

Regresión Lineal Simple

Paúl Arévalo

2024-07-04

Data Set

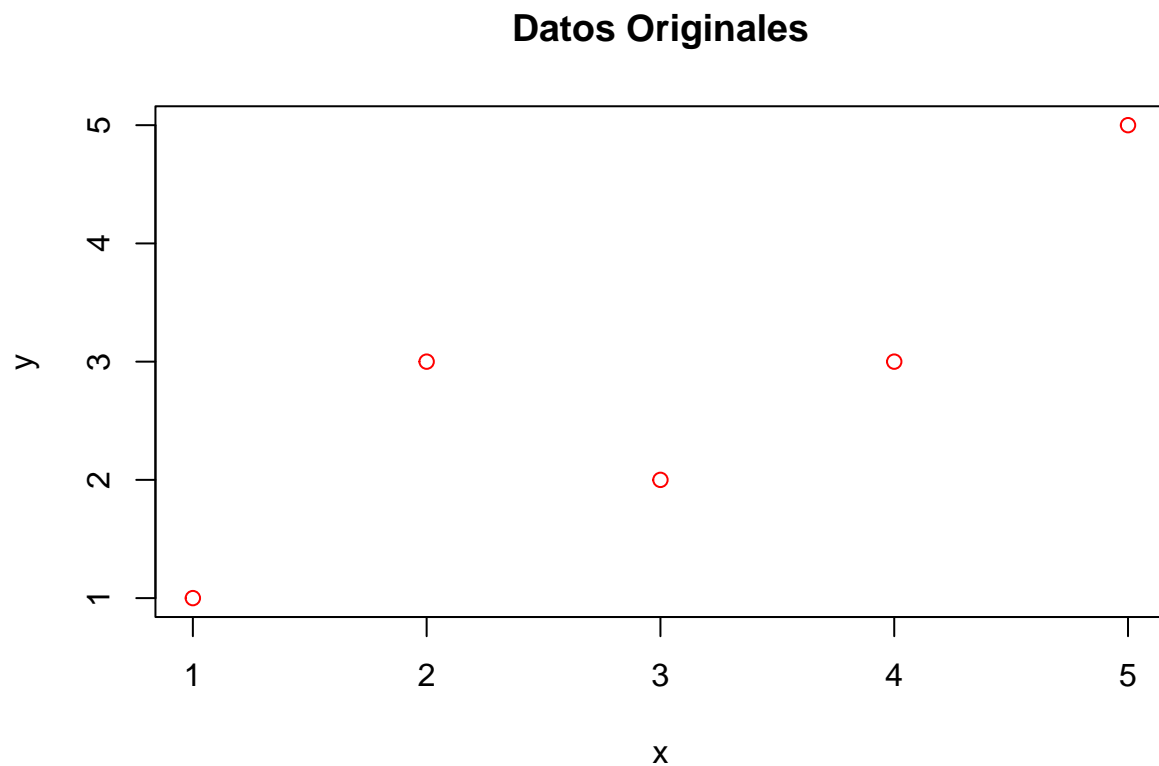
El siguiente conjunto de datos se utilizará en todo el proceso, en donde, x es la variable de entrada y y la variable de salida que se intentará predecir.

```
x <- c(1, 2, 4, 3, 5)
y <- c(1, 3, 3, 2, 5)
```

Gráfica inicial

A continuación se muestra en un gráfico los datos x versus y

```
plot(x, y, col = "red", main = "Datos Originales")
```



Regresión Lineal Simple

La fórmula para la regresión lineal se define como: $y = \beta_0 + \beta_1 x$ donde β_0 (intersección) y β_1 (pendiente) son los coeficientes a estimar.

La estimación de β_1 se define como:

La estimación de β_0 se define como:

Estimando la pendiente β_1

Primeramente se necesita calcular las medias de x y y

```
media_x <- mean(x)
media_y <- mean(y)
```

Ahora se calcula el error de cada variable con respecto a la media $x_i - \bar{x}$ y $y_i - \bar{y}$, se multiplican y se realiza la sumatoria. Además se calcula el denominador.

```
error_x <- x - media_x
error_x
```

```
## [1] -2 -1 1 0 2
```

```
error_y <- y - media_y
error_y
```

```
## [1] -1.8 0.2 0.2 -0.8 2.2
```

```
multiplicacion <- error_x * error_y
multiplicacion
```

```
## [1] 3.6 -0.2 0.2 0.0 4.4
```

```
numerador <- sum(multiplicacion)
numerador
```

```
## [1] 8
```

```
denominador <- sum(error_x^2)
denominador
```

```
## [1] 10
```

```
beta1 <- numerador / denominador
beta1
```

```
## [1] 0.8
```

