Regresión lineal simple y correlación

- **Regresión y Correlación**

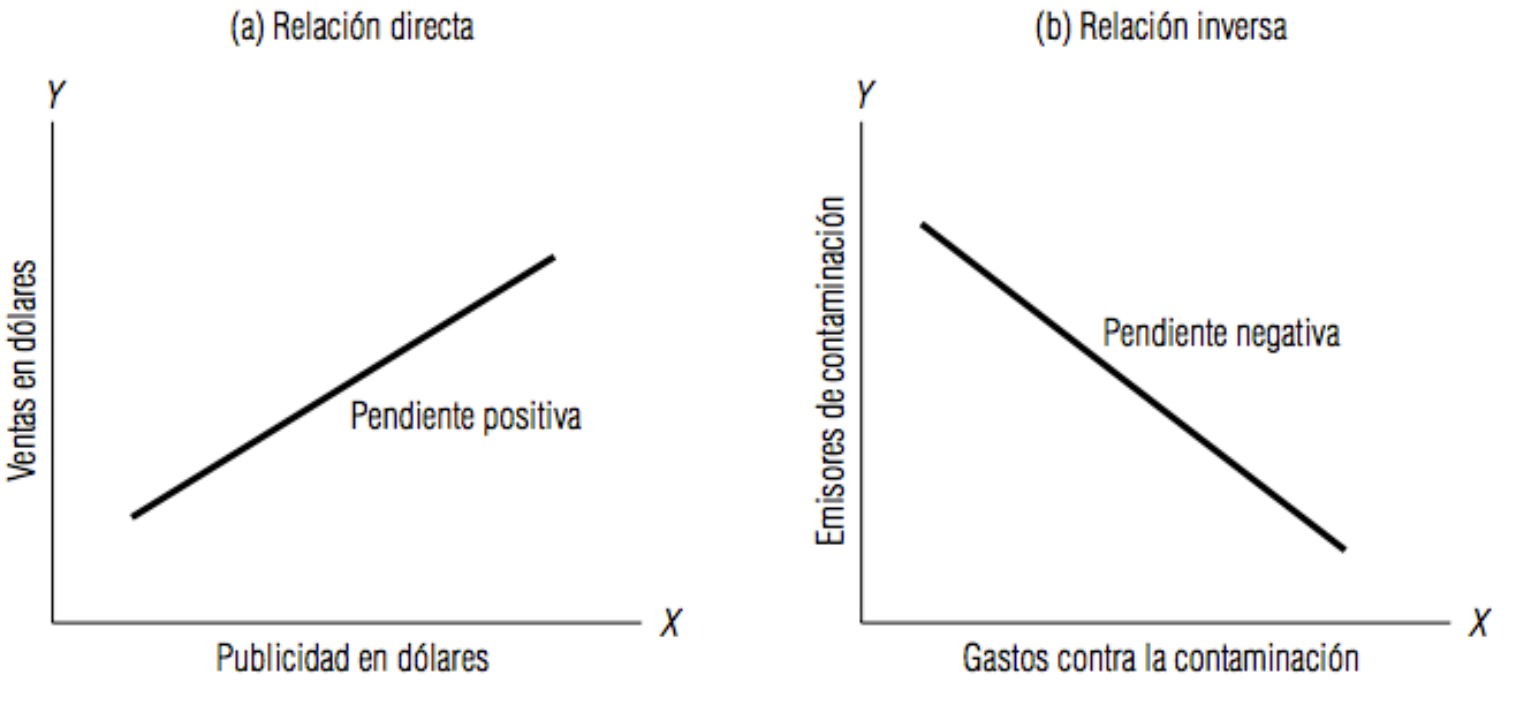
Los análisis de regresión y correlación nos mostrarán cómo determinar tanto la naturaleza como la fuerza de una relación entre dos variables. De esta forma, aprenderemos a pronosticar, con cierta precisión, el valor de una variable desconocida basándonos en observaciones anteriores de ésa y otras variables.

- **Tipos de relaciones**

Los análisis de regresión y de correlación se basan en la relación, o asociación, entre dos (o más) va- riables. La variable (o variables) conocida(s) se llaman *variable(s) independiente(s);* la que tratamos de predecir es la *variable dependiente*.

