Regresion-lineal.R

Usuario

2024-10-28

#María Fernanda González Rodríguez   
#2093400  
#28/10/2024   
  
#Regresión Lineal  
  
erupciones <- read.csv("C:/Repositorios/Met\_Est\_2024/Clase/erupciones.csv", header = T)  
  
head(erupciones)

## eruptions waiting  
## 1 3.600 79  
## 2 1.800 54  
## 3 3.333 74  
## 4 2.283 62  
## 5 4.533 85  
## 6 2.883 55

plot(erupciones$waiting, erupciones$eruptions, pch = 19, col = "lavender", xlab = "Tiempo espera (min)", ylab = "Duración (min)")  
  
range(erupciones$eruptions)

## [1] 1.6 5.1

range(erupciones$waiting)

## [1] 43 96

#Revisar si existe una relación lineal entre dos variables   
cor.test(erupciones$eruptions, erupciones$waiting)

##   
## Pearson's product-moment correlation  
##   
## data: erupciones$eruptions and erupciones$waiting  
## t = 34.089, df = 270, p-value < 2.2e-16  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## 0.8756964 0.9210652  
## sample estimates:  
## cor   
## 0.9008112

cor.test(erupciones$waiting, erupciones$eruptions)

##   
## Pearson's product-moment correlation  
##   
## data: erupciones$waiting and erupciones$eruptions  
## t = 34.089, df = 270, p-value < 2.2e-16  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## 0.8756964 0.9210652  
## sample estimates:  
## cor   
## 0.9008112

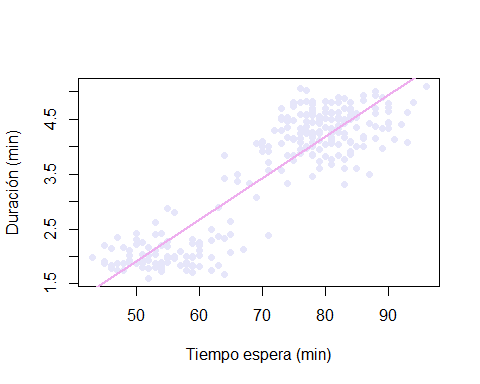
#La pregunta: Podemos predecir el tiempo de la erupción en base al tiempo de espera entre dos erupciones consecutivas.  
lm(erupciones$eruptions ~ erupciones$waiting)

##   
## Call:  
## lm(formula = erupciones$eruptions ~ erupciones$waiting)  
##   
## Coefficients:  
## (Intercept) erupciones$waiting   
## -1.87402 0.07563

gy.lm <- lm(erupciones$eruptions~ erupciones$waiting)  
  
summary(gy.lm)

##   
## Call:  
## lm(formula = erupciones$eruptions ~ erupciones$waiting)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.29917 -0.37689 0.03508 0.34909 1.19329   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -1.874016 0.160143 -11.70 <2e-16 \*\*\*  
## erupciones$waiting 0.075628 0.002219 34.09 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.4965 on 270 degrees of freedom  
## Multiple R-squared: 0.8115, Adjusted R-squared: 0.8108   
## F-statistic: 1162 on 1 and 270 DF, p-value: < 2.2e-16

#Graficar línea de tendencia central  
plot(erupciones$waiting, erupciones$eruptions, pch = 19, col = "lavender", xlab = "Tiempo espera (min)", ylab = "Duración (min)")  
abline(gy.lm, col = "plum2", lwd = 2)



#agregar una tabla el valor predicho   
  
erupciones$predicho <- gy.lm$fitted.values  
erupciones$yprima <- -1.874016 + 0.075628 \* erupciones$waiting  
erupciones$residual <- gy.lm$residuals  
erupciones$residual <- gy.lm$residuals^2  
  
  
yprima <- c(70, 75, 82)  
-1.87+0.075 \* yprima

## [1] 3.380 3.755 4.280