

Práctica 6

# Regresión lineal múltiple

---

Luis Eduardo Galindo Amaya (1274895)

Asignatura	Estadística Avanzada
Docente	Olivia Mendoza Duarte
Fecha	29-09-2022

# Regresión lineal múltiple

Luis Eduardo Galindo Amaya (1274895)

29-09-2022

## Informacion del dataset<sup>1</sup>

Breast Cancer Wisconsin (Diagnostic)

- 1. ID number
- 2. Diagnosis (M = malignant, B = benign)

Ten real-valued features are computed for each cell nucleus:

- 3. radius (mean of distances from center to points on the perimeter)
- 4. texture (standard deviation of gray-scale values)
- 5. perimeter
- 6. area
- 7. smoothness (local variation in radius lengths)
- 8. compactness ( $\text{perimeter}^2 / \text{area} - 1.0$ )
- 9. concavity (severity of concave portions of the contour)
- 10. concave points (number of concave portions of the contour)
- 11. symmetry
- 12. fractal dimension (coastline approximation 1)

---

<sup>1</sup><https://archive-beta.ics.uci.edu/ml/datasets/breast+cancer+wisconsin+diagnostic>

## Explicación del problema y solución

### Problema

Se desea conocer si existe alguna relación entre las dimensiones y la textura de un tumor. Se puede determinar mediante un modelo lineal si un tumor mas grande tiene mas textura<sup>a</sup>.

<sup>a</sup>Desviación estándar de los valores de la vista.

### Solución

Las variables evaluadas corresponden a el radio, la textura, perímetro, área y suavidad. Podemos determinar en base a la regresión NO existe una relación entre las dimensiones y las texturas del tumor, los datos están demasiado dispersos para que la linealidad sea significativa.

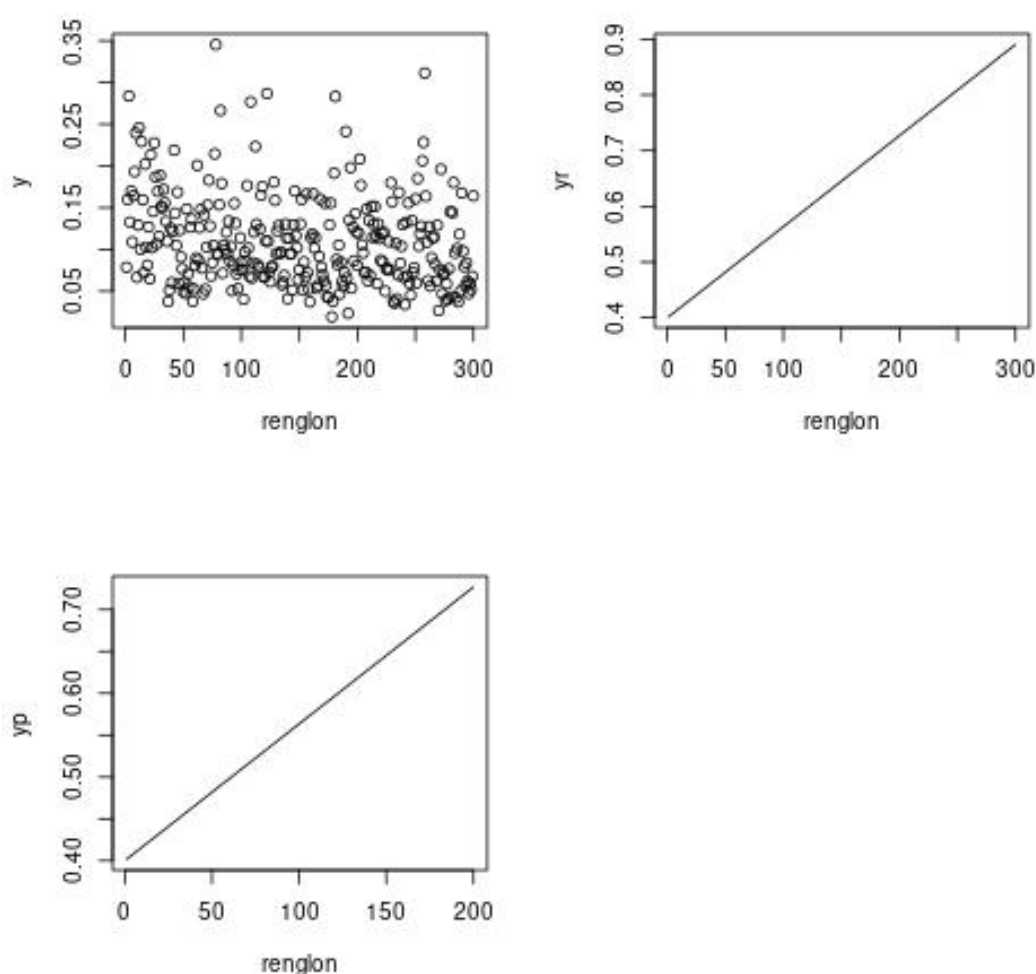


Figura 1: Resultados de la regresion lineal múltiple.

## Ejemplo de regresion lineal múltiple<sup>2</sup>

```
1 | archivo <- read.csv("data_akbilgic.csv")
2 |
3 | ## se toman los primeros 300 renglones de las columnas 2 a 7
4 | ## como datos conocidos y la columna 8 como los datos para hacer
5 | ## ajuste de la recta con la función lm
6 |
7 | renglon <- 1:300 # se genera una variable con rango del 1 a 300
8 | x <- archivo[1:300, 2:7]
9 | y <- archivo[1:300, 8]
10 |
11 | par(mfrow = c(3, 1))
12 |
13 | datos_grafica <- data.frame(renglon, y) # datos del atributo 7
14 | datos <- data.frame(x, y) # datos para la regresión
15 | datos
16 |
17 | ## se grafican todos los datos de la columna 7
18 | plot(datos_grafica)
19 |
20 | regresion <- lm(datos)
21 |
22 |
23 | b <- regresion$coefficients[1]
24 | m <- regresion$coefficients[2]
25 | b
26 | m
27 |
28 | ## se evalúa la ecuación de la recta ajustada a los primeros 300
29 | ## renglones de datos
30 | yr <- m * renglon + b
31 | yr
32 | datos_r <- data.frame(renglon, yr)
33 | datos_r
34 |
35 | ## se grafica la recta ajustada a los primeros 300 renglones de
36 | ## datos
37 | plot(datos_r, type = "l")
38 |
39 | ## se toman los renglones 301 a 500 (200 renglones) para hacer
40 | ## la predicción con la ecuación de la recta
41 | xpred <- archivo[301:500, 2:6]
42 | renglon <- 1:200
43 |
44 | ## se evalúa la ecuación de la recta ajustada a los datos de
```

<sup>2</sup><https://archive.ics.uci.edu/ml/datasets/ISTANBUL+STOCK+EXCHANGE>

```
45 | ## predicción
46 | yp <- m * renglon + b
47 | yp
48 | datos_p <- data.frame(renglon, yp)
49 | datos_p
50 | plot(datos_p, type = "l")
```