

# Clase-4.R

adela

2019-08-09

```
#Adela Garcia
#09/08/2019
#Clase 4

# Correlación -----

library(repmis)

## Registered S3 method overwritten by 'R.oo':
##   method      from
##   throw.default R.methodsS3

erupciones <- read.csv("C:/MCF202-2019/Datos/erupciones.csv", header = T)

plot(erupciones$waiting , erupciones$eruptions, pch=19,
     col = "lightgreen",
     xlab = "Tiempo de espera (min)",
     ylab = "Duracion (min)")

## A mayor tiempo de espera, mayor sera la erupción
## se observa que no hay un origen muy claro de las erupciones

library("pastecs")
stat.desc(erupciones$eruptions,basic=FALSE, norm=TRUE)

##          median          mean      SE.mean  CI.mean.0.95          var
## 4.000000e+00  3.487783e+00  6.920580e-02  1.362494e-01  1.302728e+00
##      std.dev      coef.var      skewness      skew.2SE      kurtosis
## 1.141371e+00  3.272483e-01 -4.135498e-01 -1.399854e+00 -1.511605e+00
##      kurt.2SE  normtest.W  normtest.p
## -2.567516e+00  8.459156e-01  9.036119e-16

shapiro.test(erupciones$eruptions)

##
##  Shapiro-Wilk normality test
##
## data:  erupciones$eruptions
## W = 0.84592, p-value = 9.036e-16

shapiro.test(erupciones$waiting)

##
##  Shapiro-Wilk normality test
##
## data:  erupciones$waiting
## W = 0.92215, p-value = 1.015e-10
```

```

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

## Ho no es significativa
## H1 es significativa la coorelacion
## Los datos muestran una coorrelacion significativa entre las variables.

# Regresión lineal -----

## Hipotesis = La variable tiempo de espera ayuda a predecir la duracion del geiser
## Ho no es significativa la prediccion
## H1 es significativa la prediccion

## lm "lineal model"
lm.erup <- lm(erupciones$eruptions ~ erupciones$waiting)

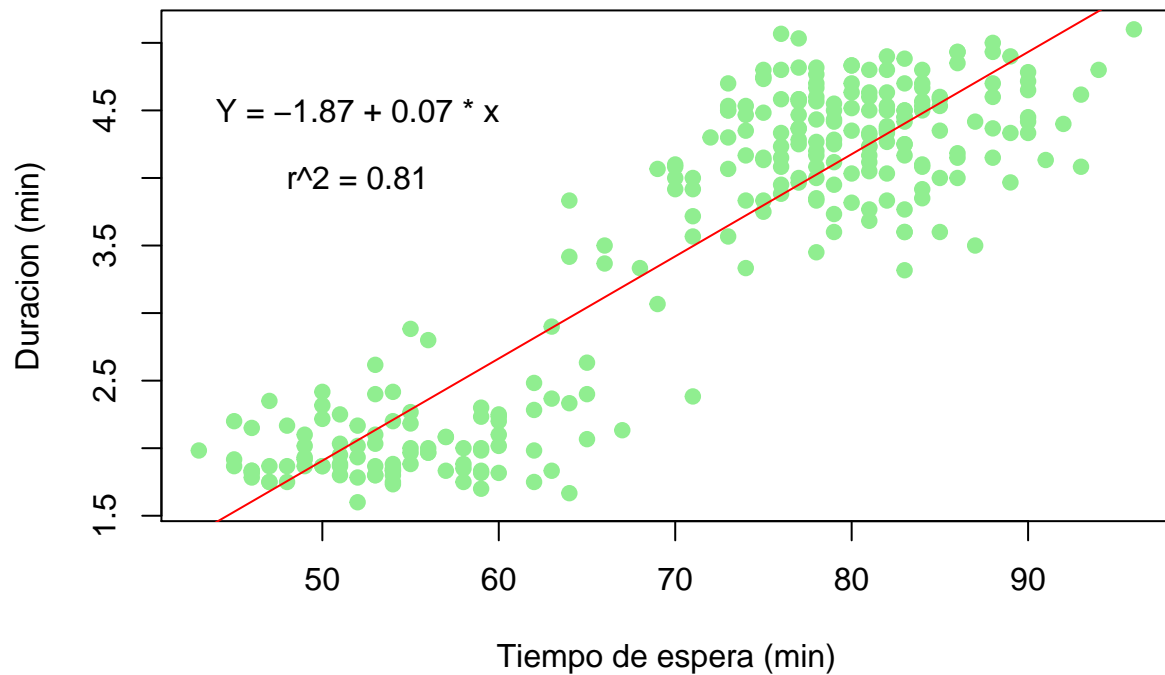
plot(erupciones$waiting , erupciones$eruptions, pch=19,
     col = "lightgreen",
     xlab = "Tiempo de espera (min)",
     ylab = "Duracion (min)")
abline(lm.erup, col = "red")

## PARA LA FORMULA LM
## X = waitin, variable independiente
## Y = eruptions, variable dependiente

## PARA LA GRÁFICA
## Y = eruptions, variable dependiente
## X = waitin, variable independiente

## Coeficientes de regresión
text(52, 4.5, "Y = -1.87 + 0.07 * x")
text(52, 4, "r^2 = 0.81")

```



```
## El tiempo de erupcion depende del tiempo de espera
```

```
## Valores de la intercepcion de alfa y beta
lm.erup
```

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

```
summary(lm.erup)
```

```
##
## 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
## coeficiente alfa = -1.87

length(erupciones$eruptions)

## [1] 272

sqrt(0.90)

## [1] 0.9486833

(0.90)^2

## [1] 0.81
y.60 <- -1.87 + 0.07*60
y.60

## [1] 2.33
# Datos regresión -----

espera <- erupciones$waiting
duracion <- erupciones$eruptions
res <- resid(lm.erup)
pre <- fitted(lm.erup)
res.2 <- res^2

cuadro <- round(data.frame(espera, duracion, pre, res,
                           res.2), 4)
SSE <- sum(cuadro$res.2)
SSE

## [1] 66.5612

vari <- SSE/(length(erupciones$waiting)-2)
vari

## [1] 0.246523

var(duracion)

## [1] 1.302728
# Prueba de hipotesis de la regresión -----

an.erup <- anova(lm.erup)
an.erup

## Analysis of Variance Table
##
## Response: erupciones$eruptions
##           Df Sum Sq Mean Sq F value    Pr(>F)
## erupciones$waiting  1 286.478  286.478  1162.1 < 2.2e-16 ***
## Residuals          270   66.562    0.247

```

```
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
## Se acepta la hipotesis alternativa, es decir, el modelo de regresion aplicado para  
## los datos de erupciones del volcan son diferentes a 0, por lo tanto es significativo.
```