Resolución del Ejercicio 6 Práctica Regresión Lineal

Un comerciante realizó un estudio para determinar la relación que hay entre los gastos de la publicidad semanal y las ventas. Registró los datos siguientes:

X: Costos de publicidad (en $): 40, 20, 25, 20, 30, 50, 40, 20, 50, 40, 25, 50.

Y: Ventas (en $): 385, 400, 395, 365, 475, 440, 490, 420, 560, 525, 480, 510.

a) Haga un gráfico de dispersión

b) Encuentre la recta de regresión estimada para pronosticar las ventas semanales, a partir de los gastos de publicidad.

c) Estime las ventas semanales cuando los costos de la publicidad sean 35$. ¿Es válido estimar las ventas semanales cuando los costos de la publicidad sean 75$?

d) Pruebe las hipótesis de que $\beta_1 = 6$ contra la alternativa de que $\beta_1 < 6$, utilice $\alpha = 0.025$.

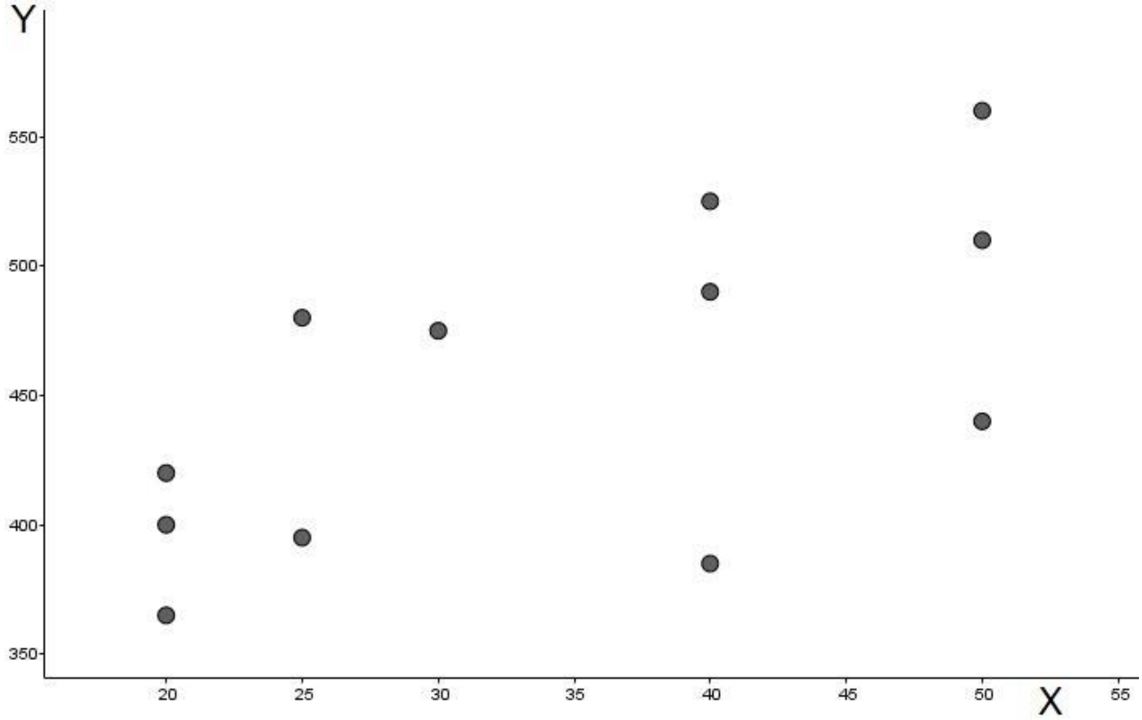
e) Construya un intervalo de confianza de 95% para la media de las ventas semanales cuando se gastan 45$ en publicidad.

f) Construya un intervalo de predicción de 95% para la media de las ventas semanales cuan do se gastan 45$ en publicidad.

g) ¿Qué proporción de la variabilidad total en las ventas está explicada por el costo en publicidad?

Solución:

a) **Diagrama de Dispersión:**



b) **Recta de regresión**: (usaré 4 decimales)

$\hat{y} = \widehat{\beta_0} + \widehat{\beta_1}\, x$ = 343,7028+ 3,2209 x

$\widehat{\beta_1} = \frac{S_{xy}}{S_{xx}}$ = 3,2209

$\widehat{\beta_0} = \bar{y} - \widehat{\beta_1}\bar{x}$= 343,7028

Cálculos necesarios para calcular S_{xy}, S_{yy} y S_{xx}
$\bar{x} = 34{,}1666$
$\bar{y}$= 453,75

$\sum_{i=1}^{n} x_i^{\,2}$= 15.650

$\sum_{i=1}^{n} y_i^{\,2}$= 2.512.925

$\sum_{i=1}^{n} x_i y_i$= 191.325

$S_{xy} = \sum_{i=1}^{n} x_i y_i - \text{n}\ \bar{x}\ \bar{y}$= 5.287,863

$S_{xx} = \sum_{i=1}^{n} x_i^{\,2}$-n $\bar{x}^2$= 15.650- 12(34,16666)^2= 1.641,7213
S_{yy}=$\sum_{i=1}^{n} y_i^{\,2}$-n $\bar{y}^2$= 2.512.925-12(453,75)^2= 42.256,25

c) Como el valor 35 pertenece al rango de valores de las x, se puede estimar las ventas semanales, esto es:

$\hat{y}_{(35)}$= 343,7028+ 3,2209. 35= 456,4343

Como 75 no es un valor dentro del rango de las x no puedo estimar las ventas semanales.

