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

Bonus Project: “Big” Data Analytics

Step1: estimation of average daily number of crimes occurring on and around the Georgia Tech campus with confidence intervals

We often receive emails from the Georgia Tech Police about crimes occurred on or around our campus. That urged me to study the average number of crimes committed 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 order for my study to be accurate, I chose to define the word crime. In my study, crime refers to burglary, strong-arm robbery, simple assault/battery and theft of automobile.

Step2: data collection source: <http://www.police.gatech.edu/crimeinfo/crimelogs/>

In our experiment, we observe the number of crimes reported to the Georgia Tech Police in a 90-day span (from 7/19/2015 to 10/16/2015). The data was collected from the Georgia Tech Police website: <http://www.police.gatech.edu/crimeinfo/crimelogs/>. The number of crimes reported was 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, a second excel file containing only the date of the crimes, the number of crimes, the nature of the crimes and the locations where those crimes occurred was made. The newly made excel file is then saved as Project.csv. From there, the R software was used to read the Project.csv file and extract important computed data about the crimes on and around our beautiful campus.

Step3: data analysis:

We learned that in a 90-day span, the maximum number of crimes reported to the GT Police was 3 which occurred on 10/2/2015 (Friday), and the minimum number of crimes reported was 0 which occurred most of the days (mode). The average number of crimes is represented by the sample mean which is 0.3, almost 1 reported crime every 3 days. These numbers are confirmed by the boxplot that also informs us that the lower quartile (Q1) is 0.0 and the upper quartile (Q3) is 0.0. What that tells us is that 25% of the number of reported crimes are lower than 0.0 and 25% of the number of reported crimes are greater than 0.0. Therefore, the inter-quartile range (IQR), which is by definition $Q3 - Q1$ is 0.0 (0.0 - 0.0). To further confirm our findings, a stem and leaf plot and histogram were also constructed. According to the stem and leaf plot, a large number of the data values are in the 0s which sort of predicts the mean value being 0.3. These results do mirror the results from the histogram which also shows that the number of crimes between 0 and 0.5 were the most frequent. The histogram is skewed to the right and that confirms the fact that the mean (0.3) is to the right of the median (0.0). We see that on the right side of the graph, there are fewer reported crimes. The number of times 2 or more number of crimes were reported is 4 which is confirmed by the stem and leaf plot and the histogram. Since the differences between the left and right sides of the histogram are very significant, the data shape can be classified as skewed to the right. We also see that the mode of the distribution is 0.

Knowing that the crime rate at and around the Georgia Tech campus is on average 0.3 per day, using Poisson's law we made the following calculations:

- a) the probability that in a given day some crimes will occur
- b) the probability that 1 or more crimes but less than 3 crimes per day will occur
- c) the probability that a crime will occur in any given hour of the day?

Considering $\mu = 0.3$, using the R software, we've obtained the following results:

- a) "Some crimes" means "1 or more crimes", $P(X > 0)$. The probability is 0.2591818 which is less than 1/3. Therefore, we are almost certain that a crime will occur on or around the Georgia Tech campus about once every 3 days.
- b) The probability of 1 or more crimes but less than 3 crimes occurring per day is $P(1 \leq X \leq 3) = 0.28915968$, which is pretty low.

c) The probability that a crime will occur in any given hour of the day is $P(X) = 0.03996623$

For curiosity and more in-depth study, we conducted a further 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. The following results were obtained:

90% confidence interval for mean value: [0.197, 0.403]

95% confidence interval for mean value: [0.177, 0.423]

99% confidence interval for mean value: [0.137, 0.463]

100% confidence interval for mean value: (-infinity, +infinity)

What the above results tell us is that we're 90% confident that the sample mean value of reported crimes will fall in the interval [0.197, 0.403], 95% confident that the sample mean value of reported crimes will fall in the interval [0.177, 0.423], 99% confident that the sample mean value of reported crimes will fall in the interval [0.137, 0.463], and 0% sure of what the exact value of the sample mean value of reported crimes will be. We notice that the more confident we are (99%), the wider the interval is ($0.463 - 0.137 = 0.326$), and the less confident we are (90%), the narrower the interval is ($0.403 - 0.197 = 0.206$).

After learning about the daily number of crimes reported to our police from these data, we are pretty confident that a crime will occur on or around the Georgia Tech campus once every 3 days. Given that piece of information extracted from our data analysis, we know that our campus and its surroundings are not as unsafe as some believe it to be. The good news is that the crime rate is pretty low and the bad news is that crimes still do occur on our campus albeit at a low rate. Therefore, when venturing off campus, let us be aware of our surroundings for our own safety and the safety of our fellow schoolmates. Since, it's almost a certainty that a crime will happen around our campus once every 3 days, let us do our best to make sure we are not the victims of those crimes. Let us have a safe a fun educational journey here at Georgia Tech, our alma mater.

Reference

<http://www.police.gatech.edu/crimeinfo/crimelogs/>