**Student’s Name: Srishti Ginjala**

**Roll Number: B19084**

**Mobile No: 9440000900**

**Branch: CSE**

# a.

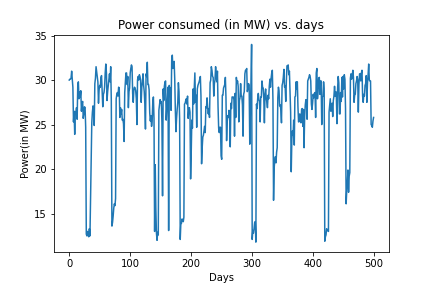


Figure 1 Power consumed (in M W) vs. days

# Inferences:

1. Yes, the days one after the other have similar power consumption.
2. It is because there is repetition in pattern in the plot after regular intervals.

**b.** The value of the Pearson’s correlation coefficient is 0.7675.

# Inferences:

1. From the value of the Pearson’s correlation coefficient, we infer that there is high degree of correlation between the two time sequences.
2. Yes, observations (here power consumption) on days one after the other are similar.
3. This is because the value of Pearson’s correlation coefficient is close to 1 which means that the two time sequences are largely correlated.

**c.**

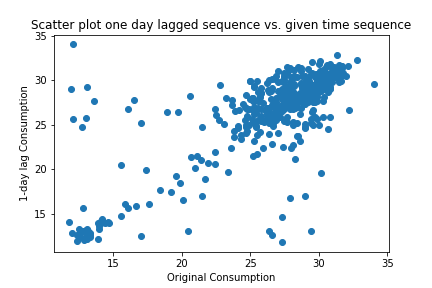


Figure 2 Scatter plot one day lagged sequence vs. given time sequence

# Inferences:

1. From the nature of spread of data points, the nature of correlation between the two sequences is mostly linear.
2. Yes, the scatter plot seems to obey the nature reflected by Pearson’s correlation coefficient calculated in 1.b.
3. The value of correlation coefficient is 0.7 which means that as one parameter increases the other also increases proportionally. This is true because if consumption on a particular day increases the Xt-1 also increases.

d.

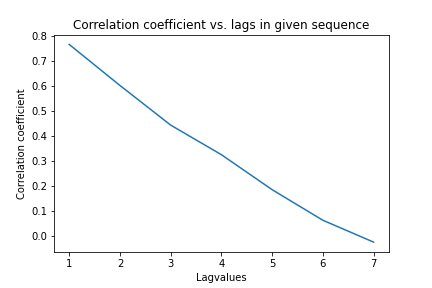


Figure 3 Correlation coefficient vs. lags in given sequence

# Inferences:

1. The correlation coefficient value decreases with respect to increase in lags in time sequence.
2. This is because the value of consumption on a particular day is more dependent on the previous day’s consumption than the consumption 2 or more days before. Hence, the correlation coefficient gradually decreases.

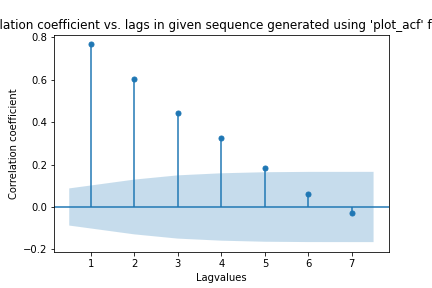


Figure 4 Correlation coefficient vs. lags in given sequence generated using 'plot\_acf' function

# Inferences:

1. The correlation coefficient value decreases as number of lags in time sequence increases.
2. This is because the value of consumption on a particular day is more dependent on the previous day’s consumption than the consumption 2 or more days before. Hence, the correlation coefficient gradually decreases.

# The RMSE between predicted power consumed for test data and original values for test data is 3.192.

# Inferences:

1. The persistent model for the given time series is not very accurate as the RMSE value is high.
2. This is because the consumption may not be the same every day as assumed by the Persistence model.

# 3a.

# 

Figure 5 Predicted test data time sequence vs. original test data sequence

# The RMSE between predicted power consumed for test data and original values for test data is 4.537.

# Inferences:

1. The model is not very accurate for the given time series as RMSE is high.
2. The power consumption may not be dependent on the previous 7-day values but it is positively correlated.
3. The model is reasonably accurate for further prediction.
4. On the basis of RMSE value, the accuracy of current model is less than the model used in question 2.

**b.**

Table 1 RMSE between predicted and original data values wrt lags in time sequence

|  |  |
| --- | --- |
| **Lag value** | **RMSE** |
| **1** | **4.5366** |
| **5** | **4.5370** |
| **10** | **4.5263** |
| **15** | **4.5558** |
| **25** | **4.5141** |

# Inferences:

1. Value of RMSE almost remains same with respect to increase in lags in time sequence.
2. This is because the power consumption is going in cycles that is it is repeating itself after regular intervals of days and hence the RMSE values remains same.

**c.** The heuristic value for optimal number of lags is 5.

The RMSE value between test data time sequence and original test data sequence is 4.5370.

# Inferences:

1. Yes, the prediction accuracy increases up to a point.
2. Correlation Coefficient is decreased from the threshold value so it is considered as optimal hence it is best fit which is why accuracy is increased.

**d.**

The optimal number of lags without using heuristics for calculating optimal lag is 25.

The optimal number of lags using heuristics for calculating optimal lag is 5.

# Inferences:

1. With calculated optimal value RMSE = 4.53701, without RMSE = 4.51413.
2. RMSE error is minimum for lag =25 which discards the fact that lag =5 is optimal value.