Unidad 6, Lección 8

Usemos el coeficiente de correlación

Examinemos más de cerca los coeficientes de correlación.

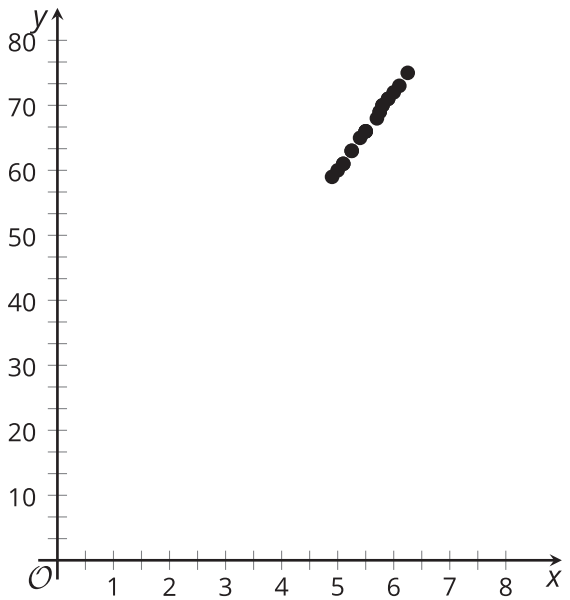
8.1 Pongamos los números en contexto

Asocia la pareja de variables de cada fila con el diagrama de dispersión al que crees que se ajustan mejor. Prepárate para explicar tu razonamiento.

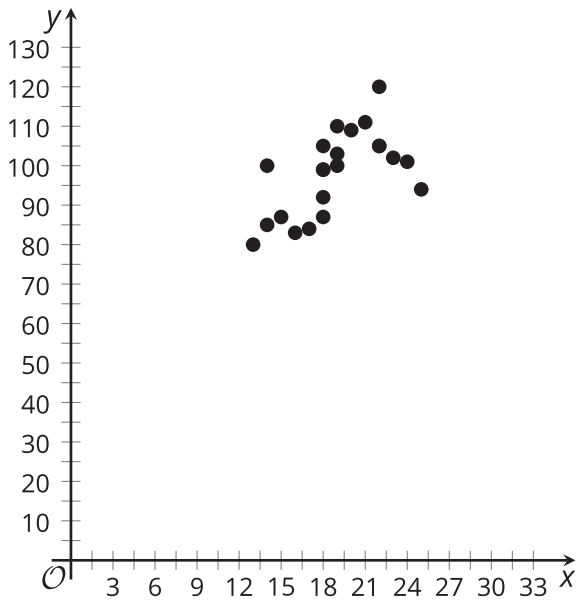
	variable x	variable y
1.	temperatura mínima diaria en grados Celsius de Denver, CO	cajas de cereal que hay en un supermercado en Miami, FL, en un día dado
2.	número de tiros libres lanzados en un partido	puntaje del equipo de baloncesto en un partido
3.	estatura de un estudiante en pies	estatura de un estudiante en pulgadas
4.	número de minutos que una persona espera a ser atendida en el hospital	calificación de satisfacción con el hospital dada por una persona

