AP Statistics Quarter 2 Project Data Report – Semester 1
Typing Speed vs Hand Size

By: James Ding, Ellie Ku, Hannah Li

December 5, 2022

1 INTRODUCTION

More and more high school students choose to use computers to complete homework. As many students uphold the belief that "due tomorrow, do tomorrow," some will not start writing essays until 2 hours before the deadline. But they are, miraculously, still able to finish writing it by the time it is due. This phenomenon makes us think about how fast people can type, and whether it is related to hand size. We conducted an experimental study within our school (California Crosspoint High School) to study the relationships between these two variables. Our hypothesis is "Students at CCA with larger hands tend to type faster." This hypothesis stems from the idea that with bigger hands, one would be able to reach keys more easily on a keyboard so that they can type faster. It also could be argued that bigger hands could reduce accuracy, as people could easily "fat finger" keys. We believe there should be a linear relationship between these two variables: hand size and typing characteristics. In our experimental study, we randomly selected 50 high school students at California Crosspoint High School and gave typing tests on their own laptop and on a school laptop.

2 DEFINITION OF TERMS

WPM (Words per Minute) – The total amount of characters in the correctly typed words including spaces, divided by 5 and normalized to 60 seconds
CPM (Characters per Minute) – The total amount of characters in correctly typed words, normalized to 60 seconds. When divided by 5, it becomes the WPM.
Raw WPM – Calculated just like WPM, but also includes incorrect words
Accuracy – The percentage of correctly pressed keys
Consistency – Based on the variance of your raw wpm. Closer to 100% is better. Calculated using the coefficient of variation of raw wpm and mapped onto a scale of 0 to 100, according to Monkeytype
Hand Length – Length of hand from top to bottom, typically from the base to the tips of the subject's middle finger
Hand Span – The distance between the tips of the subject's thumb to opposing pinky
Thumb Length – The distance from the base of the subject's thumb to the tip of their thumb
See Figure 1 for an example of Hand Length, Hand Span, and Thumb Length

3 DATA COLLECTION

3.1 SAMPLE SPACE

Our sample for this project was high school students at California Crosspoint Academy. We stratified our sample by grade and gender, first counting the number of male and female students in each grade. These numbers were divided by the total number of high school students to get the percent composition of students (see Figure 1). We split up our sample size into certain numbers of male and female students of each grade to represent the population of our school. In total, we planned to survey six 9th grade girls (11.98%), seven 9th grade boys (14.75%), three 10th grade girls (5.53%), five 10th grade boys (11.06%), six 11th grade girls (11.52%), four 11th grade boys (7.83%), eight 12th grade girls (15.67%), and eleven 12th grade boys (21.66%). This created our sample size of 50 students.

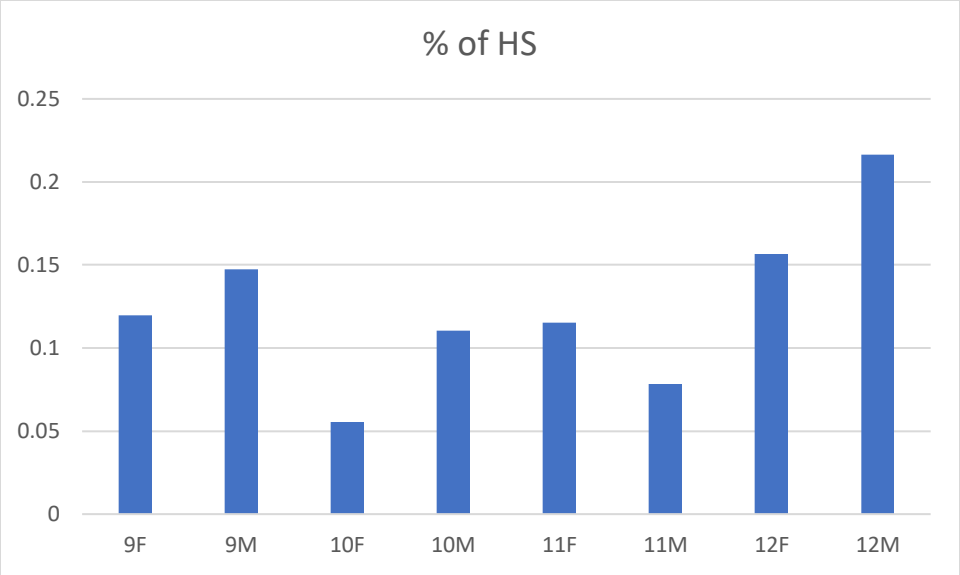


Figure 1

From there, we used a random number generator without replacement to select samples alphabetically. For example, the 9th grade girls made up rows 2-27 on the spreadsheet of all high school students, so we generated six random numbers in that range and chose students based on the corresponding row number.

Additionally, we collected data from a convenience sample, mainly consisting of juniors and seniors. These students saw us collecting data from our randomized sample and asked if they could participate as well. We categorized their data separately, analyzing it both separate from the randomized sample and together.

To minimize response bias, we also randomized which person in our group would approach each student to collect data. We believed that if students were approached by someone, they were more familiar with, they might be more likely to participate in our study, or otherwise unconsciously alter their behavior (peer pressure, anxiety, etc.) in a way that would not reflect their true typing speed. However, some data points were not gathered by the designated surveyor, as it became increasingly inconvenient for us to track down certain people. To collect data in a quick and efficient manner, a small amount of the surveyors was reassigned to other members of the group.

3.2 MONKEYTYPE

To accurately collect data on typing speed, we used Monkeytype, an open source, minimalistic, and unobtrusive typing test. For our experiment, our data used the default test from Monkeytype – the 30 second timed test, otherwise known as the "Time 30". This type of test gives users an unlimited number of words to type within the 30 second time constraint. Upon the first keystroke, the timer starts, and further inputs are not processed once the timer is over. Monkeytype calculates important typing statistics and characteristics such as WPM, raw WPM, accuracy, number of characters correct, incorrect, extra, and missed, and consistency (see definitions above). We recorded the most accurate versions of this data that Monkeytype provides, going to one or two decimal places wherever possible. Visit Monkeytype at https://monkeytype.com

3.3 Not All Keyboards are Born Equal

Because we believe keyboard familiarity may be a factor in determining typing speed and accuracy, we decided to gather two different sets of data – one from a personal laptop, one from a school laptop as control. First, we had the students type on their personal laptops. This would be a laptop that they are used to, one they have surely typed on for some time. However, factors like the space between keys or how tactile the keyboard is could affect any person's typing speed. Therefore, after instructing students to type on their own laptops, we had them type on a laptop we provided. These Lenovo 300e 2nd Gen 81M9 R91339FX laptops, borrowed from the school, have smaller than standard keyboards. By not allowing students to practice with our control laptop before taking the test, we made sure everyone had the same treatment with regards to data collection.

3.4 Measuring Hand Size

When measuring hand size, we used centimeters. Most people's right and left hands are approximately equal in size, but we measured both to account for any discrepancies. After they finished typing on both laptops, we instructed people to hold out their hands and spread them apart. We took three different measurements of each hand: length, span, and thumb length. Hand length was measured from the tip of the middle finger to the first crease when the wrist bends. Hand span was measured from the tip of the thumb to the tip of the pinky when the hand was spread out. Thumb length was measured from the tip of the thumb to the crease where the thumb bends towards the palm (See Fig. 2). We chose these measurements because they seemed most likely to be attributed to typing speed. Typing relies on extending fingers vertically and horizontally on the keyboard, so we measured hand length and span respectively. Thumb length was measured to also study mobile typing (texting) speed.

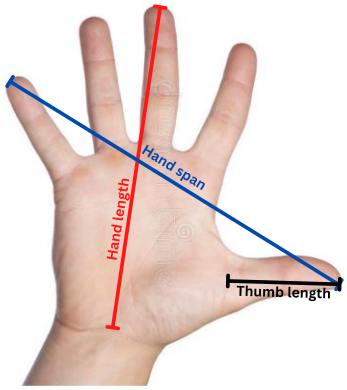


Figure 2

3.5 Unexpected Error

There were several inconsistencies and errors in our data collection that may have influenced our data analysis. Firstly, several people in our randomized sample were not able to participate in our study for a variety of reasons, such as refusal or absence from school. These data points had to be replaced, or otherwise were not collected due to time constraints. Our final data sample of 50 students comprised of both randomly selected students and convenience sampling. Other inconsistencies included the order in which we had students type on the laptops. Some students typed on their personal laptops first, others started with the school laptop. Additionally, we used normal rulers to measure hand size, meaning we could not get very accurate measurements.

