

ESTADISTICAS II

Se desea estimar las ventas semanales de una empresa, que se dedica a la venta de electrodomésticos. Para ello relaciona las ventas (en millones de pesos) con los gastos en publicidad, y con las ventas de artículos alternativos.

La información alternativa a un periodo de 12 semanas escogidas aleatoriamente desde un periodo de 2 años, como se presenta a continuación.

SEMANA	VENTAS (millones \$)	GASTOS (millones \$)	VENTAS ART. ALTERNATIVO
1	13.2	2.5	25.5
2	16.4	3	24
3	14.3	2.8	24.4
4	16.5	4.2	21.8
5	22.3	5.3	20.1
6	15.8	3.6	22.8
7	11.2	2.7	24.6
8	13.3	3.2	23.8
9	8.6	1.9	26.2
10	12.5	3.6	22.8
11	16.5	4.6	20.8
12	16	3.8	22.2

- Encuentre las ecuaciones de regresión lineal estimadas de los modelo:

- $Y_i = a + bX_i + e_i$

- $Y_i = a + bZ_i + e_i$

Interprete el valor de las pendientes.

- Pruebe la significancia de cada modelo y con base estadística elija el mejor.
- Se piensa que entre las ventas Y , y los gastos de publicidad X existe una correlación positiva de por lo menos 70%, concluya con un nivel de significación de 5%.
- Para cada modelo anterior (pregunta A), dibuje el diagrama de dispersión junto a la ecuación de regresión lineal, y según los gráficos cual ecuación se ajusta mejor.
- Se plantean los siguientes modelos que relacionan las ventas semanales con los gastos en publicidad y las ventas del artículo alternativo.

- $Y_i = a + bX_i + e_i$

- $Y_i = a + bZ_i + e_i$

Encuentre los coeficientes de cada modelo por medio de un indicador adecuado, elija cual de ellos se ajusta mejor

- Para el modelo 2 de la pregunta anterior estime las ventas para un nivel de gasto en publicidad de 2.7 millones, use un 95% de confianza.

SOLUCIÓN:

A) Modelo 1:

$$X_i = 41.2$$

$$Y_i = 176.6$$

$$X_i * Y_i = 637.45$$

$$X_i^2 = 151.48$$

$$= n * \frac{X_i * Y_i - (X_i)(Y_i)}{n * X_i^2 - (X_i)^2}$$

$$= \frac{12(637.45) - (41.2)(176.6)}{12(151.48) - 1697.44}$$

$$= \frac{12(637.45) - (41.2)(176.6)}{12(151.48) - 1697.44}$$

$$= 3.104$$

$$= 3.104$$

$$= Y_i - b_1 X_i$$

$$n$$

$$= \frac{176.6 - 3.104(41.2)}{12}$$

$$12$$

$$= 4.0596$$

Por lo tanto el modelo 1 es: $Y_i = 4.0596 + 3.104 * X_i$

De donde se puede decir que si la variable X aumenta en una unidad la variable Y varia 3.104 veces, y no tiene ninguna interpretación practica.

Modelo 2:

$$Z_i = 279$$

$$Z_i * Y_i = 4048.07$$

$$Z_i^2 = 6524.06$$

$$Y_i = 176.6$$

$$= n * \frac{Z_i * Y_i - (Z_i)(Y_i)}{n * Z_i^2 - (Z_i)^2}$$

$$= \frac{12(3646.07) - (179)(176.6)}{12(6524.06) - 77841}$$

$$= \frac{12(3646.07) - (179)(176.6)}{12(6524.06) - 77841}$$

$$12(6524.06) - 77841$$

$$= -1.5513$$

$$= \frac{\sum (Y_i - \bar{Y})(Z_i - \bar{Z})}{\sum (Z_i - \bar{Z})^2}$$

n

$$= \frac{176.6 + 1.5513(179)}{12}$$

12

$$= 50.7844$$

Por lo tanto el modelo 2 es: $Y_i = 50.7844 - 1.5513Z_i$

De donde se puede decir que si la variable Z aumenta en una unidad, la variable Y varía 1.5513 veces, y no tiene ninguna interpretación práctica.

