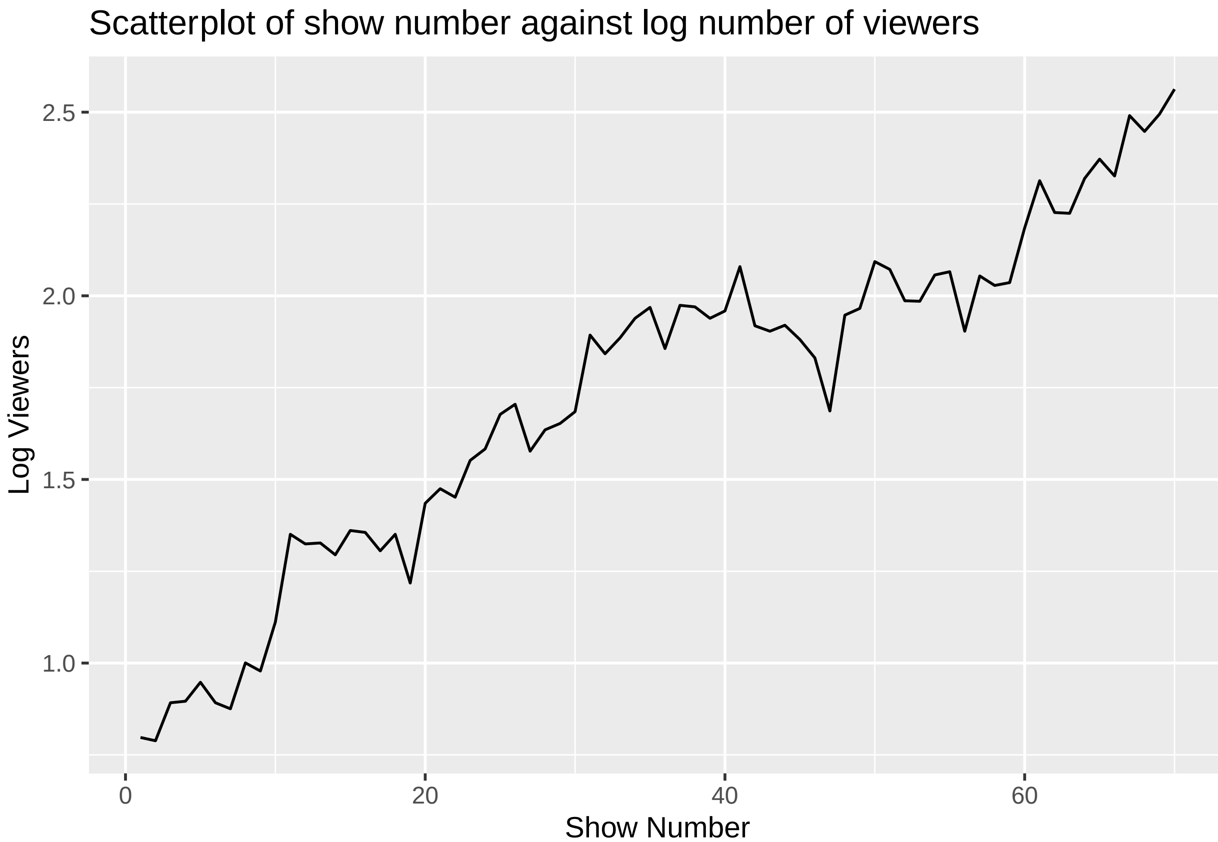
# TV Viewership

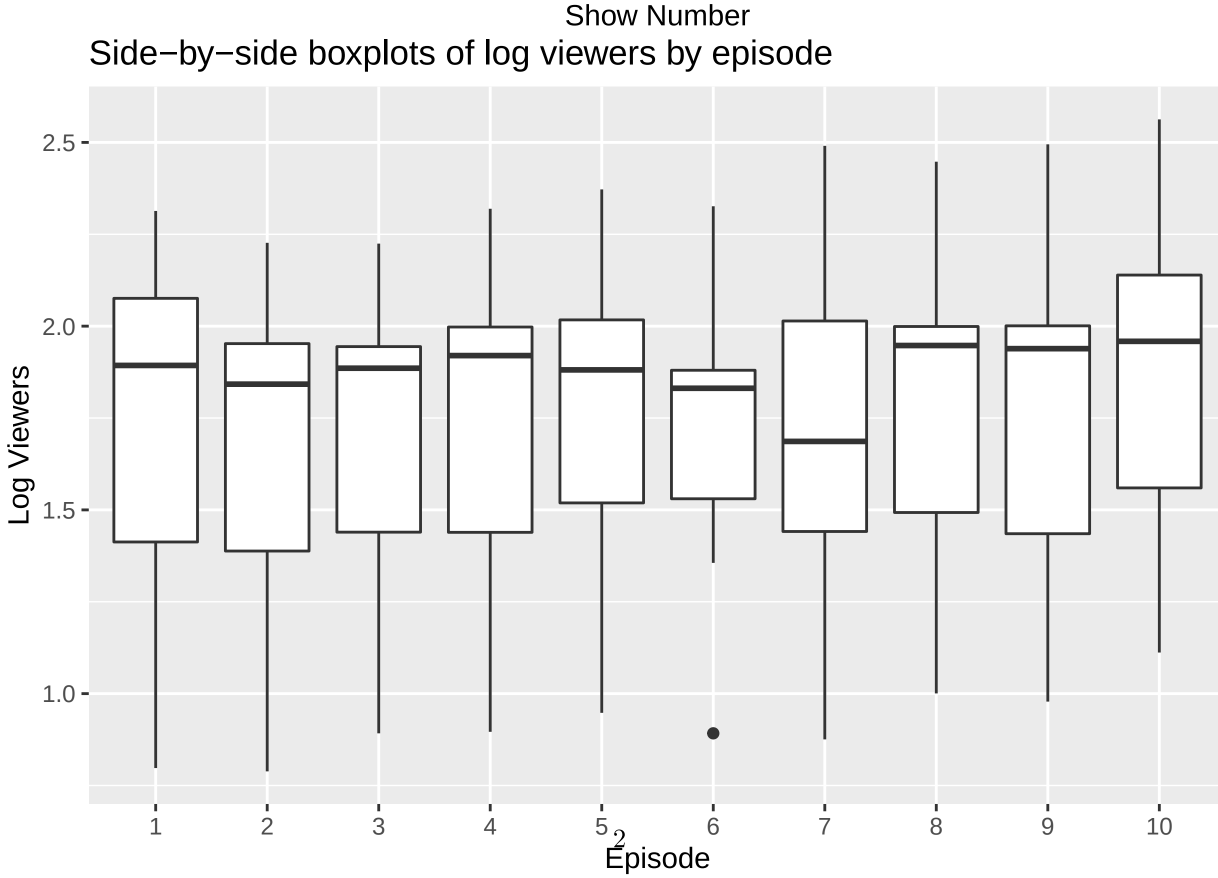
Travis Andersen

**0.**

**1.**



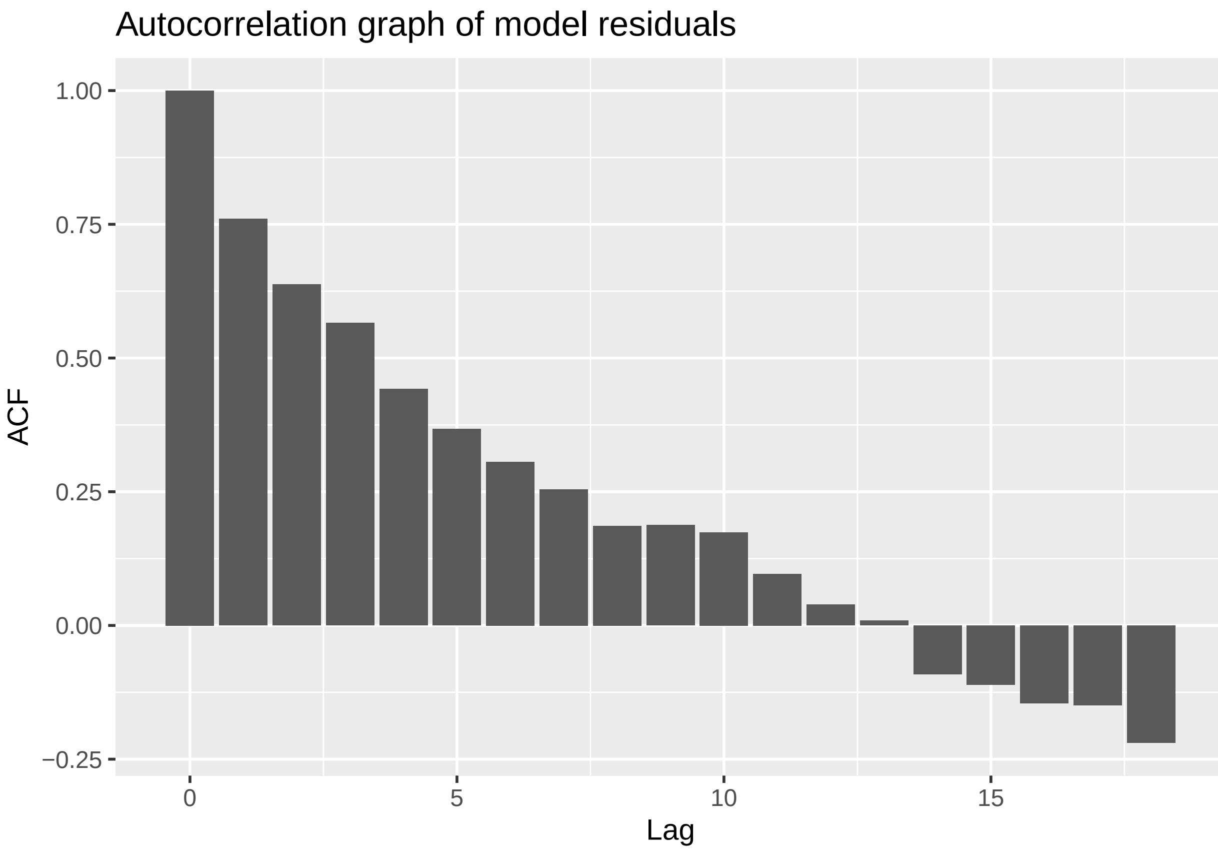
Scatterplot of show number against log number of viewers
2.5
2.0
1.5
1.0
0
20
40
60
Show Number
Log Viewers

Side−by−side boxplots of log viewers by episode
2.5
2.0
1.5
1.0
1
2
3
4
5
6
7
8
9
10
Episode
Log Viewers

From the line graph, we see that as the show goes on, log viewership increases. From the side-by-side boxplots, it seems that the episode number does not affect log viewership. The correlation between show number and log viewership is 0.9556, which is very high.

## 2.




To see if temporal correlation exists in the residuals, the autocorrelation in the residuals must be checked. From the graph above, we see that there is a significant amount of autocorrelation in the residuals. We can see that the amount of viewers who saw the last few episodes significantly affects the viewership of the current episode.

## 3.

To choose the appropriate order for the time series model, all SARIMA models with seasonality 10 and with $p \in 0, 1, 2$, $q \in 0, 1, 2$, $P \in 0, 1$, and $Q \in 0, 1$ were compared, and the model with the lowest aic was chosen. The SARIMA model which minimized aic was a SARIMA(2, 0, 0)(0, 1, 1) model.

