

PRÁCTICA 2: REGRESIÓN LINEAL SIMPLE

Alba Calvo Herrero

1ºB GII 21/10/19

EJERCICIOS PRÁCTICA 2

1. Dados los siguientes datos, que relacionan el peso y la estatura de 10 personas. Se pide:

Peso (x)	82	75	70	68	44	63	80	79	54	54
Estatura (y)	185	185	180	178	159	170	190	172	162	165

- Calcular, usando la fórmula vista en teoría, la recta de regresión del Peso en función de la Estatura.
 - Comparar los resultados obtenidos en a con los datos obtenidos de la resolución escrita de este ejercicio.
 - Realizar los pasos a) y b) pero calculando ahora la recta de regresión de la Altura en función del Peso
2. Dado el fichero cardata.sav utilizado en la práctica 1. Se pide, calcular gráficos de dispersión, rectas de regresión, coeficiente de correlación y gráficos de las rectas de regresión para los siguientes pares de valores:
- Potencia (horsepower) en función del consumo(mpg)
 - Aceleración (accel) en función del consumo (mpg)
 - Peso (weight) en función del consumo (mpg)
 - Precio(Price) en función del consumo
 - ¿Cuáles de las rectas de regresión calculadas en el punto anterior se ajustan más a los valores de los datos? ¿Porqué?

EJERCICIO 1

1.A.

En este apartado tenemos que calcular con las fórmulas vistas en teoría la recta de regresión. Estas fórmulas son las siguientes:

$$y = a + bx$$

$$b = \frac{Cov(x,y)}{S_x^2}$$

$$a = \bar{y} - b\bar{x}$$

Empezamos calculando la media de x e y:

$$x = \frac{185 + 185 + 180 + 178 + 159 + 170 + 190 + 172 + 162 + 165}{10} = 174.6cm$$

$$y = \frac{82 + 75 + 70 + 68 + 44 + 63 + 80 + 79 + 54 + 54}{10} = 66.9 kg$$

A continuación hallamos S_x^2

$$S_x^2 = \frac{185^2 + 185^2 + 180^2 + 178^2 + 159^2 + 170^2 + 190^2 + 172^2 + 162^2 + 165^2}{10} - 174.6^2 = 101.64$$

Por último, con las fórmulas anteriores calculamos a y b:

$$b = \frac{109.26}{101.64} = 1.07497$$

$$a = 66.9 - 1.07497 \cdot 174.6 = -120.7898$$

Por tanto la recta de regresión será de la forma:

$$y = -120.79 + 1.07x$$

1.B

En este apartado debemos calcular la recta de regresión del peso en función de la estatura en R. Como es el peso en función de la estatura, la variable dependiente y será el peso y la variable independiente x , la estatura.

Primero definimos los vectores peso y estatura en R con los datos del problema.

```
> estatura=c(185,185,180,178,159,170,190,172,162,165)
> view(estatura)

> peso=c(82,75,70,68,44,63,80,79,54,54)
> peso
[1] 82 75 70 68 44 63 80 79 54 54
> view(peso)
> |
```

Usamos el comando **view()** para visualizarlos en la parte superior izquierda como se ve en las siguientes capturas.

PESO		ESTATURA	
	▲ V1 ▼		▲ V1 ▼
1	82	1	185
2	75	2	185
3	70	3	180
4	68	4	178
5	44	5	159
6	63	6	170
7	80	7	190
8	79	8	172
9	54	9	162
10	54	10	165

Para calcular la recta de regresión se usan las siguientes fórmulas:

