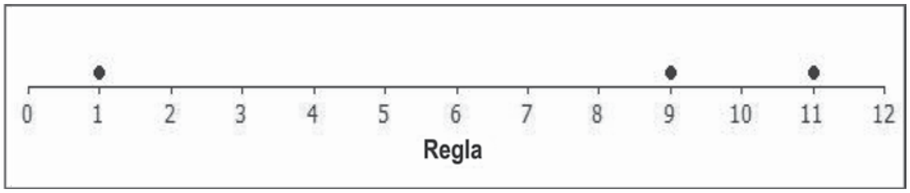
Trabajo en clase

Ejemplo

En el trabajo que hiciste anteriormente en la clase de Matemáticas, estimaste un punto de equilibrio de una distribución de datos. Vamos a repasar lo que aprendimos acerca del punto de equilibrio de una distribución. Una regla de 12 pulgadas tiene varias monedas de 25 centavos pegadas en distintas posiciones. Se coloca el lado ancho de un lápiz debajo de la regla para determinar un punto de equilibrio aproximado de la regla con las monedas.

Ejercicios 1 a 7

Considera el siguiente ejemplo de monedas de 25 centavos pegadas a una regla liviana.



1. Sam pegó 3 monedas a su regla. Las pegó en las siguientes posiciones: 1 pulgada, 9 pulgadas y 11 pulgadas. Si colocara el lápiz debajo de la posición de 5 pulgadas, ¿crees que la regla quedaría en equilibrio? ¿Por qué sí o por qué no?

2. Si la regla no quedara en equilibrio, ¿moverías el lápiz hacia la izquierda o hacia la derecha de las 5 pulgadas para lograr el equilibrio? Explica tu respuesta.

3. Estima un punto de equilibrio para la regla. Completa la siguiente tabla en función de la posición que elegiste.

Posición de la moneda	Distancia desde la moneda hasta el punto de equilibrio estimado
1	
9	
11	

4. ¿Cuál es la suma de las distancias a la derecha del punto de equilibrio que estimaste?

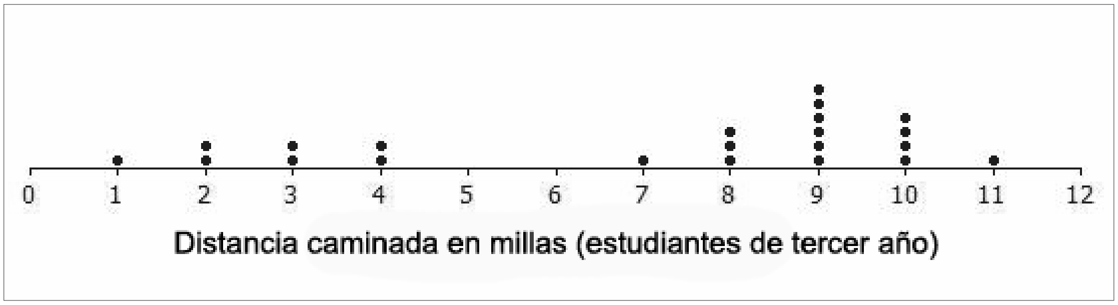
5. ¿Cuál es la suma de las distancias a la izquierda del punto de equilibrio que estimaste?

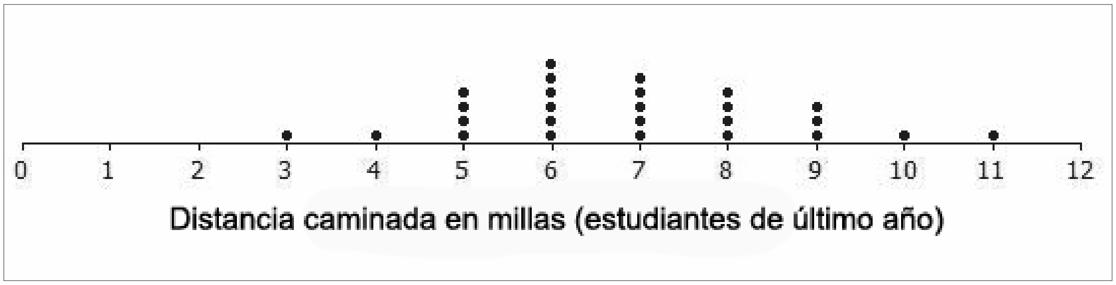
6. ¿Necesitas ajustar la posición de tu punto de equilibrio? Si es así, explica cómo.

7. Calcula la media y la mediana de la posición de las monedas de 25 centavos. ¿La media o la mediana de las posiciones proporciona una mejor estimación del punto de equilibrio para la posición de las 3 monedas pegadas a esta regla? Explica por qué elegiste esa medida.

Ejercicios 8 a 20

Veintidós estudiantes de tercer año y veintiséis estudiantes de último año de la Escuela Secundaria River City participaron en una caminata destinada a recaudar fondos para la banda de la escuela. A continuación figuran los diagramas de puntos que indican las distancias en millas que caminaron los estudiantes de cada clase.





8. Estima la media del número de millas que caminó un estudiante de tercer año y marca el valor con una X en el diagrama de puntos de ese grupo. ¿Cómo estimaste esta posición?

9. ¿Cuál es la mediana de la distribución de datos de los estudiantes de tercer año?

10. ¿La media del número de millas que caminó un estudiante de tercer año es menor que, aproximadamente igual a o mayor que la mediana? Si son diferentes, explica por qué. Si son aproximadamente iguales, explica por qué.

11. ¿Cómo describirías el número típico de millas que caminó un estudiante de tercer año en esta caminata?

12. Estima la media del número de millas que caminó un estudiante de último año y marca el valor con una X en el diagrama de puntos de ese grupo. ¿Cómo estimaste esta posición?

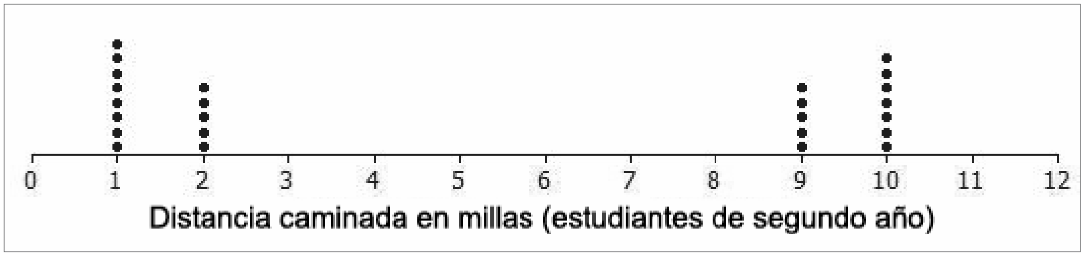
13. ¿Cuál es la mediana de la distribución de datos de los estudiantes de último año?

14. Estima la media y la mediana del número de millas que caminaron los estudiantes de último año. ¿Tu estimación de la media del número de millas que caminó un estudiante de último año es menor que, aproximadamente igual a o mayor que la mediana? Si son diferentes, explica por qué. Si son aproximadamente iguales, explica por qué.

15. ¿Cómo describirías el número típico de millas que caminó un estudiante de último año en esta caminata?

16. Un estudiante de tercer año de la Escuela Secundaria River City indicó que el número de millas que caminó un estudiante típico de tercer año fue mayor que el número de millas que caminó un estudiante típico de último año. ¿Estás de acuerdo? Explica tu respuesta.

Por último, los veinticinco estudiantes de segundo año que participaron en la caminata informaron de sus resultados. A continuación, se muestra un diagrama de puntos.



17. ¿Cuál es la diferencia entre la distribución de datos de los estudiantes de segundo año y las distribuciones de datos de los estudiantes de tercer año y de último año?

18. Estima el punto de equilibrio de la distribución de datos de los estudiantes de segundo año.

19. ¿Cuál es la mediana del número de millas que caminó un estudiante de segundo año?

20. ¿Cómo describirías la distribución de datos de los estudiantes de segundo año?

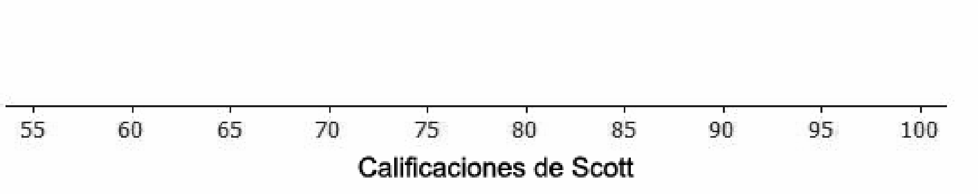
Resumen de la lección

La media de una distribución de datos representa un punto de equilibrio para esa distribución. La suma de las distancias a la derecha de la media es igual a la suma de las distancias a la izquierda de la media.

Grupo de problemas

Considera otro ejemplo de equilibrio. El señor Jackson es profesor de Matemáticas en la Escuela Secundaria Waldo. Los estudiantes de su clase realizan con frecuencia pruebas o exámenes. El señor Jackson indicó a sus estudiantes que un examen tiene un valor equivalente a 4 pruebas cuando se calcula un promedio ponderado general para determinar su calificación final. Durante un período de calificación, Scott obtuvo una calificación del 80% en un examen, del 90% en un segundo examen, del 60% en una prueba y del 70% en otra prueba.

¿Cómo podríamos representar las calificaciones de Scott? Considera la siguiente recta numérica.



1. ¿Qué valores están representados en la recta numérica?

2. Si se utiliza un símbolo "•" para representar la calificación en una prueba, ¿cómo podrías representar la calificación en un examen?

3. Utiliza símbolos "•" para representar los exámenes y las pruebas de Scott en esta recta numérica.

4. El señor Jackson indicó que los estudiantes deberían fijarse como objetivo un promedio ponderado general del 85%. ¿Crees que Scott alcanzó ese objetivo? Explica tu respuesta.

5. Coloca una X en la recta numérica en la posición en la que creas que debe estar el punto de equilibrio de todos los símbolos "•". Determina la suma de las distancias desde la X hasta cada "•" del lado derecho de la X.

6. Determina la suma de las distancias desde la X hasta cada "•" del lado izquierdo de la X.

7. ¿La distancia total del lado derecho de la X es igual a la distancia total del lado izquierdo de la X?

8. De acuerdo con tu respuesta al Problema 7, ¿cambiarías tu estimación del punto de equilibrio? De ser así, ¿dónde colocarías tu punto de equilibrio corregido? ¿De qué manera el uso de este punto corregido modifica la distancia total a la derecha de tu estimación y la distancia total a la izquierda?

9. El promedio ponderado de Scott es 81. Recuerda que la calificación de cada examen es igual a 4 veces la calificación de una prueba. Muestra los cálculos que dan como resultado este promedio ponderado.

10. ¿Cómo se compara la calificación media que calculaste con el punto de equilibrio estimado?

11. Calcula la distancia total a la derecha de la media y la distancia total a la izquierda de la media. ¿Qué observas?

12. ¿Alcanzó Scott el objetivo que estableció el señor Jackson de obtener un promedio del 85%? Explica tu respuesta.

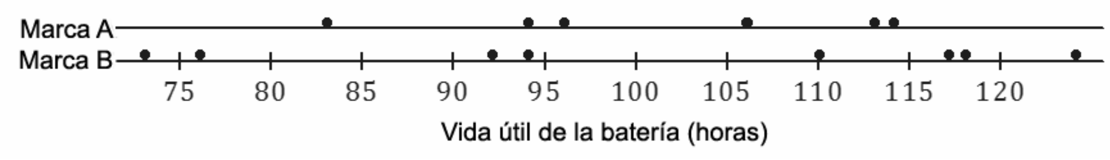
Esta página queda en blanco intencionalmente.

Lección 4: Resumir las desviaciones respecto a la media

Trabajo en clase

Ejercicios 1 a 4

Una organización de consumidores está planeando un estudio de las diversas marcas de baterías disponibles en el mercado. Como parte de la planificación, se mide la vida útil de seis baterías de Marca A y de ocho baterías de Marca B (es decir, cuánto tiempo pueden utilizarse antes de que sea necesario reemplazarlas). A continuación, se muestran diagramas de puntos que representan la vida útil de las baterías de cada marca.



1. ¿Alguna de las marcas tiende a durar más o duran aproximadamente lo mismo? ¿Qué cálculos podrías hacer para comparar la vida útil de las baterías de estas dos marcas?

2. ¿La diferencia en la vida útil de una batería a otra es mayor en la Marca A o en la Marca B?

3. ¿Preferirías una marca de baterías en la que la vida útil no varíe mucho de una batería a otra de la misma marca? ¿Por qué sí o por qué no?

La siguiente tabla muestra la vida útil (en horas) de las baterías de la Marca A.

Vida útil (horas)	83	94	96	106	113	114
Desviación respecto a la media						

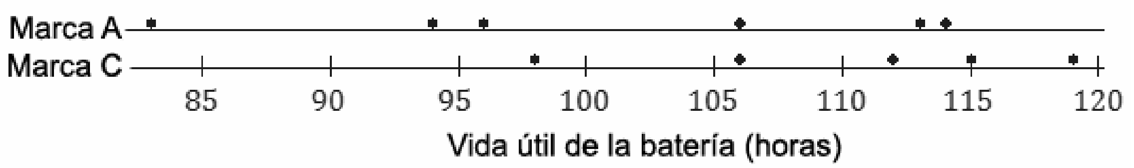
4. Calcula las desviaciones respecto a la media de los valores restantes y anota tus respuestas en los lugares correspondientes de la tabla.

La siguiente tabla muestra la vida útil y las desviaciones respecto a la media de la Marca B.

Vida útil (horas)	73	76	92	94	110	117	118	124
Desviación respecto a la media	-27.5	-24.5	-8.5	-6.5	9.5	16.5	17.5	23.5

Ejercicios 5 a 10

Se determinó la vida útil de cinco baterías de una tercera marca, la Marca C. El siguiente diagrama de puntos muestra la vida útil de las baterías de la Marca A y de la Marca C.



5. ¿Cuál de las marcas tiene la mayor media de la vida útil de las baterías? (Deberías poder responder esta pregunta sin hacer ningún cálculo).

6. ¿Cuál de las marcas muestra mayor variabilidad?

EUREKA MATH

7. ¿Cuál de las marcas esperarías que tuviera mayores desviaciones respecto a la media (sin tener en cuenta los signos de las desviaciones)?

La siguiente tabla muestra la vida útil de las baterías de la Marca C.

Vida útil (horas)	115	119	112	98	106
Desviación respecto a la media					

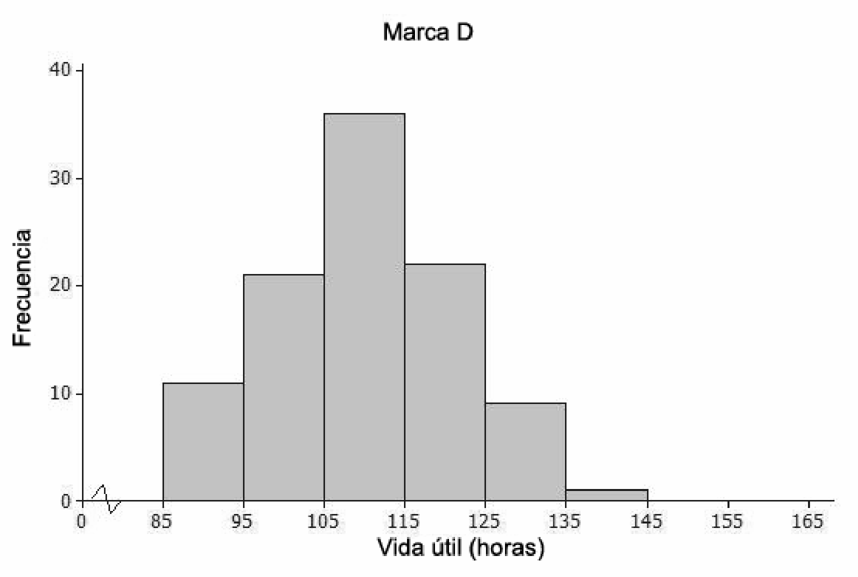
8. Calcula la media de la vida útil de las baterías de la Marca C. (Asegúrate de incluir una unidad en tu respuesta).

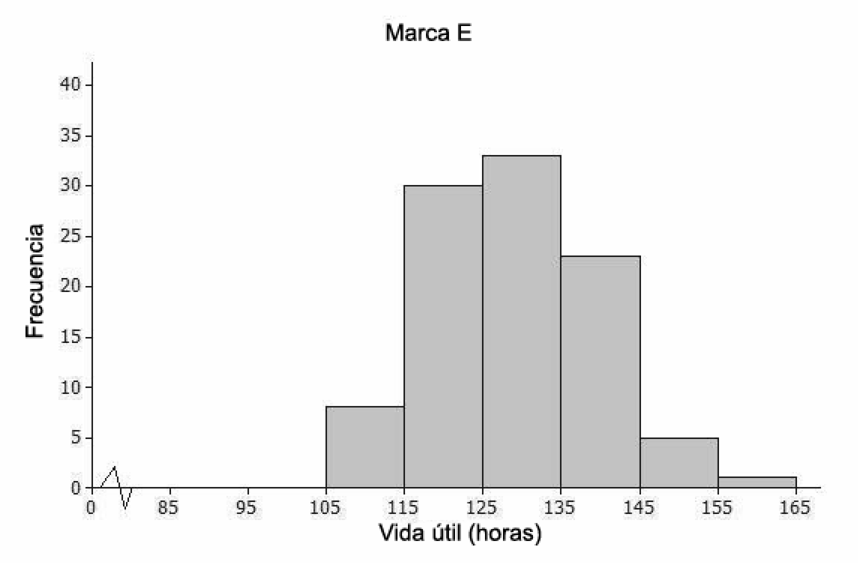
9. Escribe las desviaciones respecto a la media en las celdas vacías de la tabla para la Marca C.

10. Si no se toman en cuenta los signos, ¿las desviaciones respecto a la media son mayores, en general, para la Marca A o para la Marca C? ¿Tu respuesta concuerda con la respuesta al Ejercicio 7?

Ejercicios 11 a 15

Se determinó la vida útil de 100 baterías de la Marca D y 100 baterías de la Marca E. Los resultados se resumen en los siguientes histogramas.





EUREKA MATH

11. Estima la media de la vida útil para la Marca D. (No hagas cálculos).

12. Estima la media de la vida útil para la Marca E. (No hagas cálculos).

13. ¿Cuál de las marcas muestra mayor variabilidad en la vida útil de sus baterías: la D o la E? ¿Crees que las dos marcas son aproximadamente iguales en este sentido?

14. Estima la mayor desviación respecto a la media para la Marca D.

15. ¿Cuál considerarías que es una desviación típica respecto a la media para la Marca D?

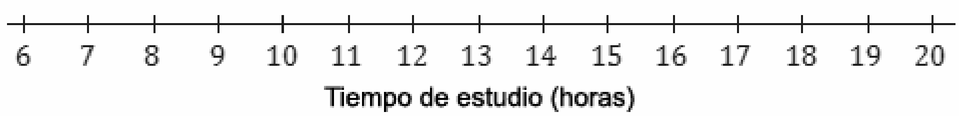
Resumen de la lección

- Para cualquier valor dado de un conjunto de datos, la desviación respecto a la media es ese valor menos la media. Si lo escribimos de forma algebraica, esto es: $x - \bar{x}$.
- Cuanto mayor sea la variabilidad (dispersión) de la distribución, mayores serán las desviaciones respecto a la media (sin tener en cuenta los signos de las desviaciones).

Grupo de problemas

1. A diez miembros del equipo de baloncesto femenino de una escuela secundaria se les preguntó cuántas horas estudian en una semana típica. Sus respuestas (en horas) fueron 20, 13, 10, 6, 13, 10, 13, 11, 11, 10.

   a. Utiliza el siguiente eje para dibujar un diagrama de puntos de estos valores. (Recuerda que, cuando hay valores repetidos, debes apilar los puntos uno sobre otro).

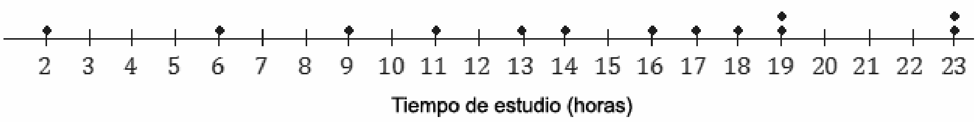


   b. Calcula la media del tiempo de estudio de estas estudiantes.

   c. Calcula las desviaciones respecto a la media para estos tiempos de estudio y anota tus respuestas en los lugares correspondientes de la tabla.

Número de horas de estudio	20	13	10	6	13	10	13	11	11	10
Desviación respecto a la media										

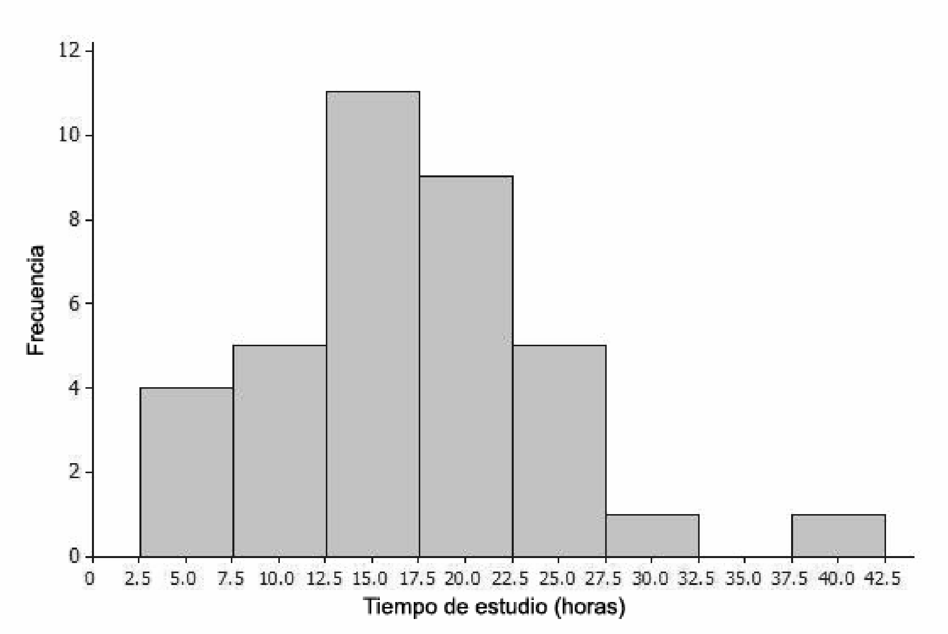
   d. En el siguiente diagrama de puntos, se muestran los tiempos de estudio de catorce chicas del equipo de fútbol de la misma escuela del ejercicio anterior.



De acuerdo con los datos, ¿las desviaciones respecto a la media (sin tener en cuenta el signo de las desviaciones) serán mayores o menores para las jugadoras de fútbol que para las jugadoras de baloncesto?

2. A todos los miembros del equipo de sóftbol de una escuela secundaria se les preguntó cuántas horas estudian en una semana típica. Los resultados se muestran en el siguiente histograma.

   (El conjunto de datos de esta pregunta proviene de las Herramientas de matemáticas del Consejo Nacional de Profesores de Matemáticas, NCTM, por sus siglas en inglés: http://www.nctm.org/Classroom-Resources/Core-Math-Tools/Data-Sets/).



   a. Según podemos ver en el histograma, cuatro estudiantes estudiaron unas 5 horas por semana. ¿Cuántos estudiantes estudiaron unas 15 horas por semana?
   b. ¿Cuántos estudiantes había en total?
   c. Supón que los cuatro estudiantes que se representan en la barra del histograma centrada en el 5 habían estudiado exactamente 5 horas, que los cinco estudiantes que se representan en la barra siguiente habían estudiado exactamente 10 horas, y así sucesivamente. Si sumaras los tiempos de estudio de todos los estudiantes, ¿qué resultado obtendrías?
   d. ¿Cuál es la media del tiempo de estudio de estos estudiantes?
   e. ¿Cuál considerarías que es una desviación típica respecto a la media para este conjunto de datos?

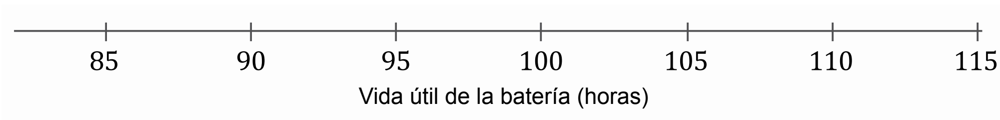
Esta página queda en blanco intencionalmente.

Lección 5: Medir la variabilidad de las distribuciones simétricas

Trabajo en clase

Ejemplo 1: Calcular la desviación estándar

Aquí se muestra un diagrama de puntos que representa la vida útil de las baterías de la Marca A de la Lección 4.



¿Cómo mides la variabilidad de este conjunto de datos? Una manera es calcular la **desviación estándar**.

- En primer lugar, halla cada desviación respecto a la media.
- Luego, eleva al cuadrado las desviaciones respecto a la media. Por ejemplo, cuando la desviación respecto a la media es -18, la desviación respecto a la media elevada al cuadrado es $(-18)^2 = 324$.

Vida útil (horas)	83	94	96	106	113	114
Desviación respecto a la media	-18	-7	-5	5	12	13
Desviación respecto a la media elevada al cuadrado	324	49	25	25	144	169

- Suma las desviaciones elevadas al cuadrado:

  $324 + 49 + 25 + 25 + 144 + 169 = 736$.

  Este resultado es la *suma* de las desviaciones elevadas al cuadrado.

El número de valores del conjunto de datos se denota mediante n. En este ejemplo, n es 6.

- Divide la suma de las desviaciones elevadas al cuadrado entre $n - 1$, que en este caso es $6 - 1$, es decir, 5.

$$\frac{736}{5} = 147.25$$

- Por último, halla la raíz cuadrada de 147.2. Redondeando a la centésima más cercana, esto da como resultado 12.13.

¡Esa es la desviación estándar! Al principio parece un proceso muy complicado, pero pronto te acostumbrarás a él.

Concluimos que una desviación típica de la vida útil de una batería de la Marca A respecto a la vida útil media de una batería de la Marca A es 12.13 horas. La unidad de la desviación estándar es siempre la misma que la unidad del conjunto de datos original. Por lo tanto, la desviación estándar a la centésima más cercana, con la unidad, es 12.13 horas. ¿Qué tan cercana es la respuesta a la desviación típica que estimaste al principio de la lección?

Ejercicios 1 a 5

Ahora, puedes calcular la desviación estándar de la vida útil de las ocho baterías de la Marca B. La media era 100.5. Ya tenemos las desviaciones respecto a la media.

Vida útil (horas)	73	76	92	94	110	117	118	124
Desviación respecto a la media	-27.5	-24.5	-8.5	-6.5	9.5	16.5	17.5	23.5
Desviación respecto a la media elevada al cuadrado								

1. Escribe en la tabla las desviaciones elevadas al cuadrado.

2. Suma las desviaciones elevadas al cuadrado. ¿Qué resultado obtienes?

3. ¿Cuál es el valor de n para este conjunto de datos? Divide la suma de las desviaciones elevadas al cuadrado entre $n-1$ y escribe tu respuesta en el siguiente espacio. Redondea tu respuesta a la milésima más cercana.

4. Calcula la raíz cuadrada para hallar la desviación estándar. Redondea tu respuesta a la centésima más cercana.

5. ¿Cómo interpretas la desviación estándar que hallaste en el Ejercicio 4? (Recuerda dar tu respuesta en el contexto de esta pregunta. Interpreta tu respuesta redondeada a la centésima más cercana).

Ejercicios 6 y 7

Jenna compró un carro híbrido nuevo. Cada semana, durante un período de siete semanas, registró la eficiencia del combustible (en millas por galón) de su carro. A continuación, se muestran los resultados.

$$45 \quad 44 \quad 43 \quad 44 \quad 45 \quad 44 \quad 43$$

6. Calcula la desviación estándar de estos resultados a la centésima más cercana. Asegúrate de mostrar el proceso.

7. ¿Qué significa la desviación estándar que hallaste en el Ejercicio 6?

Resumen de la lección

- La desviación estándar mide una desviación típica respecto a la media.
- Para calcular la desviación estándar:
  1. halla la media del conjunto de datos;
  2. calcula las desviaciones respecto a la media;
  3. eleva al cuadrado las desviaciones respecto a la media;
  4. suma las desviaciones elevadas al cuadrado;
  5. divide entre $n - 1$ (si trabajas con datos de una muestra, que es el caso más común);
  6. halla la raíz cuadrada.
- La unidad de la desviación estándar es siempre la misma que la unidad del conjunto de datos original.
- Cuanto mayor es la desviación estándar, mayor es la dispersión (variabilidad) del conjunto de datos.

Grupo de problemas

1. Una tienda pequeña de venta de carros está comprobando la eficiencia del combustible de los modelos sedán que tiene en su propiedad. La tienda elige 12 sedanes para la prueba. En la siguiente tabla, se muestran los valores de eficiencia del combustible (mi/gal) de los carros. Completa la tabla como se indica a continuación.

Eficiencia del combustible (millas por galón)	29	35	24	25	21	21	18	28	31	26	26	22
Desviación respecto a la media												
Desviación respecto a la media elevada al cuadrado												

   a. Calcula la eficiencia media del combustible para estos carros.
   b. Calcula las desviaciones respecto a la media y escribe tus respuestas en la segunda fila de la tabla.
   c. Eleva al cuadrado las desviaciones respecto a la media y escribe las desviaciones elevadas al cuadrado en la tercera fila de la tabla.
   d. Halla la suma de las desviaciones elevadas al cuadrado.
   e. ¿Cuál es el valor de n para este conjunto de datos? Divide la suma de las desviaciones elevadas al cuadrado entre $n - 1$.
   f. Calcula la raíz cuadrada de tu respuesta a la parte (e) para hallar la desviación estándar de la eficiencia del combustible para estos carros. Redondea tu respuesta a la centésima más cercana.

2. La misma tienda de venta de carros decide comprobar la eficiencia del combustible de los todoterrenos. La tienda selecciona para la prueba seis todoterrenos que tiene en su propiedad. A continuación, se muestran los valores de eficiencia del combustible (en millas por galón) de los carros.

$$21 \quad 21 \quad 21 \quad 30 \quad 28 \quad 24$$

Calcula la media y la desviación estándar de estos valores. Asegúrate de mostrar tu trabajo e incluye una unidad en tu respuesta.

3. Considera las siguientes preguntas sobre los carros que se describen en los Problemas 1 y 2.

   a. ¿Cuál es la desviación estándar de la eficiencia del combustible de los carros del Problema 1? Explica qué te indica este valor.

   b. También calculaste la desviación estándar de la eficiencia del combustible para los carros del Problema 2. ¿Cuál de los dos conjuntos de datos (Problema 1 o Problema 2) tiene la mayor desviación estándar? ¿Qué te indica esto sobre los dos tipos de carros (sedanes y todoterrenos)?

Esta página queda en blanco intencionalmente.

Lección 6: Interpretar la desviación estándar

Trabajo en clase

Ejemplo 1

Tu maestro te mostrará cómo usar una calculadora para hallar la media y la desviación estándar del siguiente conjunto de datos.

Un grupo de ocho hombres tienen las alturas (en pulgadas) que se muestran a continuación.

$$67.0 \quad 70.9 \quad 67.6 \quad 69.8 \quad 69.7 \quad 70.9 \quad 68.7 \quad 67.2$$

Indica la media y la desviación estándar que obtuviste con la calculadora a la centésima más cercana.

