Midterm Two - Statistics 153, Spring 2016

Due on 16 November, 2016

November 7, 2016

On piazza, you will find two time series datasets: `Q-1.csv` and `Q-2.csv`. Each of these datasets is of length 209 and gives the google trends data (downloaded on 06 November, 2016) for a particular query from the week of November 06, 2011 to the week of November 11, 2015. Your task is to predict the next 52 observations of these time series.

You are required to work on both of the datasets and you need to turn in the following:

1. Your predictions for the datasets that you have worked on. **These are due by 6:30 pm on 16 November, 2016**. You are required to turn in the predictions on bcourses as a txt-file for each data set. The text file should contain your predictions for the following 52 time points separated by "," or "newlines" and it should be named `Q[Number]_[Firstname]_[Lastname]_[SID].txt`. As an example, I have posted a sample submission for the dataset **Q-1** (this submission contains only 0s) on piazza: it is named `Q1_Soeren_Kuenzel_123456.txt`. We assume that you submit your values in an increasing order:
$$\hat{X}_{210}, \hat{X}_{211}, \ldots, \hat{X}_{261}$$
Please be aware that your submission must be of the right form in order to be valid.

2. A report describing your analysis. For one of the datasets that you have worked on, write a clean report describing your analysis attaching the relevant plots and R output. Include your R code as an Appendix to the report. Do not write a report for both datasets. Just write it for one of those datasets and include your R work for the other datasets as an Appendix. The length of the report including the relevant plots and R output (and excluding the R code) cannot exceed 8 pages. **The printed reports are due at the beginning of class at 6:30 pm on 16 November, 2016**.

You will be graded on the prediction accuracy as well as your report (**the report will be for 25 points and the prediction accuracy will be graded to a maximum of 10 points**). Here is a description of how your prediction accuracies will be evaluated. Let your predictions for **Q-1** be denoted by $\hat{X}_{210}, \ldots, \hat{X}_{261}$ and let the true values of **Q-1** (which we will have access to) are $X_{210}, \ldots, X_{261}$. We will first compute the sum of squares
$$\sum_{i=210}^{261} \left(\hat{X}_i - X_i\right)^2$$
This result will measure your discrepancy for **Q-1**. From here, we will compute:
$$5 * \frac{\text{best discrepancy for } \textbf{Q-1} \text{ in the class}}{\text{your discrepancy for } \textbf{Q-1}}$$
This will be your score for **Q-1**. Note that the maximum possible value for this score is 5. We will similarly compute your score for **Q-2**. To get your final points for the prediction, we will sum the score for **Q-1** and **Q-2**.

You should work completely on your own. You are allowed to use code from the lectures and the section without explicit citation. You are also allowed to consult books or online resources for your analysis but you

must credit all such sources in your report. Anyone caught cheating (which includes copying code, reports etc.) risks failing the class and being referred to the Office of Student Conduct.