$$y = a + bx$$

$$b = \frac{Cov(x, y)}{S_x^2}$$

$$a = \bar{y} - b\bar{x}$$

Podemos observar que haciéndolo con las fórmulas necesitaríamos calcular a, b, la media de x y la de y, la covarianza y la varianza. tal y como hemos hecho en el apartado a. Sin embargo, en R existe un comando que muestra directamente los parámetros a y b. Este es:

```
> mlxy = lm(peso~ estatura)
> mlxy

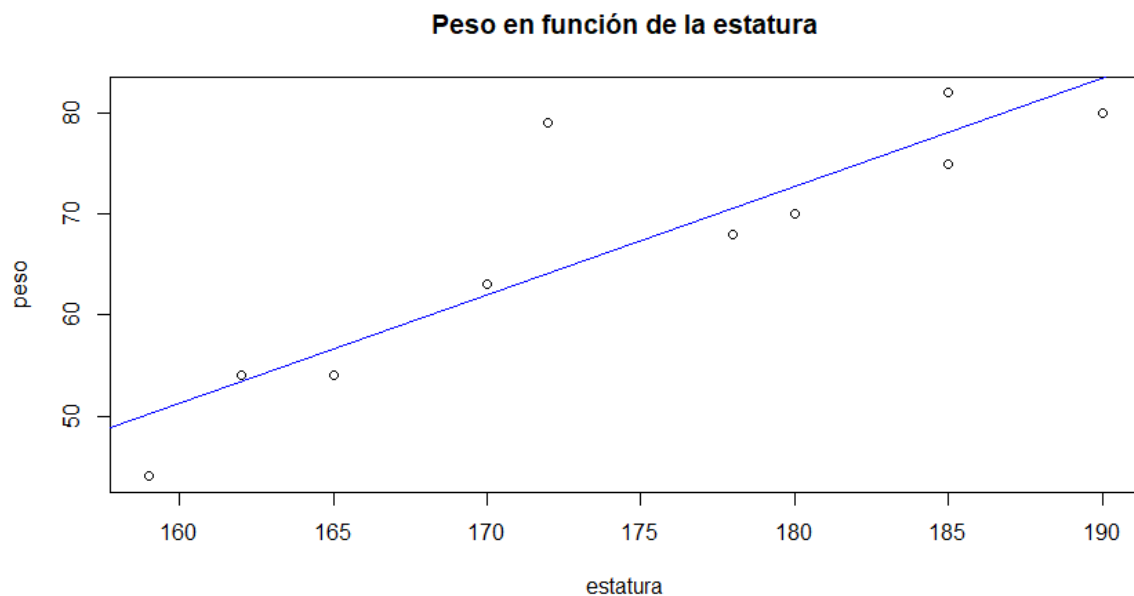
Call:
lm(formula = peso ~ estatura)

Coefficients:
(Intercept)      estatura 
   -120.962         1.076
```

Donde a=-120.962 y b=1.076. Por lo que la recta de regresión tiene la forma: $y = -120.962 + 1.076x$

A continuación, usamos el comando **plot**. Y por último, creamos la recta de regresión con el comando **abline**.

```
> plot(estatura,peso, main= "Peso en función de la estatura")
> abline(mlxy, col="blue")
```



Hemos podido observar que con el comando de R obtenemos los resultados de manera mucho más rápida y con diferencias muy pequeñas respecto al primer apartado.

1.C.

En este apartado tenemos que realizar el caso contrario, es decir, la estatura en función del peso. Para ello usamos los mismos comandos pero cambiando la estatura por el peso y viceversa.

Calculamos los parámetros a y b:

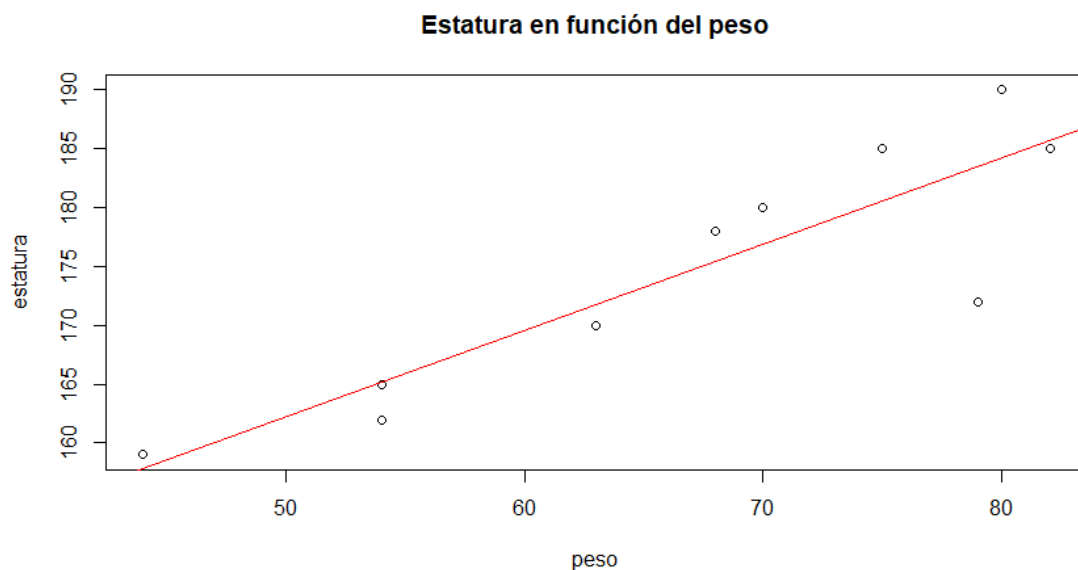
```
> mlxy = lm(estatura~ peso)
> mlxy
```

```
call:
lm(formula = estatura ~ peso)
```

```
Coefficients:
(Intercept)      peso
  125.6590      0.7316
```

