An empirical analysis says yes: do people who type fast on the keyboard also type faster on a mobile device

Cong Liu - 1006267200

2022-12-05

Abstract

A news article reports that people can achieve similar typing speed on mobile phones as on computer keyboards. To confirm this finding, we measured typing speed and typing performance factors of students enrolled in STA490. Through multiple linear regression, we show that across individuals, the most significant predictors of typing speed on a mobile phone are keyboard typing speed ($p \approx 0.0004$) and mobile phone type ($p \approx 0.0066$). People who type fast on a keyboard **also** type faster on a mobile device.

Introduction

Due to the COVID-19 pandemic, remote learning has become a new norm. Mobile phones are a lot more portable and accessible than computers and laptops for instant and seamless conversations. When using telecommunication technologies, the efficiency of communication is likely to depend on how fast people can type. For students who need efficient communication in online learning, it is ideal that they can touch type on their mobile phones with the same proficiency as when typing on a computer keyboard. The findings presented in this news article can highly encourage those who wonder how much of their keyboard typing performance can be transferred to typing on a mobile device: people can now type on their smartphones almost as fast as they can on their computer keyboards.

As statisticians, we want to collect data to formally investigate this research question: **Do people who type fast on a keyboard also type faster on their mobile phones?** Our analysis leverages data on typing speed and relevant characteristics collected from students enrolled in a fourth-year statistics course (STA490) at the University of Toronto. Using the individual-level data, we employ a multiple linear regression model to identify the correlation between computer keyboard typing speed and the speed at which people type on a mobile phone while controlling for the effect of other relevant covariates.

The remainder of the report is organized as follows. The Data Summary introduces the data sample and its collection process. Methods describe considerations behind using multiple linear regression. The Results section reports the regression results followed by the Discussion section that discusses limitations and future considerations for this analysis. Finally, the Conclusion section concludes the findings and summarizes the implications of the analysis.

Data Summary

All 39 students enrolled in STA490 tested their English typing speed using an online platform. Except for the sampling convenience, we also expect this student sample to represent other senior undergraduates in Canadian universities reasonably well. On both a keyboard and a mobile device, STA490 students tested their typing speeds three times. Later, the students reported their results from each trial in a survey form. The survey also asks for students' self-assessments of factors that we believe are relevant to typing performance (e.g. students' keyboard usage frequency, whether the student has a physical injury at the time of tests).

The online typing test has a time limit of 60 seconds. The test takers need to type out randomly generated words. For typing speed measurement purposes, each English word is standardized to be five characters or keystrokes long, including spaces and punctuation. The online typing test platform thus bases its Words Per Minute (**WPM**) calculation on the following conversion rule: **5 keystrokes equal 1 WPM**. The records of the test results entail (i) the typing speed (in *correct* WPM) and (ii) the accuracy (percentage of the correct keystrokes).

For each typing device (computer keyboard and mobile phone), we average the results from the three test trials so that each individual contributes only one value to the analysis. The averaging is necessary for the study design and the answer to our research question. The three trials were designed for getting a more precise estimate of individuals' typing speed on the two types of devices. By averaging the repeated measurements, we hope to eliminate irrelevant idiosyncrasies (e.g. measurement errors). In addition, our research question wants to understand whether a positive correlation between keyboard typing speed and mobile phone typing speed exists **across individuals** rather than validating the existence of the association within each individual. Thus the average of individuals' typing speed measurements is sufficient for this analysis. In total, we have 39 measurements of individuals' average typing speed on both a keyboard and a mobile phone, along with their self-report of other potential performance predictors available for analysis. Table A1 in the Appendix presents summary statistics of our STA490 data sample.

