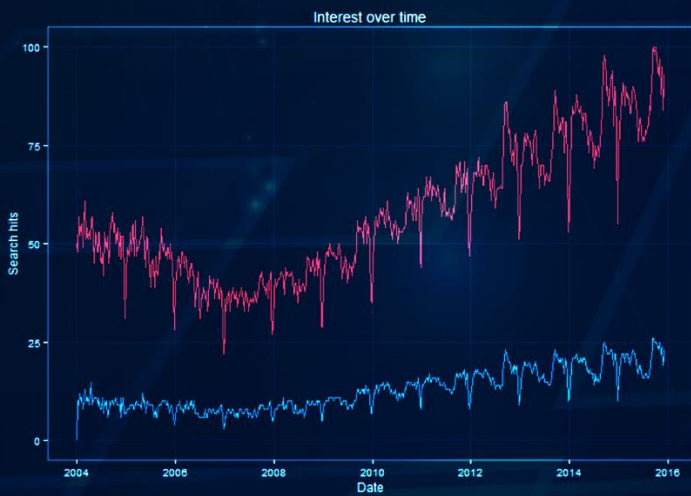


# Econometría Aplicada

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

> |
```



## SESIÓN 7: Error de especificación

## Contenido

Introducción .....	4
Error de especificación .....	5
Omisión de una variable relevante .....	5
Inclusión de una variable irrelevante (sobreajuste del modelo) .....	6
Prueba RESET .....	7
Prueba RESET en R .....	7
Bibliografía .....	9

## Introducción

Cuando se elabora un modelo econométrico es necesaria la evaluación a nivel estadístico del mismo. En esta se hace un análisis de los datos y de los supuestos estadísticos detrás del modelo de regresión lineal. Esta evaluación consta de una variedad de pruebas estadísticas que permiten averiguar si se satisfacen los supuestos.

En la siguiente sesión se verán los problemas asociados a la especificación incorrecta en un modelo de regresión lineal, los problemas que esta ocasiona en los parámetros, y la detección de la misma haciendo uso de R.

## Error de especificación

La especificación incorrecta del modelo puede deberse a una formulación no adecuada de la forma funcional o bien, a que se violan los supuestos del error aleatorio o incluso a la información empírica que se incorpora al modelo para su estimación.

La especificación del modelo incluyendo sus supuestos, conducen a que los estimadores de los parámetros satisfagan propiedades estadísticas deseables, como son el insesgamiento, la eficiencia y la consistencia.

Respecto a los errores de especificación en la selección de las variables explicativas, se tienen dos tipos:

- Omisión de variables relevante (subajuste del modelo).
- Inclusión de variables innecesarias (sobreajuste del modelo).

### Omisión de una variable relevante

---

Suponiendo que se tiene el siguiente modelo, y este es verdadero:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$$

Pero el modelo que se estima es:

$$y = \alpha_0 + \alpha_1 x_1 + v$$

Las consecuencias de omitir  $x_2$ , pueden ser:

Si la variable excluida está correlacionada con la variable que se ha incluido, el coeficiente de correlación entre las dos variables es diferente de cero, por lo tanto, los estimadores del modelo con variable la variable omitida ( $\alpha_0$  y  $\alpha_1$ ) serán sesgados e inconsistentes. Esto quiere decir,  $E(\hat{\alpha}_0)$  no es igual a  $\beta_0$  y  $E(\hat{\alpha}_1)$  no es igual a  $\beta_1$ , y el sesgo no desaparece conforme aumenta el tamaño de la muestra.

Aunque las variables del modelo verdadero no están correlacionadas, la constante eliminada será sesgada, pese a que el valor de  $\alpha_1$  sea ahora insesgado.

La varianza del término de error  $\sigma^2$  está incorrectamente estimada.

Es probable que el intervalo de confianza usual y los procedimientos de pruebas hipótesis conduzcan a conclusiones equivocadas sobre la significancia estadística de los parámetros estimados.

Otra consecuencia es que los valores pronosticados que se hacen de la variable explicada en dichos modelos sean incorrectos, ya que el estimado es inconsistente y sesgado, y los intervalos de confianza del pronóstico no son confiables.

### Inclusión de una variable irrelevante (sobreajuste del modelo)

Suponiendo que se tiene el siguiente modelo, y este es verdadero:

$$y = \alpha_0 + \alpha_1 x_1 + v$$

Pero el modelo que se estima es:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$$

Se está cometiendo un error de especificación al incluir una variable innecesaria en el modelo. Las consecuencias de dicho error son las siguientes:

Todos los estimadores de MCO de los parámetros del modelo incorrecto son sesgados e inconsistentes, o sea, los parámetros estimados del modelo incorrecto serán estadísticamente iguales a cero.

Si bien la varianza estará correctamente estimada y las pruebas de hipótesis conservan su validez los coeficientes estimados por lo general serán ineficientes, es decir, sus varianzas generalmente serán más grandes que las del verdadero modelo.

## Prueba RESET

Para probar la existencia de un error de especificación en el modelo planteado se hace uso de la prueba RESET.

Esta prueba, cuyas siglas quieren decir *Regression Equation Specification Error Test*, planteada por Ramsey (1969), sirve para detectar errores de especificación ocasionados por la omisión de variables independientes, por la existencia de correlación entre los regresores y el error, o por la forma funcional incorrecta.

Considerando el siguiente modelo:

$$y = \beta_0 + \sum_{i=1}^k \beta_i x_k$$

Este test va a probar si  $(\beta x)^2, (\beta x)^3, \dots, (\beta x)^k$  tiene algún poder para explicar a  $y$ , esto se hará realizando la siguiente regresión lineal:

$$y = \beta_0 + \sum_{i=1}^k \beta_i x_i + \sum_{i=1}^{m+1} \alpha_i \hat{y}^i + v_i$$

Donde  $m$  es el número de potencias de la variable  $y$ . Si la hipótesis de que todos los coeficientes que acompañan a  $\hat{y}$  son cero es rechazada, entonces el modelo sufre de errores de especificación.

Por tanto, la prueba RESET es utilizada para analizar si el modelo está bien o mal especificado, donde la hipótesis viene dada por lo siguiente:

$H_0$ : Forma funcional correcta

$H_1$ : Forma funcional incorrecta

La prueba para analizar esta hipótesis sigue una distribución F. Esta prueba se realiza una vez que se ha estimado el modelo.

### Prueba RESET en R

Para realizar la prueba RESET en R, se requiere el paquete **lmtest** y una vez regresionado el modelo, se realizar la prueba haciendo uso del comando **resettest()**.