Regresión lineal simple y correlación

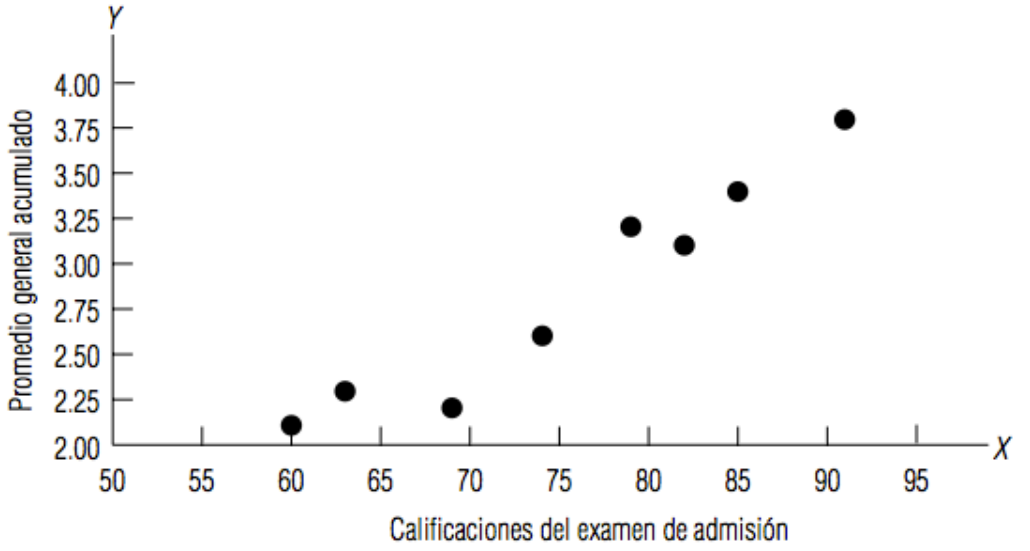
- **Tipos de relaciones**



Regresión lineal simple y correlación

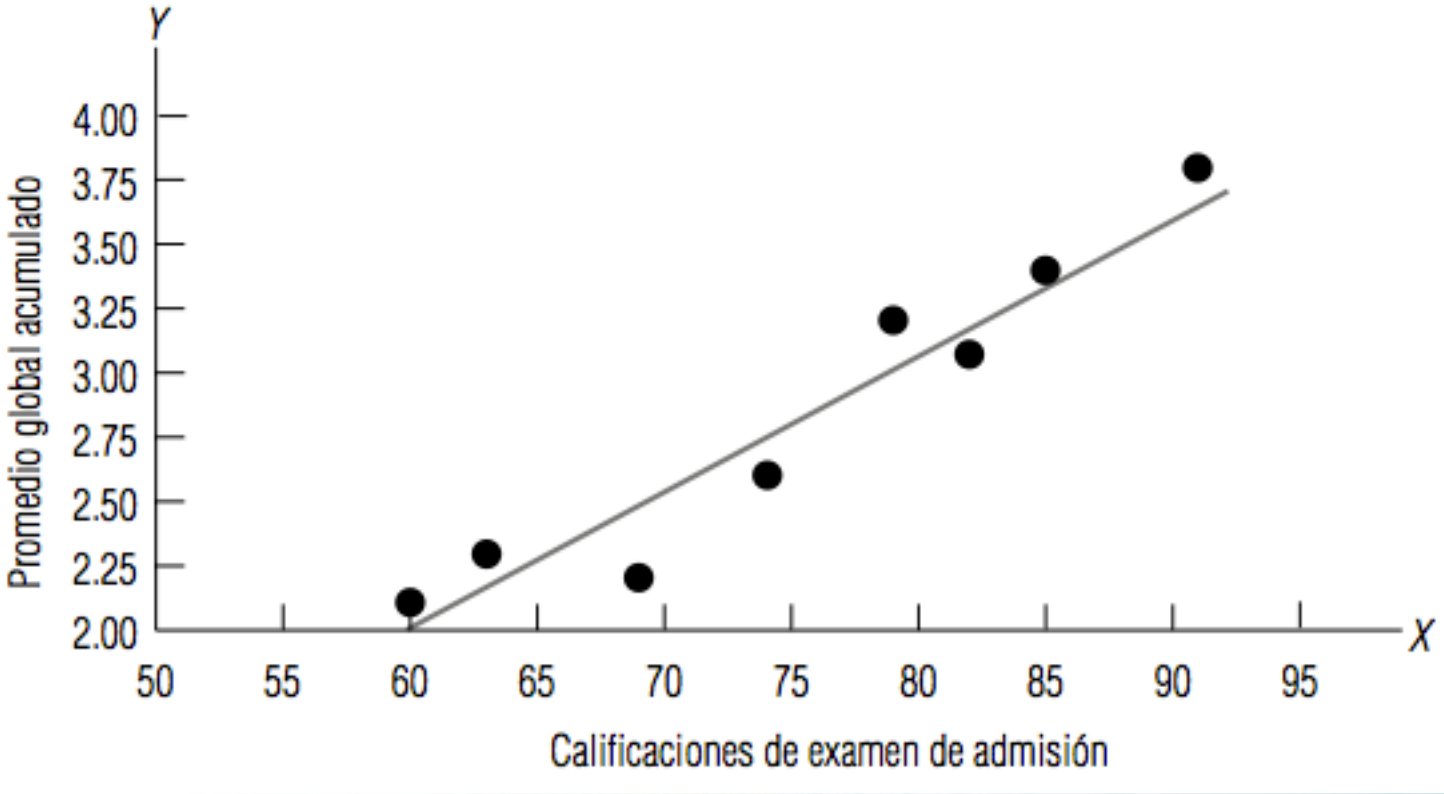
- **Diagramas de dispersión**

Estudiante	A	B	C	D	E	F	G	H
Calificaciones de examen de admisión (100 = máxima calificación posible)	74	69	85	63	82	60	79	91
Promedio general acumulado (4.0 = A)	2.6	2.2	3.4	2.3	3.1	2.1	3.2	3.8



