STAT 153 & 248 - Time Series
Homework Two

Spring 2025, UC Berkeley

Due by 11:59 pm on 24 February 2025
Total Points = 70 (STAT 153) and 90 (STAT 248)

For problems involving data analysis, you are free to use inbuilt functions from libraries in R or Python. Attach code snippets in your solutions corresponding to each problem.

1. The file "AudioNote.mp3" contains sound from a musical note.

   a) Plot the waveform of the audio signal and describe its main characteristics (**2 points**).

   b) To the waveform, fit the model:

   $$y_t = \beta_0 + \beta_1 \cos(2\pi f t) + \beta_2 \sin(2\pi f t) + \epsilon_t \qquad \text{with } \epsilon_t \overset{\text{i.i.d}}{\sim} N(0, \sigma^2).$$

   Report a point estimate of f along with an uncertainty interval. (**4 points**)

   c) What is your best estimate of the frequency (in Hertz) corresponding to the musical note from which this sound file was generated? (**2 points**)

   d) Based on your frequency estimate, identify which musical note this is. (**1 point**).

2. The dataset `lynx.csv` gives the annual numbers of lynx trappings for 1821-1934 in Canada.

   a) Plot the periodogram of the data and comment on its notable features. (**3 points**).

   b) To this dataset, fit the model:

   $$Y_t = \beta_0 + \beta_1 \cos(2\pi f t) + \beta_2 \sin(2\pi f t) + \epsilon_t \qquad \text{with } \epsilon_t \overset{\text{i.i.d}}{\sim} N(0, \sigma^2) \tag{1}$$

   Provide a point estimate and 95% uncertainty interval for the frequency f and the corresponding period of oscillation. (**4 points**).

   c) Provide point estimates and 95% uncertainty intervals for $\beta_0, \beta_1, \beta_2$ and σ (**6 points**).

   d) Comment on whether (1) is a good model for this dataset. (**2 points**)

3. Download the Google Trends Data (for the United States) for the query *mask*. This should be a monthly time series dataset that indicates the search popularity of this query from January 2004 to February 2025. To this data, fit the single change point model:

   $$Y_t = \beta_0 + \beta_1 I\{t > c\} + \epsilon_t \qquad \text{with } \epsilon_t \overset{\text{i.i.d}}{\sim} N(0, \sigma^2) \tag{2}$$

where $I\{t > c\}$ denotes 1 if $t > c$ and 0 if $t \leq c$.

a) Provide a point estimate and 95% uncertainty interval for the changepoint parameter c. Explain whether your answers make intuitive sense in the context of this dataset (**4 points**).

b) Generate $N = 1000$ samples of the parameters (c, β_0, β_1) from the posterior distribution, and plot the corresponding functions $t \mapsto \beta_0 + \beta_1 I\{t > c\}$ along with the original data. (**4 points**).

c) Comment on whether (2) is a good model for this dataset. (**2 points**)

4. In Figure 1, you will find two different time series datasets (Dataset One and Dataset Two) of the same length $n = 1000$. In Figure 2, you will find two periodograms (Periodogram A and Periodogram B). One of these periodograms corresponds to Dataset One and the other to Dataset Two. Identify the correct periodograms giving reasons for your answer. (**5 points**)

