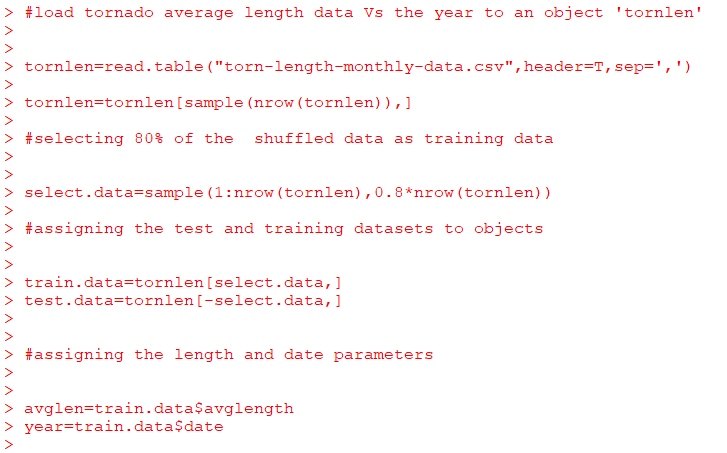
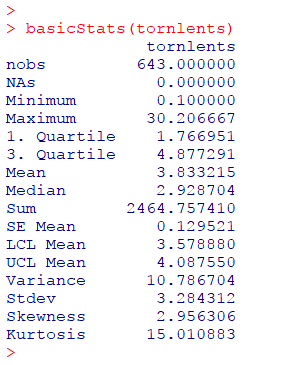
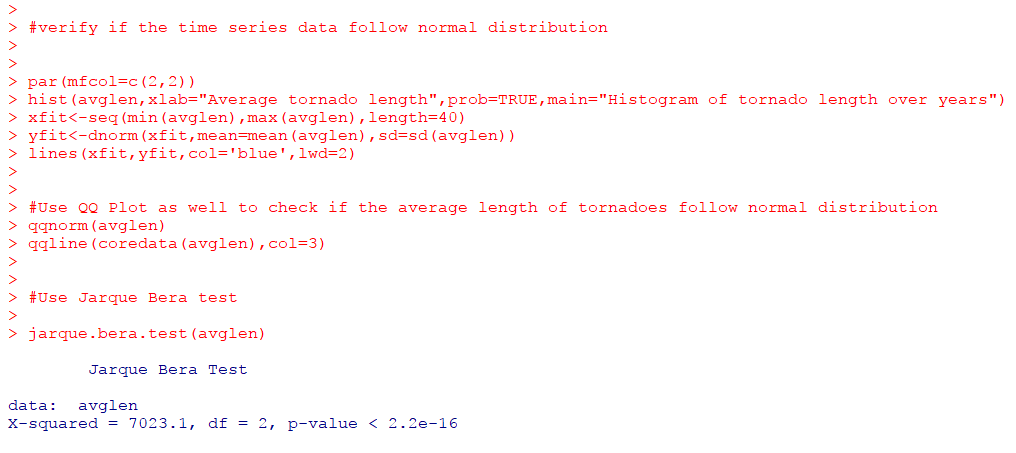
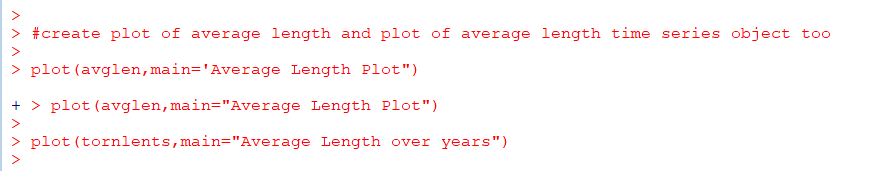
**Time series code and output screenshots**

