**DSCI 5340 – Predictive Analysis and Business Forecasting**

**Forecast of weekly average grain carload by month by US Railroads**

**TEAM MEMBERS:**

**Gollapally Shashank Reddy - 11376637**

**Jammula Keerthi Chandana - 11364726**

**Nayini Nikhil Kumar – 11374557**

**Introduction:**

This dataset includes data of weekly average carloads, containers and trailers originated by month by major U S Railroads.

This project will explore the time series plot of the data, analysis based on SAS software, and forecast of Grain carloads.

**Hypothesis:**

**Time series Auto Regression integrated Moving Averages:**

ARIMA model is used in this analysis to predict the future trends by using the seasonality. We are here using differential values in the series instead of the actual values.

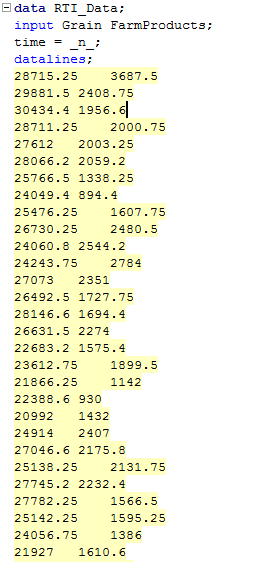
**Null Hypothesis:**

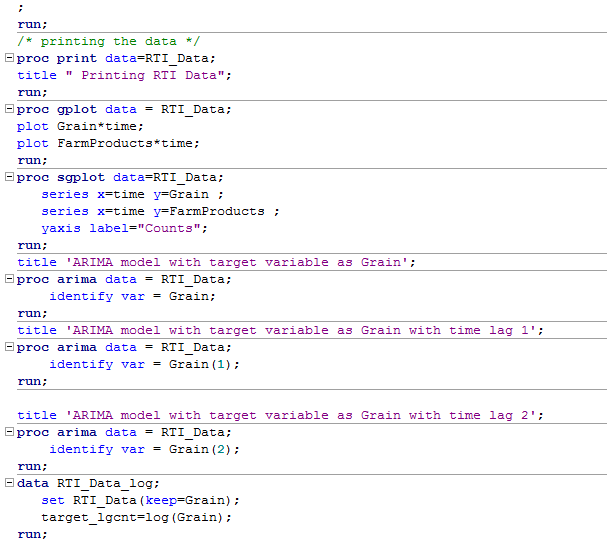
White noise value (P < 0.05)

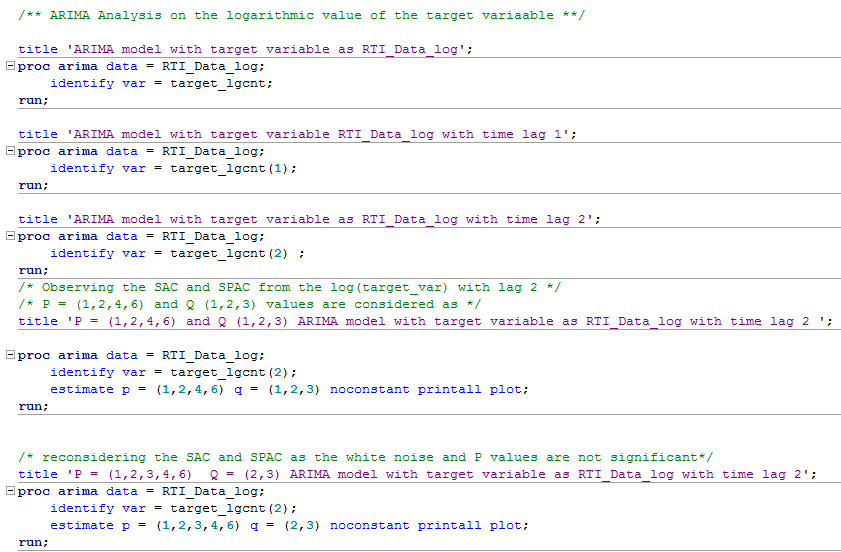
**Alternate Hypothesis:**

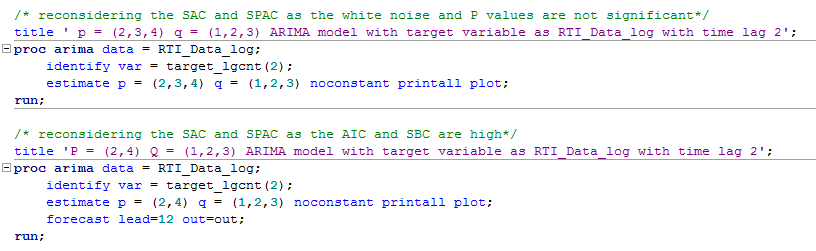
Not white noise value (P > 0.05)