## 4.

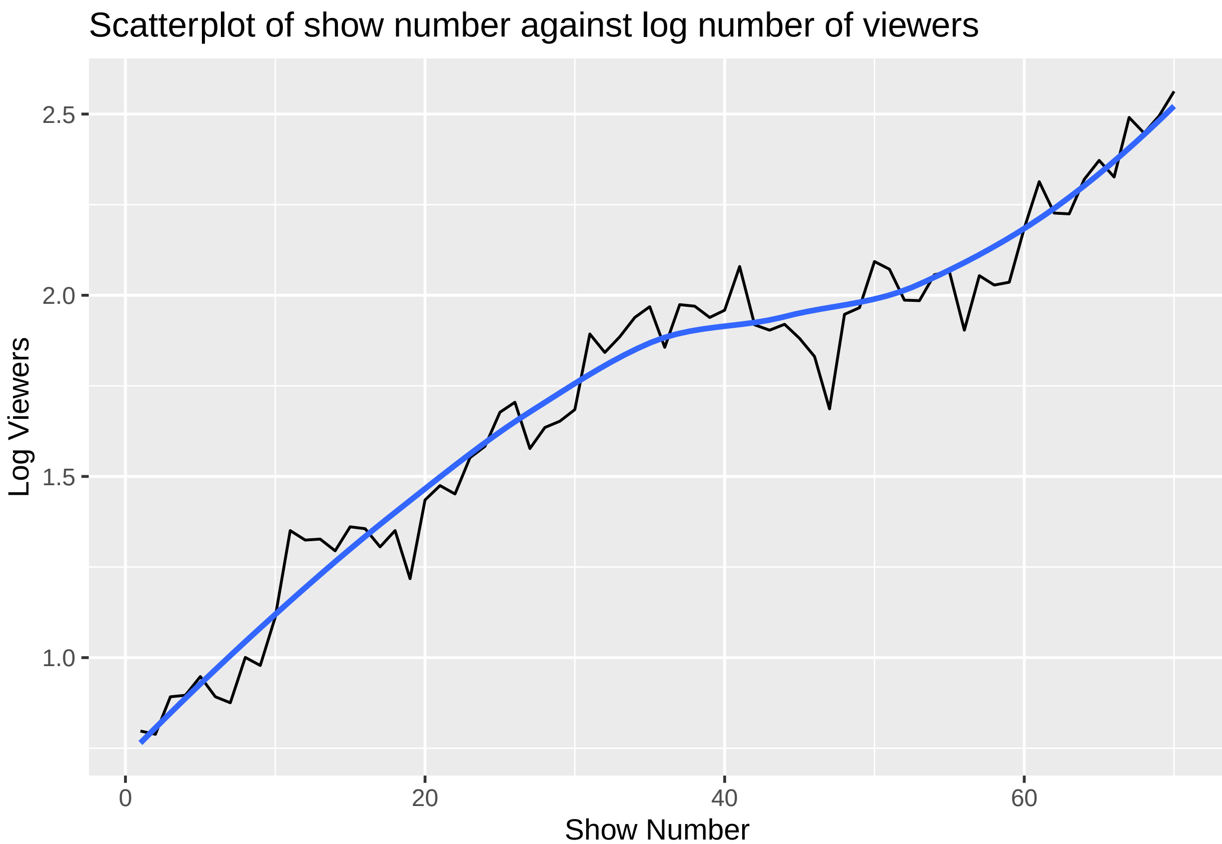
The model can be written as follows:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

