Práctica 7

Laboratorio 1 - Correlación

Luis Eduardo Galindo Amaya (1274895)

Asignatura	Estadística Avanzada
Docente	Olivia Mendoza Duarte
Fecha	06-10-2022

Laboratorio 1 - Correlación

Luis Eduardo Galindo Amaya (1274895)

06-10-2022

Informacion del dataset

Breast Cancer Wisconsin[1]

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 (perimeter^2 / 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 (çoastline approximation 1)

[1] https://archive-beta.ics.uci.edu/ml/datasets/breast+cancer+wisconsin+diagnostic

Columnas

- 1. the transaction date (for example, 2013.250=2013 March, 2013.500=2013 June, etc.)
- 2. the house age (unit: year)
- 3. the distance to the nearest MRT station (unit: meter)
- 4. the number of convenience stores in the living circle on foot (integer)
- 5. the geographic coordinate, latitude. (unit: degree)
- 6. the geographic coordinate, longitude. (unit: degree)
- 7. house price of unit area

Explicaciones

Alta correlación (Breast Cancer Wisconsin)

Se evaluaron las columnas, 3 (radius) y 5 (perimeter). el radio esta relacionado con el perímetro del tumor, si uno de los dos se incrementa el otro aumenta proporcionalmente.

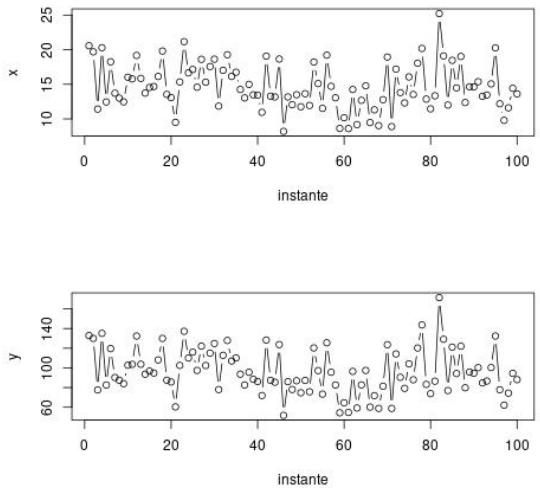

Figura 1: $r = 0{,}9964379$

Baja correlación (Breast Cancer Wisconsin)

Se evaluaron las columnas, 3 (radius) y 4 (textura). la correlación es muy baja por la textura no afecta a el radio.

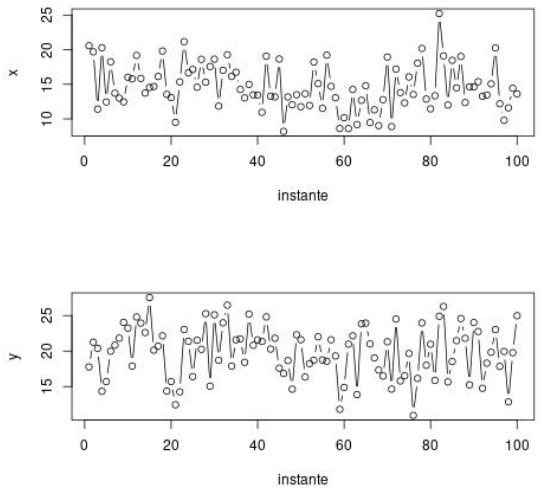

Figura 2: $r = 0{,}468663$

Correlación Inversa (Real estate valuation data set)

Se evaluaron las columnas 4 (the number of convenience stores in the living circle on foot) y 7 (house price of unit area) por lo que podemos determinar que las casa de mayor valor estan lejos de las tiendas.

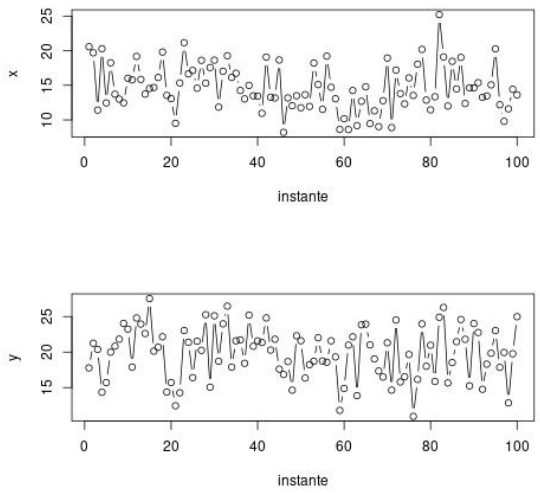

Figura 3: $r = -0{,}8917442$

Código

```
1
2  archivo <- read.csv("wdbc.data")
3
4  instante <- 1:100
5  x <- archivo[1:100, 3]
6  y <- archivo[1:100, 4]
7  n <- length(instante)
8
9  datos_x <- data.frame(instante, x)
10 datos_y <- data.frame(instante, y)
11
12 jpeg(file = "./practica 7/img/corelacion_inversa.jpeg")
13 par(mfrow = c(2, 1))
14 plot(datos_x, type = "b")
15 plot(datos_y, type = "b")
16 dev.off()
17
18 sum_xy <- 0
19 sum_x <- 0
20 sum_y <- 0
21 sum_xx <- 0
22 sum_yy <- 0
23
24
25 for (i in instante) {
26   sum_xy <- sum_xy + x[i] * y[i]
27   sum_x <- sum_x + x[i]
28   sum_y <- sum_y + y[i]
29   sum_xx <- sum_xx + x[i]^2
30   sum_yy <- sum_yy + y[i]^2
31 }
32
33 Sxy <- sum_xy - (sum_x * sum_y / n)
34 Sxx <- sum_xx - (((sum_x)^2) / n)
35 Syy <- sum_yy - (((sum_y)^2) / n)
36 r <- Sxy / (sqrt(Sxx) * sqrt(Syy))
37 r
```