**Analytical Methods – Week One Overview**

In addition to getting oriented to the course, our goal this week is to begin the path toward a fuller discussion of feature engineering in upcoming weeks. This week’s material is divided into three modules. The first contains a series of readings on feature engineering and related topics. Please read the first two in detail and skim the third, which is very technical. We will return to the third reading in later weeks. There is a required discussion with posted instructions.

The second module for the week focuses on correlation and causation. Once again, there are some readings with a required discussion. I have provided a data set consisting of 20,000 very random numbers arranged as 20 columns with 1,000 entries each or, looked at the other way, 1,000 rows with 20 entries each. You will explore in a data exercise ways correlation behaves from each of these perspectives. The goal is to raise your awareness of what correlation data are or are not telling you, especially in the absence of underlying contextual knowledge. Also, please consider finding and working with some additional datasets to explore some of the nonparametric methods for correlation discussed in this module. You should seek data that demonstrate relationships which are not necessarily linear as well as data with outliers. See if these correlation alternatives might be useful to you in some setting.

The third module for this week focuses on an idea from a number of years ago called Grainger Causality. It is meant to illustrate a middle ground between data associations which are only correlational and the strict experimental design requirements necessary to get some confidence in a truly causal relationship. Grainger won the Nobel Prize in Economics in 1969. You can find his speech via Google. You will play with some data which attempt to answer the age old question of the chicken and the egg as a way to get a sense of Grainger Causality.

So, your assignments this week are as follows:

* Compete all required readings
* Participate in the two required discussions. Add your own thoughts and comment on the ideas of others.
* Complete the two required data exercises. Comment on the ideas others offer in those discussions.
* If you wish, dive more deeply into the non-parametric options for correlation.
* Begin thinking about your project.

I’ll see you at the Webinar on Tuesday at 6:45 p.m. Central, which will be recorded.

Jim