Donde a= 125.6590 y b=0.7316.

```
> plot(peso,estatura, main= "Estatura en función del peso")
> abline(mlxy, col="red")
```



Ejercicio 2

Se pide calcular la recta de regresión y el coeficiente de correlación en estos casos:

a. **Potencia (horsepower) en función del consumo(mpg)**

Primero declaramos el vector consumo (mpg) y el vector potencia (horsepower).

```
> mpg=cardata$mpg  
> horsepower=cardata$horsepower
```

Calculamos la recta de regresión como antes.

```
> mlxy=lm(horsepower~mpg)  
> mlxy
```

call:

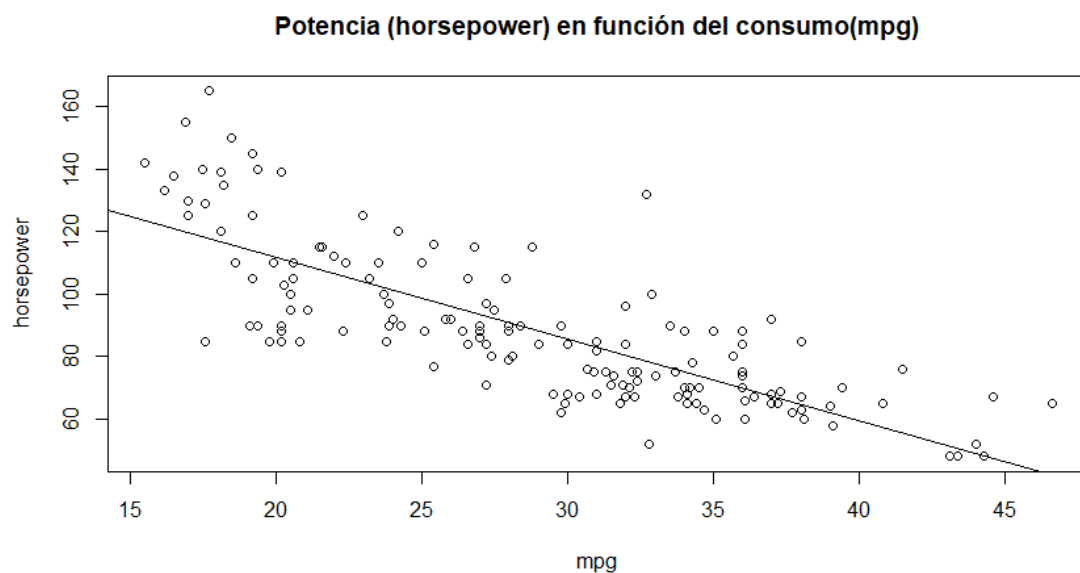
```
lm(formula = horsepower ~ mpg)
```

Coefficients:

(Intercept)	mpg
164.099	-2.617

Donde $a=164.099$ y $b=-2.167$ y por tanto la recta de regresión lineal ($y=a+bx$) es $y=164.1-2.62x$

```
> plot(mpg,horsepower,main="Potencia (horsepower) en función del consumo(mpg)")  
> abline(mlxy)
```



