Unidad 6, Lección 7

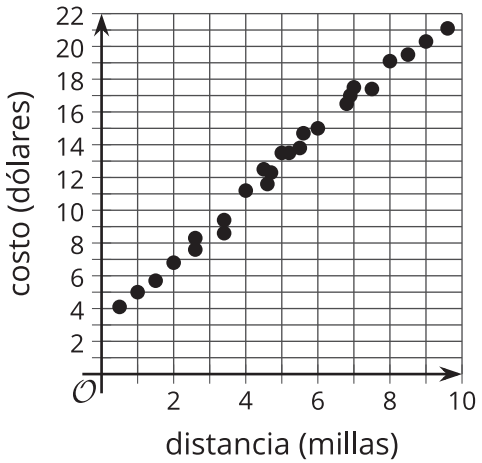
El coeficiente de correlación

Veamos qué tan bueno es un modelo lineal para un conjunto de datos.

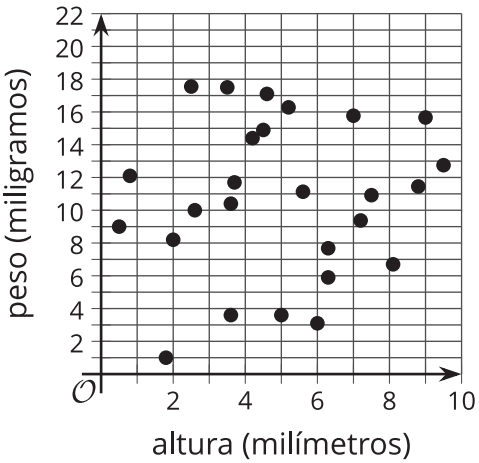
7.1 Cuáles tres van juntos: Modelos lineales

¿Cuáles tres van juntos? ¿Por qué van juntos?

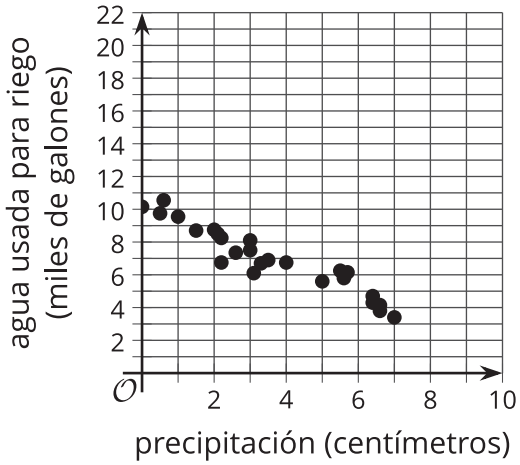
A



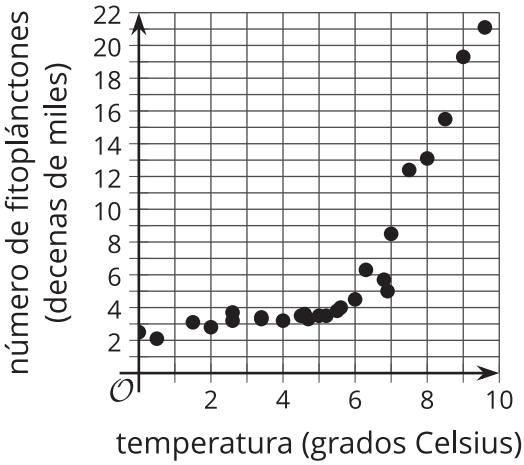
B



C



D



7.2 Clasificación de tarjetas: Diagramas de dispersión

Tu profesor te dará varias tarjetas que muestran diagramas de dispersión.