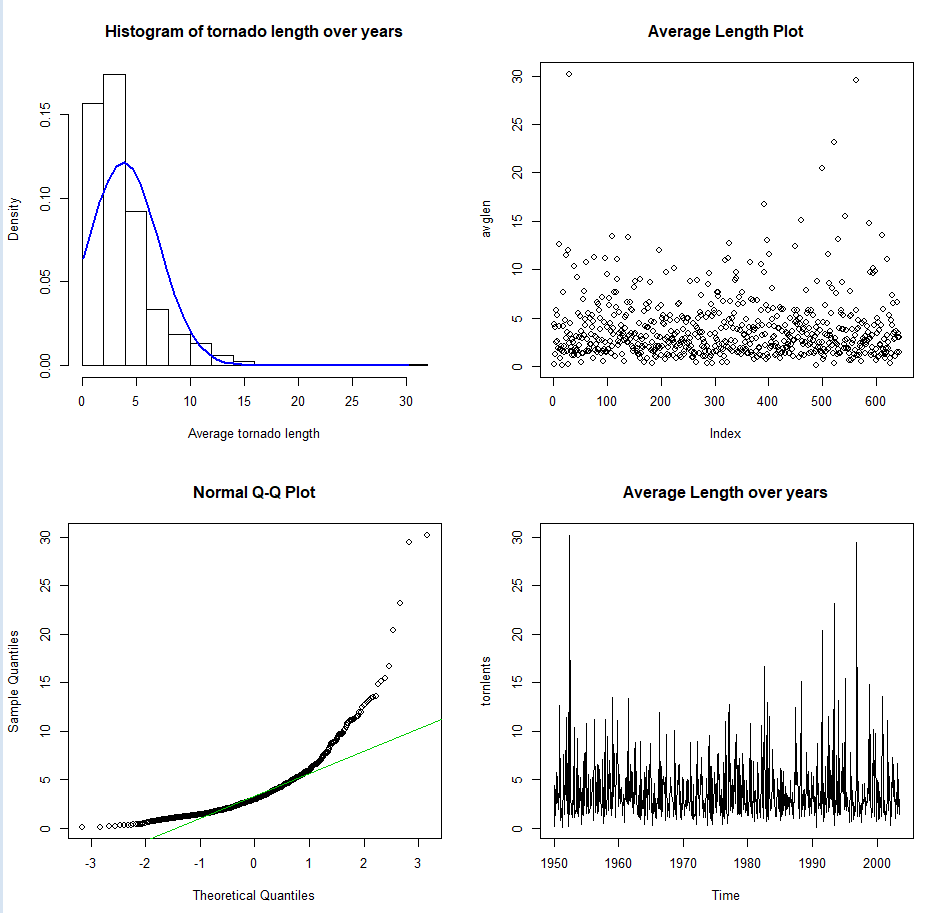




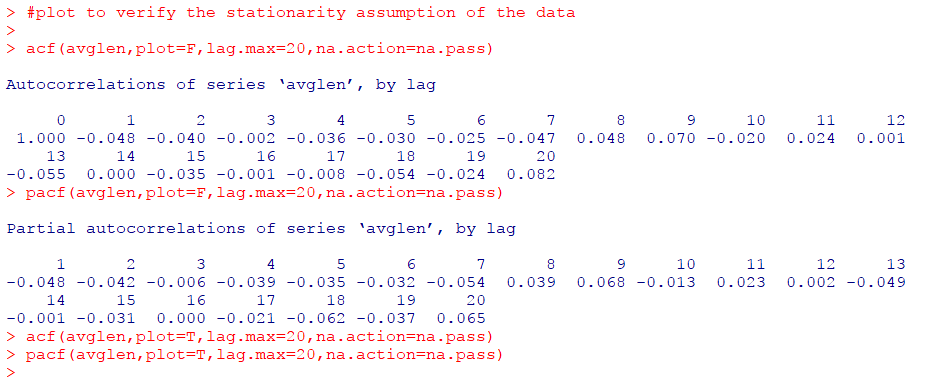


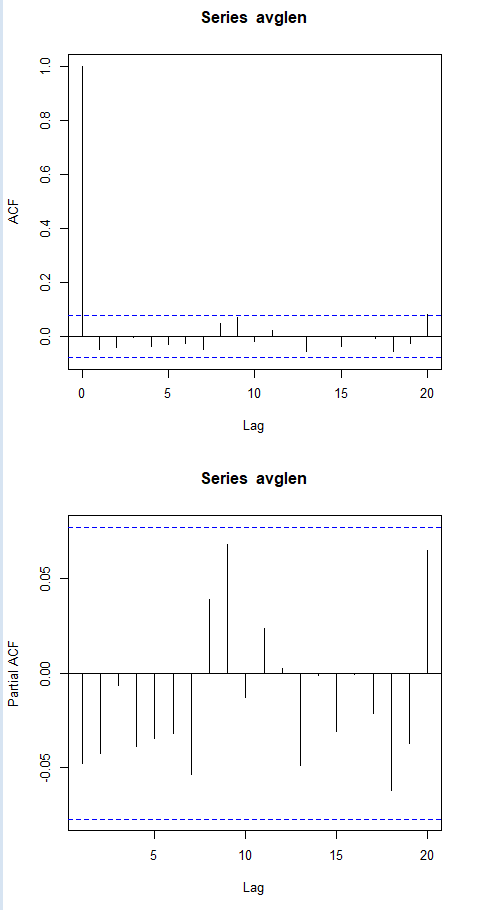


Resulting graphs and plots of Normality tests and time plots

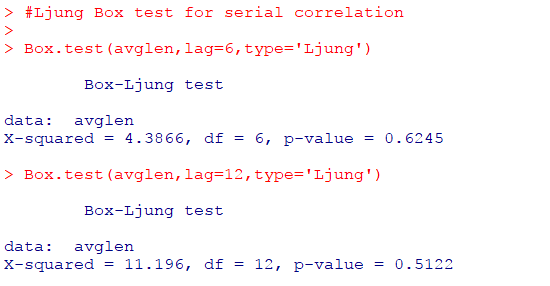


It is clear from the above histogram, QQ Plot and Jarque Bera Test that the tornadoes length data **doesn’t follow normal distribution**

