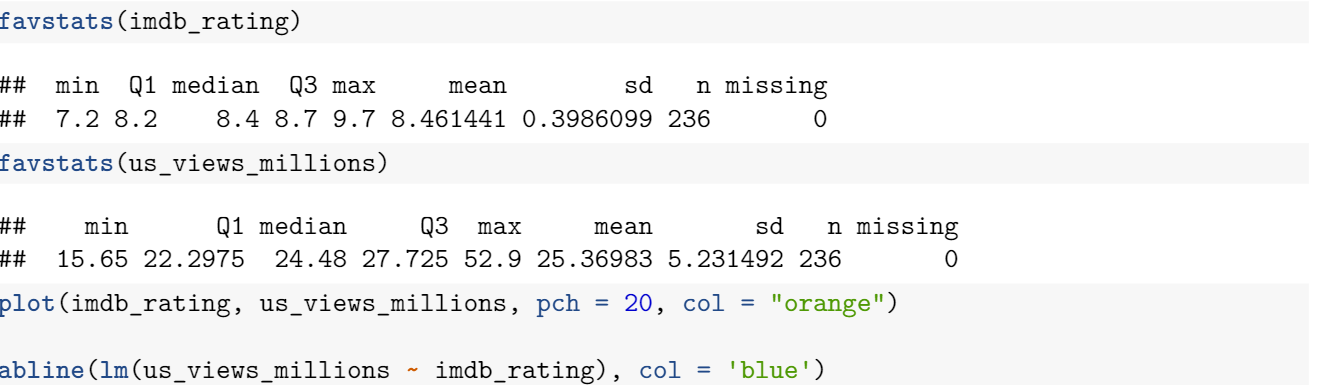
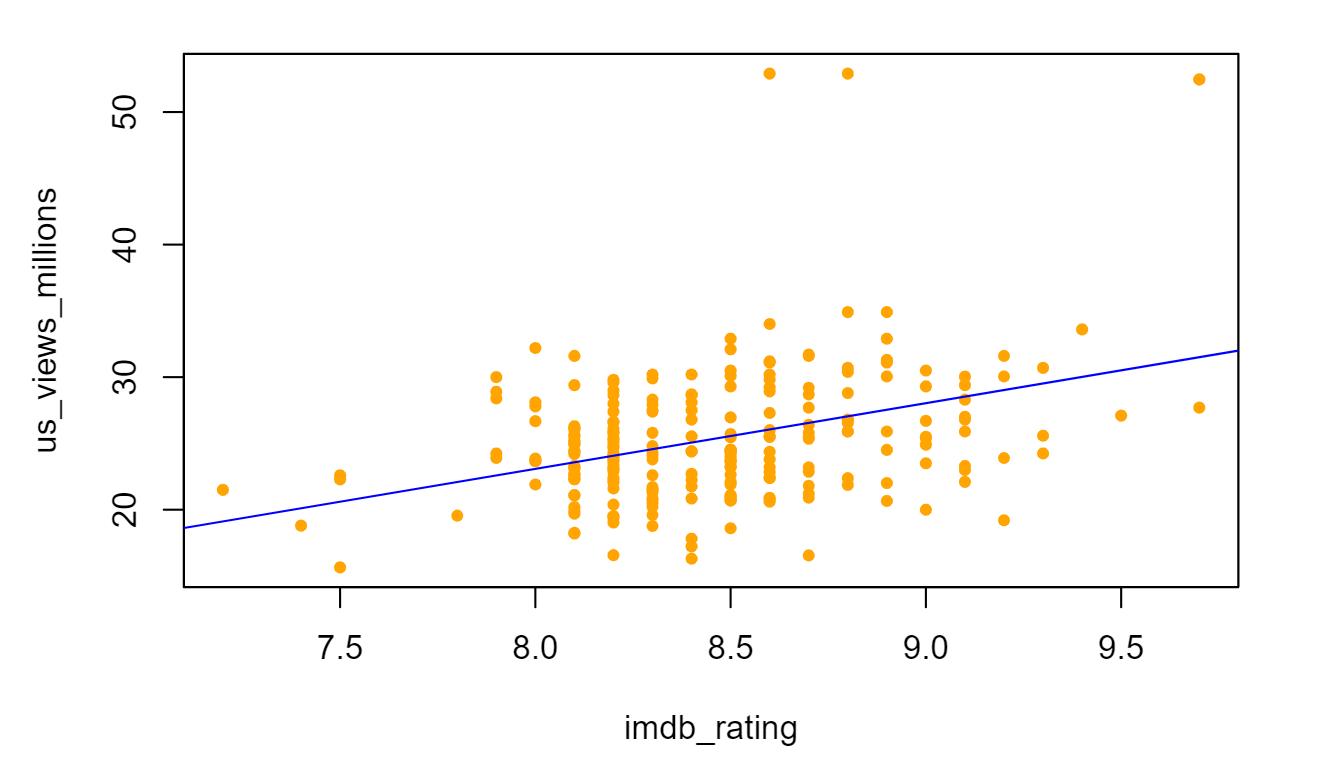
# Step 3: Statistical Inference Report

