

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) -9.3333
x 7.0000

> summary(lm_1)

Call:
lm(formula = y ~ x)

Coefficients:
(Intercept) -9.3333
x 7.0000

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: SOBREAJUSTE DEL MODELO

OBJETIVOS



El objetivo principal del ejemplo es mostrar el problema del sobreajuste de un modelo de regresión lineal y detectar dicho problema haciendo uso de la prueba RESET de Ramsey en R.

PLANTEAMIENTO



La base de datos adjunta representa características de autos en EE.UU. En el siguiente ejemplo se estimarán los parámetros de un modelo de regresión lineal y se va a probar el error en la especificación del mismo.

Desarrollo

El modelo tiene la siguiente forma:

$$price = \beta_0 + \beta_1 turn + \beta_2 length + \beta_3 mpg + \beta_4 weight + \varepsilon$$

Este modelo trata explicar el precio de los vehículos, donde los términos regresores son el grado de giro del vehículo (**turn**), la longitud (**length**), las millas por galón (**mpg**) y el peso (**weight**).

Se hace la regresión en R creando un objeto de nombre **mod1**.

```
mod1 = lm(price ~ turn + length + mpg + weight, data = auto)
summary(mod1)
```

Call:

```
lm(formula = price ~ turn + length + mpg + weight, data = auto)
```

Residuals:

Min	1Q	Median	3Q	Max
-4832.5	-1573.7	-439.6	1425.3	6230.5

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	19581.418	6005.223	3.261	0.00173	**
turn	-323.861	126.882	-2.552	0.01291	*
length	-73.147	40.212	-1.819	0.07324	.
mpg	-94.651	80.879	-1.170	0.24591	
weight	5.030	1.154	4.359	4.46e-05	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2325 on 69 degrees of freedom

Multiple R-squared: 0.4128, Adjusted R-squared: 0.3788

F-statistic: 12.13 on 4 and 69 DF, p-value: 1.606e-07

Aquí se puede notar que sólo la variable **mpg** no es significativa en el modelo, el R-cuadrado del mismo es cercano a 42% y el modelo es estadísticamente significativo a nivel global.

Para evaluar el error de especificación se debe hacer uso del comando **resettest()**.

```
resettest(mod1)
```

RESET test

data: mod1

RESET = 14.44, df1 = 2, df2 = 67, p-value = 6.106e-06

Esto indicará que la hipótesis nula es rechazada, por lo tanto, se entiende que hay un problema de especificación en el modelo.

Puede que exista un sobrejuste del mismo, que implica que hay un exceso de variables explicativas en el modelo, para eso se reduce el número de variables a dos básicamente, **turn** y **length**; esto implica que ahora se asumirá que el precio esencialmente depende del grado de giro y de la longitud del vehículo.

Se hace la estimación del modelo y una vez creado el objeto se realiza la prueba RESET:

```
mod2 = lm(price ~ turn + length, data = auto)
resettest(mod2)
```

RESET test

```
data: mod2
RESET = 0.56201, df1 = 2, df2 = 69, p-value = 0.5726
```

Los resultados del test hacen que se acepte la hipótesis nula y, por lo tanto, no se puede decir que el modelo tiene error de especificación.