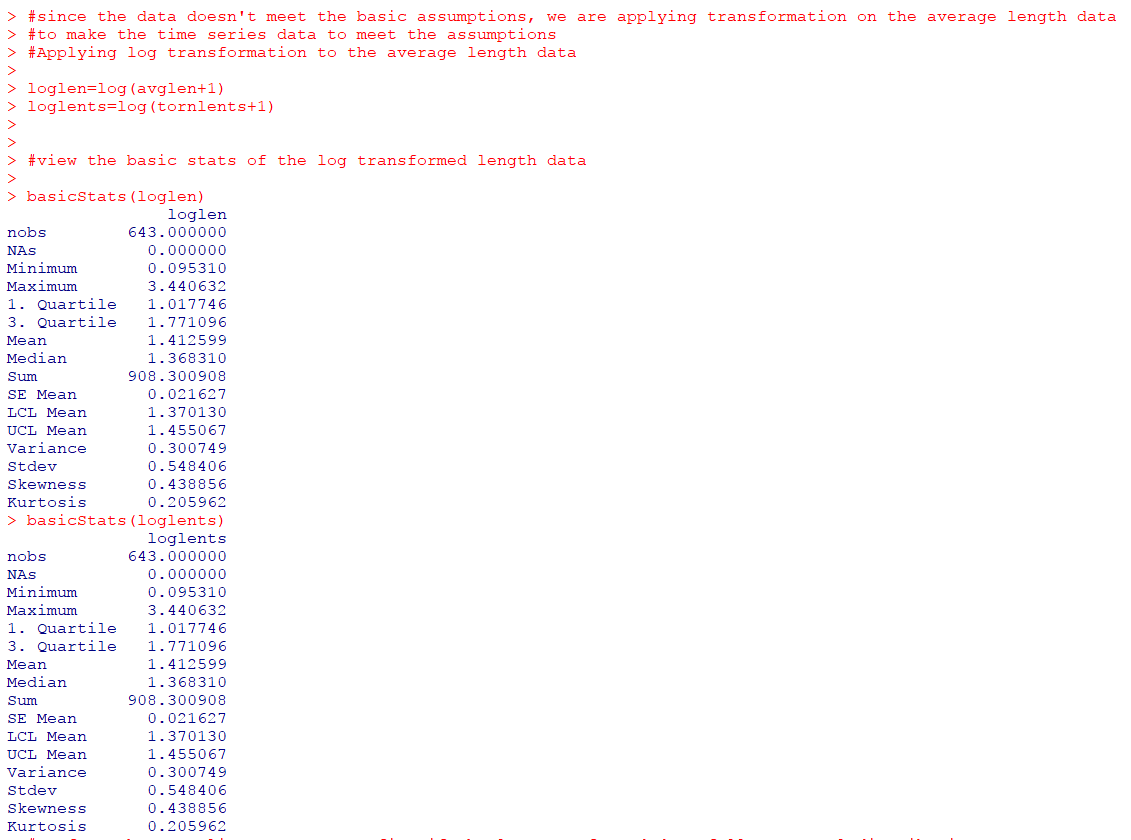




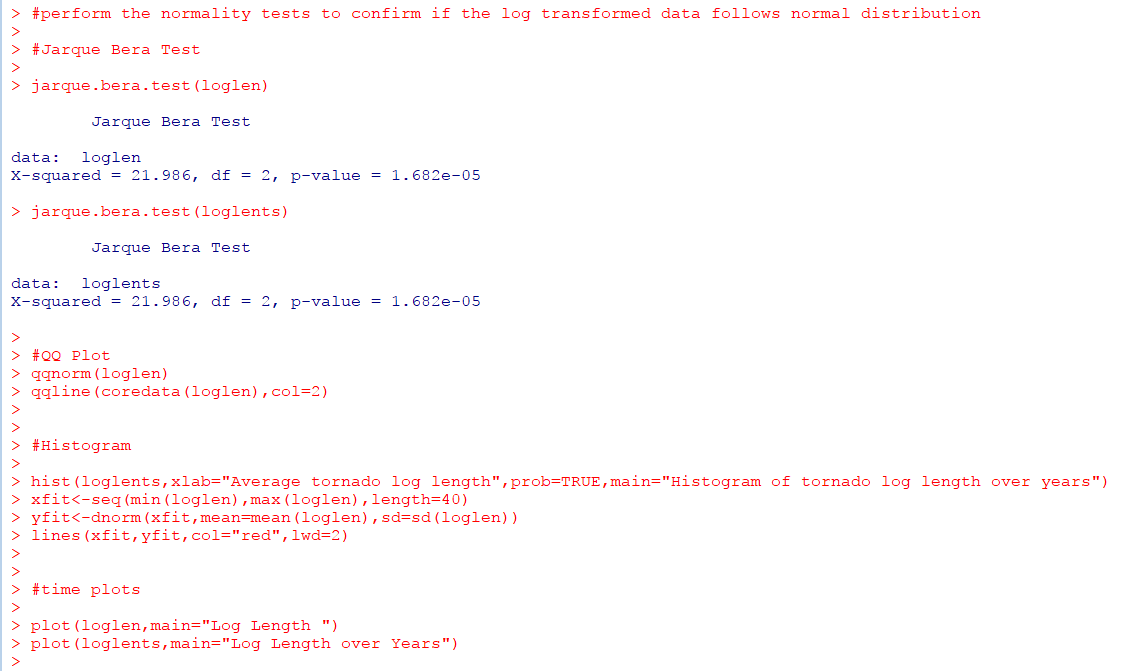
There is no single non-zero auto-correlation value present and the acf values decay slowly in the ACF plot and hence this confirms that the **tornadoes length data is not stationary**



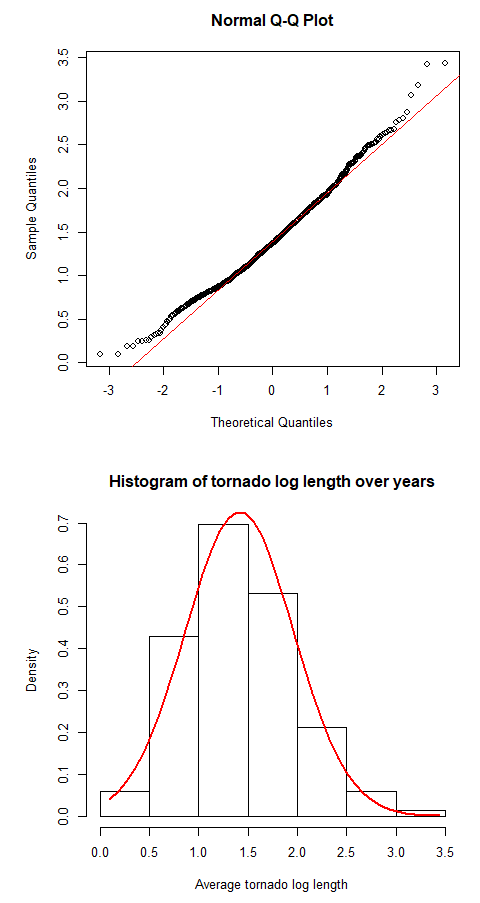
**Log transformed data:**

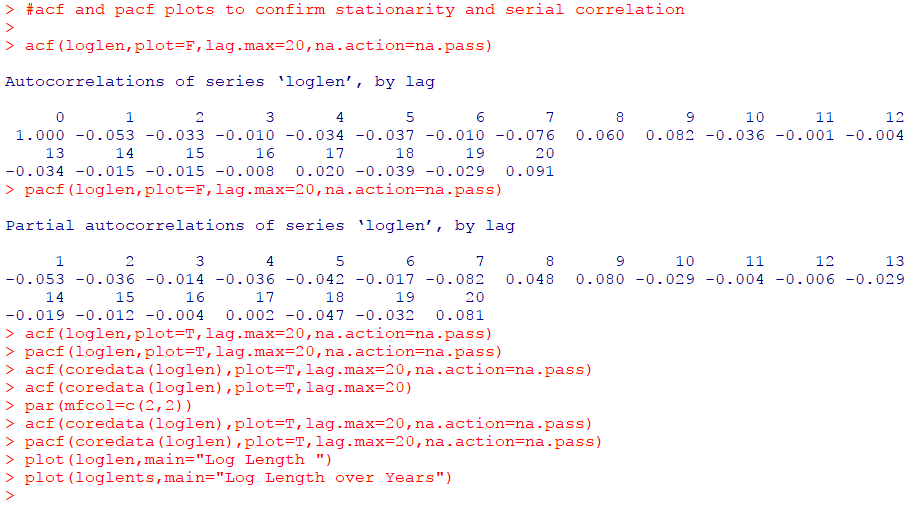


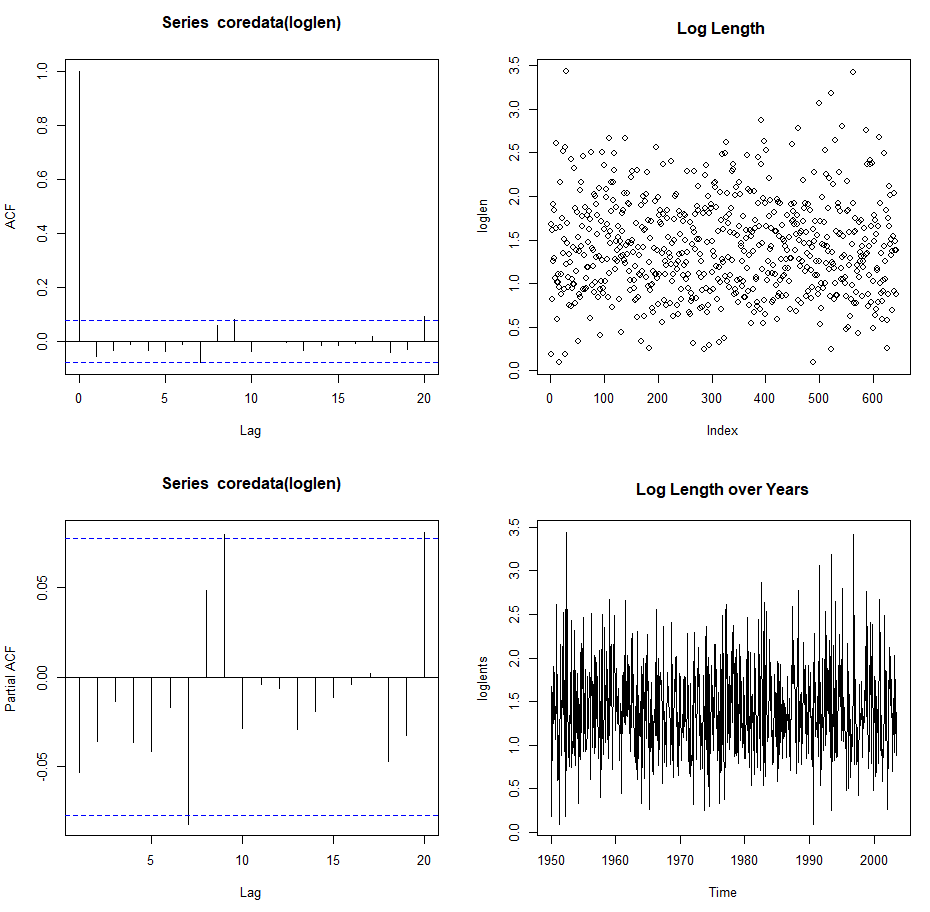
**Normality tests and time plots on log transformed data:**

