

Analisis de regresion

Ronald Bailey

2023-05-28

Laboratorio #3: Modelos de Regresión en R

Ejercicio #1: utilizando R realice una función que dado un dataframe cualquiera de dos columnas, donde la primera (índice 1) sea el valor de la variable independiente (X) y la segunda sea el valor de una variable dependiente (Y), devuelva una lista con los siguientes elementos:

1. Un arreglo con los valores de los estimadores para B0 y B1”.
2. El valor del coeficiente de determinación r^2 del modelo.
3. El coeficiente de correlación r (raíz cuadrada de r^2).
4. Un arreglo con los valores de los residuos.
5. Una gráfica con la nube de puntos y la recta de regresión del modelo.

Nota: Para este ejercicio NO está permitido utilizar la función `lm()` para calcular ninguno de los elementos solicitados (incisos 1 al 4), sin embargo puede utilizar `ggplot` para realizar la gráfica del inciso 5

La funcion lineal:

$$\bar{y} = \hat{\beta}_1 x + \hat{\beta}_0$$

Recuerde de su curso de Econometria que:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{(\sum_{i=1}^n x_i^2) - n \bar{x}^2}$$

x = valores conocido de la variable independiente

y = valores conocido de la variable dependiente

\bar{x} = promedio de valores de las x

\bar{y} = promedio de valores de las y

n = numero de datos puntuales u observaciones

i = índice de la valuacion

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

$$r^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

$$r = \sqrt{r^2}$$

Recuerde que n representa la cantidad de filas en el dataset

Leer el dataset

```
dataset = read.csv("grados.csv")
```

```
head(dataset)
```

```
##   Centígrados Farenheit
## 1           0       32.0
## 2           1       33.8
## 3           2       35.6
## 4           3       37.4
## 5           4       39.2
## 6           5       41.0
```

```
summary(dataset)
```

```
##   Centígrados      Farenheit
## Min.   : 0.0   Min.   : 32.0
## 1st Qu.:249.8   1st Qu.: 481.6
## Median :499.5   Median : 931.1
## Mean   :499.5   Mean   : 931.1
## 3rd Qu.:749.2   3rd Qu.:1380.7
## Max.   :999.0   Max.   :1830.2
```

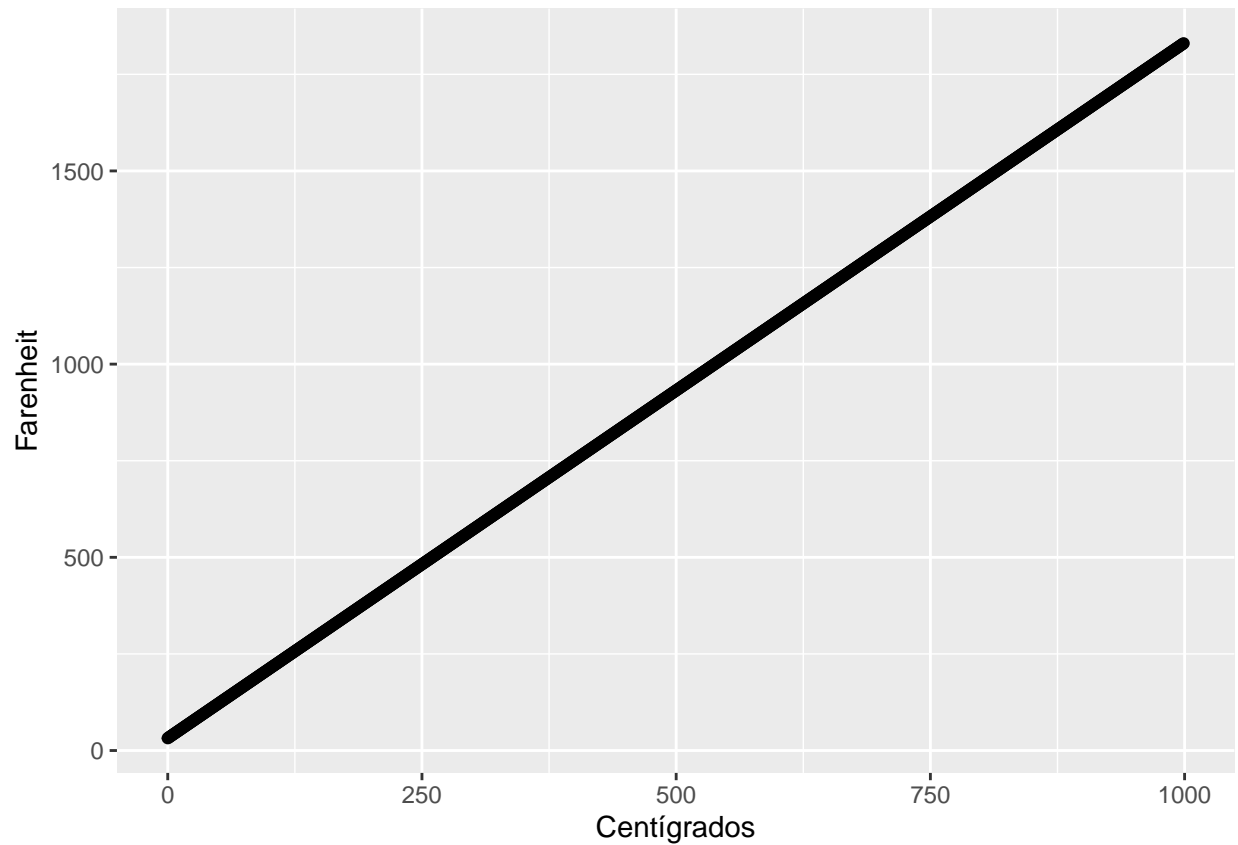
```
# Cargar la biblioteca necesaria
```