Khadeeja Naseer

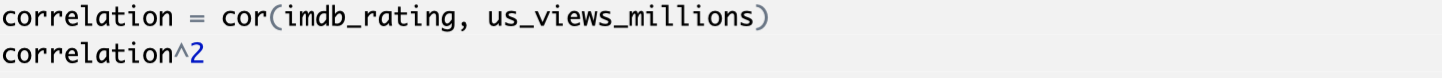
Ruby Sun

**Research Question:** Is there an association between the imdb rating of an episode and the number of people who watched the episode when it aired?

****



The scatter plot of imdb\_rating vs us\_views\_millions suggests a slight positive relationship between imdb\_rating and us\_views\_millions. This means that higher levels of IMDB rating are associated with a larger number of people who watched the episode when it aired. As IMDB rating increases, the number of viewers tend to increase. The regression equation is ŷ = -16.5421 + 4.9533x, which means that when the IMDB rating increases by 1, the number of American viewers is expected to increase by about 4.95. As the minimum IMDB rating in this context is 7.2, -16.5421 is simply the value at which the regression line intersects the y-axis.





About 14% of the variation in the number of American viewers is accounted for by variation in IMDB rating.

**Simple Linear Regression**

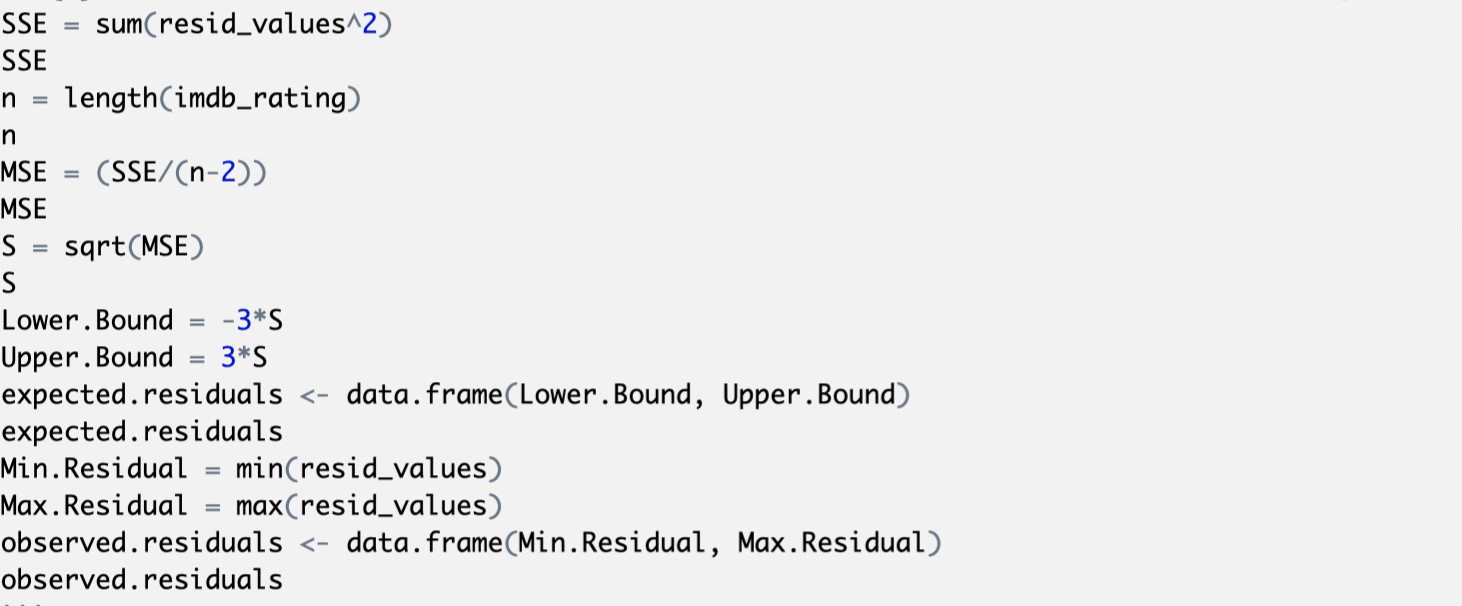
**I)**

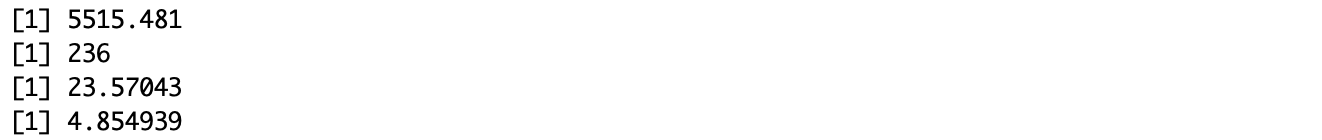
Ho: imdb\_rating contributes no information for predicting us\_views\_millions. (β=0)

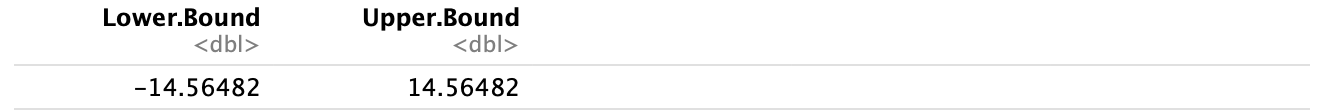
Ha: imdb\_rating is useful for predicting us\_views\_millions. (β≠0)

