

# Final project

Statistics 153 Introduction to Time Series

**Posted:** April 30, 2019

**Due:** May 13th, 2019, 2:30pm

On bCourses you find two different real time series data sets: `data1.csv` and `data2.csv`. Your task is to predict the next 10 observations of these time series.

You need to turn in the following:

1. Your **predictions** for the two data sets (see below for further instructions).
2. A **report** describing your analysis of one of the data sets (see below for further instructions).

**Instructions for the predictions:** Each group is 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 10 time points separated by “,” and it should be named:

`D[Number]_[GroupNumber]_[SID#1]_[SID#2]_[SID#3]_[SID#4\NA]_[SID#5\NA].txt`

. For example, Group 0 consists of three members: Merle, Brian, and Frank. Since there are only three members, the SID slots of members 4 and 5 in the filename are set to “NA”, and the file name would be:

`D1_0_123_234_345_NA_NA.txt`

As an example, Frank will post a sample submission for the dataset `data1.csv` (this submission contains only 0s) on bCourses: it is named `D1_0_123_234_345_NA_NA.txt`. We assume that you submit your values in an increasing order:  $\hat{X}_{N+1}, \hat{X}_{N+2}, \dots, \hat{X}_{N+10}$ . There will also be a test-reader on bCourses that you can use to check whether your submission will be read correctly. Please be aware that your submission on May 13th must be of the right form in order to be valid. **These are due by 2:30 pm on May 13th, 2019.**

**Instructions for the report:** For one of the datasets 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 each of the datasets that you worked on. Just write it for one of those datasets. **The reports have to be submitted via Gradescope by 2:30 pm on May 13th, 2019.** The length of the report including all plots (excluding the R code) **cannot exceed 6 pages** (minimal page margins are one inch per side, minimal font point size is 12). Your report must be clearly structured into the following four parts.

1. **Exploratory Data Analysis:** Transform the data appropriately such that it looks (zero-mean) stationary.
2. **Frequency Domain Analysis** You will analyze the data in the Fourier domain to see if you can detect any seasonalities.
3. **ARIMA Model Selection:** Provide convincing justification why a particular ARIMA/SARIMA/MSARIMA is suitable for the data and also compare different models.
4. **Results:** Estimate the parameters in your chosen model and perform forecasting.

The report should be written with a proper typesetting program, such as Microsoft Word or LaTeX.<sup>1</sup> On bCourses we provide a LaTeX template for a starting point. That template also includes some further details about the required sections.

**Grading scheme:** You will be graded on the prediction accuracy (maximum of 3 points) as well as your report (maximum of 21 points). The report will be graded in the following categories.

- **Technical details:** Are the technical details accurate? For example, are the plots and numerical summaries interpreted properly? We grade part A, B, C, and D separately. (3 + 3 + 6 + 3 Points)
- **Presentation:** Is the report organized? Are the plots visually appealing and informative? Does the report look professional? (3 Points)
- **Writing/Style:** Are the findings stated clearly and precisely? Is there a logical flow in the report? Is the report easy to follow? (3 Points)

**General comments:** 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.

---

<sup>1</sup>To get started with LaTeX, we highly recommend Overleaf. See <https://www.overleaf.com/edu/berkeley> for more information. Berkeley students get a free account. Using Overleaf does not require installation of any LaTeX compiler on your laptop, and makes it easy for multiple people to edit the document, like a Google doc.