MSAN 601: Linear Regression Analysis
Quiz 1

by James D. Wilson (University of San Francisco)

There are 5 questions. Each question is weighted equally. Grades will be out of 100 points. You have until 9:55. Good luck!

1. Let X be an $n \times p$ data matrix as described in class, where n is the number of samples and p is the number of features measured. Using the dimensions of X as a reference, describe the three major data paradigms: (i) *traditional*, (ii) *high dimensional*, and (iii) *big data*.

2. Briefly (in 3 sentences or less) describe the differences between *supervised* and *unsupervised* learning. Be as specific as possible here.

3. Briefly outline (in a list format) the steps involved in the statistical modeling process. Note that you can simply state the steps here and that you do not have to give any further details about each step.

4. Describe the difference between *explanatory* and *predictive* modeling. For the following examples, indicate which of these two approaches is most appropriate:

   (a) Using the demographic features *age*, *household income*, and *amount of time on Netflix*, to most accurately determine what movies a person would like to watch on Netflix.

   (b) Determine what features of a house on the market are most influential in determining its cost.

   (c) Determine what characteristics are tied to a firm's max bid in an online auction.

5. Let X be an $n \times p$ data matrix and y an n-dimensional response vector. Write down a regression model relating y with X. Is regression a supervised or unsupervised learning problem?