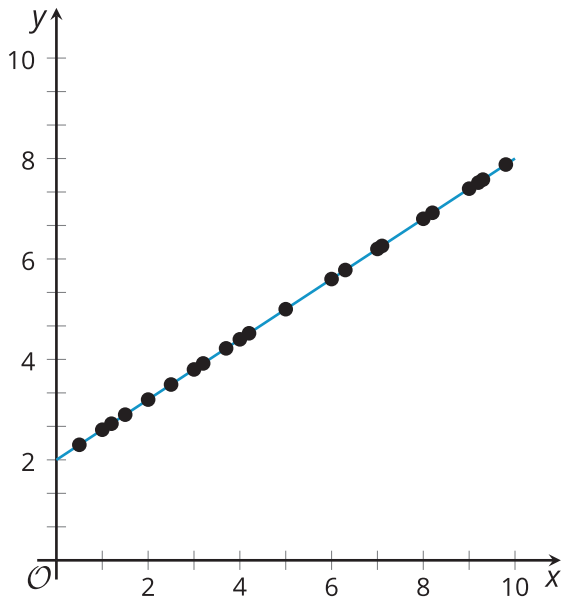
1. Clasifica las tarjetas en categorías de tu elección. Prepárate para explicar tus categorías. Haz una pausa aquí para tener una discusión con toda la clase.
2. Clasifica las tarjetas de una forma diferente, en nuevas categorías. Prepárate para explicar tus nuevas categorías.

7.3 Emparejemos coeficientes de correlación con diagramas de dispersión

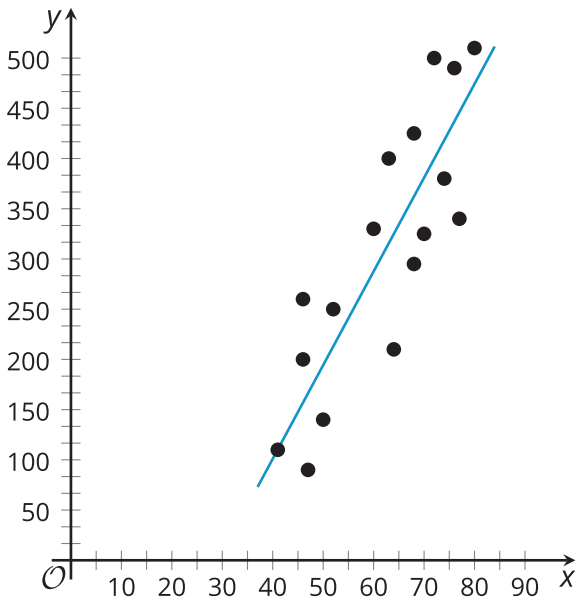
1. Por turnos, con tu compañero, empareja un diagrama de dispersión con un **coeficiente de correlación**.
2. Para cada pareja que encuentres, explícale a tu compañero cómo sabes que ese diagrama y ese coeficiente van juntos.
3. Escucha con atención la explicación de tu compañero sobre cada pareja que encuentra. Si están en desacuerdo, discutan sus ideas y trabajen para llegar a un acuerdo.

1. $r = -1$
2. $r = -0.95$
3. $r = -0.74$
4. $r = -0.06$
5. $r = 0.48$
6. $r = 0.65$
7. $r = 0.9$
8. $r = 1$