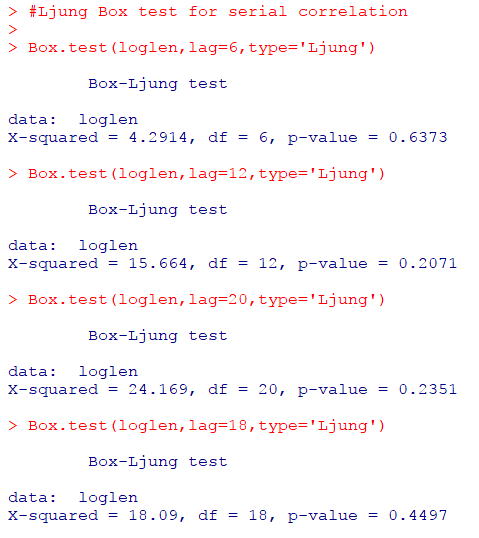


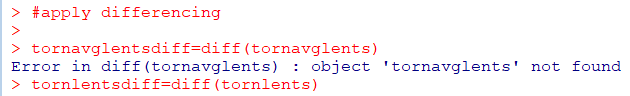
Log transformed data normality test results:

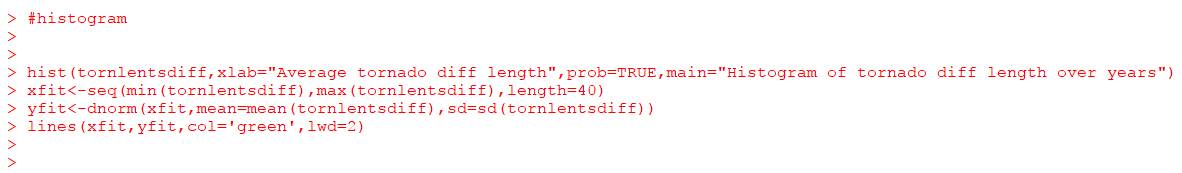


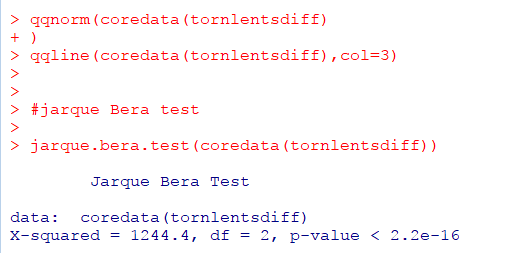


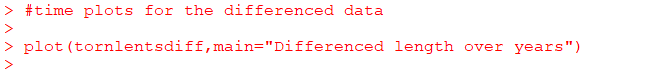




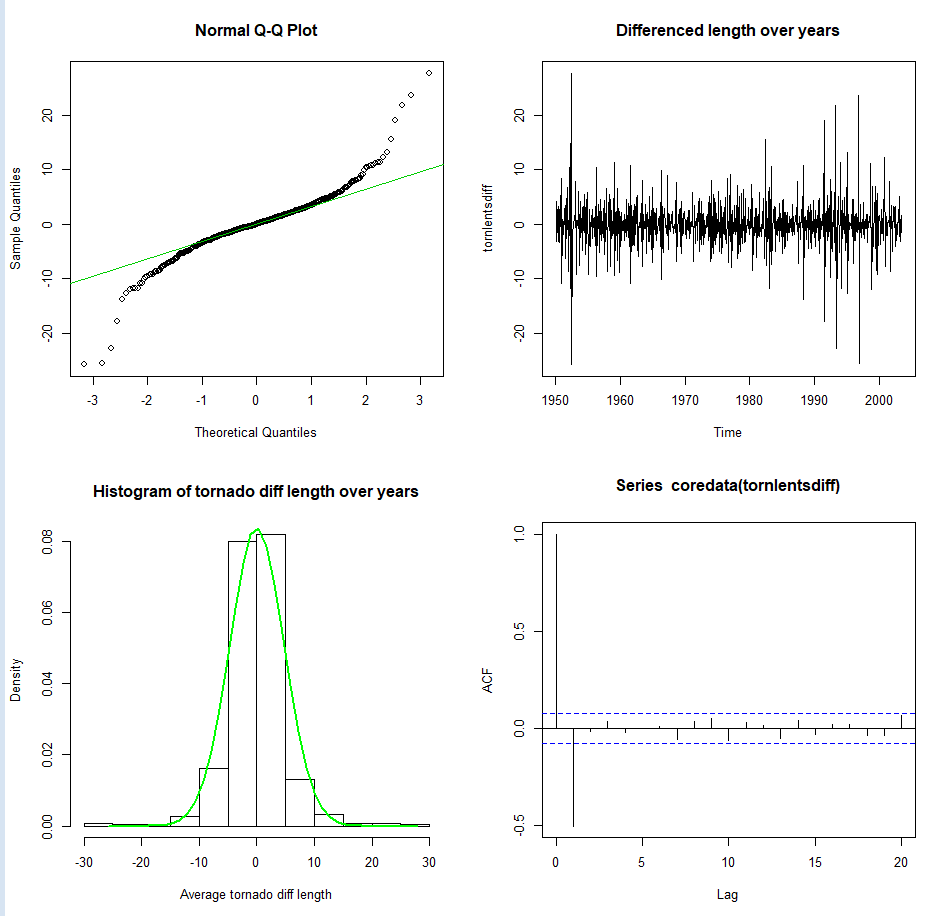