Figure 1 **Scatterplot between keyboard and mobile typing speed**

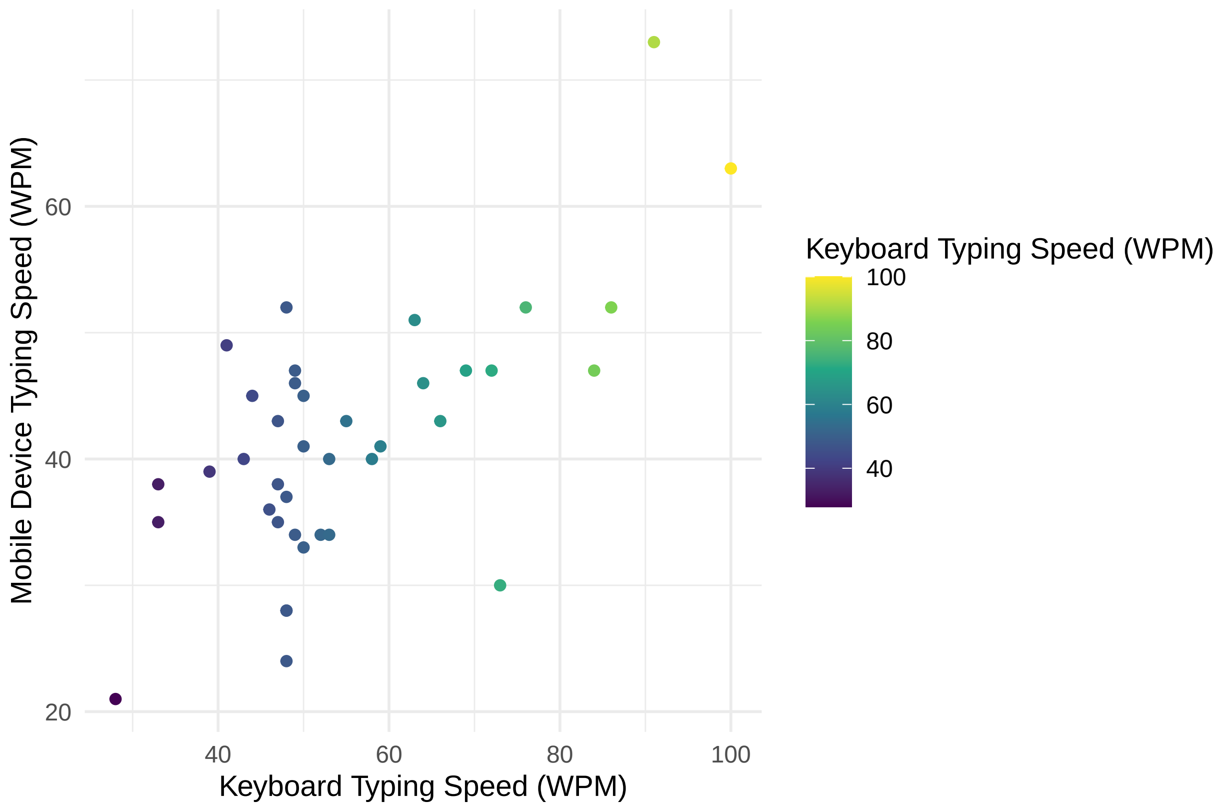


Figure 1 along with other plots of potential predictors for mobile typing performance (Figure A2 in the Appendix) motivates the method chosen in this analysis, which is to be elaborated in the Methods section.

Methods

We choose to use multiple linear regression (**MLR**) to model the effect of keyboard typing speed and other relevant covariates on how fast people type on a mobile phone.

From Figure 1 above, we can observe a positive linear trend between students' typing speed on a keyboard and typing speed on a mobile phone, after averaging measurements across the three trials. This motivates the use of a linear model for the relationship between keyboard typing speed and mobile phone typing speed. Observations from Figure A2 (in the Appendix) show that other covariates are also linearly associated with mobile phone typing speed. For example, the diagonal length of the mobile screen has a negative linear relationship with people's typing speed on a mobile phone. In addition, people with higher English proficiency tend to type faster on a mobile phone and the increment in typing speed is linear.

Incorporating graphical insights into our consideration, the most appropriate model for the data is multiple linear regression. A multiple linear regression model allows the addition of extra covariates to the linear model. Results from an MLR can help us determine whether keyboard typing speed is a significant predictor of how fast an individual can type on a mobile phone among the features collected in the student-level data, controlling for other covariates.

Since the focus of the research question is the relationship between keyboard typing speed and typing speed on a mobile device, we need to decide whether an MLR is indeed necessary. The model competing with the MLR is a simple linear regression (**SLR**) only using the keyboard typing speed to model the mobile device typing speed. To assess and compare the candidate models' goodness of fit to our data, we use the likelihood ratio test (**LRT**). At the predetermined significance level of 0.05, LRT can tell us whether at least one additional covariate is redundant in the richer MLR model. If we get a p-value smaller than 0.05 from the LRT, we can conclude that adding some of the variables is statistically significant.

Like all models, the validity of an MLR model is under the premise of satisfying assumptions. To evaluate whether our fitted model violates any of the assumptions, we look at plots of the model residuals. If our MLR model satisfies the assumptions, when plotting the residuals versus the fitted values, we would see a random scatter of the data points that is symmetric around zero. The quantile of the standardized residuals will also agree with the theoretical Normal quantiles - the quantile points follow closely to the diagonal line. The MLR assumptions and their evaluation are fully described in Appendix.