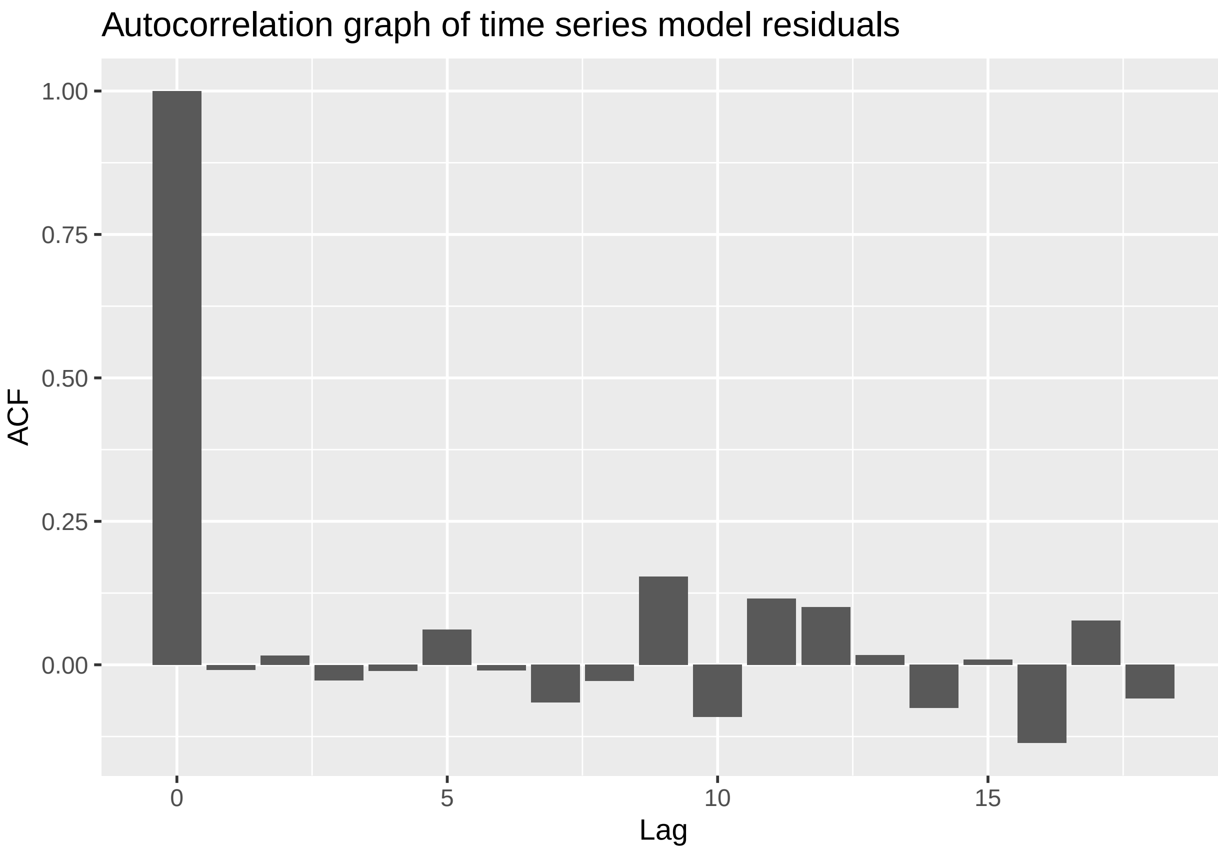
$$\epsilon \sim SARIMA(2,0,0)(0,1,1)_{10}$$

$\boldsymbol{\beta}$ is a vector containing the parameters $\beta_0$ and $\beta_1$. $\beta_0$ is the model parameter for the intercept and $\beta_1$ is the model parameter for the coefficient of the show number variable. $\epsilon$ follows a $SARIMA(2,0,0)(0,1,1)_{10}$ model. The model has a seasonality period of 10. It calculated the residuals based on the residuals of the last two episodes, and the unique part of the last episode, all after differencing the residuals by the residual of the same episode from the last season. We can predict the viewership moving forward by doing one-step ahead forecasts using the predicted residuals from the future, which will eventually converge to zero.