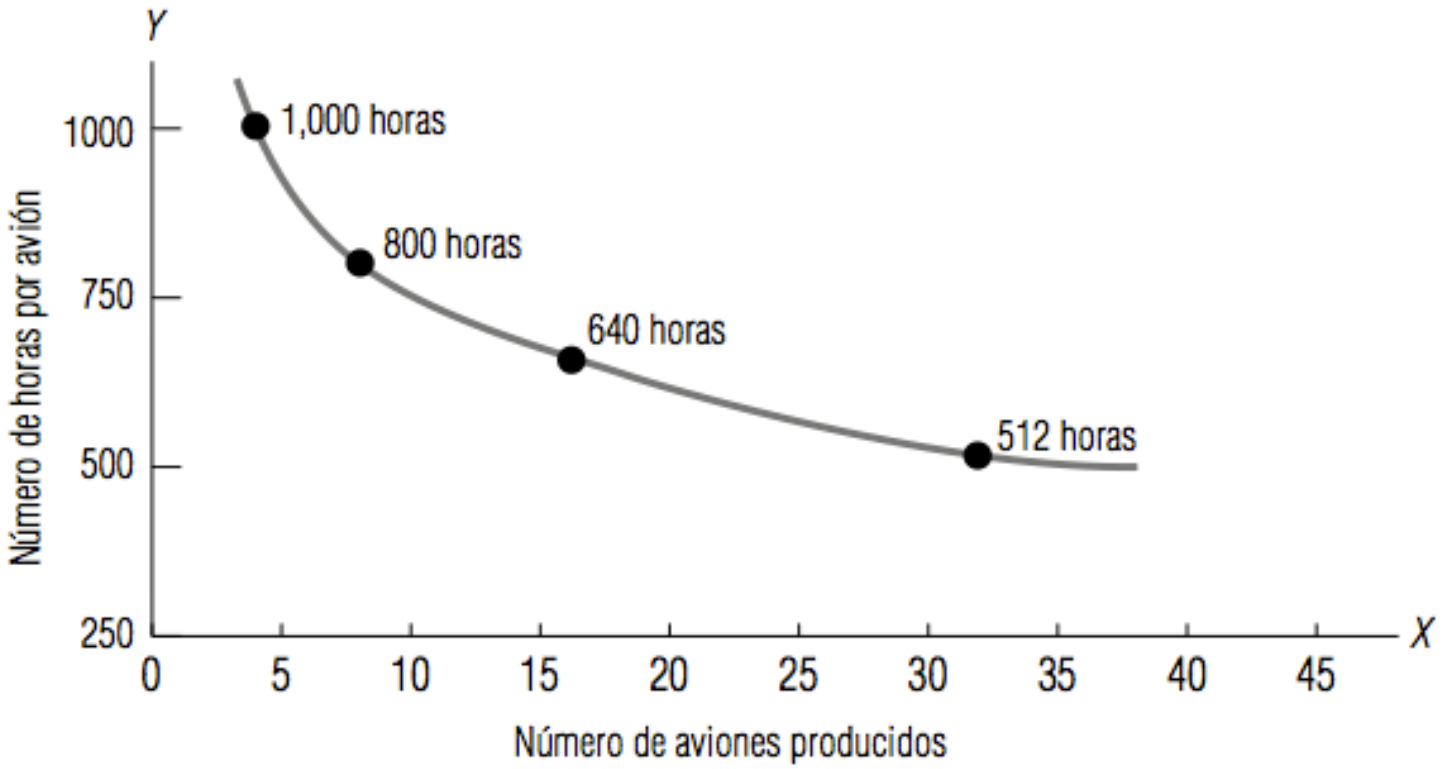
Regresión lineal simple y correlación

- **Diagramas de dispersión**



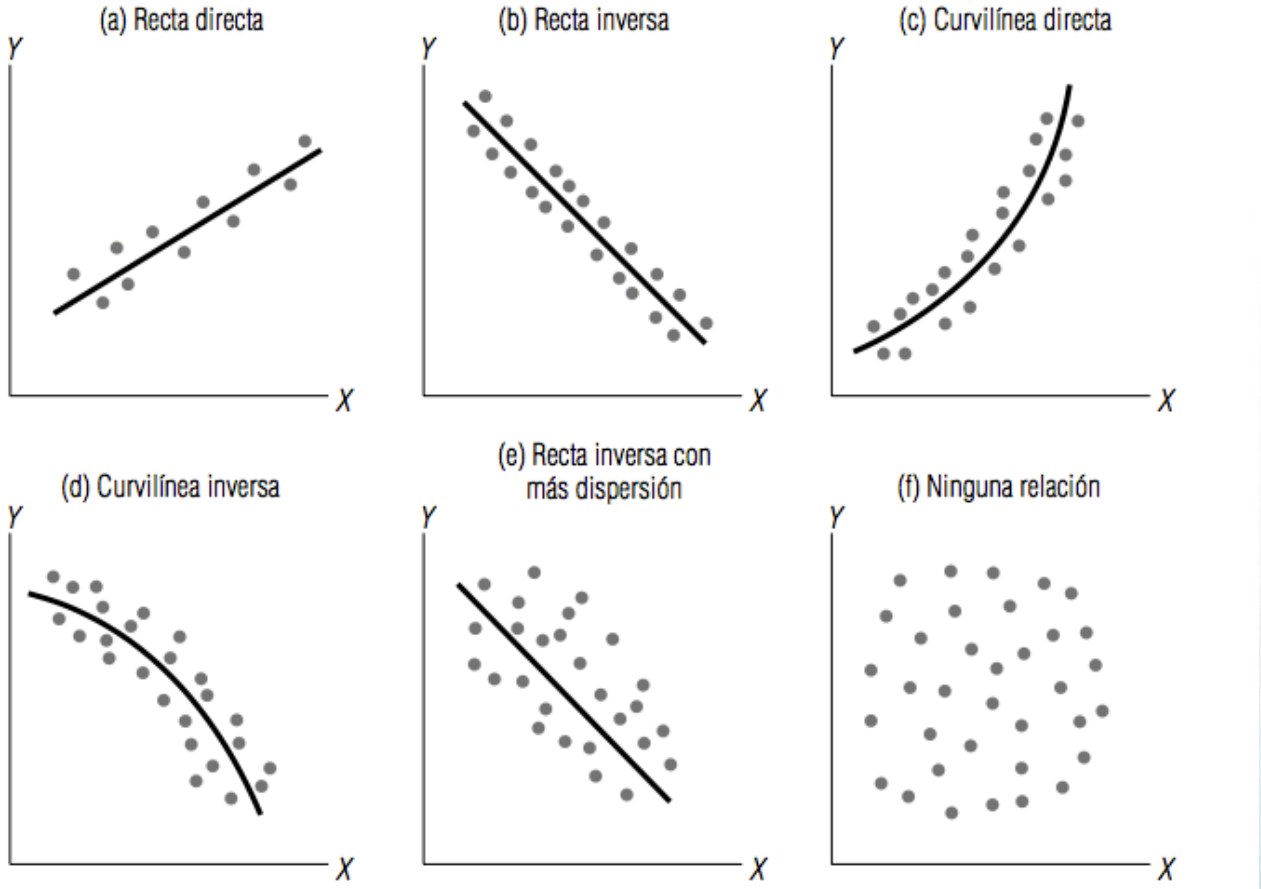
Regresión lineal simple y correlación

- **Diagramas de dispersión**



Regresión lineal simple y correlación

- **Diagramas de dispersión**



Regresión lineal simple y correlación