3.6 Summary

We randomly selected 50 people; we first asked them whether they are willing to participate in our experiment. If the person refused, we deleted them from our list and randomly selected another person from the corresponding grade and gender to replace them. We still thanked them for their time.
If a person agreed to participate in our study, we asked them to type on Monkeytype.com on their personal laptop and the laptop that we provided. We made sure that no one was around to disrupt them. After, we measured that person's hand length, hand span and thumb length, as shown above. We

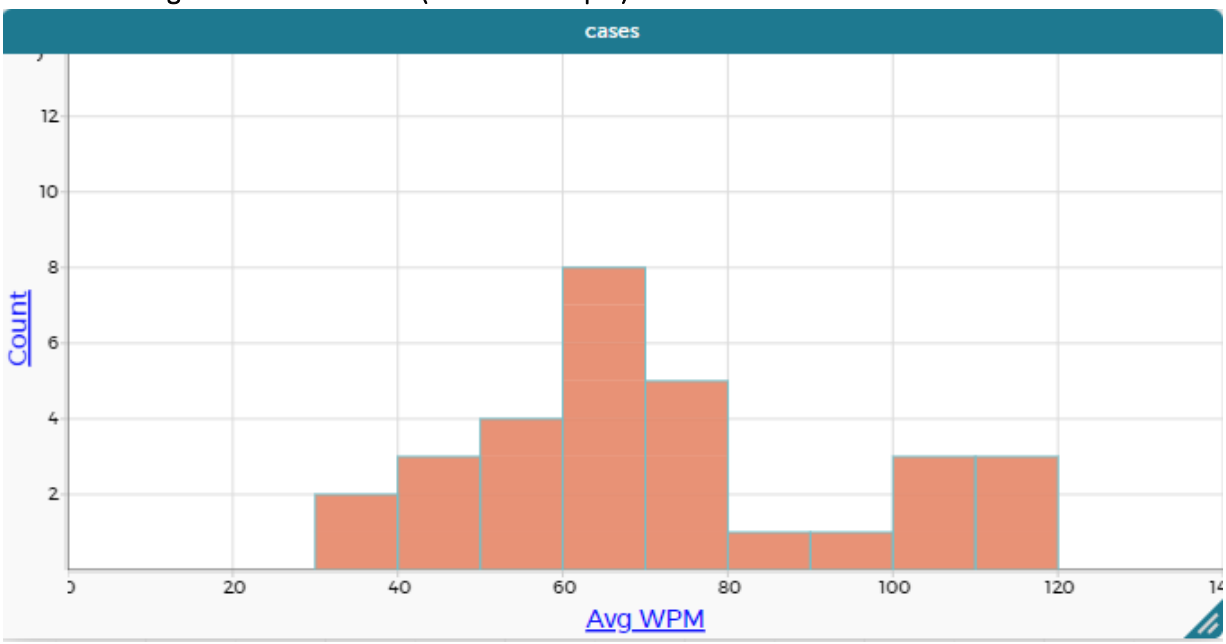
thanked them for participating in our experiments. We believe this is an experimental study because we have control over people.

4 Data Analysis

To simplify data analysis and determine correlation, we used data from each test subject's personal laptop and control laptop to create averages for each respective data field. Additionally, we averaged out the thumb length, hand size, and hand span.

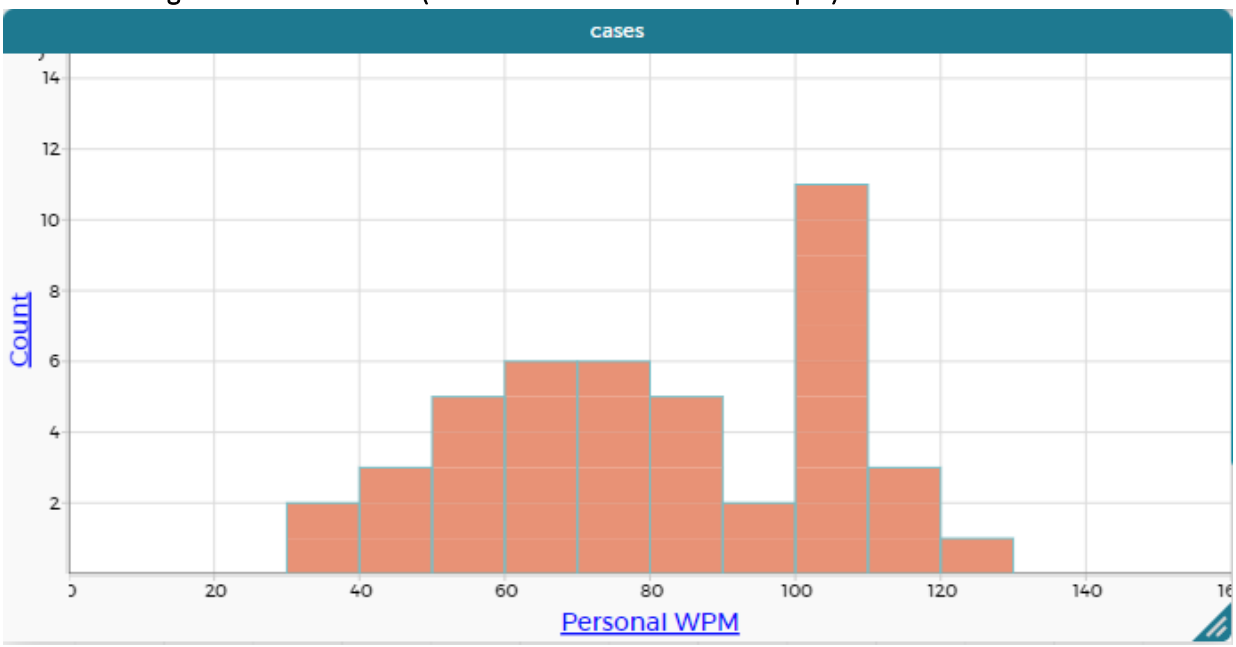
4.1 Data Distribution and Statistics

4.1.1 Average WPM Distribution (Random Sample)



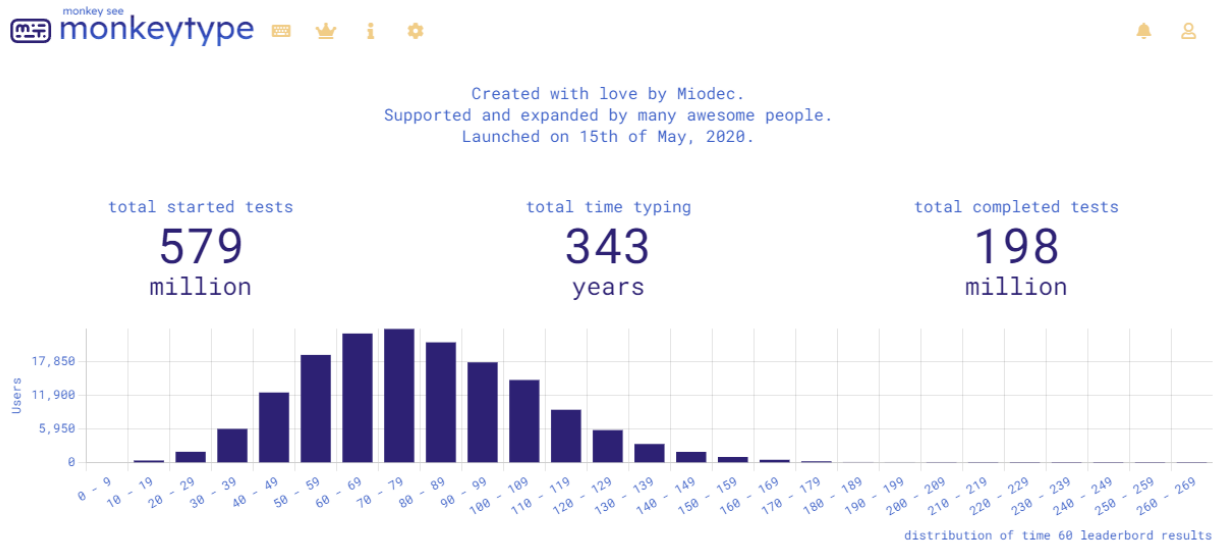
As shown in the bar chart by CODAP, the average WPM is mound shaped, with most of the random sample achieving speeds between 60-70 WPM. There is one outlier who achieved 118.56 WPM. The data from the random sample had a median of 67.3 WPM, with an mean of 71.2 WPM, with a moderate standard deviation of 23.9 WPM.

