Estimation of True Difference in Internet Latency Between Charter Internet and Kalamazoo College ISP

Daniel Michelin, Erik Hartig, and Nicholas Swain

June 11, 2018

Abstract

In this paper we estimate the true μ of the difference between ping times between Kalamazoo College and a residence within a mile located at 416 Oak St. We hypothesize that a significant difference in latency as measured by ping times is likely indicative of substantially different network architecture.

1 Introduction

A good network connection can be evaluated in a number of ways. One way it can be looked at is through a measure called latency, which is the time taken to send a packet of information from one point to another. A packet of information is a fundamental unit of data sent over packet-controlled networks such as the Internet. Another common way is to look at bandwidth, or the amount of data that can be transmitted at a single point of time. The difference between the two can be thought as analogous to delivering mail by hand. If a single packet can be thought of as a letter, then the latency can be thought of as the time it takes for the mail-person to get a single letter from point A to point B. Bandwidth by the same analogy can be thought of the number of letters that the mailperson is transmitting at a particular time. While it can be easy to confuse the two, a packet can have a low bandwidth and high latency, and vice versa. In the former case that mailperson could be transmitting mail using a high speed motorcycle: not much mail can go through at any particular time, but the mail that is on the network is rapidly transmitted. The latter case could be transmitting mail through a large truck: high capacity, low speed. Although TCP/IP, the fundamental architecture of the internet, has a maximum packet size of 64 Kilobytes the realistic maximum size of a packet is capped out usually at 1500 bytes, the maximum size of an ethernet packet. Most consumer internet connections run through ethernet, meaning that a single packet of information typically has a negligible effect on a network. Bandwidth is typically effected by traffic on the network, latency is not unless if the bandwidth has run out on

a network. On most consumer internet plans the various tiers of paid internet service have the same latency, and a differing latency between two networks often means that there is something physically different between the networks, such as a shorter physical connection or faster wires (DSL vs Fiber optic for instance.). Ping is a software utility tests the network latency of a single packet and is featured across all popular operating systems. Our hypothesis is that there is a difference in means between Kalamazoo College's and the Oak Street residence's latencies as measured by ping time to google.com. We can state this as:

$$H_0 : \mu_{kzoo} = \mu_{oak}$$
$$H_A : \mu_{kzoo} \neq \mu_{oak}$$

The alternative hypothesis is indicative of two things, the first being that the network itself is faster, and/or possibly that the network follows a different path to get to google.

2 Methods

We collected four datasets, two to estimate population parameters and two for exploratory data analysis. Samples collected from Kalamazoo College and the Oak Street residence that are used to estimate the μ of the population of the difference in means were collected at approximately the same time (within seconds of each other) to control for overall network business. These times were conducted by two laptops running a sample collection script in python that is included in the appendix. The Oak Street residence was chosen because of it's short distance from Kalamazoo College to minimize the impact of absolute physical distance being a factor in the latency difference, as large discrepancies in physical distance alone can contribute to wide differences in latency times, such as comparing a laptop in Google's headquarter's latency to google compared to a laptop in a Starbucks in Kalamazoo, MI. The script calls the Windows ping command line utility then parses the output for the correct value. The script takes 36 samples at an interval of 5 minutes each and records the location and latency. Laptops tested used the same Wifi chip to account for possible increase in latency there. Latency from router distance was controlled by being approximately the same distance from each one (about 6 feet with line of sight). Latency derived from the router itself is negligable. Tests were independent as a single packet is highly unlikely to impact the network in a measurable way.

3 Results

To get an idea of the sampling distributions we look at histograms of the two distributions and summary descriptions of each. In figures 1 and 2 the distributions appear non normal. We confirm the non normality in figures 3 and 4. Because of the non-normality it is important that for estimating the population parameter μ of the difference in means between the two sampling