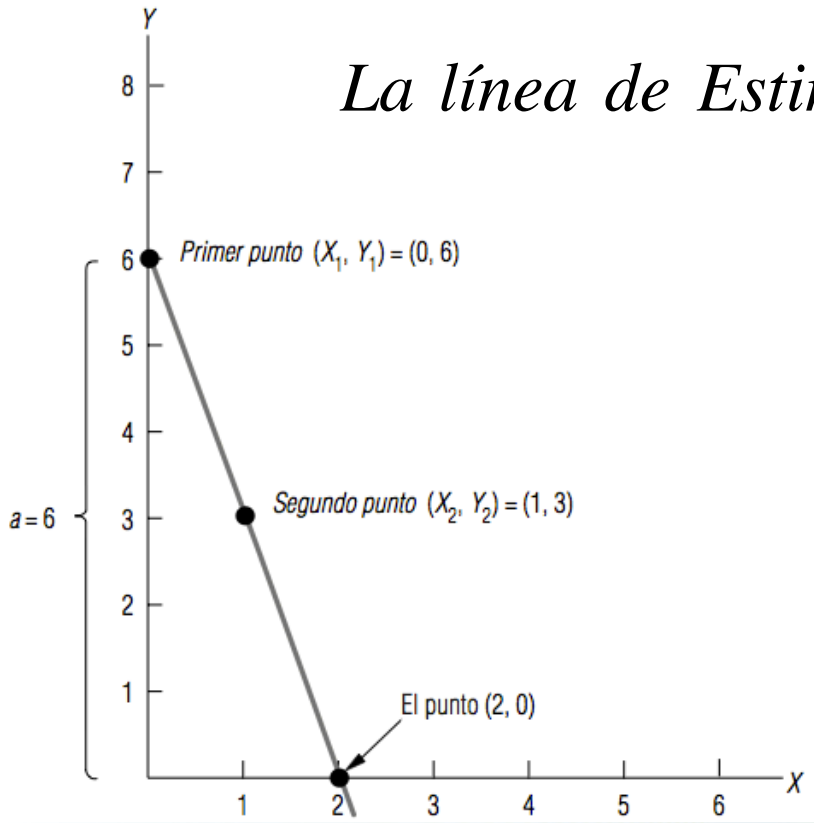
Ejemplo: Un instructor está interesado en saber cómo se relaciona el número de estudiantes ausentes con la temperatura media del día. Usó una muestra aleatoria de 10 días para el estudio. Los siguientes datos indican el número de estudiantes ausentes (AUS) y la temperatura media (TEMP) para cada día.