4.1.2 Average WPM Distribution (Random and Convenience Sample)



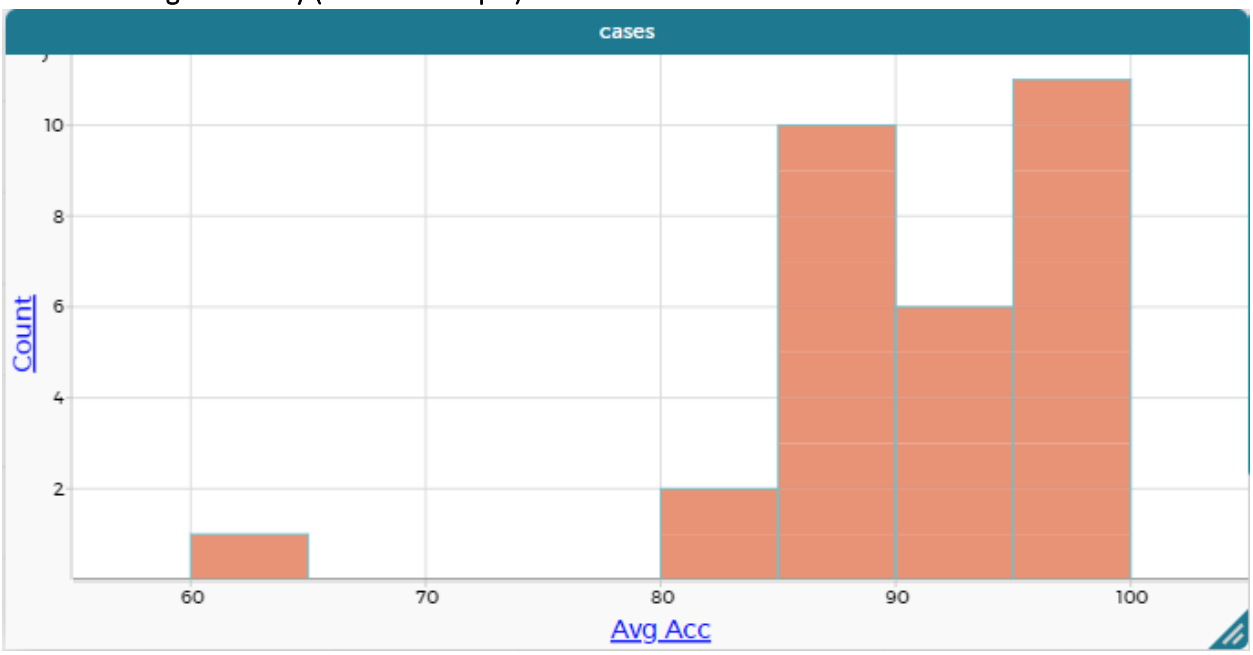
As shown in the bar chart above, most of our random sample and convenience sample achieving speeds between 100-110WPM. There are no outliers within this set of data. Additionally, the data has a median of 81.4 WPM, with a mean of 81.8 WPM, and a standard deviation of 23.9 WPM. The increase in WPM is most likely because the convenience sample primary comprises of acquaintances of James, or people who typically spend a greater amount of time with technology.

4.1.3 Monkeytype Time 60 Distribution



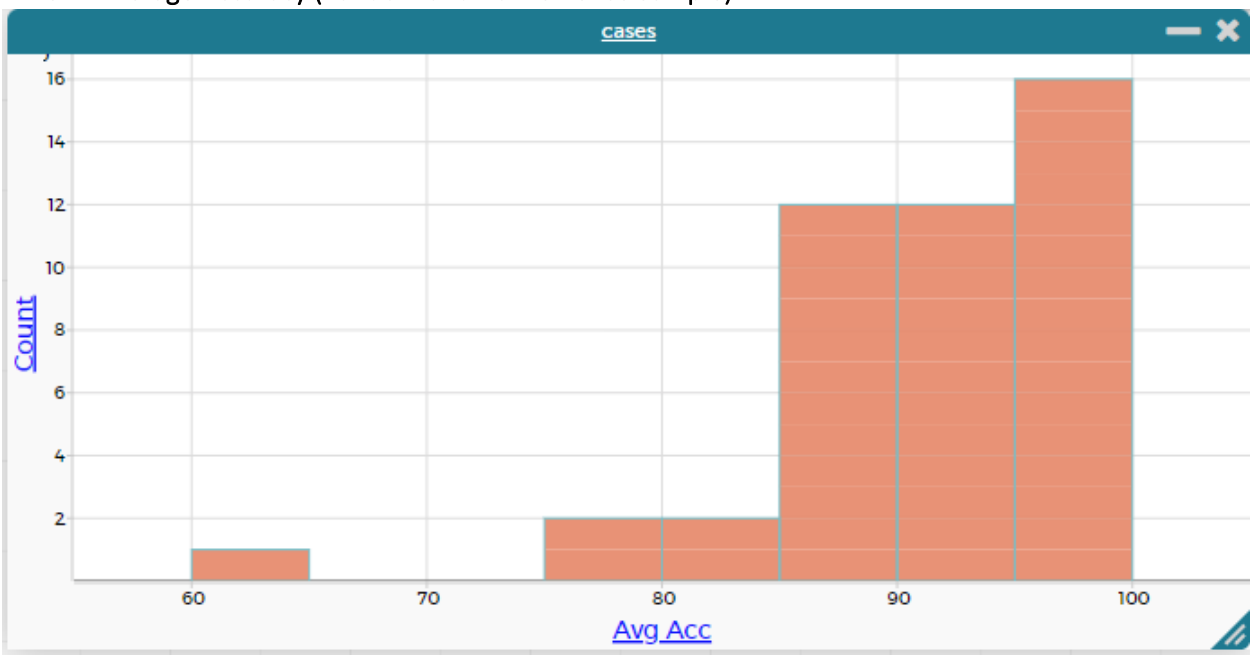
While the test our group decided to use the time 30 test, Monkeytype does publish it's time 60 WPM distributions. The data is shown to be skewed right, however when the WPM range is narrowed to 20-120 WPM, it is mound shaped. Most of Monkeytype's users achieved a WPM between the 70-79 range, which is close to our random sample. Additionally, it is shown that approximately at least 142.5K out of 161.4K (88.2%) Monkeytype users have average a WPM between from 30 to 130, with the data from https://monkeytype.com/about

4.1.4 Average Accuracy (Random Sample)



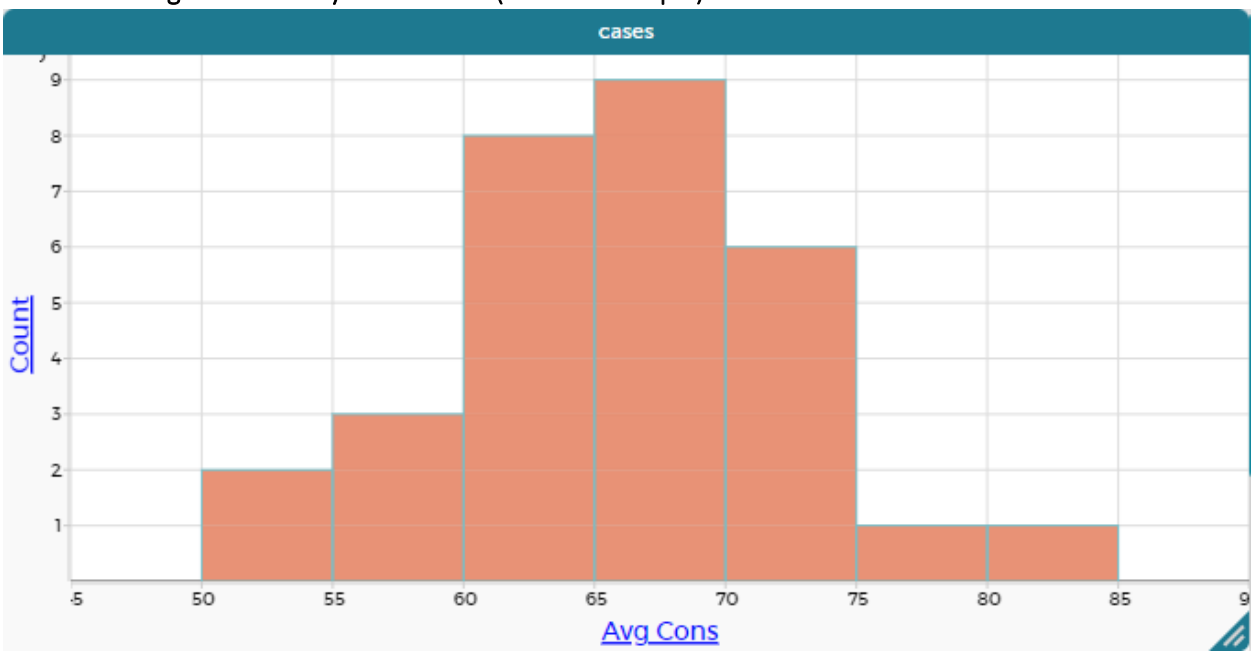
As shown in the bar chart above, the average accuracy is skewed left, with most of the random sample achieving accuracies between 95-100%. There was one outlier, who achieved a 62% accuracy. The data from the random sample had a median of 92.34%, a mean of 91.22%, and a standard deviation of 7.13%. If we remove the outlier the mean will increase, the standard deviation will decrease. Note that no one scored a perfect 100%, with the highest being a 98.18%