Media: ____________________

Desviación estándar: ____________________

Ejercicio 1

1. A continuación, se muestran las alturas (en pulgadas) de nueve mujeres.

$$68.4 \quad 70.9 \quad 67.4 \quad 67.7 \quad 67.1 \quad 69.2 \quad 66.0 \quad 70.3 \quad 67.6$$

Utiliza las herramientas de estadísticas de tu calculadora o de tu programa informático para hallar la media y la desviación estándar de estas alturas a la centésima más cercana.

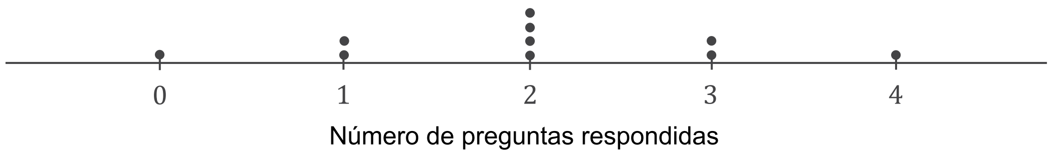
Media: ____________________

Desviación estándar: ____________________

Desafío exploratorio/Ejercicios 2 a 5

2. Un grupo de personas asistió a una charla en un congreso. Al final de la charla, diez de los asistentes recibieron un cuestionario que consistía en cuatro preguntas. Las preguntas eran opcionales, así que es posible que algunos asistentes no respondieran ninguna pregunta, mientras que otros respondieran 1, 2, 3 o las 4 preguntas (por lo tanto, los números posibles de preguntas respondidas son 0, 1, 2, 3 y 4).

   Supón que los números de preguntas que respondieron cada una de las diez personas se muestran en el siguiente diagrama de puntos.

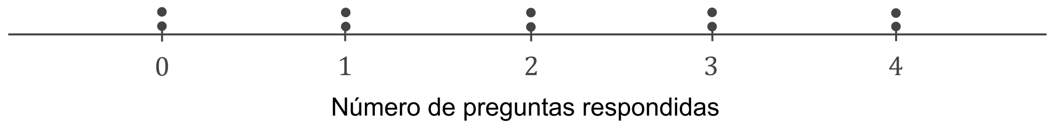


   Utiliza las herramientas de estadísticas de tu calculadora para hallar la media y la desviación estándar del conjunto de datos.

   Media: ______________________

   Desviación estándar: ______________________

3. Supón que el diagrama de puntos tuviera el siguiente aspecto:

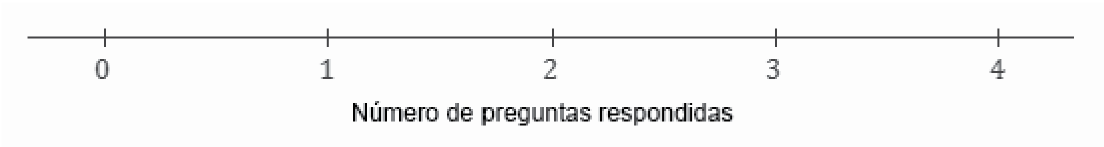


   a. Usa la calculadora para hallar la media y la desviación estándar de esta distribución.

   b. Recuerda que el tamaño de la desviación estándar está relacionado con el tamaño de las desviaciones respecto a la media. Explica por qué la desviación estándar de esta distribución es mayor que la desviación estándar del Ejercicio 2.

4. Supón que las diez personas respondieron las cuatro preguntas del cuestionario.

   a. ¿Cómo se vería el diagrama de puntos?

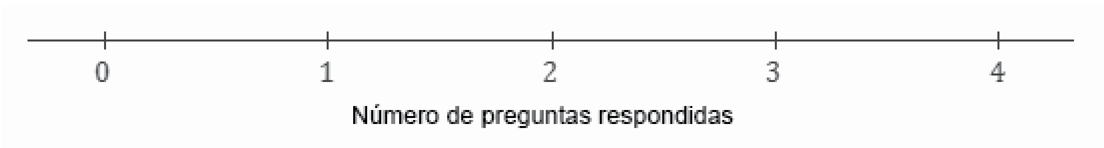


   b. ¿Cuál es la media del número de preguntas respondidas? (Deberías poder responder esta pregunta sin hacer ningún cálculo).

   c. ¿Cuál es la desviación estándar? (¡Aquí tampoco hagas ningún cálculo!).

5. Continúa pensando en la situación que se describió anteriormente, para la cual se registró el número de preguntas que respondió cada una de las diez personas.

   a. Dibuja el diagrama de puntos de la distribución de los posibles valores de los datos que tiene la mayor desviación estándar posible. (Había diez personas en la charla, así que debería haber diez puntos en tu diagrama de puntos). Utiliza la escala que se proporciona a continuación.



   b. Explica por qué la distribución que dibujaste tiene una desviación estándar mayor que la distribución del Ejercicio 4.

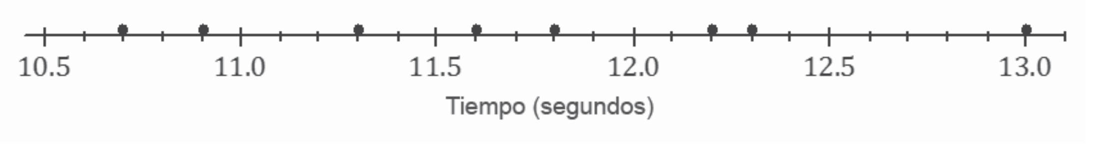
Resumen de la lección

- La media y la desviación estándar de un conjunto de datos se pueden hallar directamente utilizando las herramientas de estadísticas de una calculadora.
- El tamaño de la desviación estándar está relacionado con los tamaños de las desviaciones respecto a la media. Por lo tanto, la desviación estándar se minimiza cuando todos los números del conjunto de datos son iguales y se maximiza cuando las desviaciones respecto a la media son lo más grandes posible.

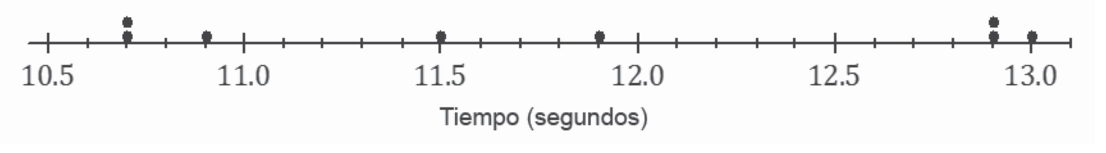
Grupo de problemas

1. En una pista de atletismo, se celebran tres carreras masculinas de 100 m. Se registran los tiempos de ocho de los corredores a la $\frac{1}{10}$ de segundo más cercana. En el siguiente diagrama de puntos se muestran los resultados de las tres carreras de estos ocho corredores.

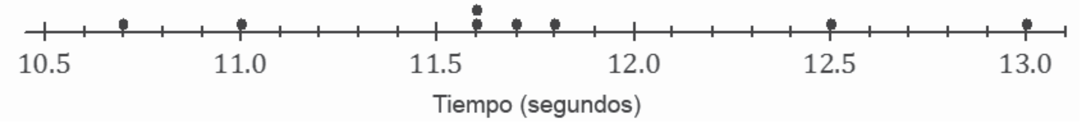
Carrera 1



Carrera 2



Carrera 3



a. Recuerda que el tamaño de la desviación estándar está relacionado con los tamaños de las desviaciones respecto a la media. Sin hacer cálculos, indica cuál de las tres carreras tiene la menor desviación estándar de los tiempos. Justifica tu respuesta.

b. ¿Qué carrera tuvo la mayor desviación estándar de los tiempos? (¡Aquí tampoco hagas ningún cálculo!). Justifica tu respuesta.

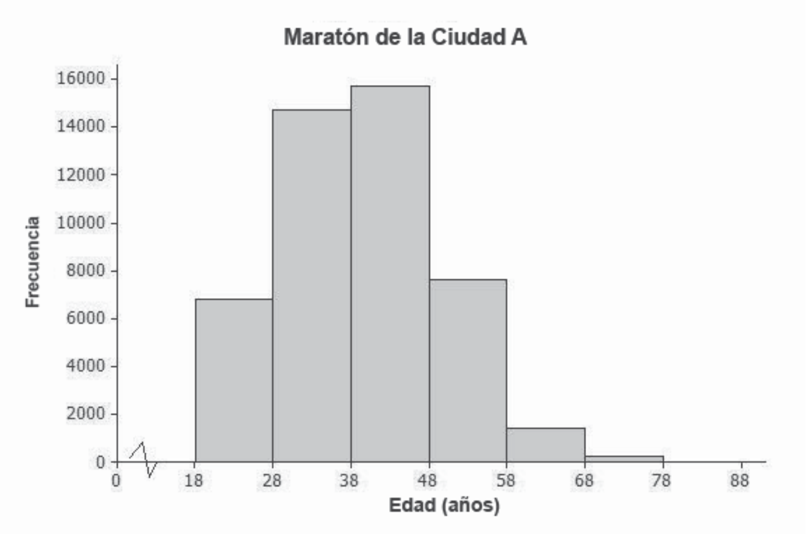
c. ¿Cuál sería aproximadamente la desviación estándar en la Carrera 1? (Recuerda que la desviación estándar es una desviación típica respecto a la media. Por tanto, aquí tienes que hallar una desviación típica respecto a la media, en segundos, para la Carrera 1).

d. Usa la calculadora para hallar la media y la desviación estándar para cada una de las tres carreras. Escribe tus respuestas a la milésima más cercana en la siguiente tabla.

	Media	Desviación estándar
Carrera 1		
Carrera 2		
Carrera 3		

e. ¿Qué tan cercanas a los valores reales fueron tus respuestas a las partes (a) a (c)?

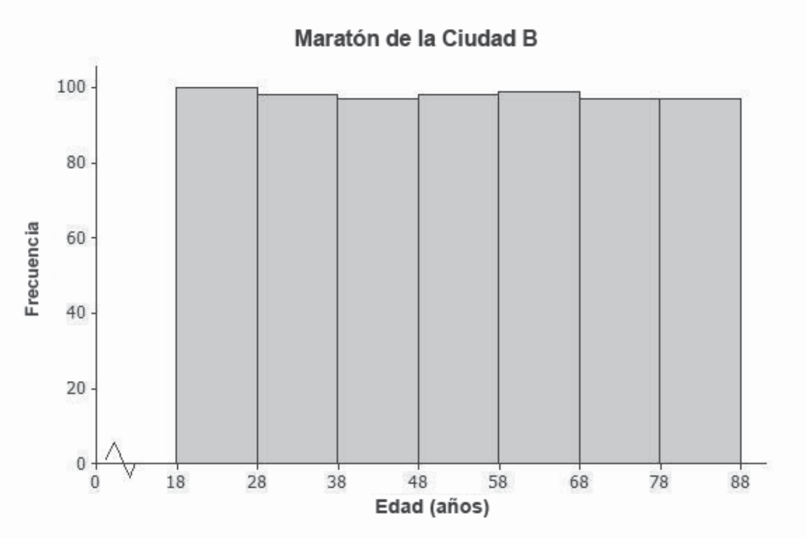
2. Una gran ciudad, a la que llamaremos Ciudad A, celebró una maratón. Supón que la edad de los participantes de la maratón que se celebró en la Ciudad A se resume en el siguiente histograma.



a. Estima la media de edad de los participantes en la maratón de la Ciudad A.

b. *Estima* la desviación estándar de la edad de los participantes en la maratón de la Ciudad A.

Una ciudad más pequeña, la Ciudad B, también celebró una maratón. Sin embargo, la Ciudad B restringió a 100 el número de personas de cada categoría de edad que podían participar. Las edades de los participantes se resumen en el siguiente histograma.



c. ¿Cuál fue aproximadamente la media de edad de los participantes en la maratón de la Ciudad B? ¿Cuál fue aproximadamente la desviación estándar de la edad de los participantes?

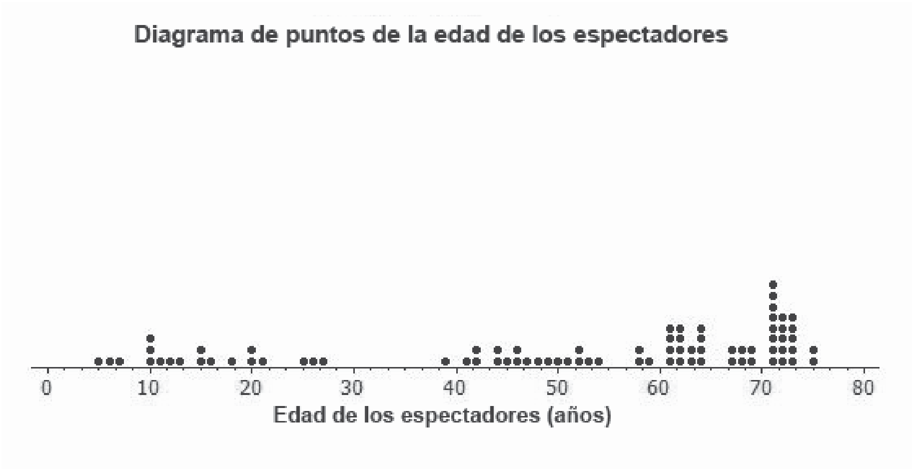
d. Explica por qué la desviación estándar de la edad de los participantes en la maratón de la Ciudad B es mayor que la desviación estándar de la edad de los participantes en la maratón de la Ciudad A.

Lección 7: Medir la variabilidad de las distribuciones sesgadas (rango intercuartil)

Trabajo en clase

Desafío exploratorio 1/Ejercicios 1 a 3: Datos sesgados y la medida de tendencia central

Considera el siguiente escenario. Un programa de televisión, *Hechos o ficción*, fue cancelado después de nueve emisiones. Muchas personas miraron las nueve emisiones y se molestaron bastante cuando el programa se dejó de emitir. Se seleccionó una muestra aleatoria de ochenta espectadores del programa. Los espectadores de la muestra respondieron varias preguntas. En el siguiente diagrama de puntos se muestra la distribución de la edad de los ochenta espectadores.



1. ¿Dónde ubicarías aproximadamente la media (el punto de equilibrio) de la distribución anterior?

2. ¿De qué manera la dirección de la cola afecta a la ubicación de la media de edad en comparación con la mediana?

3. La media de edad de la muestra anterior es aproximadamente 50. ¿Crees que esta edad describe al espectador típico de este programa? Explica tu respuesta.

Desafío exploratorio 2/Ejercicios 4 a 8: Construir e interpretar un diagrama de caja

4. Utiliza el diagrama de puntos de arriba y sigue los pasos que se indican a continuación para construir un diagrama de caja sobre el diagrama de puntos:
   i. Ubica las 40 observaciones del medio y dibuja una caja alrededor de esos valores.
   ii. Calcula la mediana y luego traza una recta dentro de la caja en la ubicación de la mediana.
   iii. Traza una recta que se extienda desde el extremo superior de la caja hasta la observación más grande del conjunto de datos.
   iv. Traza una recta que se extienda desde el borde inferior de la caja hasta el valor mínimo del conjunto de datos.

5. Recuerda que los 5 valores que se utilizaron para construir el diagrama de puntos conforman el resumen de 5 números. ¿Cuál es el resumen de 5 números para este conjunto de datos de edades?

Edad mínima: ___________

Cuartil inferior o Q1: ___________

Mediana de las edades: ___________

Cuartil superior o Q3: ___________

Edad máxima: ___________

6. ¿Qué porcentaje de los datos quedan dentro de la caja en el diagrama de caja?

7. ¿Qué porcentaje de los datos están entre el valor mínimo y Q1?

8. ¿Qué porcentaje de los datos están entre el valor máximo y Q3?

Ejercicios 9 a 14

Una agencia de publicidad investigó la edad de los espectadores que están más interesados en distintos tipos de anuncios que se emiten por televisión. Considera los siguientes resúmenes:

Edad	Productos o servicios publicitados
$30–45$	Productos electrónicos, artículos para el hogar, carros
$46–55$	Servicios financieros, electrodomésticos, muebles
$56–72$	Planes de jubilación, cruceros, servicios de atención médica

9. La media de edad de las personas encuestadas es 50 años aproximadamente. En consecuencia, los productores del programa decidieron obtener anunciantes para un espectador típico de 50 años de edad. Según la tabla, ¿qué productos o servicios crees que los productores elegirán para los anuncios publicitarios? De acuerdo con la muestra, ¿qué porcentaje de las personas encuestadas sobre el programa *Hechos o ficción* estarían interesadas en estos anuncios si la tabla sobre la publicidad es precisa?

10. El programa no logró generar el interés que los anunciantes esperaban. Como resultado, dejaron de emitir sus anuncios publicitarios en el programa y este fue cancelado. Kristin argumentó que una edad mejor para describir al espectador típico es la mediana de las edades. ¿Cuál es la mediana de las edades de la muestra? ¿Qué productos o servicios sugiere la tabla sobre publicidad para los espectadores si se considera la mediana de las edades como una descripción del espectador típico?

11. ¿Qué porcentaje de las personas encuestadas estarían interesadas en los productos o servicios que sugiere la tabla sobre publicidad si se utilizara la mediana de las edades para describir al espectador típico?

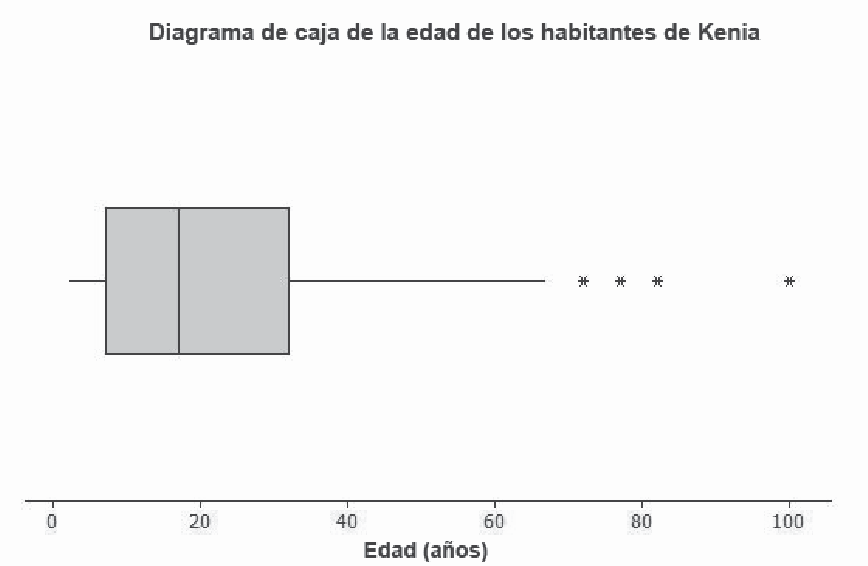
12. ¿Qué porcentaje de los espectadores tienen edades que se encuentran entre el Q1 y el Q3? La diferencia entre el Q3 y el Q1, o Q3 – Q1, se denomina "rango intercuartil", o RIQ. ¿Cuál es el RIQ de esta distribución de datos?

13. El RIQ proporciona un resumen de la variabilidad de una distribución de datos sesgada. El RIQ es un número que especifica la longitud del intervalo que contiene la mitad del medio de las edades de los espectadores. ¿Crees que los productores del programa preferirían un programa con un rango intercuartil pequeño o grande? Explica tu respuesta.

14. ¿Estás de acuerdo con el argumento de Kristin de que la mediana de las edades proporciona una mejor descripción de un espectador típico? Explica tu respuesta.

Desafío exploratorio 3/Ejercicios 15 a 20: Datos aberrantes

Los estudiantes de la Escuela Secundaria Waldo participan en un proyecto especial que consiste en comunicarse con los habitantes de Kenia. Considera un diagrama de caja que representa la edad de 200 habitantes de Kenia seleccionados aleatoriamente.



Una distribución de datos puede contener datos extremos (valores específicos de los datos que son inusualmente grandes o pequeños en relación con la mediana y con el rango intercuartil). Se puede utilizar un diagrama de caja para mostrar esos valores extremos de los datos, que se conocen como "datos aberrantes".

Cada "*" del diagrama de caja representa la edad de cuatro personas de esta muestra. Según la muestra, estas cuatro edades se consideran datos aberrantes.

15. Estima los valores de las cuatro edades que se representan con un *.

Un dato aberrante se define como cualquier valor de los datos que se encuentra a más de $1.5 \times (RIQ)$ de distancia del cuartil más cercano.

16. ¿Cuál es la mediana de las edades de la muestra de habitantes de Kenia? ¿Cuáles son los valores aproximados del Q1 y el Q3? ¿Cuál es el RIQ aproximado de esta muestra?

17. Multiplica el RIQ por 1.5. ¿Qué valor obtienes?

18. Suma $1.5 \times (RIQ)$ al tercer cuartil de las edades (Q3). ¿Qué observas acerca de las cuatro edades identificadas con un *?

19. ¿Hay algún valor de edad que sea menor que $Q1 - 1.5 \times (IQR)$? De ser así, esas edades también se considerarían datos aberrantes.

20. Explica por qué no hay ningún * en el lado inferior del diagrama de caja para las edades de la muestra de habitantes de Kenia.

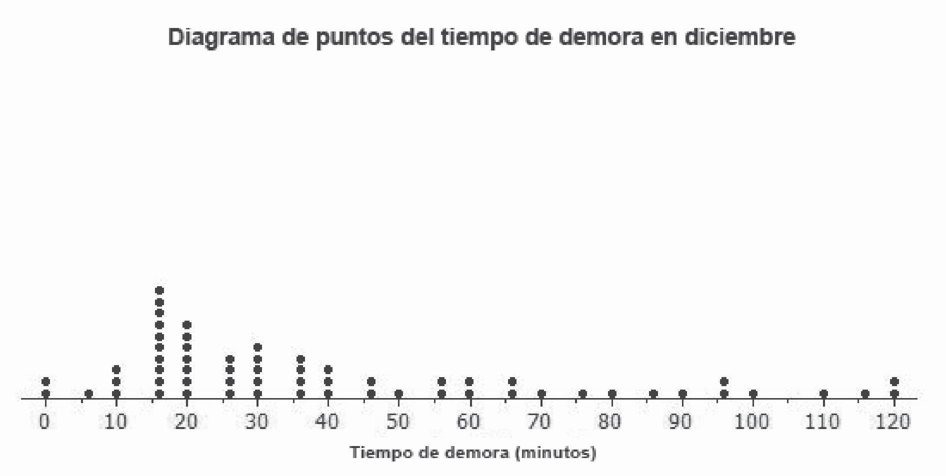
Resumen de la lección

- Las distribuciones de datos asimétricas se denominan "sesgadas".
- Cuando una distribución es sesgada hacia la izquierda, los datos tienen una mayor dispersión (como la cola de un animal) en el lado izquierdo.
- Cuando una distribución es sesgada hacia la derecha, los datos tienen una mayor dispersión (como la cola de un animal) en el lado derecho.
- La tendencia central de una distribución de datos sesgada se describe mediante la mediana.
- La variabilidad de una distribución de datos sesgada se describe mediante el rango intercuartil (RIQ).
- El RIQ describe la variabilidad al especificar la longitud del intervalo que contiene el 50% del medio de los valores de los datos.
- Los datos aberrantes de un conjunto de datos se definen como aquellos valores que se encuentran a más de $1.5 \times (RIQ)$ de distancia del cuartil más cercano. Los datos aberrantes se suelen identificar con un "*" o un "•" en un diagrama de caja.

Grupo de problemas

Considera el siguiente escenario. Los funcionarios de transporte recopilan datos sobre las demoras en los vuelos (el número de minutos de demora en el despegue de un vuelo con respecto a la hora prevista).

Considera el diagrama de puntos sobre el tiempo de demora en minutos para 60 vuelos de BigAir durante diciembre de 2012:



1. ¿Cuántos vuelos despegaron con más de 60 minutos de demora?

2. ¿Por qué esta distribución de datos se considera sesgada?

3. ¿La cola de esta distribución de datos está a la derecha o a la izquierda? ¿Cómo describirías algunos de los tiempos de demora que se encuentran en la cola?

4. Dibuja un diagrama de caja sobre el diagrama de puntos de los vuelos de diciembre.

5. ¿Cuál es el rango intercuartil, o RIQ, de este conjunto de datos?

6. La media de las demoras de los 60 vuelos es aproximadamente 42 minutos. ¿Crees que 42 minutos es un número de minutos típico para la demora de los vuelos de BigAir? ¿Por qué sí o por qué no?

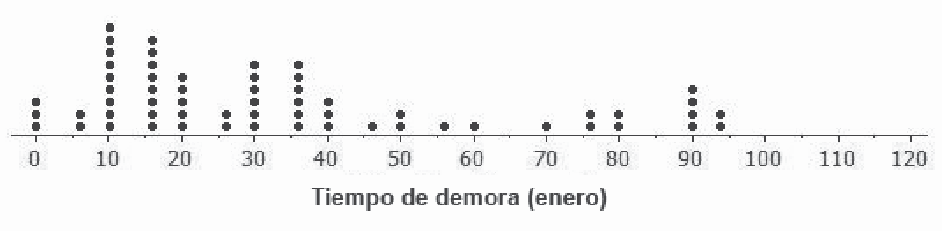
7. Basándote en los datos de diciembre, escribe una breve descripción de la distribución de los vuelos de diciembre de BigAir.

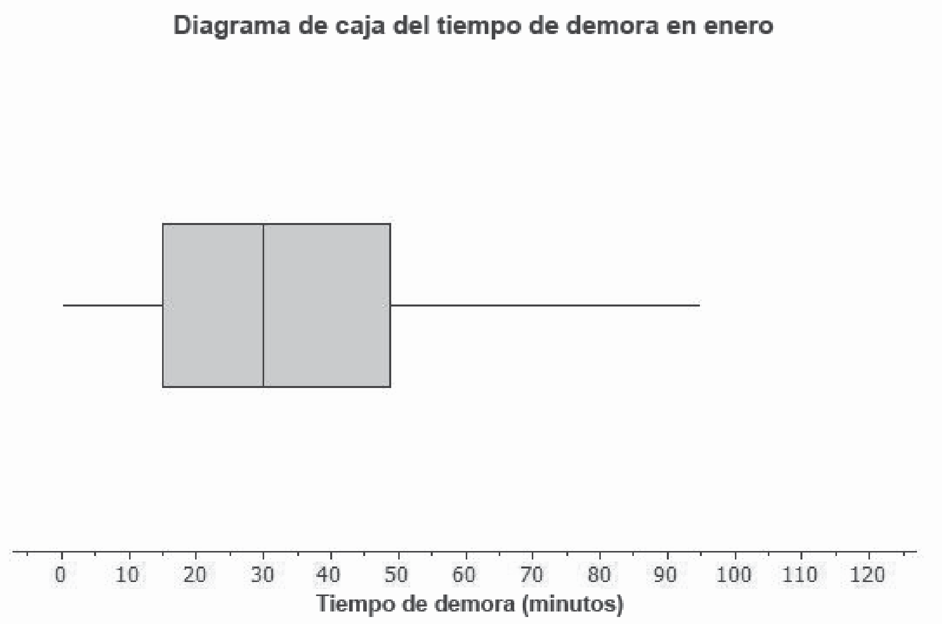
8. Calcula el porcentaje de vuelos con demoras de más de 1 hora. ¿Hubo muchos vuelos con demoras de más de 1 hora?

9. La compañía BigAir indicó más tarde que hubo una demora en un vuelo que no estaba incluido en los datos. El vuelo que no estaba registrado tuvo una demora de 48 horas. Si hubieras incluido esa demora en el diagrama de caja, ¿cómo lo habrías representado? Explica tu respuesta.

10. Considera el diagrama de puntos y el diagrama de caja sobre el tiempo de demora en minutos para 60 vuelos de BigAir durante enero de 2013.

    ¿En qué se diferencia la distribución de las demoras de los vuelos de enero respecto de la distribución que resume las demoras de los vuelos de diciembre? En términos de las demoras de los vuelos de enero, ¿la compañía BigAir mejoró, se mantuvo igual o empeoró en comparación con diciembre? Explica tu respuesta.





Esta página queda en blanco intencionalmente.

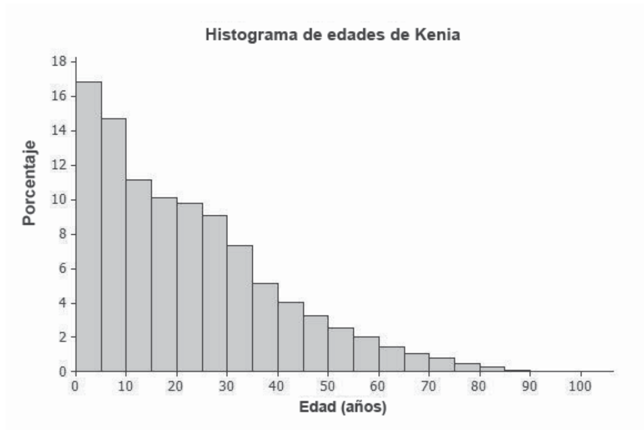
Lección 8: Comparar distribuciones

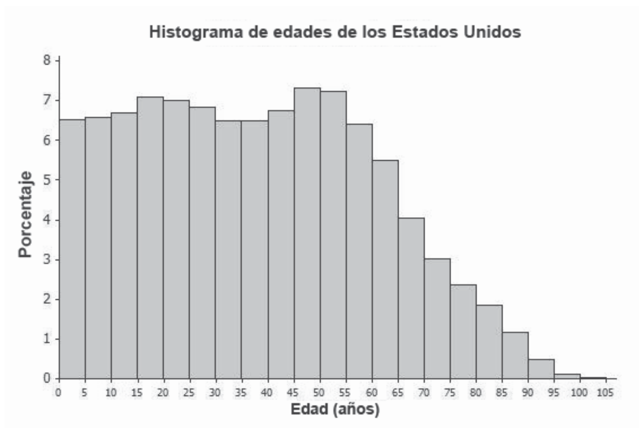
Trabajo en clase

Desafío exploratorio 1: Datos sobre países

Un museo de ciencias tiene una exposición sobre viajes alrededor del mundo. Los participantes pueden utilizar la tecnología 3D para hacer un recorrido virtual por ciudades y pueblos de todo el mundo. Los estudiantes de la Escuela Secundaria Waldo se registraron en el museo para participar en un recorrido virtual por Kenia, incluida una visita a la capital, Nairobi, y a varios pueblos pequeños. Sin embargo, antes de comenzar el recorrido, decidieron estudiar Kenia en la clase de Matemáticas y, para ello, consultaron datos demográficos de 2010 facilitados por la Oficina del Censo de los Estados Unidos. También obtuvieron datos de 2010 sobre los Estados Unidos para comparar con los datos sobre Kenia.

Los siguientes histogramas representan las distribuciones de edad de ambos países.





Ejercicios 1 a 8

1. ¿En qué se diferencian las formas de los dos histogramas?

2. ¿Aproximadamente qué porcentaje de personas en Kenia tienen entre 0 y 10 años?

3. ¿Aproximadamente qué porcentaje de personas en los Estados Unidos tienen entre 0 y 10 años?

4. ¿Aproximadamente qué porcentaje de personas en Kenia tienen 60 años o más?

5. ¿Aproximadamente qué porcentaje de personas en los Estados Unidos tienen 60 años o más?