Para calcular el coeficiente de correlación eliminamos los datos nulos con `cardata=na.omit(cardata)`. Y a continuación usamos el comando `cor`.

```
> cor(cardata$mpg, cardata$horsepower)  
[1] -0.7887461
```

b. Aceleración (accel) en función del consumo (mpg)

Calculamos la recta de regresión:

```
> accel=cardata$accel  
> mlxy=lm(accel~mpg)  
> mlxy
```

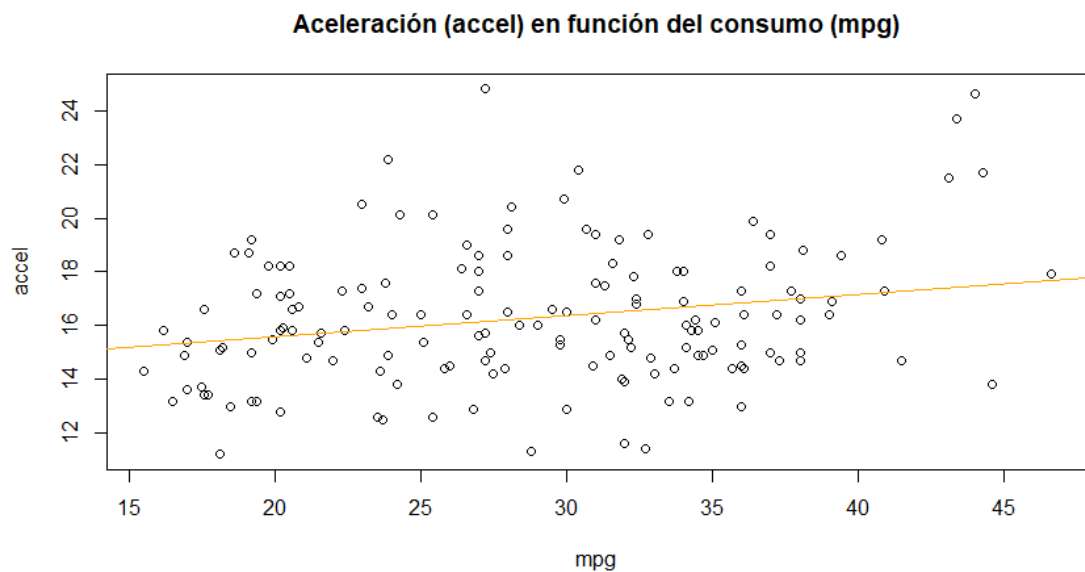
```
Call:  
lm(formula = accel ~ mpg)
```

```
Coefficients:  
(Intercept)          mpg  
    14.04237         0.07778
```

a=14.04237 y b=0.07778

y= 14.04+0.078x

```
> plot(mpg,accel,main="Aceleración (accel) en función del consumo (mpg)".  
> abline(mlxy, col="orange")
```



Calculamos el coeficiente de correlación.

```
> cardata=na.omit(cardata)  
> cor(cardata$accel, cardata$mpg)  
[1] 0.2347268
```


c. **Peso (weight) en función del consumo (mpg)**

Calculamos la recta de regresión.