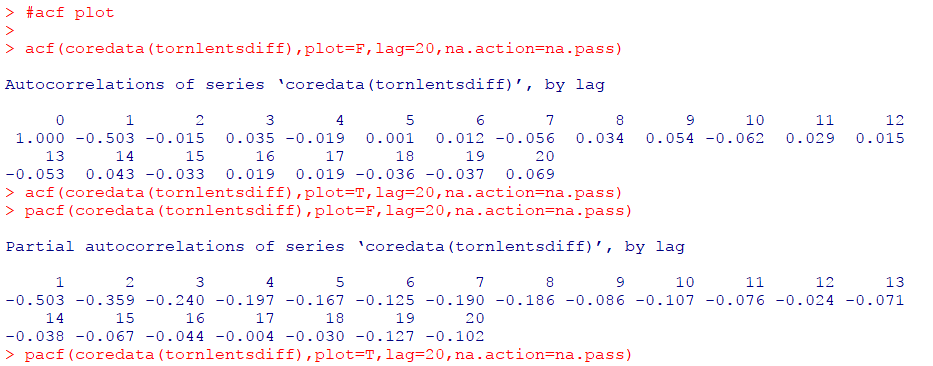


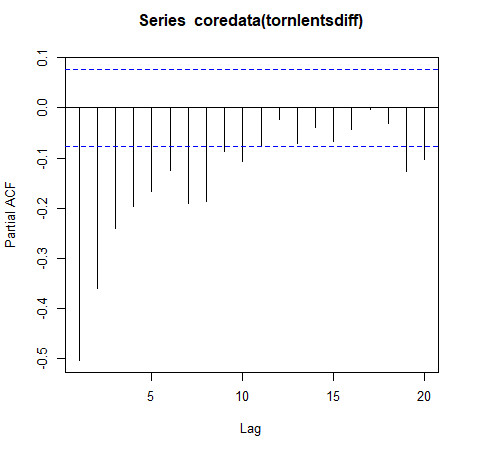


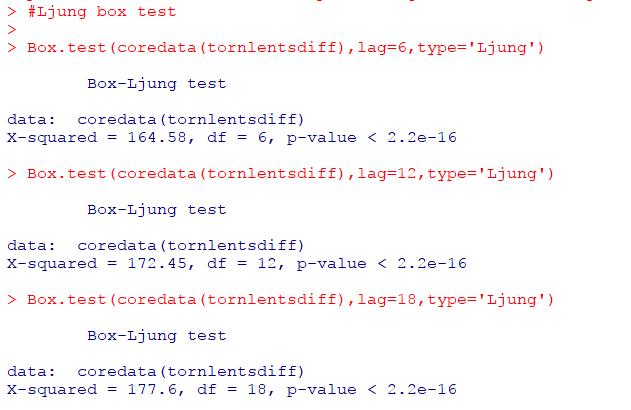


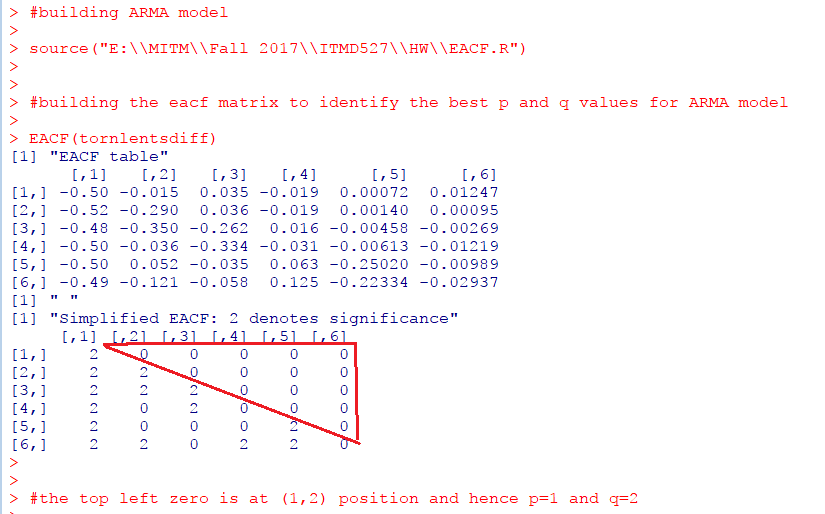
**Normality test results:**





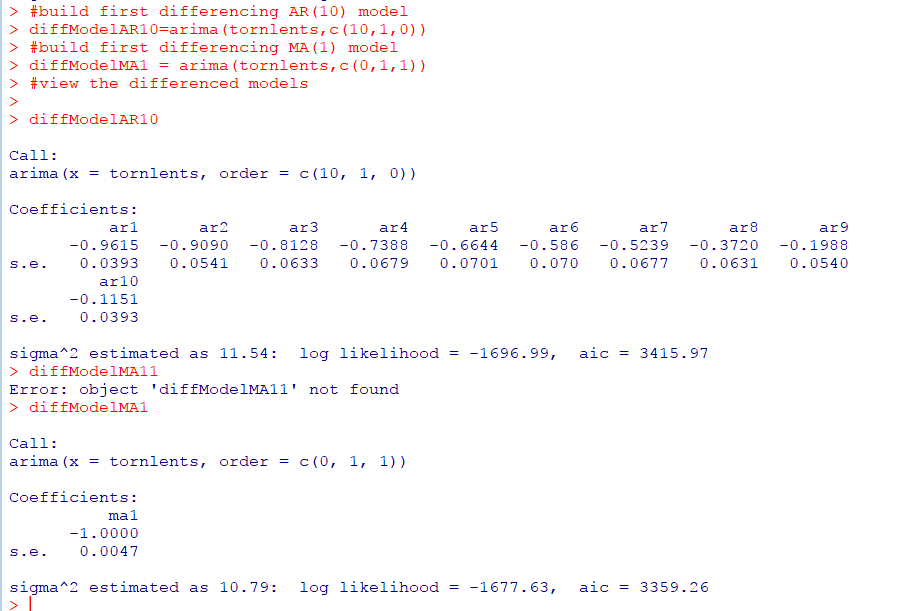




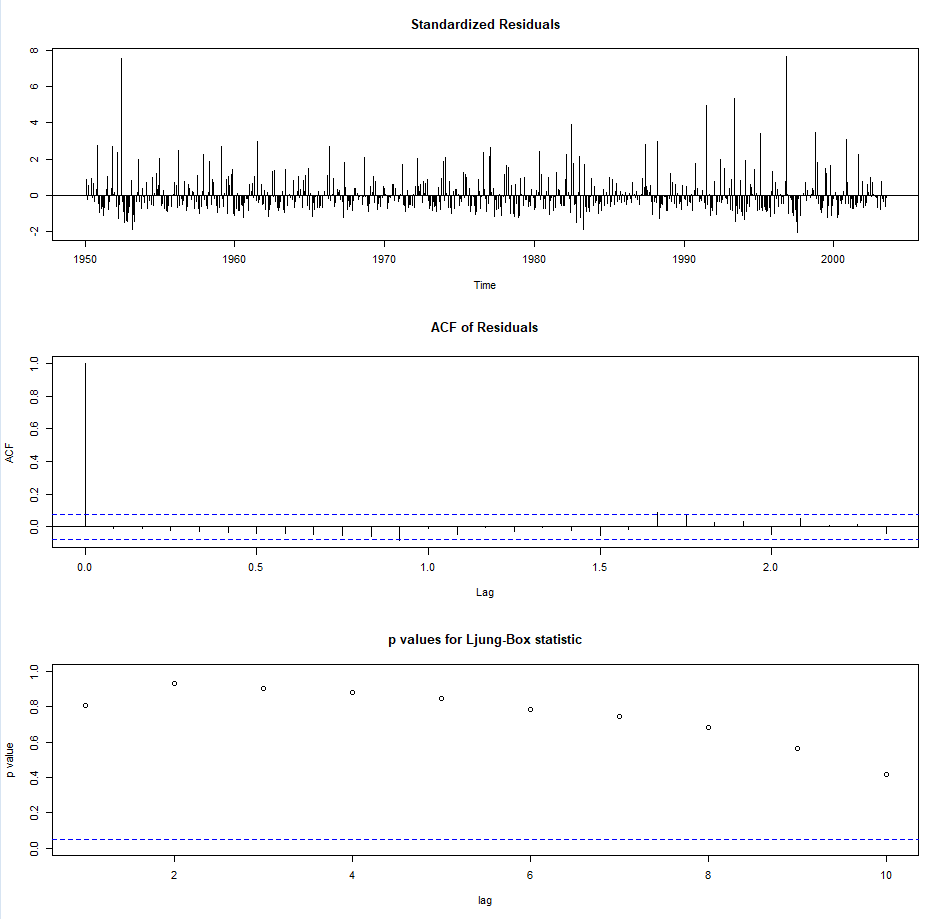


