

Econometría Aplicada con



```
R Console (32-bit)
Archivo Editar Misc. Ejecutar Ventanas Ayuda

> x <- c(1,2,3,4,5,6)
> y <- x^2
> print(y)
[1] 1 4 9 16 25 36
> mean(y)
[1] 15.16667
> var(y)
[1] 178.9444
> lm_1 <- lm(y ~ x)
> print(lm_1)

Call:
lm(formula = y ~ x)

Coefficients:
(Intercept) -9.333
x 7.000

> summary(lm_1)

Call:
lm(formula = y ~ x)

Coefficients:
(Intercept) -9.333
x 7.000

Residuals:
1 2 3 4 5 6
3.3333 -0.6667 -2.6667 -2.6667 -0.6667 3.3333

Coefficients:
(Intercept) Estimate Std. Error t value Pr(>|t|)
1 -9.3333 2.8441 -3.282 0.030453 *
x 7.0000 0.7303 9.585 0.000662 ***
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.055 on 4 degrees of freedom
Multiple R-squared: 0.9583, Adjusted R-squared: 0.9478
F-statistic: 91.87 on 1 and 4 DF, p-value: 0.000662

> |
```



EJEMPLO 1: REGRESIÓN LINEAL SIMPLE

OBJETIVOS



El objetivo principal del ejemplo es realizar un análisis de regresión haciendo uso de R para estimar el valor de los parámetros planteados en el modelo.

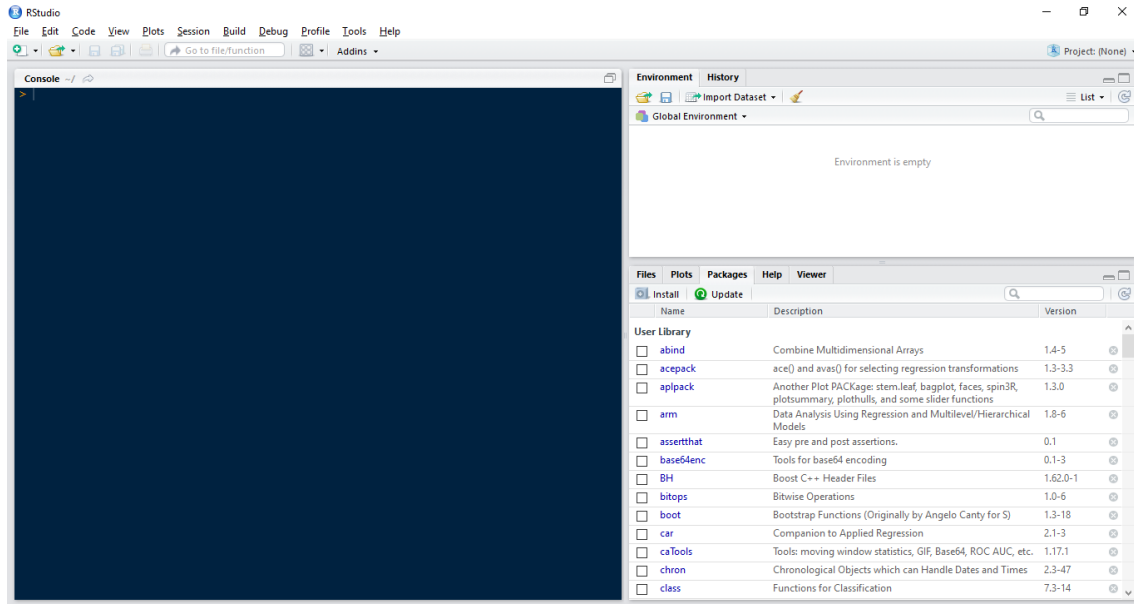
PLANTEAMIENTO



La base de datos adjunta muestra información de 2000 mujeres, en el siguiente ejemplo se estimarán los parámetros de un modelo que tratará de explicar los salarios por hora en función de los años de educación.

Desarrollo

Se tiene el programa RStudio abierto:



El modelo tiene la siguiente forma:

