

# Statistics for Engineering, 4C3/6C3

## Assignment 4

Kevin Dunn, [kevin.dunn@mcmaster.ca](mailto:kevin.dunn@mcmaster.ca)

Due date: 27 February 2014

**Note:** I strongly recommend you submit this [assignment electronically](#) (see instructions on the course website), so that you can practice using the system for the course project.

### Question 1 [8]

The [Paper basis dataset](#) contains data, sampled 30 seconds apart, of the basis weight (a measure of paper density) from an industrial source.

Show the autocorrelation plot for the data, and interpret the plot.

600-level students: also confirm that your interpretation is correct by sub-sampling the vector and repeating the autocorrelation test. *Hint:* use the `seq(start_from, end_at, step_size)` command in R to subsample a vector.

### Question 2 [6, for 600-level students only]

Another interesting data set is the [Aeration rate](#) is the amount of air added to a [sparging tank](#).

Use the autocorrelation function on this data set, show the plot, and carefully interpret what the results imply. *Hint:* you will notice there is a missing value in the data set, so use the `na.action=na.omit` as the second input into the `acf(...)` function.

### Question 3 [16]

This question uses two data sets. You may answer the question using either one of the data sets (your choice), however 600-level students are expected to use both data sets and compare the results side-by-side (i.e. don't repeat your analysis a second time below the first, do your analysis on both datasets simultaneously, making comparisons between the two data sets). Even 400-level students are encouraged to examine both data sets. Your answer may not exceed 4 pages.

1. Data from [CHEM ENG 4M3, 2013 class](#)
2. Data from [CHEM ENG 4N4, 2013 class](#)

The data is related to the time duration of students writing midterm tests, and it also records the grade the student achieved on the test. One column in the data is labelled as `Grade` [percentage] and the other as `Time` [minutes].

This question is of an exploratory nature. You may consider the ideas below, but also feel free to add to these:

- Explore the data: is there a relationship between the grade and time taken to write the test? (Consider learning about and using the `lowess` function in R.) How would you describe the relationship?
- Should a regression model use `Time` or `Grade` as the input variable?
- Build a suitable regression model using these two variables. What conclusions do you draw from the model?
- Investigate whether the assumptions for regression models hold true.
- What advice would you give to students based on these results?
- What result(s) do you learn from these data that is(are) useful for course instructors to know?