**Build Models:**

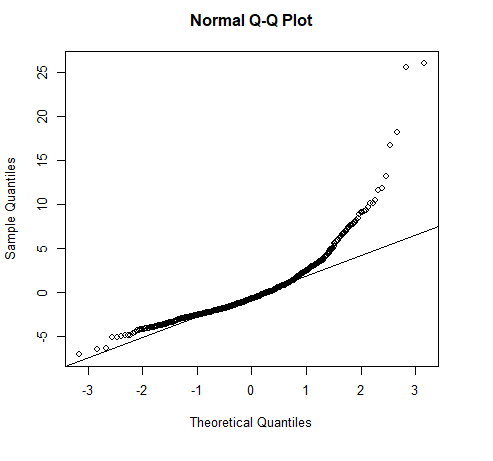
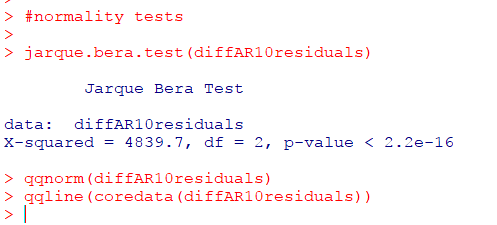
**AR (10) and MA (1) models:**



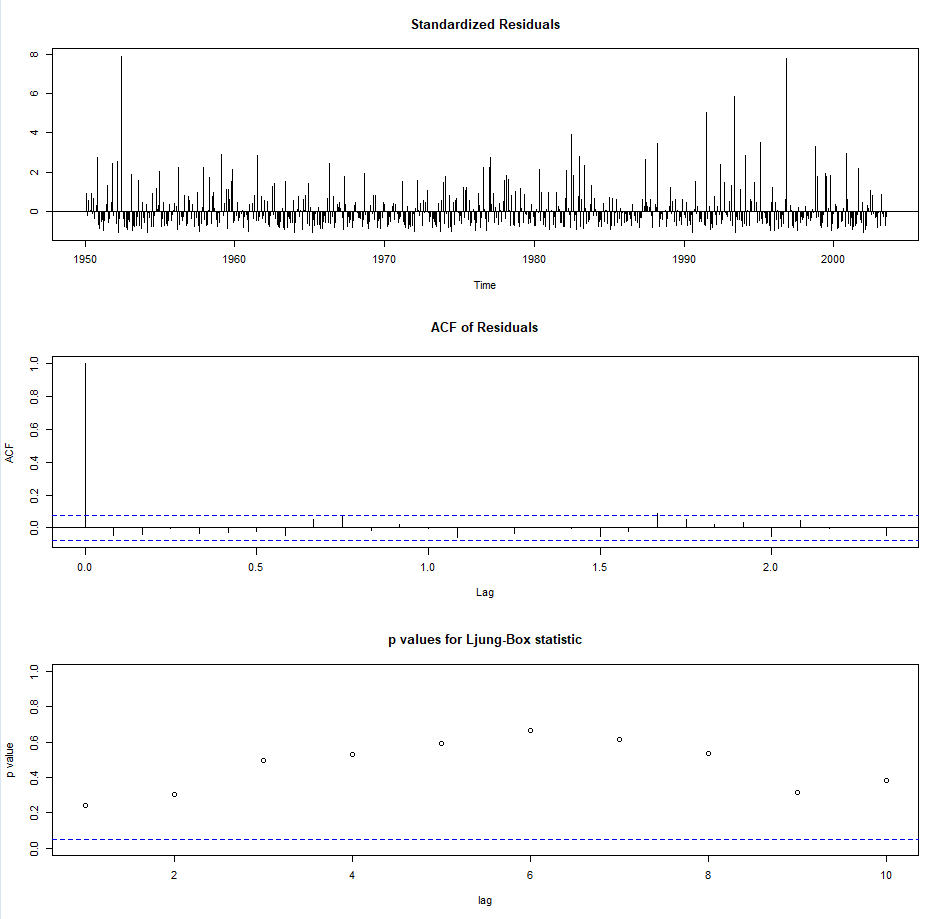
**Residual plots of AR (10) model:**



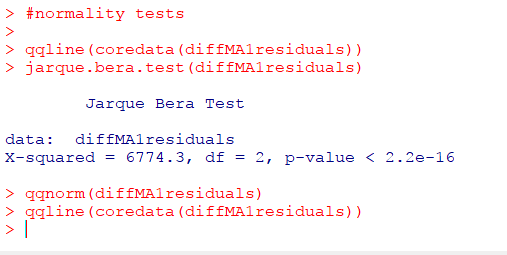
**Normality tests of AR (10) model residuals:**