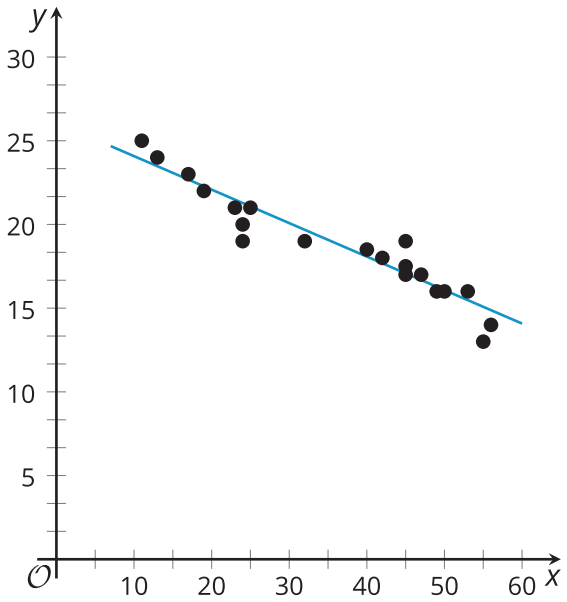
A


y
10
8
6
4
2
O
2
4
6
8
10
x

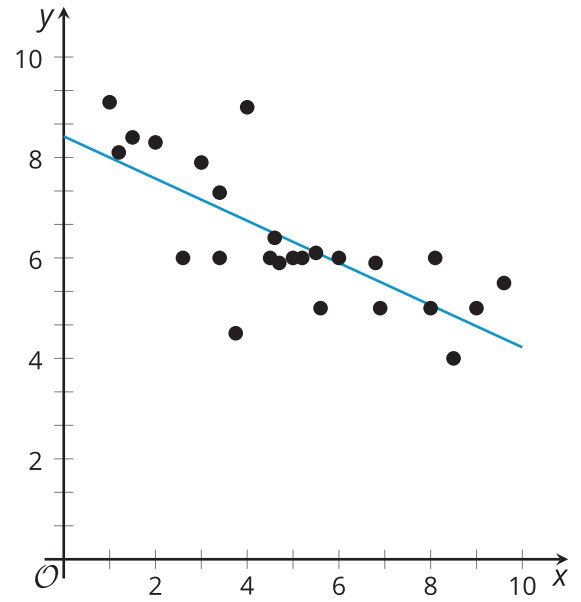
B


y
500
450
400
350
300
250
200
150
100
50
O
10
20
30
40
50
60
70
80
90
x

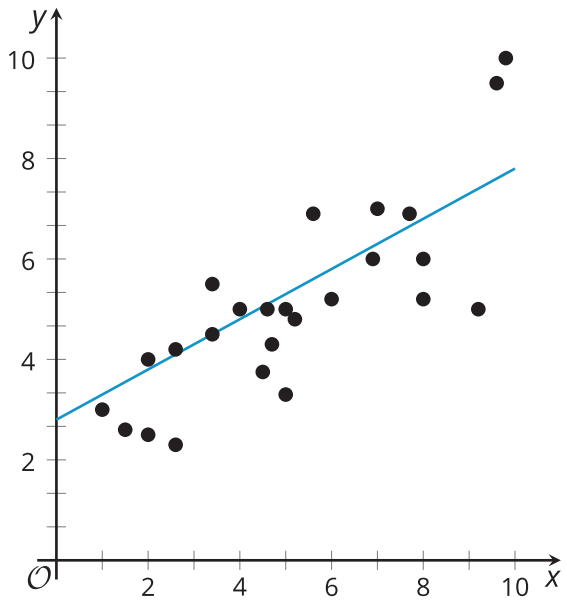
C


y
30
25
20
15
10
5
O
10
20
30
40
50
60
x

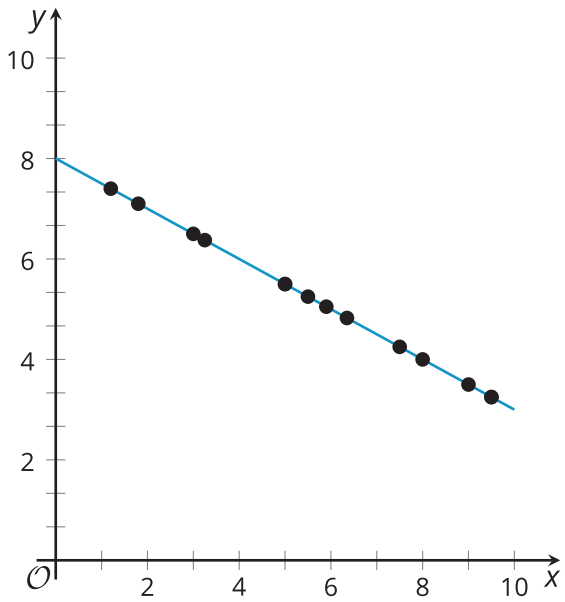
D


y
10
8
6
4
2
O
2