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.2.3
```

```
# generar gráfico
```

```
ggplot(dataset, aes(x=Centígrados, y=Farenheit)) +  
  geom_point()
```



```
# Crear algunos datos de ejemplo

x <- dataset$Centígrados
y <- dataset$Farenheit

# Calcular las medias
x_bar <- mean(x)
y_bar <- mean(y)

# Calcular las desviaciones
x_dev <- x - x_bar
y_dev <- y - y_bar

# Calcular los coeficientes de regresión
beta_1 <- sum(x_dev * y_dev) / sum(x_dev^2)
beta_0 <- y_bar - beta_1 * x_bar

# Crear un arreglo con los valores de los estimadores para B0 y B1
estimadores <- c(beta_0, beta_1)

# Calcular los valores predichos
y_predicha <- beta_0 + beta_1 * x

# Calcular los residuos
residuos <- y - y_predicha
```

```

# Calcular R cuadrado
r_cuadrado <- 1 - sum(residuos^2) / sum((y - y_bar)^2)

# Calcular el coeficiente de correlación r
r <- sqrt(r_cuadrado)

# Crear una lista con los resultados
resultados <- list(estimadores = estimadores,
                  r_cuadrado = r_cuadrado,
                  r = r,
                  residuos = residuos)

# Imprimir los resultados
print(resultados)

```

```

## $estimadores
## [1] 32.0 1.8
##
## $r_cuadrado
## [1] 1
##
## $r
## [1] 1
##
## $residuos
##      [1] 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00
##      [6] 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00
##     [11] 0.000000e+00 0.000000e+00 0.000000e+00 -7.105427e-15 0.000000e+00
##     [16] 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00
##     [21] 0.000000e+00 -1.421085e-14 0.000000e+00 0.000000e+00 0.000000e+00
##     [26] 0.000000e+00 -1.421085e-14 0.000000e+00 0.000000e+00 0.000000e+00
##     [31] 0.000000e+00 -1.421085e-14 0.000000e+00 0.000000e+00 0.000000e+00
##     [36] 