Results

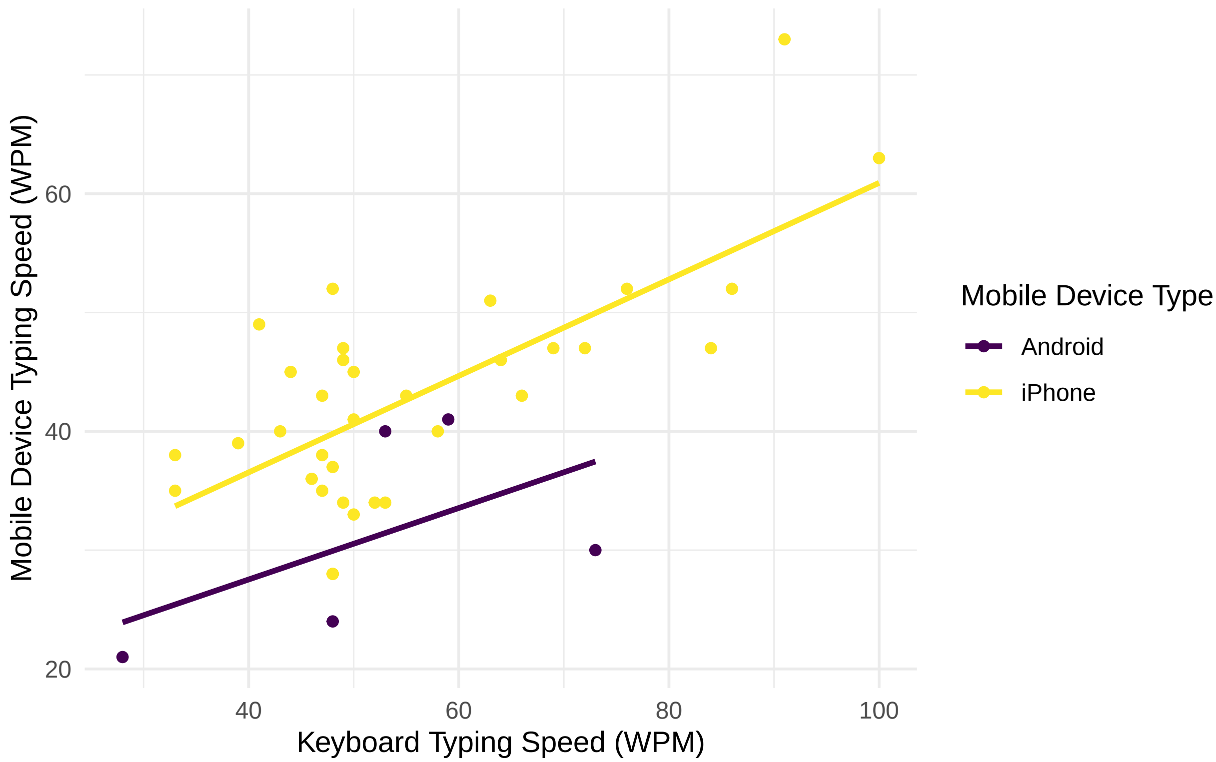
After conducting LRT, the test statistic we get is $\chi^2_7 \approx 17.236$, which is associated with a p-value less than 0.05 ($p \approx 0.01541$). Our richer MLR model fits better to typing test data sampled from the STA490 students, compared with the SLR. The diagnostics plots (Figure A4-A6 in Appendix) show that our MLR model does not have a major violation of assumptions. Thus we can proceed with interpreting the results from the MLR model.

Table 1: Significant Result From Regression on Mobile Phone Typing Speed

Variable	Estimate	Std. Error	t-statistic	95% Confidence interval
Keyboard typing speed (WPM)	0.3387	0.0843	4.0186	(0.1666, 0.5109)
Mobile device type: iPhone	10.5396	3.6083	2.9209	(3.1704, 17.9087)

Figure 2: Keyboard versus Mobile Typing Speed across Mobile Device Types

iPhone users are going to win the touch typing race?



We first highlight the predictors that have a statistically significant effect on mobile phone typing speed. In terms of statistical significance, the best predictor of STA490 students' mobile typing performance is the average keyboard typing speed from the three test trials, which has the smallest p-value ($p \approx 0.0004$) among all model variables. The 95% confidence interval of the coefficient estimate for the keyboard typing speed is (0.1666, 0.5109). Connecting back to our research question, the statistical significance supports us to answer it with a "yes": those who type fast on a keyboard **also** type faster on their mobile phones. An STA490 student who types 10 words per minute faster on a keyboard is estimated to type at least 3 WPM faster on a mobile phone, holding all other model covariates constant. Other than an individual's typing speed on a keyboard, the type of mobile phone being used also has a statistically significant impact on how fast one can type on a mobile phone (95% CI 3.1704 to 17.9087). The MLR model estimates that, when controlling for all other predictors in the model, iPhone users can type at least 10 WPM faster than Android users. iPhone users have a significant advantage in touch typing. When all other model variables are held constant, the improvement in phone typing speed by using an iPhone is roughly equivalent to a speed increase brought by a person typing 33 WPM faster on the keyboard.

Table 2: Comparing Direction of Regression Results

Variable	Estimate	Coefficient direction	Graph direction
Age	-1.6423	Negative	Negative
Gamer: Yes	1.1650	Positive	Positive
English fluency: Professional	1.4491	Positive	Positive
English fluency: Full	2.8090	Positive	Positive
Keyboard type: Raised	4.9906	Positive	Positive
Mobile device screen size (cm)	0.9399	Positive	Negative

Based on our MLR model, none of the other covariates have a statistically significant effect on mobile phone typing speed. Yet it is reassuring to see that most of the coefficient estimates have the same direction as the trend seen from graphical exploration (Figure A2 in Appendix). English fluency level is indeed positively correlated with how fast a person can type on a mobile phone. Compared with an individual with conversational fluency in English, someone who is professionally fluent can type around 1 to 2 WPM faster on a mobile phone if all other factors are held constant. The improvement in phone typing speed doubles for those who are fully fluent in English. Controlling for all other model variables, people with full fluency in English are estimated to type 2 to 3 WPM faster on their phones. Being a gamer also increases peoples' phone typing speed, while aging slows people down when typing on a phone unsurprisingly.

Yet there is one unexpected estimate. The MLR model yields a positive coefficient estimate for the relationship between mobile phone screen size and typing speed on a mobile phone, while graphically we observe a negative trend between screen size and phone typing speed (Figure A2 in Appendix). This leads to concerns about the potential limitations of the MLR model, which are discussed in more depth in the Discussion section below.

Discussion

Limitation

The MLR model estimates a positive correlation between the screen size of the phone and typing speed on mobile phones, contrary to the negative trend shown in Figure A2 (Appendix). The lack of variation in data shown in Table A1 in the Appendix may explain the inconsistency between the model estimate and the graphical trend. The popularity of the iPhone in STA490 class not only shrinks variations in mobile phone type but also affects the variation of screen sizes. iPhones have similar sizes despite the fact that students may have different iPhone models. Less variation in the screen sizes makes it difficult for the model to identify the true relationship between screen size and the speed at which people type on their phones.

