ITAM
Departamento de Estadística

Inferencia Estadística– Laboratorio #10
Modelo de Regresión Lineal Simple

Recordar:

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} \quad (1)$$

$$S_{XY} = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) \quad (2)$$

$$S_{JJ} = \sum_{i=1}^{n}(j_i - \bar{j})^2 \;\; para\ J = x, y \quad (3)$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x} \quad (4)$$

Y la recta de mínimos cuadrados es:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x \quad (5)$$

1. Conteste cada una de las siguientes preguntas:
   (a) ¿Qué es un modelo de regresión lineal simple?
   (b) ¿Cuáles son los objetivos específicos que persigue este modelo?

2. Si $\hat{\beta}_0$ y $\hat{\beta}_0$ son las estimaciones de MCO para un modelo de regresión lineal simple. Demuestre que la recta de MCO siempre pasa por el punto $(\bar{x}, \bar{y})$.

3. Ajuste la recta de MCO. Para los iguientes datos:

y	3.0	2.0	1.0	1.0	0.5
x	−2.0	−1.0	0.0	1.0	2.0

4. Es frecuente que a los auditores se les exija comparar el valor auditado (o de lista) de un artículo de inventario contra el valor en libros. Si una empresa está llevando su inventario y libros actualizados, debería haber una fuerte relación lineal entre los valores auditados y en libros. Una empresa muestreó diez artículos de inventario y obtuvo los valores auditado y en libros que se dan en la tabla siguiente. Ajuste el modelo a estos datos.

Artículo	Valor auditado (y_i)	Valor en libros (x_i)
1	9	10
2	14	12
3	7	9
4	29	27
5	45	47
6	109	112
7	40	36
8	238	241
9	60	59
10	170	167

Datos :

- $S_{xy} = 54243$, $S_{xx} = 54714$
- $\sum_1^n x_i = 720$, $\sum_1^n y_i = 721$

(a) ¿Cuál es su estimación para el cambio esperado en valor auditado para un cambio de unidad en el valor de libros?

(b) Si el valor en libros es $x = 100$, ¿qué usaría para estimar el valor auditado?

5. Suponga que postulamos el siguiente modelo

$$y_i = \beta_1 x_1 + \epsilon_i, \quad i = 1, 2, ..., n.$$

Encuentre el estimador de MCO de β_1.[*Hint:* SCR=$\sum_1^n [y_i - \hat{y}]^2$]

6. Encuentre:

(a) $E[y_i]$, $V[y_i]$, $Cov(y_i, y_j)$

(b) $E[\hat{\beta}_1]$, $V[\hat{\beta}_1]$

(c) $E[\hat{\beta}_0]$, $V[\hat{\beta}_0]$

(d) $Cov(\hat{\beta}_1, \hat{\beta}_0)$

7. (a) Demuestre que $SCR = S_{yy} - \hat{\beta}_1 S_{xy}$

(b) Use el inciso anterior para demostrar que $SCR \leq S_{yy}$

8. Suponga que $Y_1, Y_2, ...Y_n \overset{v.a.i.i.d}{\sim} N(\beta_0 + \beta_1 x_i, \sigma^2)\ \forall\, i$. Demuestre que los estimadores de máxima probabilidad (MLE) de β_0, β_1 son iguales a los estimadores de MCO.

laboratorio 10 Inferencia Estad Parte R

Luis Martinez

Otoño 20201

Vamos a correr un modelo de regresión lineal simple en R.

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --

## v ggplot2 3.3.3     v purrr   0.3.4
## v tibble  3.1.0     v dplyr   1.0.4
## v tidyr   1.1.1     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.5.0

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(ggplot2)
library(ggcorrplot)
library(dplyr)
library(ggpubr)
```

En este caso vamos a usar los datos *mtcars*, ya cargados a R y realizaremos inferencia sobre los mismos.