0.000000e+00 0.000000e+00 -1.421085e-14 0.000000e+00 0.000000e+00
##     [41] 0.000000e+00 0.000000e+00 -1.421085e-14 0.000000e+00 0.000000e+00
##     [46] 0.000000e+00 0.000000e+00 -1.421085e-14 0.000000e+00 0.000000e+00
##     [51] 0.000000e+00 0.000000e+00 -1.421085e-14 0.000000e+00 0.000000e+00
##     [56] 0.000000e+00 0.000000e+00 -2.842171e-14 0.000000e+00 0.000000e+00
##     [61] 0.000000e+00 0.000000e+00 -2.842171e-14 0.000000e+00 0.000000e+00
##     [66] 0.000000e+00 0.000000e+00 -2.842171e-14 0.000000e+00 0.000000e+00
##     [71] 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 -2.842171e-14
##     [76] 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 -2.842171e-14
##     [81] 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 -2.842171e-14
##     [86] 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 -2.842171e-14
##     [91] 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 -2.842171e-14
##     [96] 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 -2.842171e-14
##    [101] 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 -2.842171e-14
##    [106] 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 -2.842171e-14
##    [111] 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 -2.842171e-14
##    [116] 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 -2.842171e-14
##    [121] 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 -2.842171e-14
##    [126] 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 -5.684342e-14
##    [131] 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 -5.684342e-14
##    [136] 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 -5.684342e-14

```

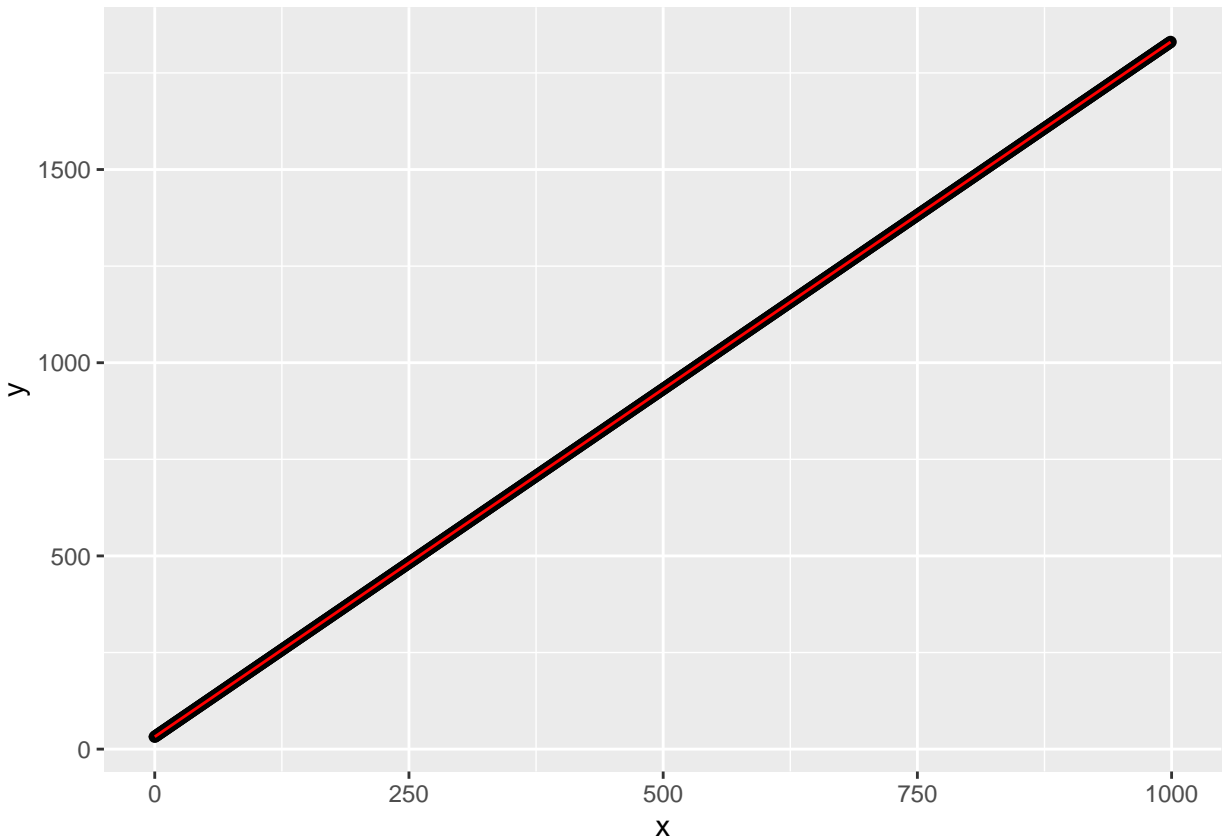
[illegible]

[illegible]

[illegible]