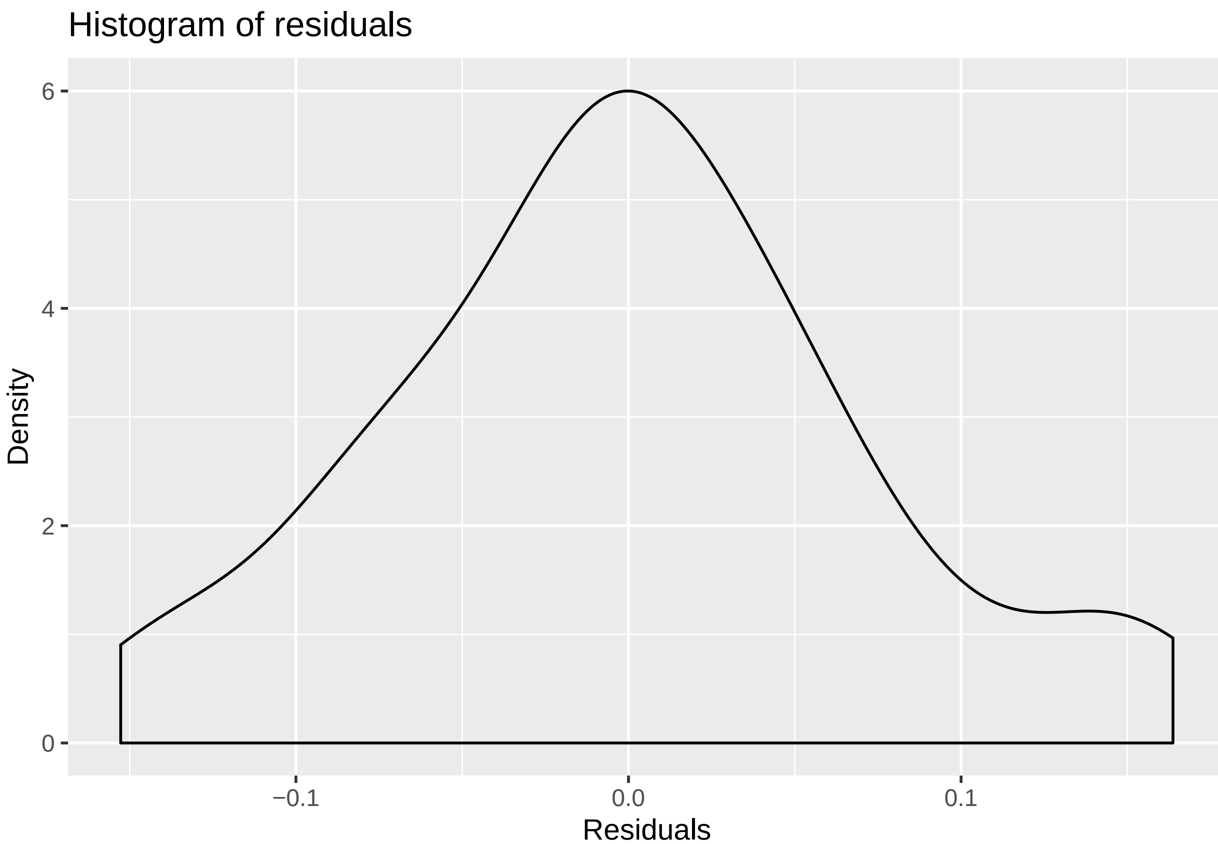
## 5.