El costo de publicidad, hace referencia a la variable X, toma valores entre 20 y 50.

d) Probar las **hipótesis**:

H_0: $\beta_1 = 6$ vs H_a: $\beta_1 < 6$ (test unilateral por izquierda)

Estadístico de Prueba:

$$T = \frac{\widehat{\beta_1} - 6}{s_r / \sqrt{S_{xx}}} \sim T(n-2) \text{ bajo } H_0$$

Región de Rechazo:

Rechazo H_0 a favor de H_a si: $t_O < -t_{\alpha, n-2}$

No rechazo H_0 a favor de H_a si: $t_O \geq -t_{\alpha, n-2}$

En este problema t_O= -2,2421 y $-t_{\alpha, n-2} = -t_{0,025,10}$= -2,228

Conclusión: Como se cumple que -2,2421<-2,228, por lo tanto: -2,2421 cayó en la región de rechazo, hay suficiente evidencia contra H_0 a favor de H_a, es decir, que con un nivel de significancia del 0,025 puedo afirmar que $\beta_1 < 6$.

Cálculos auxiliares:

$\hat{\sigma} = S_r = \sqrt{\frac{S_{rr}}{n-2}} = \sqrt{\frac{25224{,}434}{10}}$ = 50,2239

$SS_R = S_{rr} = S_{yy} - \frac{{S_{xy}}^2}{S_{xx}}$ = 25224,434

e) Como x_0=45 está dentro de los valores que puede tomar la variable X, puedo calcular el intervalo pedido.

Función Pivote a utilizar y distribución:

$$\frac{\hat{\beta}_0 + \hat{\beta}_1 x_0 - (\beta_0 + \beta_1 x_0)}{\sqrt{\hat{\sigma}^2 \left[\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right]}} \text{ tiene distribución Student con } n-2 \text{ grados de libertad}$$

El intervalo resultante es:

$$\left[\hat{\beta}_0 + \hat{\beta}_1 x_0 - t_{\frac{\alpha}{2},n-2}\sqrt{\hat{\sigma}^2\left[\frac{1}{n}+\frac{(x_0-\bar{x})^2}{S_{xx}}\right]};\ \hat{\beta}_0 + \hat{\beta}_1 x_0 + t_{\frac{\alpha}{2},n-2}\sqrt{\hat{\sigma}^2\left[\frac{1}{n}+\frac{(x_0-\bar{x})^2}{S_{xx}}\right]}\right]$$

Haciendo las cuentas:

$t_{\alpha/2,\, n-2}=t_{0,025,10}=2,228$

$\widehat{\beta_0} + \widehat{\beta_1}$ x_0= 343,7028+ 3,2209. 45= 488,433

$\hat{\sigma}$ =S_r =$\sqrt{\frac{S_{rr}}{n-2}}$= $\sqrt{\frac{25224,434}{10}}$= 50,2239

Reemplazo los datos en el intervalo:

[488,433-2,228. 50,2239$\sqrt{0,1548}$; 488,433+2,228. 50,2239 $\sqrt{0,1548}$]

[488,433-44,0291; 488,433+44,0291]

El intervalo pedido es: [444,4039; 532,4621]

f) Como x_0=45 está dentro de los valores que puede tomar la variable X, puedo calcular el intervalo pedido.

Función Pivote a utilizar y distribución:

$$\frac{Y_0 - \hat{Y}_0}{\sqrt{\hat{\sigma}^2\left[1+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{S_{xx}}\right]}} \sim t_{n-2}$$

El intervalo resultante es:

$$\left[\widehat{Y_0} - t_{\alpha/2,\, n-2}\, S_r\sqrt{1+\frac{1}{n}+\frac{(x_o-\bar{x})^2}{S_{xx}}};\ \widehat{Y_0} + t_{\alpha/2,\, n-2}\, S_r\sqrt{1+\frac{1}{n}+\frac{(x_o-\bar{x})^2}{S_{xx}}}\right]$$

Haciendo las cuentas:

$t_{\alpha/2,\, n-2}=t_{0,025,10}=2,228$

$\widehat{Y_0} = \widehat{\beta_0} + \widehat{\beta_1}$ x_0= 343,7028+ 3,2209. 45= 488,433

$\hat{\sigma}$ = S_r=$\sqrt{\frac{S_{rr}}{n-2}}$= $\sqrt{\frac{25224,434}{10}}$= 50,2239

[488,433-2,228. 50,2239$\sqrt{1,1548}$; 488,433+2,228. 50,2239$\sqrt{1,1548}$]

[488,433-120,2483; 488,433+120,2483]

El intervalo resultante es: [368,1847; 608,6813]

Notar que este intervalo es más ancho que el del ejercicio e)

g) Debemos calcular el Coeficiente de determinación

$$R^2 = 1 - \frac{SS_R}{S_{YY}}$$

Siendo $SS_R = S_{rr}$

$R^2 = 1 - S_{rr}/S_{yy}$ =1- 25224,434/42.256,25= 1-0,5969= 0,4031

La proporción de la variabilidad total en las ventas está explicada por el costo en publicidad es aproximadamente del 40%.