IC 272: DATA SCIENCE - III
LAB ASSIGNMENT – VI
Auto-regression

Student's Name: **Mobile No:**

Roll Number: **Branch:**

1   a.

Figure 1 Power consumed (in MW) vs. days

Inferences:
1.   Infer from the plot whether the days one after the other have similar power consumption?
2.   State the reason behind inference 1.

b. The value of the Pearson's correlation coefficient is

Inferences:
1.   From the value of the Pearson's correlation coefficient, what do you infer about degree of correlation between the two time sequences?
2.   We generally expect observations (here power consumption) on days one after the other to be similar. To what extent does it hold true? Answer with respect to the value of Pearson's correlation coefficient.
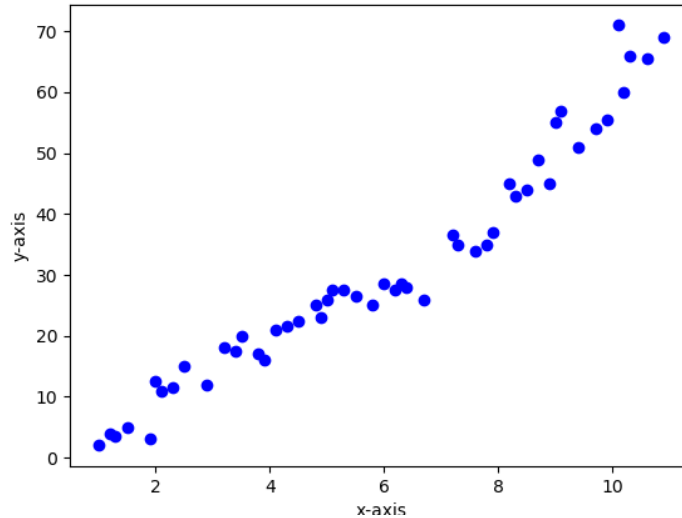3.   State the reason behind Inferences 1 and 2.

c.

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

Inferences:
1. From the nature of spread of data points, what do you infer about the nature of correlation between the two sequences?
2. Does the scatter plot seem to obey the nature reflected by Pearson's correlation coefficient calculated in 1.b?
3. State reason behind inference 2.

   Note: The scatter plot above is for illustration purpose. Replace it with the scatter plot obtained by you. Suitably rename x-axis and y-axis legends.

d.

Figure 3 Correlation coefficient vs. lags in given sequence

Inferences:

1. Infer the trend of correlation coefficient value with respect to increase in lags in time sequence.
2. Explain the reason behind the observed trend.

e.

Figure 4 Correlation coefficient vs. lags in given sequence generated using 'plot_acf' function

Inferences:

1. Infer the trend of correlation coefficient value with respect to lags in time sequence.
2. Explain the reason behind the observed trend.

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

Inferences:

1. From the value of RMSE value comment how accurate is persistent model for the given time series.
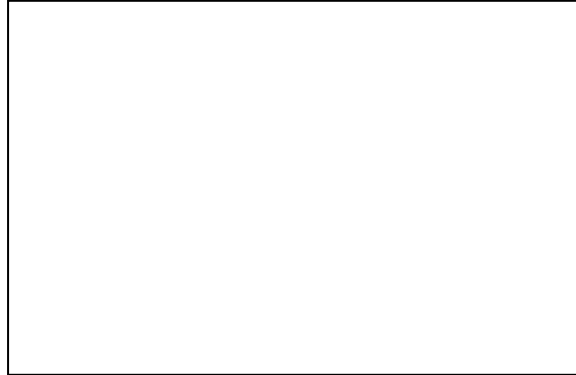2. State the reason behind Inference 1.

3   a.

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 .

Inferences:

1. From the value of RMSE value comment how accurate is the model for the given time series.
2. State the reason behind Inference 1.
3. From the plot of predicted test data time sequence vs. original test data sequence comment how reliable is the model for future predictions with suitable reasons.
4. On the basis of RMSE value, compare the accuracy between the current model and model used in question 2.

b.

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

Lag value	RMSE

Inferences:

1. Infer the trend of RMSE with respect to increase in lags in time sequence.
2. State the reason behind Inference 1.

c. The heuristic value for optimal number of lags is

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

Inferences:
1. Based upon the RMSE value, comment did using heuristics for calculating optimal number of lags improve the prediction accuracy of the model?
2. State the reason behind Inference 1.

d.

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

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

Inferences:
1. Compare the prediction accuracies obtained without and with heuristic for calculating optimal lag with respect to RMSE values.
2. State the reason behind Inference 1.

Guidelines for Report (Delete this while you submit the report):

- **The plot/graph/figure/table should be centre justified with sequence number and caption.**
- **Inferences should be written as a numbered list.**
- **Use specific and technical terms to write inferences.**
- **Values observed/calculated should be rounded off to three decimal places.**
- **The quantities which have units should be written with units.**