4
6
8
10
x

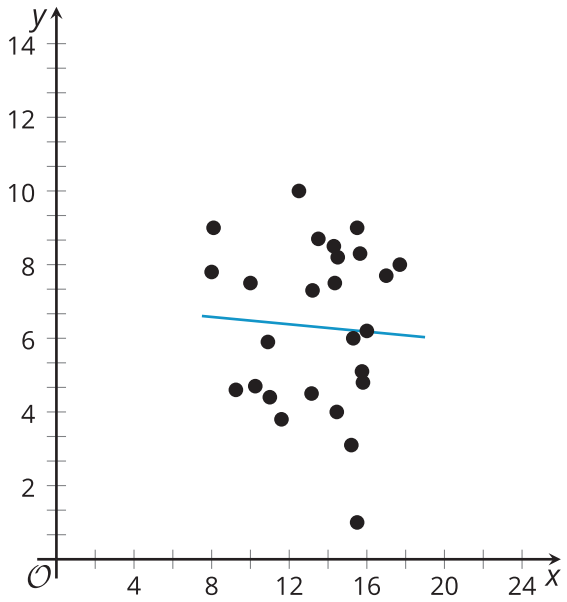
E


y
10
8
6
4
2
O
2
4
6
8
10
x

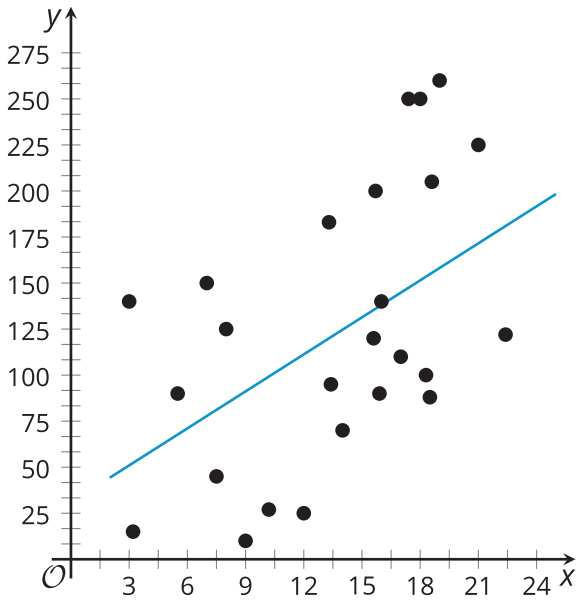
F


y
10
8
6
4
2
O
2
4
6
8
10
x

G


y
14
12
10
8
6
4
2
O
4
8
12
16
20
24
x

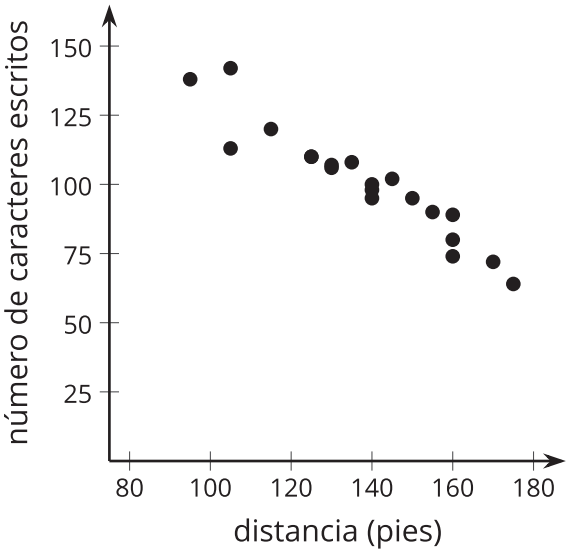
H


y
275
250
225
200
175
150
125
100
75
50
25
O
3
6
9
12
15
18
21
24
x

¿Estás listo para más?