There is evidence that interactions between typing performance predictors and keyboard typing speed should have been included in our MLR model (Figure A3 in Appendix). For example, the slopes of the trend line between keyboard typing speed and phone typing speed are notably different for typing on a mechanical (raised) keyboard versus a laptop (low-profile) keyboard. The positive association between using a raised keyboard and phone typing speed could be due to a correlation between keyboard type and keyboard typing speed rather than the type of keyboard directly affecting how fast a person types on a phone. Yet in the final model, we still exclude the interaction terms since our data is likely to be insufficient to justify the interaction effects, and there is a threat of model overfitting if we fit additional variables in the model.

Future Direction

As discussed, an inadequate model results from insufficient data. One potential solution to facilitate our analysis is to create "bootstrapped" samples from our STA490 data. That is, we can have a bigger dataset consisting of new samples from our small data at hand with replacement. Bootstrapping provides us with a larger sample to train our model without distorting the original sample. By building models on the bootstrapped data, we have a better chance of capturing the relationship between collected variables and typing speed on the phone.

More ideally, we can collect more data using an improved study design. The population of interest for the research question can be broadly described as "people who type fast on a keyboard". In our investigation, we restrict the study population to fourth-year undergraduate students at a Canadian university. Yet it is arguable whether our convenience sample from the STA490 class at the University of Toronto can represent this restricted study population. A better sample would be, for example, 100 fourth-year students at UofT from different programs. Instead of using the same online test platform, we can test the typing speed using randomly generated article excerpts instead. Usually, we type logically coherent sentences rather than randomly generated words. The coherence allows us to anticipate the words coming next and hence we are likely to be able to type at a higher speed. Typing randomly generated words does not properly simulate real-world typing for the test takers. Students' test results from typing the excerpts can better reflect their actual typing performance. It would be beneficial for the analysis to fix the time lapse between each typing test trial and to increase the number of trials so that the repeated measurements can help capture more information on people's typing performance.

Conclusion

To conclude, our analysis says "Yes" to the research question "do people who type fast on a keyboard also type faster on their mobile phones?". The analysis focuses on typing test results and information relevant to typing performance collected from STA490 students. A multiple linear regression model is fit to the data to investigate the relationship between students' keyboard typing speed and mobile phone typing speed. From the model, we find that the positive correlation between keyboard typing speed and mobile phone typing speed is statistically significant. Controlling for other model variables, someone who types 10 WPM faster on a keyboard is estimated to type at least 3 WPM faster on a mobile phone on average. The MLR also identifies the mobile phone type as the other significant predictor. On average, people who use an iPhone can type at least 10 WPM faster than Android users when all other model variables are held constant.

Appendix

Summary Statistics

Table A1: Summary Statistics of the STA490 Sample

Variable	N	Mean	Std. Dev	Min	Max
Averaged keyboard typing speed (WPM)	39	55.359	16.052	28	100
Mobile typing posture in tests	39				
... One finger	1	2.6%			
... Two fingers	38	97.4%			
Mobile device type used in tests	39				
... Android	5	12.8%			
... iPhone	34	87.2%			
Screen size of mobile device used in tests (cm)	39	15.706	1.268	11.9	17.876
Keyboard type used in tests	39				
... Low-profile	33	84.6%			
... Raised	6	15.4%			
Age at the time of tests (years)	39	21.462	0.996	20	24
Keyboard usage frequency at the time of tests	39				
... Every day	37	94.9%			
... Weekly	2	5.1%			
Gamer at the time of tests	39				
... No	26	66.7%			
... Yes	13	33.3%			
Plays a keyboard instrument at the time of tests	39				
... No	30	76.9%			
... Yes	9	23.1%			
Physical limitation related to typing at the time of tests	39				
... No	35	89.7%			
... Yes	4	10.3%			
English fluency level at the time of tests	39				
... Conversational	13	33.3%			
... Professional	15	38.5%			
... Full	11	28.2%			
Averaged mobile device typing speed (WPM)	39	41.462	9.949	21	73