4.1.5 Average Accuracy (Random and Convenience Sample)



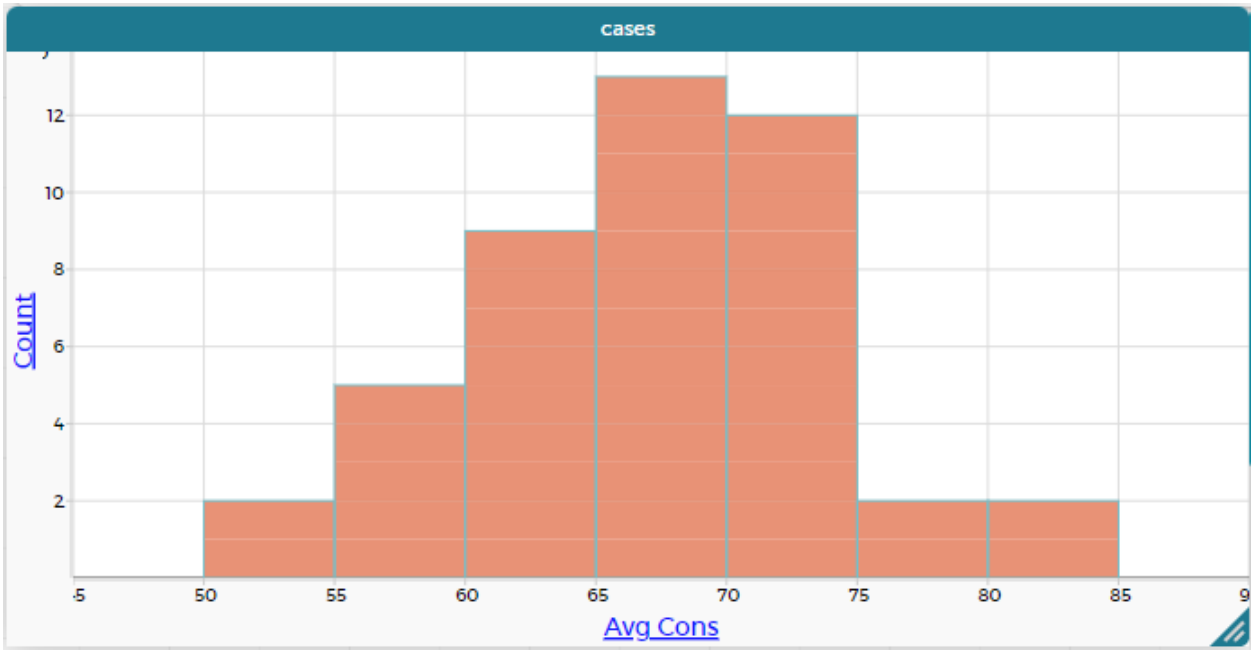
As shown in the bar chart, the average accuracy is skewed left, with most of the random sample achieving accuracies between 95-100%. There were two outliers at 62% and 78.305%. The data from the random and convenience sample had a median of 92.28%, a mean of 91.31%, and a standard deviation of 6.67%. If we remove the outlier the mean will increase, the standard deviation will decrease. Note that again, no one scored a perfect 100%. Additionally, with the inclusion of the convenience sample, there was a noticeable increase in accuracy in the 90%-95% range, but without much effect to the median and mean.

4.1.6 Average Consistency Distribution (Random Sample)



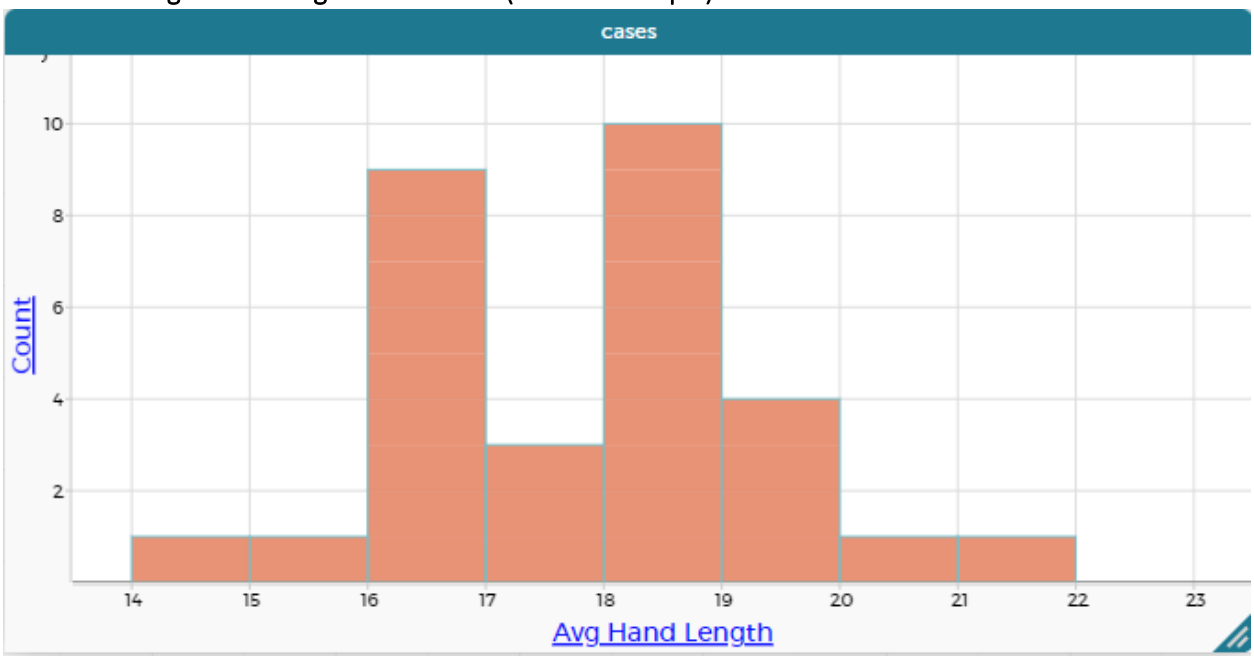
As shown in the bar chart, the average consistency is slightly skewed left, with most of the random sample achieving consistencies between 65%-70%. There are no outliers. The data from the random sample had a median of 65.81%, a mean of 65.94%, and a small standard deviation of 6.54%.

4.1.7 Average Consistency Distribution (Random Sample and Convenience Sample)



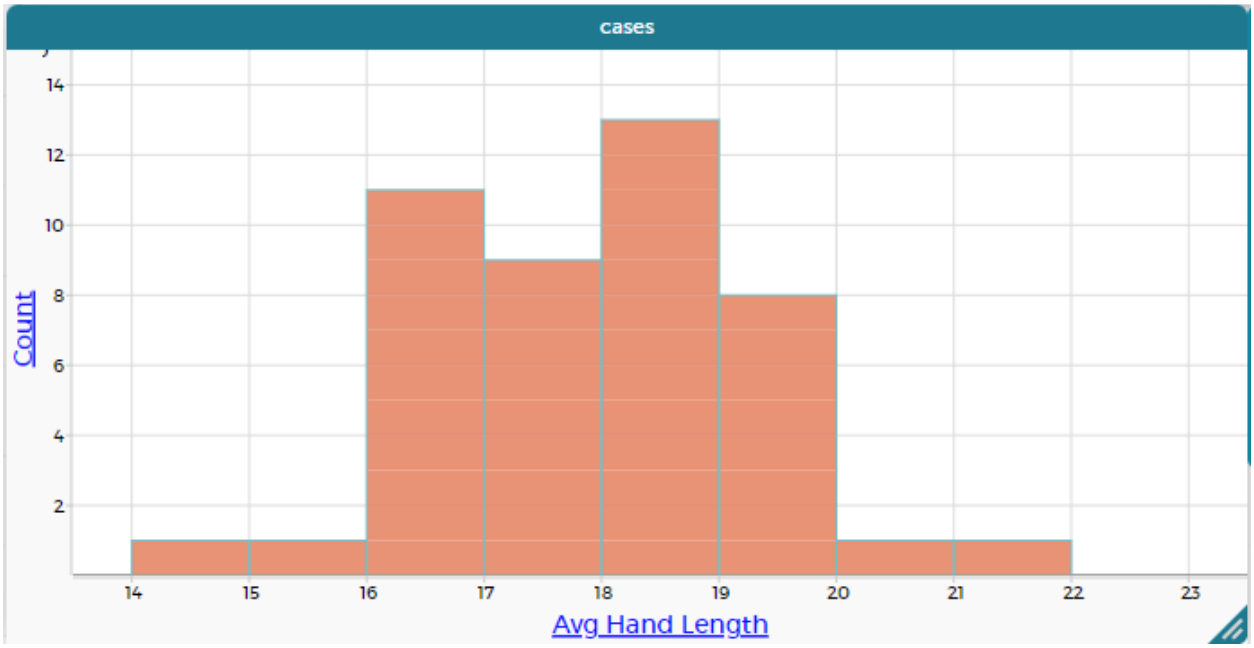
As shown in the bar chart, the average consistency is skewed left, with most of the random sample and convenience sample between 65%-70%. There are no outliers. The data from the random and convenience sample has a median of 66.95%, a mean of 67.01%, and a small standard deviation of 6.68%. With the addition of the convenience sample, there is a noticeable increase in consistency within the 70%-75% range, most likely because of the same reason mentioned in 4.1.2