CC BY NC 2024
Illustrative Mathematics®

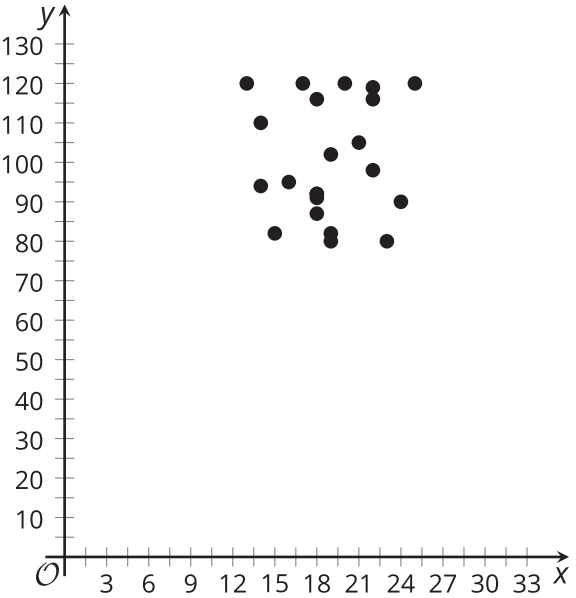
A


y
80
70
60
50
40
30
20
10
O
1 2 3 4 5 6 7 8
x

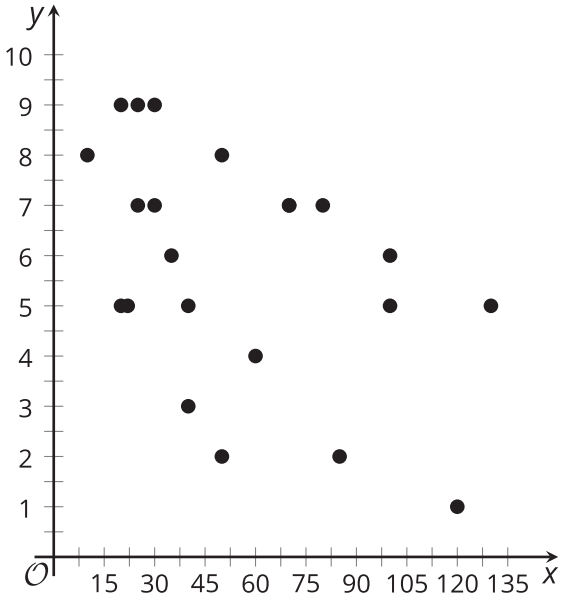
B


y
130
120
110
100
90
80
70
60
50
40
30
20
10
O
3 6 9 12 15 18 21 24 27 30 33
x

C


y
130
120
110
100
90
80
70
60
50
40
30
20
10
O
3 6 9 12 15 18 21 24 27 30 33
x

D


y
10
9
8
7
6
5
4
3
2
1
O
15 30 45 60 75 90 105 120 135
x

8.2 Nunca sabes qué tan lejos llegarás

Priya anota la distancia que recorre su automóvil en varios trayectos y el tiempo que tarda en llegar a su destino.

distancia (mi) (x)	tiempo de recorrido (min) (y)
2	4
5	7
10	11
10	15
12	16
15	22
20	23
25	25
26	28
30	36
32	35
40	37
50	51
65	70
78	72

1. La distancia es uno de los factores que influyen en el tiempo de los recorridos de Priya en su automóvil. ¿Qué otros factores influyen?

2. ¿Cuáles de estos factores (incluyendo la distancia) puede tener la influencia más consistente en todos los recorridos del automóvil? Explica tu razonamiento.

3. Usa tecnología para crear un diagrama de dispersión de los datos y agrega la recta de mejor ajuste al gráfico.
4. En esta situación, ¿qué significan la pendiente de la recta de mejor ajuste y su intersección con el eje y?

5. Usa tecnología para encontrar el coeficiente de correlación de estos datos. Con base en este valor, ¿cómo describirías la intensidad de la relación lineal entre los datos?

6. ¿Cuánto tiempo crees que tardará Priya en hacer un recorrido de 90 millas si la relación lineal continúa? Si ella conduce 90 millas, ¿crees que la predicción que hiciste va a estar cerca del valor real? Explica tu razonamiento.

8.3 Zoológico de correlaciones

En cada situación, describe la relación entre las variables de acuerdo al coeficiente de correlación. Asegúrate de mencionar si hay una **relación fuerte** o una **relación débil**, y si es una **relación positiva** o una **relación negativa**.

1. Número de pasos dados cada día y número de kilómetros caminados cada día. $r = 0.92$

2. Temperatura de una banda elástica y distancia que la banda elástica se puede estirar. $r = 0.84$

3. Peso de un automóvil y distancia que el automóvil recorre si usa toda la gasolina del tanque. $r = -0.86$

4. Consumo promedio de grasa de un ciudadano de un país y tasa promedio de cáncer de un país. $r = 0.73$

5. Puntaje en un examen de Ciencias y número de palabras escritas en la pregunta tipo ensayo. $r = 0.28$

6. Tiempo promedio dedicado a escuchar música en un día y tiempo promedio dedicado a ver televisión en un día. $r = -0.17$

¿Estás listo para más?

Un biólogo trata de determinar si un grupo de delfines es de una nueva especie de delfines o si es un nuevo grupo de individuos de una especie conocida de delfines. El biólogo mide el ancho, en milímetros, de la parte más grande del cráneo (ancho cigomático) y la longitud, en milímetros, del hocico (longitud del rostro) de 10 delfines del mismo grupo de individuos.