**CODE:**

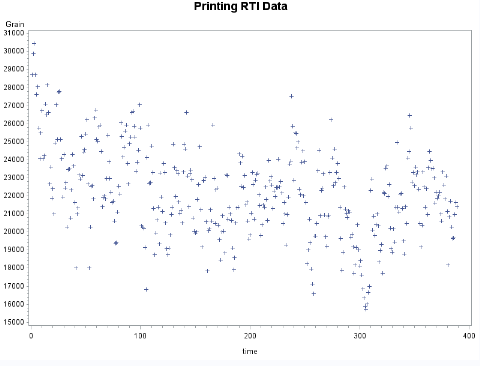


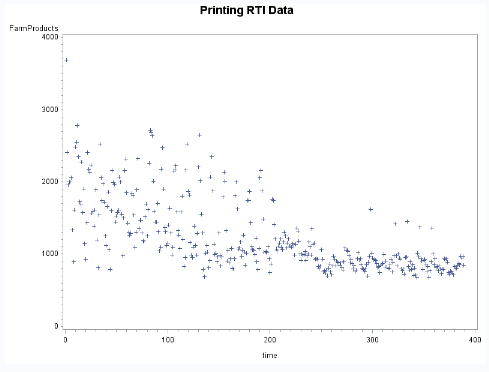




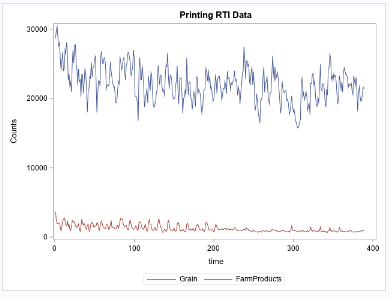


**GPlot:**

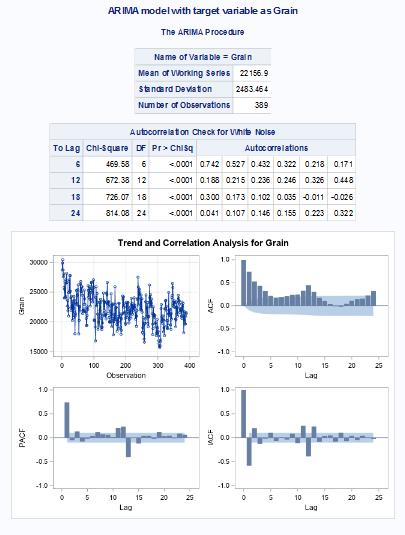




**Time Series Plot:**



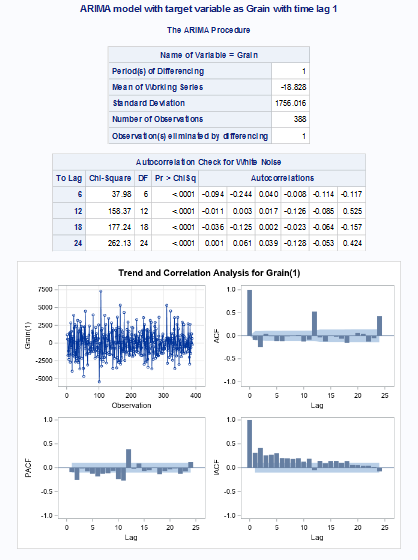
**Time series ARIMA model preliminary analysis:**



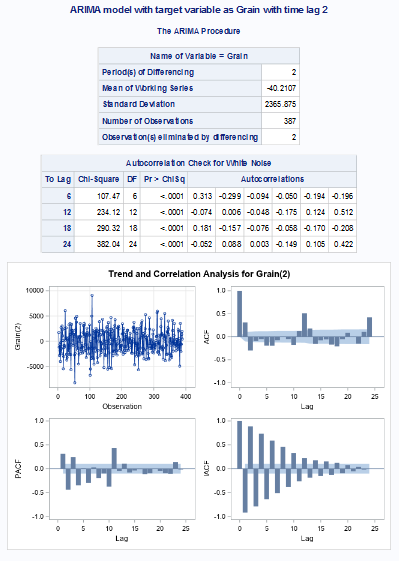
Here, we observed the Trend and correlation of the variable.