- Modelo 1:

– Hipótesis: $H_0: \beta_0 = 0$ (el modelo no se ajusta a los datos)

$H_1: \beta_0 \neq 0$ (el modelo se ajusta a los datos)

- Estadístico de prueba:

$$F = \frac{CMR}{CME} = \frac{SMR}{SCE} \sim F(1; n-2; 1-\alpha)$$

$$CME = SCE / (n-2)$$

$$SMR = \sum (Y_i - \bar{Y})^2 \quad (\text{el subíndice } p \text{ indica promedio})$$

$$SMR = 3.10412 \sum (Y_i - 14.72)^2$$

$$SMR = 96.6126$$

$$SCE = SCT - SMR$$

$$SCT = \sum (Y_i - \bar{Y})^2$$

$$SCT = 128.6968$$

$$SCE = 128.6968 - 96.6126$$

$$SCE = 32.0842$$

$$F_{obs} = \frac{96.6126}{32.0842} = 3.01122$$

$$32.0842/2$$

– RC: $F(1; 10; 0.95) = 4.9646$

- Como F_{obs} pertenece a la región crítica existe suficiente evidencia para rechazar H_0 por lo que con

un 5% de significación se puede decir que el modelo es significativo.

- P-value

$$* = \text{Prob}(F(1;10) > F_{\text{obs}})$$

$$* = \text{Prob}(F(1;10) > 30.1122)$$

$$* = 1 - \text{Prob}(F(1;10) < 30.1122)$$

$$* = 1 - 0.9997335$$

$$* = 0.0002664$$

Modelo 2:

– Hipótesis: $H_0: \beta = 0$ (el modelo no se ajusta a los datos)

$H_1: \beta \neq 0$ (el modelo se ajusta a los datos)

- Estadístico de prueba:

$$F = \frac{\text{CMR}}{\text{SCE}} = \frac{\text{SMR}}{N-2} \sim F(1; n-2; 1-\alpha)$$

$$\text{CME} = \text{SCE} / (N-2)$$

$$\text{SMR} = \sum_{i=1}^n (Z_i - \bar{Z})^2 \quad (\text{el subíndice } p \text{ indica promedio})$$

$$\text{SMR} = (-1.5513)^2 \sum_{i=1}^n (Y_i - 23.25)^2$$

$$\text{SMR} = 89.7877$$

$$\text{SCE} = \text{SCT} - \text{SMR}$$

$$\text{SCT} = \sum_{i=1}^n (Y_i - \bar{Y})^2$$

$$\text{SCT} = 128.6968$$

$$\text{SCE} = 128.6968 - 89.7877$$

$$\text{SCE} = 38.9091$$

$$F_{\text{obs}} = \frac{89.7877}{38.9091} = 23.0763$$

$$38.9091/2$$

– RC: $F(1;10;0.95) = 4.9646$

– Como F_{obs} pertenece a la región crítica existe suficiente evidencia para rechazar H_0 por lo que con un 5% de significación se puede decir que el modelo es significativo.

– P-value

$$* = \text{Prob}(F(1;10) > F_{obs})$$

$$* = \text{Prob}(F(1;10) > 23.07)$$

$$* = 1 - \text{Prob}(F(1;10) > 30.1122)$$

$$* = 1 - 0.9992797$$

$$* = 0.00072$$

Coeficiente de determinación:

Modelo 1:

$$SCR = (n-1)b_1^2 S_x^2$$

$$96.612 = 11(3.104)^2 S_x^2$$

$$S_x^2 = 0.91158$$

$$SCT = (n-1)S_y^2$$

$$128.6968 = 11S_y^2$$

$$S_y^2 = 11.6997$$

$$R^2 = b_1^2 \underline{S_x^2}$$

$$S_y^2$$

$$R^2 = (3.104)^2 * \underline{0.91158}$$

$$11.6997$$

$$R^2 = 0.75069$$

Modelo 2:

$$SCR = (n-1) S_z^2$$

$$89.7877 = 11(-1.5513)^2 S_z^2$$

$$S_z^2 = 3.3918$$

$$SCT = (n-1)S_y^2$$

$$128.6968 = 11S_y^2$$

$$S_y^2 = 11.6997$$

$$R^2 = \hat{E} * \underline{S_z^2}$$

Sy2

$$R^2 = (-1.5513)^2 * \underline{3.3918}$$

11.6997

$$R^2 = 0.6976$$

Como el percentil es mayor en el modelo 1, el porcentaje de ajuste es mayor en 1.

