Assignment 1

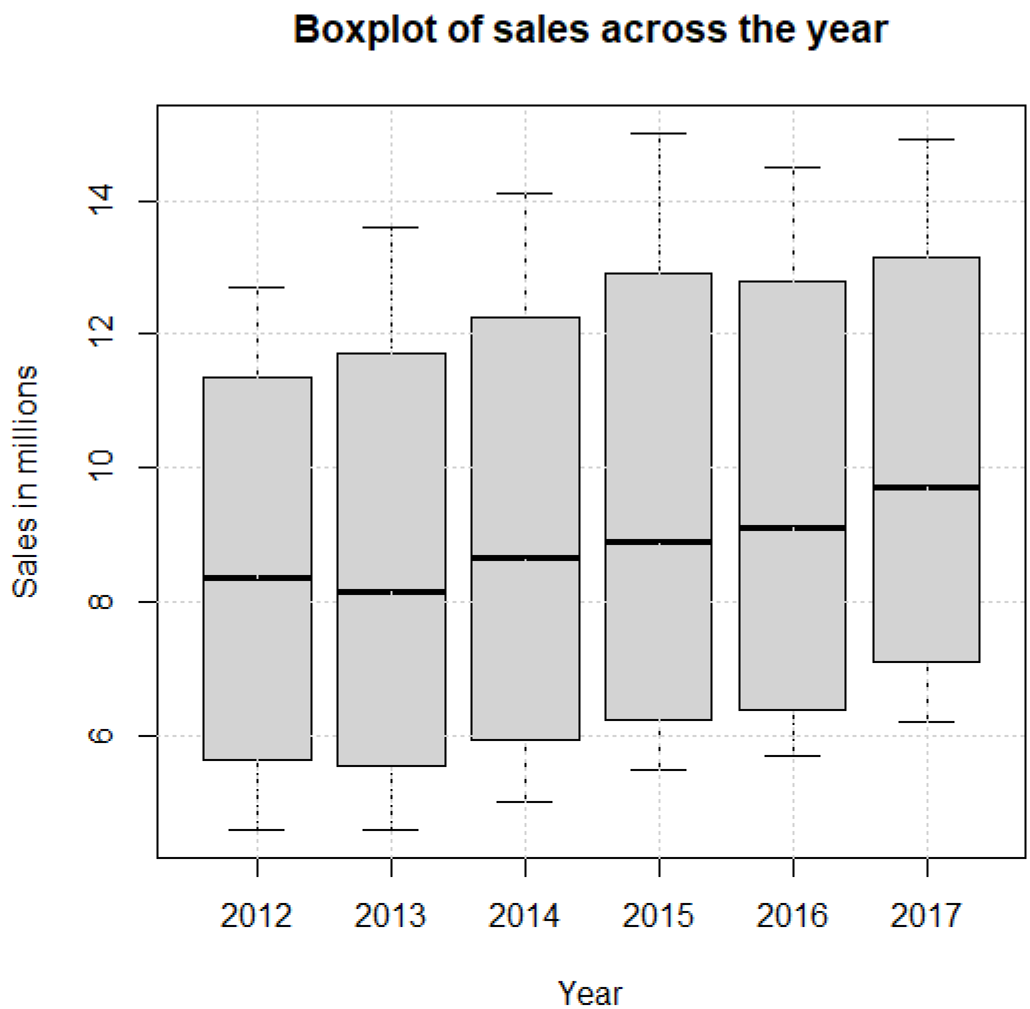
DSA 511: Time Series Analysis & Forecasting

Student Name: Tahmid Hussain

Student ID: 2023-1-83-021

Q1: Draw a boxplot of sales across the year. [to find out any changes in sales over the year]

