
ISyE 2028 – Basic Statistical Methods - Fall 2015

Bonus Project: “Big” Data Analytics

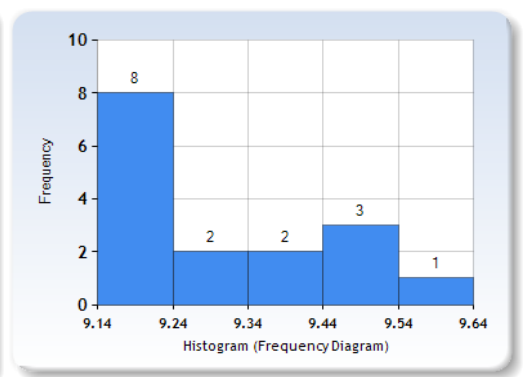
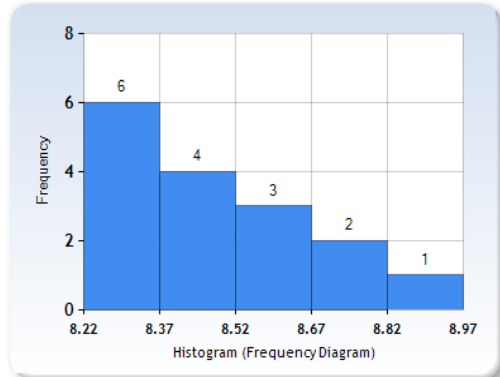
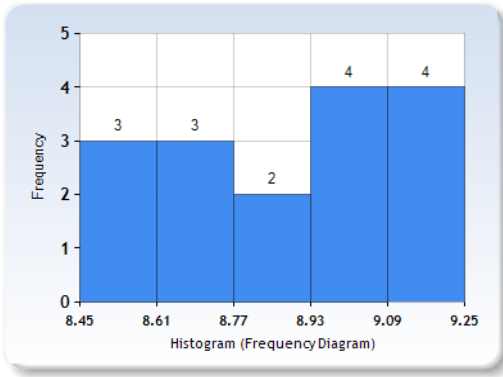
Final Report

My proposal for this project involved the efficiency of various routes to the CULC from Tech Square, a common distance Tech students must walk. With the opening of Square on Fifth and University housing, as well as the many restaurants students enjoy eating lunch at, this route is becoming increasingly common. Though a trolley stop may be centrally located in tech square, the trolley system is often unreliable or overcrowded, so having an efficient way to walk is important. The following report details my findings regarding three different routes I walked personally.

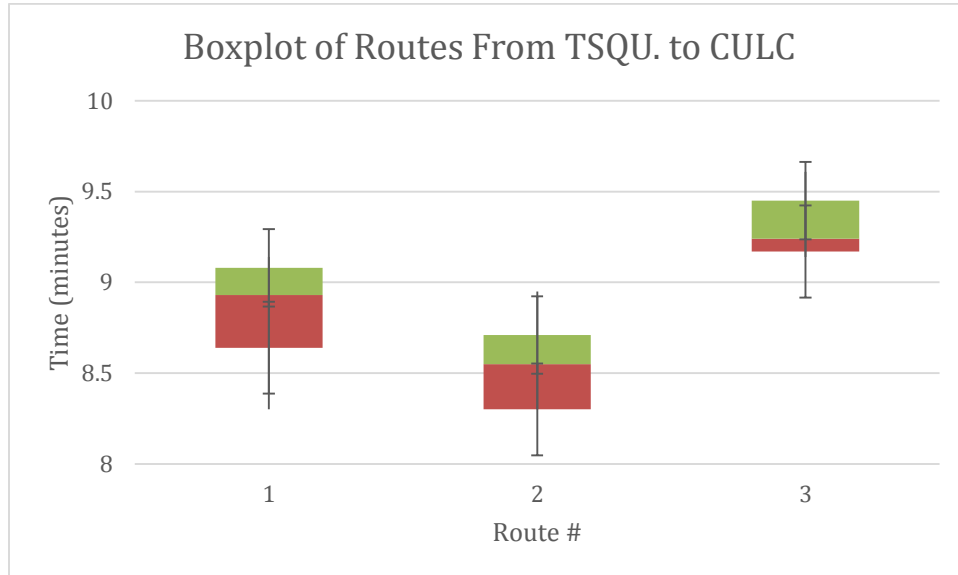
Before we can get into the data, I must first list the routes I took. The first route is down 5th St. NW, then turning left at the Klaus Advanced Computing Building, electing to cut through the College of Architecture Building. After that, it is a simple walk to the northern entrance of the CULC. The second route is down 5th St. NW, then turning left at Fowler St. NW. When Fowler ends, I turn right onto 4th St. NW, then taking it almost directly to the northern entrance of the CULC. The final route is taking 5th St. NW to Klaus, then cutting through Klaus and turning left for a straight shot south to the northern entrance of the CULC. I have designated these routes as Route 1, Route 2, and Route 3 for the purposes of this report. I timed my walk on each route 16 times, meeting the goal I set for myself in the proposal.

First, I calculated some relevant descriptive statistics for each of the three routes. The following are the results, listing in this order: maximum, minimum, mean, median, and standard deviation. The times are all in minutes. For Route 1: 8.45, 9.14, 8.86, 8.93, and .2393. For the Route 2: 8.22, 8.87, 8.473, 8.47, and .1946. For Route 3: 9.14, 9.61, 9.303, 9.24, .1505. From these statistics, I already got a sense of what I thought the answer to my question would be. Route 2 appears to most efficient. However, I decided to continue onto a series of graphical displays, to further explore the subject.

The first graphical display I constructed was a histogram. Here are the histograms for each Route, starting with Route 1 on the left, and ending with Route 3 on the right:



As you can see, the three histograms are certainly not normal, which does not bode well for any hypothesis testing, or confidence intervals. The first appears to be fairly flat, with a high variance. The second is skewed heavily to the right. The third is also heavily skewed to the right, with an inordinate amount of data entries to the left. Unfortunately, the amount of data is also too low to assume normality, but none-the-less, I will create some confidence intervals to show for certain which route is most efficient. However, first comes the second graphical display I created: boxplots. Here is the boxplot comparing each route to each other:



This boxplot shows pretty clearly what I already suspected: Route 2 is the most efficient route. At the very least, we can rule out Route 3 as the most efficient route, as its minimum data point is higher than almost every other data point in Route 1, and over half the data points in Route 2. So, it comes down to Routes 1 and 2: which one is most efficient? To answer this question, I

decided to create confidence intervals for each Route, so I can conclude with a reasonable degree of confidence which route is most efficient.

I decided on a 95% confidence interval for the mean, using t-values since we don't know the population standard deviation. For Route 1, my results were $8.86 \pm .12$, or $[8.74, 8.98]$. For the Route 2, my results were $8.47 \pm .1$, or $[8.37, 8.57]$. For Route 3, my results were $9.3 \pm .07$, or $[9.23, 9.37]$. Using these results, it is clear that Route 2 is most efficient, since the upper bound on Route 2, 8.57, is smaller than the lower bound on Route 1, 8.74. Route 3 still lags far behind both, with a minimum bound of 9.23.

In conclusion, based on the data, Route 2 is clearly the most efficient route to the CULC from Tech Square. Route 1 was relatively close, but the confidence intervals made it clear in the end which Route was superior. Route 3 was fairly clearly from the beginning not the most efficient route, though it did have a lower variance than the other routes. This project was an interesting experience, and I certainly appreciate statistics more after completing it.