On further analysis we get,

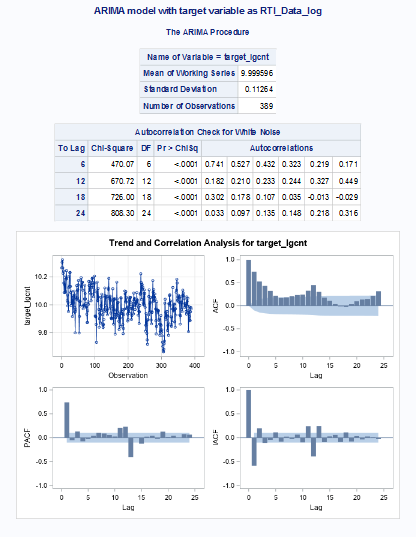
ARIMA with target variable with lag 1



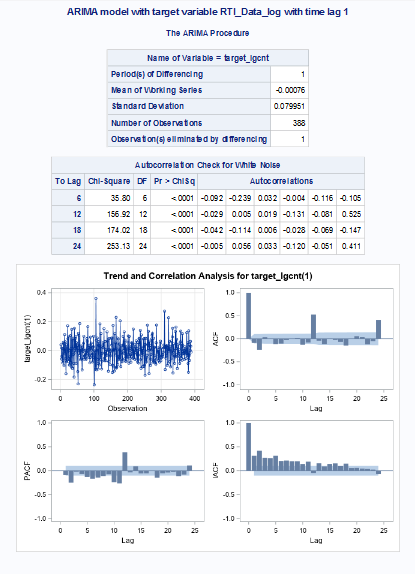
ARIMA with target variable with lag 2



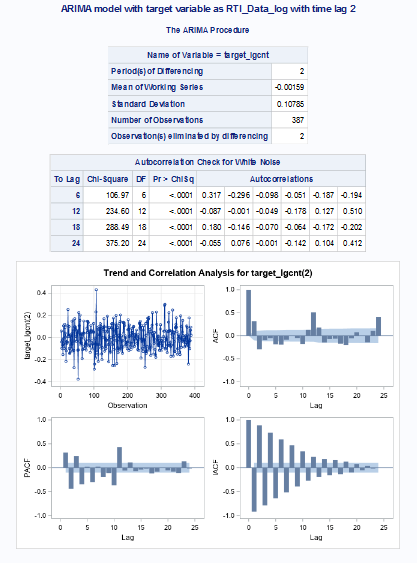
ARIMA with log target variable



ARIMA with log target variable with lag 1



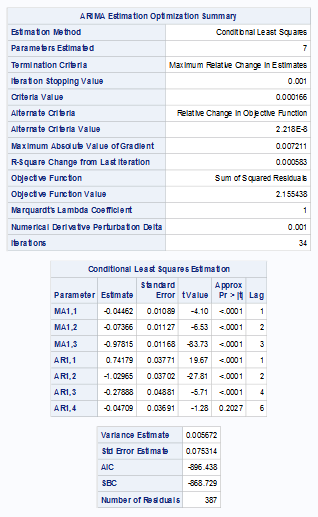
ARIMA with log target variable with lag 2