AUS	8	7	5	4	2	3	5	6	8	9
TEMP	10	20	25	30	40	45	50	55	59	60

a. Establezca la variable dependiente (Y) y la variable independiente (X).
b. Dibuje un diagrama de dispersión para estos datos.
c. ¿La relación entre las variables parece lineal o curvilínea?
d. ¿Qué tipo de curva puede dibujar a través de los datos?
e. ¿Cuál es la explicación lógica para la relación observada?

Regresión lineal simple y correlación

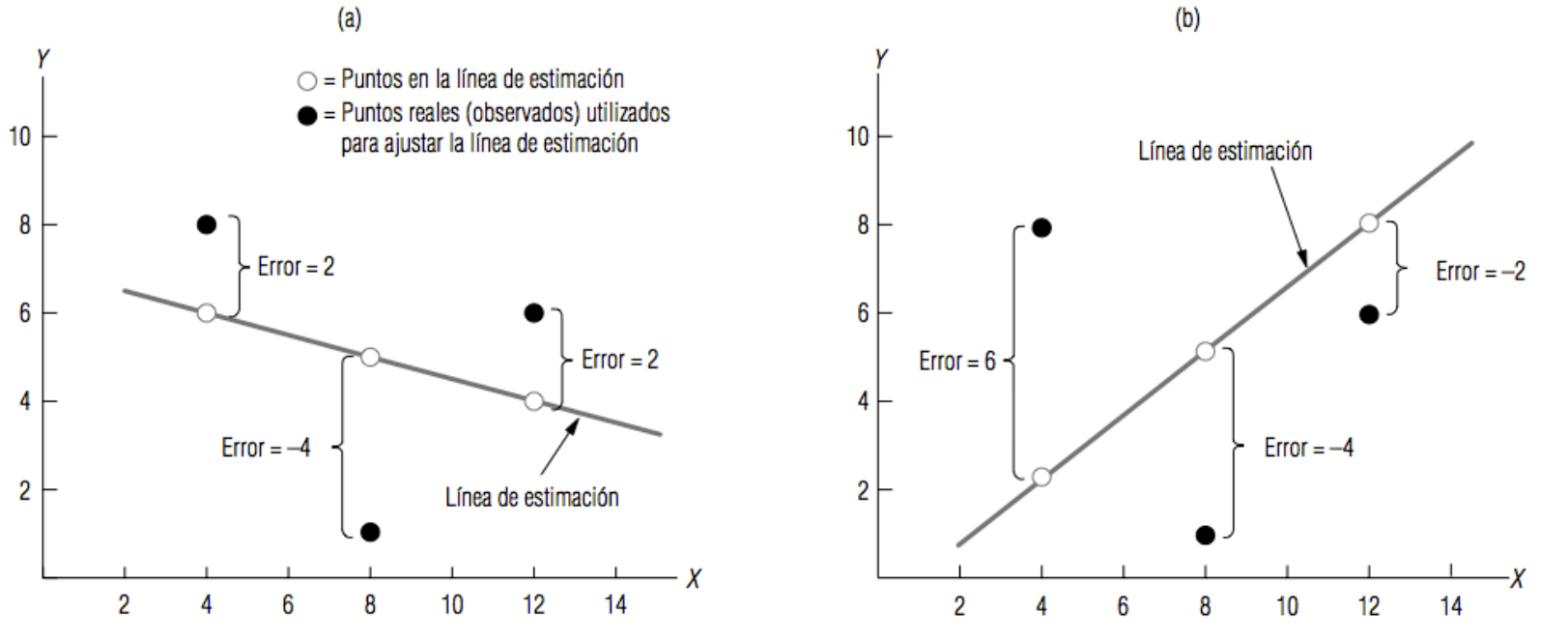
- **El método de mínimos cuadrados**



$$La\ línea\ de\ Estimación: \quad \hat{Y} = a + bX$$

Regresión lineal simple y correlación

- **Forma de "medir el error"**



Regresión lineal simple y correlación

- **Pendiente de la recta de regresión de mejor ajuste**

$$b = \frac{\sum XY - n\bar{X}\bar{Y}}{\sum X^2 - n\bar{X}^2}$$

donde,

- b = pendiente de la línea de estimación de mejor ajuste
- X = valores de la variable independiente
- Y = valores de la variable dependiente
- $\bar{X}$ = media de los valores de la variable independiente
- $\bar{Y}$ = media de los valores de la variable dependiente
- n = número de puntos (es decir, el número de pares de valores de las variables independiente y dependiente)