Coeficiente de correlación:

$$= x \text{ ---}$$

y

$$= Sx \text{ ---}$$

Sy

Modelo 1:

$$= (3.104) * \underline{0.9547}$$

3.4205

$$= 0.86636$$

Modelo 2:

$$= (-1.5513) * \underline{1.8416}$$

3.4205

$$= -0.83522$$

Para el modelo 2 existe una relación lineal inversa, ya que es negativo

Por lo tanto el modelo 1 es mejor que el modelo 2.

C) Modelo 1:

– Hipótesis: $H_0: \rho = 0.7$ (no existe regresión)

$H_1: \rho > 0.7$ (existe regresión)

• Estadístico de prueba:

$$Z_{obs} = (Z_r - Z_c) * (n-3)^{1/2} N(0;1)$$

$$r = (SIG)^{b1} (SCR/SCT)^{1/2}$$

$$r = (9.612/128.6968)^{1/2}$$

$$r = 0.8664$$

$$Z_r = 1/2(\ln((1+r)/(1-r)))$$

$$Z_r = 0.5(\ln((1+0.8664)/(1-0.8664)))$$

$$Z_r = 1.3184$$

$$Z = 1/2(\ln((1+)/(1-)))$$

$$Z = 1/2(\ln((1+0.7)/(1-0.7)))$$

$$Z = 0.8673$$

$$Z_{obs} = (1.3184 - 0.8673) \cdot 3$$

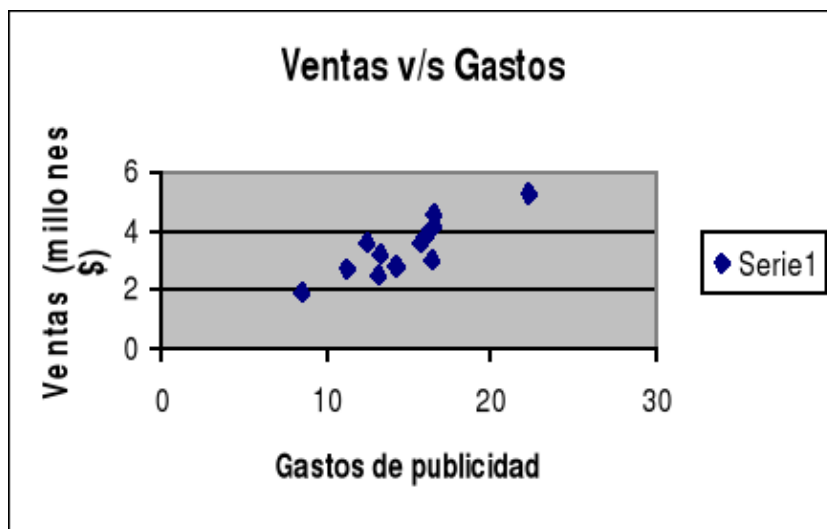
$$Z_{obs} = 1.3532$$

- RC:

Distribuye normal con $\mu = 0$ y $\sigma = 1$, con 5% de significación

$$Z = 1.644 \text{ (Z de tabla)}$$

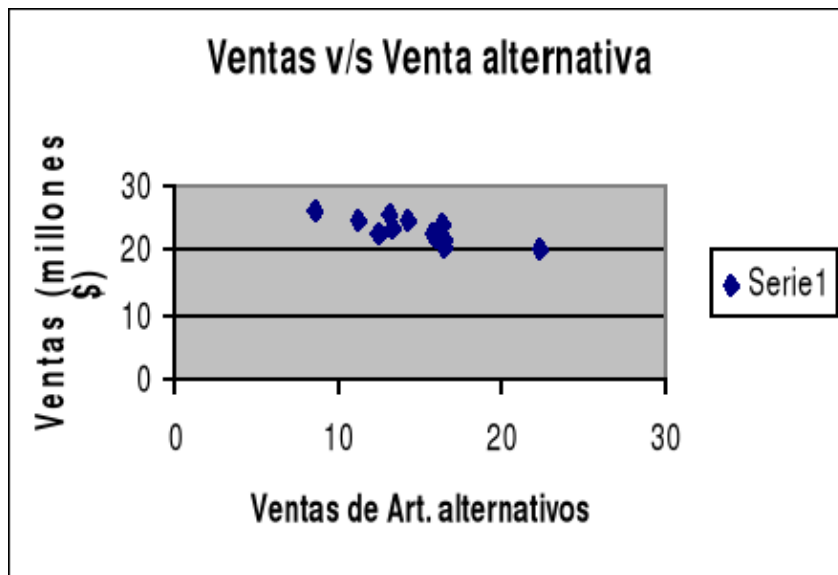
- Como Z observado no pertenece a la región crítica no existe suficiente evidencia para rechazar H_0 . Por lo tanto para la variabilidad de los datos no existe una correlación positiva del 70%.