Relevant Visualizations

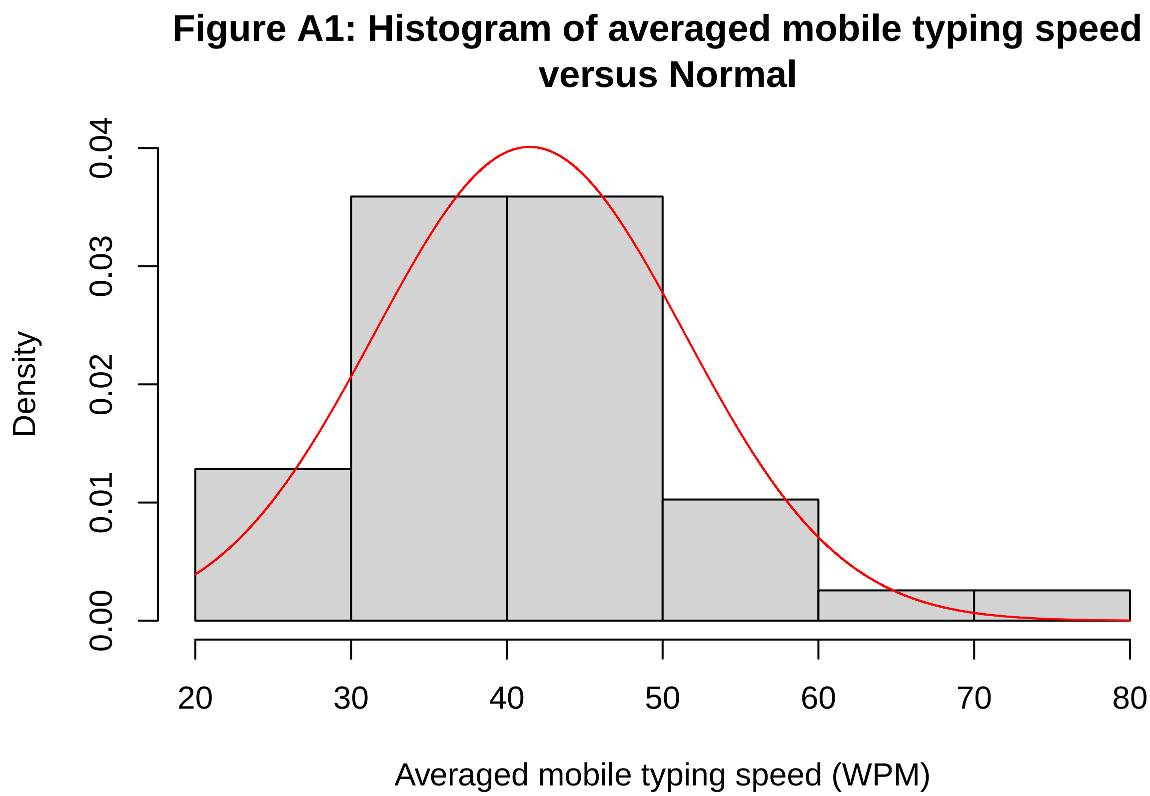


Figure A2

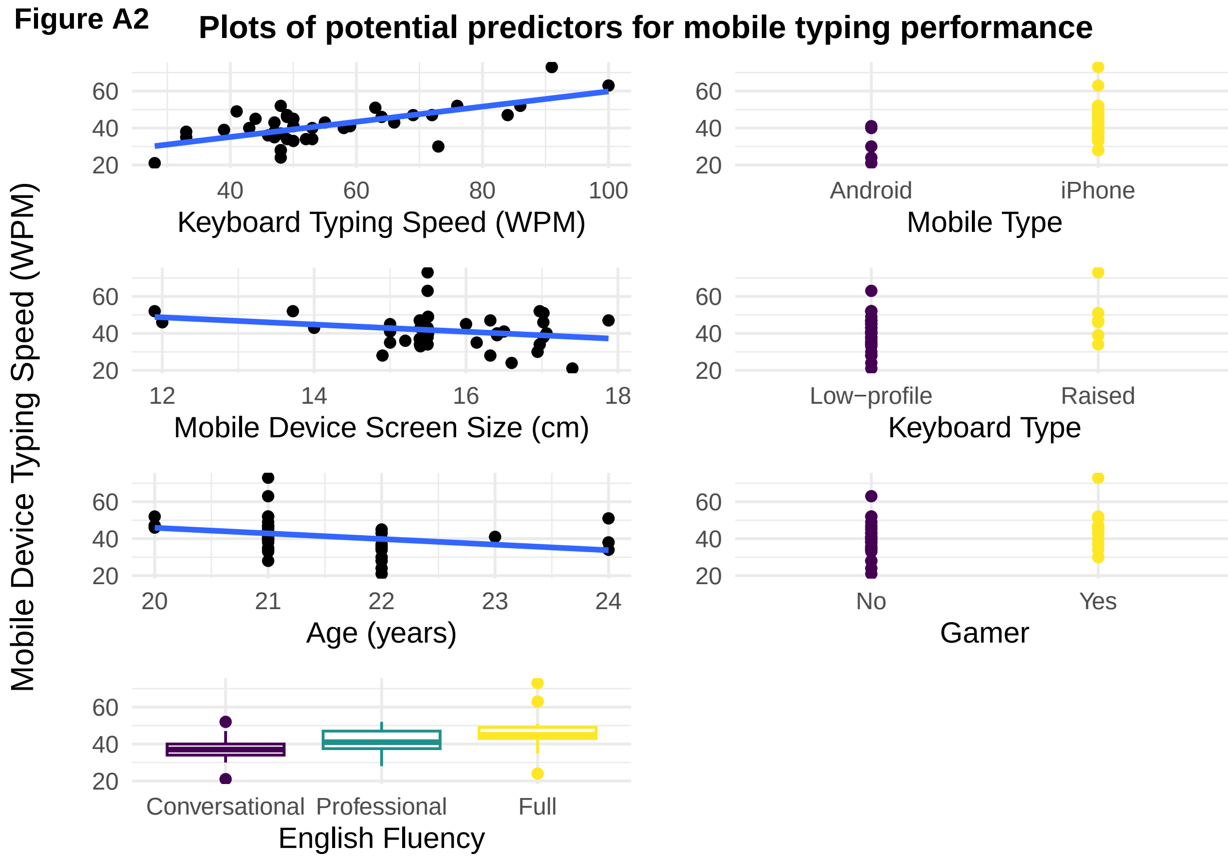
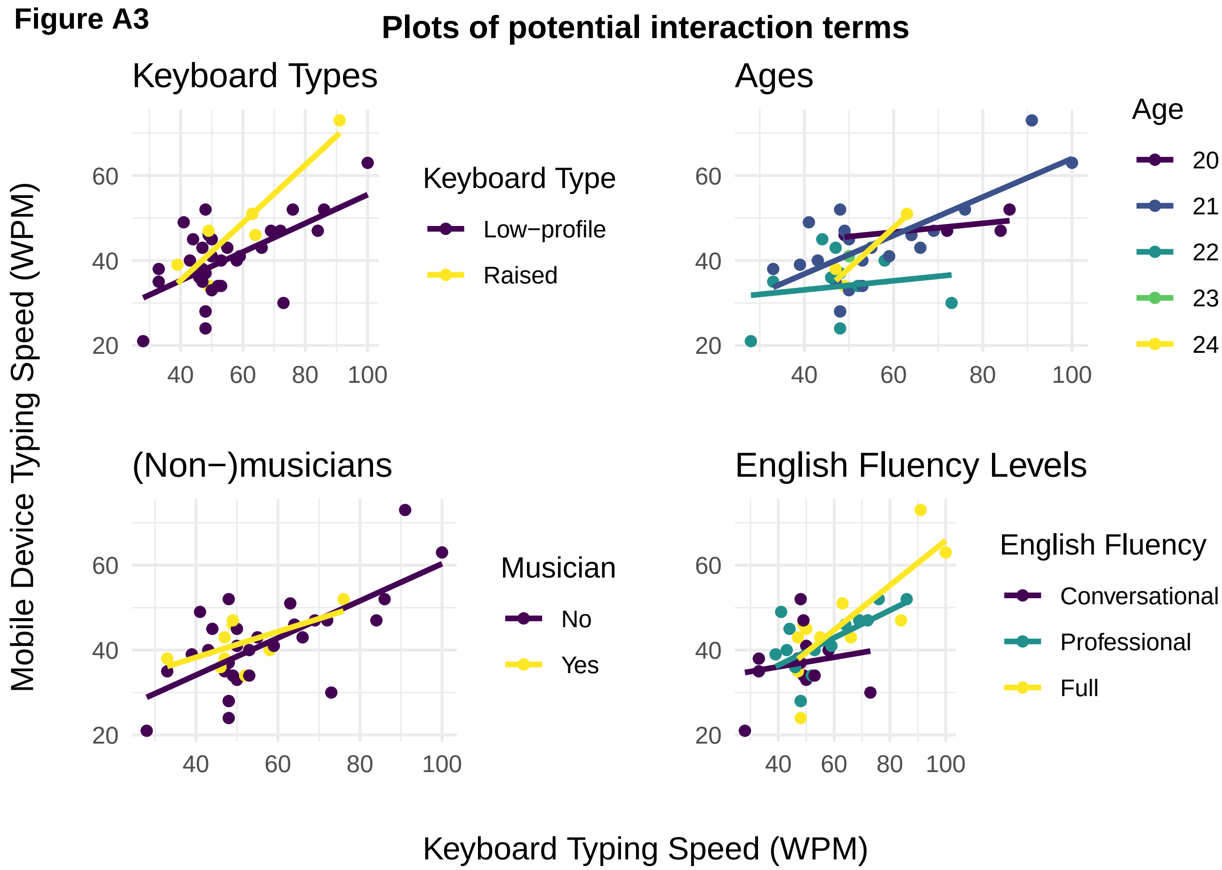

