
ISyE 2028 – Basic Statistical Methods - Fall 2015
Bonus Project: “Big” Data Analytics
Proposal Report

Bonus Project: “Big” Data Analytics

We often receive emails from the Georgia Tech Police about crimes occurred on or around our campus. That incited me to study the average number of crimes reported to our police in order to have an idea on how safe our campus and its surroundings are and what actions can be taken in order for students and the school personnel not to fall victims of those reported crimes.

In our experiment, we will observe the number of crimes reported to the Georgia Tech Police for the past 90 days (from 7/19/2015 to 10/16/2015). The data will be collected from the Georgia Tech Police website: <http://www.police.gatech.edu/crimeinfo/crimelogs/>. The number of crimes reported will be collected from the 2015 crime log excel file that contains the incidents numbers, dates, times, locations, descriptions, and case statuses. Since, we are only interested in the number of crimes reported each day for 90 days, we will make a second excel file containing only the days and the number of crimes reported each day by counting the number of reported crimes per day and summing them up together in order to get the total number of crimes reported each day. The R software will be used to read the newly made excel file and extract important computed data about the crimes on and around our beautiful campus.

We will learn what the day with the maximum number of reported crimes and the day with the minimum number of reported crimes are and what those data values are. We will derive the sample mean through the summary of the data. We will construct the boxplot, the stem and leaf plot and histogram of the data in order to extract specific information about the collected data. Finally, we will conduct a further data analysis of the data by constructing the 4 following confidence intervals for the mean value of reported crimes (90%, 95%, 99% and 100%) using The R software. With the results, we will state how confident we are the sample mean will fall within a certain range.