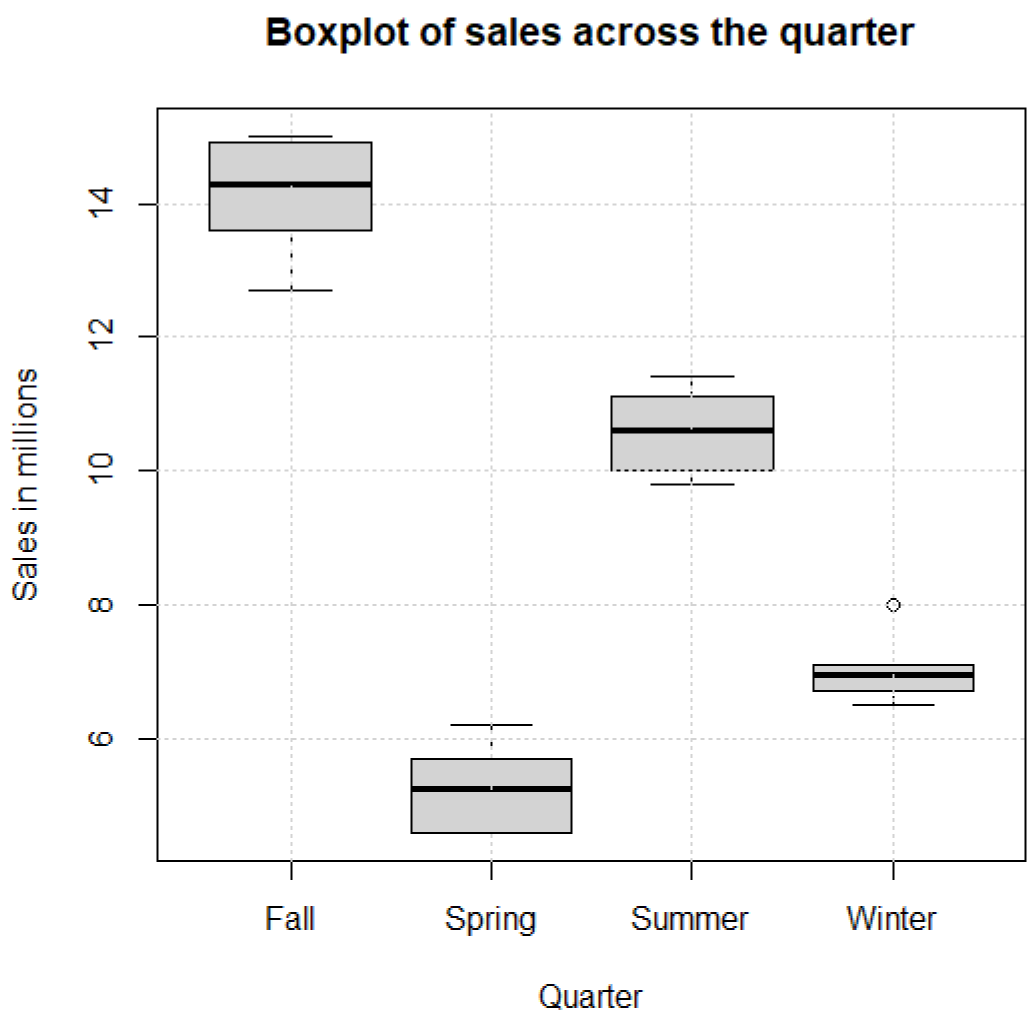
Ans: The boxplot is shown below.



From the above boxplot of the sales across the year, we can see that the median sales in each year do not vary largely; almost all the median sales across the years are within the range of 8–9 million. As the median sales across the years do not vary significantly, we can’t say that there is a strong presence of trends in the data. We can also confirm this fact from the ADF test, from which we see that a low p-value is obtained. This indicates that the data is, in fact, stationary.

Q2: Draw a boxplot of sales across the quarter. [to find out any changes in sales over the quarter]

Ans: The boxplot is shown below.

