

## Hoja de Trabajo – Regresión

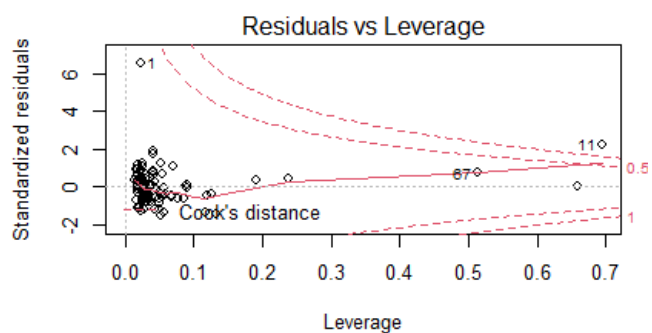
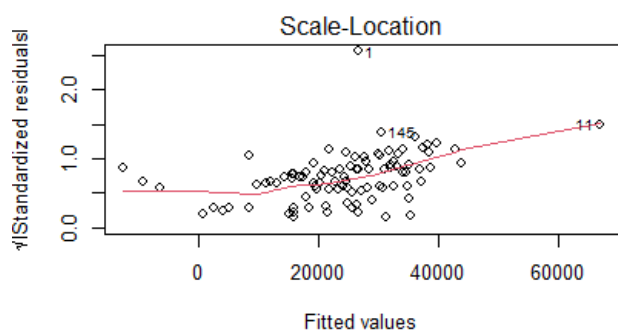
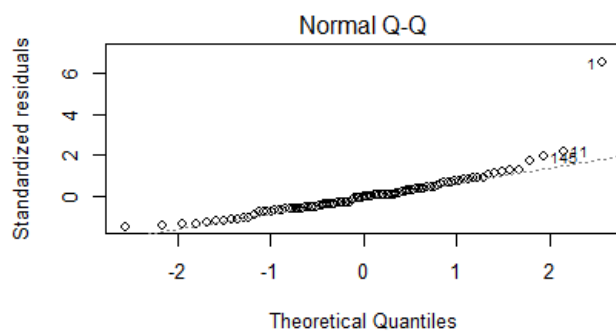
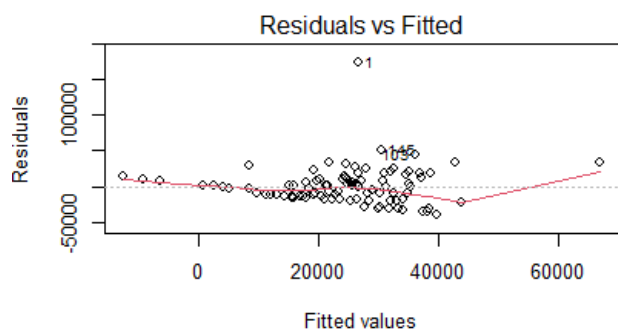
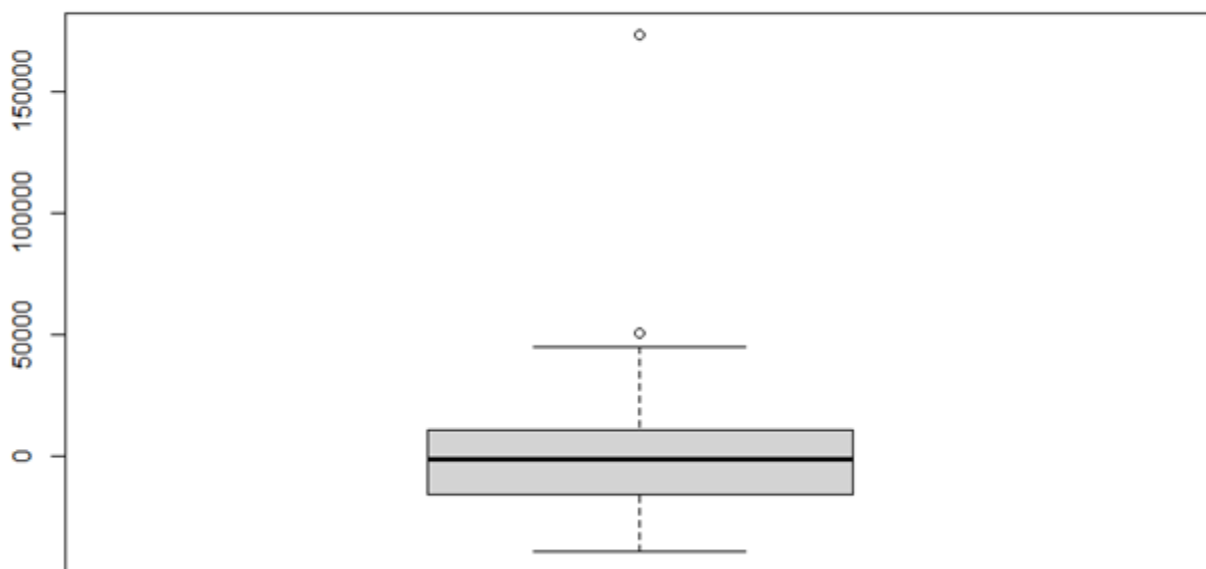
En la siguiente hoja de deberá utilizar las fuentes de datos indicadas para analizar la información usando R

Para cada uno de los ejercicios:

- Suba el archivo .R indicado
- Realice un análisis basado en los resultados estimados, el plot con los gráficos de aceptación y el RMSE obtenido.

### 1. Ejercicio # 1

- Utilice los métodos de regresión aprendidos en clase para pronosticar cuantas veces será compartido un artículo en redes sociales.
- Cree un nuevo script .R llamado “Regresión”
- Lea el archivo subido al portal llamado “articulos\_ml”
  - Title: Titulo del Artículo
  - url: ruta al artículo
  - Word count: la cantidad de palabras del artículo,
  - # of Links: los enlaces externos que contiene,
  - # of comments: cantidad de comentarios,
  - # Images video: suma de imágenes (o videos),
  - Elapsed days: la cantidad de días transcurridos (al momento de crear el archivo)
  - # Shares: nuestra columna de salida que será la “cantidad de veces que se compartió el artículo”.
- Debe de crear y adjuntar el plot que muestre la compatibilidad del modelo con el algoritmo de regresión lineal.



# **Análisis**

## **Residual-Location**

La gráfica determina la varianza no constante de la distribución de errores. Al observar la composición de los puntos, se puede observar que se ajusta a la línea hasta cierta medida, esto nos da la alerta de que probablemente no exista un comportamiento lineal, por lo que el modelo no sería útil.

## **Normal Q-Q**

Esta gráfica determina si tiene una distribución normal. Al observar el ajuste de los puntos, no cabe duda de que el comportamiento de los datos persigue una distribución normal.

## **Scale-Location**

La gráfica representa los residuos estandarizados. Se observa que los datos tienen comportamiento similar.

## **Residuals-Leverage**

El objetivo de la gráfica es detectar influenciadores en los modelos, basados en la distancia de cook. Define una frontera entre elementos que ejercen una influencia superior en el modelo. En este caso en específico, se detectan pocos influenciadores importantes, por lo que es poco probable que puedan sesgar el modelo.