```
## [951] 0.000000e+00 0.000000e+00 -2.273737e-13 0.000000e+00 0.000000e+00
## [956] 0.000000e+00 0.000000e+00 -2.273737e-13 0.000000e+00 0.000000e+00
## [961] 0.000000e+00 0.000000e+00 -2.273737e-13 0.000000e+00 0.000000e+00
## [966] 0.000000e+00 0.000000e+00 -2.273737e-13 0.000000e+00 0.000000e+00
## [971] 0.000000e+00 0.000000e+00 -2.273737e-13 0.000000e+00 0.000000e+00
## [976] 0.000000e+00 0.000000e+00 -2.273737e-13 0.000000e+00 0.000000e+00
## [981] 0.000000e+00 0.000000e+00 -2.273737e-13 0.000000e+00 0.000000e+00
## [986] 0.000000e+00 0.000000e+00 -2.273737e-13 0.000000e+00 0.000000e+00
## [991] 0.000000e+00 0.000000e+00 -2.273737e-13 0.000000e+00 0.000000e+00
## [996] 0.000000e+00 0.000000e+00 -2.273737e-13 0.000000e+00 0.000000e+00
```

```
# Crear el gráfico
library(ggplot2)
dataset <- data.frame(x = x, y = y, y_predicha = y_predicha)
grafico <- ggplot(dataset, aes(x = x, y = y)) +
  geom_point() +
  geom_line(aes(x = x, y = y_predicha), color = "red")
grafico
```



```
head(data.frame(B0 = beta_0,
                B1 = beta_1,
                R_Cuadrado = r_cuadrado,
                R = r,
                Residuos = list(residuos)) )
```

```
##   B0  B1 R_Cuadrado R
```



```
## 1 32 1.8      1 1
## 2 32 1.8      1 1
## 3 32 1.8      1 1
## 4 32 1.8      1 1
## 5 32 1.8      1 1
## 6 32 1.8      1 1
## c.0..0..0..0..0..0..0..0..0..0..0..0..0..0..0...7.105427357601e.15..
## 1                                                    0
## 2                                                    0
## 3                                                    0
## 4                                                    0
## 5                                                    0
## 6                                                    0
```

```
summary(data.frame(B0 = beta_0,
                   B1 = beta_1,
                   R_Cuadrado = r_cuadrado,
                   R = r,
                   Residuos = list(residuos)) )
```

```
##      B0      B1      R_Cuadrado      R
## Min.   :32   Min.   :1.8   Min.    :1   Min.    :1
## 1st Qu.:32   1st Qu.:1.8   1st Qu.:1   1st Qu.:1
## Median :32   Median :1.8   Median :1   Median :1
## Mean   :32   Mean    :1.8   Mean    :1   Mean    :1
## 3rd Qu.:32   3rd Qu.:1.8   3rd Qu.:1   3rd Qu.:1
## Max.   :32   Max.    :1.8   Max.    :1   Max.    :1
## c.0..0..0..0..0..0..0..0..0..0..0..0..0..0..0...7.105427357601e.15..
## Min.    :-2.274e-13
## 1st Qu.: 0.000e+00
## Median : 0.000e+00
## Mean    :-2.893e-14
## 3rd Qu.: 0.000e+00
## Max.    : 0.000e+00
```

Resumen

1. Un arreglo con los valores de los estimadores para B0 y B1”.

```
arreglo = c(beta_0, beta_1)
arreglo
```

```
## [1] 32.0 1.8
```

2. El valor del coeficiente de determinación r^2 del modelo.