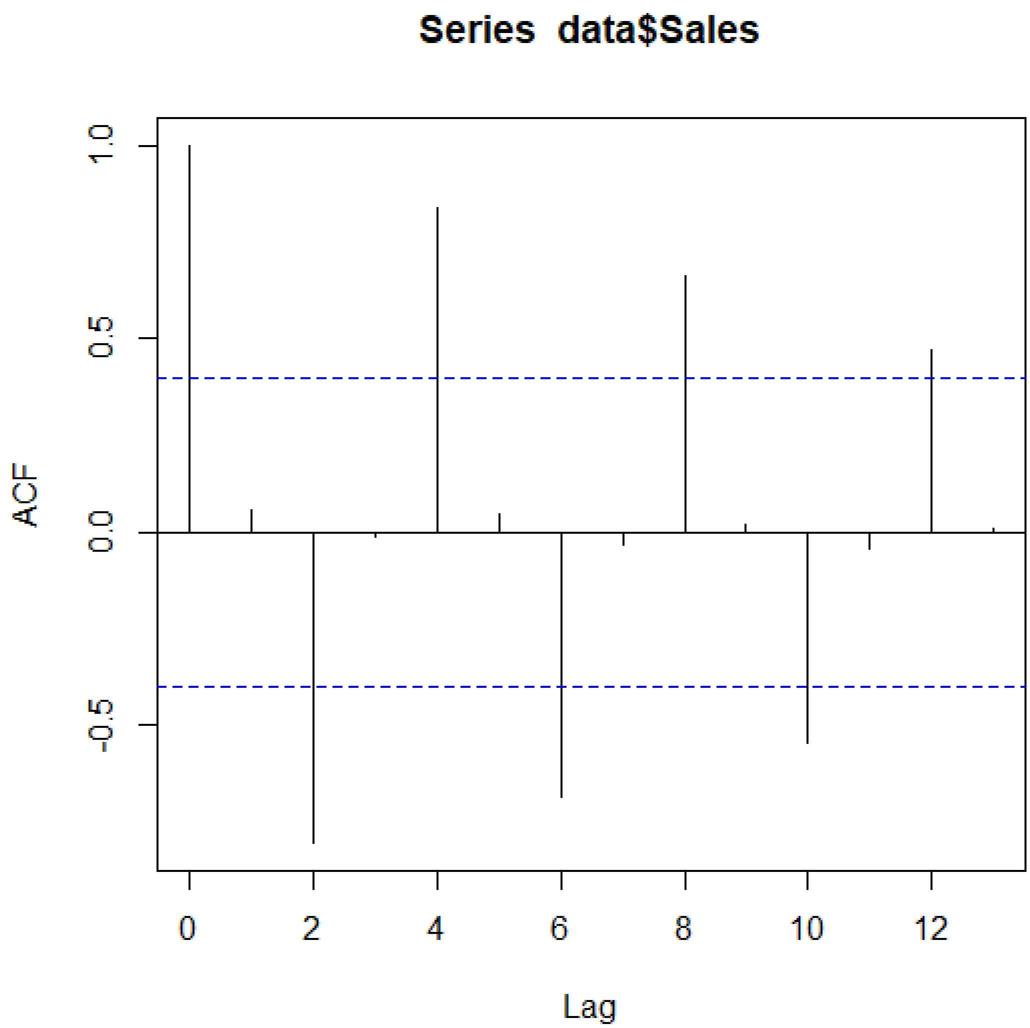


From the above boxplot of the sales across the year, we can see that there is significant difference in the median sales across the quarters. Also, the interquartile ranges of the boxes are in largely different regions in the graph, which indicates a strong presence of seasonality in the data.

Q3: Report the lags where significant autocorrelations are found.

Ans: To find out the lags where significant autocorrelations are found, we perform three tests: the ACF plot, the Ljung-Box test, and the Box-Pierce test.

ACF plot



We find the presence of significant autocorrelations at lags 2, 4, 6, 8, 10, and 12.

Ljung-Box Test

Null Hypothesis, H_0 : $\rho_1 = \rho_2 = \rho_3 = \ldots\ldots. = \rho_m = 0$

Alternative Hypothesis, H_a : At least one is different from 0.

If we perform the Ljung-Box test for lags 1 to 23, we find that for lags 2 to 23, the p-value is less than 0.05. From this result, we can say that, for lags 2 to 23, there are significant autocorrelations up to each lag.