4.1.8 Average Hand Length Distribution (Random Sample)



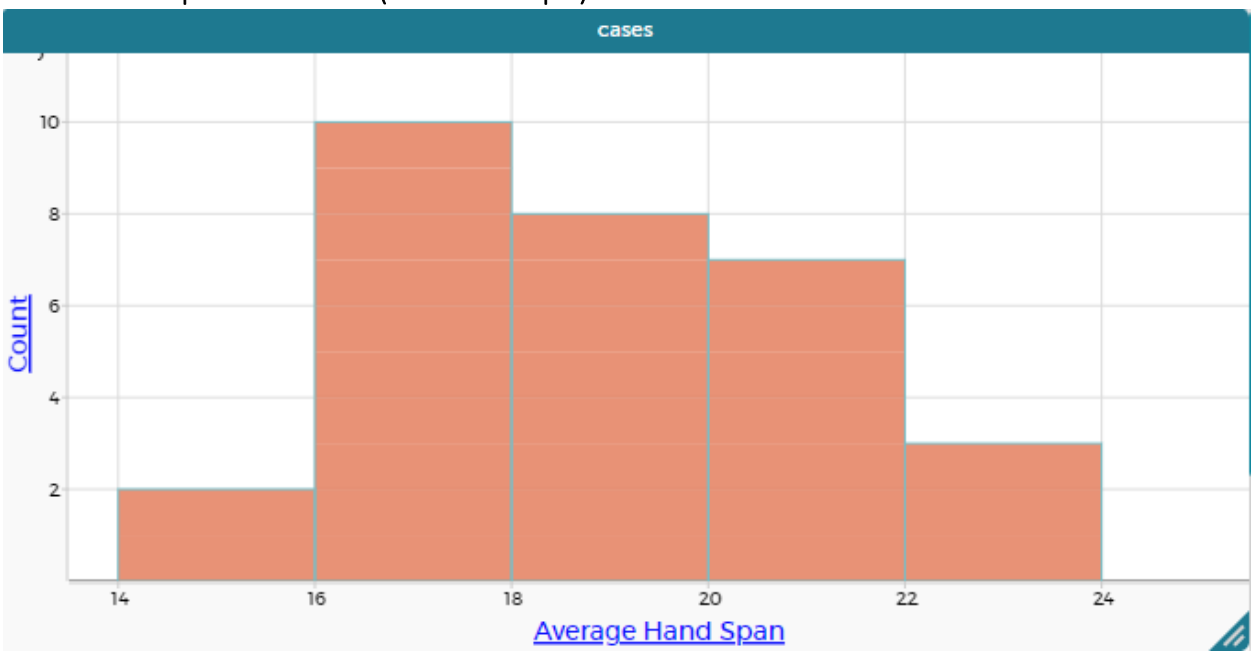
As shown in the bar chart above, the average hand length is bimodal, with most of the random sample having hand lengths between 16cm-17cm and 18cm-19cm. There is one outlier at 21.59cm. Additionally, the median is 18cm, the mean is 17.61cm, and a small standard deviation of 1.57cm. If we remove the outlier, the mean and standard deviation will decrease.

4.1.9 Average Hand Length Distribution (Random and Convenience Sample)



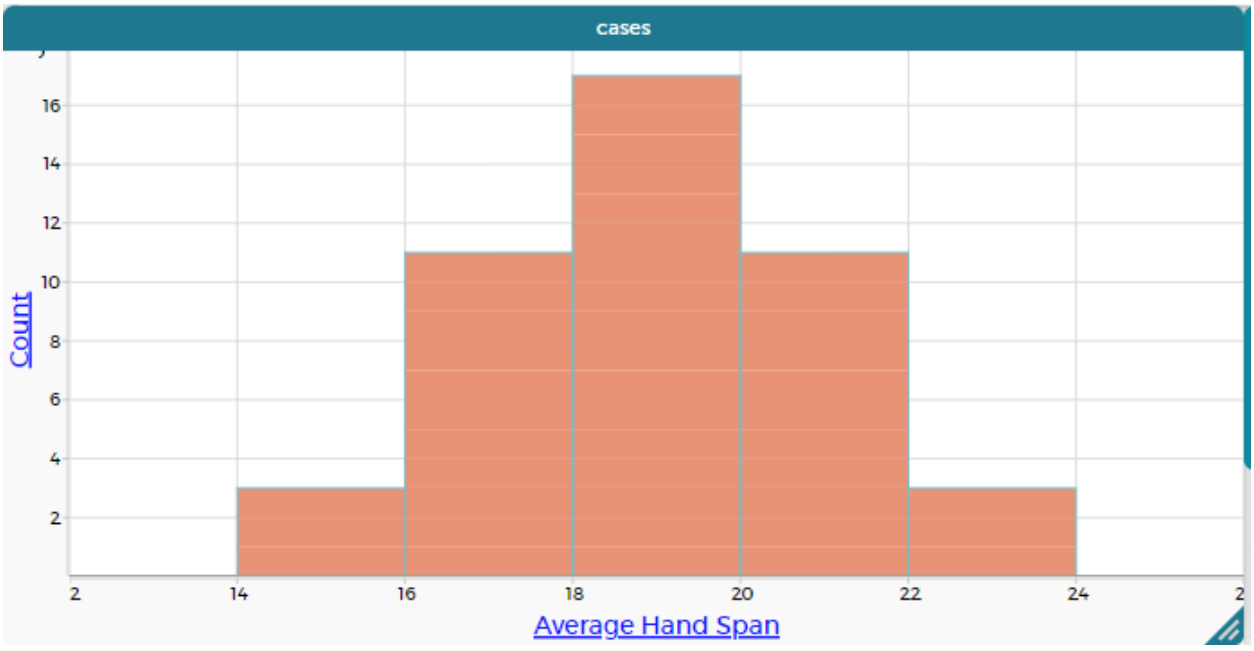
As shown in the bar chart, most of the random sample and convenience sample have hand lengths between 18cm-19cm. There are two outliers at 14cm and 21.59cm. Additionally, the median is 17.73cm, with a mean of 17.73, and the standard deviation is a small 1.38cm.

4.1.10 Hand Span Distribution (Random Sample)



As shown in the chart above, the average hand span is right skew, with most of the random sample having hand spans between 16cm-18cm. There are no outliers, the median is 18.25cm, the mean is 18.6cm, and the standard deviation is a moderate 2.24cm.

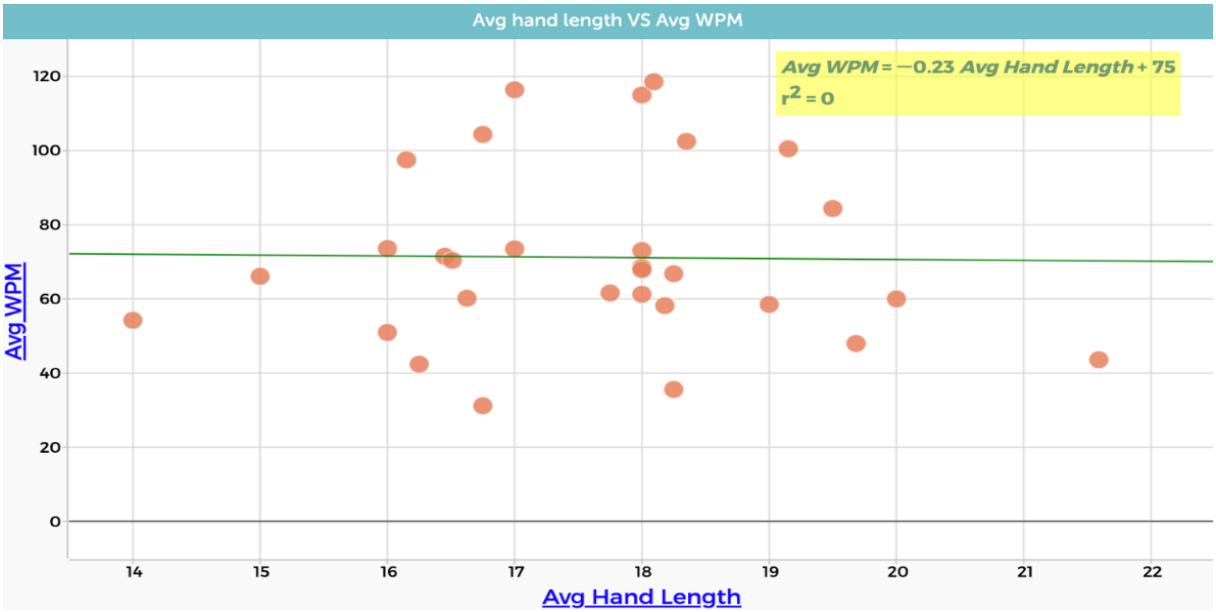
4.1.11 Hand Span Distribution (Random and Convenience Sample)