Por lo tanto $\beta_1 = 8 / 10 = 0.8$

Estimando la intercepción β_0

Procedemos a calcular $\beta_0 = \bar{y} - \beta_1 \bar{x}$

```
beta0 <- media_y - beta1 * media_x
beta0
```

```
## [1] 0.4
```

Realizando predicciones

Calculados los coeficientes de la regresión lineal, la ecuación $Y = 0 + 1 x$ queda de la siguiente manera: $y = 0.4 + 0.8 x$, la predicción con los datos originales es:

```
x
```

```
## [1] 1 2 4 3 5
```

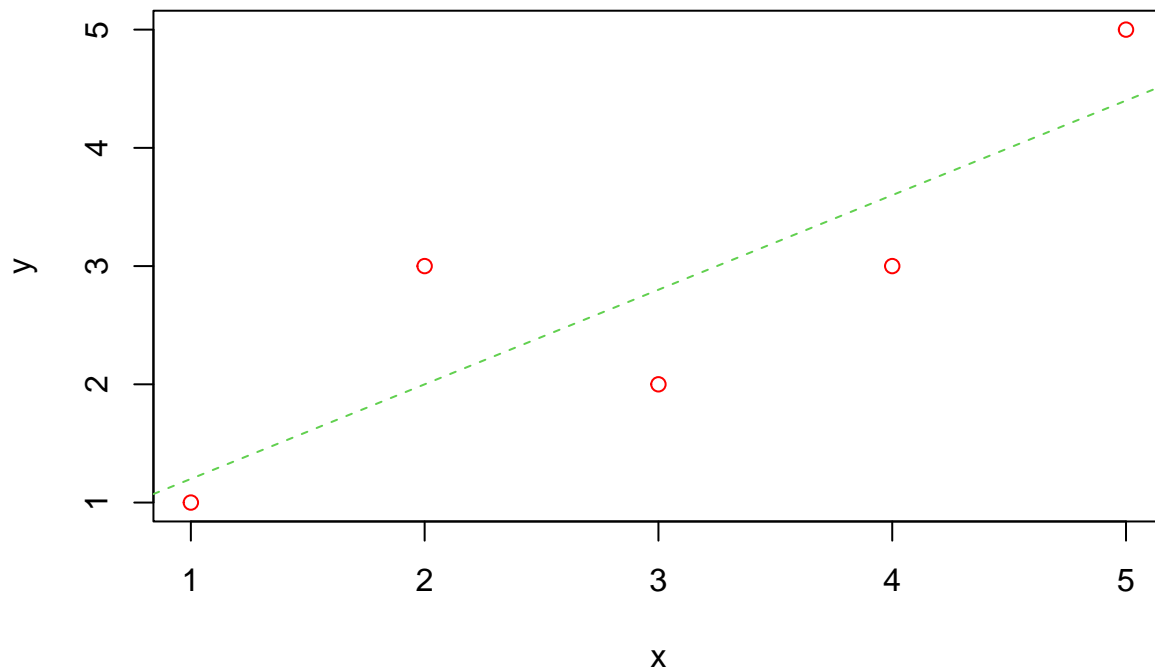
```
prediccion_y <- beta0 + beta1 * x  
prediccion_y
```

```
## [1] 1.2 2.0 3.6 2.8 4.4
```

A continuación una gráfica de los datos originales y la recta de regresión

```
plot(x, y, col = "red", main = "Datos Originales/Regresión")  
abline(beta0, beta1, lty = 2, col = 3)
```

Datos Originales/Regresión



Estimación del Error

Se puede calcular el error de las predicciones a través de el error cuadrático medio (Root Mean Squared Error - RMSE), definido como:

```
prediccion_y
```

```
## [1] 1.2 2.0 3.6 2.8 4.4
```

```
y
```

```
## [1] 1 3 3 2 5
```

```
error_prediccion <- prediccion_y - y
error_prediccion
```

```
## [1]  0.2 -1.0  0.6  0.8 -0.6
```

```
squared_error <- error_prediccion^2
squared_error
```

```
## [1] 0.04 1.00 0.36 0.64 0.36
```

```
suma_error <- sum(squared_error)
suma_error
```

```
## [1] 2.4
```

```
RMSE <- sqrt(suma_error / length(x))
RMSE
```

```
## [1] 0.6928203
```

Por lo tanto, cada predicción en promedio es incorrecta en aproximadamente 0.6928203