Box-Pierce Test

Null Hypothesis, $H_0 : \rho_1 = \rho_2 = \rho_3 = \ldots\ldots = \rho_m = 0$

Alternative Hypothesis, H_a : At least one is different from 0.

If we perform the Box-Pierce test for lags 1 to 23, we also find the same result. That is, for lags 2 to 23, the p-value is less than 0.05. From this result, we can say that, for lags 2 to 23, there are significant autocorrelations up to each lag.

Q4: Separate the last four quarter data as test data and the rest as training data.

Ans: We separate the last four quarters of data as test data and the rest as training data in the way demonstrated in the attached code file.

Q5: Fit the additive and multiplicative model on the training data set.

Ans: Let, T_t be the trend component, S_t be the seasonality component and e_t be the residuals of the models.

As we all know, the equation of the additive model can be expressed in the following way:

$Y_t = T_t + S_t + e_t = a + b * t + S_t + e_t$

Fitting an additive model on the training data, we find that the intercept of the model, a, has a value of 8.26951, and slope, b, has a value of 0.07719. S_t indicates the respective seasonal indices of the quarters.

So, the additive model can be expressed as:

$Y_t = 8.27 + 0.077 * t + S_t$

As we all know, the equation of the multiplicative model can be expressed in the following way:

$Y_t = T_t * S_t + e_t = [a + b * t] * S_t + e_t$

Fitting a multiplicative model on the training data, we find that the intercept of the model, a, has a value of 8.18054, and slope, b, has a value of 0.08115. S_t indicates the respective seasonal indices of the quarters.

So, the multiplicative model can be expressed as:

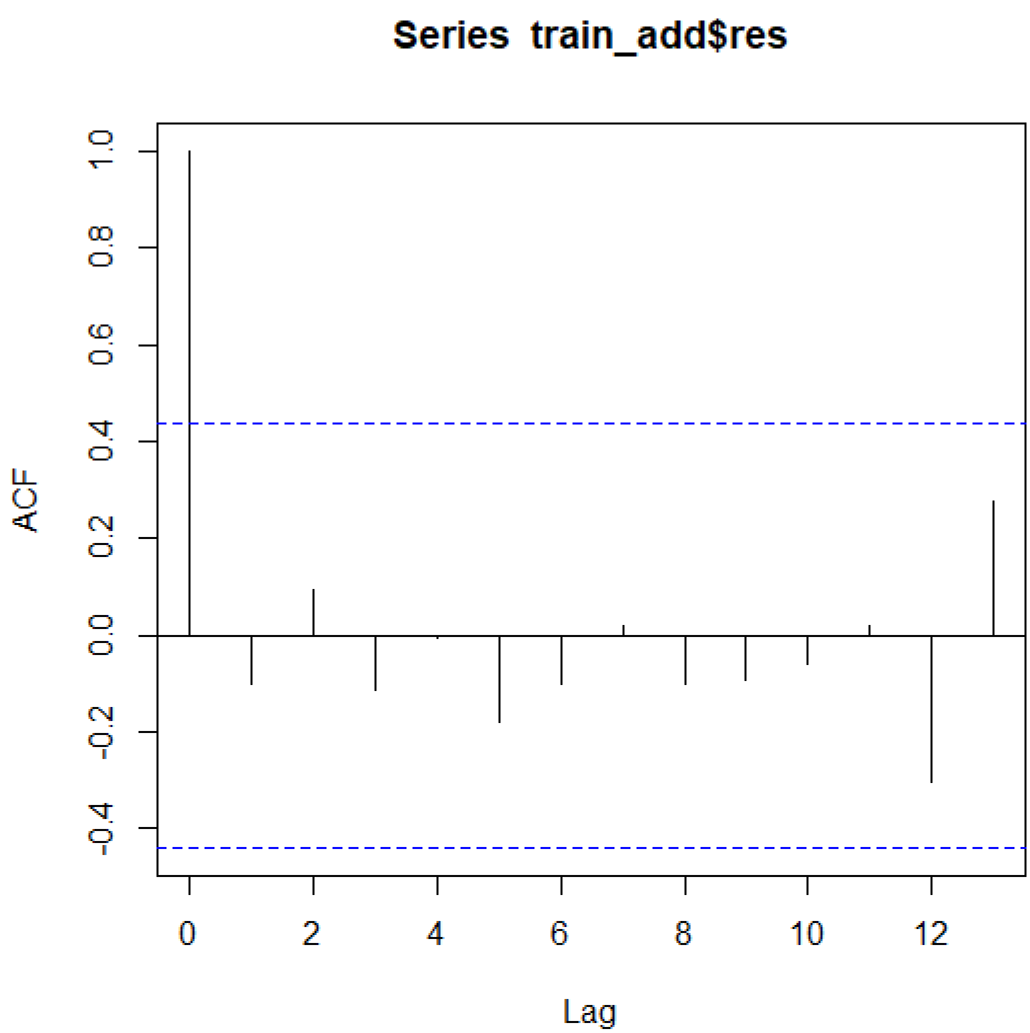
$Y_t = [8.18 + 0.08 * t] * S_t$

Q6: Check whether the models produce white noise residuals.

Ans: To find out whether the models have produced white noise residuals, we will check whether there are significant autocorrelations at any of the lags. We perform three tests for this: the ACF plot, the Ljung-Box test, and the Box-Pierce test.

White noise checking for the additive model

ACF plot



From the ACF plot of the residuals of the additive model, we do not find the presence of significant autocorrelation at any of the lags. So, we can say that the model has produced white noise residuals according to the ACF plot.

Ljung-Box Test

Null Hypothesis, $H_0 : \rho_1 = \rho_2 = \rho_3 = \ldots\ldots = \rho_m = 0$

Alternative Hypothesis, H_a : At least one is different from 0.

If we perform the Ljung-Box test for lags 1 to 19, we find that for all of the lags, the p-value is greater than 0.05. From this result, we can say that there are no significant autocorrelations at any of the lags. So, the additive model has produced white noise residuals according to the Ljung-Box test.