```
attach(mroz)
library(lmtest)
mod1 = lm(lwage ~ educ + exper , data = mroz)
summary(mod1)
```

Los resultados de la regresión serán los siguientes:

**Call:**

```
lm(formula = lwage ~ educ + exper, data = mroz)
```

Residuals:

Min	1Q	Median	3Q	Max
-3.05608	-0.30524	0.05599	0.38846	2.27384

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-0.400174	0.190368	-2.102	0.036132 *
educ	0.109489	0.014167	7.728	7.94e-14 ***
exper	0.015674	0.004019	3.900	0.000112 ***

---

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

Residual standard error: 0.669 on 425 degrees of freedom  
(325 observations deleted due to missingness)

Multiple R-squared: 0.1484, Adjusted R-squared: 0.1444

F-statistic: 37.02 on 2 and 425 DF, p-value: 1.512e-15

Estos resultados muestran que el modelo es estadísticamente significativo de forma conjunta y también de forma individual para cada uno de sus parámetros.

Para probar problemas de errores de especificación, se aplica el test RESET de Ramsey, con el comando **resettest()**.

```
resettest(mod1)
```

RESET test

data: mod1

RESET = 0.88725, df1 = 2, df2 = 423, p-value = 0.4126

De los resultados se puede entender que como la hipótesis nula es aceptada, por lo tanto, no se puede detectar una especificación errónea en dicho modelo estimado.



## Bibliografía

- Kleiber, C. & Zeileis, A. (2008). *Applied econometrics with R*. Springer Science & Business Media.
- Quintana, L. & Mendoza, A. (2016). *Econometria aplicada usando R*. Universidad Nacional Autónoma de México.