Regresión lineal simple y correlación

- **Ordenada Y de la recta de regresión de mejor ajuste**

$$a = \bar{Y} - b\bar{X}$$

donde,

- a = ordenada Y
- b = pendiente de la ecuación
- $\bar{Y}$ = media de los valores de la variable dependiente
- $\bar{X}$ = media de los valores de la variable independiente

Regresión lineal simple y correlación

- **Ejemplo:** A menudo, quienes hacen la contabilidad de costos estiman los gastos generales con base en el nivel de producción. En Standard Knitting Co. han reunido información acerca de los gastos generales y las unidades producidas en diferentes plantas, y ahora desean estimar una ecuación de regresión para predecir los gastos generales futuros.

Gastos generales	191	170	272	155	280	173	234	116	153	178
Unidades	40	42	53	35	56	39	48	30	37	40

a. dibuje un diagrama de dispersión.

b. Desarrolle una ecuación de regresión para contabilidad de costos.

c. Pronostique los gastos generales cuando se producen 50 unidades.

Regresión lineal simple y correlación

- **Método abreviado para encontrar el error estándar de la estimación**

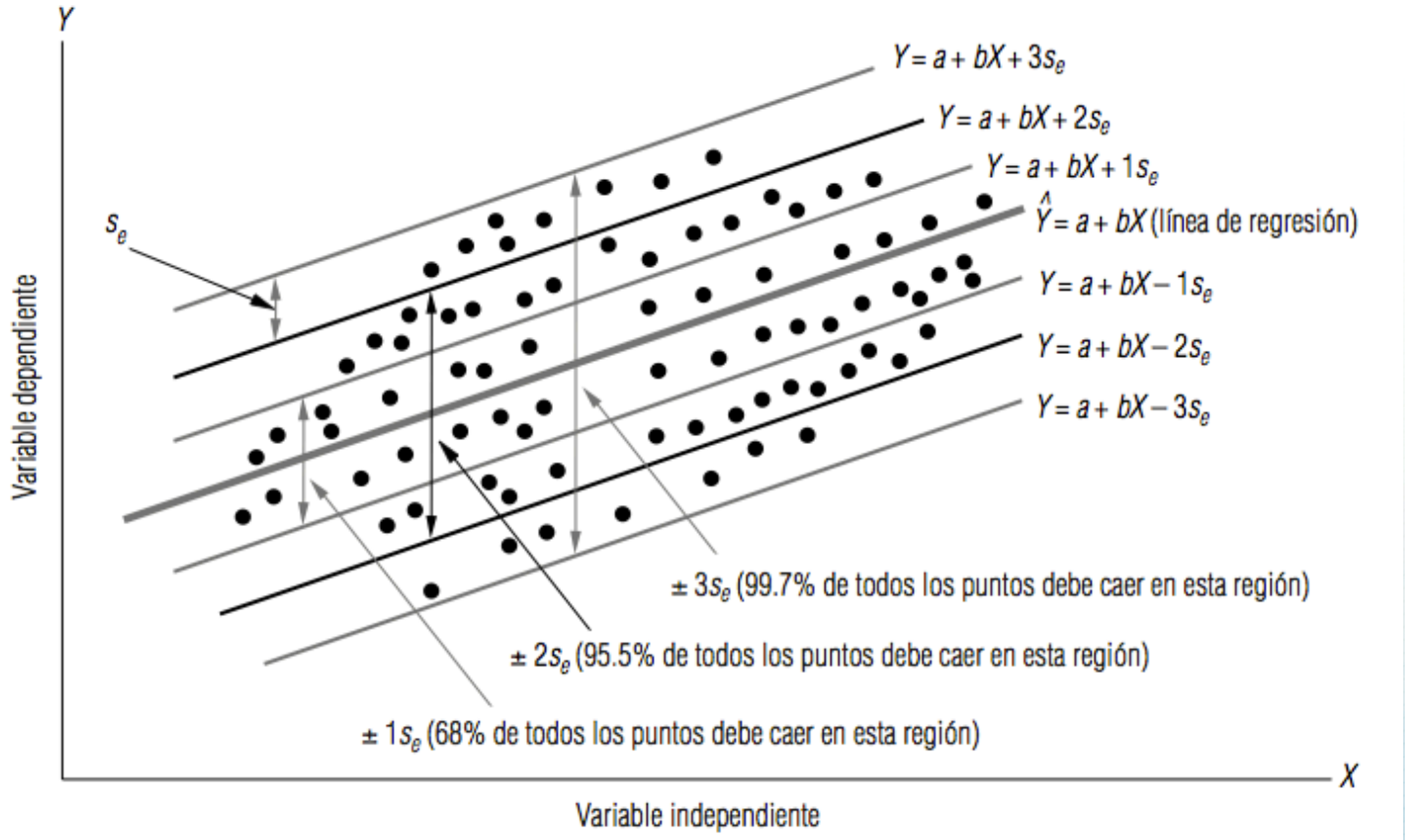
$$s_e = \sqrt{\frac{\sum Y^2 - a\sum Y - b\sum XY}{n-2}}$$

