Steven Monje

Dr. Huson

IB Mathematics SL

23 April 2018

Subway Commute Study IA

Introduction:

The aim of my investigation is to calculate and evaluate the most efficient transportation route to school on average, in a fixed number of days using sorted data and graphs. The two transportation routes used are very similar as both include the use of trains. The difference is number train taken and getting off the train to either take the bus or walk. The time starts as soon the train doors close. The time is measured in seconds because it is comfortably accurate, given the situation. The main graph I used was whisker plot because I am analyzing data concerning efficiency on average. Whisker plots can help spot outliers and more common numbers within a fixed set of data.

Background:

The reason why I chose to partake in this investigation is because transportation is very interesting, especially in New York City. There are many ways to get to one specific location, but the most efficient routes are the most popular. Perhaps you are an hour early and decide to take the long route. You can do that but, in the morning, everyone is trying to get to work. I have data collected from multiple days of routes taken to school. It is important that I know the most efficient route to school on average, so I do not arrive to school late.

Information Collected:

To decide what measurements were necessary, I studied the similarities and differences between the two alternative routes. On either route, I first walk to 86th Street and catch the first train. At the 149th Street stop, the two routes differ. On the 4 train, I continue to 161st Street and then walk to school. For the 5 train, I get off at 149th Street and catch the bus to Grand Concourse and 163rd Street, from which point I walk the short remaining distance to school (part of the same walk as when taking the 4 train). Since I want to compare the differences in the two routes, I identified the following four points in my travel.

<u>Time Collection Point</u>

149th Street

This is the stop where I get off the 5 train. On the 4 train, I pass this stop.

161st Street (4 Train Only)

This is the stop where I get off the 4 train and begin to walk.

Catch the Bus (5 Train Only)

This is where I catch the bus near the 149th Street stop.

163rd Street & Grand Concourse

This is where both routes intersect to reach the destination.

Method of data collection:

Each day I took the first train that came and used my cell phone to determine the exact time of each of the Time Collection Points, to an accuracy of one second. I recorded the times in a log, which is reproduced in Appendix A: Travel Log Raw Data. I collected times for 19 trips on each route.

From the complete set of raw data in Appendix A, the Total Travel Time durations for the 38 trips are summarized in table 1, below. The times have been sorted from least to greatest. This is the data that is analyzed in the next section to answer the question of which the superior route is.

As was mentioned in the introduction, unusual data points require special consideration. For example, an extremely long delay caused by train mechanical problems might deceptively affect the mean calculation. Our textbook defines an outlier as data values +/- 1.5 times the IQR beyond the lower/upper quartile (Blythe 67). As is shown in Table 1 and the calculations below, one trip qualifies as an outlier, and another comes close. Both were slow trips on the 4 train. The minimum time for the 4 train was 14 minutes and 3 seconds. The number close to the minimum on the 4 train was 14 minutes and 8 seconds. The small difference between both suggests that the 4 train, at its fastest time on average, is 14 minutes. The minimum time for the 5 train was 7 minutes and 27 seconds. The number close to the minimum on the 5 train was 7 minutes and 41 seconds. The small difference between both suggests that the 5 train, at its fastest time on average, is 7 minutes and 30 seconds.