Figure 1: Histogram of observed Kalamazoo College Ping Times
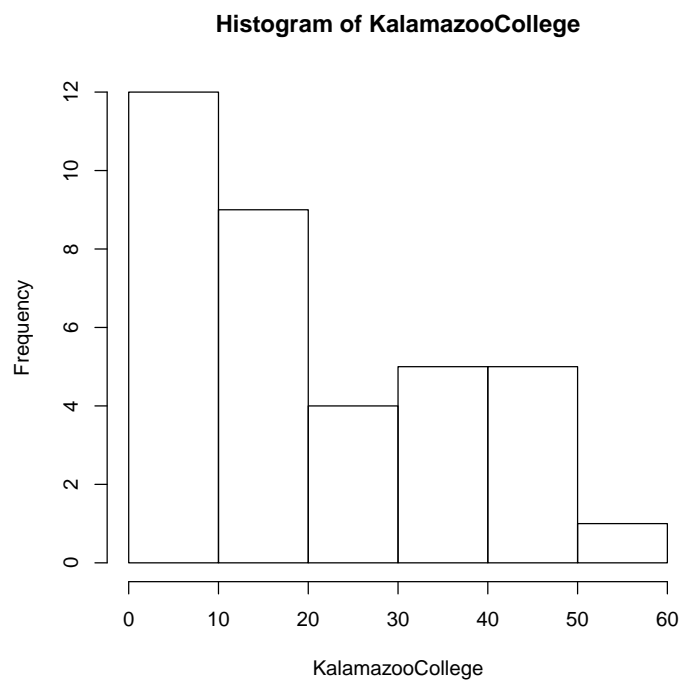
```
> hist(KalamazooCollege)
```

Histogram of KalamazooCollege

Figure 2: Histogram of observed Oak Street residence Ping Times

```
> hist(PersonalResidence)
```
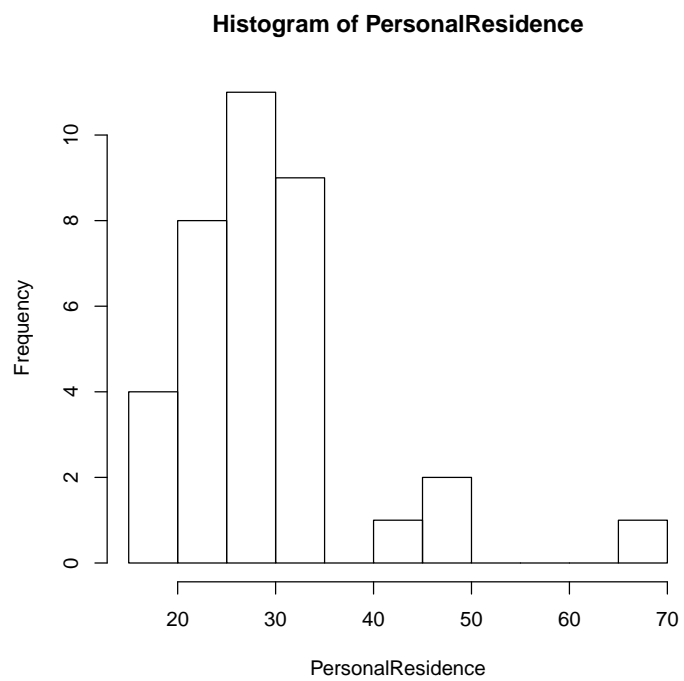
Histogram of PersonalResidence

PersonalResidence
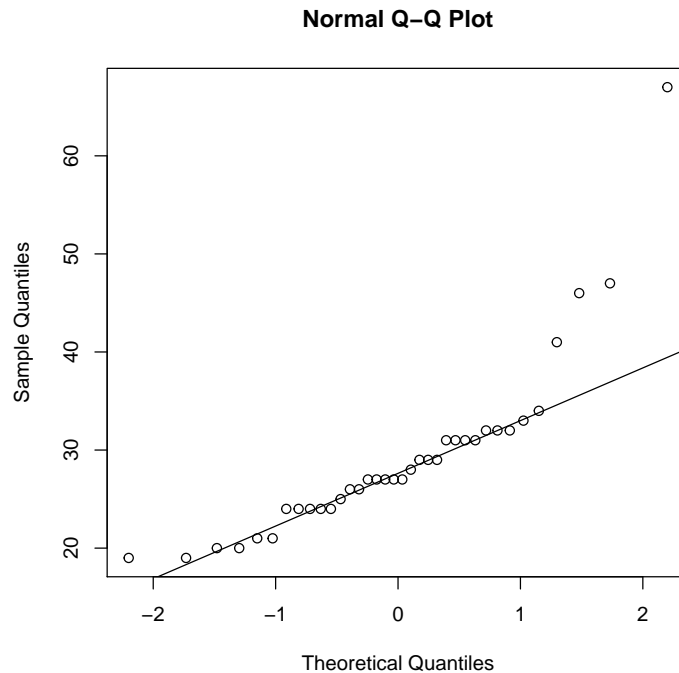
Figure 3: QQNorm compared with normal line of the data at the Oak Street residence

```
> qqnorm(PersonalResidence)
> qqline(PersonalResidence)
```

Normal Q–Q Plot

distributions that we perform tests on a bootstrapped distribution.

