Data Science Methods for Clean Energy Research

Week 2, Lecture 2: Statistics

January 4, 2017

UNIVERSITY of WASHINGTON

Outline

- Goals for statistics lectures
- Topics to be covered
- Warm up
  - Developing some intuition about sample sizes and expectation
  - Setting the stage
- Basic descriptive statistics and a bit of practice in python

Goals: why teach probability and statistics?

- Spend some time reading answers online to "what is the difference between statistics and machine learning?" (or also *statistical modeling*)
- Not going to belabor this point, because there is no good answer
- An imperfect qualitative answer is:
  - Statistics is more concerned with quantifying uncertainty
  - Machine learning is more concerned with making predictions
- Statistics and statistical modeling existed long before machine learning, we begin there

Topics to be covered in next 1.5–2.5 lectures

- **Descriptive statistics**
  - **Given some (*data, measurements, observations, calculations*) , what can we directly calculate, and what does it mean?**
- **Statistical distributions**
  - **Do we expected our data to fall any regular ordered pattern – and do they?**
- **Hypothesis testing**
  - **What can we say about our data and with what level of certainty**
  - **What the hell is a p-value, for real?**
- **How can we do more with a given data set** (resampling) **or** (bootstrapping)

Rest of lecture – create a directory for W2L2 and a new Jupyter notebook

- Warm up
- Basic descriptive statistics and a bit of practice in python

UNIVERSITY *of* WASHINGTON