

# 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:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -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: MULTICOLINEALIDAD I

### OBJETIVOS

---



El objetivo principal del ejemplo es mostrar la detección del problema de multicolinealidad en el modelo de regresión haciendo uso de R.

## PLANTEAMIENTO

---



La base de datos adjunta representa características de autos en EE.UU. En el siguiente ejemplo se detectará el problema de la multicolinealidad estimando una matriz de dispersiones y haciendo el análisis del factor inflador de varianzas.

## Desarrollo

El modelo tiene la siguiente forma:

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

Este modelo trata explicar el precio de los vehículos, donde los términos regresores son el espacio en el maletero (**trunk**), 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 ~ mpg + weight + length + trunk,
           data = auto)
summary(mod1)
```

La regresión mostrará los siguientes resultados:

```
Call:
lm(formula = price ~ mpg + weight + length + trunk, data = auto)

Residuals:
    Min       1Q   Median       3Q      Max
-4794.8 -1627.5  -513.4  1400.1  5912.3

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 14896.450   6080.278   2.450 0.016828 *
mpg          -86.162    84.540  -1.019 0.311675
weight         4.388     1.178   3.723 0.000398 ***
length       -109.062    43.035  -2.534 0.013542 *
trunk         25.594     97.070   0.264 0.792825
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

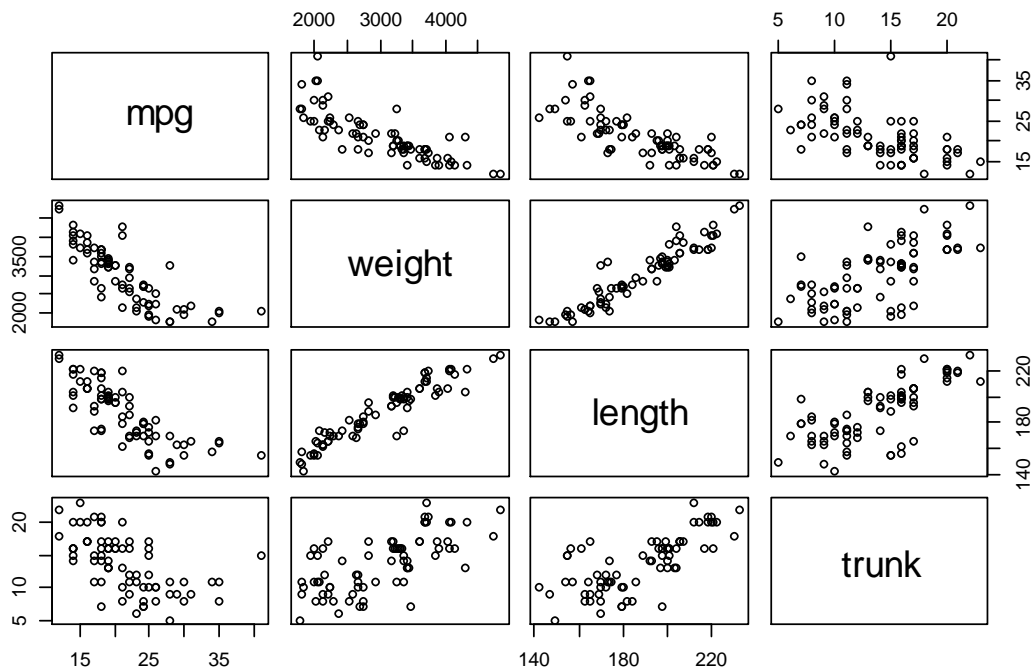
Residual standard error: 2431 on 69 degrees of freedom
Multiple R-squared:  0.358,    Adjusted R-squared:  0.3208
F-statistic:  9.62 on 4 and 69 DF,  p-value: 3.055e-06
```

Este modelo tiene a dos variables que no son significativas, como es el caso de las millas por galón y de la variable que representa el espacio en el maletero.

Se va a generar un gráfico de dispersión entre los términos regresores, para observar cual es el grado de asociación que pueden existir entre los mismos.

```
attach(auto)
pairs(~ mpg + weight + length +trunk)
```

El gráfico resultante será el siguiente:



El gráfico muestra alta correlación entre regresores como **mpg** y **trunk**, **length** y **mpg**; pero principalmente la relación **weight-length** es la que se ve más fuerte.

Con la sospecha de multicolinealidad en el modelo se procede a hacer la estimación del factor inflador de varianza para cada variable. Primero se debe activar el uso de la librería **car**, y luego de eso se hace uso del comando **vif()**.

```
library(car)
vif(mod1)
```

Se mostrarán los siguientes resultados:

```
> vif(mod1)
      mpg    weight    length    trunk 
2.955592 10.363737 11.344351  2.129927
```

Estos resultados darán razón de que las variables **weight** y **length** son las que generan la multicolinealidad. Se podrían plantear soluciones en las cuales se elimine una de estas variables, ya que la pérdida de la capacidad explicativa del modelo no será tan significativa.