donde,

- X = valores de la variable independiente
- Y = valores de la variable dependiente
- a = ordenada Y de la ecuación 12-5
- b = pendiente de la ecuación de estimación
- n = número de puntos

Regresión lineal simple y correlación

- **Interpretación del error estándar de la estimación Intervalos de confianza para la estimación**



Regresión lineal simple y correlación

Apéndice tabla 2

*Áreas combinadas de ambos extremos para formar la distribución *t* de Student

Ejemplo:

Para encontrar el valor de *t* que corresponde a un área de 0.10 en ambos extremos de la distribución, cuando existen 19 grados de libertad, busque en la columna encabezada con 0.10 hasta el renglón correspondiente a 19 grados de libertad; el valor apropiado de *t* es 1.729.

Grados de libertad	Área combinada de ambos extremos			
	0.10	0.05	0.02	0.01
1	6.314	12.706	31.821	63.657
2	2.920	4.303	6.965	9.925
3	2.353	3.182	4.541	5.841
4	2.132	2.776	3.747	4.604
5	2.015	2.571	3.365	4.032
6	1.943	2.447	3.143	3.707
7	1.895	2.365	2.998	3.499
8	1.860	2.306	2.896	3.355
9	1.833	2.262	2.821	3.250
10	1.812	2.228	2.764	3.169
11	1.796	2.201	2.718	3.106
12	1.782	2.179	2.681	3.055
13	1.771	2.160	2.650	3.012
14	1.761	2.145	2.624	2.977
15	1.753	2.131	2.602	2.947
16	1.746	2.120	2.583	2.921
17	1.740	2.110	2.567	2.898
18	1.734	2.101	2.552	2.878
19	1.729	2.093	2.539	2.861
20	1.725	2.086	2.528	2.845
21	1.721	2.080	2.518	2.831
22	1.717	2.074	2.508	2.819
23	1.714	2.069	2.500	2.807
24	1.711	2.064	2.492	2.797
25	1.708	2.060	2.485	2.787
26	1.706	2.056	2.479	2.779
27	1.703	2.052	2.473	2.771
28	1.701	2.048	2.467	2.763
29	1.699	2.045	2.462	2.756
30	1.697	2.042	2.457	2.750
40	1.684	2.021	2.423	2.704
60	1.671	2.000	2.390	2.660
120	1.658	1.980	2.358	2.617
Distribución normal	1.645	1.960	2.326	2.576

*Tomado de la tabla III de Fisher y Yates, *Statistical Tables for Biological, Agricultural and Medical Research*, publicado por Longman Group, Ltd., Londres (publicado anteriormente por Oliver & Boyd, Edimburgo) y con licencia de los autores y los editores.

$$I.\ de\ Confianza: \begin{cases} límite\ \text{s}: Y = a + bX + t(s_e) \\ límite\ i: Y = a + bX - t(s_e) \end{cases}$$

Regresión lineal simple y correlación

- **Ejemplo**: Las ventas de línea blanca varían según el estado del mercado de casas nuevas: cuando las ventas de ca- sas nuevas son buenas, también lo son las de lavaplatos, lavadoras de ropa, secadoras y refrigeradores. Una asociación de comercio compiló los siguientes datos históricos (en miles de unidades) de las ventas de línea blanca y la construcción de casas.

Construcción de casas (miles)	Ventas de línea blanca (miles)
2.0	5.0
2.5	5.5
3.2	6.0
3.6	7.0
3.3	7.2
4.0	7.7
4.2	8.4
4.6	9.0
4.8	9.7
5.0	10.0

Regresión lineal simple y correlación

a. Desarrolle una ecuación para la relación entre las ventas de línea blanca (en miles) y la construcción de casas (en miles).

b. Interprete la pendiente de la recta de regresión.

c. Calcule e interprete el error estándar de la estimación.

d. La construcción de casas durante el año próximo puede ser mayor que el intervalo registrado; se han pronosticado estimaciones hasta de 8.0 millones de unidades. Calcule un intervalo de predicción de 90% de confianza para las ventas de línea blanca, con base en los datos anteriores y el nuevo pronós- tico de construcción de casas.