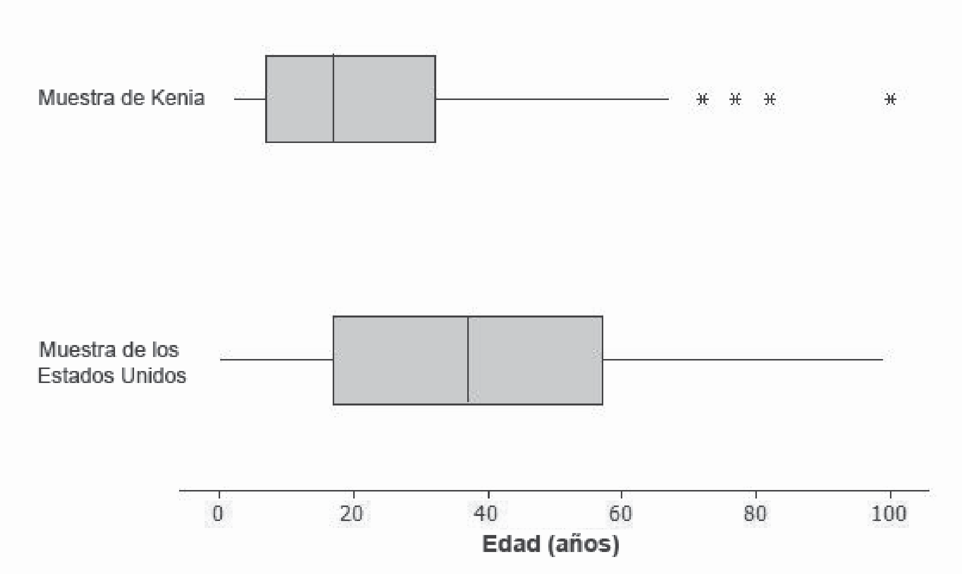
6. La población de Kenia en 2010 era de aproximadamente 41 millones de personas. ¿Cuál es el número aproximado de personas en Kenia que tienen entre 0 y 10 años?

7. La población de los Estados Unidos en 2010 era de aproximadamente 309 millones de personas. ¿Cuál es el número aproximado de personas en los Estados Unidos que tienen entre 0 y 10 años?

8. Los estudiantes de la Escuela Secundaria Waldo empezaron a planificar su visita virtual a los vecindarios de Nairobi y a varios pueblos de Kenia. ¿Crees que verán muchos adolescentes? ¿Verán muchas personas mayores que tengan 70 años o más? Explica tu respuesta basándote en el histograma.

Desafío exploratorio 2: Aprender más sobre los países mediante diagramas de caja e histogramas

En lecciones anteriores, trabajamos con una muestra aleatoria de 200 personas de Kenia en 2010. También está disponible para analizar una muestra aleatoria de 200 personas de los Estados Unidos. A continuación, se muestran diagramas de caja construidos utilizando la edad de las personas de estas dos muestras.



Ejercicios 9 a 16

9. Adrián, un estudiante de último año de la Escuela Secundaria Waldo, dijo que los diagramas de caja indican que, en comparación con Kenia, hay muchas personas mayores en los Estados Unidos. ¿Estás de acuerdo? ¿Cómo describirías la diferencia en la edad de las personas de estos dos países de acuerdo con los diagramas de caja de arriba?

10. Utiliza los diagramas de caja para estimar la mediana de las edades de los habitantes de Kenia y la mediana de las edades de los habitantes de los Estados Unidos.

11. Basándote en el diagrama de caja, ¿el 25% de las personas de los Estados Unidos son menores de qué edad? ¿Cómo determinaste esa edad?

12. Basándote en el diagrama de caja, ¿aproximadamente qué porcentaje de habitantes de Kenia son menores de 18 años?

13. ¿Podrías haber estimado la edad media de los habitantes de Kenia utilizando el diagrama de caja? Explica tu respuesta.

14. La edad media de los habitantes de los Estados Unidos es aproximadamente 38 años. Utilizando el histograma, estima el porcentaje de personas de los Estados Unidos que son menores que la edad media de este país.

15. Si la mediana de la edad se utiliza para describir a una persona típica de Kenia, ¿qué porcentaje de personas en Kenia son menores que la mediana de la edad? ¿Cuál es una mejor descripción de una persona típica de Kenia: la media o la mediana de las edades? Explica tu respuesta.

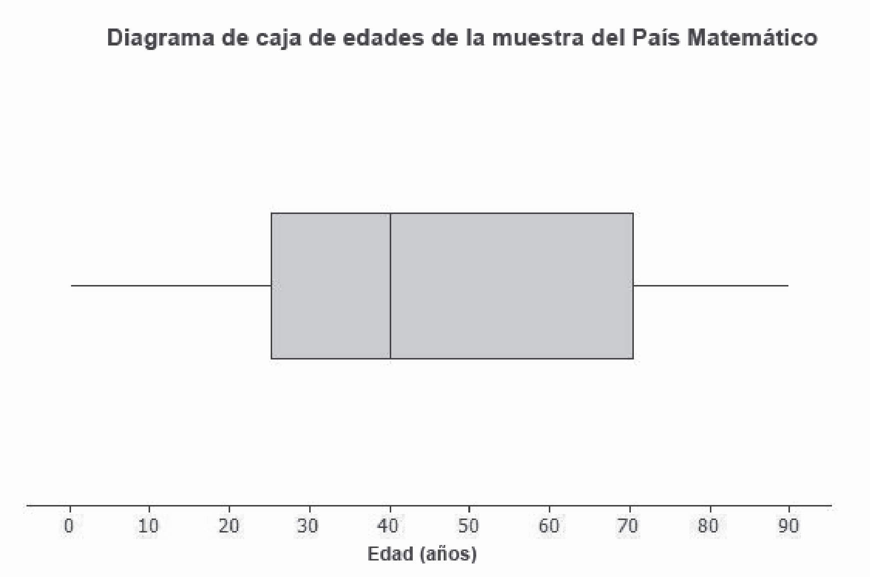
16. ¿Cuál es el RIQ de las edades de la muestra de los Estados Unidos? ¿Cuál es el RIQ de las edades de la muestra de Kenia? Si los RIQ se utilizan para comparar países, ¿qué indica un RIQ más bajo acerca de un país? Utiliza Kenia y los Estados Unidos para explicar tu respuesta.

Resumen de la lección

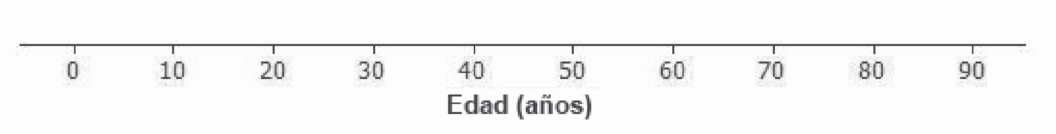
- Los histogramas muestran la forma general de una distribución.
- Los diagramas de caja se crean a partir del resumen de 5 números de un conjunto de datos.
- Un diagrama de caja identifica la mediana, los valores mínimo y máximo y los cuartiles superior e inferior.
- El rango intercuartil (RIQ) describe de qué manera los datos están dispersos alrededor de la mediana; es la longitud del intervalo que contiene el 50% de los valores de los datos.
- La mediana se utiliza como medida de la tendencia central cuando una distribución está sesgada o contiene datos aberrantes.

Grupo de problemas

El siguiente diagrama de caja resume las edades de una muestra aleatoria de un país inventado llamado País Matemático.



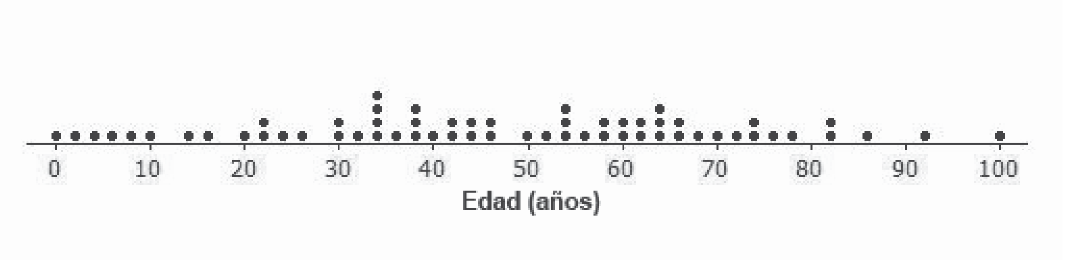
1. Inventa tu propia muestra de cuarenta edades que podrían representarse mediante el diagrama de caja del País Matemático. Utiliza un diagrama de puntos para representar la edad de las cuarenta personas del País Matemático.



2. ¿La muestra de cuarenta edades que se representa en tu diagrama de puntos del País Matemático es la única muestra que se podría representar mediante el diagrama de caja? Explica tu respuesta.

3. El siguiente es un diagrama de puntos de sesenta edades de una muestra aleatoria de habitantes de Japón en 2010.

   Dibuja un diagrama de caja sobre este diagrama de puntos.



4. Basándote en tu diagrama de caja, ¿la mediana de las edades de los habitantes de Japón estaría más cerca de la mediana de las edades de los habitantes de Kenia o de los habitantes de los Estados Unidos? Justifica tu respuesta.

5. ¿Qué indica el diagrama de caja de esta muestra de Japón acerca de las posibles diferencias entre las distribuciones de edades de las personas de Japón y de Kenia?

Lección 9: Resumir datos categóricos bivariados

Trabajo en clase

Recuerda, a partir del trabajo que hiciste en 6.° y 8.° grado, que los datos categóricos son datos no numéricos. Los datos categóricos bivariados se obtienen al recopilar datos de dos variables categóricas. En esta lección, verás ejemplos que incluyen datos categóricos recopilados a partir de dos preguntas de una encuesta.

Desafío exploratorio 1: Poderes de superhéroe

Los superhéroes han sido personajes populares de las películas, la televisión, los libros y los cómics a lo largo de muchas generaciones. Superman fue una de las series más populares en la década de 1950, mientras que Batman fue una serie muy apreciada en la década de 1960. Estos dos personajes también fueron populares por aparecer en películas estrenadas entre 1990 y 2013. Otros personajes importantes que han sido representados en las películas a lo largo de las últimas décadas son el Capitán América, She-Ra y los Cuatro Fantásticos. ¿Qué tiene de especial un superhéroe? ¿Existe algún poder especial de superhéroe que haga que estos personajes sean tan populares?

En 2010, se invitó a estudiantes de escuela secundaria de los Estados Unidos a completar una encuesta en línea. Parte de la encuesta incluía preguntas sobre los poderes de los superhéroes. Más de $1{,}000$ estudiantes respondieron esta encuesta, que incluía una pregunta sobre su poder favorito de superhéroe. Los investigadores eligieron aleatoriamente 450 de las encuestas completadas. A partir de las 450 encuestas, se compiló un desglose un poco confuso de los datos por género:

- 100 estudiantes indicaron que su poder favorito era volar. 49 de esos estudiantes eran mujeres.
- 131 estudiantes eligieron el poder de congelar el tiempo como su poder favorito. 71 de esos estudiantes eran varones.
- 75 estudiantes eligieron la invisibilidad como su poder favorito. 48 de esos estudiantes eran mujeres.
- 26 estudiantes dijeron que la fuerza sobrehumana era su poder favorito. 25 de esos estudiantes eran varones.
- Y, por último, 118 estudiantes indicaron que su poder favorito era la telepatía. 70 de esos estudiantes eran mujeres.

Ejercicios 1 a 4

Muchos superhéroes representados en las películas y las series de televisión tienen al menos un poder extraordinario. Algunos tienen más de un poder especial. ¿El poder de volar de Superman era el poder favorito de sus seguidores? ¿O lo era su fuerza sobrehumana? ¿Las mujeres considerarían el poder de volar de manera distinta que los varones o lo verían de la misma manera? Utiliza la información de la encuesta que se brinda en el Ejemplo 1 para responder las siguientes preguntas.

1. ¿Cuántas mujeres más que varones indicaron que su poder favorito era la telepatía?

2. ¿Cuántos varones más que mujeres indicaron que su poder favorito era volar?

3. Escribe preguntas de encuesta que crees que pudieron utilizarse para recopilar estos datos.

4. ¿Cómo crees que se pudieron elegir las 450 encuestas que se utilizaron en el Ejemplo 1? Puedes suponer que hubo $1{,}000$ encuestas entre las que elegir.

Desafío exploratorio 2: Un estudio estadístico con una tabla de frecuencia de doble entrada

Los datos del Ejemplo 1 condujeron a los estudiantes de una clase de Matemáticas a plantear la siguiente pregunta estadística: "¿Los varones de las escuelas secundarias tienen preferencias diferentes respecto a los poderes de los superhéroes que las mujeres de las escuelas secundarias?". Para responder esta pregunta estadística, es necesario recopilar datos, así como anticipar la variabilidad de los datos recopilados.

Los datos consisten en dos respuestas de cada estudiante que completa la encuesta. La primera respuesta indica el género del estudiante y la segunda respuesta indica el superpoder favorito del estudiante. Por ejemplo, los datos recopilados de un estudiante fueron *varón* y *volar*. Estos datos son datos categóricos bivariados.

El primer paso para analizar la pregunta estadística planteada por los estudiantes de la clase de Matemáticas consiste en organizar estos datos en una tabla de frecuencia de doble entrada.

Se puede utilizar una tabla de frecuencia de doble entrada para organizar los datos categóricos que se muestran a continuación. Las letras de abajo representan los conteos de frecuencia de las celdas de la tabla.

	Volar	**Congelar el tiempo**	**Invisibilidad**	**Fuerza sobrehumana**	**Telepatía**	**Total**
Mujeres	(a)	(b)	(c)	(d)	(e)	(f)
Varones	(g)	(h)	(i)	(j)	(k)	(l)
Total	(m)	(n)	(ñ)	(o)	(p)	(q)

- Las celdas sombreadas se denominan *frecuencias marginales*. Se encuentran alrededor de los márgenes de la tabla y representan los totales de las filas o columnas de la tabla.
- Las celdas no sombreadas *dentro* de la tabla se conocen como *frecuencias conjuntas*. Cada celda conjunta es el conteo de frecuencia de las respuestas a las dos variables categóricas y se ubica en la intersección de una fila y una columna.

Ejercicios 5 a 12

5. Describe los datos que se contarían en la celda (a).

6. Describe los datos que se contarían en la celda (j).

7. Describe los datos que se contarían en la celda (l).

8. Describe los datos que se contarían en la celda (n).

9. Describe los datos que se contarían en la celda (q).

10. La celda (i) es el número de estudiantes varones que eligieron la *invisibilidad* como su superpoder favorito. Utilizando la información que se proporciona en el Ejemplo 1, ¿cuál es el valor de este número?

11. La celda (d) es el número de estudiantes mujeres cuyo superpoder favorito es la fuerza sobrehumana. Utilizando la información que se proporciona en el Ejemplo 1, ¿cuál es el valor de este número?

12. Para completar la siguiente tabla, determina un conteo de frecuencia para cada celda de acuerdo con el resumen de los datos.

	Volar	Congelar el tiempo	Invisibilidad	Fuerza sobrehumana	Telepatía	Total
Mujeres						
Varones						
Total						

Resumen de la lección

- Los *datos categóricos* son datos que adoptan valores que son categorías en lugar de números. Los ejemplos incluyen “varón” o “mujer” para la variable categórica de género o las cinco categorías de superpoderes para la variable categórica de superpoder favorito.
- Para resumir los datos categóricos bivariados, se utiliza una *tabla de frecuencia de doble entrada*.
- En una tabla de frecuencia de doble entrada, el número que se encuentra en la intersección de una fila y una columna de la respuesta a dos variables categóricas representa una *frecuencia conjunta*.
- El número total de respuestas para cada valor de una variable categórica de la tabla representa la *frecuencia marginal* de ese valor.

Grupo de problemas

Varios estudiantes de la Escuela Secundaria Rufus King debatían acerca de si los varones o las mujeres participaban más en las actividades extracurriculares. Hay tres actividades organizadas en el programa de actividades extracurriculares: baloncesto escolar, club de ajedrez y banda de jazz. Debido a restricciones presupuestarias, cada estudiante solo puede elegir una de estas actividades. Los estudiantes no pudieron preguntar a todos los estudiantes de la escuela si participaron en el programa de actividades extracurriculares o, en caso de que participaran, qué actividad eligieron.

1. Escribe preguntas que podrían incluirse en la encuesta para investigar la pregunta sobre la que debaten los estudiantes. Las preguntas que se podrían utilizar en este estudio incluyen las siguientes:

2. La Escuela Secundaria Rufus King tiene aproximadamente 1,500 estudiantes. Sam sugirió que los 100 primeros estudiantes que entraran a la cafetería para almorzar servirían como muestra aleatoria para analizar. Janet sugirió que eligieran 100 estudiantes basándose en el número de identificación escolar. ¿Quién plantea la mejor estrategia para elegir una muestra aleatoria? ¿Cómo crees que se podría elegir a 100 estudiantes al azar para completar la encuesta?

3. Considera los siguientes resultados a partir de los 100 estudiantes elegidos aleatoriamente:

   - De las 60 estudiantes mujeres que se seleccionaron, 20 jugaban al baloncesto escolar, 10 jugaban al ajedrez y 10 estaban en la banda de jazz. El resto no participaba en el programa de actividades extracurriculares.
   - De los estudiantes varones, 10 no participaban en el programa de actividades extracurriculares, 20 jugaban al baloncesto escolar, 8 tocaban en la banda de jazz y el resto jugaba al ajedrez.

   Se comenzó a completar una tabla de frecuencia de doble entrada para resumir los datos de la encuesta. Indica qué rótulo debe ir en la celda de la tabla que se identifica con ???.

	Baloncesto escolar	**Club de ajedrez**	**Banda de jazz**	**???**	**Total**
Mujeres					
Varones					
Total					

4. Completa la tabla de arriba con los datos de los 100 estudiantes que fueron encuestados.

5. La tabla muestra las respuestas a la pregunta sobre las actividades extracurriculares para varones y mujeres. ¿Crees que existe una diferencia entre las respuestas de los varones y de las mujeres? Explica tu respuesta.

Lección 10: Resumir datos categóricos bivariados con frecuencias relativas

Trabajo en clase

Esta lección amplía el trabajo que hiciste con tablas de frecuencia de doble entrada en la Lección 9.

Desafío exploratorio 1: Extender la tabla de frecuencias a una tabla de frecuencias relativas

Determinar el número de estudiantes de cada celda representa el primer paso para organizar los datos categóricos bivariados. Otra manera de analizar los datos de la tabla consiste en calcular la *frecuencia relativa* de cada celda. Las frecuencias relativas relacionan cada conteo de frecuencia con el número total de observaciones. Para cada celda de esta tabla, la *frecuencia relativa* se halla dividiendo la frecuencia de esa celda entre el número total de respuestas.

Considera la tabla de frecuencia de doble entrada de la lección anterior.

Tabla de frecuencia de doble entrada:

	Volar	**Congelar el tiempo**	**Invisibilidad**	**Fuerza sobrehumana**	**Telepatía**	**Total**
Mujeres	49	60	48	1	70	228
Varones	51	71	27	25	48	222
Total	100	131	75	26	118	450

Para hallar la tabla de frecuencias relativas, se divide el valor de cada una de las celdas de arriba entre 450. Por ejemplo, la frecuencia relativa de mujeres que seleccionan volar es $\frac{49}{450}$, o aproximadamente 0.109 a la milésima más cercana. En la siguiente tabla de frecuencias relativas se muestran algunas de las otras frecuencias relativas a la milésima más cercana:

	Volar	**Congelar el tiempo**	**Invisibilidad**	**Fuerza sobrehumana**	**Telepatía**	**Total**
Mujeres	$\frac{49}{450} \approx 0.109$					$\frac{228}{450} \approx 0.507$
Varones			$\frac{27}{450} \approx 0.060$			
Total		$\frac{131}{450} \approx 0.291$			$\frac{118}{450} \approx 0.262$	

Ejercicios 1 a 7

1. Calcula las frecuencias relativas restantes en la siguiente tabla. Escribe el valor en la tabla en forma de número decimal redondeado a la milésima más cercana o en forma de porcentaje.

   Tabla de frecuencia de doble entrada:

	Volar	Congelar el tiempo	Invisibilidad	Fuerza sobrehumana	Telepatía	Total
Mujeres						
Varones						
Total						

2. Basándote en el trabajo que hiciste anteriormente con tablas de frecuencia, ¿qué celdas de esta tabla representarían las frecuencias relativas conjuntas?

3. ¿Qué celdas de la tabla de frecuencias relativas representarían las frecuencias relativas marginales?

4. ¿Cuál es la frecuencia relativa conjunta de las mujeres que eligieron la invisibilidad como su superpoder favorito?

5. ¿Cuál es la frecuencia relativa marginal de congelar el tiempo? Interpreta el significado de este valor.

6. ¿Cuál es la diferencia entre las frecuencias relativas conjuntas de los varones y las mujeres que eligieron volar como su superpoder favorito?

7. ¿Existe alguna diferencia notable entre los géneros y sus superpoderes favoritos?

Desafío exploratorio 2: Interpretar los datos

En la Escuela Secundaria Rufus King, continúa el interés por los superhéroes. Los estudiantes que analizaron los datos en la lección anterior decidieron crear una historieta para el sitio web de la escuela en la que aparezca un superhéroe. Pensaron que los resúmenes que crearon a partir de los datos les serían útiles para diseñar la historieta.

Solo le darán un poder al superhéroe. Surgió un debate acerca de qué poder debería tener el superhéroe de la escuela. Los estudiantes utilizaron la tabla de frecuencia de doble entrada y la tabla de frecuencias relativas para continuar la discusión. Da otro vistazo a esas tablas.

Al principio, Scott dijo que el personaje que crearan debería tener fuerza sobrehumana como poder especial. Esta sugerencia no fue bien recibida por los demás estudiantes que planificaban el proyecto. Jill, en particular, argumentó: “Bueno, si no quieres ignorar a más de la mitad de nuestros lectores, sugiero que la telepatía es un poder mejor para nuestro personaje”.

Ejercicios 8 a 10

Scott reconoció que la fuerza sobrehumana probablemente no sería la mejor elección, según los datos. Y agregó: “Los datos indican que congelar el tiempo es el poder más popular para un superhéroe”. Sin embargo, Jill, una vez más, no estuvo de acuerdo con Scott respecto a que esa fuera una buena opción. Argumentó que la telepatía era una mejor opción.

8. ¿De qué manera los datos apoyan la afirmación de Scott? ¿Por qué crees que eligió congelar el tiempo como el poder especial para el superhéroe de la historieta?

9. ¿De qué manera los datos apoyan la afirmación de Jill? ¿Por qué crees que eligió la telepatía como el poder especial para el superhéroe de la historieta?

10. Elige uno de los dos poderes especiales, congelar el tiempo o telepatía, y justifica por qué crees que es una mejor opción de acuerdo con los datos.

Resumen de la lección

- Los *datos categóricos* son datos que adoptan valores que son categorías en lugar de números. Los ejemplos incluyen "varón" o "mujer" para la variable categórica de género o las cinco categorías de superpoderes para la variable categórica de superpoder favorito.
- Para resumir los datos categóricos bivariados, se utiliza una *tabla de frecuencia de doble entrada*.
- Una *frecuencia relativa* compara un conteo de frecuencia con el número total de observaciones. Se puede escribir en forma de número decimal o en forma de porcentaje. Una tabla de doble entrada que resume las frecuencias relativas de cada celda se denomina *tabla de frecuencias relativas*.
- Las celdas marginales de una tabla de frecuencias relativas de doble entrada se llaman *frecuencias relativas marginales*, mientras que las celdas conjuntas se denominan *frecuencias relativas conjuntas*.

Grupo de problemas

1. Considera los datos sobre la Escuela Secundaria Rufus King de la lección anterior respecto a las actividades extracurriculares:

	Baloncesto escolar	**Club de ajedrez**	**Banda de jazz**	**No participan**	**Total**
Varones	20	2	8	10	40
Mujeres	20	10	10	20	60
Total	40	12	18	30	100

Calcula las frecuencias relativas para cada una de las celdas a la milésima más cercana. Escribe las frecuencias relativas en las celdas de la siguiente tabla. (Se completó la primera celda a modo de ejemplo).

	Baloncesto escolar	**Club de ajedrez**	**Banda de jazz**	**No participan**	**Total**
Varones	$\frac{20}{100} = 0.200$				
Mujeres					
Total					

2. Basándote en tu tabla de frecuencias relativas, ¿cuál es la frecuencia relativa de los estudiantes que indicaron que juegan al baloncesto?

3. Basándote en tu tabla, ¿cuál es la frecuencia relativa de varones que juegan al baloncesto?

4. Si se eligiera aleatoriamente un estudiante de la escuela, ¿crees que el estudiante elegido sería varón o mujer?

5. Si se eligiera aleatoriamente un estudiante de la escuela, ¿crees que el estudiante participaría en alguna actividad extracurricular? Explica tu respuesta.

6. ¿Por qué alguien podría dudar de que los estudiantes que completaron la encuesta hayan sido elegidos aleatoriamente? Si los estudiantes que completaron la encuesta fueron elegidos al azar, ¿qué indicarán las frecuencias relativas marginales sobre la escuela? Explica tu respuesta.

7. ¿Por qué es posible que las mujeres crean que participan más en las actividades extracurriculares que los varones? Explica tu respuesta.

Lección 11: Frecuencias relativas condicionales y asociación

Trabajo en clase

Tras discutirlo un poco más, los estudiantes encargados de diseñar la historieta sobre el superhéroe decidieron que, antes de tomar ninguna decisión, debían observar con más atención los datos sobre los poderes especiales que el personaje del superhéroe debía poseer. Existe una asociación entre el género y las respuestas sobre los superpoderes si las respuestas de los varones no son las mismas que las respuestas de las mujeres. Examinar cada una de las filas de la tabla puede ayudar a determinar si existe o no una asociación.

Desafío exploratorio 1: Frecuencias relativas condicionales

Recuerda la tabla de doble entrada de la lección anterior.

	Volar	**Congelar el tiempo**	**Invisibilidad**	**Fuerza sobrehumana**	**Telepatía**	**Total**
Mujeres	49	60	48	1	70	228
Varones	51	71	27	25	48	222
Total	100	131	75	26	118	450

Una *frecuencia relativa condicional* compara un conteo de frecuencia con el total marginal que representa la condición de interés. Por ejemplo, la condición de interés en la primera fila son las mujeres. La frecuencia relativa condicional por filas de las mujeres que responden que la invisibilidad es su superpoder favorito es $\frac{48}{228}$, o aproximadamente 0.211. Esta frecuencia relativa condicional indica que aproximadamente el 21.1% de las mujeres eligieron la invisibilidad como su superpoder favorito. Del mismo modo, $\frac{27}{222}$, o aproximadamente 0.122 o el 12.2%, de los varones eligieron la invisibilidad como su superpoder favorito.

Ejercicios 1 a 5

1. Utiliza los conteos de frecuencia de la tabla del Desafío exploratorio 1 para calcular las frecuencias relativas condicionales por filas que faltan. Redondea las respuestas a la milésima más cercana.

	Volar	**Congelar el tiempo**	**Invisibilidad**	**Fuerza sobrehumana**	**Telepatía**	**Total**
Mujeres			$\frac{48}{228} \approx 0.211$			
Varones	$\frac{51}{222} \approx 0.230$					$\frac{222}{222} = 1.000$
Total						

2. Supón que se selecciona aleatoriamente a uno de los estudiantes que completaron la encuesta. ¿Cuál crees que será el género del estudiante seleccionado? ¿Qué respuesta predices que daría este o esta estudiante a la pregunta sobre los superpoderes?

3. Supón que se selecciona aleatoriamente a uno de los estudiantes que completaron la encuesta. Si el estudiante seleccionado es varón, ¿cuál crees que fue su respuesta a la pregunta sobre el superpoder favorito? Explica tu respuesta.

4. Supón que se selecciona aleatoriamente a uno de los estudiantes que completaron la encuesta. Si el estudiante seleccionado es mujer, ¿cuál crees que fue su respuesta a la pregunta sobre el superpoder favorito? Explica tu respuesta.

5. ¿Qué superpoder eligieron aproximadamente un tercio de las mujeres? ¿Qué superpoder eligieron aproximadamente un tercio de los varones? ¿Cómo determinaste cada una de las respuestas a partir de la tabla de frecuencias relativas condicionales?

Desafío exploratorio 2: Posible asociación basada en las frecuencias relativas condicionales

Dos variables categóricas están asociadas si las frecuencias relativas condicionales por filas (o por columnas) son diferentes para las filas (o columnas) de la tabla. Por ejemplo, si la elección de superpoderes por parte de las mujeres es diferente de la elección de superpoderes por parte de los varones, entonces el género y el superpoder favorito están asociados. Esta diferencia indica que conocer el género de una persona de la muestra nos da pistas sobre sus preferencias en cuanto al superpoder.

Los indicios de una asociación son más evidentes cuando las frecuencias relativas condicionales son considerablemente distintas. Si las frecuencias relativas condicionales son casi iguales en todas las categorías, entonces probablemente no exista una asociación entre las variables.

Ejercicios 6 a 10

Examina las frecuencias relativas condicionales de la tabla de doble entrada que creaste en el Ejercicio 1. Observa que, para cada superpoder, las frecuencias relativas condicionales de los varones y de las mujeres son diferentes.

6. ¿Para qué superpoderes dirías que las frecuencias relativas condicionales de los varones y de las mujeres son muy diferentes?

7. ¿Para qué superpoderes las frecuencias relativas condicionales de los varones y de las mujeres son casi iguales?

8. Supón que se selecciona aleatoriamente a uno de los estudiantes que completaron la encuesta. ¿Sería útil conocer el género del estudiante para predecir qué superpoder seleccionó? Explica tu respuesta.

9. ¿Hay evidencias de una asociación entre el género y el superpoder favorito? Explica por qué sí o por qué no.

10. ¿Qué superpoder les recomendarías a los estudiantes de la Escuela Secundaria Rufus King para su personaje de superhéroe? Justifica tu elección.

Desafío exploratorio 3: Asociación y causa y efecto

Les dieron a los estudiantes la oportunidad de asistir a un curso de repaso para prepararse para una prueba de nivel de matemáticas para la universidad. No todos los estudiantes aprovecharon esta oportunidad. Los siguientes resultados se obtuvieron a partir de una muestra aleatoria de estudiantes que hicieron la prueba de nivel.

	Ubicados en Matemáticas 200	**Ubicados en Matemáticas 100**	**Ubicados en Matemáticas 50**	**Total**
Asistieron al curso de repaso.	40	13	7	60
No asistieron al curso de repaso.	10	15	15	40
Total	50	28	22	100

Ejercicios 11 a 16

11. Construye una tabla de frecuencias relativas condicionales por filas con los datos de arriba.

	Ubicados en Matemáticas 200	Ubicados en Matemáticas 100	Ubicados en Matemáticas 50	Total
Asistieron al curso de repaso.				
No asistieron al curso de repaso.				
Total				

12. De acuerdo con las frecuencias relativas condicionales, ¿hay evidencias de una asociación entre si un estudiante asistió al curso de repaso y el nivel de matemáticas en el cual fue ubicado? Explica tu respuesta.

