Unidad 3 - Modulo 1

Pablo Anicich

6 de julio de 2016

Ejercicio adicional

Se desea estimar el costo total de la construcción de una vivienda suponiendo que tiene una fuerte relación con el tamaño del lote. Se toma una muestra aleatoria de 12 casas construidas el año pasado y la información recopilada se presenta en el cuadro siguiente:

Observation	Size	Cost
1	5.5	34.76
2	7.7	35.64
3	11.0	45.87
4	11.0	55.22
5	13.2	50.82
6	22.0	64.35
7	24.2	65.23
8	16.5	53.24
9	33.0	70.07
10	44.0	93.83
11	13.2	58.74
12	16.5	59.95

Se pide:

a. Indique cuál es la variable explicativa y cuál es la variable explicada.

b. Construya un diagrama de dispersión muestral.

c. Estime e interprete, en los términos de este problema, la ordenada al origen y la pendiente de la recta de regresión poblacional.

d. Calcule e interprete el coeficiente de determinación.

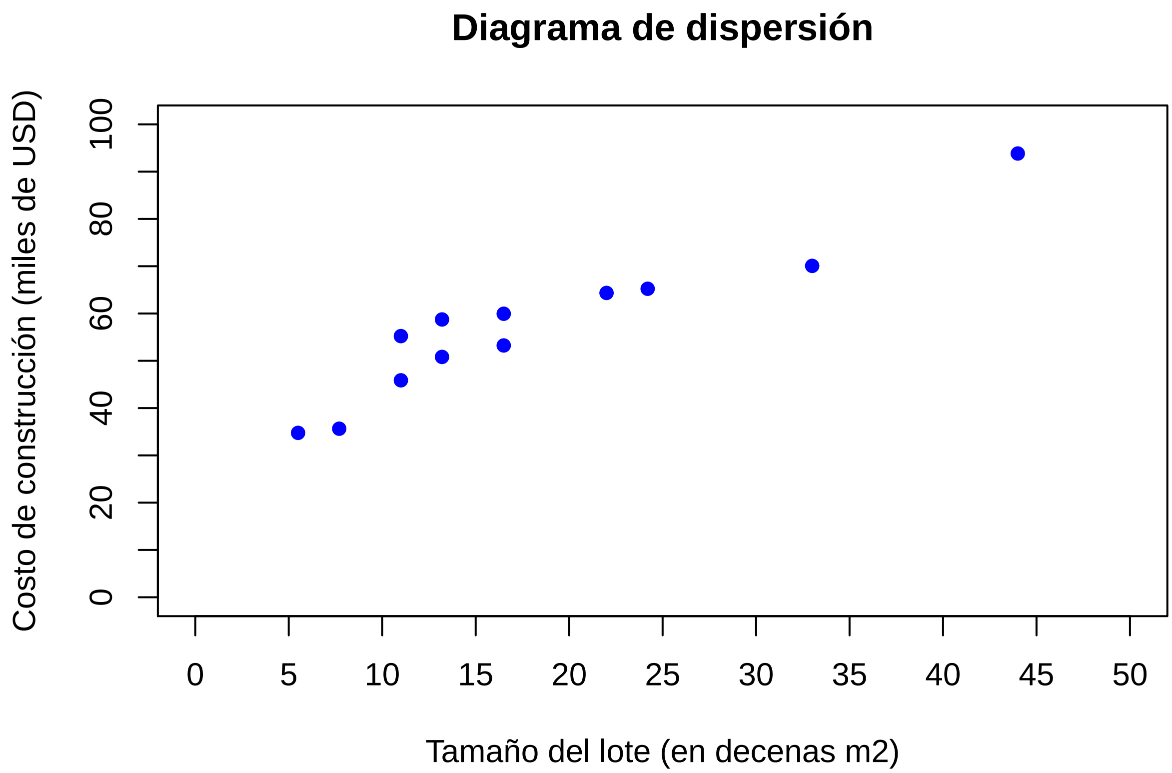
e. Estime, con una confianza del 95% el costo de construcción para un lote de 150 m².

f. Estime el coeficiente de correlación

Solución

a. La variable explicativa es el tamaño del lote a construir y la variable explicada es el costo de la construcción de una vivienda en tal lote.

b. Diagrama de dispersión.



c. Estimación de la recta de cuadrados mínimos $(y = ax + b)$. Primero estimo σ_{jl} usando que $\sigma_{jl} = \sum_{i=1}^{n} x_i^j y_i^l$:

$\sigma_{01} = 217.8.$

$\sigma_{10} = 687.72.$

$\sigma_{11} = 1.4326521 \times 10^4$

$\sigma_{20} = 5319.16.$

El sistema de ecuaciones resultante se puede escribir como:

$$\begin{pmatrix} \sigma_{20} & \sigma_{10} \\ \sigma_{10} & N \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} \sigma_{11} \\ \sigma_{01} \end{pmatrix}$$

cuya solución es:

$$a = \frac{\sigma_{10}\sigma_{01} - N\sigma_{11}}{\sigma_{10}^2 - N\sigma_{20}}$$

$$b = \frac{\sigma_{01} - a\sigma_{10}}{N}$$

de modo que se obtiene para nuestro caso que:

$$a = 1.350$$

$$b = 32.805$$

La recta de regresión estimada es: $\hat{y} = 1.350x_i + 32.805$.
La ordenada al origen b representa el costo fijo del lote. La pendiente a representa la variación de Y, cuando X varía en una unidad. En este problema representa el costo medio variable del costo de construcción. La parte del costo que varia por unidad de tamaño del lote.

d. El coeficiente de determinación es:

$$R^2 = \frac{\sum_{k=1}^{n}(x_i - x_{medio})^2 a^2}{\sum_{k=1}^{n}(y_i - y_{medio})^2} = 0.8924137$$

Expresado en porcentaje, el 89.3% de la variación del costo de construccion está explicada por la variación del tamaño del lote.

e. Para un lote de 150 m^2 se puede usar el resultado de la regresión lineal efectuada en puntos anteriores, para predecir el costo promedio de construcción:

$x = 15$ (recordar que el tamaño del lote se mide en decenas) corresponde a la estimación $\hat{y} = 53.055$.

Para determinar los valores con un intervalo de confianza de 95% primero debemos estimar la varianza residual muestral (VRM), que se define como:

$$VRM = \frac{\sum_{k=1}^{n}(y_i - \hat{y}_i)^2}{n-2} = 29.96596$$

La varianza estimada de un valor puntual de la recta es:

$$V[\hat{y}(x = x_0)] = VRM(1 + \frac{1}{n} + \frac{(x_0 - E(x))^2}{\sum_{k=1}^{n}(X_i - E(x))^2})$$

por lo que:

$$V[\hat{y}(x = 15] = 32.4632827$$

y

$$\sigma_{\hat{y}} = V[\hat{y}(x = 15]^{1/2} = 5.6976559$$

Para obtener el intervalo de confianza del 95%, al tratarse de un test de dos colas se debe evaluar el fractil de 97.5% de la distribución t-Student con $(n-2) = 10$ grados de libertad, es decir:

$$t = 2.2281389$$

Lo que permite estimar el intervalo buscado:

$$I_{\hat{y}}^{95} = (\hat{y} - \sigma_{\hat{y}} * t; \hat{y} + \sigma_{\hat{y}} * t) = (40.3598315, 65.7501685)$$

f. El coeficiente de correlación es 0.9447695.