```
#Cargamos los datos
data("mtcars")
datos.coches<- mtcars
#Glimpse al data frame
head(mtcars)
```

```
##                    mpg cyl disp  hp drat    wt  qsec vs am gear carb
## Mazda RX4         21.0   6  160 110 3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag     21.0   6  160 110 3.90 2.875 17.02  0  1    4    4
## Datsun 710        22.8   4  108  93 3.85 2.320 18.61  1  1    4    1
## Hornet 4 Drive    21.4   6  258 110 3.08 3.215 19.44  1  0    3    1
## Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02  0  0    3    2
## Valiant           18.1   6  225 105 2.76 3.460 20.22  1  0    3    1
```

```
# Manipulación de variables
x<- datos.coches$wt
y<- datos.coches$mpg


#Correlaciones
(cor(x, y, method="pearson"))
```

```
## [1] -0.8676594
```

```r
(cor(x,y, method = "spearman"))
```

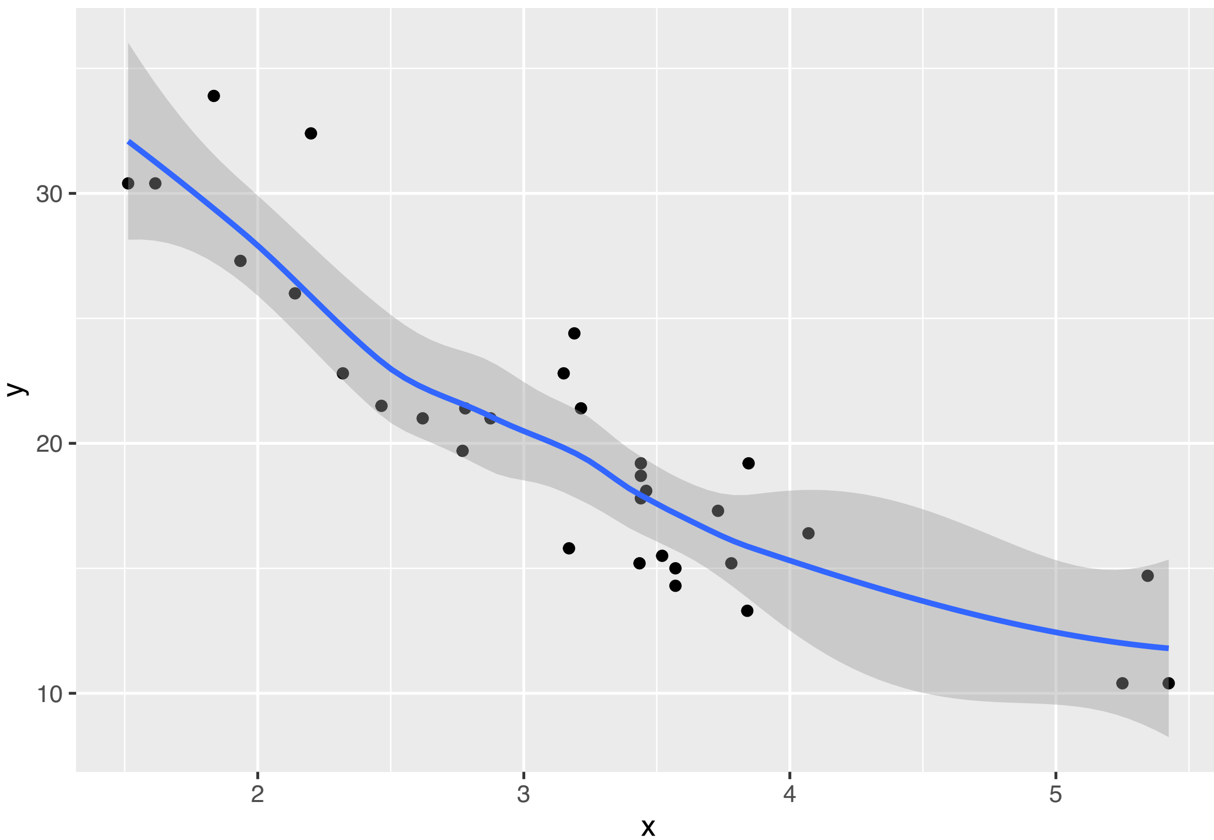
```
## [1] -0.886422
```

```r
(cor(x,y, method = "kendall"))
```

```
## [1] -0.7278321
```

```r
#Sugerencia de una relacion negativa entre miles per galon y weight
ggplot(datos.coches, aes(x,y))+geom_point()+stat_smooth()
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```




```r
#Construimos el modelo de regresion lineal simple
(modelo<- lm(y~x, data=datos.coches))
```