13. Al observar las frecuencias relativas condicionales, vemos que la proporción de estudiantes que fueron ubicados en Matemáticas 200 es mucho mayor entre los estudiantes que asistieron al curso de repaso que entre los que no asistieron. Una explicación posible es que asistir al curso de repaso mejorara las calificaciones en la prueba de nivel. ¿Cuál es otra explicación posible?

Ahora, considera el siguiente estudio estadístico:

Se seleccionó aleatoriamente a cincuenta estudiantes de una escuela intermedia grande. Se clasificó a cada uno de esos estudiantes de acuerdo con su consumo de azúcar (alto o bajo) y su nivel de actividad física (alto o bajo). Los datos de los resultados se resumen en la siguiente tabla de frecuencias.

		Nivel de actividad física		
		Alto	Bajo	Total
Consumo de azúcar	Alto	14	18	32
	Bajo	14	4	18
	Total	28	22	50

14. Calcula las frecuencias relativas condicionales por filas y muéstralas en una tabla de frecuencias relativas condicionales por filas.

		Nivel de actividad física		
		Alto	**Bajo**	**Total**
Consumo de azúcar	**Alto**			
	Bajo			
	Total			

15. ¿Hay evidencias de una asociación entre la categoría de consumo de azúcar y el nivel de actividad física? Utiliza las frecuencias relativas condicionales para apoyar tu respuesta.

16. ¿Es razonable concluir que el alto consumo de azúcar es la causa de las diferencias que se observan en las frecuencias relativas condicionales? ¿Qué otras explicaciones podrían justificar la diferencia en las frecuencias relativas condicionales? Explica tu respuesta.

Resumen de la lección

- Una frecuencia relativa condicional compara un conteo de frecuencia con el total marginal que representa la *condición* de interés.
- Las diferencias en las frecuencias relativas condicionales se utilizan para evaluar si existe o no una asociación entre dos variables categóricas.
- Cuanto mayores son las diferencias en las frecuencias relativas condicionales, más evidentes son los indicios de que existe una asociación.
- El hecho de que se observe una asociación entre dos variables no significa necesariamente que exista una relación de causa y efecto entre las dos variables.

Grupo de problemas

Vuelve a considerar el resumen de los datos de la investigación sobre las actividades extracurriculares y el género de los 100 estudiantes que se seleccionaron aleatoriamente en la Escuela Secundaria Rufus King.

	Baloncesto escolar	**Club de ajedrez**	**Banda de jazz**	**No participan**	**Total**
Mujeres	20	10	10	20	60
Varones	20	2	8	10	40
Total	40	12	18	30	100

1. Construye una tabla de frecuencias relativas condicionales por filas con estos datos. Se indican los valores decimales redondeados a la milésima más cercana.

	Baloncesto escolar	**Club de ajedrez**	**Banda de jazz**	**No participan**	**Total**
Mujeres					60
Varones					40
Total					

2. ¿En cuáles actividades extracurriculares crees que las frecuencias relativas condicionales por filas para las mujeres y para los varones son muy diferentes? ¿Qué dato podría explicar por qué los varones y las mujeres eligen actividades diferentes?

3. Si John, un estudiante varón de la Escuela Secundaria Rufus King, completó la encuesta sobre las actividades extracurriculares, ¿qué predicción harías sobre cuál fue su respuesta? Explica tu respuesta.

4. Si Beth, una estudiante mujer de la Escuela Secundaria Rufus King, completó la encuesta sobre las actividades extracurriculares, ¿qué predicción harías sobre cuál fue su respuesta? Explica tu respuesta.

5. Observa que 20 estudiantes mujeres y 20 estudiantes varones participan en el baloncesto escolar. ¿Es preciso decir que las mujeres y los varones participan en el baloncesto escolar en igual medida? Explica tu respuesta.

6. ¿Crees que existe una asociación entre el género y la elección de un programa extracurricular? Explica tu respuesta.

Las *frecuencias relativas condicionales por columnas* también se pueden calcular dividiendo cada frecuencia de una tabla de frecuencias entre el total de la columna correspondiente para crear una tabla de frecuencias relativas condicionales por columnas. Las frecuencias relativas condicionales por columnas indican las proporciones, o las frecuencias relativas, basadas en los totales de las columnas.

7. Si quisieras saber la frecuencia relativa de las mujeres encuestadas que participan en el club de ajedrez, ¿utilizarías una frecuencia relativa condicional por filas o una frecuencia relativa condicional por columnas?

8. Si quisieras saber la frecuencia relativa de los miembros de la banda encuestados que son mujeres, ¿utilizarías una frecuencia relativa condicional por filas o una frecuencia relativa condicional por columnas?

9. Para los datos de la encuesta sobre los superpoderes, escribe una pregunta que se respondería utilizando una frecuencia relativa condicional por filas.

10. Para los datos de la encuesta sobre los superpoderes, escribe una pregunta que se respondería utilizando una frecuencia relativa condicional por columnas.

Esta página queda en blanco intencionalmente.

Lección 12: Relaciones entre dos variables numéricas

Trabajo en clase

Un diagrama de dispersión es una manera informativa de representar datos numéricos con dos variables. En el trabajo que hiciste anteriormente en 8.° grado, viste cómo construir e interpretar diagramas de dispersión. Recuerda que, si dos variables numéricas se denotan mediante x y y, el diagrama de dispersión de los datos es una representación gráfica de los pares de datos (x, y).

Ejemplo 1: Buscar patrones en un diagrama de dispersión

El Centro Nacional de Datos Climáticos recopila datos sobre las condiciones meteorológicas en diversos lugares. Clasifican cada día como despejado, parcialmente nublado o nublado. A partir de los datos reunidos a lo largo de varios años, proporcionan datos sobre las siguientes variables.

x representa la elevación sobre el nivel del mar (en pies).

y representa la media del número de días despejados por año.

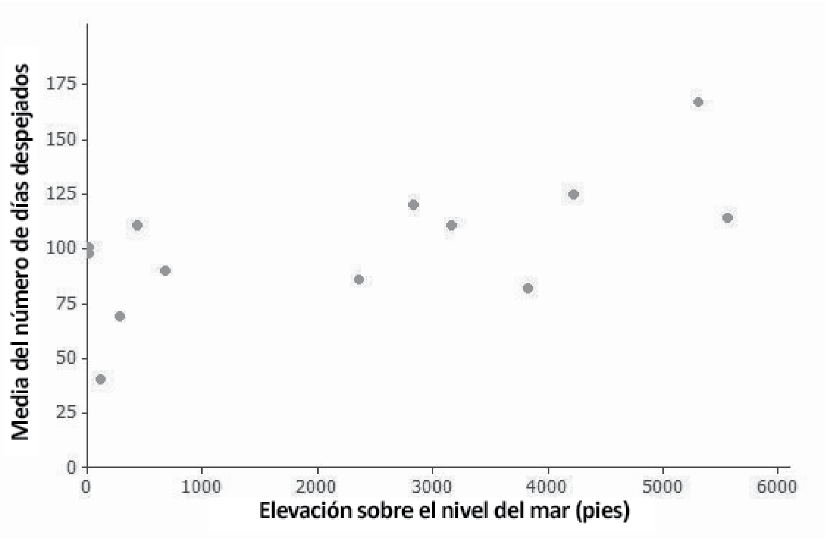
w representa la media del número de días parcialmente nublados por año.

z representa la media del número de días nublados por año.

En la siguiente tabla, se muestran los datos correspondientes a 14 ciudades estadounidenses.

Ciudad	**x (elevación sobre el nivel del mar en pies)**	**y (media del número de días despejados por año)**	**w (media del número de días parcialmente nublados por año)**	**z (media del número de días nublados por año)**
Albany, NY	275	69	111	185
Albuquerque, NM	5,311	167	111	87
Anchorage, AK	114	40	60	265
Boise, ID	2,838	120	90	155
Boston, MA	15	98	103	164
Helena, MT	3,828	82	104	179
Lander, WY	5,557	114	122	129
Milwaukee, WI	672	90	100	175
Nueva Orleans, LA	4	101	118	146
Raleigh, NC	434	111	106	149
Rapid City, SD	3,162	111	115	139
Salt Lake City, UT	4,221	125	101	139
Spokane, WA	2,356	86	88	191
Tampa, FL	19	101	143	121

Este es un diagrama de dispersión de los datos sobre la elevación y la media del número de días despejados.



Fuente de datos: www.ncdc.noaa.gov

Ejercicios 1 a 3

1. ¿Observas un patrón en el diagrama de dispersión o los puntos de datos parecen estar dispersos?

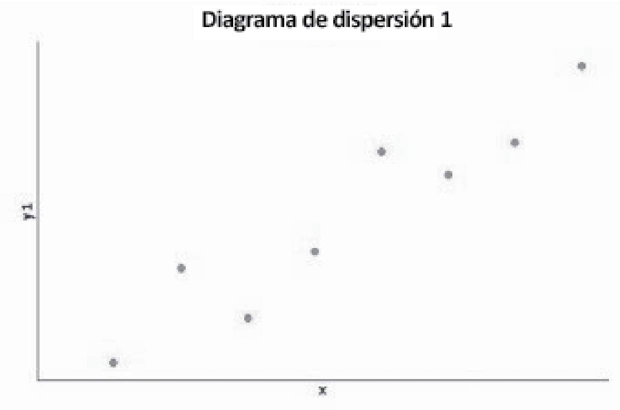
2. ¿Cómo describirías la relación entre la elevación y la media del número de días despejados para estas 14 ciudades? Es decir, ¿la media del número de días despejados tiende a aumentar a medida que aumenta la elevación, o la media del número de días despejados tiende a disminuir a medida que aumenta la elevación?

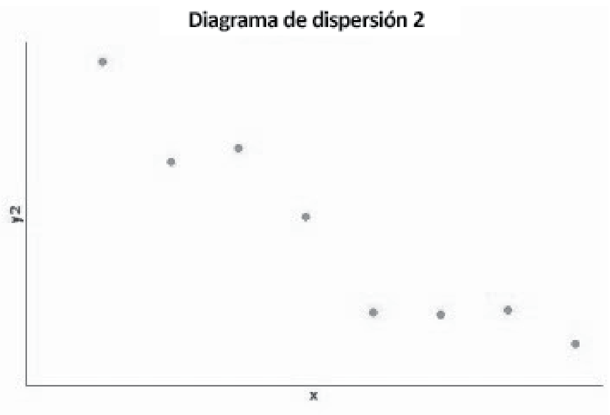
3. ¿Crees que una línea recta sería una buena manera de describir la relación entre la media del número de días despejados y la elevación? ¿Por qué crees que es así?

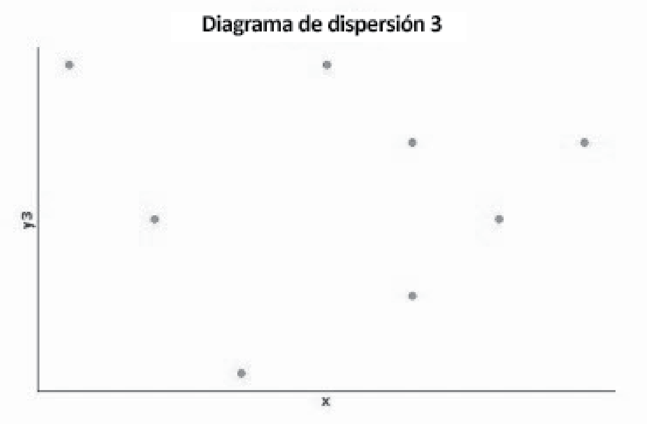
Ejercicios 4 a 7: Pensar en las relaciones lineales

A continuación hay tres diagramas de dispersión. Cada uno representa un conjunto de datos con ocho observaciones.

Las escalas del eje x y del eje y se han omitido intencionalmente en estos diagramas, de modo que debes pensar atentamente en las relaciones.







4. Si uno de estos diagramas de dispersión representa la relación entre la altura y el peso de ocho adultos, ¿cuál de los diagramas de dispersión crees que es? ¿Por qué?

5. Si uno de estos diagramas de dispersión representa la relación entre la altura y la calificación de matemáticas en el Examen de Aptitud Académica (SAT, por sus siglas en inglés) de ocho estudiantes de último año de la escuela secundaria, ¿cuál de los diagramas de dispersión crees que es? ¿Por qué?

6. Si uno de estos diagramas de dispersión representa la relación entre el peso de un carro y la eficiencia del combustible de ocho carros, ¿cuál de los diagramas de dispersión crees que es? ¿Por qué?

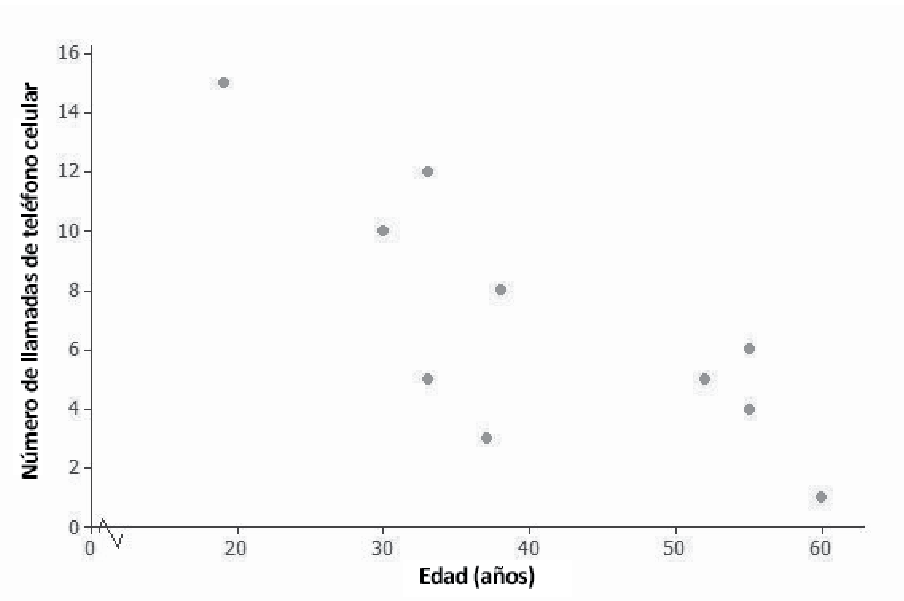
7. ¿Cuál de estos tres diagramas de dispersión *no* parece representar una relación lineal? Explica el razonamiento de tu decisión.

Ejercicios 8 a 13: No todas las relaciones son lineales

Cuando una línea recta proporciona un resumen razonable de la relación entre dos variables numéricas, decimos que las dos variables están *relacionadas linealmente* o que existe una *relación lineal* entre las dos variables.

Observa los diagramas de dispersión de abajo y responde las preguntas que siguen.

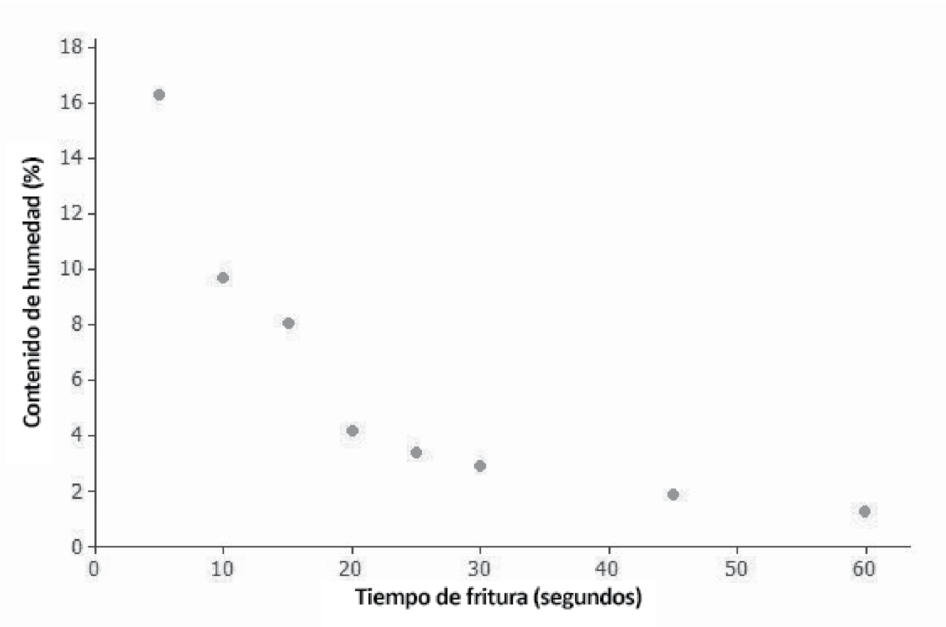
Diagrama de dispersión 1



8. ¿Existe una relación entre el número de llamadas de teléfono celular y la edad, o los puntos de datos parecen estar dispersos?

9. Si existe una relación entre el número de llamadas de teléfono celular y la edad, ¿la relación parece ser lineal?

Diagrama de dispersión 2

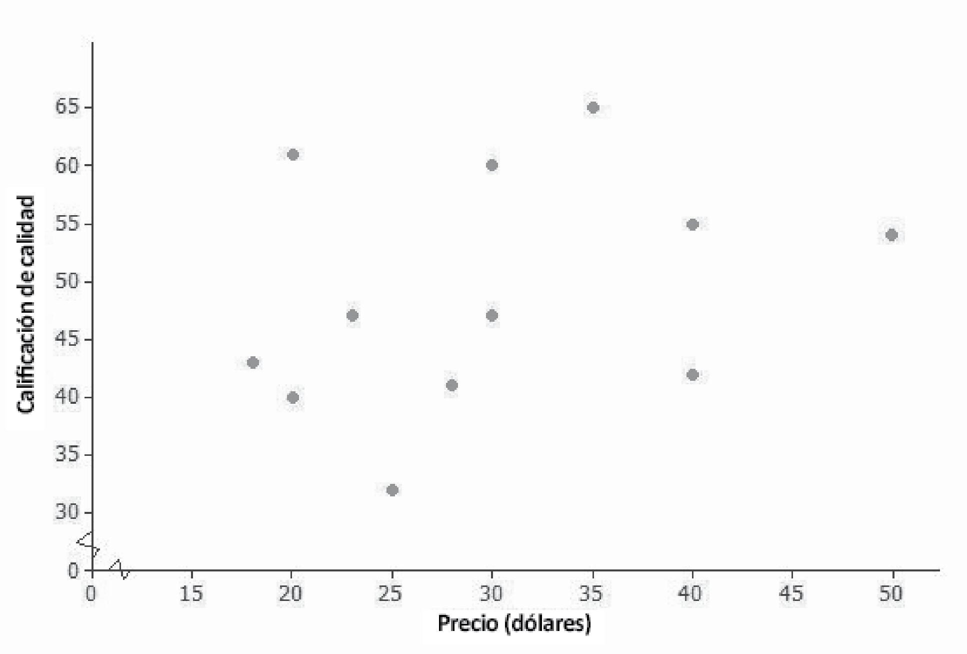


Fuente de datos: R.G. Moreira, J. Palau, V.E. Sweat y X. Sun, "Thermal and Physical Properties of Tortilla Chips as a Function of Frying Time", *Journal of Food Processing and Preservation*, 19 (1995): 175.

10. ¿Existe una relación entre el contenido de humedad y el tiempo de fritura, o los puntos de datos parecen estar dispersos?

11. Si existe una relación entre el contenido de humedad y el tiempo de fritura, ¿la relación parece lineal?

Diagrama de dispersión 3



Fuente de datos: www.consumerreports.org/health

12. En el Diagrama de dispersión 3, se muestran los datos de precios de cascos para bicicleta y las calificaciones de calidad de los cascos (basadas en una escala que estima la calidad de los cascos). ¿Existe una relación entre la calificación de calidad y el precio, o los puntos de datos están dispersos?

13. Si existe una relación entre la calificación de calidad y el precio de los cascos para bicicletas, ¿la relación parece ser lineal?

Resumen de la lección

- Se puede utilizar un diagrama de dispersión para investigar si existe o no una relación entre dos variables numéricas.
- Una relación entre dos variables numéricas se puede describir como lineal o no lineal.

Grupo de problemas

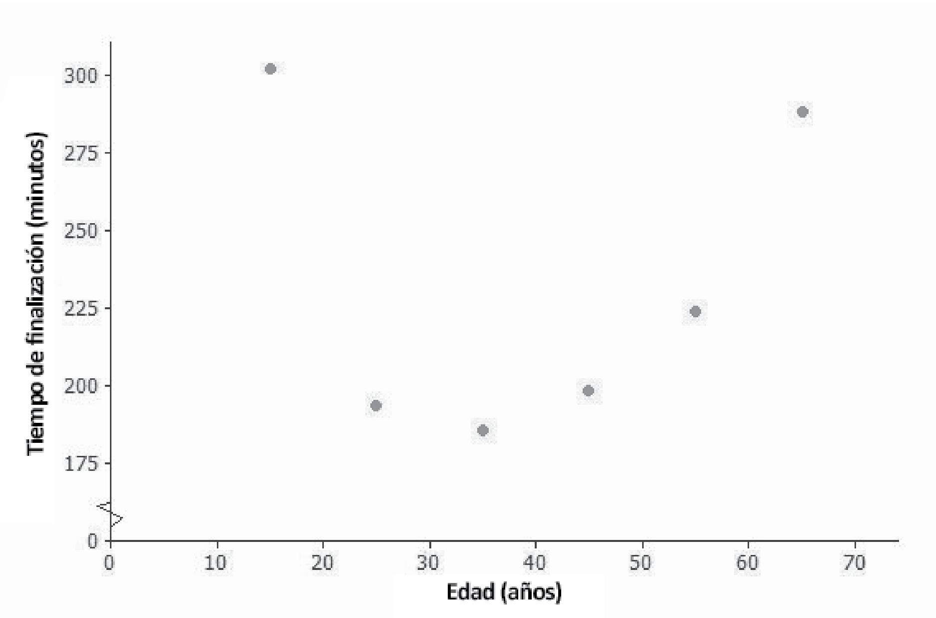
1. Construye un diagrama de dispersión que muestre los datos de x (elevación sobre el nivel del mar en pies) y w (media del número de *días parcialmente nublados por año*).

Ciudad	x (elevación sobre el nivel del mar en pies)	y (media del número de días despejados por año)	w (media del número de días parcialmente nublados por año)	z (media del número de días nublados por año)
Albany, NY	275	69	111	185
Albuquerque, NM	5,311	167	111	87
Anchorage, AK	114	40	60	265
Boise, ID	2,838	120	90	155
Boston, MA	15	98	103	164
Helena, MT	3,828	82	104	179
Lander, WY	5,557	114	122	129
Milwaukee, WI	672	90	100	175
Nueva Orleans, LA	4	101	118	146
Raleigh, NC	434	111	106	149
Rapid City, SD	3,162	111	115	139
Salt Lake City, UT	4,221	125	101	139
Spokane, WA	2,356	86	88	191
Tampa, FL	19	101	143	121

2. Según el diagrama de dispersión que construiste en el Problema 1, ¿existe una relación entre la elevación y la media del número de días parcialmente nublados por año? De ser así, ¿cómo describirías esa relación? Explica tu razonamiento.

Considera el siguiente diagrama de dispersión para los Problemas 3 y 4.

Diagrama de dispersión 4

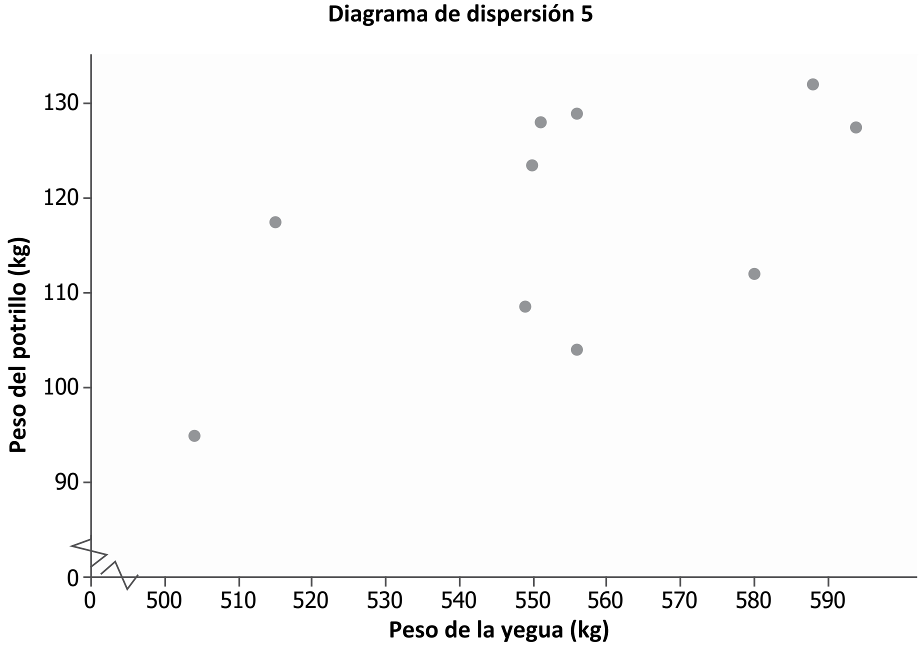


Fuente de datos: Muestra de seis mujeres que corrieron la maratón de la Ciudad de Nueva York de 2003

3. ¿Existe una relación entre el tiempo de finalización y la edad, o los puntos de datos están dispersos?

4. ¿Crees que existe una relación entre el tiempo de finalización y la edad? Der ser así, ¿parece lineal?

Considera el siguiente diagrama de dispersión para los Problemas 5 y 6.



Fuente de datos: Elissa Z. Cameron, Kevin J. Stafford, Wayne L. Linklater y Clare J. Veltman, "Suckling behaviour does not measure milk intake in horses, equus caballus", *Animal Behaviour,* 57 (1999): 673.

5. Una yegua es un caballo hembra y un potrillo es la cría del caballo. ¿Existe una relación entre el peso de un potrillo al nacer y el peso de la yegua, o los puntos de datos están dispersos?

6. Si existe una relación entre el peso de la cría al nacer y el peso de la madre, ¿la relación parece lineal?

Esta página queda en blanco intencionalmente.

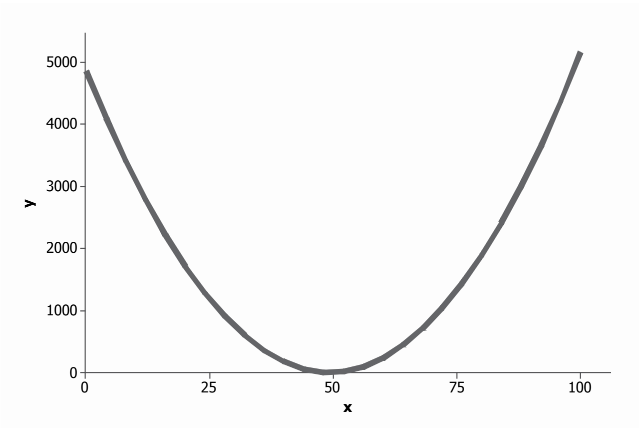
Lección 13: Relaciones entre dos variables numéricas

Trabajo en clase

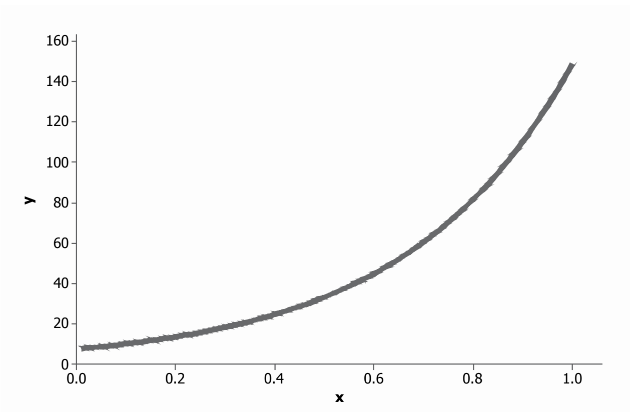
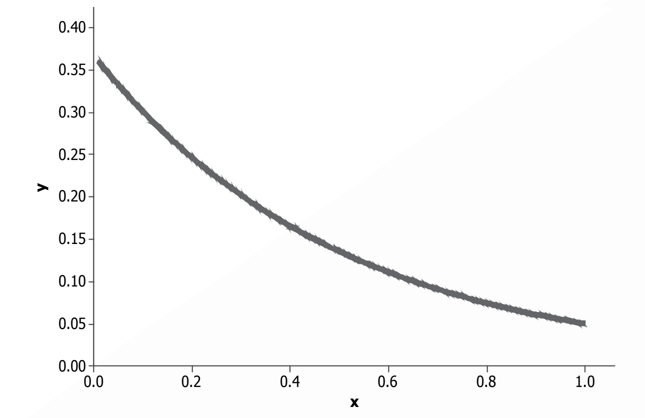
No todas las relaciones entre dos variables numéricas son *lineales*. Existen muchas situaciones en las que sería más apropiado describir el patrón del diagrama de dispersión como una curva. Se suelen utilizar dos tipos de funciones para representar las relaciones no lineales: funciones *cuadráticas* y funciones *exponenciales*.

Ejemplo 1: Representar relaciones

A veces, el patrón de un diagrama de dispersión parece una gráfica de una función cuadrática (los puntos dibujan una forma parecida a una *U* que se abre hacia arriba o hacia abajo), tal como se muestra en la siguiente gráfica.



En otras ocasiones, el patrón del diagrama de dispersión puede parecerse a las gráficas de las funciones exponenciales, ya sea las de pendiente ascendente (Gráfica 1) o las de pendiente descendente (Gráfica 2).

Gráfica 1: Exponencial, pendiente ascendente	Gráfica 2: Exponencial, pendiente descendente
	

Ejercicios 1 a 6

Vuelve a considerar los cinco diagramas de dispersión que aparecieron en la lección anterior.