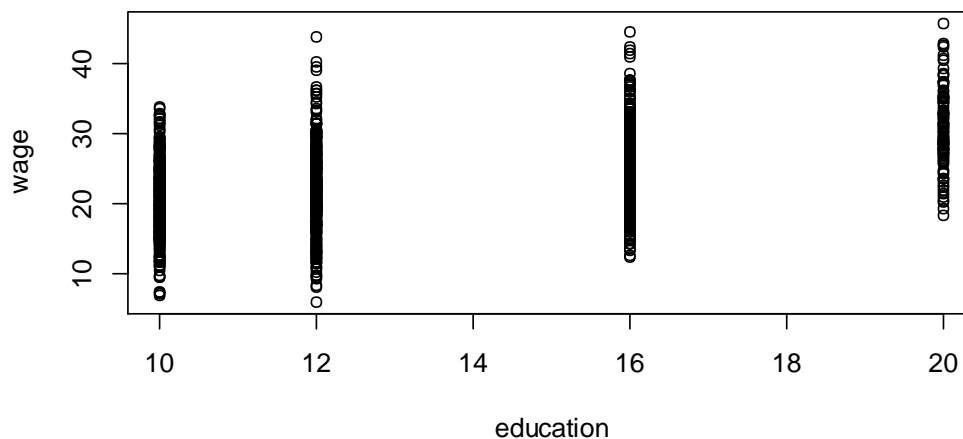
$$wage = \beta_0 + \beta_1 education + \varepsilon$$

En primer lugar, se creará un gráfico de regresión con ambas variables, el código es el siguiente:

```
attach(womenwk)
plot(education , wage, main = "Dispersión entre educación y salario")
```

Esto mostrará el siguiente gráfico:

Dispersión entre educación y salario



En el modelo a estimarse **wage** (salario), será la variable dependiente y **education** (años de educación), será la variable independiente. Luego de crear el objeto del modelo de regresión se digitará el comando **summary()**, para mostrar los resultados de dicho modelo

```
modelo = lm(wage ~ education, data = womenwk)
summary(modelo)
```

Los resultados son los siguientes:

```
Call:
lm(formula = wage ~ education, data = womenwk)

Residuals:
    Min       1Q   Median       3Q      Max
-16.3781  -3.7422  -0.1346   3.7202  21.7162

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  10.6190     0.6905   15.38  <2e-16 ***
education     0.9703     0.0500   19.41  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.573 on 1341 degrees of freedom
(657 observations deleted due to missingness)
Multiple R-squared:  0.2193,    Adjusted R-squared:  0.2187
F-statistic: 376.7 on 1 and 1341 DF,  p-value: < 2.2e-16
```

Estos resultados muestran que el parámetro que acompaña a la variable de años de educación es estadísticamente significativo a más del 99% de confianza, muestran que a medida que se incremente un año de educación el salario por hora aumentará en 0.9703 dólares.

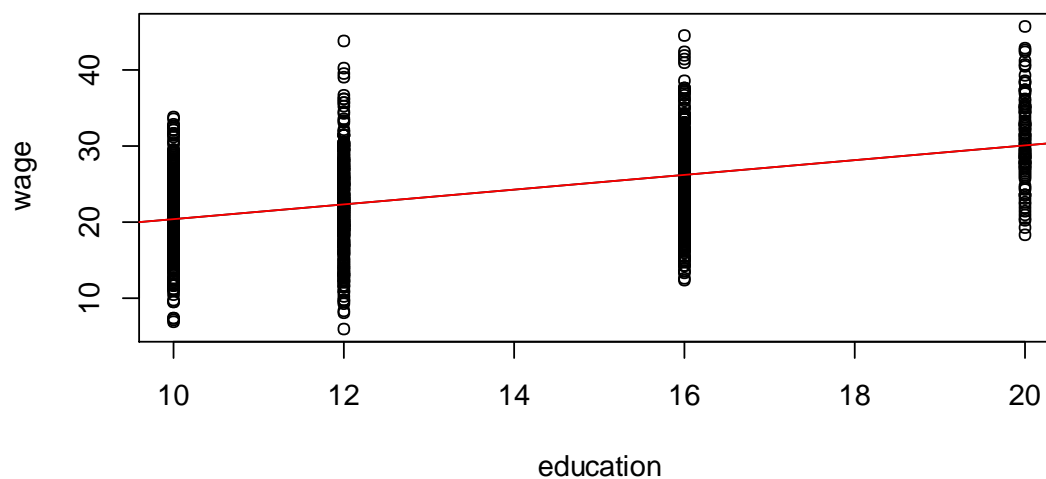
Se debe tener en cuenta en el siguiente modelo que el R-cuadrado asociado a dicha estimación es un valor muy bajo, por lo que el modelo no tiene gran capacidad explicativa; esto indica que se puede mejorar dicho modelo introduciendo más variables explicativas o generar valores logarítmicos para la variable dependiente.

Para colocar la línea de regresión en el gráfico generado se usará la función **abline()**, indicando el modelo con el cuál se trabajará:

```
abline(modelo , col = "red")
```

El gráfico será el siguiente:

Dispersión entre educación y salario



Este gráfico dará razón de la relación directa entre las variables analizadas en el modelo de regresión.