**II)**

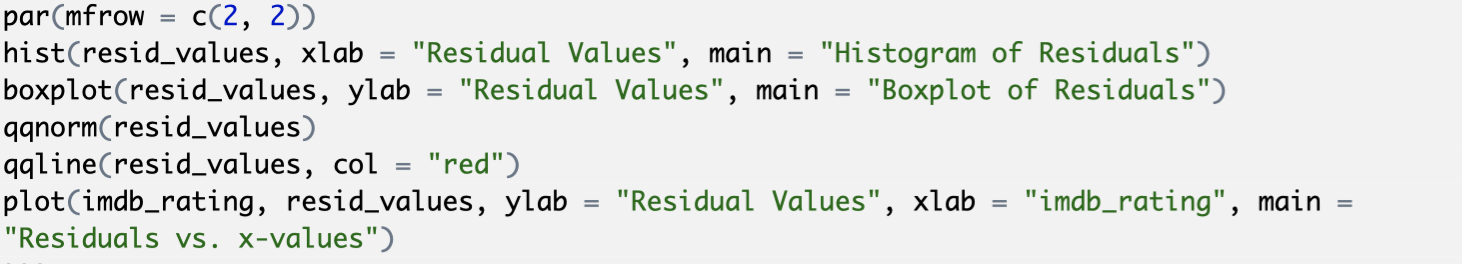
**Assumptions Check**

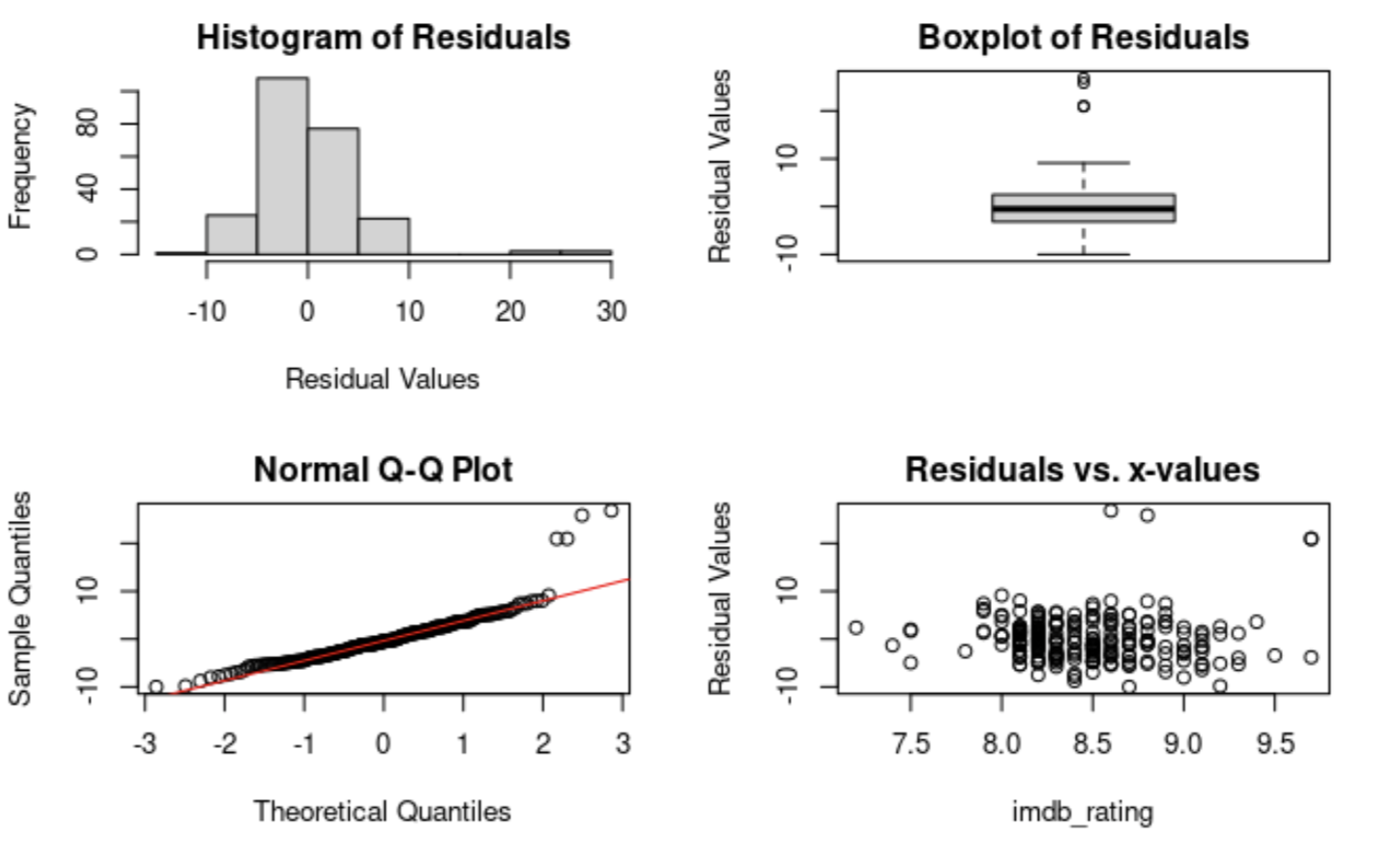
****

****

****

****





**Histogram of Residuals**: The histogram plot of residuals show that residuals appear to have approximate a unimodal, relative symmetric distribution. Thus, the assumption of normality of random errors is met.

