

Clase_8.R

Usuario

2024-12-02

```
#Clase 8
#Alejandra Elizondo Trejo
#1847945
#28/10/2024

#Regresión lineal
geyser<- read.csv("C:/Repositorios/Met_Est_2024/Clase/erupciones.csv",
                  header = T)
#primero se identifica cual es la independiente y la dependiente, despues
se
#coloca primero la independiente (x) y despues la dependiente (y)
plot(geyser$waiting, geyser$eruptions,
     xlab="Tiempo de espera (min)",
     ylab= "Duración (min)",
     col="blue")

#Rango de duración de erupciones en minutos
range(geyser$eruptions)

## [1] 1.6 5.1

#Rango de duración de espera entre erupción-erupción en minutos
range(geyser$waiting)

## [1] 43 96

#hay que saber si existe una relación lineal entre los 2 valores
cor.test(geyser$eruptions,geyser$waiting)

##
## Pearson's product-moment correlation
##
## data: geyser$eruptions and geyser$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

#se obtuvo un valor de 0.9, casi 1, es un valor significativo y positivo
```

```

#Nuestra pregunta es ¿podemos predecir el tiempo de la erupción en base
al
#tiempo de espera entre dos erupciones consecutivas?

#REQUISITOS PARA PODER APLICAR LA REGRESIÓN LINEAL
#1 para cada observación en x hay que tener una en y, si no tengo en
alguna de las 2 variables no
#podre realizar esto
#2 Los datos son idnependientes entre ellos, pueden estar relacionados
pero no dependientes
#3 que tenga una correlacion siginficativa para que se pueda obtener una
linealidad
#4 Que Los datos tengan una distribución normal

#para colocar "~" es AltGr + La tecla donde esta
#primero va la variable y y despues x, es para obtener la línea de
tendencia central
lm(geyser$eruptions~ geyser$waiting)

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

#guardamos Los datos en un objeto
gy.lm <- lm(geyser$eruptions~ geyser$waiting)

#Para ver Los datos de lm usar "summary"
summary(gy.lm)

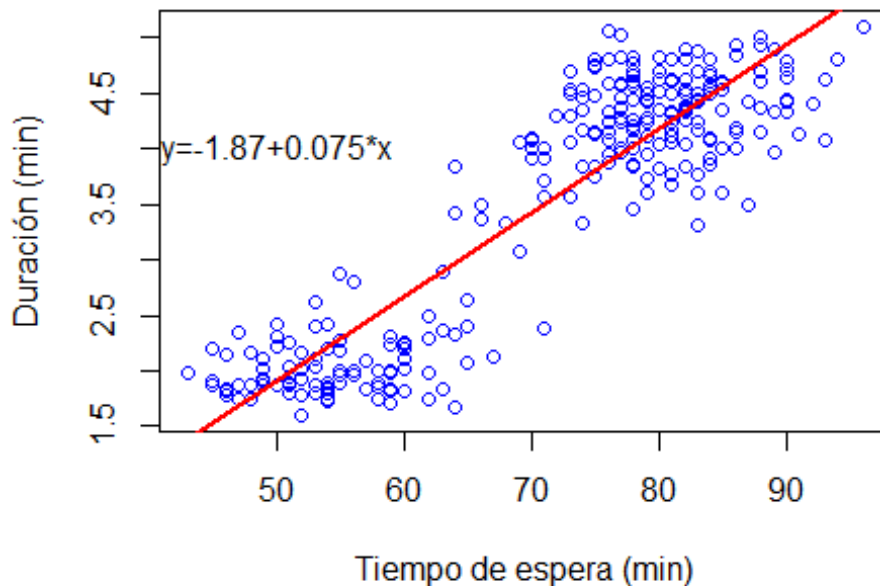
##
## Call:
## lm(formula = geyser$eruptions ~ geyser$waiting)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.29917 -0.37689  0.03508  0.34909  1.19329
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -1.874016   0.160143  -11.70  <2e-16 ***
## geyser$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

#Graficaas la línea de tendencia central
plot(geyser$waiting, geyser$eruptions,
     xlab="Tiempo de espera (min)",
     ylab= "Duración (min)",
     col="blue")
abline(gy.lm, col = "red", lwd = 2)

text (50,4, "y=-1.87+0.075*x")
```



```
#Agregar en la tabla el valor predicho
#"eruptions" son los valores observado y "predicho" son los valores
predichos
geyser$predicho <- gy.lm$fitted.values

geyser$yprima <- -1.874016 + 0.075628 * geyser$waiting
# la fórmula de arriba es la fórmula de "y= alpha + B (x)"

geyser$residual <- gy.lm$residual^2

#ahora para calcular la duración del geyser dependiendo de el tiempo de
espera
#de 70 min, 75 min y 82 min
```

```
yprima <- c(70,75,82)
-1.87 + 0.075* yprima
## [1] 3.380 3.755 4.280
```