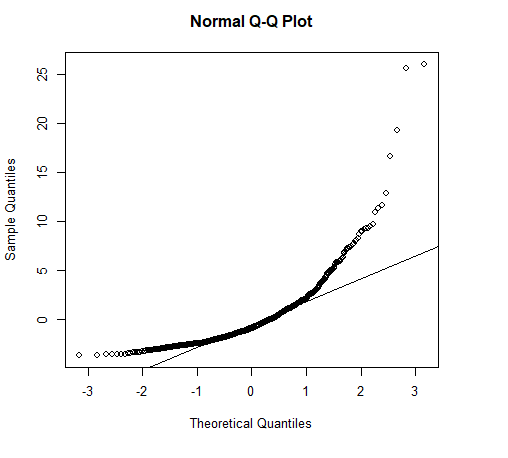


**Residual plots of MA (1) model:**

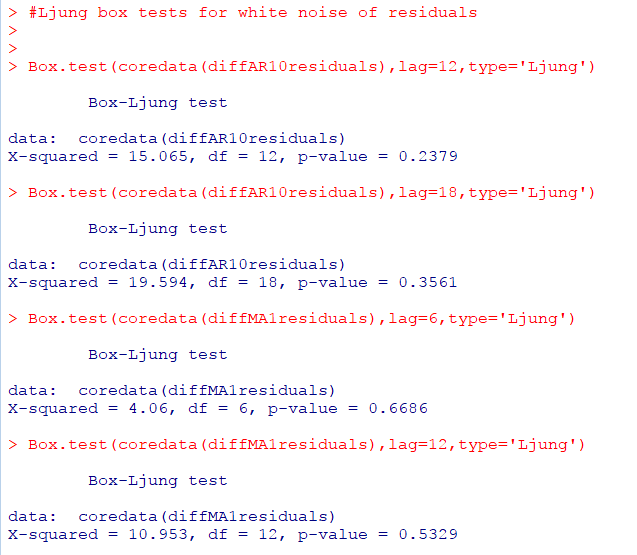


**Normality tests of MA (1) model**

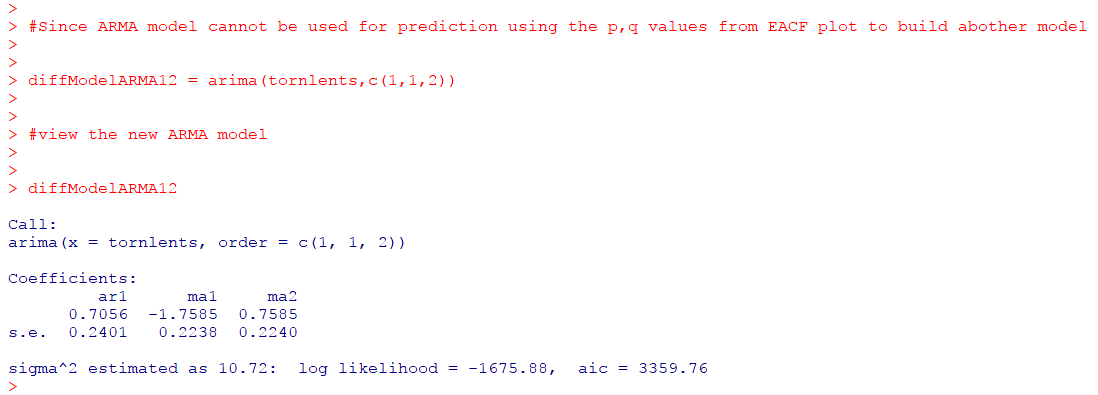




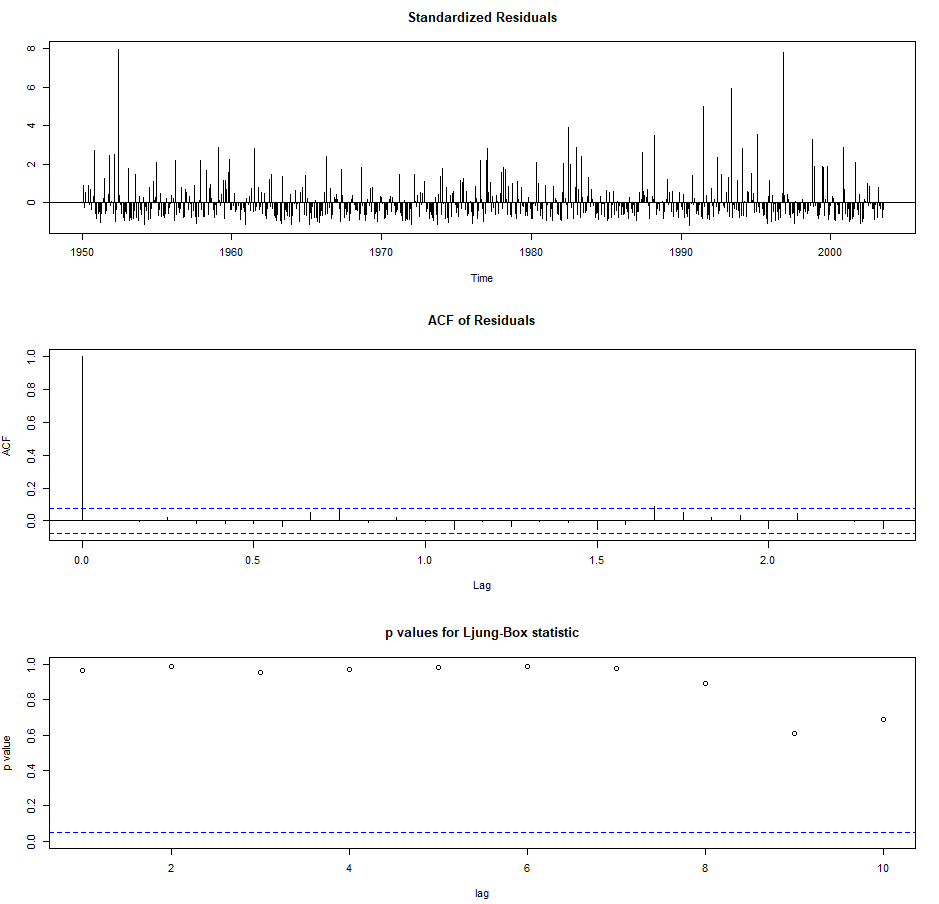
**Ljung Box test for residuals white noise:**