As shown in the chart above, the average hand span is unimodal, with most of the random and convenience sample having hand spans between 18cm-20cm. There are no outliers, the median is 19cm, the mean is 18.76cm, and the standard deviation is a moderate 2.01cm. Compared to the shape of the random sample, there was a noticeable increase in hand spans within the 18cm-22cm range.

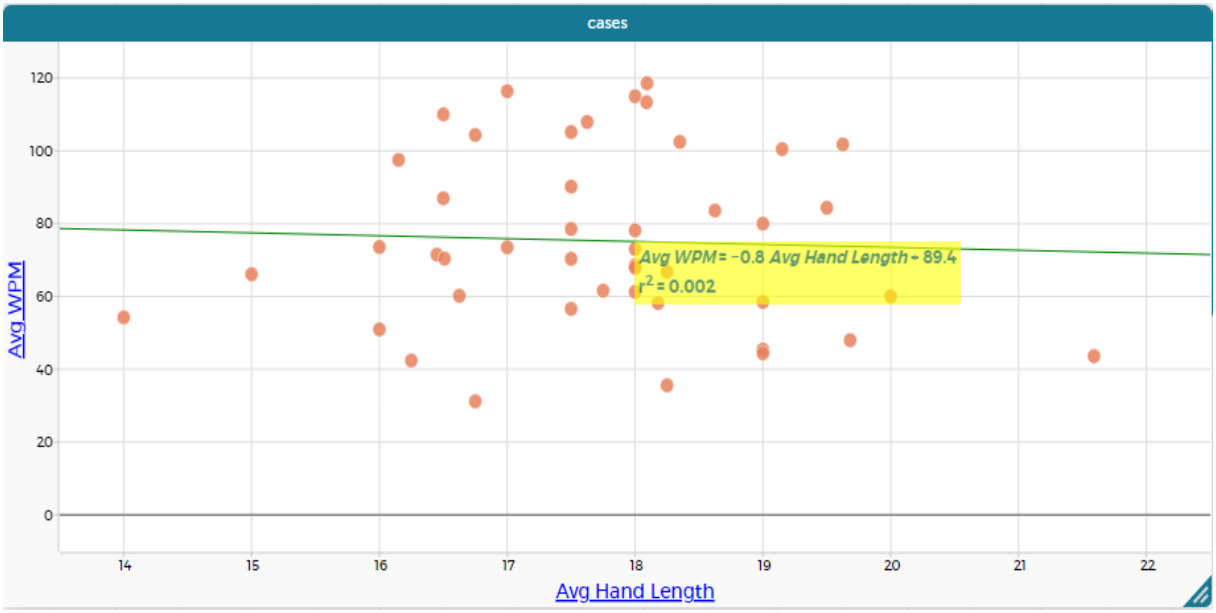
4.2 Hand Attributes vs. Typing Characteristics

4.2.1 Average Hand Length vs. Average WPM (Random Sample)



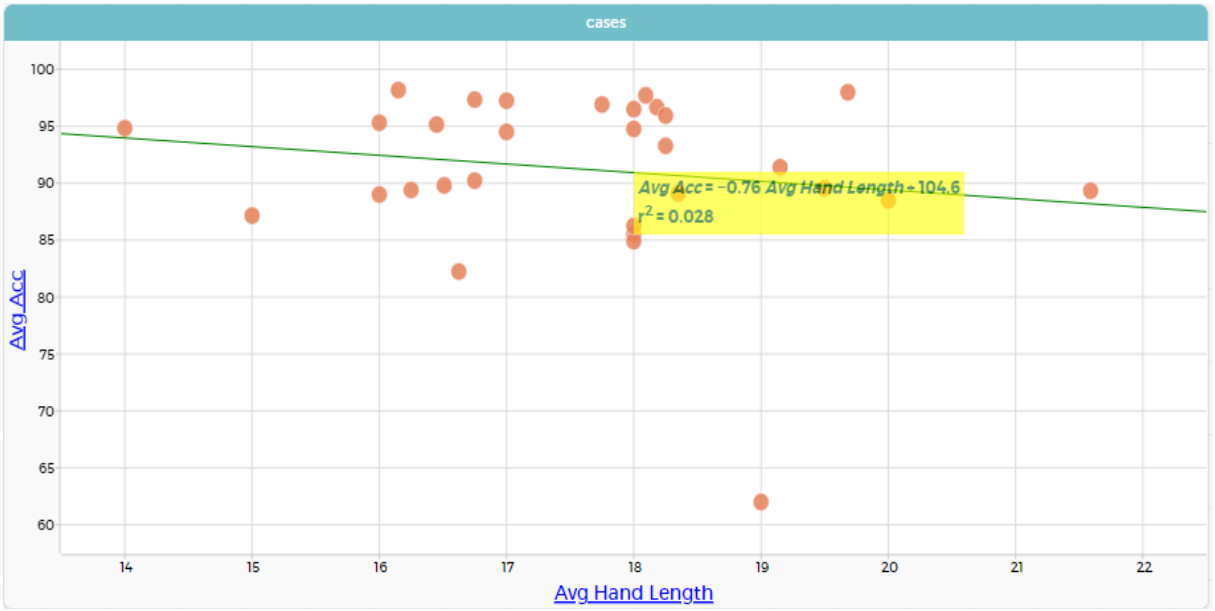
For the random sample, CODAP calculated r^2 to be equal to 0, meaning approximately 0% of variability in average WPM can be explained by the linear relationship between average hand length and average WPM. This means the correlation coefficient (r) of 0 indicates perfectly random relationship between the average hand length and average WPM.

4.2.2 Average Hand Length vs. Average WPM (Random and Convenience Sample)



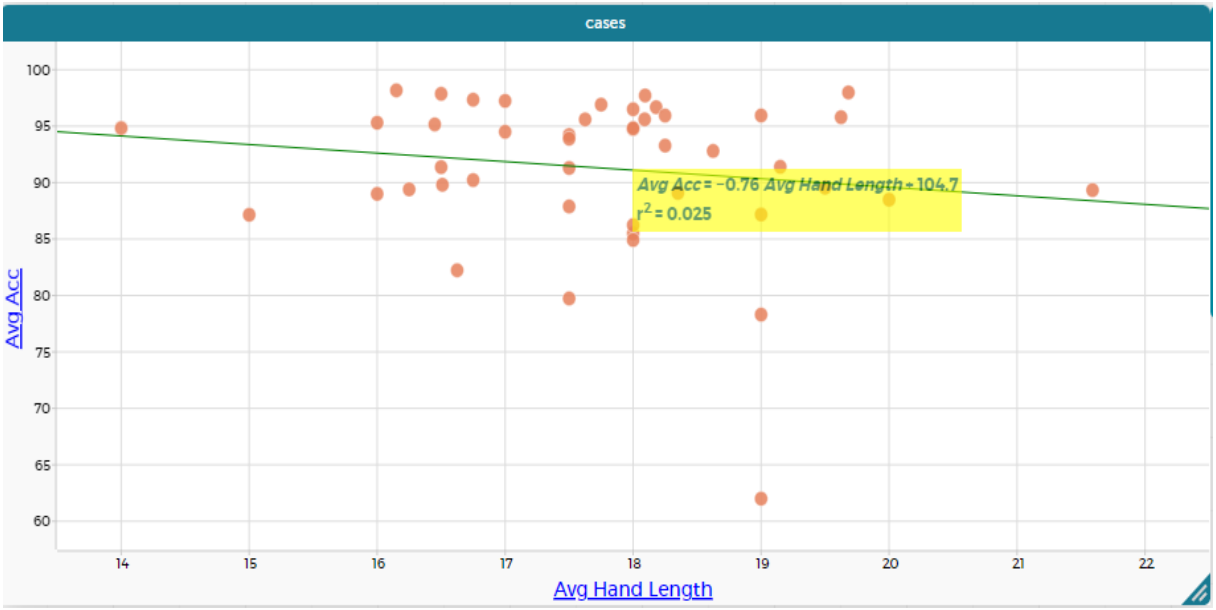
For the random sample and convenience sample, CODAP calculated r^2 to be equal to 0.002, meaning approximately 0.2% of variability in average WPM can be explained by the linear relationship between average hand length and average WPM. This means the correlation coefficient (r) of approximately 0.044 indicates a negligible negative linear relationship between the average hand length and average WPM.

