**Ejercicio Propuesto1**

**La siguiente tabla recoge información sobre el diámetro (en pulgadas), la altura (en pies) y el volumen (en pies cúbicos) del tronco de distintos cerezos en una determinada región**

|  |  |  |
| --- | --- | --- |
| **Diámetro** | **Altura** | **Volumen** |
| 8.3 | 70 | 10.3 |
| 8.6 | 65 | 10.3 |
| 8.8 | 63 | 10.2 |
| 10.5 | 72 | 16.4 |
| 10.5 | 81 | 18.8 |
| 10.8 | 83 | 19.7 |
| 11 | 66 | 15.6 |
| 11 | 75 | 16.3 |

[Tabla4. Datos del Ejemplo Propuesto 1](https://wpd.ugr.es/~bioestad/wp-content/uploads/datos4_Propu-1.txt)

**Se pide:**

**a) Crear un data frame de nombre Cerezos que almacene los datos de los árboles**

**b) Dibujar el diagrama de dispersión de las variables altura y volumen y determinar si puede existir una cierta relación lineal entre ambas**

**c) ¿Cuál es la recta de regresión lineal simple que considera a la altura como variable dependiente y al volumen como variable independiente? Interpreta los parámetros de esa recta**

**d) ¿Son significativos estos parámetros? ¿Qué puede decirse del ajuste del modelo a los datos?**

**e) ¿Cuál es la correlación lineal de Pearson entre ambas variables? ¿Es significativa?**

**f) ¿Cuál es la recta de regresión lineal si se considera también como variable independiente el diámetro?**

**g) Calcular los valores de las estimaciones** $\hat{β}\_{0}, \hat{β}\_{1} y \hat{β}\_{2}$ en el siguiente modelo de regresión:
$alt\_{i}= \hat{β}\_{0}+\hat{β}\_{1}vol\_{i} + \hat{β}\_{2}diam\_{i}^{2}$?

**Solución:**

1. **Crear un data frame de nombre Cerezos que almacene los datos de los árboles**

**> diam <- c(8.3, 8.6, 8.8, 10.5, 10.5, 10.8, 11, 11)**

**> alt <- c(70, 65, 63, 72, 81, 83, 66, 75)**

**> vol <- c(10.3, 10.3, 10.2, 16.4, 18.8, 19.7, 15.6, 16.3)**

**> Cerezos <- data.frame(diam, alt, vol)**

**> Cerezos**

 diam alt vol

1 8.3 70 10.3

2 8.6 65 10.3

3 8.8 63 10.2

4 10.5 72 16.4

5 10.5 81 18.8

6 10.8 83 19.7

7 11.0 66 15.6

8 11.0 75 16.3

**b) Dibujar el diagrama de dispersión de las variables altura y volumen y determinar si puede existir una cierta relación lineal entre ambas**

**c) ¿Cuál es la recta de regresión lineal simple que considera a la altura como variable dependiente y al volumen como variable independiente? Interpreta los parámetros de esa recta**

**d) ¿Son significativos estos parámetros? ¿Qué puede decirse del ajuste del modelo a los datos?**

Para resolver estos tres apartados, primero cargamos el paquete BrailleR

> library("BrailleR")

> OnePredictor(alt, vol)

**The term which is significant to 1% is**

**vol with an estimate of 1.601271 and P-Value of 0.007679516**

**Call:**

**lm(formula = alt ~ vol, data = alt.vol)**

**Residuals:**

 **Min 1Q Median 3Q Max**

**-7.3161 -1.9013 0.3668 2.6995 5.1706**

**Coefficients:**

 **Estimate Std. Error t value Pr(>|t|)**

**(Intercept) 48.3363 6.1662 7.839 0.000228 \*\*\***

**vol 1.6013 0.4071 3.934 0.007680 \*\***

**---**

**Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1**

**Residual standard error: 4.211 on 6 degrees of freedom**

**Multiple R-squared: 0.7206, Adjusted R-squared: 0.674**

**F-statistic: 15.47 on 1 and 6 DF, p-value: 0.00768**

La recta de regresión lineal simple es $alt\_{i}=48.3363+1.6013 vol\_{i}$. Los dos parámetros de la recta se interpretan del siguiente modo: 48.336 es la altura esperada para un cerezo que tiene un volumen de 0 pies cúbicos. Por otra parte, por cada pie cúbico de incremento en el volumen del tronco del cerezo, se espera un aumento en su altura de 1.601 pies.

Considerando un nivel de significación del 5%, ambos parámetros son significativamente distintos de 0, ya que los p-valores asociados a los contrastes t de Student de los dos parámetros son inferiores a 0.05.

El ajuste del modelo a los datos es aceptable, ya que el valor de R2 es de 0.7206.

**e) ¿Cuál es la correlación lineal de Pearson entre ambas variables? ¿Es significativa?**

> cor(alt, vol)

[1] 0.8488805

> cor.test(alt, vol)

 Pearson's product-moment correlation

data: alt and vol

t = 3.9338, df = 6, p-value = 0.00768

alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval:

 0.3588886 0.9720747

sample estimates:

 cor

0.8488805

Con un p-valor de 0.00768 < 0.05, rechazamos la hipótesis de que el coeficiente de correlación lineal de Pearson entre ambas variables pueda considerarse 0.

**f) ¿Cuál es la recta de regresión lineal si se considera también como variable independiente el diámetro?**

> reg\_lin\_mul<- lm(alt ~ vol + diam)

> reg\_lin\_mul

Call:

lm(formula = alt ~ vol + diam)

Coefficients:

(Intercept) vol diam

 85.340 3.231 -6.135

En este caso, la recta de regresión lineal múltiple es $alt\_{i}=85.340+3.231vol\_{i}-6.135diam\_{i}$

**g) Calcular los valores de las estimaciones** $\hat{β}\_{0}, \hat{β}\_{1} y \hat{β}\_{2}$ en el siguiente modelo de regresión:
$alt\_{i}= \hat{β}\_{0}+\hat{β}\_{1}vol\_{i} + \hat{β}\_{2}diam\_{i}^{2}$?

> diam2 <- diam^2

> reg\_cuad <- lm(alt ~ vol + diam2)

> reg\_cuad

Call:

lm(formula = alt ~ vol + diam2)

Coefficients:

(Intercept) vol diam2

 55.830 3.158 -0.304

En este caso, el modelo de regresión cuadrático puede escribirse del siguiente modo:

$$alt\_{i}=55.830+3.158vol\_{i}-0.304diam\_{i}^{2}$$