

# Ciencia de datos, práctica 2

Juan Casado Ballesteros, Samuel García Gonzalez, Iván Anaya Martín

October 25, 2019

**Abstract**

## Contents

<b>1</b>	<b>EJ1</b>	<b>3</b>
<b>2</b>	<b>EJ2</b>	<b>3</b>
2.1	Residuos . . . . .	3
2.2	Coefficientes . . . . .	3
2.3	Error estándar . . . . .	4
2.4	Correlación cuadrada . . . . .	4
<b>3</b>	<b>EJ3</b>	<b>5</b>

## 1 EJ1

## 2 EJ2

Creamos un .txt con los datos proporcionados sobre el radio y densidad de los planetas y lo leemos.

```
> datos2 <- read.table("datos2.txt")
> datos2
```

	Nombre	Radio	Densidad
1	Mercurio	2.4	5.4
2	Venus	6.1	5.2
3	Tierra	6.4	5.5
4	Marte	3.4	3.9

Calculamos la regresión sobre dichos datos para obtener la recta que más se aproxime a los puntos que tenemos.

```
> regresion2 <- lm(Densidad~Radio, data=datos2)
```

Podemos ver los valores que adopta la ecuación de la recta que se generará.

```
> # y = ax + b
> regresion2$coefficients[1] # b
```

```
(Intercept)
4.362396
```

```
> regresion2$coefficients[2] # a
```

```
Radio
0.1393669
```

Cuando calculamos la recta de regresión sobre unos datos es necesario evaluar la calidad de esta. Debemos analizar cómo de bien se ajusta a nuestros datos. Podemos ver esta información mediante summary.

### 2.1 Residuos

Diferencias entre cada valor de y real y cada valor de y obtenido mediante la función de regresión.

```
> summary(regresion2)$residuals
```

1	2	3	4
0.70312301	-0.01253452	0.24565541	-0.93624389

### 2.2 Coeficientes

Coefficientes estimados para y error estándar para cada uno de ellos.

```
> summary(regresion2)$coefficients
```

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	4.3623964	1.2049754	3.6203201	0.06854492
Radio	0.1393669	0.2466205	0.5651067	0.62893696

### 2.3 Error estándar

Podemos comprobar que coincide con nuestra implementación. Cuanto más próximo a 0 sea el error estándar mejor será la recta de regresión.

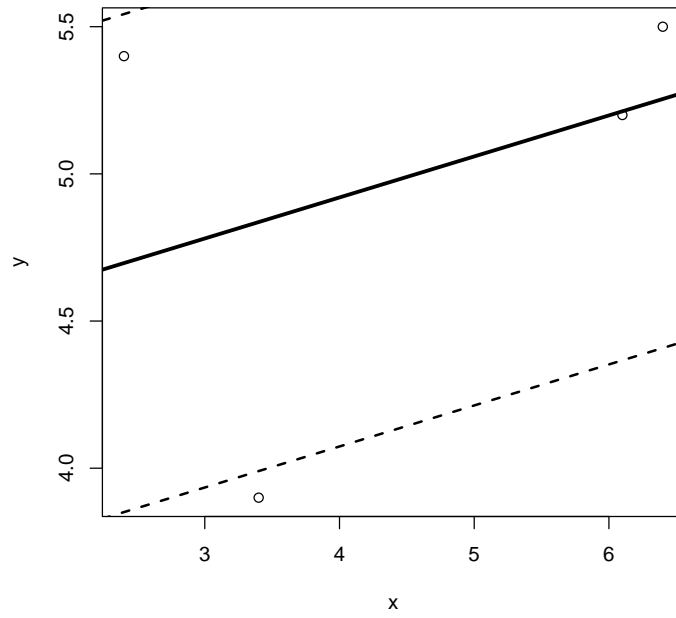
```
> summary(regresion2)$sigma  
[1] 0.8460019  
  
> errorEstandar(datos2$Radio, datos2$Densidad, regresion2)  
  
Radio  
0.8460019
```

### 2.4 Correlación cuadrada

Podemos comprobar que coincide con nuestra implementación. Este valor está entre 0 y 1 siendo mejor cuanto más próximo a 1 sea (idealmente a partir de 0.8).

```
> summary(regresion2)$r.squared  
[1] 0.1376878  
  
> correlacionCuadrada(datos2$Radio, datos2$Densidad)  
  
Radio  
0.1376878
```

Para finalizar dibujaremos una gráfica en la que se representarán los datos junto a la recta de regresión. Paralela a la recta de regresión dibujaremos las rectas que marcan el error estándar entorno a la recta de regresión. En trazado gris grueso la que marca la región en la que estarán el 66% de los datos y en gris fino la



que marca el 95%.

### 3 EJ3