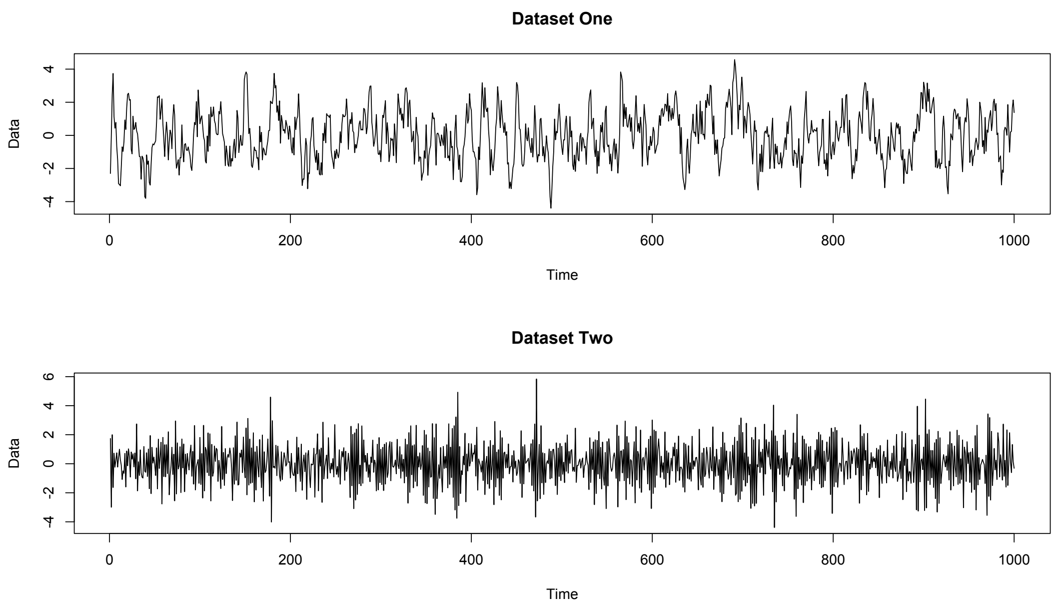

Figure 1: Two Time Series Datasets

5. **DFT of a periodic series**:

a) Suppose $y_0, y_1, \ldots, y_{n-1}$ is periodic with period h i.e., $y_{t+hu} = y_t$ for all integers t and u. Let n be an integer multiple of h i.e., $n = kh$ for an integer k. Suppose that the DFT the data $y_0, \ldots, y_{n-1}$ is $b_0, b_1, \ldots, b_{n-1}$. Also suppose that the DFT of the data in the first cycle (i.e., $y_0, y_1, \ldots, y_{h-1}$) is $\beta_0, \ldots, \beta_{h-1}$. Show that $b_0 = k\beta_0, b_k = k\beta_1, b_{2k} = k\beta_2, \ldots, b_{(h-1)k} = k\beta_{h-1}$ and that all other b_js are zero. (**6 points**)

b) Download the Google Trends Data (for the United States) for the query *halloween*. This should be a monthly time series dataset that indicates the search popularity of this query from January 2004 to February 2025. Remove the data for the last two months so that the data length is 252 (i.e., work with the data from January 2004 to December 2024). Plot the data and describe its main characteristics. (**2 points**)

