

Este taller se divide en dos secciones, en la primera se trabajará lo relacionado a la validación del modelo. Posterior a esto, se considera un ejercicio en el que se realiza la prueba de falta de ajuste a un modelo.

En primer lugar considere el siguiente conjunto de datos.

Cuadro 1: Presentación de los datos

y	x
36.460743	5.878002
6.075999	-0.624800
47.402151	-3.727324
62.778036	-4.232765
28.971238	-2.318757
6.675766	3.712115

El día de hoy, la misión será realizar los siguientes ejercicios, claro está, haciendo uso de R.

1. Genere la base de datos que se muestra previamente usando el siguiente código.

```
gen_dat <- function(n, seed = 7) {  
  varianza <- 16  
  set.seed(seed)  
  x <- runif(n=n, min=-5, max=6)  
  media <- 4 - 6 * x + 2 * x^2  
  set.seed(seed^2)  
  y <- rnorm(n=n, mean=media, sd=sqrt(varianza))  
  marco_datos <- data.frame(y=y, x=x)  
  return(marco_datos)  
}  
  
datos <- gen_dat(75)
```

2. Ajuste el modelo de regresión lineal simple

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \varepsilon_i \stackrel{iid}{\sim} N(0, \sigma^2); 1 \leq i \leq 75$$

3. Determine que parámetros son significativos y cuales no en el modelo, hágalo de manera rápida aprovechando alguna de las funciones de R usadas hasta el momento.

4. Extraiga los residuales del modelo y verifique que estos tengan media igual a 0, dé un argumento de por qué este supuesto siempre se cumple.
5. Determine si los residuales tienen varianza constante, argumente por qué esto es o no es así, además, si nota algún patrón o algo que considere anormal, coméntelo.
6. Evalúe el supuesto de normalidad de los residuales, hágalo usando un histograma, un gráfico cuantil - cuantil y finalmente una prueba de hipótesis.
7. Finalmente verifique si los residuales son o no independientes, hágalo de manera gráfica. Los valores de las variables están ingresados en la base de datos por orden cronológico.
8. Con la base de datos `table.b3` del paquete **MPV**, realice la prueba de falta de ajuste, del modelo

$$y_i = \beta_0 + \beta_1 x_{4i} + \varepsilon_i, \varepsilon_i \stackrel{iid}{\sim} N(0, \sigma^2); 1 \leq i \leq 32$$

para ello use la función **rsm** del paquete **rsm**.

Nota: se propone como ejercicio realizar la validación del modelo.