4.2.3 Average Hand Length vs. Average Accuracy (Random Sample)



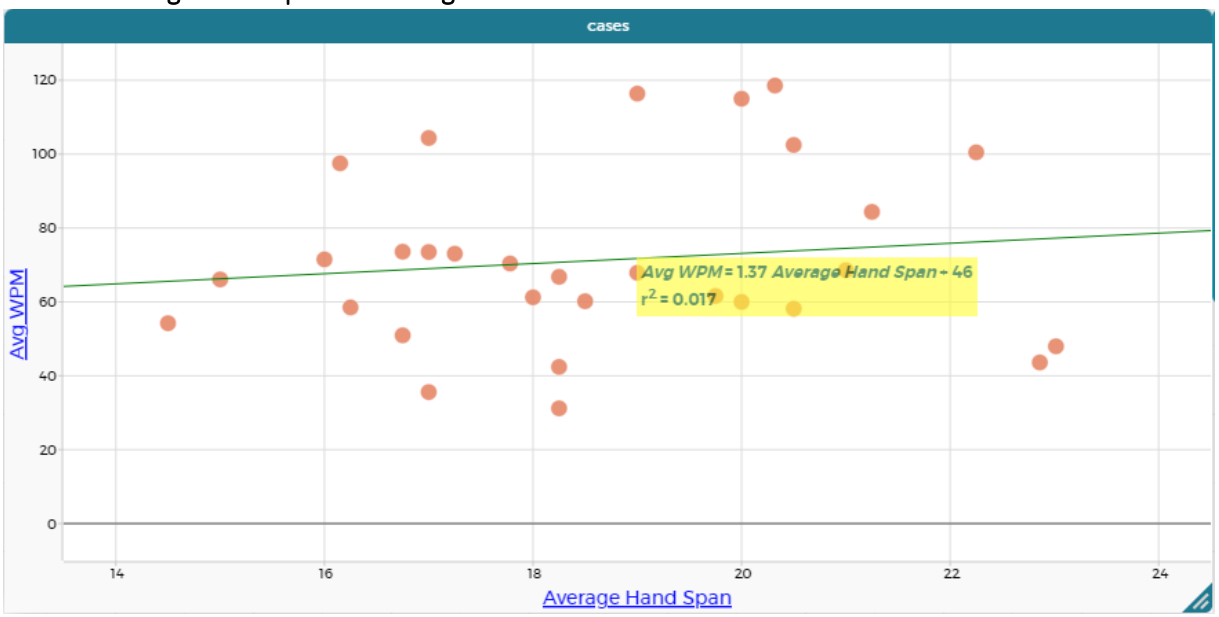
For the random sample, CODAP calculated r^2 to be equal to 0.028, meaning approximately 2.8% of variability in average WPM can be explained by the linear relationship between average accuracy and average hand length. This means the correlation coefficient (r) of approximately 0.167 indicates a negligible negative relationship between the WPM and average hand length. There is one outlier at (19,62) If we remove it, r^2 will decrease to 0.004. meaning approximately 0.4% of variability in average WPM can be explained by the linear relationship between average accuracy and average hand length. It won't change the model we chose because the out conclusion we got remain the same. Hand Length has no correlation with accuracy.

4.2.4 Average Hand Length vs. Average Accuracy (Random and Convenience Sample)



For the random sample and convenience sample, CODAP calculated r^2 to be equal to 0.025, meaning approximately 2.5% of variability in average accuracy can be explained by the linear relationship between average accuracy and average hand length. This means the correlation coefficient (r) of approximately 0.159 indicates a negligible negative relationship between the average accuracy and average hand length. There is one outlier at (19,62) If we remove it, r^2 will decrease to 0.007. meaning approximately 074% of variability in average WPM can be explained by the linear relationship between average accuracy and average hand length. It won't change the model we chose because the out conclusion we got remain the same. Hand Length has no correlation with accuracy.

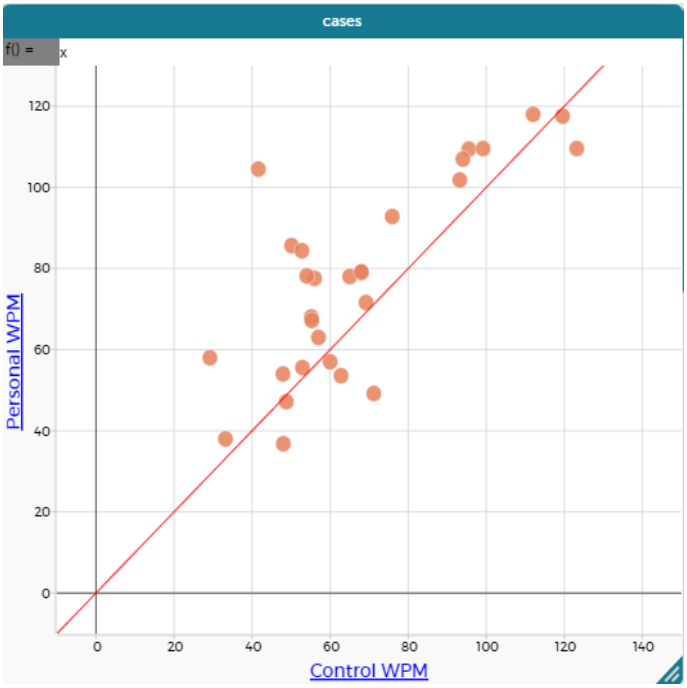
4.2.5 Average Hand Span vs. Average WPM



For the random sample, CODAP calculated r^2 to be equal to 0.017, meaning approximately 1.7% of variability in average WPM can be explained by the linear relationship between average WPM and average hand span. This means the correlation coefficient (r) of approximately 0.13 indicates a negligible positive relationship between the WPM and average hand length.

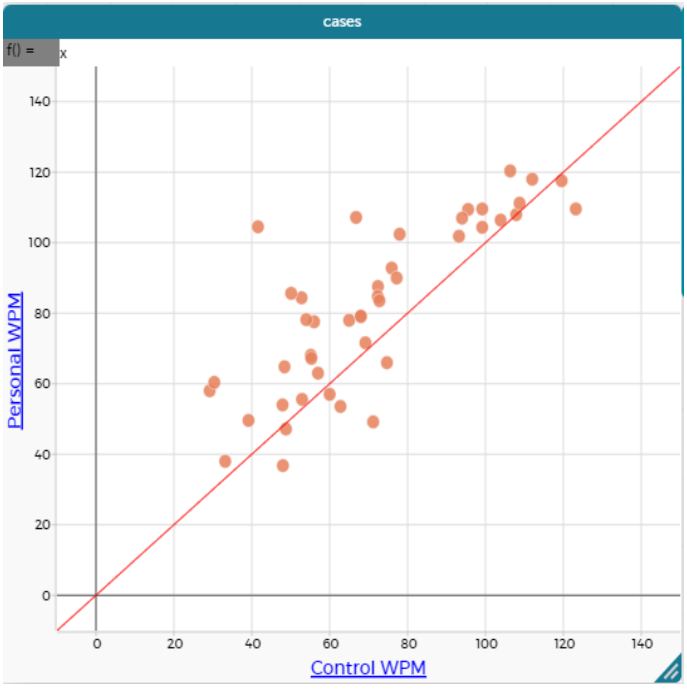
4.3 Control Typing Characteristics vs. Personal Typing Characteristics

4.3.1 Control WPM vs. Personal WPM (Random)



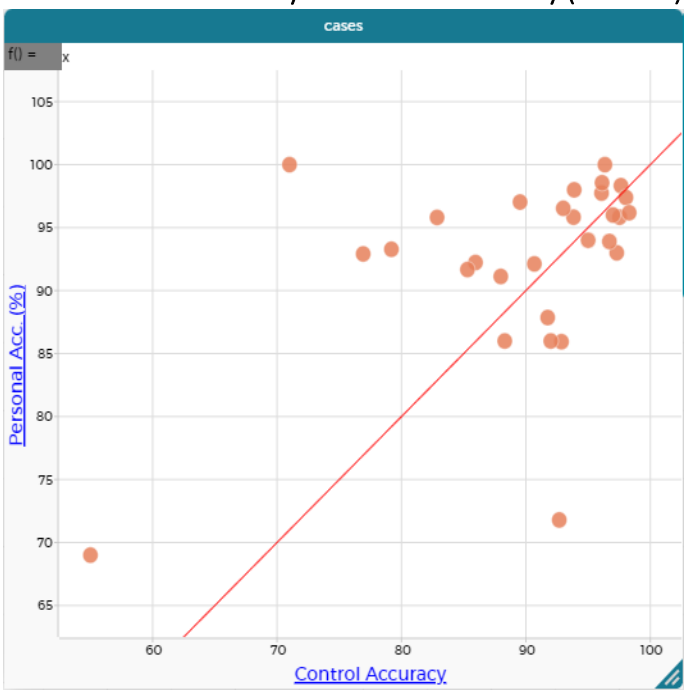
The equation of the line is y=x has been displayed. The graph very clearly shows that there are more points are above the line, indicating that our random sample tend to have higher WPM on their personal laptop.