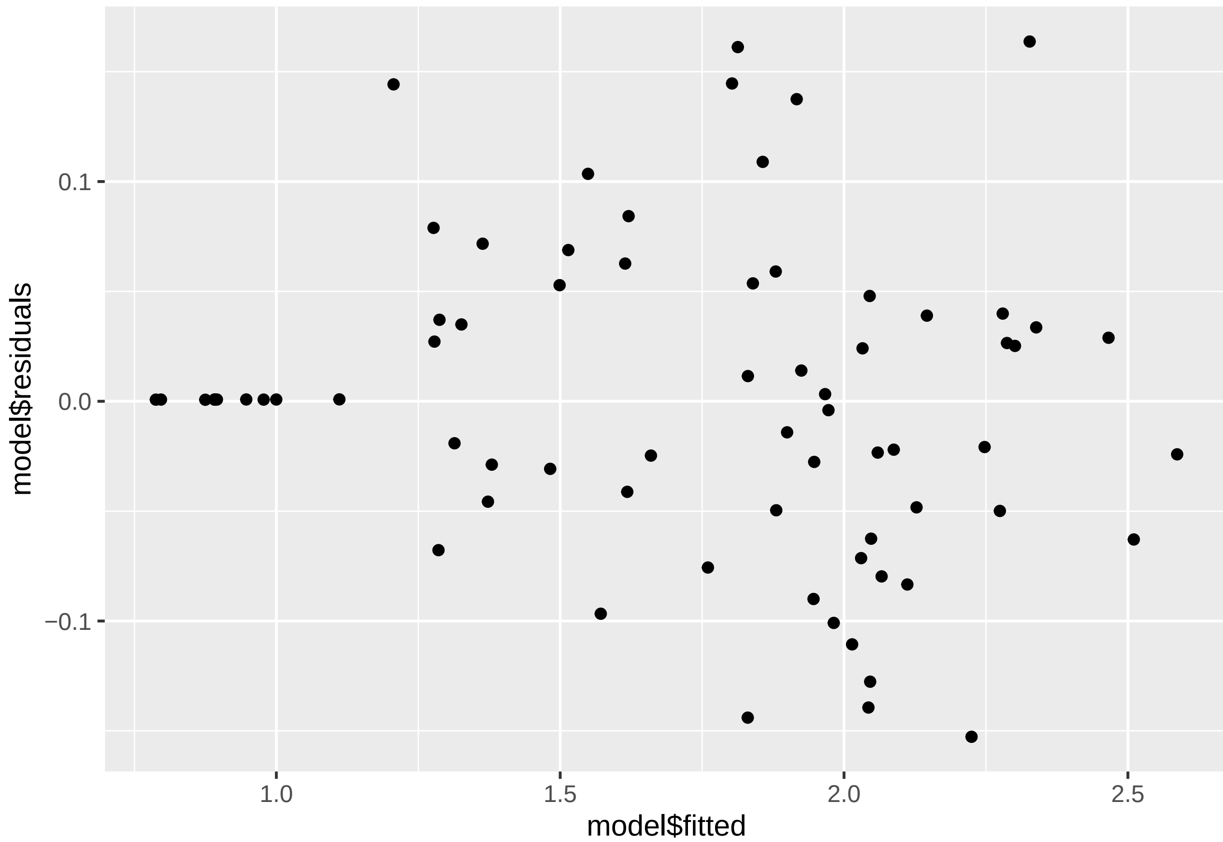

From the line graph of the time series, we see that there is a fairly linear relationship.




From an autocorrelation graph of the residuals from the time series model, we can see that there is no longer a significant correlation between the residuals. Therefore, the assumption of independence is met.