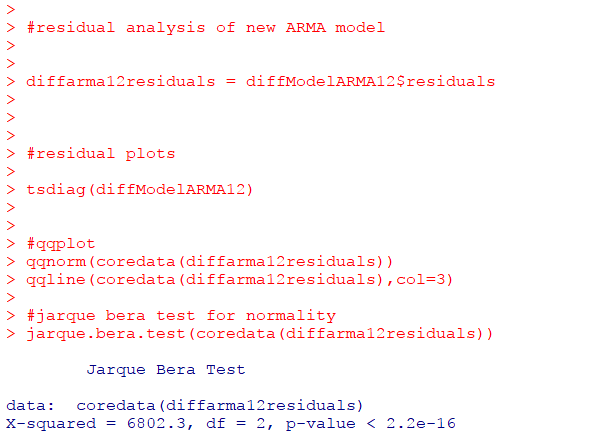
**ARMA (1,2) model**

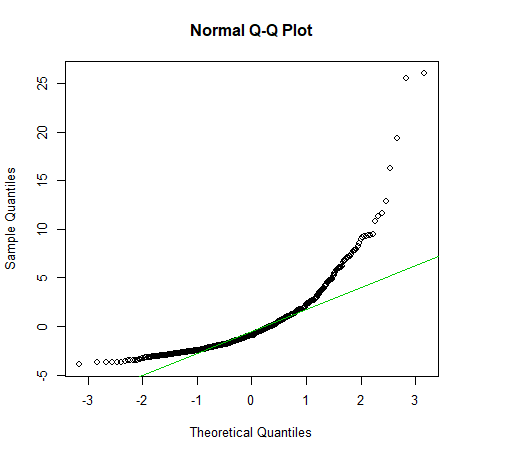


**Residual plots of ARMA (1,2) model**

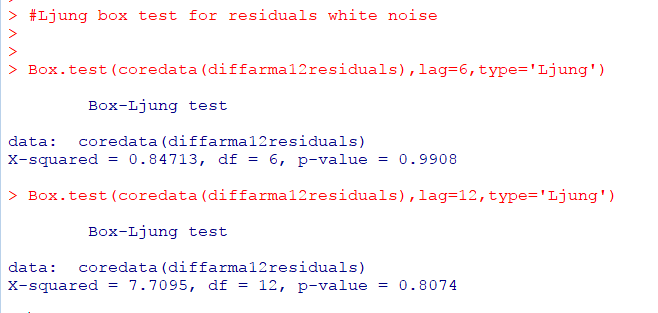


**Normality tests of ARMA (1,2) residuals**



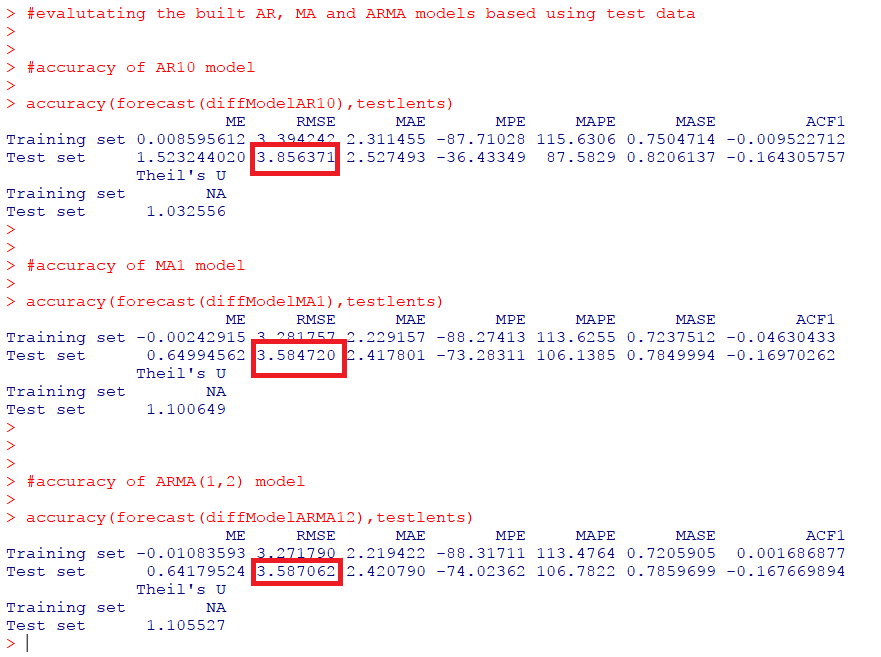


**Ljung Box test for ARMA (1,2) model residuals white noise**



A detailed summary of all the 3 models built with the training data is as given below:

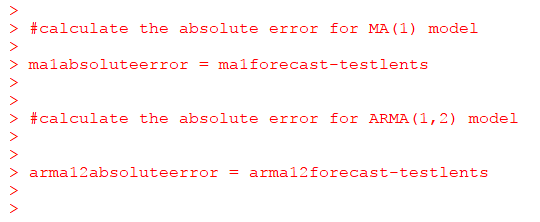
|  |  |  |  |
| --- | --- | --- | --- |
| Criteria | AR (10) | MA (1) | ARMA (1,2) |
| AIC | 3415.97 | 3359.26 | 3359.76 |
| Residuals white noise? | yes | Yes | Yes |
| Residuals normally distributed | Yes | Yes | Yes |
| Qualified Model? | Yes | Yes | Yes |



We could see that the RMSE values of MA (1) model and ARMA (1,2) model are the lowest compared to that of AR (10) model

But MA (1) and ARMA (1,2) model have very close RMSE values and hence we have a doubt if both these RMSEs are the same or not. So, to confirm this, we will perform two paired sample two-tailed hypothesis testing on absolute errors and make a conclusion on the best time series model for prediction.

Calculating absolute errors for MA (1) model and ARMA (1,2) models as shown below



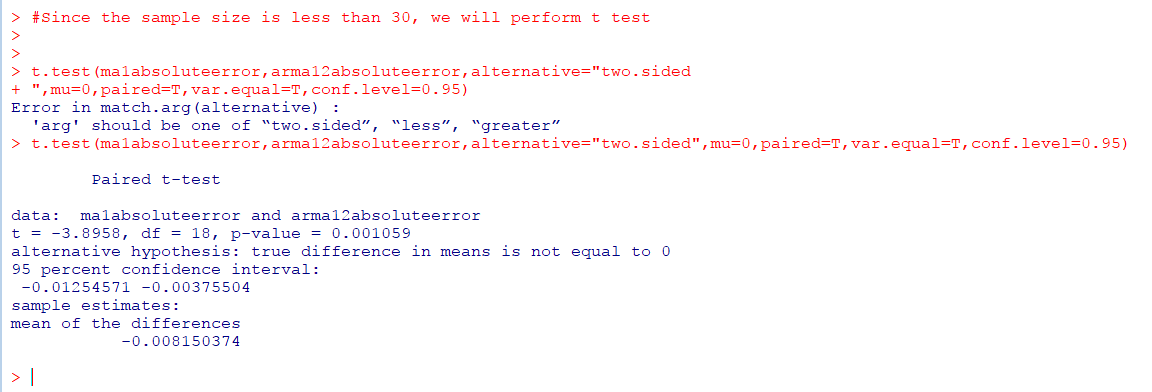
Since both the models are built on the same data and the mean errors are also for the same data, we will carry out two sample paired hypothesis.

Next step is to perform the two paired samples two tailed hypothesis testing. Since the sample size here is less than 30, we perform t-test.

The Hypothesis statements are:

*H0: Errors of models MA (1) and ARMA (1,2) are the same*

*Ha: Errors of models MA (1) and ARMA (1,2) are not the same*

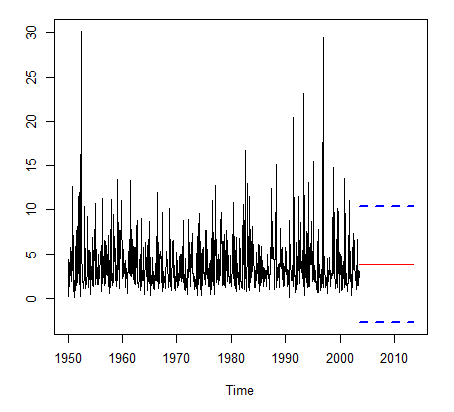


From the above hypothesis testing screenshot, we see that the p-value is less than 0.05. Hence, we do not have enough evidence to accept the null hypothesis at 95% confidence level. Hence, we can conclude that the errors of models MA (1) and ARMA (1,2) are not the same.

Hence, we confirm that **model MA (1) is the best model for tornado length prediction for different months in future**

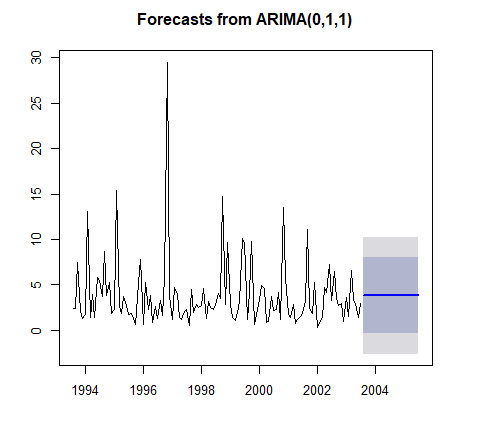
The predictions graph and the forecast graph for the next few years are as given below

**Prediction Graph:**



The lines in blue represent the prediction interval and the one in red indicates the predicted tornado length for the years from 2004.

**Forecast Graph**



The line in blue represents the forecasted tornado length for the years after 2004, the dark gray portion represent the 80% confidence interval and the light gray portion represents the 95% confidence interval.