4.3.2 Control WPM vs. Personal WPM (Random and Convenience Sample)



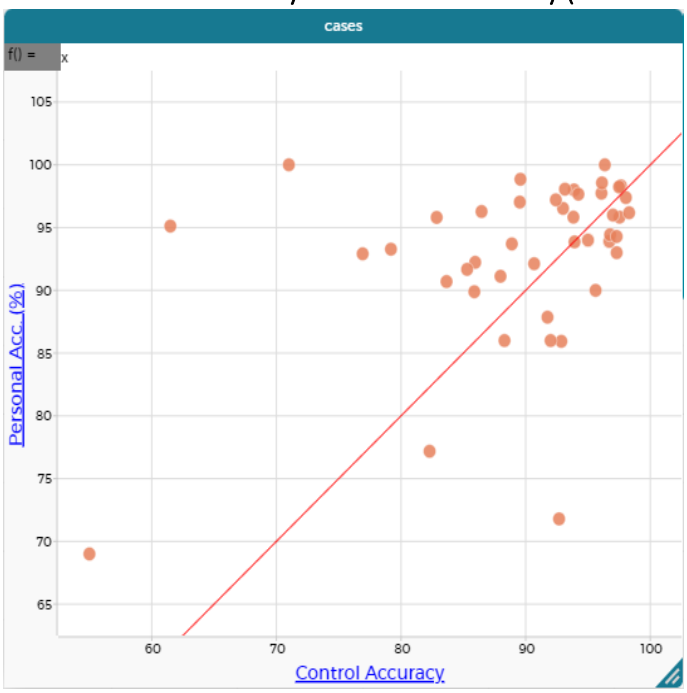
The equation of the line is y=x has been displayed. The graph very clearly shows that there are more points are above the line, indicating that our random sample and convenience sample tend to have higher WPM on their personal laptop.

4.3.3 Control Accuracy vs. Personal Accuracy (Random)



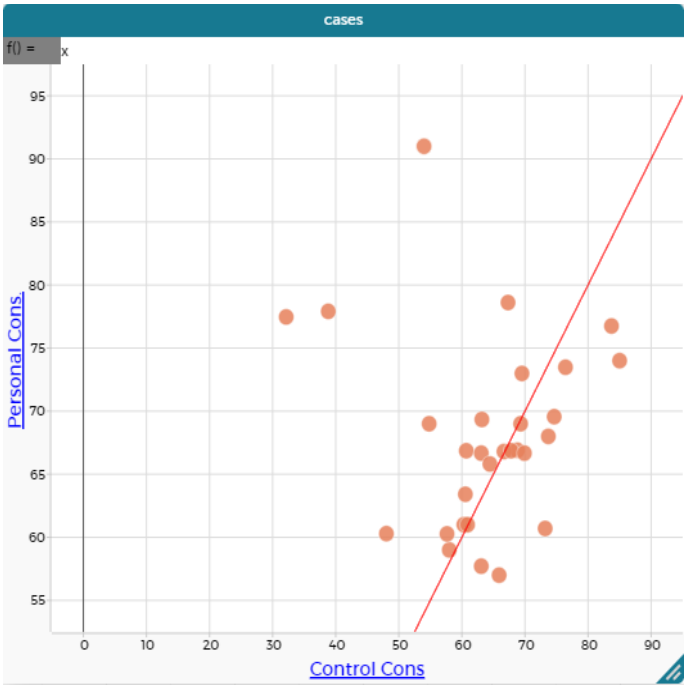
The equation of the line is y=x has been displayed. The graph shows that there are slightly more points are above the line, indicating that our random sample tend to be more accurate on their personal laptop. There are some cases where the accuracy on the personal laptop is significantly higher than the control laptop.

4.3.4 Control Accuracy vs. Personal Accuracy (Random and Convenience Sample)



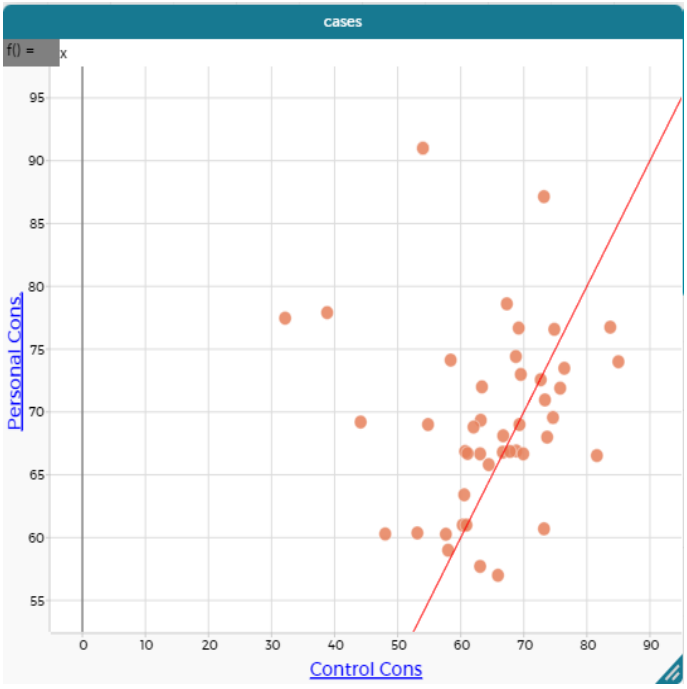
The equation of the line is y=x has been displayed. The graph shows that there are slightly more points are above the line, indicating that our random sample and convenience sample tend to be more accurate on their personal laptop. There are some cases where the accuracy on the personal laptop is significantly higher than the control laptop.

4.3.5 Control Consistency vs. Personal Consistency (Random)



The equation of the line y=x has been displayed. While there are slightly more points above the line, it is shown that in a couple cases that the random sample have a significantly higher consistency with their personal laptop. There are multiple cases where the consistency on the personal laptop is much higher than the control laptop.

4.3.6 Control Consistency vs. Personal Consistency (Random and Convenience Sample)



The equation of the line y=x has been displayed. While there are more points above the line, it is shown that in a couple cases that the random sample and convenience sample have a significantly higher consistency with their personal laptop. There are multiple cases where the consistency on the personal laptop is much higher than the control laptop.

5 Discussion

Our data analysis reports a correlation coefficient of 0, indicating that our data sample is completely random. The average typing speed in our data is 75 WPM. Typing.com, a typing school website, says that the average speed for high schoolers is 30-40 WPM. High school students at California Crosspoint Academy generally have a higher-than-average typing speed. The average accuracy in our data is 91.3%. Typing.com cites the average accuracy for high school students is 95%. While students at our school are faster than average at typing, they are less accurate than average.

There is no linear correlation between hand size and typing speed. This is most likely due to the fact that typing requires dexterity, which is a not a given with big hands. Having larger hands may make it easier or faster to reach the keys, but it does not necessarily allow a person to type faster. Typing speed is most likely correlated more closely with the amount of typing practice a person has. This means that typing speed is a learned skill, rather than a natural one.

https://typingcom.helpscoutdocs.com/article/240-wpm-averages-grade-level

6 Conclusion

If we were to repeat our study, we would start collecting data even earlier because data collection was the hardest part of the study. In our case, there is no relationship between hand size and typing speed. We think the frequency of computer usage is the main factor. For example, in our convenience sample, we tested many tech people/coders; they use their laptops more often than others, so they

type faster compared to our other sample. Even though some people have large hands, if they do not use computers often, their typing speed is not as fast. Our results do not match with our hypothesis "Students at CCA with larger hand tend to type faster." After detailed collection and analysis of data such as WPM, Accuracy, Consistency, and defining hand characteristics, our results show that sometimes people with smaller hands can type faster than people with larger hands.