Plots of potential predictors for mobile typing performance
Mobile Device Typing Speed (WPM)
Keyboard Typing Speed (WPM)
Mobile Type
Android
iPhone
Mobile Device Screen Size (cm)
Keyboard Type
Low−profile
Raised
Age (years)
Gamer
No
Yes
English Fluency
Conversational
Professional
Full

Figure A3



Model Assumptions and Diagnostics

We can be assured that the multiple linear regression model gives us reliable coefficient estimates if the data satisfies, or at least does not have a major violation of, the assumptions listed below.

1. Independence of observations

Other than enrolling in the same course, there is no known inter-correlation among the individuals relevant to typing speed (e.g. the course enrollment is not based on any evaluation of typing proficiency). Thus we do have **39 independent observations**.

2. A linear relationship

The scatterplot between keyboard typing speed and mobile typing speed shows an approximately linear trend. From Figure A2, there is no sufficient evidence suggesting that other variables have a non-linear effect on the mobile device typing speed.

3. No perfect correlation between model predictors

There is some concern since some factors (e.g. students' age, physical limitation and English fluency) affect the typing speed on both devices simultaneously.

4. Error terms have a constant variance.

5. Error terms follow a Normal distribution.

The multiple linear regression also assumes that error terms follow a Normal distribution and have a constant variance. Error terms with a constant variance ensure the stability of our coefficient estimates. Others can be convinced by the model results if the variability of the coefficient estimates is well-controlled. The normal distribution is a nice-to-have. Mathematically, normality gives us convenient properties when we try to dig deeper into the coefficient estimates.

From Figure A1, the approximately close-to-normal distribution of the mobile typing speed give us some confidence that our data do not have a major violation of the assumptions. In addition, three residual plots are closely examined to check whether error terms follow a Normal distribution and have a constant variance.