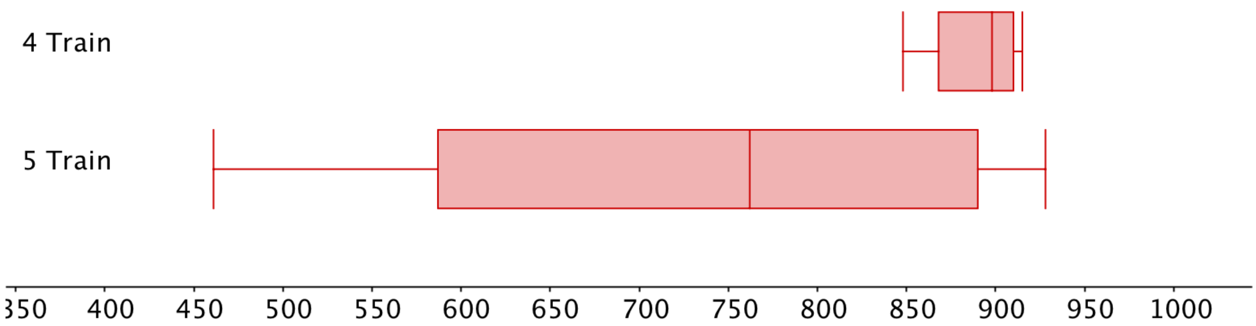
The mean of the travel times represents one criterion to judge the two commute alternatives. A lower time in seconds means a faster commute, which is good. Taking the 4 train had a mean time of 888 seconds while the 5 train alternative had a mean time three minutes faster of 708 seconds. Similarly, the median times for the 4 train and 5 train were 898 and 720 respectively, again favoring the 5 train by roughly 3 minutes. Clearly, in terms of central tendency—by either measure—the 5 train is better.

Total Travel Times (Seconds)
Sorted in increasing order

	4 Train plus Walk		5 Train plus Bus	
Min	**1**	**843**	**1**	**447**
	2	848	2	461
	3	851	3	494
	4	854	4	557
Q1	**5**	**866**	**5**	**568**
	6	868	6	587
	7	887	7	630
	8	891	8	672
	9	898	9	689
M	**10**	**898**	**10**	**720**
	11	898	11	762
	12	905	12	771
	13	907	13	771
	14	907	14	795
Q3	**15**	**908**	**15**	**865**
	16	910	16	890
	17	910	17	916
	18	912	18	920
Max	**19**	**915**	**19**	**928**

Min	843	447
Q1	866	568
Med	898	720
Q3	908	865
Max	915	928

Mean	888	708
Sigma	24.1	154



The box and whisker plots were created using the Geogebra mathematics software package

---

The dispersion of times tells a different story, with the 4 train being much more reliable. The calculations and results are discussed below.

The formulas for two measures of dispersion, range and inter-quartile range ("IQR"), are

$$range = x_{max} - x_{min}$$

$$IQR = Q_3 - Q_1$$

where x are the travel times in seconds and Q_1 and Q_3 are the first and third quartile values, respectively. The calculation for standard deviation is

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}}$$

4 Train

Range: 1 minute and 12 seconds

Interquartile Range: 42 seconds

Standard Deviation: 24.1 seconds

5 Train

Range: 8 minutes and 1 second

Interquartile Range: 4 minutes and 57 seconds

Standard Deviation: 2 minutes and 34 seconds

This set of data supports the 4 train because it is more consistent than the 5 train. Even if the 5 train is faster than the 4 train, consistency is important as well.

The formula to calculate the mean travel time, or average, is

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

where *x* represents the travel times in seconds and *n* is the number of data.

Conclusion:

In conclusion, I have succeeded in my investigation to figure out which route is more efficient on average. The 5 train route was much more efficient but the 4 train is more consistent. Through data analysis and whisker plots, the 5 train is the most efficient route on average. There are some limitations, however, that cannot be ignored.

The first limitation is sometimes when I start the timer as the door closes, the train may take longer than usual to depart from the current location. This may add some unnecessary seconds, but it is something that is out of my control. Another limitation is how long the bus may take between the smaller stops. This adds more time than needed but it is something out of my control as well. The third limitation is how much of an effect does either route's time change under certain conditions such as weather, traffic, and maintenance. Which route would be faster if it was raining? What route would be faster under heavy traffic? What about when maintenance is involved?

My investigation does not answer these questions, but it is the same reason why I recorded data multiple times. This is the best way to eliminate any of these minor situations from affecting my investigation. Keeping all of this in mind, the 5 train is more efficient than the 4 train but the 4 train is more consistent. I found this very interesting because this means that the 5 train has a higher risk of being affected by the conditions I mentioned earlier. If you want to play it safe and take a secure route, the 4 train is your best bet. If you want the fastest route on average, the 5 train route is the way to go.

Works Cited

International Geogebra Institute. *Geogebra*. Version 5. www.geogebra.org. Accessed 20 April 2018.