MATH 70096 - Exploratory Data Analysis & Visualisation

Spring 2022 - Assessed Coursework 2

Deadline: 15 February 2022, 23:59 (UK time)

You should submit an R Markdown notebook, containing your answers to these questions, via the Imperial College Blackboard VLE, by the deadline stated above. Your submission should include both the .Rmd source file for your notebook, as well as the corresponding, compiled PDF document.

This coursework should involve no more than **2 hours** of effort. The available marks are indicated in square brackets for each part of the question.

This coursework counts for 10% **of your total mark for EDAV.**

Plagiarism: Your submission should be *your own work*. Note that software tools are used for plagiarism detection.

Q1) Consider the time series data provided in the accompanying file `time-series.csv`. Load this into R. You may assume this data contains daily observations of some variable of interest.

a) Estimate and report the parameters for the linear trend component of this time series. Remove this component and plot the de-trended data. **[2]**

b) Establish whether the time series contains a seasonal component. If so, estimate its frequency, interpret the seasonality, and use seasonal differencing to remove this component from the data. **[3]**

After removing any trend and seasonal components from the time series, you will be left with a residual time series.

c) Using R, implement the standard Dickey-Fuller test (i.e. using a lag order of `k=1`) for this residual time series. State the conclusion of the test. **[2]**

d) By examining the internal dependence structure of the residual time series, comment on the suitability of the test performed in part (c). **[3]**