

DUE: April 26th, 2019 at 11:59 PM

QMM 1002 Case Study 2

SCENARIO

You have been hired by an environmental management agency to monitor the levels of Lake Erie to ensure they stay within the normal range. The “*LakeErie.csv*” data gives the year, month and water levels in tens of metres above sea level for Lake Erie from 1967 to 2016. Levels have been somewhat high lately and if water levels get too high, there is a risk of flooding. On the other hand, if water levels get too low, shorelines start to recede. The agency wants you to answer the following questions:

1. What is the average water level in Lake Erie?
2. What is the predicted water level of Lake Erie for the next five months? Are water levels increasing or decreasing?
3. A high water level is defined as 15 tens of metres or greater above sea level and a low water level is defined as less than 15 tens of metres above sea level. If a random sample of 100 water level readings is taken, is the count of high and low water levels independent of season (winter, spring, summer and fall)?

CASE STUDY 2

For the scenario above, answer the **three given questions** by creating statistical models and completing statistical tests in R. You must include the following components and explain all of the tests, results, and interpretations thoroughly in your report:

- Plot the Lake Erie time series data, find the average water level, and decompose the time series data.
- Find the best moving average model for the Lake Erie time-series data and plot the model.
- Find the best exponential smoothing model for the Lake Erie time-series data and plot the model.
- Use the best overall model to forecast the water level in Lake Erie for the next five months. Report and plot the forecasts.
- Assess the quality of the best overall time series model and create a plot.
- Take a random sample of 100 water levels. Perform a statistical test to determine if the count of high and low water levels is independent of season.
- Create a plot to determine where differences (if any) arise in high/low water levels by season.

DUE: April 26th, 2019 at 11:59 PM

In your report you must have three sections:

1. Introduction: describe the problem and the three questions that you intend to answer
2. Data analysis: introduce each test or model, explain why you chose the test or model and why it is best, explain the results of the test or model, and interpret the test or the results of the model. Also, include and interpret all plots created.
3. Conclusion: summarize your answers to each of the three key questions.

There is no required length for the report but you must make sure you have all of the required components. Please consult the rubric on the next page for further details.

EVALUATION

A rubric is provided on the next page that details ALL of the requirements of the report.

You must submit a Word document of your report (a template is provided).

You must also submit an R script that includes all of your code. This code will only be run to ensure your models and tests are the same as those detailed in the report. All interpretations and results must be explained and interpreted WITHIN the report.

	0 marks	1 mark	2 marks	3 marks	Gr.
Data Analysis Report	No moving average model is included	A moving average model is included with more than 1 error	A moving average model is included with 1 error	A correct moving average model is included	
	No exponential smoothing model is included	An exponential smoothing model is included with more than 1 error	An exponential smoothing model is included with 1 error	A correct exponential smoothing model is included	
	No random sample is taken	A random sample is taken with at least one error	A correct random sample is taken from the population	N/A	
	No chi-square test is included	A chi-square test is included with more than 1 error	A chi-square test is included with 1 error	A correct chi-square test is included	
	The key questions are not answered	The key questions are answered with some missing detail	The key questions are clearly answered in full detail	N/A	X3
	No correct hypotheses are included for the test conducted	Hypotheses are included for the test conducted with at least 1 error	Correct hypotheses are included for the test conducted	N/A	
	No results of the statistical test and an interpretation are included with error	The results of the statistical test and an interpretation are included with error	The results of the statistical test and a correct interpretation are included	N/A	
	No assumptions and conditions are checked	Conditions and assumptions are checked with 1 error	All conditions and assumptions are checked correctly and in detail	N/A	
	No plot included	Plots are included and interpreted with at least 1 error for: time series data, MA model, ES model, forecasts of model, quality of model, differences in statistical test	Plots are included and correctly interpreted for: time series data, MA model, ES model, forecasts of model, quality of model, differences in statistical test	N/A	X6
Format	Incorrect or no time series models are chosen with	The best time series model is chosen with more than 1 error and/or explanation missing	The best time series model is chosen with 1 error and/or explanation missing	The best time series model is chosen and an explanation is given for MA and ES model	X2
	The model forecasts are not included	The model is used to forecast 5 periods into the future and an interpretation of the values is given with error	The best model is used to forecast 5 periods into the future and an interpretation of the values is given	N/A	
	Numerous spelling errors, incorrect punctuation, and/or severe errors in grammar so that the report is hard to understand.	Frequent spelling errors, incorrect punctuation, and grammar problems that sometimes interfere with understanding.	Occasional lapses in spelling, punctuation, grammar, but not enough to seriously distract the reader.	Very few spelling errors, correct punctuation, grammatically correct, complete sentences.	
Bonus	Incorrect notation and terminology used throughout	More than 2 errors in notation and terminology	1-2 minor errors in notation and terminology	No errors in notation or terminology	
	There are not 3 sections of the report	There are 3 sections of the report as given in the template	N/A	N/A	
	No additional test(s) are conducted		N/A	Additional time series model is included and interpreted correctly	0 or 3 Bonus