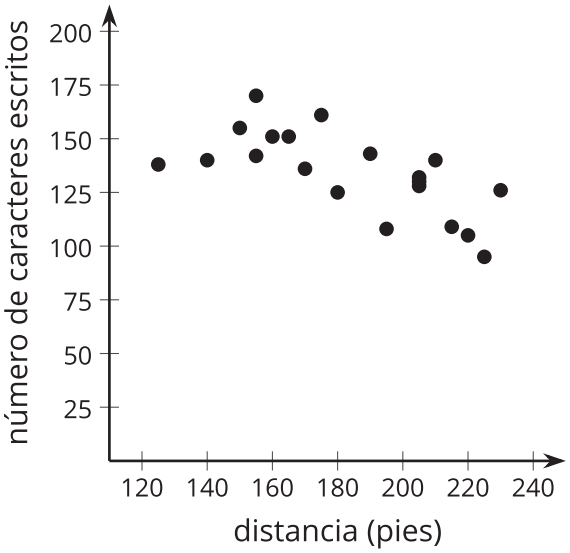
Jada quiere saber si la velocidad a la que caminan las personas está correlacionada con la velocidad a la que escriben mensajes de texto. Para investigarlo, ella midió la distancia, en pies, que 5 de sus amigos caminaron en 30 segundos y el número de caracteres que escribieron durante ese mismo tiempo. Cada uno de los 5 amigos caminó 4 veces, para un total de 20 recorridos. Estos son los resultados de los primeros 20 recorridos.

distancia (pies)	número de caracteres escritos	distancia (pies)	número de caracteres escritos
105	142	95	138
125	110	125	110
115	120	160	80
140	98	175	64
145	102	130	106
160	89	140	95
170	72	150	95
140	100	155	90
130	107	160	74
105	113	135	108



Durante los siguientes días, los mismos 5 amigos estuvieron practicando para ver si podían caminar más rápido y escribir más caracteres. Ellos no recolectaron ningún dato mientras practicaban. Después de practicar, cada uno de los 5 amigos caminó otras 4 veces. Estos son los resultados de los últimos 20 recorridos.

distancia (pies)	número de caracteres escritos	distancia (pies)	número de caracteres escritos
140	140	165	151
150	155	170	136
160	151	190	143
155	170	205	132
180	125	205	128
205	130	210	140
225	95	215	109
175	161	220	105
195	108	230	126
155	142	225	138



1. ¿Qué observas acerca de los dos diagramas de dispersión?

2. Jada observó que en los últimos 20 recorridos, sus amigos caminaron una mayor distancia y escribieron más rápido que en los primeros 20 recorridos. Como ellos hicieron ambas cosas más rápido, ella predice que el coeficiente de correlación de la recta de mejor ajuste de los últimos 20 recorridos va a estar más cerca de -1 que el coeficiente de correlación de la recta de mejor ajuste de los primeros 20 recorridos. ¿Estás de acuerdo con Jada? Explica tu razonamiento.