**Boxplot of Residuals**: The boxplot of residuals show that residuals appear to have approximate a normal distribution. Although there are a few residual points that are plotted individually, the overall pattern is symmetric about median. Thus, the assumption of normality of random errors is met.

**Normal Q-Q Plot**: The normal probability plot of residuals show that residuals appear to have an approximate normal distribution. Although, there are some residual points that are plotted away from the overall pattern, the majority of residual points are normally distributed. Thus, the assumption of normality of residuals is met.

**Residuals vs. x-values**: Residual points are randomly located around the zero line. Thus, the assumption about random errors having mean zero is met. No obvious pattern is observed. Thus, the assumption about random errors having constant variance is met.

**III)**

t = = = 6.234, df = 234

**IV)**

p-value = 2.1 x ≈ 0.000

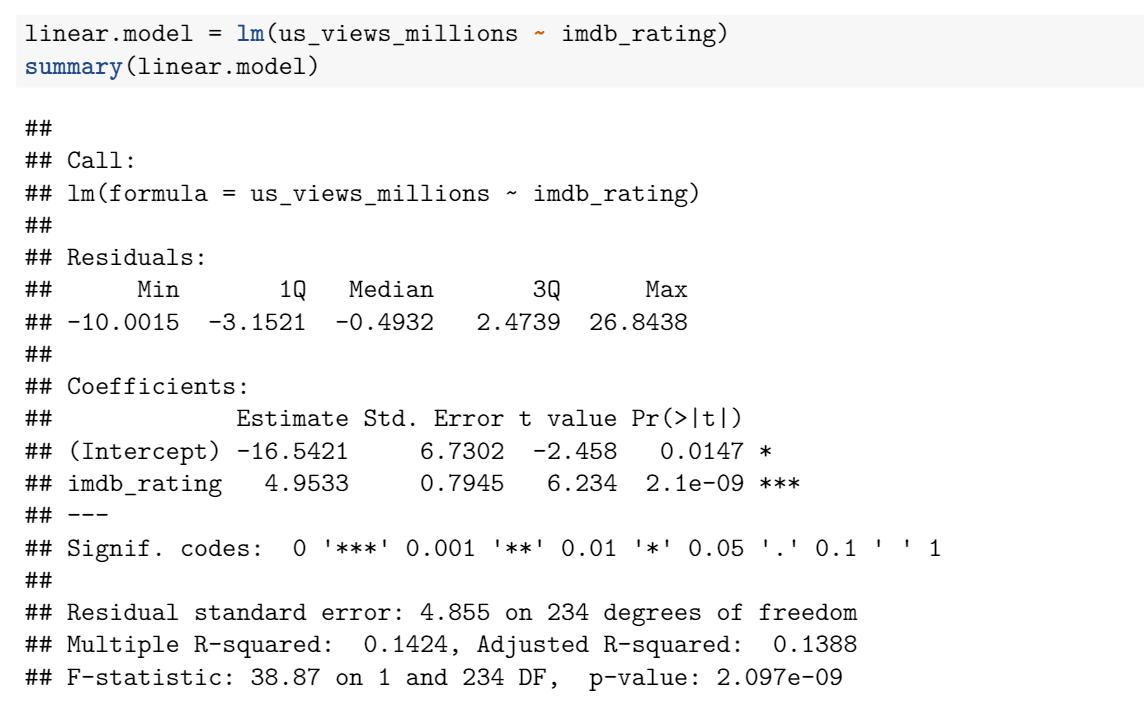
**V)**

p-value < α = 0.05

Reject Ho. We have a strong evidence to claim that imdb\_rating is useful for predicting us\_views\_millions.

**VI)**

We can conclude that the higher imdb rating an episode has, the higher US views it received when it aired, on average.

****