ISyE 2028 – Basic Statistical Methods - Fall 2015

Bonus Project: “Big” Data Analytics

Big data is about the volume of the data, the variety and the velocity of the data. But it is also about what can we do, and how we to with the data: to extract the information that people were not able to find from a piece of data, or demonstrate the novel use of data to achieve something new. In this project, we will perform some “big” data analysis using methods you learned in class. The dataset you use does not necessarily have to be really large. But you should come up with a novel way of using data.

In the following we provide a guideline for the project.

Step 1: identify a problem, which can be an estimation problem with confidence interval, a hypothesis you would like to test, making prediction using a linear regression model, or anything you would like to study or verify using data.

Example statistical problems:

(1) Test whether or not there is a difference in the daily coffee consumption by engineering student versus non-engineering student.
(2) Estimate the expected difference in the hours spending in front of computers for male versus female student, for different majors, etc.
(3) Estimate the number of hours of study towards the end of the semester.
(4) Fitting a model for the travel time on major Atlanta road during rush hour.
(5) Find whether or not there has been an increasing trend in campus crimes reported by the Georgia Tech police.

Step 2: ahead to collect data yourself, find data from Internet by crawling data yourself (if you are really good at programming), or identify an existing dataset from some source.

Example data source:

(1) Handing out and collecting survey at Tech Square.
(2) Online survey with classmates.
(3) Health or life style data recorded using smart phone apps by friends, or dorm mates.
(4) Travel times recorded by yourself.
(5) Expenditures.
(6) Data you find from Internet.

Step 3: analyze data using methods we did in class. Interpret your results, whether or not it makes sense or are according to what you expect. Do you have a large enough sample size? Is your finding statistically significant? Do you have a representative sample or its biased? How would you improve your result? Can you use it for anything real?

Evaluation: Initial project proposal (1 page using the word template provided), due **Nov. 1,** 2015. Final report (minimum 2 pages, using the word template provided) about your project, describe the 3 steps, due **Dec. 10**.

We will post your project the findings on a wordpress blog for our class

https://isye2028fall15.wordpress.com

This project accounts for 40 bonus points.

And finally, be creative, be interesting, and be bold.