Diagrama de dispersión 1

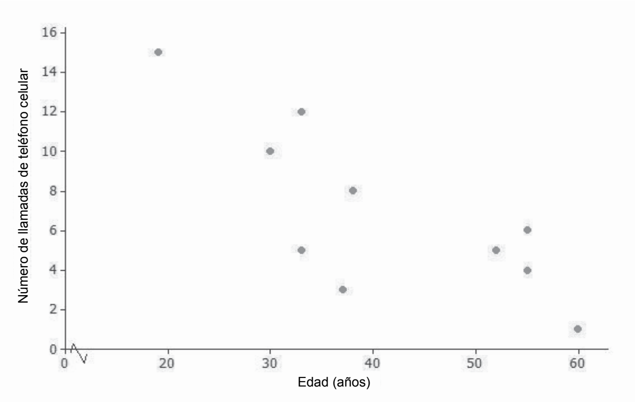


Diagrama de dispersión 2

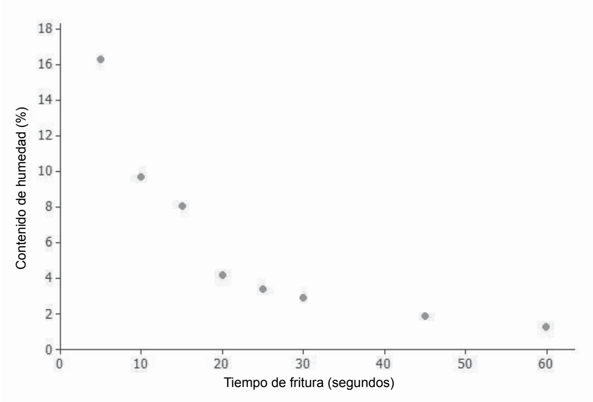


Diagrama de dispersión 3

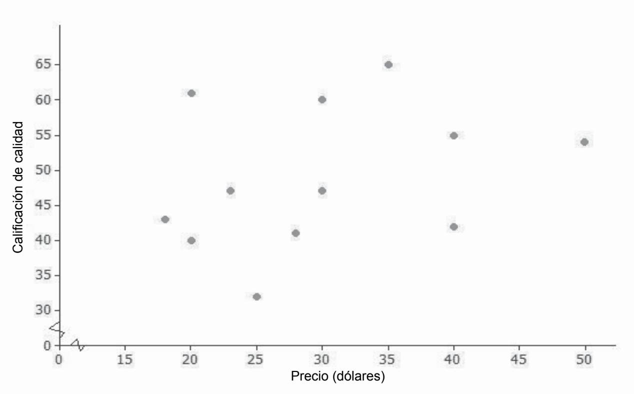


Diagrama de dispersión 4

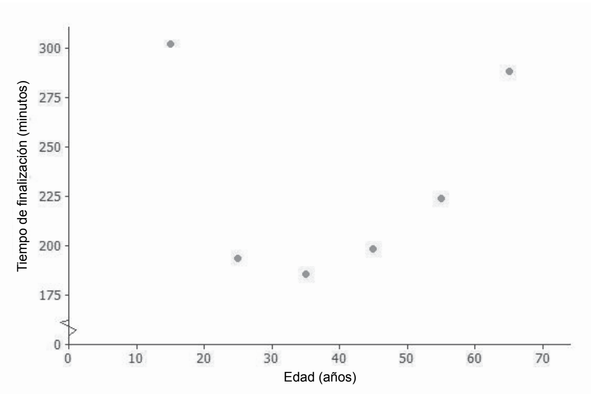
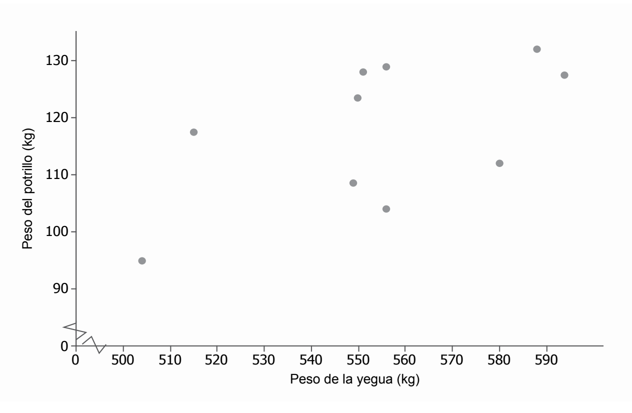


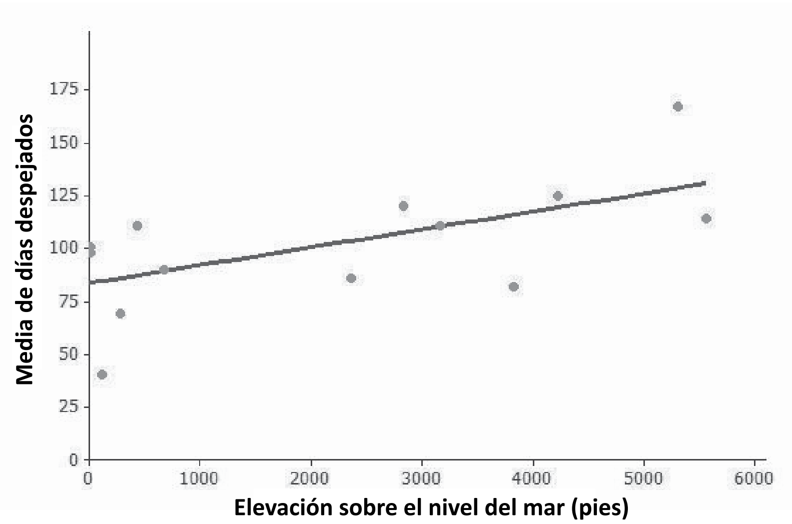
Diagrama de dispersión 5



1. ¿Cuál de los cinco diagramas de dispersión de la Lección 12 muestra un patrón que podría describirse razonablemente con una curva cuadrática?

2. ¿Cuál de los cinco diagramas de dispersión muestra un patrón que podría describirse razonablemente con una curva exponencial?

Repasemos los datos sobre la elevación (en pies sobre el nivel del mar) y la media de días despejados por año. A continuación, se muestra el diagrama de dispersión de estos datos. El diagrama también muestra una línea recta que se puede utilizar para representar la relación entre la elevación y la media de días despejados. (En 8.° grado, ajustaste una línea recta de manera informal para representar la relación entre dos variables. En la próxima lección, aprenderás una manera más formal de ajustar una línea recta). La ecuación de esta recta es $y = 83.6 + 0.008x$.



3. Si suponemos que las 14 ciudades que se utilizaron para este diagrama de dispersión representan las ciudades de todos los Estados Unidos, ¿debería haber más días despejados por año en Los Ángeles, que está cerca del nivel del mar, o en Denver, que se conoce como la ciudad a una milla de altura? Justifica tu elección mediante una recta que muestre la relación entre la elevación y la media de días despejados.

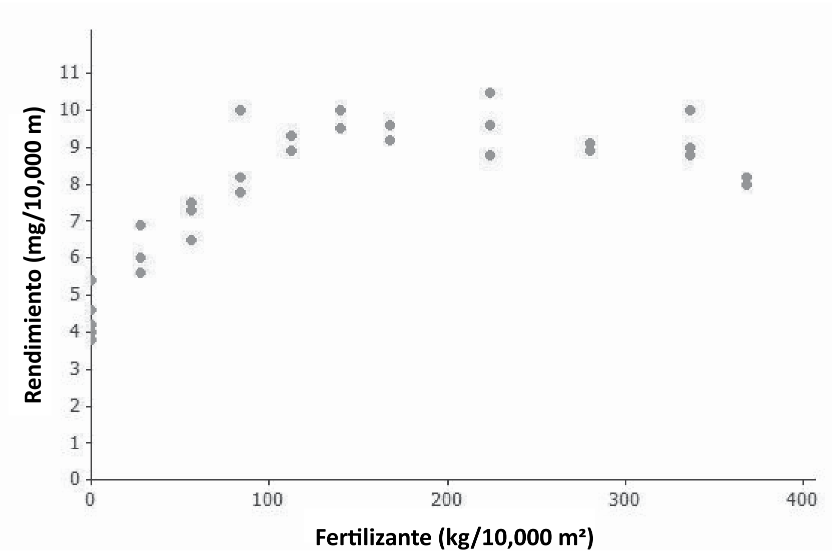
4. Una de las ciudades que conforman el conjunto de datos es Albany, Nueva York, que se encuentra a una elevación de 275 ft. Si no supieras la media de días despejados en Albany, ¿cuál número predices según la recta que describe la relación entre la elevación y la media de días despejados?

5. Otra de las ciudades incluidas en el conjunto de datos es Albuquerque, Nuevo México. Albuquerque se encuentra a una elevación de 5,311 ft. Si no supieras la media de días despejados en Albuquerque, ¿cuál sería tu predicción según la recta que describe la relación entre la elevación y la media de días despejados?

6. Tu predicción de la media de días despejados según la recta, ¿estuvo más cerca del valor real para Albany, con 69 días despejados, o para Albuquerque, con 167 días despejados? ¿Cómo podrías saberlo si observas el diagrama de dispersión con la recta que se muestra arriba?

Ejemplo 2: Un modelo cuadrático

A veces, los agricultores utilizan fertilizantes para incrementar el rendimiento de sus cultivos, pero suelen preguntarse cuánta cantidad deben utilizar. Los datos que se muestran en el siguiente diagrama de dispersión provienen de un estudio sobre los efectos del fertilizante en el rendimiento de maíz.



Fuente de datos: M.E. Cerrato y A.M. Blackmer, "Comparison of Models for Describing Corn Yield Response to Nitrogen Fertilizer" *Agronomy Journal*, 82 (1990): 138

Ejercicios 7 a 9

7. Los investigadores que llevaron a cabo este estudio decidieron que utilizarían una curva cuadrática para describir la relación entre el rendimiento y la cantidad de fertilizante. Explica por qué tomaron esa decisión.

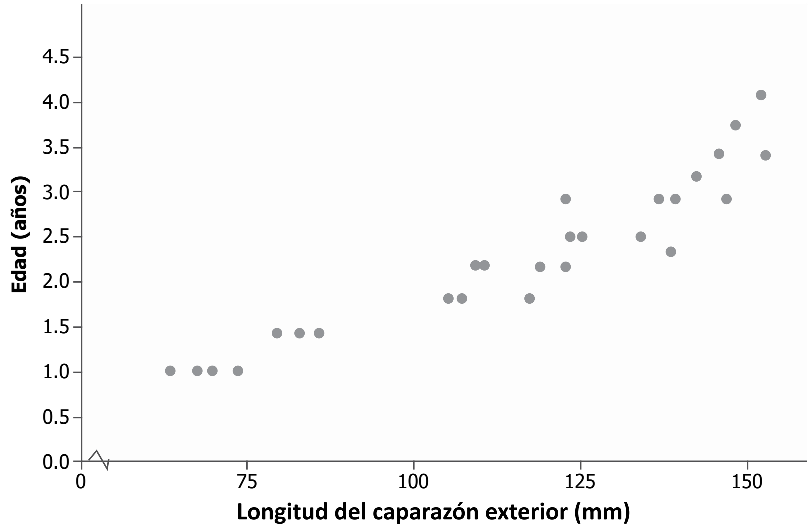
8. El modelo que utilizaron los investigadores para describir la relación fue $y = 4.7 + 0.05x - 0.0001x^2$, donde x representa la cantidad de fertilizante (kg por $10{,}000\ \mathrm{m}^2$) y y representa el rendimiento de maíz (mg por $10{,}000\ \mathrm{m}^2$). Utiliza este modelo cuadrático para completar la siguiente tabla. Luego, dibuja la gráfica de esta ecuación cuadrática en el diagrama de dispersión.

x	y
0	
100	
200	
300	
400	

9. Según este modelo cuadrático, ¿cuánto fertilizante por $10{,}000\ \mathrm{m}^2$ recomendarías que use un agricultor en sus campos de maíz para maximizar el rendimiento de su cultivo? Justifica tu elección.

Ejemplo 3: Un modelo exponencial

¿Cómo sabes cuántos años tiene una langosta? Esta pregunta es importante para los biólogos y las personas que regulan la pesca de langostas. Para responderla, un grupo de investigadores recogieron datos sobre la longitud del caparazón de 27 langostas que habían sido criadas en un laboratorio y cuya edad era conocida.

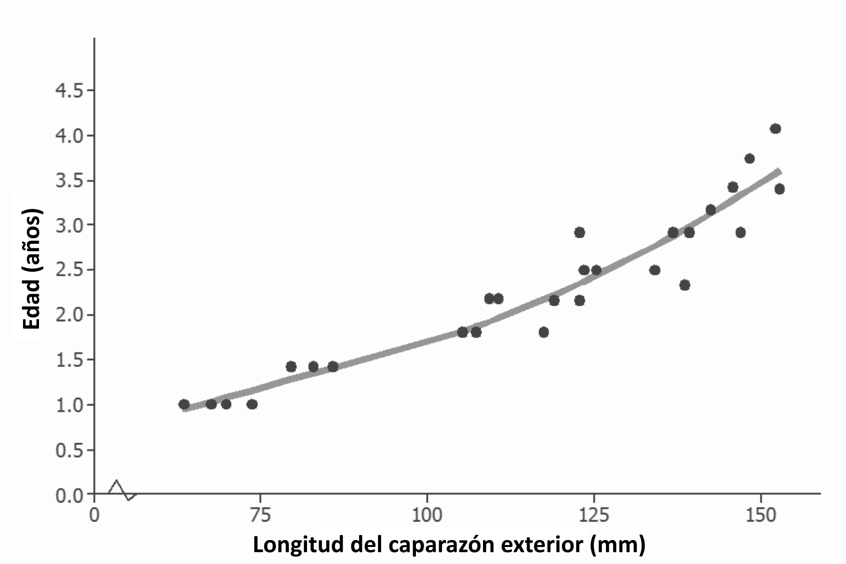


Fuente de datos: Kerry E. Maxwell, Thomas R. Matthews, Matt R.J. Sheehy, Rodney D. Bertelsen y Charles D. Derby, "Neurolipofuscin is a Measure of Age in *Panulirus argus*, the Caribbean Spiny Lobster, in Florida" *Biological Bulletin*, 213 (2007): 55

Ejercicios 10 a 13

10. Los investigadores que llevaron a cabo este estudio decidieron que utilizarían una curva exponencial para describir la relación entre la edad y la longitud del caparazón exterior. Explica por qué tomaron esa decisión.

11. El modelo que utilizaron los investigadores para describir esta relación es $y = 10^{-0.403 + 0.0063x}$, donde x representa la longitud del caparazón exterior (mm) y la y representa la edad de la langosta (en años). La curva exponencial se muestra en el siguiente diagrama de dispersión. ¿Es este modelo una buena descripción de la relación entre la edad y la longitud del caparazón exterior? Explica por qué sí o por qué no.



12. Según este modelo exponencial, ¿qué edad tiene una langosta cuyo caparazón exterior tiene una longitud de 100 mm?

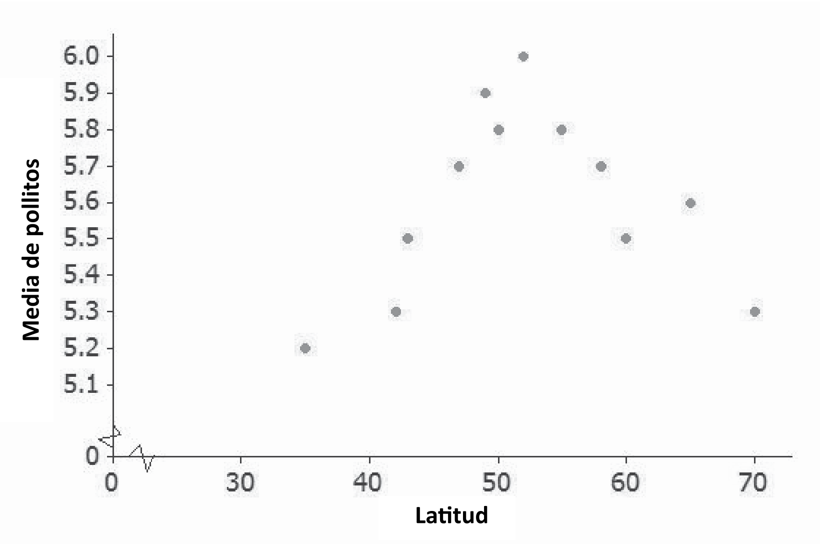
13. Supón que la normativa de la pesca indica que se debe liberar a cualquier langosta que tenga un caparazón exterior de una longitud menor que 75 mm o mayor que 150 mm. Según el modelo exponencial, ¿qué edad tienen las langostas cuyo caparazón exterior tiene una longitud menor que 75 mm? ¿Qué edad tienen las langostas cuyo caparazón exterior tiene una longitud mayor que 150 mm? Explica cómo llegaste a tu respuesta.

Resumen de la lección

- Se puede utilizar un diagrama de dispersión para investigar si existe o no una relación entre dos variables numéricas.
- Las funciones lineales, cuadráticas y exponenciales son modelos comunes que se pueden utilizar para describir la relación entre las variables.
- Se pueden utilizar modelos para responder preguntas sobre la manera en que se relacionan dos variables.

Grupo de problemas

Un grupo de biólogos llevó a cabo un estudio sobre el comportamiento de anidación de un tipo de ave llamada papamoscas. Examinaron un gran número de nidos y registraron la latitud de la ubicación de cada nido y el número de pollitos que había en cada uno.



Fuente de datos: Juan José Sanz, "Geographic variation in breeding parameters of the pied flycatcher *Ficedula hypoleuca*" *Ibis*, 139 (1997): 107

1. ¿Qué tipo de modelo (lineal, cuadrático o exponencial) describiría mejor la relación entre la latitud y la media de pollitos?

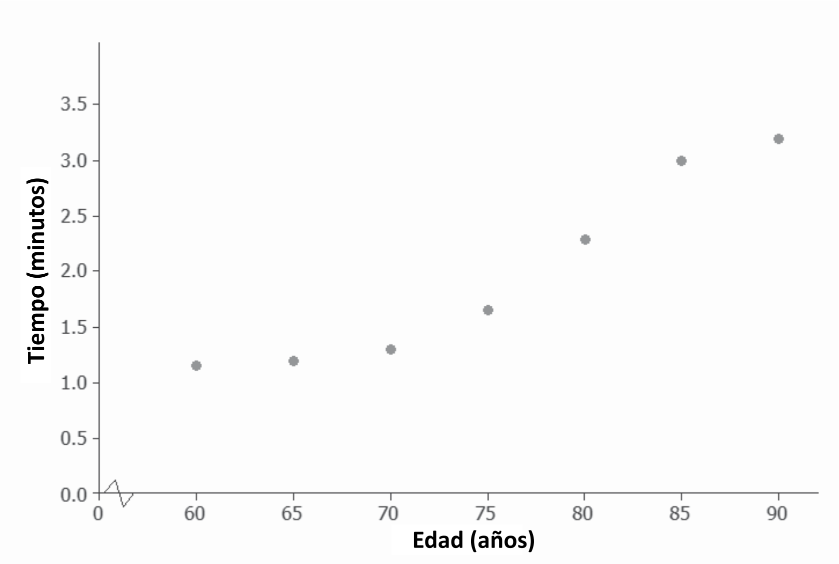
2. Un modelo que podría utilizarse para describir la relación entre la media de pollitos y la latitud es $y = 0.175 + 0.21x - 0.002x^2$, donde x representa la latitud de la ubicación del nido y la y representa el número de pollitos en el nido. Utiliza el modelo cuadrático para completar la siguiente tabla. Luego, dibuja la gráfica de la curva cuadrática en el diagrama de dispersión que se proporciona al comienzo del Grupo de problemas.

x (grados)	y
30	
40	
50	
60	
70	

3. Según este modelo cuadrático, ¿cuál es la mejor latitud para empollar la mayor cantidad de pollitos? Justifica tu elección.

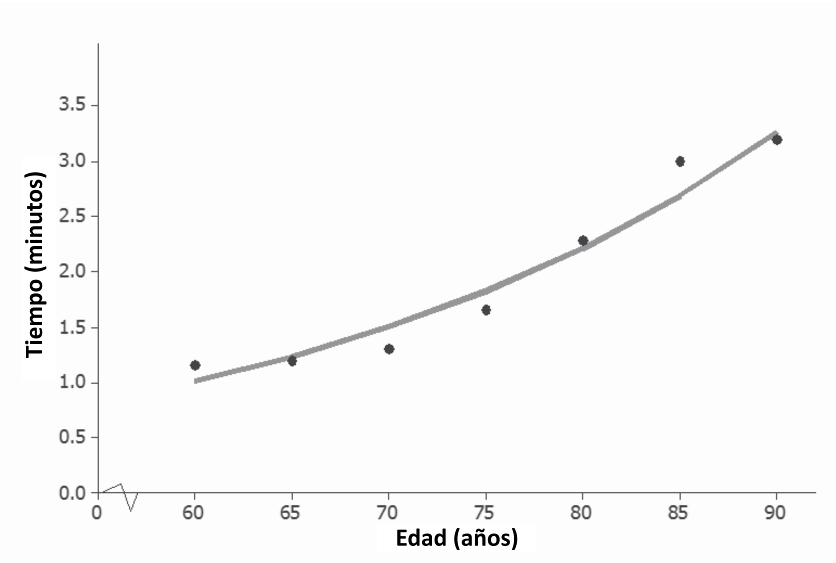
Supón que un grupo de científicos sociales llevan a cabo un estudio sobre adultos mayores para ver cómo cambia, con la edad, la cantidad de tiempo (en minutos) que se requiere para resolver un crucigrama. El siguiente diagrama de dispersión muestra los datos de este estudio.

Sea x igual a la edad de los adultos y y igual al tiempo (en minutos) que requieren para resolver un crucigrama siete participantes del estudio.



4. ¿Qué tipo de modelo (lineal, cuadrático o exponencial) utilizarías para describir la relación entre la edad y el tiempo que se requiere para completar el crucigrama?

5. Un modelo que podría representar la relación entre la edad y el tiempo que se requiere para completar el crucigrama es $y = 10^{-1.01\,+\,0.017x}$. Esta curva exponencial se muestra en el siguiente diagrama de dispersión. ¿Es este modelo un método efectivo para describir la relación entre la edad y el tiempo que se requiere para completar el crucigrama? Explica por qué sí o por qué no.



6. Según este modelo exponencial, ¿cuánto tiempo predices que tardaría una persona de 78 años en completar el crucigrama?

Lección 14: Representar relaciones con una recta

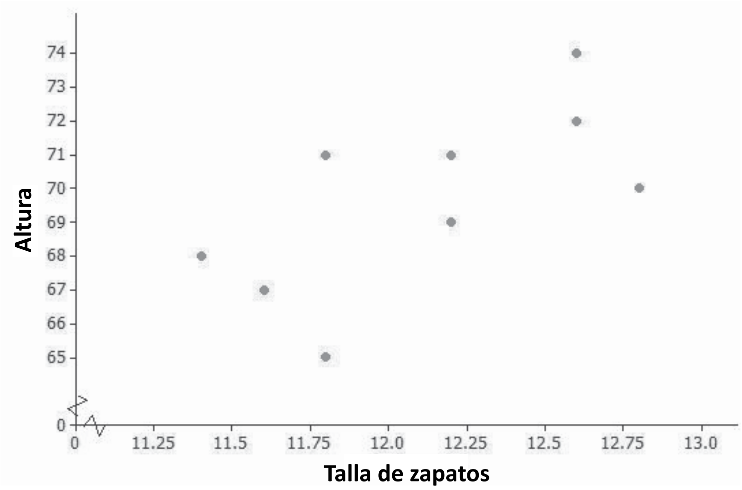
Trabajo en clase

Ejemplo 1: Utilizar una recta para describir una relación

A Kendra le gusta mirar programas de investigación de escenas de crímenes por televisión. Miró un programa donde los investigadores utilizaban una huella de zapato para identificar a un sospechoso en un caso. Kendra se preguntó cómo era posible predecir la altura de una persona a partir de la huella de su zapato.

Para investigarlo, recopiló datos sobre la talla de zapatos (en pulgadas) y la altura (en pulgadas) de 10 hombres adultos. Sus datos aparecen en la siguiente tabla y diagrama de dispersión.

x (talla de zapatos)	y (altura)
12.6	74
11.8	65
12.2	71
11.6	67
12.2	69
11.4	68
12.8	70
12.2	69
12.6	72
11.8	71



Ejercicios 1 y 2

1. ¿Existe alguna relación entre la talla de zapatos y la altura?

2. ¿Cómo describirías esa relación? ¿Tienden a ser más altos los hombres con mayor talla de zapatos?

Ejemplo 2: Utilizar modelos para hacer predicciones

Cuando dos variables x y y tienen una relación lineal, puedes utilizar una recta para describir su relación. También puedes utilizar la ecuación de la recta para predecir el valor de la variable y según el valor de la variable x.

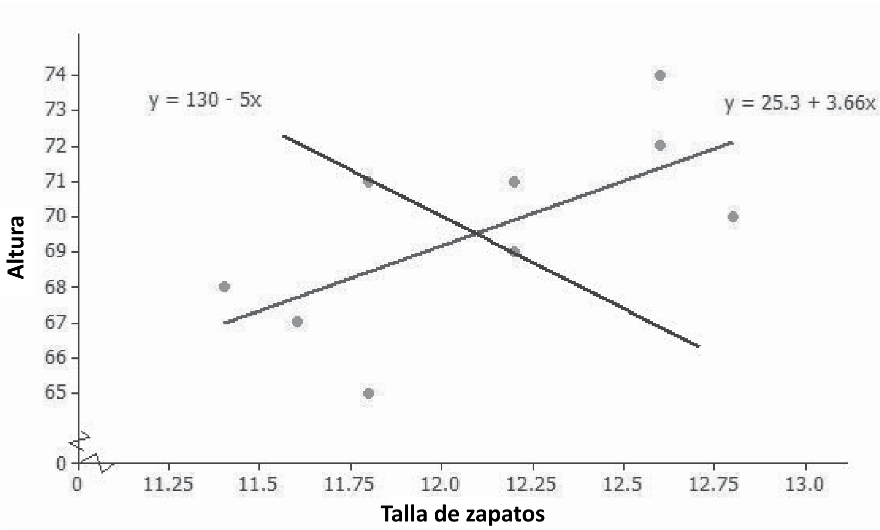
Por ejemplo, la recta $y = 25.3 + 3.66x$ puede utilizarse para describir la relación entre la talla de zapatos y la altura, donde x representa la talla de zapatos y y representa la altura. Para predecir la altura de un hombre con una talla de zapatos de 12 in, debes reemplazar la x por 12 en la ecuación de la recta y calcular el valor de y.

$$y = 25.3 + 3.66x = 25.3 + 3.66(12) = 69.22$$

La predicción de la altura de un hombre con una talla de zapatos de 12 in es 69.22 in.

Ejercicios 3 a 7

3. A continuación, se presenta un diagrama de dispersión de los datos con dos modelos lineales, $y = 130 - 5x$ y $y = 25.3 + 3.66x$. ¿Cuál de estos dos modelos es más efectivo para describir la relación que existe entre la talla de zapatos (x) y la altura (y)? Explica tu elección.



4. Uno de los hombres de la muestra tiene una talla de zapatos de 11.8 in y una altura de 71 in. Encierra en un círculo el punto en el diagrama de dispersión del Ejercicio 3 que representa a este hombre.

5. Supón que no conoces la altura de este hombre, pero sí sabes que su talla de zapatos es 11.8 in. Si utilizaras el modelo $y = 25.3 + 3.66x$, ¿cuál sería tu predicción sobre su altura? Si utilizaras el modelo $y = 130 - 5x$, ¿cuál sería tu predicción sobre su altura?

6. ¿Qué modelo está más cerca de 71 in, la altura real? ¿Es ese modelo el que mejor se ajusta a los datos? Explica tu respuesta.

7. ¿Existe una mejor forma de decidir qué recta proporciona una mejor descripción de una relación (en lugar de comparar el valor predicho con el valor real de uno de los puntos de datos de la muestra)?

Ejemplo 3: Residuos

Una forma de pensar sobre qué tan útil es una recta para describir la relación entre dos variables es utilizar la recta para predecir los valores de y para los puntos del diagrama de dispersión. Luego, estos valores predichos se pueden comparar con los valores reales de y.

Por ejemplo, el primer punto de datos de la tabla representa a un hombre con una talla de zapatos de 12.6 in y una altura de 74 in. Si utilizas la recta $y = 25.3 + 3.66x$ para predecir la altura de este hombre, obtienes lo siguiente:

$$\begin{aligned} y &= 25.3 + 3.66x \\ &= 25.3 + 3.66(12.6) \\ &= 71.42 \end{aligned}$$

Se predice que su altura es 71.42 in. Dado que su altura real es 74 in, puedes calcular el error de predicción si restas el valor predicho del valor real. Este error de predicción se denomina *residuo*. Para el primer punto de datos, el residuo se calcula de la siguiente manera:

$$\begin{aligned} \text{Residuo} &= \text{valor real de } y - \text{valor predicho de } y \\ &= 74 - 71.42 \\ &= 2.58 \end{aligned}$$

Ejercicios 8 a 10

8. Para la recta $y = 25.3 + 3.66x$, calcula los valores faltantes y añádelos a la tabla para completarla.

x (talla de zapatos)	y (altura)	Valor predicho de y	Residuo
12.6	74	71.42	2.58
11.8	65		-3.49
12.2	71		
11.6	67	67.76	-0.76
12.2	69	69.95	-0.95
11.4	68	67.02	
12.8	70	72.15	-2.15
12.2	69		-0.95
12.6	72	71.42	0.58
11.8	71	68.49	2.51

9. ¿Por qué el residuo de la primera fila de la tabla es positivo y el residuo de la segunda fila es negativo?

10. ¿Cuál es la suma de los residuos? ¿Por qué obtuviste un número cercano a cero en esta suma? ¿Esto significa que todos los residuos eran cercanos a 0?

Ejercicios 11 a 13

Cuando utilizas una recta para describir la relación entre dos variables numéricas, la *mejor* recta es la que hace que los residuos sean lo más bajos posibles en general.

11. Si los residuos tienden a ser bajos, ¿qué te indica esto sobre cuánto se ajusta la recta a los datos?

La elección más común para la *mejor* recta es la recta que hace que la suma de los residuos *elevados al cuadrado* sea lo más baja posible. Añade una columna a la derecha de la tabla del Ejercicio 8. Calcula el cuadrado de cada residuo y escribe la respuesta en la nueva columna.

12. ¿Por qué utilizamos la suma de los residuos elevados al cuadrado en lugar de simplemente sumar los residuos (sin elevarlos al cuadrado)? Pista: piensa si la suma de los residuos para la recta puede ser baja incluso si los errores de predicción son altos. ¿Puede ocurrir esto con los residuos elevados al cuadrado?

13. ¿Cuál es la suma de los residuos elevados al cuadrado para la recta $y = 25.3 + 3.66x$ y los datos del Ejercicio 11?

Ejemplo 4: La recta de mínimos cuadrados (recta de mejor ajuste)

La recta que tiene la suma de residuos elevados al cuadrado más baja para este conjunto de datos en comparación con el resto de las rectas se denomina *recta de mínimos cuadrados*. Esta recta también se llama *recta de mejor ajuste* o *recta que mejor se ajusta* (o recta de regresión).

Para los datos de talla de zapatos y altura de la muestra de 10 hombres, la recta $y = 25.3 + 3.66x$ es la recta de mínimos cuadrados. Ninguna otra recta tiene una suma menor de residuos elevados al cuadrado para este conjunto de datos.

Se pueden utilizar algunas ecuaciones para calcular el valor de la pendiente y el intercepto de la recta de mínimos cuadrados, pero estas fórmulas requieren muchos cálculos tediosos. Afortunadamente, se puede utilizar una calculadora gráfica para hallar la ecuación de la recta de mínimos cuadrados.

Tu maestro te mostrará cómo ingresar datos en la calculadora gráfica u otro programa de estadísticas para obtener la ecuación de la recta de mínimos cuadrados.

Ejercicios 14 a 17

14. Ingresa los datos de la talla de zapatos y la altura y, luego, utiliza tu calculadora para hallar la ecuación de la recta de mínimos cuadrados. ¿Obtuviste $y = 25.3 + 3.66x$? (La pendiente y el intercepto de y se redondean a la centésima más cercana).

15. Si suponemos que los 10 hombres de la muestra representan a los hombres adultos en general, ¿qué altura predecirías que tiene un hombre cuya talla de zapatos es 12.5 in? ¿Qué altura predecirías que tiene un hombre cuya talla de zapatos es 11.9 in?

Una vez que hayas hallado la ecuación de la recta de mínimos cuadrados, los valores de la pendiente y el intercepto de y de la recta suelen revelar datos interesantes sobre la relación que estás representando.

La pendiente de la recta de mínimos cuadrados es el cambio en el valor predicho de la variable y asociada con un incremento de uno en el valor de la variable x.