We see a difference in the sample means in figures 5 and 6 . We then performed a bootstrap T test to see if the results are statistically significant.

```
[1] 0.013
```

Since our p-value is less than .05, the difference in means is statistically significant at 95% confidence. Given statistical significant of the difference in means we then found a confidence interval for μ. For our confidence interval we bootstrap our quantiles. Metrics in our estimator are in figure **??**.

We find our 95% confidence interval to be in support of our hypothesis that the true difference in means is nonzero. Because $H_0 : \mu_{kzoo} = \mu_{oak}$ is not contained in the confidence interval we reject the null hypothesis and accept $H_A : \mu_{kzoo} \neq \mu_{oak}$. We use the bootstrapped sample means as a hypothesis for the true mean.

5

Figure 4: QQNorm compared with normal line of the data at the Oak Street residence

```
> qqnorm(KalamazooCollege)
> qqline(KalamazooCollege)
```
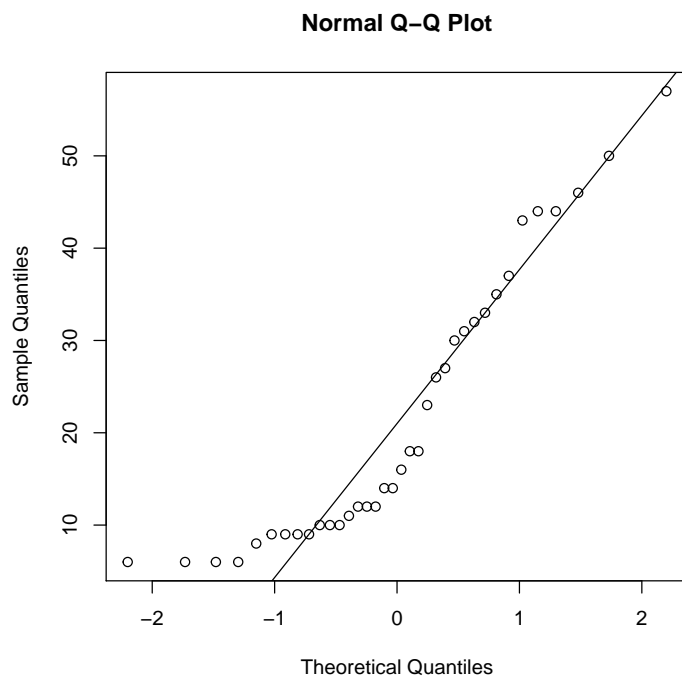
Normal Q–Q Plot

Figure 5: Kalamazoo College

```
> summary(KalamazooCollege)
```

```
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  6.00    9.75   15.00   21.75   32.25   57.00
```

Figure 6: Personal Residence Data

```
> summary(PersonalResidence)
```

```
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 19.00   24.00   27.00   29.31   31.25   67.00
```

Figure 7: Bootstrap Evaluation

```
> obs

[1] -7.555556

> bias

[1] 0.004922222

> variance

[1] 8.542067

> #mean squared error
> variance - bias^2

[1] 8.542043

> quantile(times.diff.mean,c(0.025,0.975))

      2.5%       97.5%
-13.250694  -1.833333

??
```
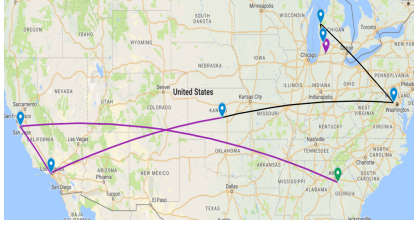
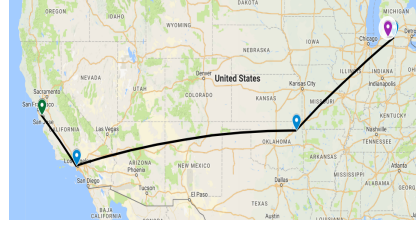Figure 8: Packet network trace, Kalamazoo College to Google

Figure 9: Network packet trace graph, 416 Oak St to Google

4   Conclusion

We conclude that Kalamazoo college's internet has lower latency as measured by ping than the Oak Street residence and that theoretically that there is a chance that the two networks are on different underlying connections. To confirm this, we traced a packet as it went from Kalamazoo College and the Oak Street residence to google. In figures 8 and 9 we see that a packet going to google.com consistently took a separate path. Furthermore, running the packet trace utility showed that Kalamazoo College and the Oak Street residence were running on different ISPs.