STATS 531 Homework 3

Due Sunday 2/8, 11.59pm

This homework is a small data analysis project designed as preparation for the midterm project. In particular, you are required to write your report using qmd format, using the project template at https://github.com/ionides/531w26/tree/main/template. A recommended workflow is to edit the qmd file in vscode using the Quarto extension.

Submit the compiled pdf file and the source qmd file. Also submit any other files, such as `references.bib`*, needed to compile the qmd to pdf. You do not need to submit the data files from the course repository. If the grader runs* `quarto render` *on your qmd file, it should reproduce the pdf.*

You will need to know some Latex to write equations in markdown. Many tutorials exist online, e.g. http://www.latex-tutorial.com/tutorials. It may be helpful to identify equations in the notes that you would like to modify, and then dig out the source code from the course GitHub repository.

Report requirements. Try out some of the ARMA techniques studied in class on the Ann Arbor January Low temperature time series that we saw in Chapter 1 of the notes. This is an open-ended assignment, but you're only expected to do what you can in a reasonable amount of time.

Page limit and format requirements. The main body of the report is limited to 5 pages. The reference list and Supplementary Material sections are not included in this limit, and there is no constraint on their length. You should not change the font size or other format specifications of the template. The strict page limit is a new requirement in STATS 531 this year. Projects will also have a strict page limit, but longer. The rationale for a page limit is: (i) it forces an emphasis on quality over quantity, and this is especially valuable when quantity can be generated easily with AI; (ii) readers have limited attention span, so it is helpful to focus on the most important results for the main body of the report, leaving longer supporting analysis in a supplementary appendix; (iii) professional research reports, e.g. scientific journal articles, have advisory or mandatory page limits for these reasons. If your report is constrained by the page limit, note that figures can often be compressed without loss of information. Make sure that the axis labels remain legible.

Bibliography and source requirements. Your report should contain a reference section listing sources. The grader should be able to clearly identify where the sources were used, using a Bibtex reference in the text. Anything and anyone consulted while you are working on the homework counts as a source and should be credited. The homework will be graded following the posted rubric.

Some advice follows.

1. You can download the data from https://github.com/ionides/531w26/tree/main/hw03/ann_arbor_weather.csv and read it in with

```
import pandas as pd
import matplotlib.pyplot as plt
y = pd.read_csv("ann_arbor_weather.csv", sep='\t', comment='#')
plt.plot(y['Year'], y['Low'])
plt.xlabel('Year'); plt.ylabel('Low'); plt.show()
```

2. Your report should involve model equations and hypotheses, and should define the notation used. Please be careful to distinguish between symbols representing random variables (usually using capital letters) and numbers. You are welcome to follow the notation in the course notes, and if you deviate from this notation it is especially necessary to define the notation that you choose.

3. You are advised to try a few things from the notes, spot something that catches your attention, and try a few more things to investigate it. Write up what you found, and you're finished!

4. To tell Quarto not to include the code in the pdf document, use the `echo: false` chunk option. See the template for examples of `echo`. Presentation of raw code and unprocessed code output is usually best avoided in a report. Focus on discussing the specific numbers that are relevant to your conclusions. The page limit prevents excessive code and computer output in the main body of the report.

5. When you have got everything you can out of the Ann Arbor January Low temperature time series, consider it in the context of the global mean annual temperature time series on the class github site:

```
global_temp = pd.read_csv("Global_Temperature.txt",sep='\s+',comment='#')
plt.plot(global_temp['Year'],global_temp['Annual'], '-')
plt.xlabel('Year'); plt.ylabel('Temperature anomaly (°C)')
plt.tight_layout(); plt.show()
```