16. Brinda una interpretación de la pendiente de la recta de mínimos cuadrados $y = 25.3 + 3.66x$ para predecir la altura a partir de la talla de zapatos de hombres adultos.

El intercepto de y de una recta es el valor predicho de y cuando x es igual a cero. Cuando utilizas una recta como modelo de la relación entre dos variables numéricas, no suele tener sentido interpretar el intercepto de y porque un valor cero para x puede no tener ningún sentido.

17. Explica por qué no tiene sentido interpretar el intercepto de y de 25.3 como la altura predicha de un varón adulto cuya talla de zapatos es cero.

Resumen de la lección

Cuando la relación entre dos variables numéricas x y y es lineal, se puede utilizar una línea recta para describir la relación. Luego, esta recta se puede utilizar para predecir el valor de y según el valor de x. Cuando se hace una predicción, el error de predicción es la diferencia entre el valor real de y y el valor predicho de y. El error de predicción se denomina *residuo*. El residuo se calcula de esta manera: *residuo* $=$ *valor real de* y $-$ *valor predicho de* y. La *recta de mínimos cuadrados* es la recta que se utiliza para representar una relación lineal. La recta de mínimos cuadrados es la *mejor* recta en el sentido de que tiene la suma de residuos elevados al cuadrado más baja de todas las rectas.

Grupo de problemas

Kendra se pregunta si la relación entre la talla de zapatos y la altura puede ser diferente para hombres y para mujeres. Para investigarlo, también recopiló datos sobre la talla de zapatos (en pulgadas) y la altura (en pulgadas) de 12 mujeres.

x (talla de zapatos de mujeres)	y (altura de mujeres)
8.9	61
9.6	61
9.8	66
10.0	64
10.2	64
10.4	65
10.6	65
10.6	67
10.5	66
10.8	67
11.0	67
11.8	70

1. Construye un diagrama de dispersión con estos datos.

2. ¿Existe alguna relación entre la talla de zapatos y la altura de estas 12 mujeres?

3. Halla la ecuación de la recta de mínimos cuadrados. (Redondea los valores a la centésima más cercana).

4. Supón que estas 12 mujeres representan a las mujeres adultas en general. Según la recta de mínimos cuadrados, ¿qué altura predecirías que tiene una mujer cuya talla de zapatos es 10.5 in? ¿Qué altura predecirías que tiene una mujer cuya talla de zapatos es 11.5 in?

5. Una de las mujeres de la muestra tenía una talla de zapatos de 9.8 in. Según la recta de regresión, ¿qué altura predecirías que tiene esa mujer?

6. ¿Cuál es el valor del residuo asociado con la observación sobre la mujer cuya talla de zapatos era 9.8 in?

7. Añade el valor predicho y el residuo que acabas de calcular a la siguiente tabla. Luego, calcula la suma de los residuos elevados al cuadrado.

x (talla de zapatos de mujeres)	y (altura de mujeres)	Altura predicha (in)	Residuo (in)	Residuo elevado al cuadrado
8.9	61	60.72	0.28	
9.6	61	62.92	−1.92	
9.8	66			
10.0	64	64.18	−0.18	
10.2	64	64.81	−0.81	
10.4	65	65.44	−0.44	
10.6	65	66.07	−1.07	
10.6	67	66.07	0.93	
10.5	66	65.76	0.24	
10.8	67	66.7	0.3	
11.0	67	67.33	−0.33	
11.8	70	69.85	0.15	

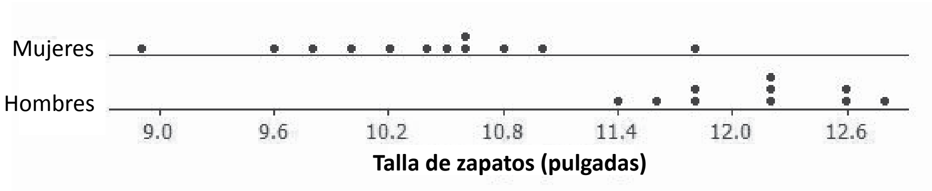
8. Brinda una interpretación de la pendiente de la recta de mínimos cuadrados.

9. ¿Tiene sentido interpretar el intercepto de y de la recta de mínimos cuadrados en este contexto? Explica por qué sí o por qué no.

10. ¿La suma de los residuos elevados al cuadrado para la recta $y = 25 + 2.8x$ sería mayor que, casi igual a o menor que la suma que calculaste en el Problema 7? Explica cómo lo sabes. Deberías poder responder esta pregunta sin calcular la suma de los residuos elevados al cuadrado para esta nueva recta.

11. Para los hombres, la recta de mínimos cuadrados que describe la relación entre x, que representa la talla de zapatos (en pulgadas), y la y, que representa la altura (en pulgadas), era $y = 25.3 + 3.66x$. ¿Cuál es la diferencia entre esta y la ecuación de la recta de mínimos cuadrados para las mujeres? ¿Utilizarías $y = 25.3 + 3.66x$ para predecir la altura de una mujer según su talla de zapatos? Explica por qué sí o por qué no.

12. A continuación, se presentan diagramas de puntos de la talla de zapatos de las mujeres y la talla de zapatos de los hombres. Supón que hallaste una huella de zapato y que cuando mediste la talla, obtuviste 10.8 in. ¿Quién crees que dejó esa huella: un hombre o una mujer? Explica tu elección.



13. Supón que hallaste una huella de zapato y que su talla es 12 in. ¿Cuál sería tu predicción de la altura de la persona que dejó esa huella? Explica cómo llegaste a esa respuesta.

Esta página queda en blanco intencionalmente.

Lección 15: Interpretar residuos de una recta

Trabajo en clase

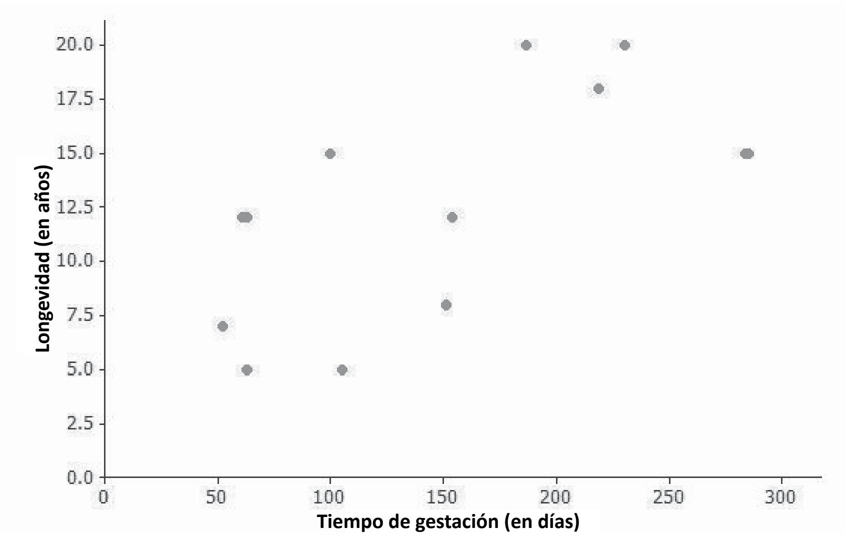
Ejemplo 1: Calcular el error de predicción

El tiempo de gestación de un animal es el lapso de tiempo típico que transcurre entre su concepción y el nacimiento. La longevidad es la expectativa de vida típica para ese animal. En la tabla a continuación, se muestra el tiempo de gestación (representado en días) y la longevidad (representada en años) para 13 tipos de animales diferentes.

Animal	Tiempo de gestación (en días)	Longevidad (en años)
Babuino	187	20
Oso negro	219	18
Castor	105	5
Bisonte	285	15
Gato	63	12
Chimpancé	230	20
Vaca	284	15
Perro	61	12
Zorro rojo	52	7
Cabra	151	8
León	100	15
Oveja	154	12
Lobo	63	5

Fuente de datos: *Core Math Tools*, http://nctm.org

Este es el diagrama de dispersión para el conjunto de datos:

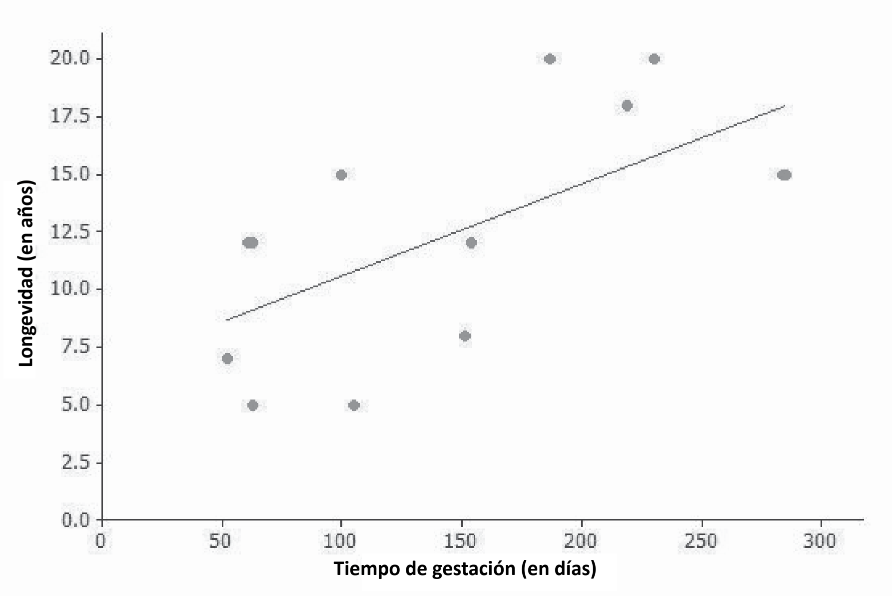


Ejercicios 1 a 4

Al hallar la ecuación para la recta de mínimos cuadrados que relacione la longevidad con el tiempo de gestación para estos animales, hallarás la ecuación para predecir la longevidad. ¿Qué tan buena es la recta? En otras palabras, si se da el tiempo de gestación de otro tipo de animal que no está incluido en la lista original, ¿qué tan precisa resultará la recta de mínimos cuadrados para predecir la longevidad de ese animal?

1. Usando una calculadora gráfica, verifica que la ecuación de la recta de mínimos cuadrados sea $y = 6.642 + 0.03974x$, donde x represente el tiempo de gestación (en días), y y represente la longevidad (en años).

   Se agregó la recta de mínimos cuadrados al siguiente diagrama de dispersión.



2. Supón que un tipo de animal en particular tiene un tiempo de gestación de 200 días. Aproximadamente, ¿qué valor predice la recta para la longevidad de ese tipo de animal?

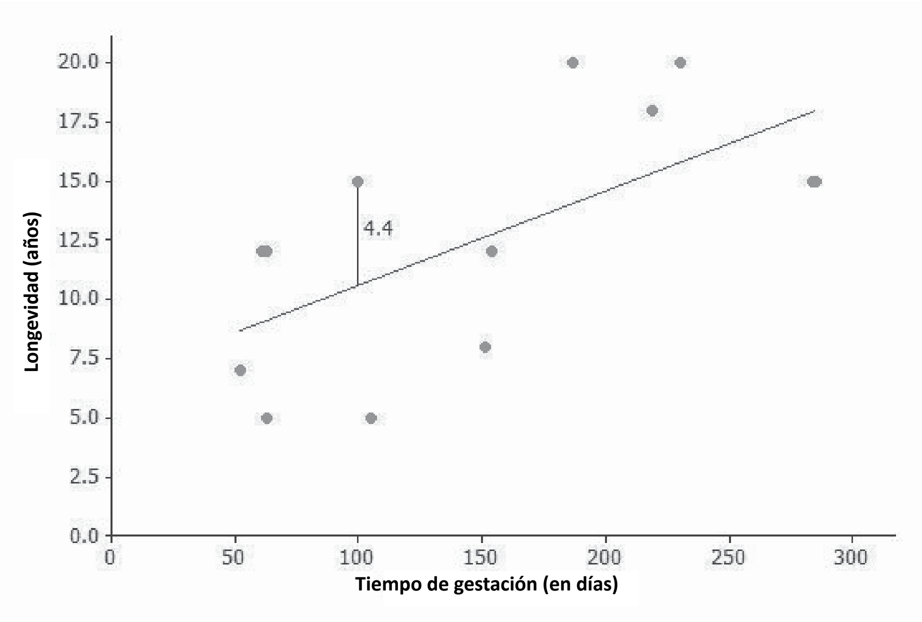
3. ¿El valor que obtuviste en la predicción del Ejercicio 2 será necesariamente el valor exacto para la longevidad de ese tipo de animal? ¿La longevidad para ese tipo de animal podría ser más larga que lo predicho? ¿Podría ser más corta?

Puedes investigar más a fondo observando los tipos de animales incluidos en el conjunto de datos original. Tomemos al león, por ejemplo. Su periodo de gestación es de 100 días. También sabemos que su longevidad es de 15 años, pero ¿cuál es la *predicción* de la recta de mínimos cuadrados para la longevidad del león?

Si sustituyes $x = 100$ días en la ecuación, obtienes que $y = 6.642 + 0.03974(100)$ o aproximadamente 10.6. La recta de mínimos cuadrados predice que la longevidad del león es de 10.6 años aproximadamente.

4. ¿Qué tan cerca está esto de ser correcto? Más precisamente, ¿cuánto debes agregar a 10.6 para obtener la longevidad real del león que es de 15 años?

Puedes representar gráficamente el error de predicción de 4.4 años de la siguiente forma:



Ejercicios 5 y 6

5. Continuemos pensando en el tiempo de gestación y longevidad de los animales. Específicamente, investiguemos qué tan precisa puede resultar la recta de mínimos cuadrados para predecir la longevidad del oso negro.

   a. ¿Cuál es el tiempo de gestación para un oso negro?

b. Observa la gráfica. A simple vista, ¿cuál es la predicción de la recta de mínimos cuadrados para la longevidad del oso negro?

c. Utiliza el tiempo de gestación de la parte (a) y la recta de mínimos cuadrados $y = 6.642 + 0.03974x$ para predecir la longevidad del oso negro. Redondea tu respuesta a la décima más cercana.

d. ¿Cuál es la longevidad real para el oso negro?

e. ¿Cuánto debes agregar al valor predicho para obtener la longevidad real del oso negro?

f. Muestra tu respuesta de la parte (e) en la gráfica como un segmento de recta vertical.

6. Repite esta actividad para la oveja.

a. Sustituye x por el tiempo de gestación de la oveja en la ecuación para hallar el valor predicho para la longevidad de la oveja. Redondea tu respuesta a la décima más cercana.

b. ¿Qué deberías agregar al valor predicho para obtener el valor real de la longevidad de la oveja? (Pista: la respuesta tendrá un valor negativo).

c. Muestra tu respuesta de la parte (b) en la gráfica como un segmento de recta vertical. Escribe un enunciado que describa los puntos en la gráfica para los que deberías agregar un número negativo al valor predicho para obtener el valor real.

Ejemplo 2: Residuos como errores de predicción

En los ejercicios anteriores, averiguaste cuánto debías agregar al valor predicho a fin de hallar el valor real de la longevidad de un animal. Para hacerlo, calculaste

el valor real − el valor predicho.

Esa cantidad se denomina residuo. Se resume como

residuo = valor real de y − valor predicho de y.

Ahora puedes calcular los residuos para todos los puntos del ejemplo de longevidad de los animales. Los valores de los residuos se muestran en la tabla a continuación.

Animal	**Tiempo de gestación (en días)**	**Longevidad (en años)**	**Residuo (en años)**
Babuino	187	20	5.9
Oso negro	219	18	2.7
Castor	105	5	-5.8
Bisonte	285	15	-3.0
Gato	63	12	2.9
Chimpancé	230	20	4.2
Vaca	284	15	-2.9
Perro	61	12	2.9
Zorro rojo	52	7	-1.7
Cabra	151	8	-4.6
León	100	15	4.4
Oveja	154	12	-0.8
Lobo	63	5	-4.1

Estos residuos muestran que la longevidad real de un animal debe estar dentro de los seis años con respecto al valor predicho de la recta de mínimos cuadrados.

Supón que seleccionaste un tipo de animal que no está incluido en el conjunto de datos originales y que el tiempo de gestación para este tipo de animal es de 270 días. Si sustituyes $x = 270$ en la ecuación de la recta de mínimos cuadrados, obtienes

$$\begin{aligned} y &= 6.642 + 0.03974(270) \\ &= 17.4. \end{aligned}$$

La longevidad predicha para este animal es de 17.4 años.

Ejercicios 7 y 8

Piensa cuál podría ser la longevidad *real* para este animal.

7. ¿Podría ser de 30 años? ¿Y de 5 años?

8. A juzgar por el tamaño del residuo en tu tabla, ¿qué valores crees que serían razonables para representar la longevidad de este tipo de animal?

Ejercicios 9 y 10

Continuemos pensando en el tiempo de gestación y la longevidad de los animales. Se sabe que el tiempo de gestación para el ocelote es de 85 días.

La recta de mínimos cuadrados predice que la longevidad del ocelote es de 10.0 años.

$$y = 6.642 + 0.03974(85) = 10.0$$

9. Si te basas en los residuos del Ejemplo 3, ¿te sorprendería descubrir que la longevidad del ocelote es de 2 años? ¿Por qué sí o por qué no? ¿Cuál sería un rango de valores sensato para la longevidad real del ocelote?

10. Se sabe que la longevidad real del ocelote es de 9 años. ¿Cuál es el residuo para el ocelote?

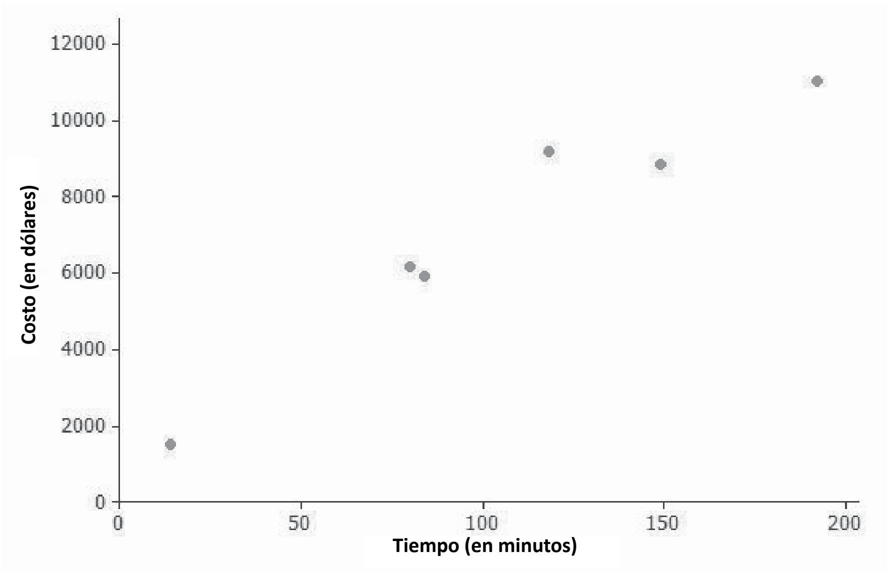
Resumen de la lección

- Cuando la recta de mínimos cuadrados se usa para calcular un valor predicho, el error de predicción puede medirse mediante
$$\text{residuo} = \text{valor real de } y - \text{valor predicho de } y.$$
- En la gráfica, los residuos están representados por las distancias verticales de los puntos respecto a la recta de mínimos cuadrados.
- Los residuos dan una idea de cuán próxima puede ser una predicción cuando la recta de mínimos cuadrados se usa para hacer la predicción de un valor que no está incluido en el conjunto de datos.

Grupo de problemas

Se registró el tiempo que seis pacientes estuvieron en cirugía y el costo de esas cirugías. Los resultados y el diagrama de dispersión se muestran a continuación.

Tiempo (en minutos)	Costo ($)
14	1,510
80	6,178
84	5,912
118	9,184
149	8,855
192	11,023



1. Calcula la ecuación de la recta de mínimos cuadrados que relacione el costo y el tiempo. (Indica la pendiente a la décima más cercana y el intercepto de y al número entero más cercano).

2. Traza la recta de mínimos cuadrados en la gráfica de arriba. (Pista: sustituye $x = 30$ en la ecuación para hallar el valor predicho y. Marca el punto $(30$, tu respuesta$)$ en la gráfica. Luego, sustituye $x = 180$ en la ecuación y marca el punto. Une los dos puntos con una regla).

3. ¿Cuál es la predicción de la recta de mínimos cuadrados para el costo de una cirugía que dura 118 min? (Calcula el costo al centavo más cercano).

4. ¿Cuánto debes agregar a tu respuesta del Problema 3 para obtener el costo real de la intervención quirúrgica para una cirugía que dura 118 min? (Este es el residuo).

5. Muestra tu respuesta al Problema 4 en forma de recta vertical entre el punto para esa persona en el diagrama de dispersión y la recta de mínimos cuadrados.

6. Recuerda que el residuo es el valor real de y menos el valor predicho de y. Calcula el residuo para una cirugía que duró 149 min y tuvo un costo de $\$8{,}855$.

7. Calcula los otros residuos y escribe todos los resultados en la tabla a continuación.

Tiempo (en minutos)	Costo ($)	Valor predicho ($)	Residuo ($)
14	1,510		
80	6,178		
84	5,912		
118	9,184		
149	8,855		
192	11,023		

8. Supón que una cirugía duró 100 min.
   a. ¿Cuál es la predicción de la recta de mínimos cuadrados para el costo de esta cirugía?
   b. ¿Te sorprendería si el costo real fuera $\$9{,}000$? ¿Por qué sí o por qué no?
   c. Analiza la pendiente de la recta de mínimos cuadrados.

Lección 16: Más sobre representar relaciones con una recta

Trabajo en clase

Ejemplo 1: Calcular residuos

El peso en vacío de un carro constituye el peso de ese carro sin equipaje ni pasajeros. La tabla a continuación muestra el peso en vacío (en cientos de libras) y el rendimiento del combustible (en millas por galón) de cinco carros compactos.

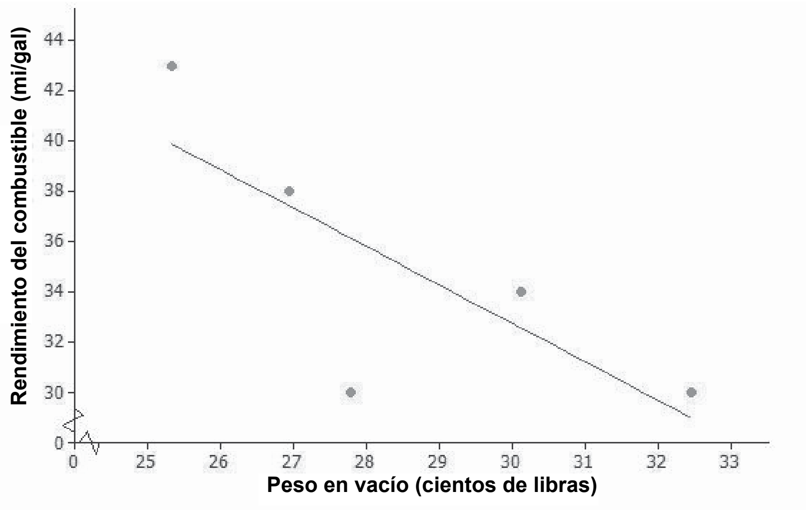
Peso en vacío (cientos de libras)	Rendimiento del combustible (millas por galón)
25.33	43
26.94	38
27.79	30
30.12	34
32.47	30

Con una calculadora, se halló que la ecuación para la recta de mínimos cuadrados de este conjunto de datos es:

$$y = 78.62 - 1.5290x,$$

donde x es el peso en vacío (en cientos de libras) y y es el valor predicho del rendimiento de combustible (en millas por galón).

A continuación, se muestra el diagrama de dispersión de este conjunto de datos, y la recta de mínimos cuadrados está representada en la gráfica.



Deberás calcular los residuos para los cinco puntos en el diagrama de dispersión. Observa el diagrama de dispersión antes de realizar los cálculos.

Ejercicios 1 y 2

1. ¿El residuo del carro cuyo peso en vacío es 25.33 (en cientos de libras) será negativo o positivo? Aproximadamente, ¿cuál es el valor del residuo para este punto?

2. ¿El residuo del carro cuyo peso en vacío es 27.79 (en cientos de libras) será negativo o positivo? Aproximadamente, ¿cuál es el valor del residuo para este punto?

El residuo para ambos pesos en vacío se calcula de la siguiente manera:

Sustituye $x = 25.33$ en la ecuación de la recta de mínimos cuadrados para hallar el valor predicho del rendimiento de combustible. $y = 78.62 - 1.5290(25.33)$ $= 39.9$ Ahora calcula el residuo. Residuo = valor real de y − valor predicho de y $= 43 \text{ mi/gal} - 39.9 \text{ mi/gal}$ $= 3.1 \text{ mi/gal}$	Sustituye $x = 27.79$ en la ecuación de la recta de mínimos cuadrados para hallar el valor predicho del rendimiento de combustible. $y = 78.62 - 1.5290(27.79)$ $= 36.1$ Ahora calcula el residuo. Residuo = valor real de y − valor predicho de y $= 30 \text{ mi/gal} - 36.1 \text{ mi/gal}$ $= -6.1 \text{ mi/gal}$

Estos dos residuos están registrados en la tabla a continuación.

Peso en vacío (cientos de libras)	**Rendimiento del combustible (mi/gal)**	**Residuo (mi/gal)**
25.33	43	3.1
26.94	38	
27.79	30	-6.1
30.12	34	
32.47	30	

Ejercicios 3 y 4

Continuemos con el Ejemplo 1 del peso en vacío del carro y el rendimiento de combustible.

3. Calcula los tres residuos restantes y anótalos en la tabla.

4. Supón que un carro tiene un peso en vacío de 31 (en cientos de libras).
   a. ¿Cuál es la predicción de la recta de mínimos cuadrados para el rendimiento del combustible de este carro?

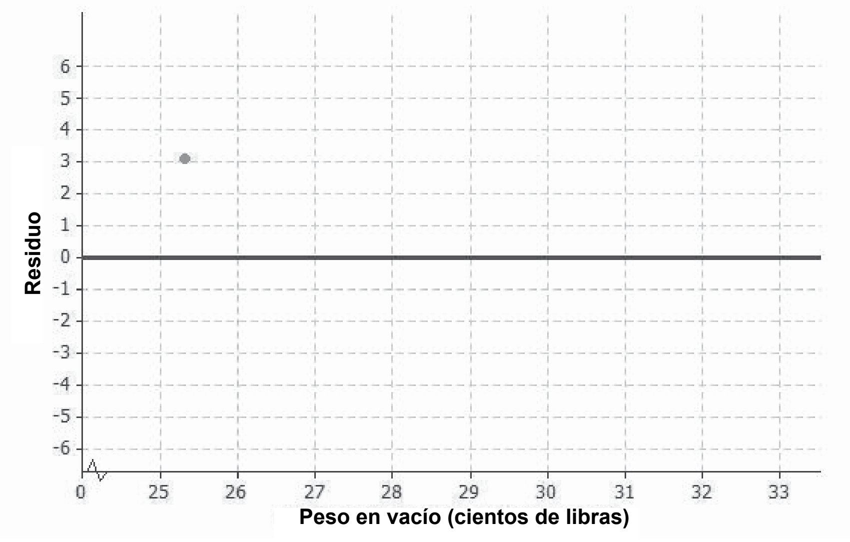
   b. ¿Te sorprendería si el rendimiento real del combustible de este carro fuera de 29 millas por galón? Explica tu respuesta.

Ejemplo 2: Hacer una gráfica de residuos para evaluar una recta

Con frecuencia, es útil realizar una gráfica de los residuos, denominada gráfica de residuos. Realiza la gráfica de residuos para el conjunto de datos del carro compacto.

Representa gráficamente la variable original de x (el peso en vacío, en este caso) sobre el eje horizontal y los residuos sobre el eje vertical. Para este ejemplo, necesitarás trazar un eje horizontal que vaya de 25 a 32 y un eje vertical con una escala que incluya los valores de los residuos que calculaste. Luego, marca el punto para el primer carro. El peso en vacío para el primer carro es 25.33 (en cientos de libras) y el residuo es 3.1 mi/gal. Marca el punto $(25.33, 3.1)$.

A continuación, se muestran los ejes y este primer punto.



Ejercicios 5 y 6

5. Marca los cuatro residuos restantes en la gráfica de residuos del Ejemplo 2.

6. ¿De qué manera se relaciona el patrón de los puntos en la gráfica de residuos con el patrón del diagrama de dispersión original? ¿Puedes saber cuál será el patrón en la gráfica de residuos al observar el diagrama de dispersión original?

Resumen de la lección

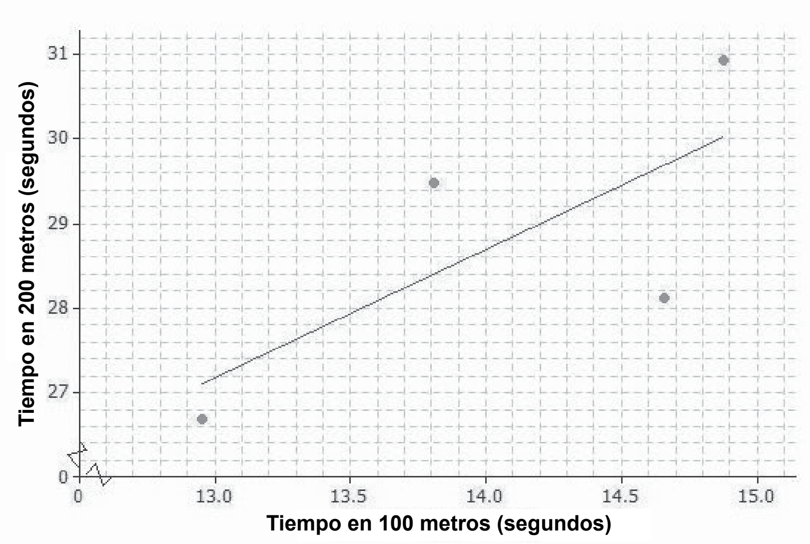
- El valor predicho de y se calcula mediante la ecuación de la recta de mínimos cuadrados.
- El residuo se calcula mediante
$$\text{residuo} = \text{valor real de } y - \text{valor predicho de } y.$$
- La suma de los residuos nos proporciona una idea del grado de exactitud cuando usamos la recta de mínimos cuadrados para realizar predicciones.
- Para hacer una gráfica de residuos, ubica los valores de x sobre el eje horizontal y los residuos sobre el eje vertical.

Grupo de problemas

Cuatro atletas de un equipo de pista comparan sus marcas personales en las carreras de 100 y 200 metros. A continuación, se muestra una tabla con sus mejores marcas.

Atleta	Tiempo en 100 metros (segundos)	Tiempo en 200 metros (segundos)
1	12.95	26.68
2	13.81	29.48
3	14.66	28.11
4	14.88	30.93

A continuación, se muestra un diagrama de dispersión de estos resultados (incluida la recta de mínimos cuadrados).