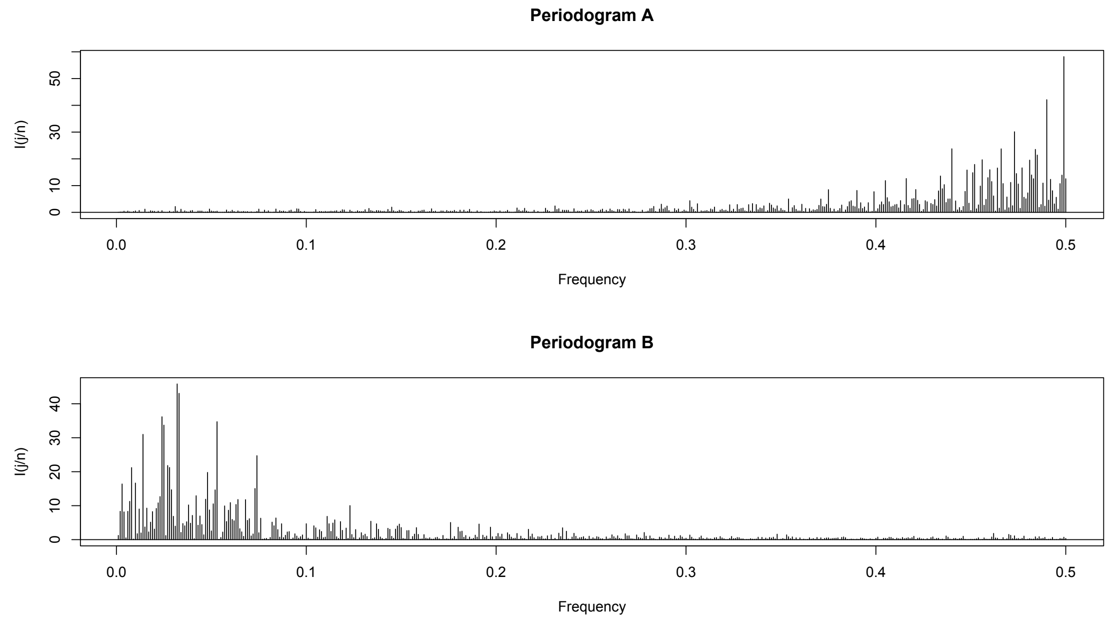

Figure 2: Two Periodograms

c) Plot the periodogram of the data. Explain how it aligns with the mathematical fact derived in Part (a) above. (**4 points**).

6. Download the Google Trends Data (for the United States) for the query *golf*. This should be a monthly time series dataset that indicates the search popularity of this query from January 2004 to February 2025. To this data, fit the model:

$$Y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 \cos(2\pi f t) + \beta_4 \sin(2\pi f t) + \epsilon_t \qquad \text{with } \epsilon_t \overset{\text{i.i.d}}{\sim} N(0, \sigma^2). \quad (3)$$

a) Provide a point estimate and a 95% uncertainty interval for the unknown frequency parameter f. (**4 points**)

b) On a scatter plot of the data, plot your best estimate of the fitted function:

$$t \mapsto \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 \cos(2\pi f t) + \beta_4 \sin(2\pi f t)$$

along with appropriate uncertainty quantification. (**5 points**).

c) Comment on whether model (3) is appropriate for this dataset. (**2 points**).

7. Download the FRED dataset on Total Construction Spending in the United States from `https://fred.stlouisfed.org/series/TTLCONS`. This gives monthly seasonally adjusted data on total construction spending in the United States in millions of dollars from January 1993 to December 2024. To this dataset, fit the model:

$$Y_t = \beta_0 + \beta_1 t + \beta_2 (t - s_1)_+ + \beta_3 (t - s_2)_+ + \epsilon_t \quad (4)$$

with $\epsilon_t \overset{\text{i.i.d}}{\sim} N(0, \sigma^2)$. The unknown parameters in this model are $\beta_0, \beta_1, \beta_2, \beta_3, s_1, s_2, \sigma$.

a) Provide point estimates for the change of slope parameters s_1 and s_2. (**4 points**)

b) On a scatter plot of the data, plot your best estimate of the fitted function: (**2 points**)

$$t \mapsto \beta_0 + \beta_1 t + \beta_2 (t - s_1)_+ + \beta_3 (t - s_2)_+.$$

c) Comment on whether model (4) is appropriate for this dataset. (**2 points**).

8. [**This question is only for students taking STAT 248**] For two datasets $x_0, x_1, \ldots, x_{n-1}$ and $y_0, y_1, \ldots, y_{n-1}$, their *convolution* is the dataset $z_0, z_1, \ldots, z_{n-1}$ defined by

$$z_i = \sum_{j=0}^{n-1} x_{i-j} y_j \qquad \text{for } i = 0, \ldots, n-1$$

where $x_{-k} = x_{-k+n}$. Find the DFT of $z_0, \ldots, z_{n-1}$ in terms of the DFTs of $x_0, \ldots, x_{n-1}$ and $y_0, \ldots, y_{n-1}$. (**6 points**)

9. [**This question is only for students taking STAT 248**] The file "DataProblem9HomeworkTwo248Spring2025.csv" contains a time series dataset having two columns. The first column is the time index t (here time is not equally spaced), and the column is the value of the time series y_t. This data was generated according to the model:

$$\begin{aligned} y_i = {} & \beta_0 + \beta_1 \cos(2\pi f_1 t_i) + \beta_2 \sin(2\pi f_1 t_i) \\ & + \beta_3 \cos(2\pi f_2 t_i) + \beta_4 \sin(2\pi f_2 t_i) + \epsilon_t \end{aligned}$$

with $\epsilon_t \sim N(0, \sigma^2)$.

a) Provide point estimates and 95% uncertainty intervals for f_1 and f_2. (**5 points**)

b) Provide points estimates and 95% uncertainty intervals for each $b_j, 0 \leq j \leq 4$ as well as σ. (**9 points**).