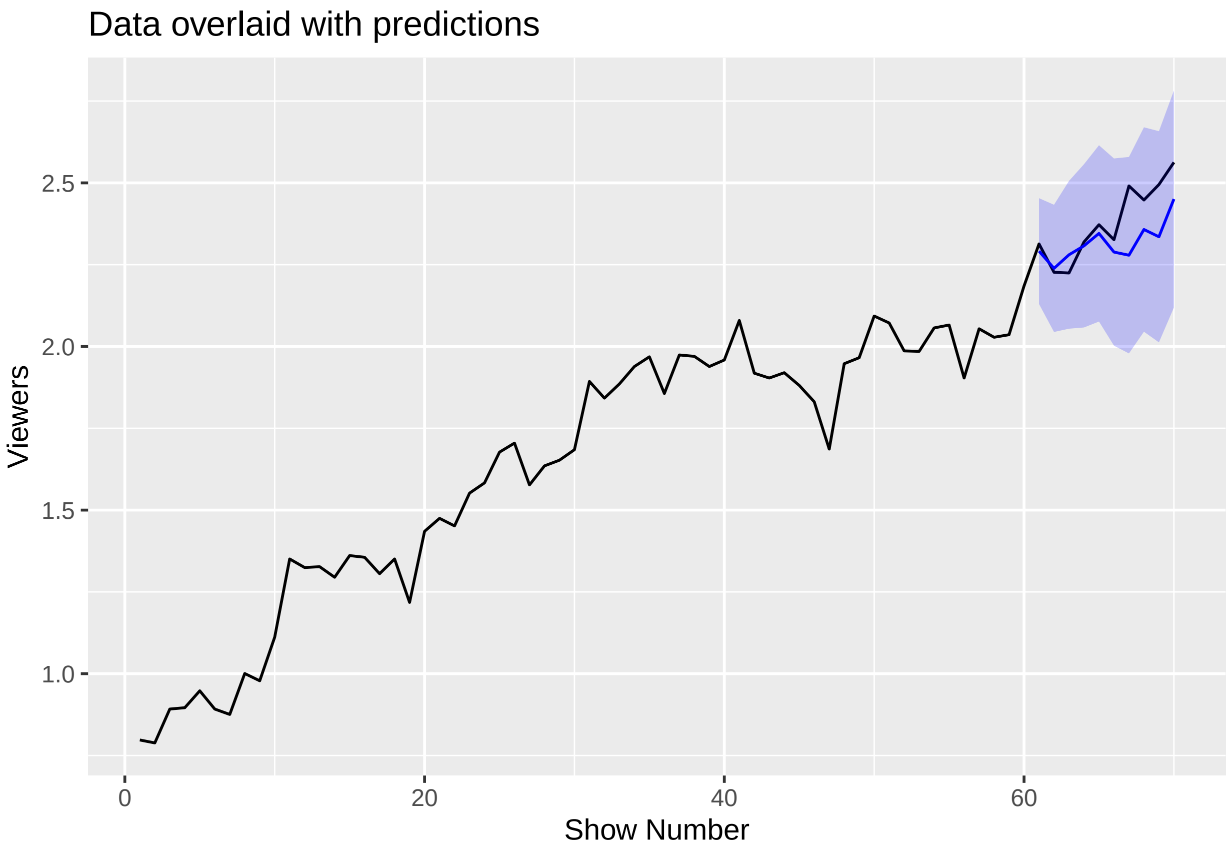



From the density plot above, we see that the residuals follow an approximately normal distribution.




Disregarding the fitted values from the first season where they are fit exactly, the variance of the residuals does not depend on the fitted value amount, so equal variance is met.

## 6.




A cross validation of the model over the last season of data was performed, and the calculated root predictive mean square error over the test period was 0.0985.

## 7.

A t-test with $H_0 : \beta_{ShowNum} = 0$ and $H_A : \beta_{ShowNum} > 0$ returns a p-value of 0. A 95% confidence interval for $\beta_{ShowNum}$ is (0.0169, 0.034). Therefore, we conclude that viewership is increasing.

## 8.

Table 1: Season 8 viewership forecast

| Point Forecast | Lo 95 | Hi 95 | Episode |
|---:|---:|---:|---:|
| 2.6845 | 2.5263 | 2.8426 | 1 |
| 2.6070 | 2.4195 | 2.7945 | 2 |
| 2.6314 | 2.4143 | 2.8485 | 3 |
| 2.6763 | 2.4378 | 2.9148 | 4 |
| 2.7095 | 2.4530 | 2.9659 | 5 |
| 2.6434 | 2.3721 | 2.9147 | 6 |
| 2.6822 | 2.3983 | 2.9660 | 7 |
| 2.7283 | 2.4338 | 3.0228 | 8 |

| Point Forecast | Lo 95 | Hi 95 | Episode |
|---|---|---|---|
| 2.7258 | 2.4222 | 3.0294 | 9 |
| 2.8283 | 2.5169 | 3.1396 | 10 |

Executives can use these forecasts to gauge if the show should continue for a ninth season. These correspond to a percentage increase in viewership.