Box-Pierce Test

Null Hypothesis, $H_0 : \rho_1 = \rho_2 = \rho_3 = \ldots\ldots = \rho_m = 0$

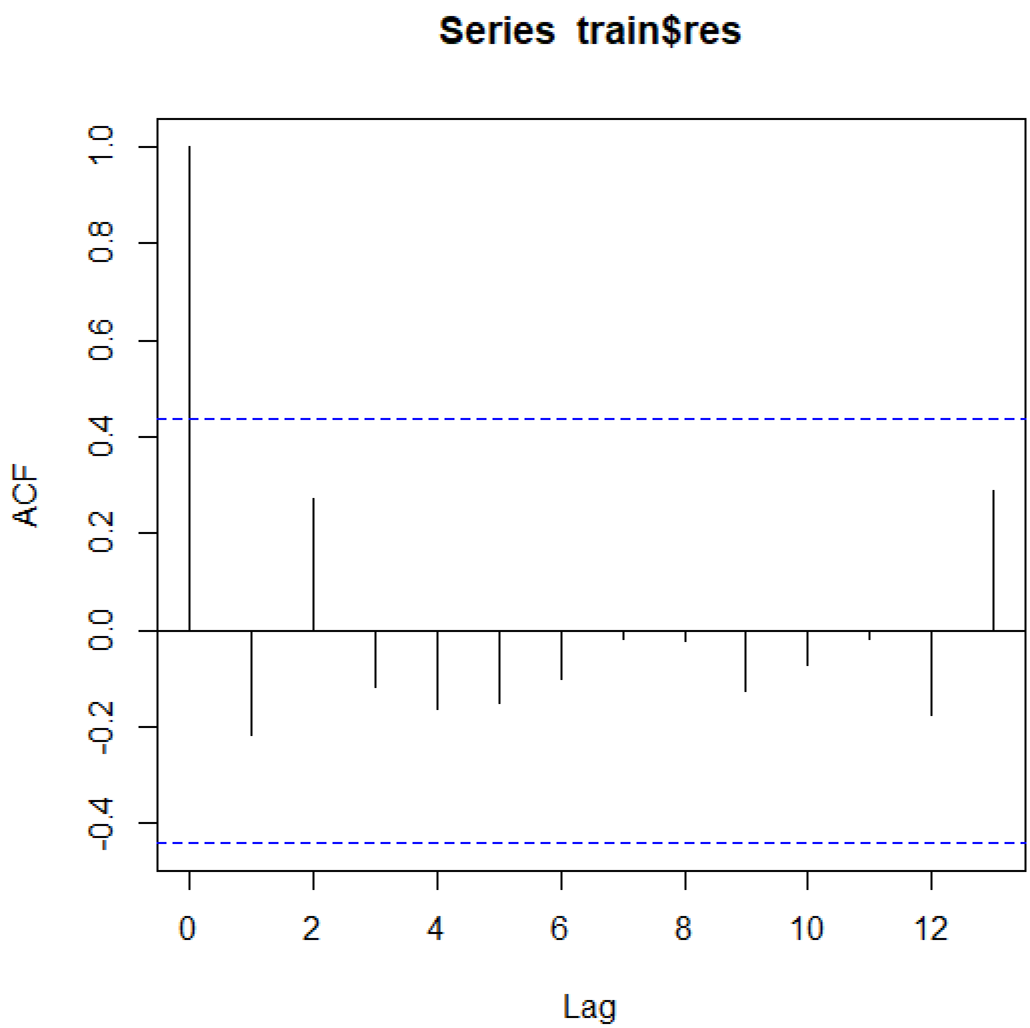
Alternative Hypothesis, H_a : At least one is different from 0.

If we perform the Box-Pierce test for lags 1 to 19, we find that for all of the lags, the p-value is greater than 0.05. From this result, we can say that there are no significant autocorrelations at any of the lags. So, the additive model has produced white noise residuals according to the Box-Pierce test.

So, we can say that the additive model has produced white noise residuals.

White noise checking for the multiplicative model

ACF plot



From the ACF plot of the residuals of the multiplicative model, we do not find the presence of significant autocorrelation at any of the lags. So, we can say that the model has produced white noise residuals according to the ACF plot.

Ljung-Box Test

Null Hypothesis, $H_0 : \rho_1 = \rho_2 = \rho_3 = \ldots\ldots = \rho_m = 0$

Alternative Hypothesis, H_a : At least one is different from 0.

If we perform the Ljung-Box test for lags 1 to 19, we find that for lag 19, the p-value is less than 0.05. From this result, we can say that there are presences of significant autocorrelations up to lag 19. So, the multiplicative model has not produced white noise residuals according to the Ljung-Box test.

Box-Pierce Test

Null Hypothesis, $H_0 : \rho_1 = \rho_2 = \rho_3 = \ldots\ldots = \rho_m = 0$

Alternative Hypothesis, H_a : At least one is different from 0.

If we perform the Box-Pierce test for lags 1 to 19, we find that for all of the lags, the p-value is greater than 0.05. From this result, we can say that there are no significant autocorrelations at any of the lags. So, the multiplicative model has produced white noise residuals according to the Box-Pierce test.

Overall, we can say that the multiplicative model has produced white noise residuals.

Q7: Find out the model that predicts the test data set better.

Ans: We employ RMSE (Root Mean Square Error) for comparing the performance of the additive and multiplicative models.

We find that the additive model gives us an RMSE value of 0.1865539 and the multiplicative model gives an RMSE value of 0.4924547 on the test set.

So, on the basis of the comparison of RMSE values, we can say that the additive model performs better than the multiplicative model in predicting the test set, i.e., in modeling our given dataset of quarterly sales. This is perhaps because we observed that the seasonal variation is relatively constant over time in our data. Using an additive model for modeling the time series data is helpful when the seasonal variation is relatively constant over time. So, the additive model gives a better result than the multiplicative model for the test set.