Understanding why Keyboards Result in Faster Typing Speeds Compared to Phones and Investigating Contributing Influential Factors

Name: Dhanraj Patel

Date: Nov 18 2022

TA: George Stefan

Abstract

In this study, 40 students students were tasked with preformed three trails of a typing test on both and phone and keyboard and answer a series of questions relating to their typing abilities. Using this data this study investigated its primary research question of which typing method resulted in a faster typing speed. It was found that on average using a keyboard resulted in much faster typing speed compared to phones. In addition, the study also analyzed which quantitative and qualitative factors most influence typing speeds on phones and keyboards. To accomplish this linear regression was utilized with mean typing scores used as a response variable for both phones and keyboards. To conclude, for keyboard typing it was found that being a gamer significantly increased the participant's typing speed and age and if the participant was injured significantly decreased typing speeds. While for phones we found that having a smaller phone screen size, having high typing accuracy, having a apple device as apposed to an android device, being fluent in English and being a gamer significantly improved typing speeds on phones.

Introduction

The main purpose of this study is to investigate the typing speeds when comparing typing on keyboards vs. mobile phones. As typing has become a vital skill in today's technological driven world, comparing and contrast the two typing methods give us key insights on which typing method is superior and the key factors that influence typing speeds. This will help us identify the qualities that increase tying acumen which can be utilized to help people become better at typing. For this study 40 university students took the same typing test three times on both keyboards and phones and were also asked a series of questions about themselves which are relevant to their typing abilities. Using this information, the main goal is to analyze both the primary and secondary research questions with the primary research question being which of the two typing methods results in faster typing speeds and the secondary research questions involving investigating quantitative and qualitative factors (if they type regularly/screen size) which most strongly contribute to the typing speeds.

In this report we will first look at at the important variables gathered from the data collection process. In this section we will identify what variables we have in our data set to work with and exactly what information was collected in the study. Moving on from there we will move onto the methodology section. In this section we will briefly overview the design of the study and explore the statistical methods were used in the study. In the next section we will examine the results of the study attained by our methodology. In the last section we will conclude by discussing the main takeaways from the results, discuss the limitations of our study and discuss ways to reduce our limitations in a reduplication study.

Important Variables

- Trial 1 results keyboard - Number of words typed on keyboard in trial 1 on
- Trial 1 accuracy results keyboard - percentage accuracy of words correctly typed on keyboard in trial 1
- Trial 1 results phone - Number of words typed on phone in trial 1
- Trial 1 accuracy results phone - percentage accuracy of words correctly typed on phone in trial 1

(please note that each one of the four variables above exists for all three three trials with 12 variables in total)

- Number of fingers used - Number of fingers used to type
- Size screen - size of the phone screen used for the trials
- Age - Age of participant
- Keyboard usage frequency - How often the participant uses a keyboard in daily lives
- Is participant a gamer - If the participant self identifies as a gamer
- Is participant a musician - If the participant self identifies as a musician
- Is participant injured - If the participant was injured when doing the trials
- English fluency level - How fluent the participant is in English

Methods

In this section we will be discuss the methodologies used to investigate our research question.

Study design

Before selecting which methods to use to answer the research question it is first important to understand the how the data selection portion of the study was designed and ultimately carried out. This is important as it effects which methods we can use in our analysis and if the assumptions in the chosen methods are . In this study, roughly 40 students who are in the same upper level statistics course were asked to take a typing test using both a keyboard and mobile phones, they all took three trials with each method of typing. For each trial they were asked to report their resulting words per minute score and their percent of typing accuracy. Lastly, all participants were asked a series of qualitative and qualitative questions relating to their typing speed relevant to their typing abilities and the specific devices they used for the typing test.