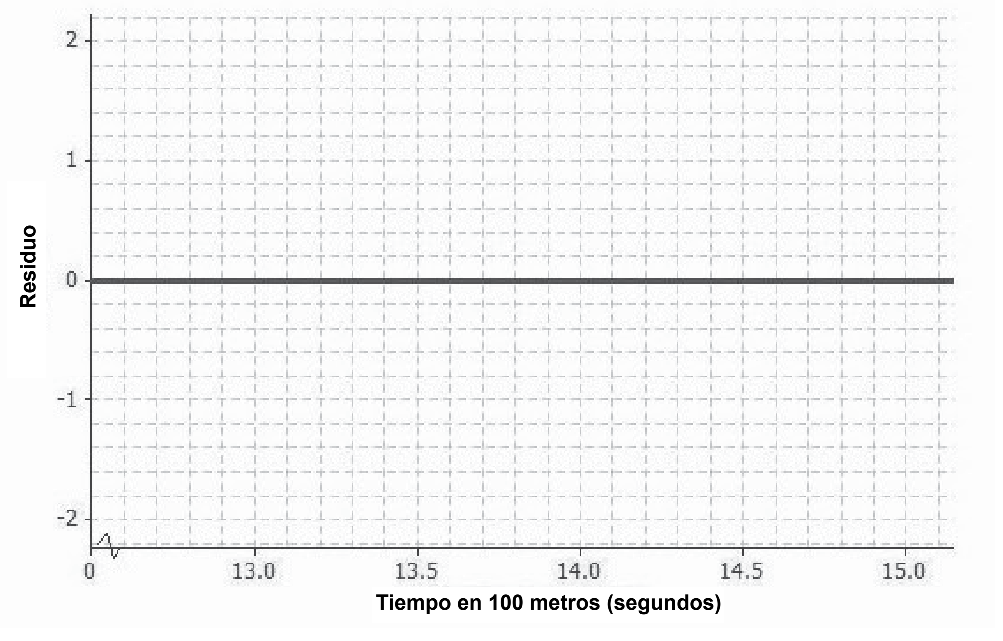
1. Utiliza tu calculadora o computadora para hallar la ecuación de la recta de mínimos cuadrados.

2. Utiliza esa ecuación para hallar la predicción de tiempo en 200 metros para el corredor cuyo tiempo en los 100 metros fue de 12.95 segundos. ¿Cuál es el residuo para este atleta?

3. Calcula los residuos para los otros tres atletas. Escribe los residuos en la tabla a continuación.

Atleta	Tiempo en 100 metros (segundos)	Tiempo en 200 metros (segundos)	Residuo (segundos)
1	12.95	26.68	
2	13.81	29.48	
3	14.66	28.11	
4	14.88	30.93	

4. Construye una gráfica de residuos para este conjunto de datos a partir de los ejes a continuación.



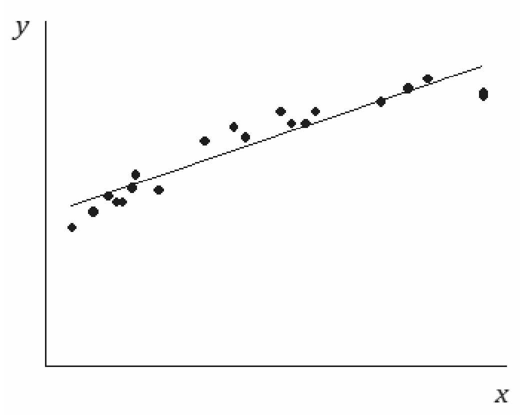
EUREKA MATH

Lección 17: Analizar residuos

Trabajo en clase

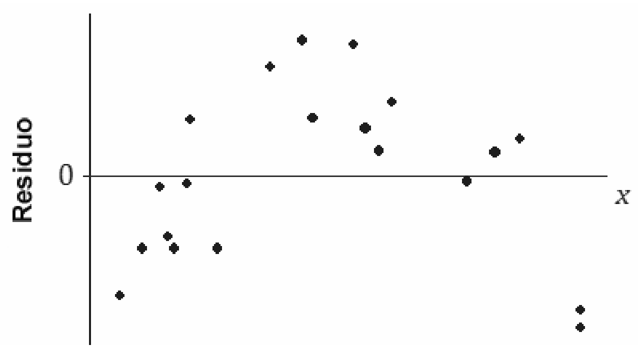
Ejemplo 1: Predecir el patrón en una gráfica de residuos

Supón que tienes un diagrama de dispersión y una recta de mínimos cuadrados que se ve así:



Describe cómo crees que se verá la gráfica de residuos.

La gráfica de residuos tiene una forma de arco como la siguiente:

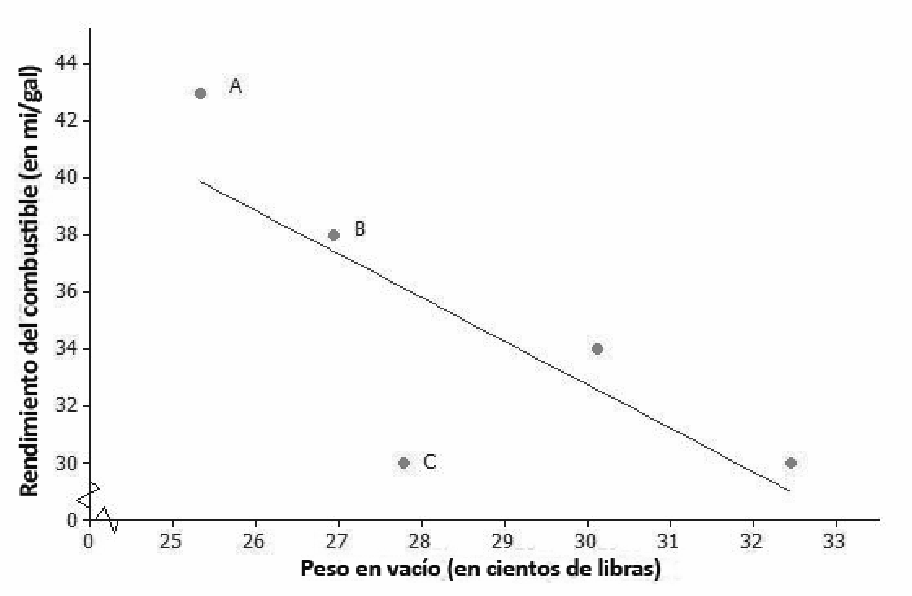


¿Por qué es importante observar el patrón en la gráfica de residuos?

Ejemplo 2: El significado de los residuos

Supón que tienes un diagrama de dispersión y que has trazado la recta de mínimos cuadrados. Recuerda que el residuo de un punto en el diagrama de dispersión es la distancia vertical de ese punto respecto de la recta de mínimos cuadrados.

En la lección anterior, vimos un diagrama de dispersión que mostraba el rendimiento del combustible en relación con el peso en vacío de cinco carros compactos. A continuación, se muestran el diagrama de dispersión y la recta de mínimos cuadrados.



Considera las siguientes preguntas:

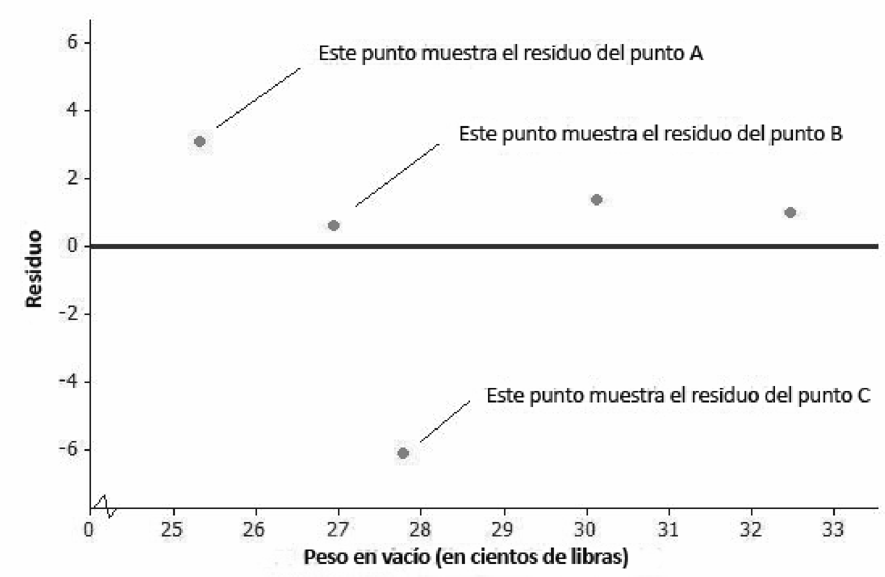
- ¿Qué tipo de residuo tiene el Punto A?

- ¿Qué tipo de residuo tiene el Punto B?

- ¿Qué tipo de residuo tiene el Punto C?

EUREKA MATH

Observa también la gráfica de residuos para este conjunto de datos:



El maestro mostrará cómo usar una calculadora gráfica o un programa gráfico para construir un diagrama de dispersión y una gráfica de residuos. Considera el siguiente ejemplo.

Ejemplo 3: Usar una calculadora gráfica para construir una gráfica de residuos

En una lección anterior, observamos un conjunto de datos que representaba la talla de zapatos y la altura de 12 mujeres adultas. El conjunto de datos se muestra en la tabla a continuación.

x **(talla de zapatos)**	y **(altura)**
pulgadas	pulgadas
8.9	61
9.6	61
9.8	66
10.0	64
10.2	64
10.4	65
10.6	65
10.6	67
10.5	66
10.8	67
11.0	67
11.8	70

Utiliza una calculadora para construir el diagrama de dispersión (con la recta de mínimos cuadrados) y la gráfica de residuos para este conjunto de datos.

Resumen de la lección

- Luego de ajustar una recta, se puede construir la gráfica de residuos utilizando una calculadora gráfica.
- La presencia de un patrón en la gráfica de residuos indica que la relación en el conjunto de datos original no es lineal.

Grupo de problemas

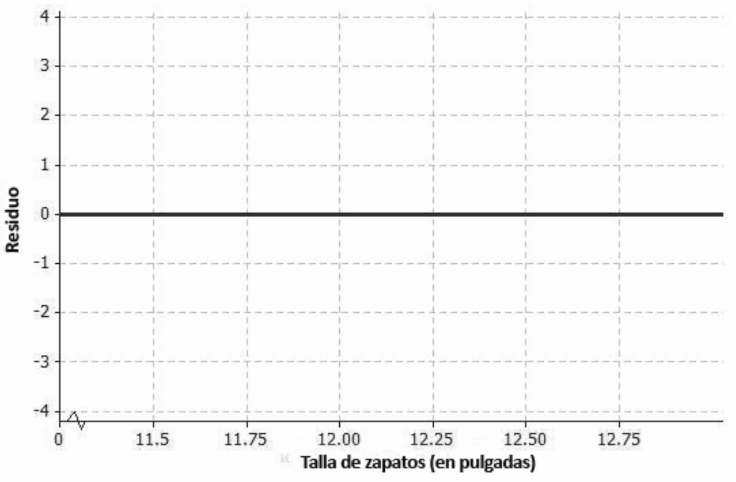
Considera nuevamente un conjunto de datos con la talla de zapatos y la altura de 10 hombres adultos. El conjunto de datos se muestra en la tabla a continuación.

x (talla de zapatos)	y (altura)
pulgadas	pulgadas
12.6	74
11.8	65
12.2	71
11.6	67
12.2	69
11.4	68
12.8	70
12.2	69
12.6	72
11.8	71

1. Utiliza tu calculadora o un programa gráfico para construir el diagrama de dispersión para este conjunto de datos. Incluye en tu gráfica la recta de mínimos cuadrados. Explica qué indica la pendiente de la recta de mínimos cuadrados con respecto a la talla de zapatos y la altura.

2. Utiliza tu calculadora para construir la gráfica de residuos para este conjunto de datos.

3. Dibuja la gráfica de residuos sobre los ejes que se dan a continuación. ¿Indica la dispersión de los puntos en la gráfica de residuos una relación lineal en el conjunto de datos original? Explica tu respuesta.



EUREKA MATH

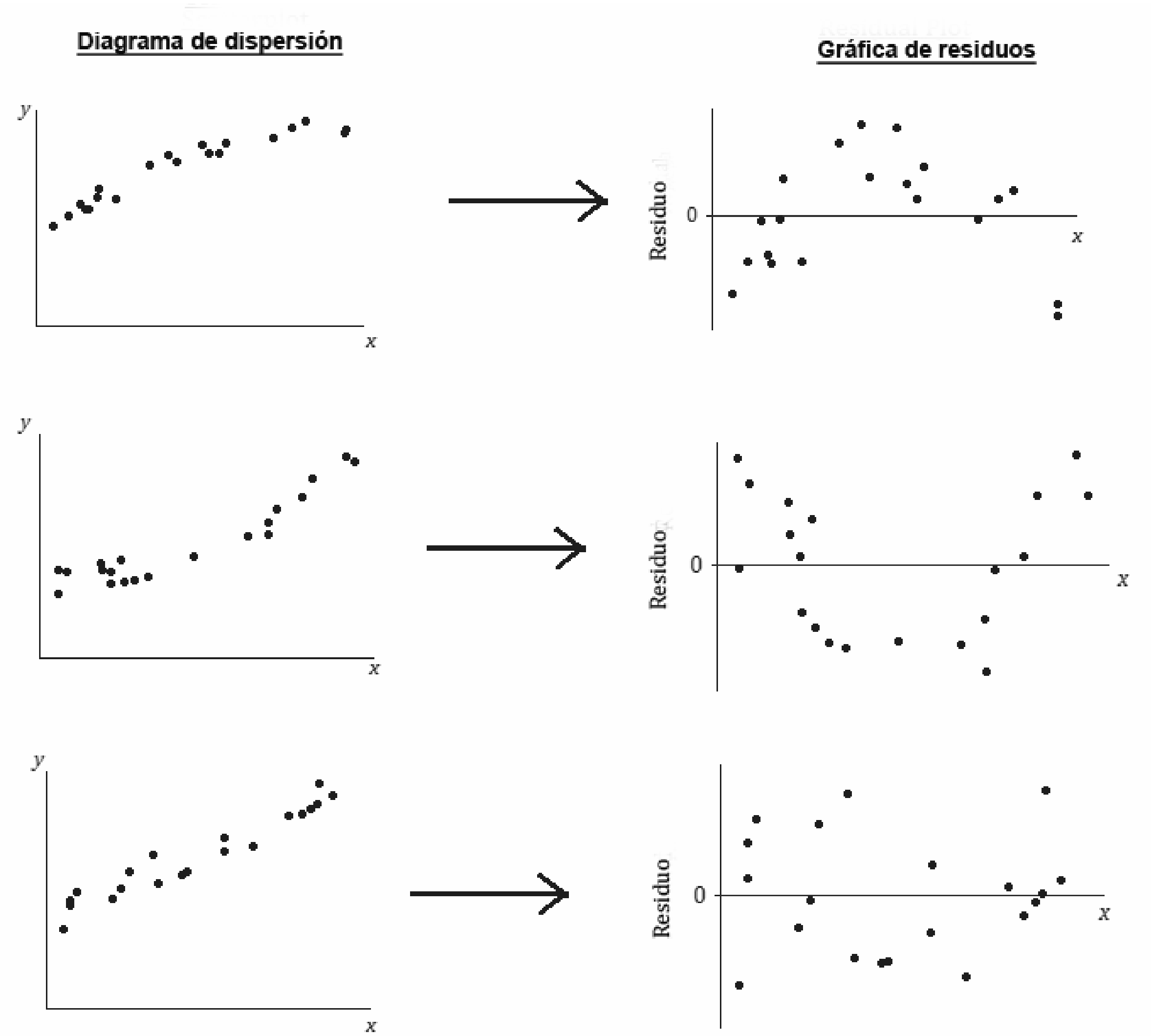
Lección 18: Analizar residuos

Trabajo en clase

La lección anterior muestra que, cuando los datos se ajustan a una recta, un diagrama de dispersión con un patrón curvo produce una gráfica de residuos que muestra un patrón claro. También viste que, cuando se ajusta una recta, el diagrama de dispersión donde los puntos muestran un patrón de línea recta da como resultado una gráfica de residuos donde los puntos están dispersos de manera aleatoria.

Ejemplo 1: La relevancia del patrón en la gráfica de residuos

Nuestros hallazgos previos se resumen en los siguientes diagramas:



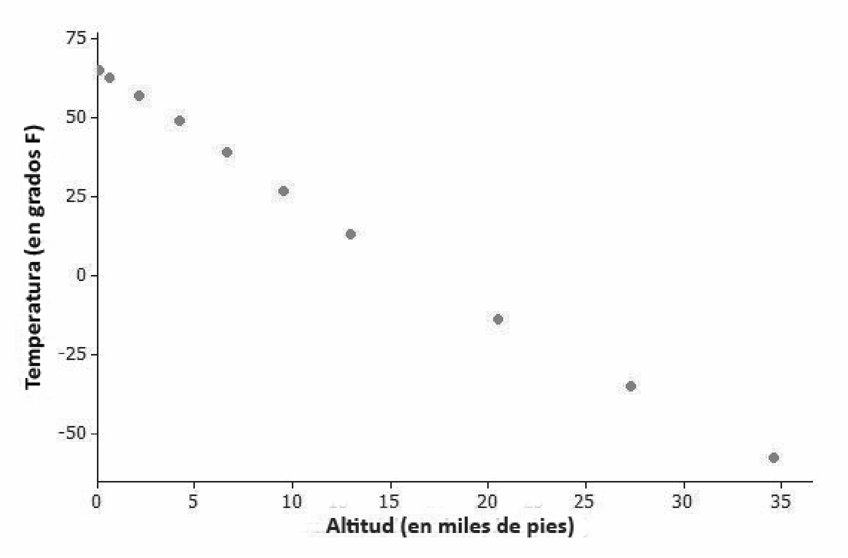
¿Qué significa que haya un patrón curvo en la gráfica de residuos?

¿Qué significa que los puntos de la gráfica de residuos parezcan estar dispersos aleatoriamente sin un patrón visible?

¿Por qué no se podría simplemente mirar el diagrama de dispersión del conjunto de datos original? ¿Por qué era necesaria la gráfica de residuos? El siguiente ejemplo responde a estas preguntas.

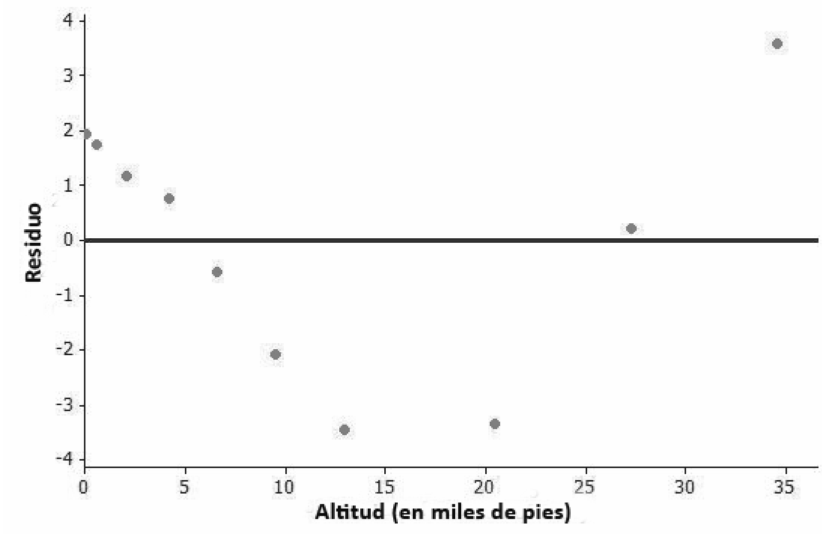
Ejemplo 2: ¿Por qué necesitas una gráfica de residuos?

Se midió la temperatura (en grados Fahrenheit) en distintas altitudes (en miles de pies) de Los Ángeles. El diagrama de dispersión (a continuación) parece mostrar una relación lineal (línea recta) entre estas dos cantidades.



Fuente de datos: *Core Math Tools*, http://nctm.org

Sin embargo, mira la gráfica de residuos:



Hay una curva clara en la gráfica de residuos. Por lo tanto, lo que parecía una relación lineal en el diagrama de dispersión original era, de hecho, una relación no lineal.

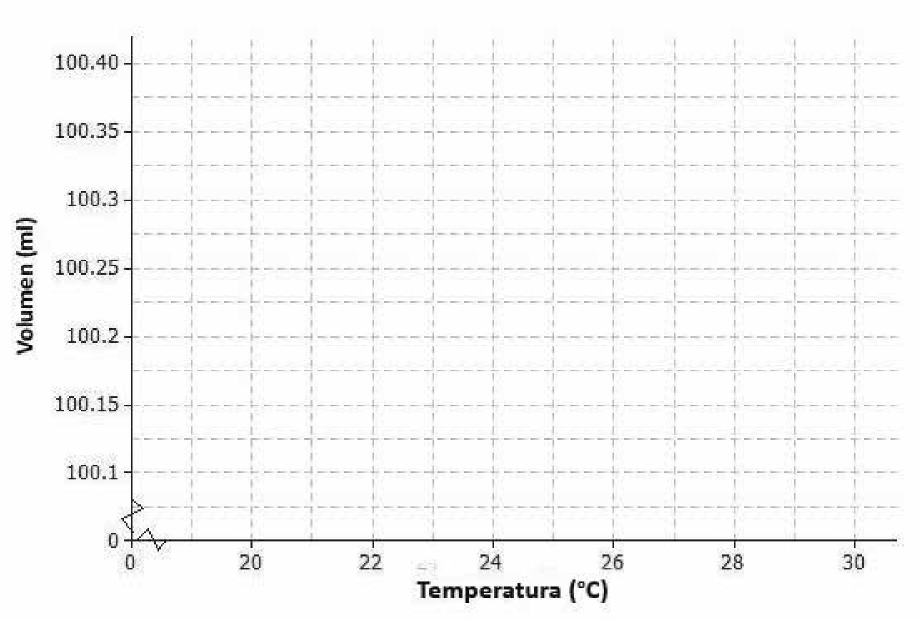
¿De qué manera se derivó esta gráfica de residuos del diagrama de dispersión original?

Ejercicios 1 a 3: Volumen y temperatura

El agua se expande cuando se calienta. Los investigadores midieron el volumen (en milímetros) de agua a distintas temperaturas. Los resultados se muestran a continuación.

Temperatura (°C)	Volumen (ml)
20	100.125
21	100.145
22	100.170
23	100.191
24	100.215
25	100.239
26	100.266
27	100.290
28	100.319
29	100.345
30	100.374

1. Utiliza la calculadora gráfica para construir el diagrama de dispersión de este conjunto de datos. Incluye en tu gráfica la recta de mínimos cuadrados. Dibuja el diagrama de dispersión e incluye la recta de mínimos cuadrados en los ejes a continuación.



2. Utiliza la calculadora para construir una gráfica de residuos para este conjunto de datos. Dibuja la gráfica de residuos en los ejes que se dan a continuación.

3. ¿Ves una curva clara en la gráfica de residuos? ¿Qué te dice esto sobre el conjunto de datos original?

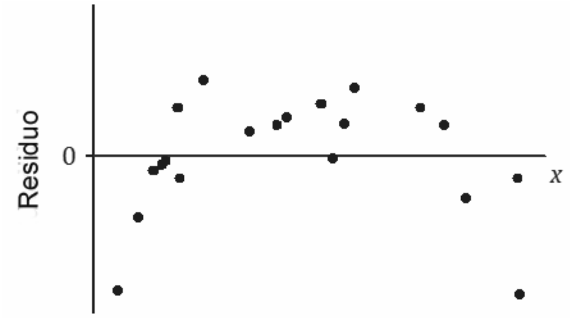
Resumen de la lección

- Luego de ajustar una recta, se puede construir la gráfica de residuos utilizando una calculadora gráfica.
- Una curva o patrón en la gráfica de residuos indica una relación no lineal en el conjunto de datos original.
- Una dispersión de puntos aleatoria en la gráfica de residuos indica una relación lineal en el conjunto de datos original.

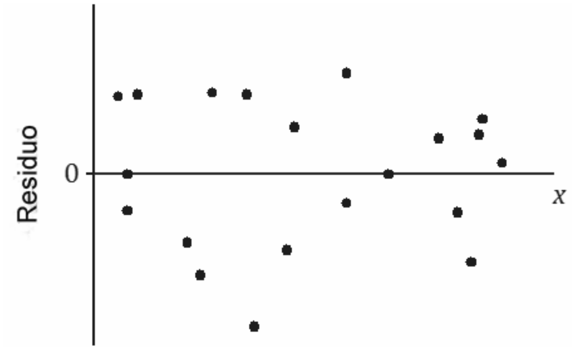
Grupo de problemas

1. Para cada una de las siguientes gráficas de residuos, ¿qué conclusión sacarías sobre la relación entre las variables del conjunto de datos original? Indica si los valores se representarían mejor con una relación lineal o no lineal.

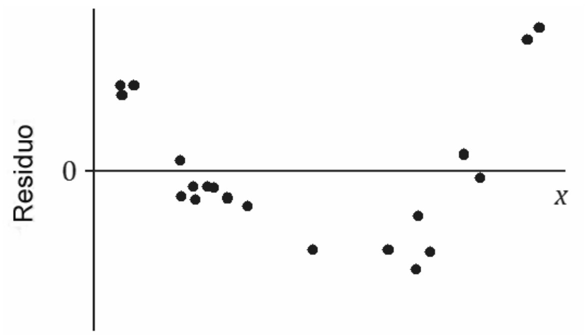
a.



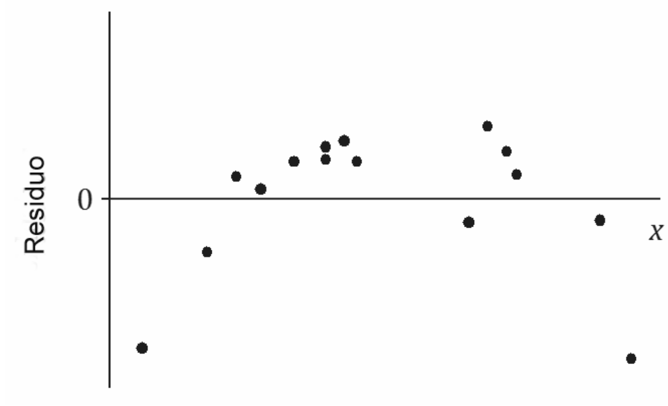
b.



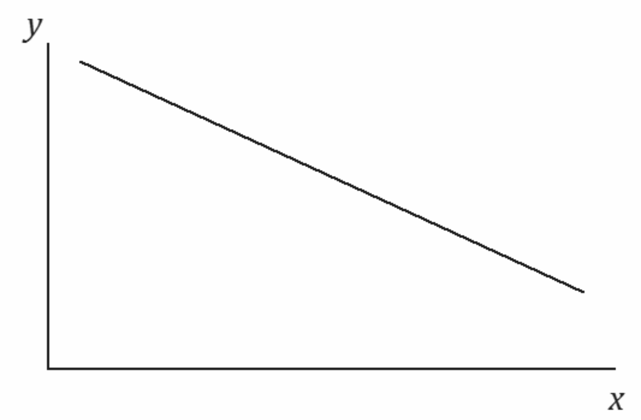
c.



2. Supón que, luego de ajustar una línea, un conjunto de datos produce la gráfica de residuos que se muestra a continuación.



A continuación, se muestra un diagrama de dispersión del conjunto de datos original. Se muestra la recta de mínimos cuadrados, pero los puntos del diagrama de dispersión se han borrado. Calcula aproximadamente las ubicaciones de los puntos originales y crea una aproximación del diagrama de dispersión a continuación.



EUREKA MATH

Lección 19: Interpretar la correlación

Trabajo en clase

Ejemplo 1: Relaciones lineales positivas y negativas

Las relaciones lineales pueden describirse como positivas o negativas. A continuación, se presentan dos diagramas de dispersión que muestran una relación lineal entre dos variables numéricas x y y.

Diagrama de dispersión 1

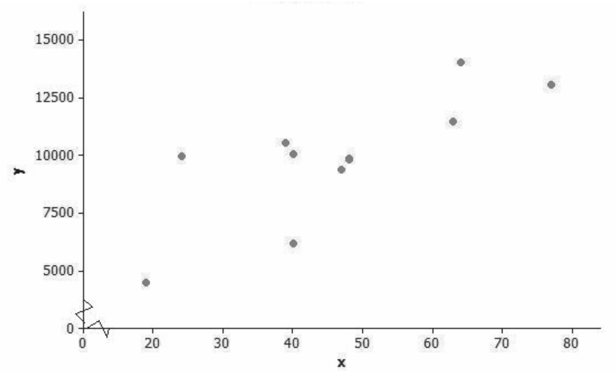
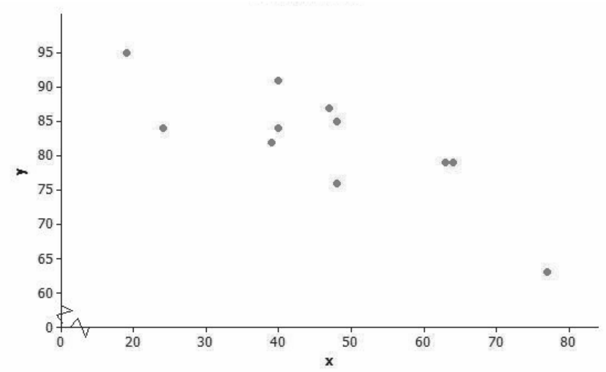


Diagrama de dispersión 2



Ejercicios 1 a 4

1. La relación que se muestra en el Diagrama de dispersión 1 es una relación lineal positiva. ¿El valor de la variable y tiende a disminuir o aumentar a medida que el valor de x aumenta? Si tuvieras que describir esta relación con una recta, ¿la recta tendría una pendiente positiva o negativa?

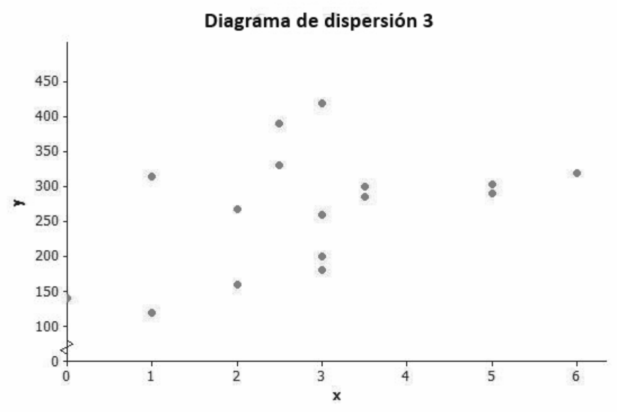
2. La relación que se muestra en el Diagrama de dispersión 2 es una relación lineal negativa. A medida que el valor de una de las variables aumenta, ¿qué sucede con el valor de la otra variable? Si tuvieras que describir esta relación con una recta, ¿la recta tendría una pendiente positiva o negativa?

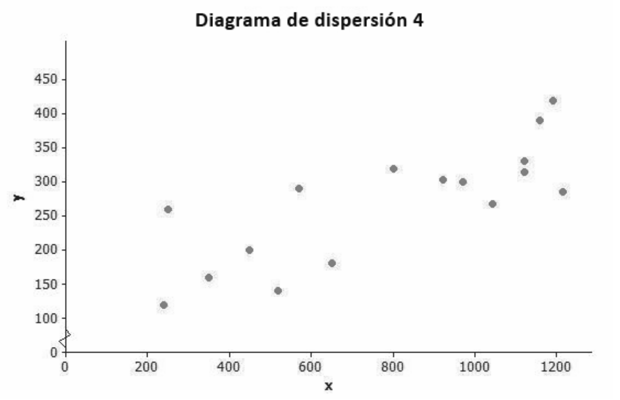
3. ¿Qué significa que hay una relación lineal positiva entre dos variables?

