G13TSA/MATH3026 & G14TFG/MATH4022

Time Series   2019/20

MATH3026/G13TSA students: Question 1 is ASSESSED and is worth 10% of the total assessment for MATH3026. Coursework should be submitted via the coursework submission area on the Moodle page by Thursday ~~2 April~~ 9 April, 3pm. Please submit a single html file that has been produced by R Markdown in R Studio.

MATH4022/G14TFG students: Question 1 and 2 are both ASSESSED and worth 10% each. Coursework should be submitted via the coursework submission area on the Moodle page by Thursday ~~2 April~~ 9 April, 3pm. Please submit one html file that has been produced by R Markdown for Question 1, one html file produced by R Markdown for the Statistical Report in Question 2 and one pdf or Word file for the Executive Summary for Question 2.

Plagiarism and Academic Misconduct. For all assessed coursework it is very important that you submit your own work. Some information about plagiarism is given on the Moodle webpage.

1. Analyse the following time series datasets which are available in R.

    (1) Nile;

    (2) treering;

    (3) co2.

    Write a short report in html from R Markdown describing the identification, estimation and checking of the proposed models for each dataset. Provide appropriate graphs and clearly justify your choice of model in each case. Do not include any simulations. Note that there is no unique correct answer to each of these problems.

    In case it is useful you can use LaTeX in the comment areas between R chunks (in the white areas), e.g.

    ```
    $X_t - 0.8 X_{i-1} = Z_t - 0.2 Z_{t-1}$
    ```

    becomes $X_t - 0.8X_{i-1} = Z_t - 0.2Z_{t-1}$

2. (**MATH4022 only**) You work as a statistician for an environmental agency and your task is to provide a Statistical Report and an Executive Summary for an analysis of a dataset of lake levels for Mystery Lake. The aim of the project is to forecast the lake level above mean sea level (MSL) for the next three years, and provide uncertainty information. The data are available on the Moodle page in the file `lakelevel.txt` and consist of the MSL levels in months starting in January 1860 and ending in December 2019. The

`lakelevel.txt` file contains a single time series of length $n = 1920$ presented in row order. In particular the first few values are 580.294 580.994 582.1426 584.1317 583.7293 583.0988 582.4333 ...

The Statistical Report should be produced in html from R Markdown and should be suitable for your peer statisticians to read and understand.

The Executive Summary should be maximum of one page of text that is suitable for a broad set of readers, including the CEO of the agency. Please make sure the Executive Summary contains only text and no figures.

Please ensure that you explain clearly what you are aiming to do and that you present, interpret and discuss your results in a professional manner. It is important to explain with comments how you (i) analysed the data (ii) fitted the models (iii) performed model selection (iv) checked the modelling assumptions and (v) calculated the forecasts.

Report length. The reports should not be too long. You should aim to convey the important details in a way which is easy to follow, but not excessively long. Think about someone reading it through and try to help make it easy for them. Make it clear, without too much repetition and avoid long items of numerical output.

Grading Each question will be marked out of 10:

- 5 marks for technical content, use of R and appropriate methods

- 5 marks for presentation and interpretation of results