Los datos parecen seguir un patrón lineal. La ecuación de la recta de mejor ajuste es $y = 0.201x + 110.806$, y el valor de r es 0.201.

x, longitud del rostro (mm)	y, ancho cigomático (mm)
288	147
247	147
268	171
278	177
258	168
272	184
272	161
258	159
273	168
277	166

1. Después de revisar los datos, el biólogo se da cuenta de que el primer ancho cigomático de 147 mm de la lista es un error. En realidad debe ser 180 mm. Usa tecnología para encontrar la ecuación de la recta de mejor ajuste y el coeficiente de correlación de los datos corregidos. ¿Cuál es la ecuación de la recta de mejor ajuste y cuál es el coeficiente de correlación?

2. Compara la nueva ecuación de la recta de mejor ajuste con la original. ¿Cómo el cambio de ese dato influyó en la pendiente de la recta de mejor ajuste, en la intersección de la recta de mejor ajuste con el eje y y en el coeficiente de correlación?

3. ¿Por qué crees que la asociación débil y positiva se volvió una asociación moderadamente fuerte? Explica tu razonamiento.

4. Usa tecnología para cambiar el valor de y de la primera y de la segunda entrada de la tabla.

   a. ¿Cómo cambia el coeficiente de correlación cuando cambiamos cada valor de y de los puntos?

   b. ¿Puedes cambiar dos valores para obtener un coeficiente de correlación que esté más cerca de 1? Usa datos para justificar tu respuesta.

   c. Si no cambias el par $(288, 180)$, ¿puedes cambiar otro valor para hacer que la relación cambie de ser positiva a ser negativa? Usa datos para justificar o refutar tu respuesta.

Resumen de la lección 8

El valor del coeficiente de correlación se puede usar para determinar la intensidad de la relación entre dos variables que están representadas en los datos.

En general, si las variables aumentan juntas, podemos decir que tienen una **relación positiva**. Si un aumento en los datos de una variable tiende a estar acompañado de una disminución en los datos de la otra variable, las variables tienen una **relación negativa**. Si los datos están muy cerca de la recta de mejor ajuste, decimos que hay una **relación fuerte**. Si los datos están esparcidos y no están tan cerca de la recta de mejor ajuste, decimos que hay una **relación débil**.

Un coeficiente de correlación que tiene un valor cercano a 1 sugiere que hay una relación fuerte y positiva entre las variables. Esto significa que la mayoría de los datos están agrupados muy cerca de una recta, y que cuando el valor de una de las variables aumenta, el valor de la otra también aumenta. El número de escuelas de una comunidad y la población de la comunidad son un ejemplo de variables que tienen una correlación fuerte y positiva. Usualmente, las poblaciones grandes tienen una gran cantidad de escuelas y las comunidades pequeñas tienden a tener menos escuelas, así que la correlación es positiva. Estas variables están estrechamente conectadas, así que la correlación es fuerte.

De forma similar, un coeficiente de correlación que está cerca de -1 sugiere una relación fuerte y negativa entre las variables. Nuevamente, la mayoría de los datos tienden a estar muy cerca de una recta, pero ahora, cuando un valor aumenta, el otro valor disminuye. Las variables "tiempo que transcurre desde que sales de casa" y "distancia que te falta para llegar a la escuela" tienen una correlación fuerte y negativa. Cuando el tiempo de recorrido aumenta, la distancia a la escuela disminuye, así que la correlación es negativa. Nuevamente, las variables están estrechamente conectadas, así que la correlación es fuerte.

Cuando la correlación es débil, concluimos que puede haber otras razones por las cuales los datos cambian aparte de la conexión entre las dos variables. Por ejemplo, las variables "número de mascotas" y "número de hermanos y hermanas" tienen una correlación débil. Puede que haya cierta relación, pero hay muchos otros factores (distintos del número de hermanos y hermanas) que explican la variabilidad del número de mascotas.

Para determinar si el valor de la correlación es fuerte o débil, se debe considerar el contexto de la situación. En física, donde las cosas se miden con instrumentos precisos, es posible que un coeficiente de correlación de 0.8 no se considere fuerte. En ciencias sociales, donde se recolectan datos usando encuestas, puede que un coeficiente de correlación de 0.8 sea muy fuerte.