

Q1.4

- To assess whether the distribution of the errors is Homoscedastic, there are two ways. One is through the “scale-location” plot and the second way is through the Breusch-Pagan test (bp test). Homoscedasticity would show up on the scale-location plot, if the red line on the plot is roughly horizontal. From Figure 1 plot, we see that the spread of the residuals is roughly equal at all the fitted values, highlighting evidences for Homoscedasticity. This plot does not show any strong evidence for heteroscedasticity.

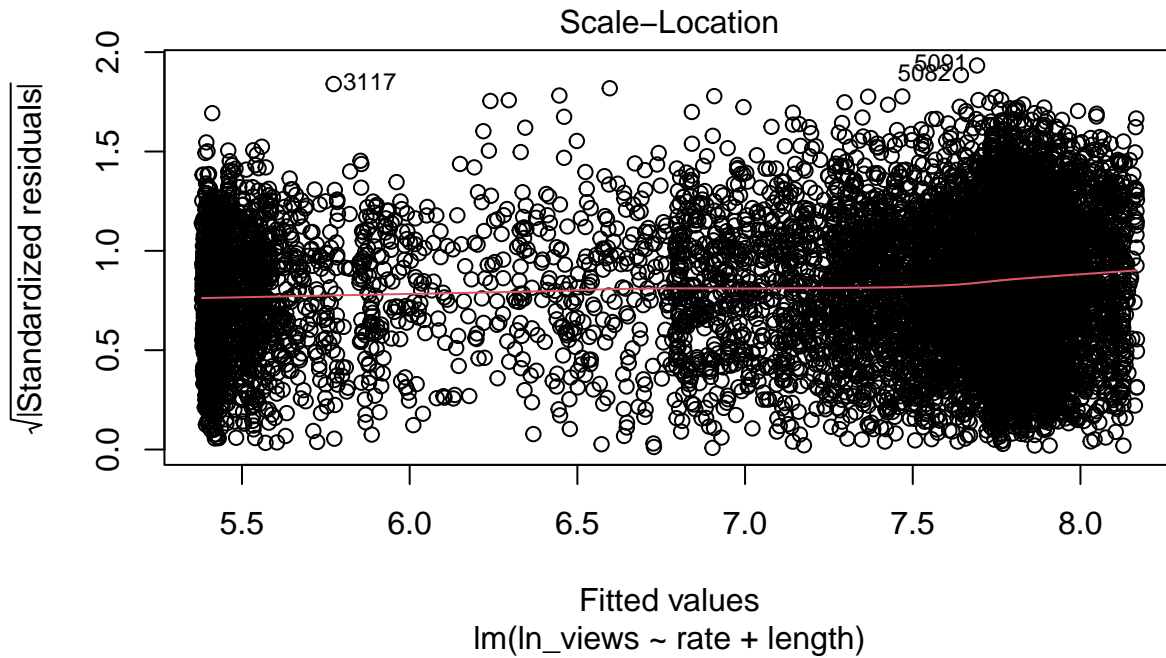


Figure 1: Predicted Vs Residuals

- The `bptest()` function in R is used to perform the Breusch-Pagan test. The null hypothesis (H_0) of the Breusch-Pagan test is that the residuals are Homoscedastic. The alternate hypothesis (H_a) is that the residuals are heteroscedastic. If the p-value from the test is less than the chosen significance level (e.g., 0.05), we can reject the null hypothesis of homoscedasticity. From the `bptest` results, it is seen that the p-value ($2.2e-161$) is less than 0.05 and we can reject the null hypothesis and conclude that there is strong evidence of heteroscedasticity in the linear regression model.
- The Breusch-Pagan test is a statistical test and a sensitive test and no statistical tests can give perfect results. In this case, the Breusch-Pagan test shows evidences of heteroskedasticity. In this case, due to the large sample size, we should still continue with the model and use effective solutions like robust standard errors to handle heteroskedasticity.
- Note: The columns having null values are dropped in the dataset. To make the visualization better, the length column in seconds is transformed to minutes. The median views for videos greater than 11 min and videos smaller than 11 mins are very different. This is evident from histogram distribution of the video length variable which is right skewed, hence, the video length greater than 11 min are not significant and are dropped from dataset.

Breusch-Pagan Test Results

```
## studentized Breusch-Pagan test data: model BP = 124.91, df = 2, p-value < 2.2e-16
```