```
r_cuadrado
```

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

3. El coeficiente de correlación r (raíz cuadrada de r^2).

```
r
```

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

4. Un arreglo con los valores de los residuos.

```
residuos
```

```
## [1] 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00
## [6] 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00
## [11] 0.000000e+00 0.000000e+00 0.000000e+00 -7.105427e-15 0.000000e+00
## [16] 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00
## [21] 0.000000e+00 -1.421085e-14 0.000000e+00 0.000000e+00 0.000000e+00
## [26] 0.000000e+00 -1.421085e-14 0.000000e+00 0.000000e+00 0.000000e+00
## [31] 0.000000e+00 -1.421085e-14 0.000000e+00 0.000000e+00 0.000000e+00
## [36] 0.000000e+00 0.000000e+00 -1.421085e-14 0.000000e+00 0.000000e+00
## [41] 0.000000e+00 0.000000e+00 -1.421085e-14 0.000000e+00 0.000000e+00
## [46] 0.000000e+00 0.000000e+00 -1.421085e-14 0.000000e+00 0.000000e+00
## [51] 0.000000e+00 0.000000e+00 -1.421085e-14 0.000000e+00 0.000000e+00
## [56] 0.000000e+00 0.000000e+00 -2.842171e-14 0.000000e+00 0.000000e+00
## [61] 0.000000e+00 0.000000e+00 -2.842171e-14 0.000000e+00 0.000000e+00
## [66] 0.000000e+00 0.000000e+00 -2.842171e-14 0.000000e+00 0.000000e+00
## [71] 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 -2.842171e-14
## [76] 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 -2.842171e-14
## [81] 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 -2.842171e-14
## [86] 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 -2.842171e-14
## [91] 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 -2.842171e-14
## [96] 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 -2.842171e-14
## [101] 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 -2.842171e-14
## [106] 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 -2.842171e-14
## [111] 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 -2.842171e-14
## [116] 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 -2.842171e-14
## [121] 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 -2.842171e-14
## [126] 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 -5.684342e-14
## [131] 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 -5.684342e-14
## [136] 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 -5.684342e-14
## [141] 0.000000e+00 0.000000e+00 0.000000e+00 -5.684342e-14 0.000000e+00
## [146] 0.000000e+00 0.000000e+00 0.000000e+00 -5.684342e-14 0.000000e+00
## [151] 0.000000e+00 0.000000e+00 0.000000e+00 -5.684342e-14 0.000000e+00
## [156] 0.000000e+00 0.000000e+00 0.000000e+00 -5.684342e-14 0.000000e+00
## [161] 0.000000e+00 0.000000e+00 0.000000e+00 -5.684342e-14 0.000000e+00
## [166] 0.000000e+00 0.000000e+00 0.000000e+00 -5.684342e-14 0.000000e+00
## [171] 0.000000e+00 0.000000e+00 0.000000e+00 -5.684342e-14 0.000000e+00
## [176] 0.000000e+00 0.000000e+00 0.000000e+00 -5.684342e-14 0.000000e+00
## [181] 0.000000e+00 0.000000e+00 0.000000e+00 -5.684342e-14 0.000000e+00
## [186] 0.000000e+00 0.000000e+00 0.000000e+00 -5.684342e-14 0.000000e+00
## [191] 0.000000e+00 0.000000e+00 0.000000e+00 -5.684342e-14 0.000000e+00
## [196] 0.000000e+00 0.000000e+00 0.000000e+00 -5.684342e-14 0.000000e+00
## [201] 0.000000e+00 0.000000e+00 0.000000e+00 -5.684342e-14 0.000000e+00
## [206] 0.000000e+00 0.000000e+00 0.000000e+00 -5.684342e-14 0.000000e+00
## [211] 0.000000e+00 0.000000e+00 0.000000e+00 -5.684342e-14 0.000000e+00
```

[illegible]

[illegible]

[illegible]

5. Una gráfica con la nube de puntos y la recta de regresión del modelo.

grafico

