# Math 545: Data analysis assignment 1
# Due on: Monday, February 27, 5pm.

The goal of this data assignment is to become familiar with the basic techniques of time series analysis and preprocessing. *Please implement all the functions you need.* Using high-level built-in functions (e.g., for evaluating moving averages) won't qualify for credit; it is good to do it by hand once to assure that you understand the mechanics.

Please submit the files via Gradescope.

**Task 1, 0 pts**

Open the file 'electricity.xls'. The first tab called 'State Assignment List' and find the state assignment on the line with your name (e.g., CA for California). On the tab 'Monthly-States', select your state from the drop down menu. Copy the 'Residential/Sales' column into a separate .xls or .csv file or tab (see 'Example' tab; you need only 'Residential/Sales' data for analysis). The 'Sales' are expressed in megawatthours, so we will refer to the data as 'Megawatthours' below. *.csv/*.xls data can imported/transformed into the format of your preference that is compatible with the software you are using.

**Task 2, 20 pts**

Plot the 'Megawatthours' data. Plot the logarithms of 'Megawatthours' data. Describe your empirical observation of the difference between 'Megawatthours' and its logarithmic plot (in terms of variability as time increases). Which of these time series better fit the classical decomposition model?

**Task 3, 20 pts**

*In what follows, 'Megawatthours' and 'original data' refer either to the raw sales data, or to their logarithms, depending on your answer in Task 2.*

Estimate the trend in 'Megawatthours' data via a 12-term simple moving average and plot it on top of the original 'Megawatthours' data (for the first and last 6 points, use 'constant interpolation' – e.g., replace the moving average for observations 1-6 by the moving average for the 7th observation).

**Task 4, 20 pts**

Estimate the seasonal adjustment for each month and present results as a 12x2 table (month-adjustement).

**Task 5, 20 pts**

Subtract the trend and the seasonal component from the original time series and plot the resulting data.

**Task 6, 20 pts**

Estimate the trend by fitting the (global) quadratic polynomial to the original data *minus the seasonal component.* Plot it on top of the 12-term moving average (you can consult Chapter 1 section 4, e.g. equation 1.4.3, in the textbook; this is not a local polynomial estimator!). Plot it on top of the 12-term moving average, and compare the results.

**Task 7, 20 points**

Find the coefficients of the moving average that corresponds to a local polynomial estimator of order 1 (“local linear estimator”) when the window of size 3 is used (meaning that the moving average is of order 7). Subtract the seasonal component *only* from the original time series, and estimate the trend via applying the filter you found to deseasonalized data. Compare the result to the trend estimate obtained via a simple moving average of order 12.

**Your submission should include a (preferably *.pdf) file containing**

1. your code in the software of your choice;
2. the plot of ‘Sales’ in Megawatthours;
3. natural logarithm of ‘Sales’;
4. ‘Megawatthours’ (original or the logs, depending on Task 2), moving average, and fitted quadratic polynomial (all on the same graph);
5. de-trended and de-seasonalized data;
6. the table with seasonal adjustments for 12 months;
7. plot from task 6;
8. short comments required by tasks 1-7, and the (short, no need to explain every step) derivation of the local polynomial filter from task 7.