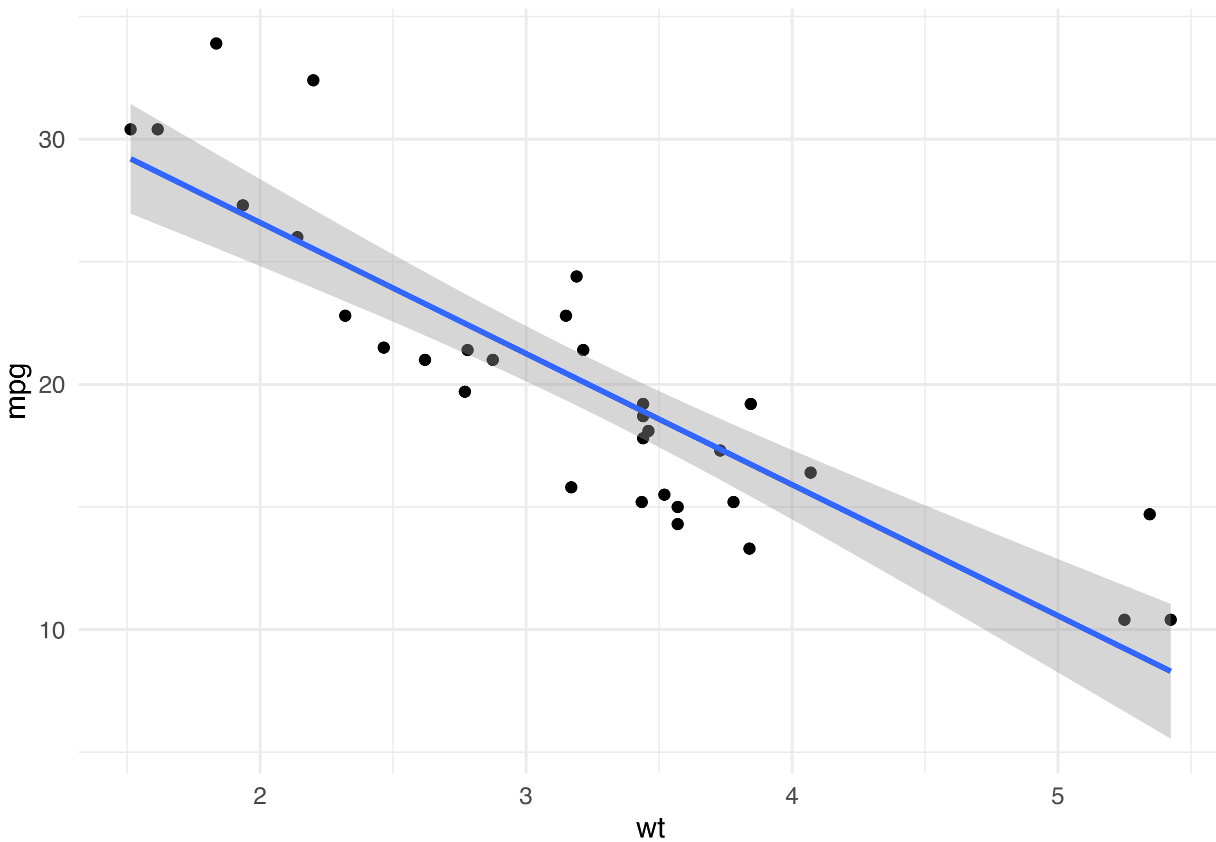
```
##
## Call:
## lm(formula = y ~ x, data = datos.coches)
##
## Coefficients:
## (Intercept)            x
##      37.285       -5.344
```

```r
summary(modelo)
```

```
##
## Call:
## lm(formula = y ~ x, data = datos.coches)
##
```

```
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.5432 -2.3647 -0.1252  1.4096  6.8727
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  37.2851     1.8776  19.858  < 2e-16 ***
## x            -5.3445     0.5591  -9.559 1.29e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.046 on 30 degrees of freedom
## Multiple R-squared:  0.7528, Adjusted R-squared:  0.7446
## F-statistic: 91.38 on 1 and 30 DF,  p-value: 1.294e-10
```

```r
#Corremos la Regresion
ggplot(datos.coches)+ geom_point(aes(x= wt, y= mpg))+
                      stat_smooth(aes(x= wt, y= mpg), method= "lm",
                      formula= y ~ x, se=TRUE)+theme_minimal()
```




```r
#Intervalos de confianza para estimadores del modelo al 97.5%
confint(modelo)
```

```
##                 2.5 %    97.5 %
## (Intercept) 33.450500 41.119753
## x           -6.486308 -4.202635
```

```r
#Coeficiente de correlación R^2
#Este coeficiente mide cuanta proporcion del modelo es explicada por la
```

```r
#regresion
(R.cuadrada<- (cor(x, y, method="pearson"))^2)
```

```
## [1] 0.7528328
```