1. Plot of residuals versus fitted values
2. Plot of residuals versus leverage
3. Normal Quantile-Quantile plot of the standardized model residuals' quantile against the theoretical Normal quantiles

If the model is suitable for the data, we can see:

- From Plot 1 – Majority of the values fall in the rough range of [-2, 2] – Random scatter of the values, no observable pattern/trend – Rough symmetry around 0
- Plot 2 – Falls inside the boundary defined by the Cook's distance, a measurement of how influential an observation is to the fitted model
- Plot 3 – Quantiles of the standardized residuals agree well with the theoretical Normal quantiles. – The quantile points follow closely to the diagonal

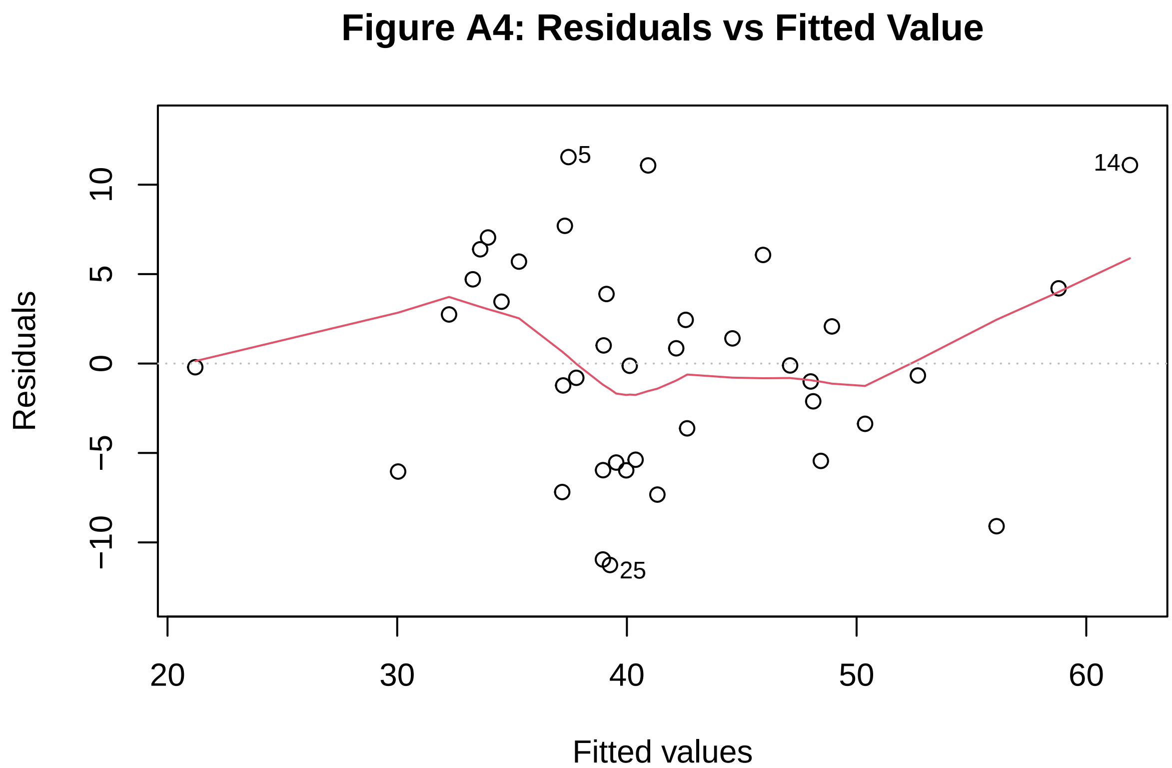


Figure A4: By visual inspection,

- no systematic pattern in the residuals, linearity of the relationship
- no cluster of residuals, independence of errors
- no fanning pattern, constant variance