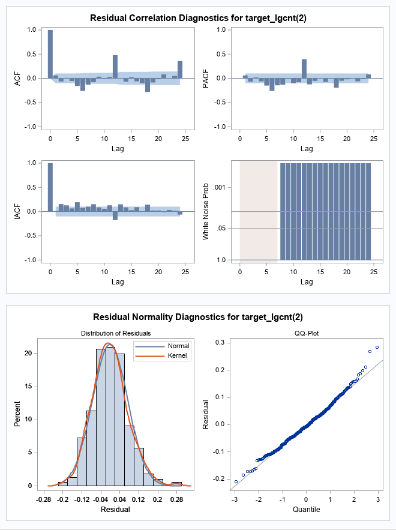


**Time series ARIMA Model with iterations using SAC and SPAC:**

P = (1,2,3,4) Q = (1,2,3) ARIMA model

Target variable: RTA\_Data\_log with lag 2

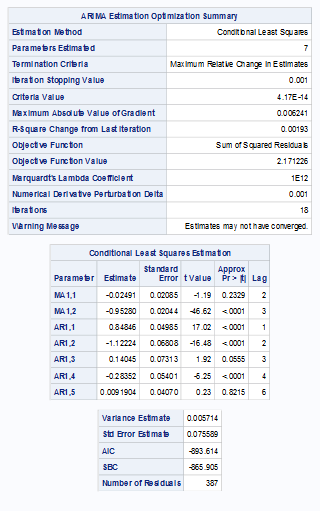


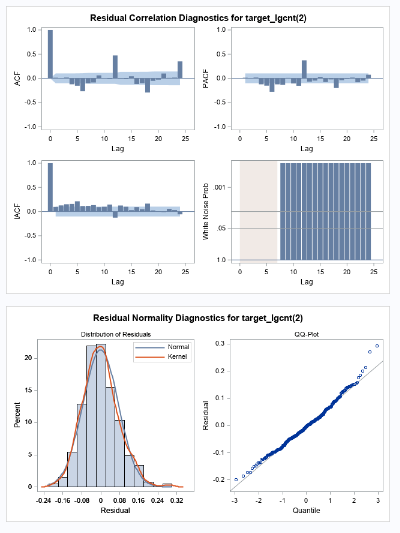


Here the white noise hypothesis failed and the lags did not die down.

P = (1,2,4,6) Q = (1,2,3) ARIMA model

Target variable: RTA\_Data\_log with lag 2

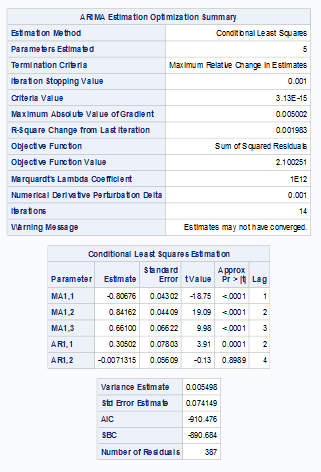


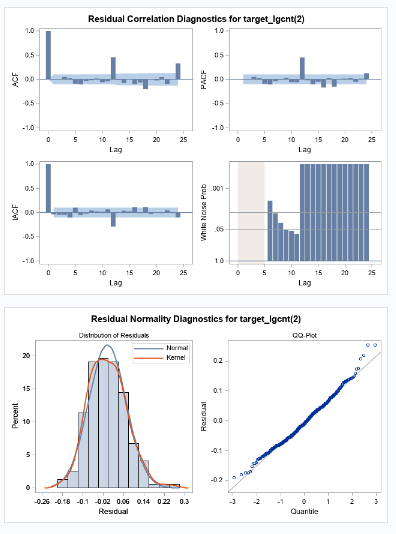


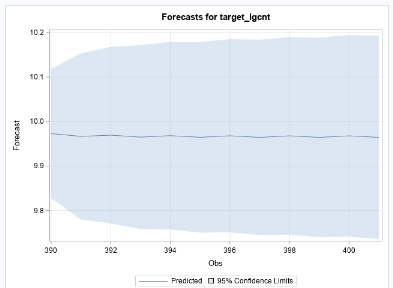
Here the white noise hypothesis failed and the lags did not die down.

P = (2,4) Q = (1,2,3) ARIMA model

Target variable: RTA\_Data\_log with lag 2





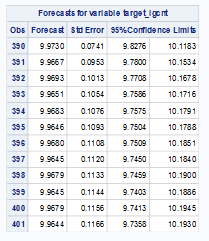


The final model produces P-value P > 0.05 which are significant and rejects the Null Hypothesis.

The standard error is less when compared to the previous models and shows more accuracy.

The Residual ACF and PACF plots satisfy the stationarity for the residuals.

Here are the forecast results.

.