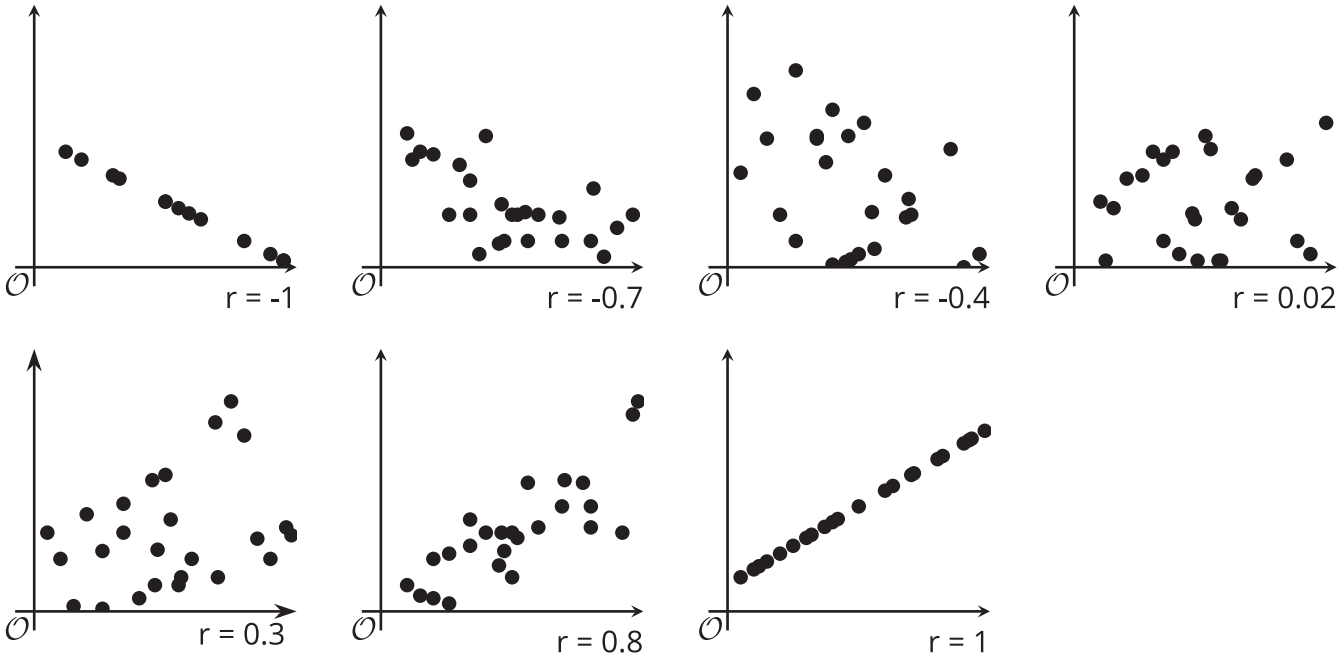
3. Usa tecnología para encontrar la ecuación de la recta de mejor ajuste y el coeficiente de correlación de ambos conjuntos de datos. ¿Tu respuesta a la pregunta anterior fue correcta?

4. ¿Por qué crees que los coeficientes de correlación de los dos conjuntos de datos son tan diferentes? Explica tu razonamiento.

Resumen de la lección 7

Los residuos pueden ayudar a escoger, entre todas las rectas, la recta que mejor se ajusta a los datos. Sin embargo, necesitamos una forma de determinar la intensidad de una relación lineal. Los diagramas de dispersión de datos que están cerca de la recta de mejor ajuste se modelan mejor con esa recta que los diagramas de dispersión de datos que están lejos de la recta de mejor ajuste.

El **coeficiente de correlación** es un número que se usa para describir la intensidad y la dirección de una relación lineal. Usualmente, el coeficiente de correlación se representa con la letra r y toma valores desde -1 hasta 1. El signo del coeficiente de correlación es el mismo signo de la pendiente de la recta de mejor ajuste. Cuanto más cerca de 0 está el coeficiente de correlación, más débil es la relación lineal. Cuanto más cerca de 1 o de -1 está el coeficiente de correlación, el modelo lineal se ajusta mejor a los datos.



Aunque es posible tratar de ajustar un modelo lineal a cualquier conjunto de datos, siempre debemos observar el diagrama de dispersión para ver si hay una posible tendencia lineal. El coeficiente de correlación y los residuos también pueden ayudar a decidir si tiene sentido usar el modelo lineal para estimar la situación. En algunos casos, otro tipo de función puede ajustarse mejor a los datos, o puede que las dos variables que estamos estudiando no estén correlacionadas y debamos buscar conexiones usando otras variables.

CC BY NC 2024
Illustrative Mathematics®