4. ¿Qué significa que hay una relación lineal negativa entre dos variables?

Ejemplo 2: Algunas relaciones lineales son más fuertes que otras

A continuación, se presentan dos diagramas de dispersión que muestran una relación lineal entre dos variables numéricas x y y.





Ejercicios 5 a 9

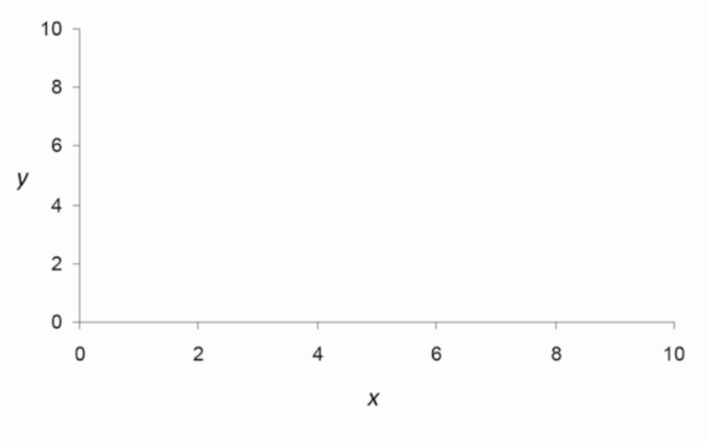
5. ¿La relación lineal del Diagrama de dispersión 3 es positiva o negativa?

6. ¿La relación lineal del Diagrama de dispersión 4 es positiva o negativa?

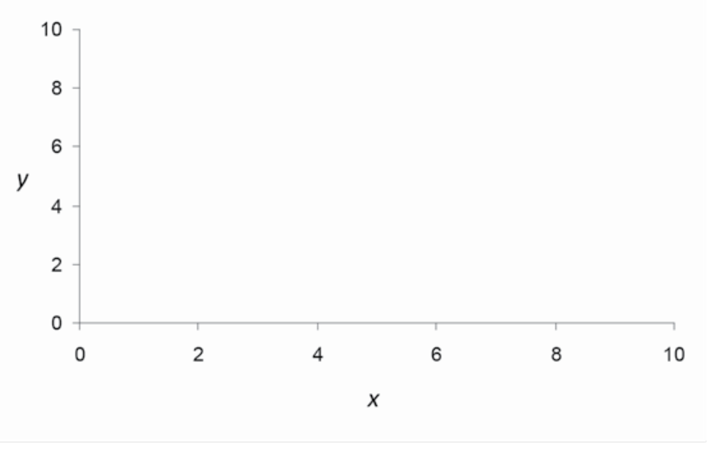
También se suele describir la fuerza de una relación lineal. Diríamos que la relación lineal del Diagrama de dispersión 3 es más débil que la relación lineal del Diagrama de dispersión 4.

7. ¿Por qué crees que se considera que la relación lineal del Diagrama de dispersión 3 es más débil que la relación lineal del Diagrama de dispersión 4?

8. ¿Cómo crees que sería un diagrama de dispersión con la relación lineal más fuerte posible si la relación es positiva? Dibuja un diagrama de dispersión con cinco puntos para ilustrarlo.



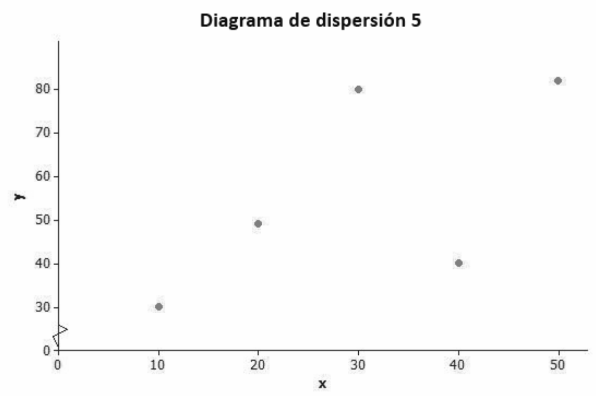
9. ¿En qué se diferenciaría un diagrama de dispersión que muestra la relación lineal negativa más fuerte posible del diagrama de dispersión que dibujaste en la pregunta anterior?

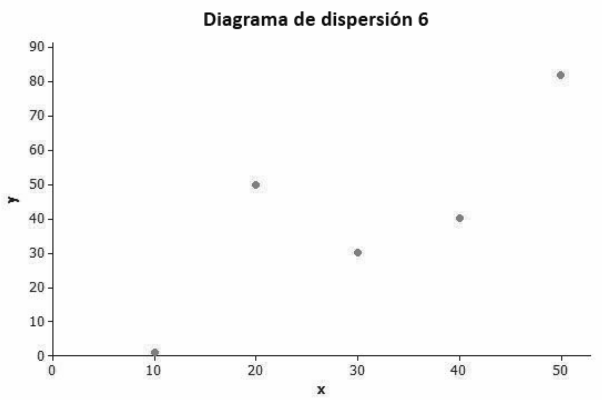


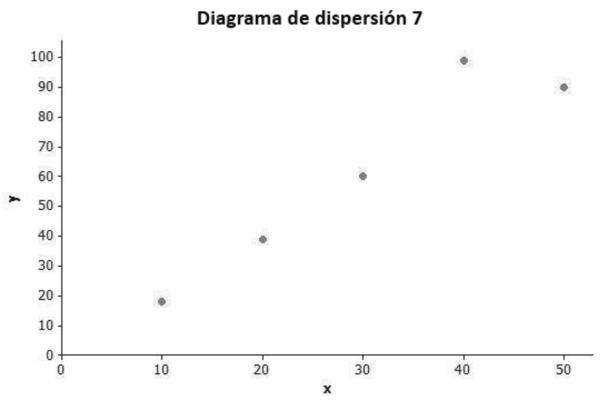
Ejercicios 10 a 12: Fuerza de las relaciones lineales

10. Considera los siguientes tres diagramas de dispersión. Ordénalos desde el que muestra la relación lineal más fuerte hasta el que muestra la relación lineal más débil.

Más fuerte	→	Más débil



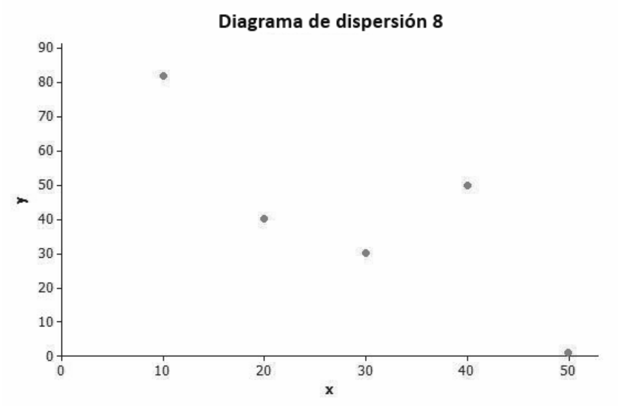


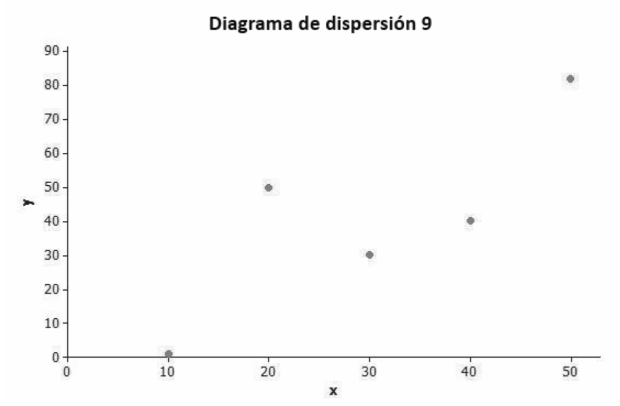


11. Explica tu razonamiento para elegir el orden del Ejercicio 10.

EUREKA MATH

12. ¿Cuál de los siguientes dos diagramas de dispersión muestra la relación lineal más fuerte? (¡Piensa tu respuesta con atención!).

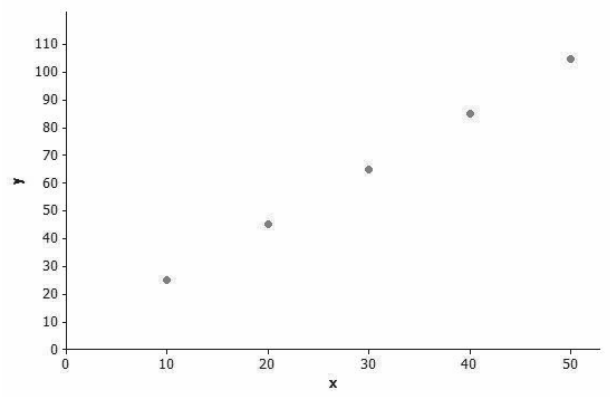


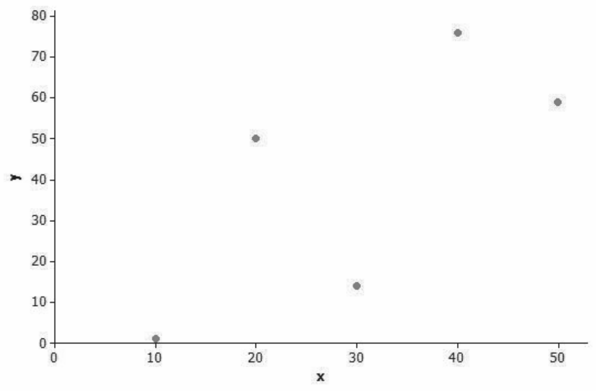
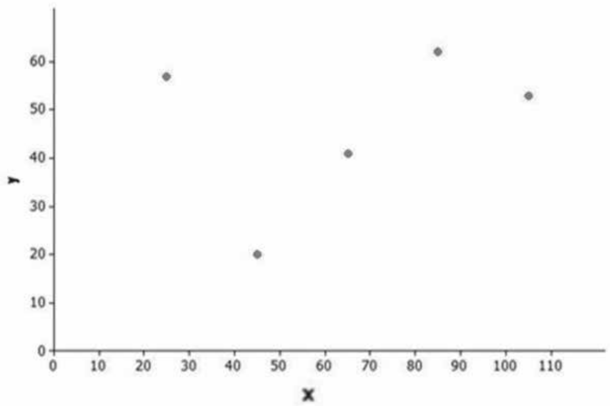
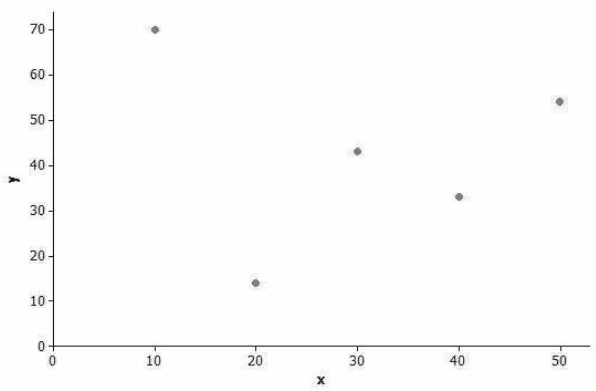


Ejemplo 3: El coeficiente de correlación

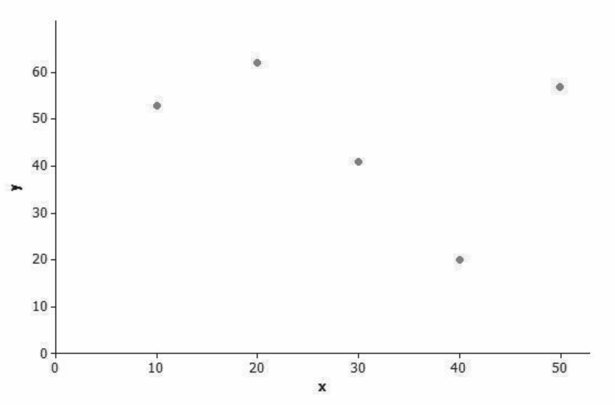
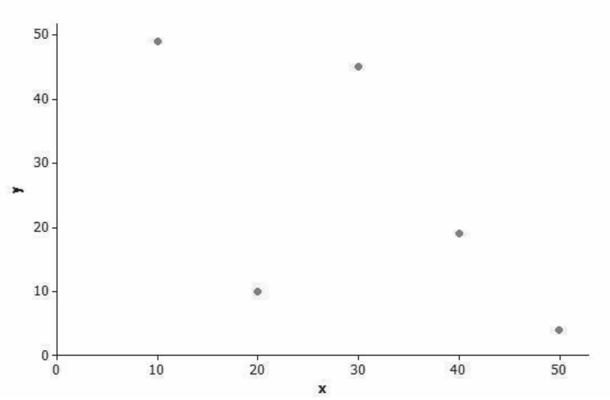
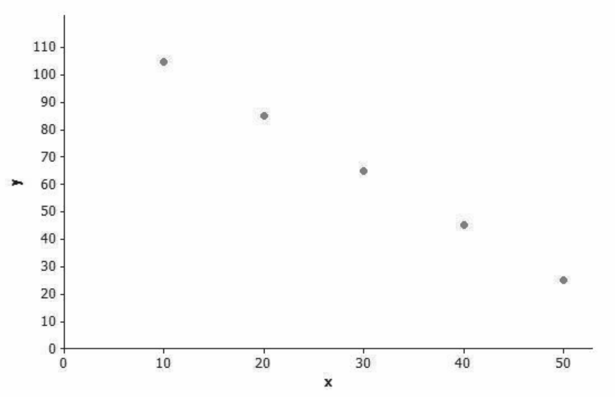
El *coeficiente de correlación* es un número entre -1 y $+1$ (incluidos -1 y $+1$) que mide la fuerza y la dirección de una relación lineal. El coeficiente de correlación se denota con la letra r.

A continuación, se muestran varios diagramas de dispersión. También se muestra el valor del coeficiente de correlación de los datos que se muestran en cada diagrama.

	$r = 1.00$

	$r = 0.71$
	$r = 0.32$
	$r = -0.10$

EUREKA MATH

	$r = -0.32$
	$r = -0.63$
	$r = -1.00$

Ejercicios 13 a 15

13. ¿Cuándo es positivo el valor del coeficiente de correlación?

14. ¿Cuándo es negativo el valor del coeficiente de correlación?

15. ¿La relación lineal es más fuerte cuando el coeficiente de correlación está más cerca de 0 o de 1 (o -1)?

Al observar los diagramas de dispersión del Ejemplo 3, probablemente hayas descubierto las siguientes propiedades del coeficiente de correlación:

Propiedad 1: El signo de r (positivo o negativo) corresponde a la dirección de la relación lineal.

Propiedad 2: Un valor de $r = +1$ indica una relación lineal positiva perfecta, ya que todos los puntos del diagrama de dispersión se encuentran exactamente sobre una línea recta.

Propiedad 3: Un valor de $r = -1$ indica una relación lineal negativa perfecta, ya que todos los puntos del diagrama de dispersión se encuentran exactamente sobre una línea recta.

Propiedad 4: Cuanto más cerca esté el valor de r de $+1$ o -1, más fuerte será la relación lineal.

Ejemplo 4: Calcular el valor del coeficiente de correlación

Hay una ecuación que puede utilizarse para calcular el valor del coeficiente de correlación dados los datos de dos variables numéricas. Utilizar esta fórmula requiere hacer muchos cálculos tediosos que se abordarán en los próximos grados. Afortunadamente, se puede utilizar la calculadora gráfica para hallar el valor del coeficiente de correlación una vez ingresados los datos.

Tu maestro te mostrará cómo ingresar los datos y cómo usar una calculadora gráfica para obtener el valor del coeficiente de correlación.

Estos son los datos de una lección anterior de la talla de zapatos en pulgadas y la altura en pulgadas de 10 hombres.

x (talla de zapatos) pulgadas	y (altura) pulgadas
12.6	74
11.8	65
12.2	71
11.6	67
12.2	69
11.4	68
12.8	70
12.2	69
12.6	72
11.8	71

Ejercicios 16 y 17

16. Ingresa los datos de la talla de zapatos y la altura en la calculadora. Halla el valor del coeficiente de correlación entre la talla de zapatos y la altura. Redondea a la décima más cercana.

La siguiente tabla muestra cómo puedes interpretar informalmente el valor de un coeficiente de correlación.

Si el valor del coeficiente de correlación está entre...	puedes decir que...
$r = 1.0$	Hay una relación lineal positiva perfecta.
$0.7 \leq r < 1.0$	Hay una relación lineal positiva fuerte.
$0.3 \leq r < 0.7$	Hay una relación lineal positiva moderada.
$0 < r < 0.3$	Hay una relación lineal positiva débil.
$r = 0$	No hay una relación lineal.
$-0.3 < r < 0$	Hay una relación lineal negativa débil.
$-0.7 < r \leq -0.3$	Hay una relación lineal negativa moderada.
$-1.0 < r \leq -0.7$	Hay una relación lineal negativa fuerte.
$r = -1.0$	Hay una relación lineal negativa perfecta.

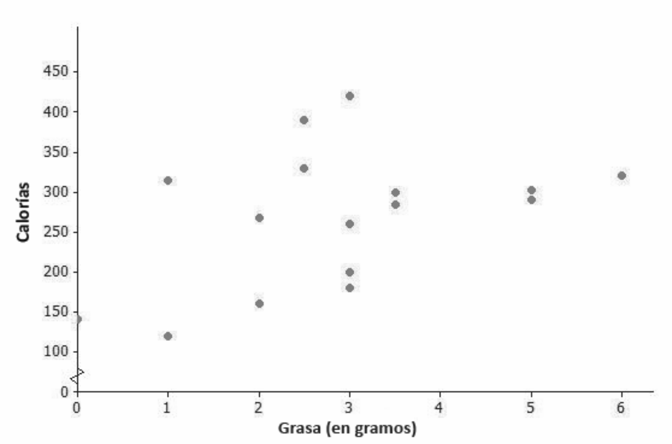
17. Interpreta el valor del coeficiente de correlación entre la talla de zapatos y la altura para los datos de arriba.

Ejercicios 18 a 24: Practicar calcular e interpretar los coeficientes de correlación

Consumer Reports publicó un estudio sobre productos de comida rápida. La tabla y el diagrama de dispersión a continuación muestran el contenido de grasa (en gramos) y el número de calorías por porción de 16 productos de comida rápida.

Grasa (g)	Calorías (cal)
2	268
5	303
3	260
3.5	300
1	315
2	160
3	200
6	320
3	420
5	290
3.5	285
2.5	390
0	140
2.5	330
1	120
3	180

Fuente de datos: *Consumer Reports*



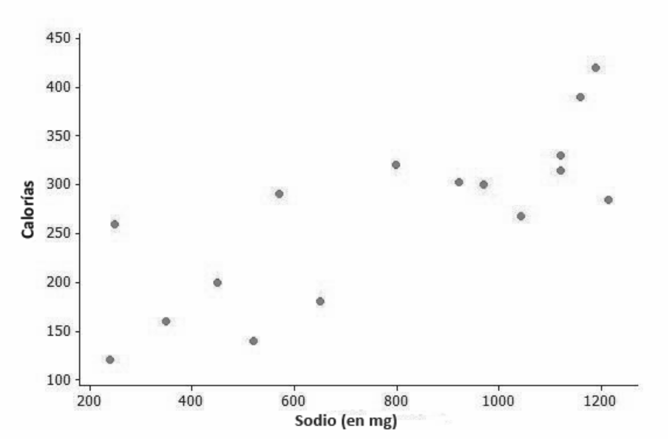
18. Teniendo en cuenta el diagrama de dispersión, ¿crees que el valor del coeficiente de correlación entre el contenido de grasa y las calorías por porción será positivo o negativo? Explica por qué hiciste esa elección.

19. Teniendo en cuenta el diagrama de dispersión, calcula aproximadamente el valor del coeficiente de correlación entre el contenido de grasa y las calorías.

20. Calcula el valor del coeficiente de correlación entre el contenido de grasa y las calorías por porción. Redondea a la centésima más cercana. Interpreta este valor.

El estudio de *Consumer Reports* también recopiló datos sobre el contenido de sodio (en mg) y el número de calorías por porción para los mismos 16 productos de comida rápida. Los datos se representan en la tabla y el diagrama de dispersión a continuación.

Sodio (mg)	Calorías (cal)
1,042	268
921	303
250	260
970	300
1,120	315
350	160
450	200
800	320
1,190	420
570	290
1,215	285
1,160	390
520	140
1,120	330
240	120
650	180



21. Teniendo en cuenta el diagrama de dispersión, ¿crees que el valor del coeficiente de correlación entre el contenido de sodio y las calorías por porción será positivo o negativo? Explica por qué hiciste esa elección.

22. Teniendo en cuenta el diagrama de dispersión, calcula aproximadamente el valor del coeficiente de correlación entre el contenido de sodio y las calorías por porción.

23. Calcula el valor del coeficiente de correlación entre el contenido de sodio y las calorías por porción. Redondea a la centésima más cercana. Interpreta este valor.

24. Para estos 16 productos de comida rápida, ¿la relación lineal entre el contenido de grasa y el número de calorías es más fuerte o más débil que la relación lineal entre el contenido de sodio y el número de calorías? ¿Te sorprende? Explica por qué si o por qué no.

Ejemplo 5: La correlación no significa que haya una relación de causa y efecto entre las variables

A veces es tentador llegar a la conclusión de que, si hay una relación lineal fuerte entre dos variables, una de las variables está causando el aumento o la disminución del valor de la otra variable. Pero deberías evitar cometer este error. Cuando hay una relación lineal fuerte, significa que las dos variables tienden a variar juntas de una manera predecible, y el motivo podría no ser una relación de causa y efecto.

Por ejemplo, el valor del coeficiente de correlación entre el contenido de sodio y el número de calorías para los productos de comida rápida del ejemplo anterior era $r = 0.79$, lo que indica una relación positiva fuerte. Esto significa que los productos que tienen un mayor contenido de sodio suelen tener más calorías. Pero la causa del número alto de calorías no es el alto contenido de sodio. De hecho, el sodio no tiene ninguna caloría. Lo que podría estar sucediendo es que los productos de comida rápida que tienen un contenido alto de sodio también sean los productos que tienen un contenido alto de azúcar o de grasas, y ese es el motivo por el cual esos productos tienen más calorías.

De manera similar, hay una correlación positiva fuerte entre la talla de zapatos y la capacidad de lectura de los niños. Pero no estaría bien pensar que tener los pies más grandes hace que los niños puedan leer mejor. Solo significa que las dos variables varían al mismo tiempo de una manera predecible. ¿Puedes pensar en una razón para explicar por qué los niños que tienen pies más grandes también suelen obtener una puntuación más alta en las pruebas de lectura?

Resumen de la lección

- Las relaciones lineales suelen describirse en función de la fuerza y la dirección.
- El coeficiente de correlación es una medida de la fuerza y la dirección de una relación lineal.
- Cuanto más cercano sea el valor del coeficiente de correlación a $+1$ o -1, más fuerte será la relación lineal.
- Que haya una correlación fuerte entre dos variables no significa que haya una relación de causa y efecto.

Grupo de problemas

1. ¿Cuál de los siguientes diagramas de dispersión muestra la relación lineal más fuerte? ¿Cuál muestra la relación lineal más débil?

Diagrama de dispersión 1

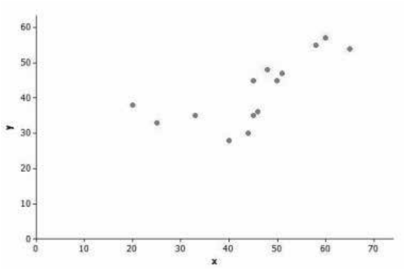


Diagrama de dispersión 2

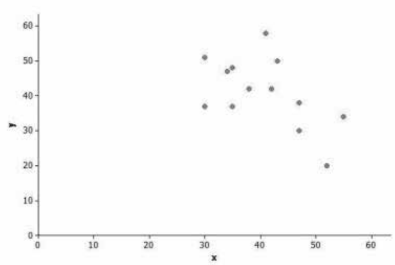
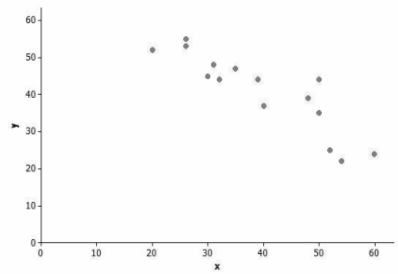
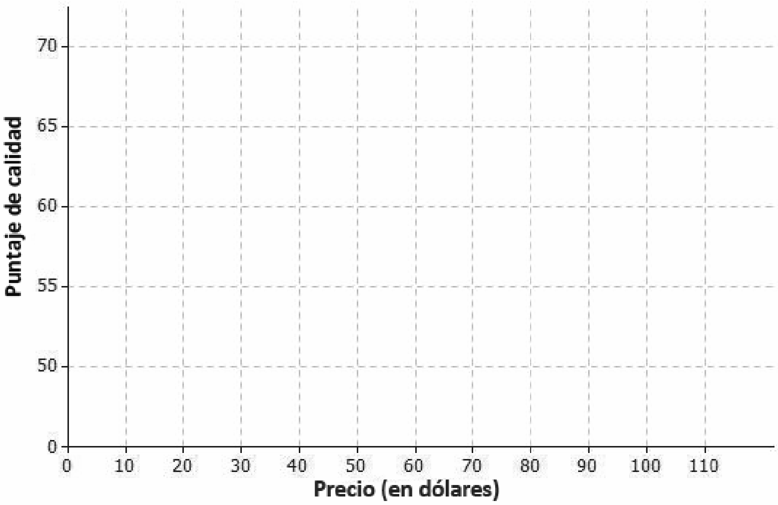


Diagrama de dispersión 3



2. *Consumer Reports* publicó los datos de los precios (en dólares) y el puntaje de calidad (en una escala de 0 a 100) de 10 marcas distintas de zapatos deportivos para hombres.

Precio ($)	Puntaje de calidad
65	71
45	70
45	62
80	59
110	58
110	57
30	56
80	52
110	51
70	51



a. Utiliza la cuadrícula anterior para construir un diagrama de dispersión de estos datos.

b. Calcula el valor del coeficiente de correlación entre el precio y el puntaje de calidad e interpreta ese valor. Redondea a la centésima más cercana.

c. ¿Te sorprende que el valor del coeficiente de correlación sea negativo? Explica por qué si o por qué no.

d. ¿Es razonable concluir que los zapatos de precio más alto son de mejor calidad? Explica tu respuesta.

e. La correlación entre el precio y el puntaje de calidad es negativa. ¿Es razonable concluir que aumentar el precio provoca una disminución en el puntaje de calidad? Explica tu respuesta.

3. La revista *The Princeton Review* publica información sobre escuelas y universidades. A continuación, se presentan los datos de seis universidades públicas con carreras de 4 años ubicadas en Nueva York. La tasa de graduación es el porcentaje de estudiantes que se gradúan en seis años o menos. La razón estudiante a profesor es el número de estudiantes por profesor a tiempo completo.

Universidad	Número de estudiantes a tiempo completo	Razón estudiante a profesor	Tasa de graduación
CUNY Bernard M. Baruch College	11,477	17	63
CUNY Brooklyn College	9,876	15.3	48
CUNY City College	10,047	13.1	40
SUNY, Albany	14,013	19.5	64
SUNY, Binghamton	13,031	20	77
SUNY College, Buffalo	9,398	14.1	47

a. Calcula el valor del coeficiente de correlación entre el número de estudiantes a tiempo completo y la tasa de graduación. Redondea a la centésima más cercana.

b. ¿La relación lineal entre la tasa de graduación y el número de estudiantes a tiempo completo es débil, moderada o fuerte? ¿En qué basaste tu decisión?

c. ¿El siguiente enunciado es verdadero o falso? Teniendo en cuenta el valor del coeficiente de correlación, es razonable concluir que el hecho de que haya un mayor número de estudiantes es la causa de que la tasa de graduación sea mayor.

d. Calcula el valor del coeficiente de correlación entre la razón estudiante a profesor y la tasa de graduación. Redondea a la centésima más cercana.

e. ¿Qué relación lineal es más fuerte: la tasa de graduación y el número de estudiantes a tiempo completo o la tasa de graduación y la razón estudiante a profesor? Justifica tu respuesta.

Lección 20: Analizar los datos recopilados en dos variables

Trabajo en clase

En las Lecciones 12 a 19, se incluyeron varios conjuntos de datos que se utilizaron para averiguar de qué manera pueden relacionarse dos variables numéricas. Recuerda los datos de la elevación sobre el nivel del mar y el número de días despejados por año de 14 ciudades de los Estados Unidos. ¿Podría usarse la elevación de una ciudad sobre el nivel del mar para predecir el número de días despejados por año de una ciudad? Luego de observar el diagrama de dispersión de los datos, un modelo lineal (o el modelo lineal de mínimos cuadrados obtenido con una calculadora o programa informático) brindó una descripción razonable de la relación entre estas dos variables. Para evaluar este modelo lineal, se consideró qué tan cerca estaban los puntos de datos de la gráfica correspondiente de la recta. La ecuación del modelo lineal se utilizó para responder la pregunta estadística sobre la elevación y el número de días despejados.

También se brindaron varios conjuntos de datos para ilustrar otros modelos posibles, específicamente modelos cuadráticos y exponenciales. Hallar un modelo que describe la relación entre dos variables nos permite responder preguntas estadísticas sobre cómo varían dos variables numéricas. Por ejemplo, en la Lección 13, se hizo la siguiente pregunta estadística sobre la latitud y la media de pollitos papamoscas de un nido: ¿cuál es la mejor latitud para criar pollitos papamoscas? Para responder esta pregunta estadística, se usó un modelo cuadrático de la latitud y la media de pollitos papamoscas de un nido.

En las Lecciones 12 a 19, trabajaste con varios conjuntos de datos y modelos para responder preguntas estadísticas. Selecciona uno de los conjuntos de datos que se presentan en las Lecciones 12 a 19 y crea un póster que resuma cómo se responde una pregunta estadística que contiene dos variables numéricas. Tu póster debería incluir lo siguiente: un resumen breve de los datos, la pregunta estadística sobre la relación entre las dos variables numéricas, un diagrama de dispersión de los datos y un resumen breve que indique qué tan bien se ajustan los datos a un modelo específico.

Una vez que identifiques uno de los problemas para resumir en un póster, considera las siguientes preguntas para planificar el póster:

1. ¿Qué dos variables contiene este problema?
2. ¿Cuál era la pregunta estadística? Recuerda que una pregunta estadística involucra datos. La pregunta también anticipa que los datos variarán.
3. ¿Qué modelo se utilizó para describir la relación entre las dos variables?
4. Teniendo en cuenta el diagrama de dispersión de los datos, ¿el modelo era adecuado?
5. ¿Se utilizó la gráfica de residuos para evaluar el modelo? ¿Qué indicaba la gráfica de residuos sobre el modelo?
6. ¿Cómo utilizarías el modelo para predecir los valores que no se incluyen en el conjunto de datos?
7. ¿El modelo responde la pregunta estadística?

Puedes encontrar ejemplos de pósters de dos variables numéricas en el sitio web de la Asociación de Estadísticas de los Estados Unidos (www.amstat.org/education/posterprojects/index.cfm).

Esta página queda en blanco intencionalmente.