

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 *
2              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 2: MULTICOLINEALIDAD II

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 muestra información sobre 2000 mujeres y los determinantes de sus salarios, en el siguiente ejemplo se estimará un modelo para explicar los determinantes del salario por hora, se detectará la multicolinealidad en caso exista.

Desarrollo

El modelo tiene la siguiente forma:

$$wage = \beta_0 + \beta_1 age + \beta_2 education + \beta_3 children + \varepsilon$$

El modelo trata de explicar el salario por hora, con determinantes, como la educación, la edad y el número de niños menores a 12 años que habitan en el hogar.

Se regresa el primer modelo de acuerdo a las variables explicativas dadas:

```
mod1 = lm(wage ~ children + education + age,
          data = womenwk)
summary(mod1)
```

Los resultados de la regresión son los siguientes:

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

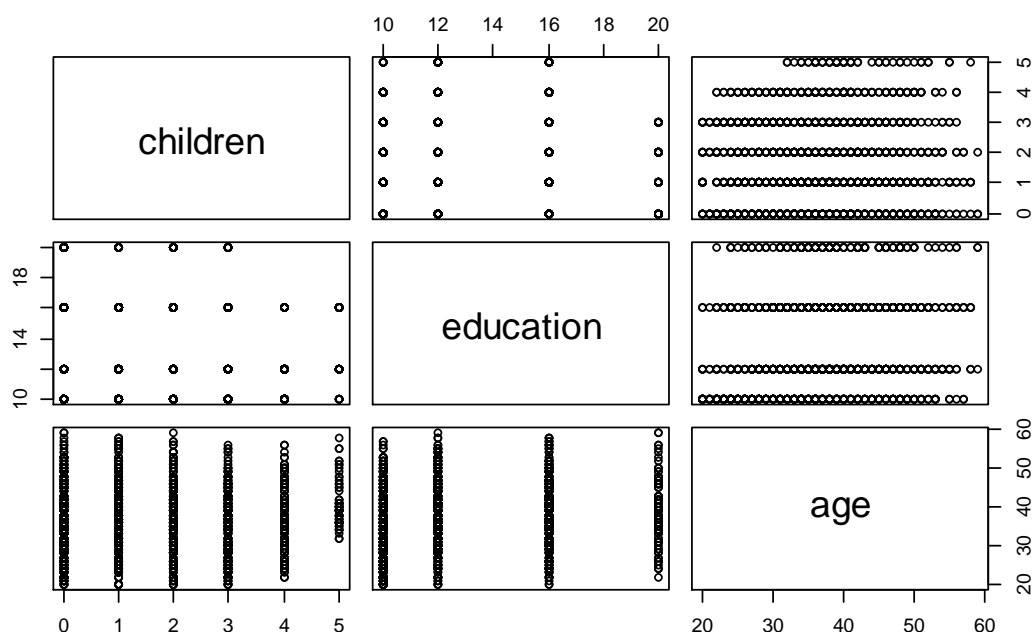
Residuals:
    Min       1Q   Median       3Q      Max
-15.5114  -3.5730  -0.0894   3.5698  19.9133

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  8.01952    0.92517   8.668 < 2e-16 ***
children    -0.66783    0.10255  -6.512 1.04e-10 ***
education     0.86074    0.04936  17.437 < 2e-16 ***
age           0.14298    0.01844   7.754 1.75e-14 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.37 on 1339 degrees of freedom
(657 observations deleted due to missingness)
Multiple R-squared:  0.2764,    Adjusted R-squared:  0.2748
F-statistic: 170.5 on 3 and 1339 DF,  p-value: < 2.2e-16
```

Estos resultados muestran a todos los coeficientes significativos, cosa que también lo es el modelo a nivel global. Se nota una capacidad explicativa muy baja por el bajo valor del coeficiente R-Cuadrado.

Para probar la multicolinealidad, se puede hacer un gráfico de dispersiones que quedará de la siguiente manera:



Este gráfico muestra poca relación entre los regresores, por lo que no habrá sospecha de multicolinealidad, en todo caso para comprobarla de forma más exacta se estima el factor inflador de varianzas para cada variable que será dado por la siguiente expresión.

```

> vif(mod1)
  children education    age
  1.015300  1.050123  1.037999

```

Estos resultados muestran que no existen multicolinealidad entre los términos independientes del modelo planteado.