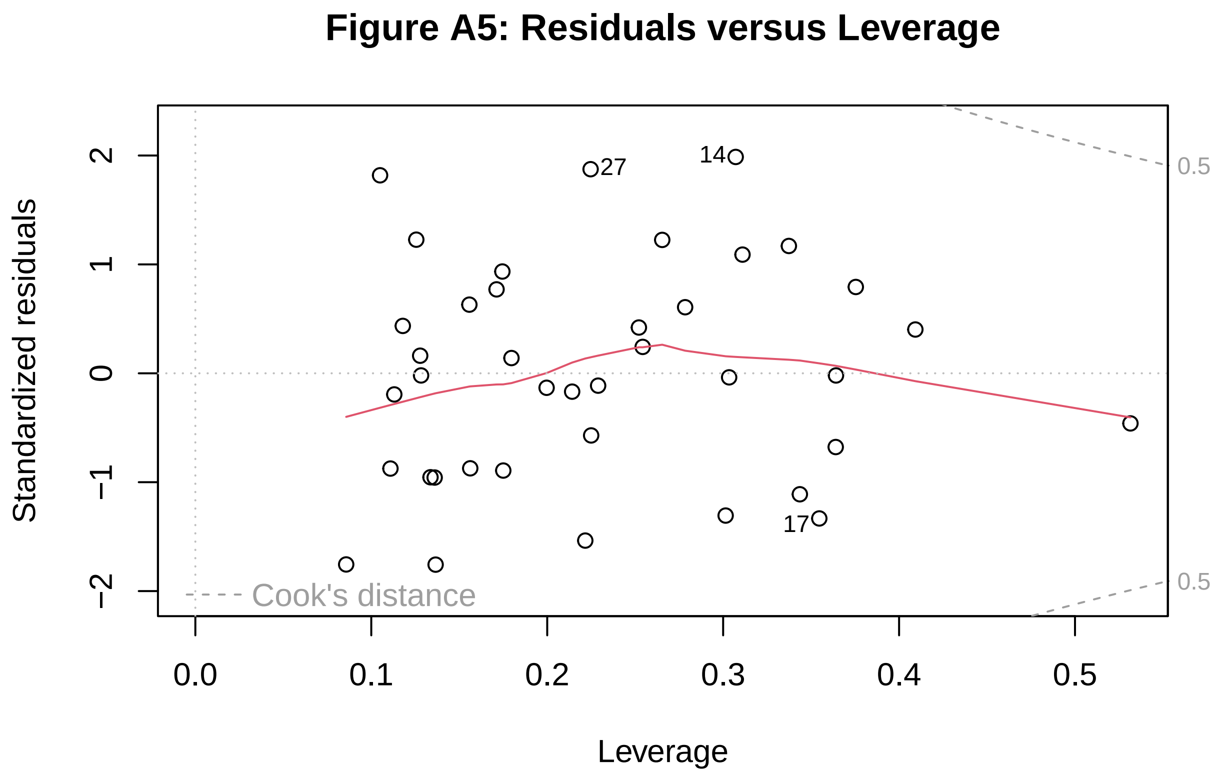


Figure A5: Observation #17 lies closest to the border of Cook's distance, but it doesn't fall outside of the dashed line. This means there are not any influential points in our regression model.

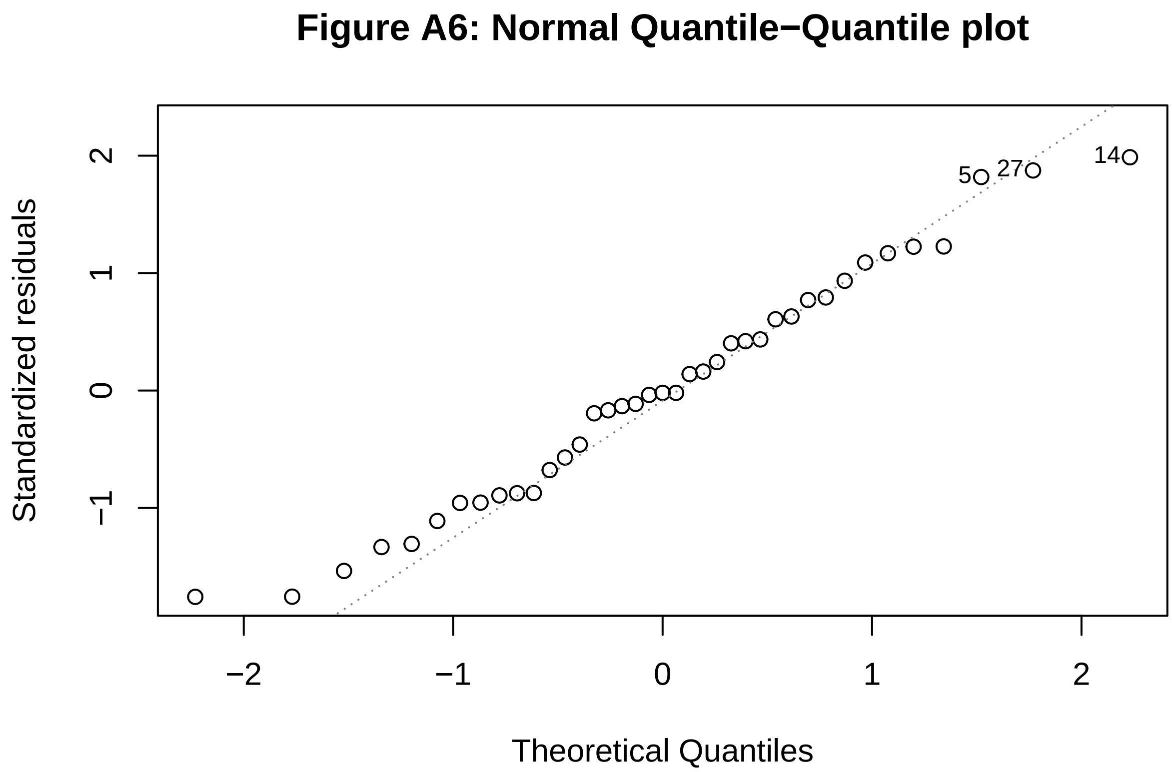


Figure A6: Majority of the standardized residuals stay close enough with the theoretical normal quantiles.

Full Regression Results

Table A2: Regression Analysis of Mobile Phone Typing Speed on Selected Predictors

Term	Estimate	Std. Error	t-statistic	p-value
Intercept	31.5007	33.8860	0.9296	0.3600
Keyboard typing speed (WPM)	0.3387	0.0843	4.0186	0.0004
Mobile device type: iPhone	10.5396	3.6083	2.9209	0.0066
Mobile device screen size (cm)	0.9399	1.0606	0.8862	0.3826
Keyboard type: Raised	4.9906	3.5977	1.3872	0.1756
Age	-1.6423	1.2787	-1.2844	0.2088
Gamer: Yes	1.1650	2.8232	0.4126	0.6828
English fluency: Professional	1.4491	2.9447	0.4921	0.6262
English fluency: Full	2.8090	3.0746	0.9136	0.3682