Regresión lineal simple y correlación

- **Coeficiente de determinación muestral r^2**

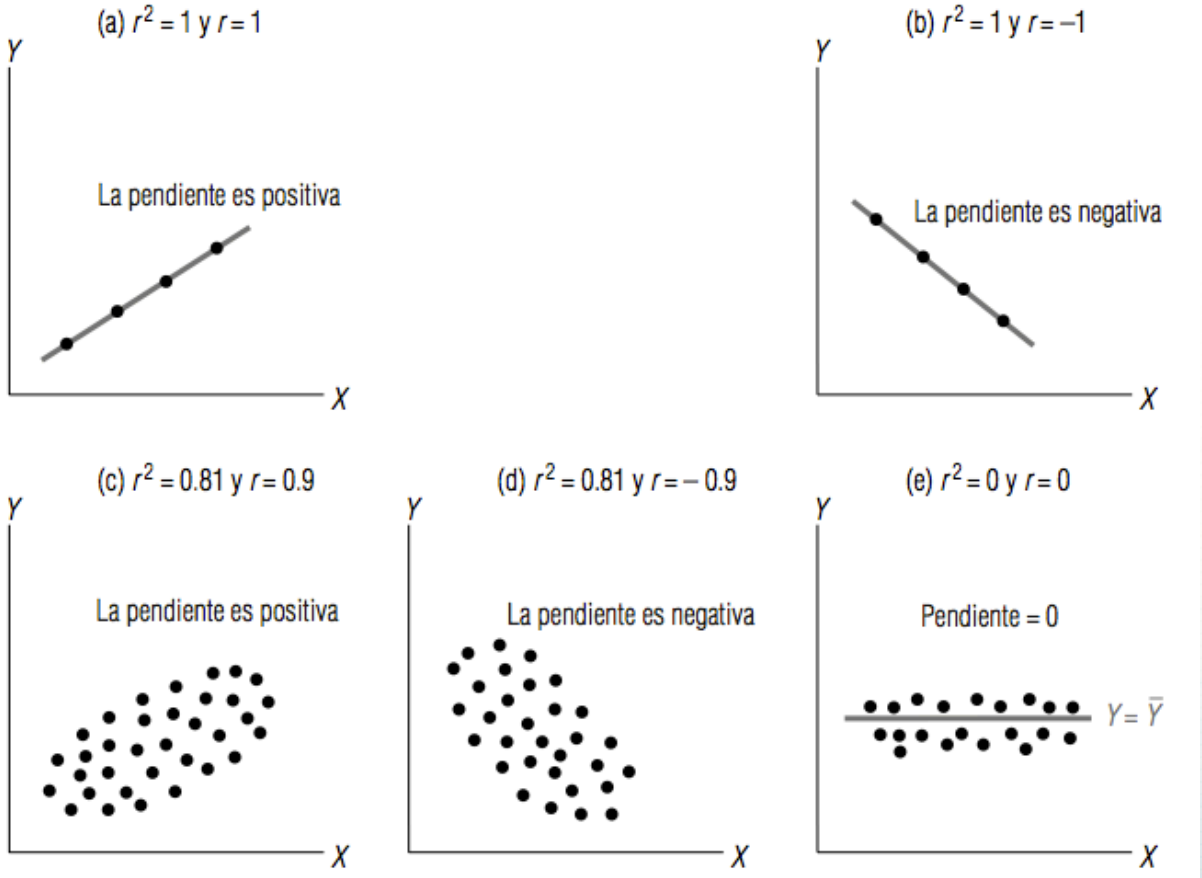
$$r^2 = \frac{a\sum Y + b\sum XY - n\bar{Y}^2}{\sum Y^2 - n\bar{Y}^2}$$

donde,

- r^2 = coeficiente de determinación de la muestra
- a = ordenada Y
- b = pendiente de la línea de estimación de mejor ajuste
- n = número de puntos de datos
- X = valores de la variable independiente
- Y = valores de la variable dependiente
- $\bar{Y}$ = media de los valores observados de la variable dependiente

Regresión lineal simple y correlación

- **Coeficiente de correlación de la muestra** $r = \sqrt{r^2}$



Regresión lineal simple y correlación

Ejemplo: Zippy Cola está estudiando el efecto de su última campaña publicitaria. Se escogieron personas al azar y se les llamó para preguntarles cuántas latas de Zippy Cola habían comprado la semana anterior y cuántos anuncios de Zippy Cola habían leído o visto durante el mismo periodo.

X (número de anuncios)	3	7	4	2	0	4	1	2
Y (latas compradas)	11	18	9	4	7	6	3	8

a. Desarrolle la ecuación de estimación que mejor ajuste los datos.

b. Calcule el coeficiente de determinación de la muestra y el coeficiente de correlación.