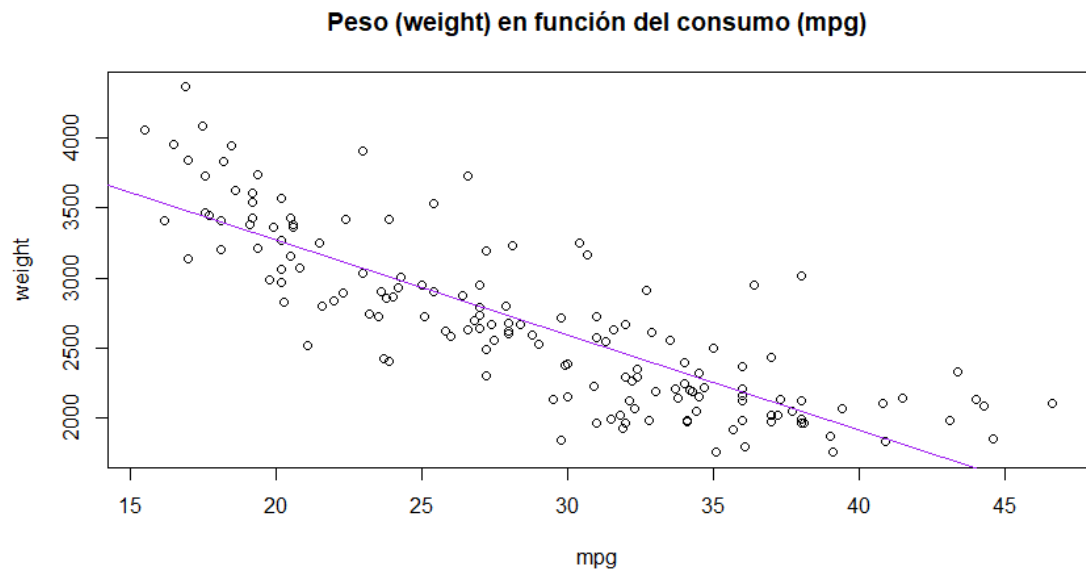
```
> mpg=cardata$mpg  
> weight=cardata$weight  
> mlxy=lm(weight~mpg)  
> mlxy
```

```
Call:  
lm(formula = weight ~ mpg)
```

```
Coefficients:  
(Intercept)          mpg  
    4623.52         -67.77
```

```
> plot(mpg,weight,main="Peso (weight) en función del consumo (mpg)")  
> abline(mlxy, col="purple")
```

$y=4623.52-67.77x$



Calculamos el coeficiente de correlación:

```
> cor(cardata$weight, cardata$mpg)  
[1] -0.8257621
```

d. Precio(Price) en función del consumo

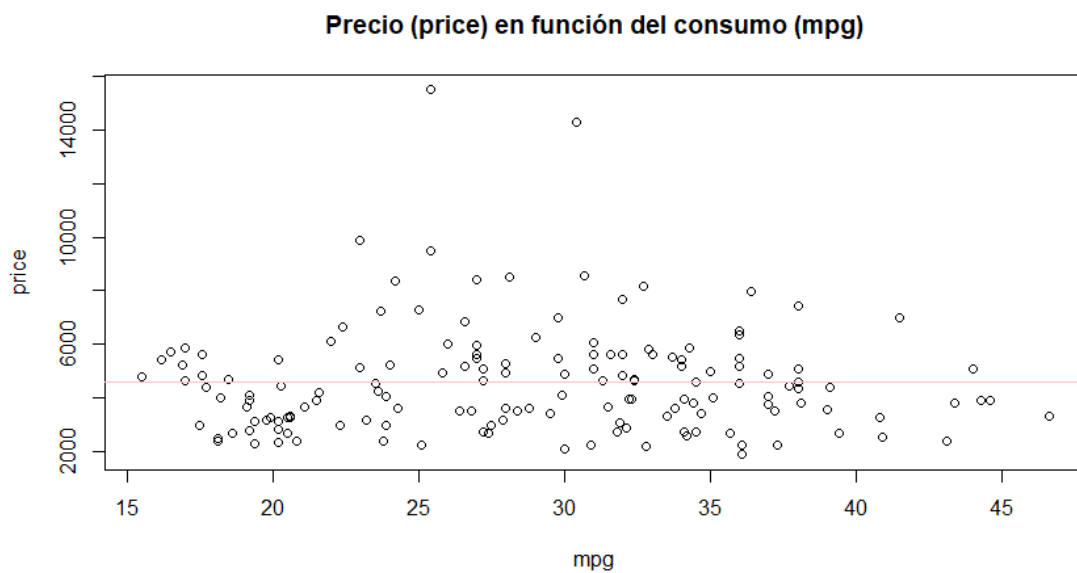
```
> price=cardata$price
> mlxy=lm(price~mpg)
> mlxy

call:
lm(formula = price ~ mpg)

Coefficients:
(Intercept)      mpg
  4617.1930      -0.2588

> plot(mpg,price,main="Precio (price) en función del consumo (mpg)")
> abline(mlxy, col="pink")
> |
```

$$y=4617.19-0.26x$$



Calculamos el coeficiente de correlación:

```
> cor(cardata$price, cardata$mpg)
[1] 0.01070007
```