D) Modelo 1:

$$Y_i = 4.0596 + 3.104 \cdot X_i$$

Modelo 2:



$$Y_i = 50.7844 - 1.5513Z_i$$

Como se puede apreciar los datos de el modelo 1 se ajustan de una mejor forma al modelo que se propuso, teniendo una variación de las ventas con respecto a los gastos en publicidad positiva.

- Modelo 1:

$$Y_i = X_i + i$$

Para este caso $b_1 = 1$ y $b_0 = 0$, la ecuación del modelo pasa por el origen.

$$S = \sum e_i^2 = \sum (Y_i - Y_{ie})^2 \text{ (el subíndice e indica estimación)}$$

$$\sum (Y_i - X_i)^2$$

$$\frac{dS}{dX_i} = 0$$

$$d$$

$$2 \sum (Y_i - X_i)(-X_i) = 0 \quad \text{---} \quad \frac{1}{2}$$

$$\sum (-X_i * Y_i) + \sum X_i^2 = 0$$

$$\sum X_i^2 = \sum X_i * Y_i$$

$$= \sum \underline{X_i * Y_i}$$

$$\sum X_i^2$$

datos:

$$\sum X_i * Y_i = 637.45$$

$$\sum X_i^2 = 151.48$$

Por lo tanto

$$=4.208$$

Lo que implica que el modelo seria $Y_i = 4.208X_i + i$

Modelo 2:

$$Y_i = A * Z_i^c$$

Tenemos que $Y_i = a + c * Z_i$

Donde " $Z_i = \ln(Z_i) = 37.72$

" $Y_i = \ln(Y_i) = 31.96$

" $Z_i * Y_i = \ln(Z_i * Y_i) = 100.298$

$Y_p = \ln(Y_i)/n = 2.663$

$Z_p = \ln(Z_i)/n = 3.143$

" $Z_i^2 = 6524.06$

Tenemos que $c = \frac{Z_i * Y_i - ("Z_i")("Y_i")/n}{$

$(("Z_i^2") - ("Z_i")^2/n)$

$c = \frac{100.298 - (37.72)(31.96)/12}{$

$6524.06 - 37.72^2/12$

$c = -2.5436 * 10^{-5}$

$a = Y_p - c * Z_p$

$a = 2.663 + 2.5436 * 10^{-5}(3.143)$

$a = 2.663$

Por lo tanto $A = e^a$

$A = e^{2.663}$

$A = 14.3403$

- $Y_i = A * Z_i^c$

Gasto=2,7 millones, con 95% de confianza.

Reemplazando 2,7 en la ecuación anterior:

$$Y_0 = (14,3403) \cdot (2,7) - 0.000025436$$

$$Y_0 = 14,339$$

$$Y_0 \pm [Y_0 \pm T_{n-2, 1-\alpha/2} \cdot r](1 -)$$

$$r = [CME \cdot (1 + 1/n + (Z_0 - Z_P)^2 / (n(Z_i - Z_P)))^{1/2}]$$

$$Z_P = \sum Z_i / n = 23.25$$

$$CME = SCE / (n - 2)$$

$$SCE = SCT - SCR$$

$$SCR = c^2 \cdot \sum (Z_i - Z_P)^2 = (-2.5436 \cdot 10^{-5})^2 \cdot 37.31$$

$$SCR = 2.4139 \cdot 10^{-8}$$

$$SCT = \sum (Y_i - Y_P)^2 = 128.6968$$

$$SCE = 128.6968 - 2.4139 \cdot 10^{-8}$$

$$SCE = 128,6967$$

$$CME = 128.6967 / 10 = 12,8696$$

$$r = [12.89696(1 + 1/12 + (2.7 - 23.25)^2 / 37.31)]$$

$$r = 12.6336$$

$$Y_0 \pm [14.339 \pm T_{100.975} \cdot 12.6336]$$

$$Y_0 \pm [14.339 \pm 2.2281 \cdot 12.6336]$$

$$Y_0 \pm [-13.809 ; 42.487](95\%)$$