Methodologies for primary research question

With the study design understood, we are able to select appropriate methods to answer the primary research question which involves understanding if keyboards or phones result in faster typing score. Thus to identify which typing method resulted in a faster typing speed we can compare the mean number of words per minute across all three trials for both keyboard and phones for all the participants. It is appropriate to take the mean across all three trials to get the most accurate representation of the participants typing ability while minimizing bias attributed to taking the test multiple times. With the mean number of words per minute for both phone and keyboard calculated for all participants we can then calculate numerical (i.e. Minimum value, mean value) and graphical summaries (i.e. variance) for the data and compare the values between phones and keyboards. This methodology will aid us in analyzing which is the faster typing method by allowing us to directly compare typing speeds using statistical summaries and by better understanding statistical features of the typing speed data such as the variance of the data. From this we will be given a complete understanding of which typing speed resulted in faster typing speeds.

Methodoligies for secondary reaserch question

For the secondary research question we want to identify the the quantitative and qualitative factors that impact the typing scores for both phones and keyboards. All the participants were asked a series of questions relating to their typing speeds (e.i. if their work involves a lot of typing) and using these questions we want to identify which factors relate to their typing scores and the degree the factors influence their typing speeds. In order to accomplish this we will be using two linear regression models, one model for phones and one model for one model for keyboards. It is appropriate to use linear regression model as the statistical method allows us to compare and understand the relationship between the participants typing speeds with both quantitative and qualitative factors. In addition, the coefficients values will help us understand the degree which the factors impact the typing speeds allowing us to understand what is most influential.

Moving on to the components of the linear regression model used in this study, for both phones and keyboards the response variable will be the mean typing scores across all three trials for the keyboard and phone models respectively. Also it is reasonable to put it as the response as we want to understand the factors that influence typing against their scores in the trials. The main motivation for using two linear regression models (one for each typing method) is that will enable us to determine which factors most significantly effect typing speeds through model selection and to what degree each factor effected typing speeds by analyzing the coefficient values. By using two models instead of one we do lose information on how the two methods relate to each other, but instead we do retain information about the individual typing methods that would be lost by using only one method. To counteract the lose of information, we will be directly comparing the model variables and coefficient values between the two variables to understand how one typing method compares to the other.

As for the predictor variables for each model it is important to identify the best set variables to include in the models that helps us best understand what influences the participants typing abilities. We can not simply include all the predictor variables as it is not desirable due to many factors. Some of these factors include the fact that some variables might be irrelevant or redundant to the response variable. We can determine, irrelevant and redundant variables by using the adjusted R square of the model. The adjusted R square can be thought of as the value of the information that the predictors have on the response variable. If the value doesn't significantly change after adding a variable, we know that the variable doesn't add a lot of information to the model and that it should not be included. Also, there is an issue of colinearity with some variables being linked to each other (i.e. not independent) which would cause issues with the results of the linear regression models and an issue of over-fitting which causes the model to be unusable outside the scope of this specific study. To identify which predictor variables would be included in both models in this study we used the step-wise selection method. We first created models with mean typing score across all three trials for each model and then included all possible predictors relevant to the typing method and then we used the step-wise function. The way the step-wise function works is that it starts off with all the predictors and then cycles through every possible combination of predictors adding and subtracting variables in a step wise fashion. It cycles through all possible number of predictors and all combinations of predictors and calculates the adjusted R squared score for each model. Using this method we attain the a good combination predictors to use for our models. The final models are the following:

Keyboard model:

$$Mean\_Keyboard\_score_i = \beta_0 + \beta_1 age_i + \beta_2 is\_gamer_i + \beta_3 is\_injured_i + \epsilon_i$$

Phone model:

$$Mean\_Phone\_score_i = \beta_0 + \beta_1 mean\_phone\_accuracy_i + \beta_2 device_i + \beta_3 size\_screen_i + \beta_4 is\_gamer_i + \beta_5 fluency_i + \epsilon_i$$

After finalizing the models its important to check that the models do not violate the linear regression assumptions so that their results are accurate. With the use of linear regression we need to assume the following about our data:

- Normality: Response variable is distributed normally for any fixed value of the predictors,
- Linearity: The relationship the mean of the response variable and the predictor variables is linear.
- Independence: All observations are independent.
- Homoscedasticity: The variance of residual is the same for any value of the predictor variables.

To check if the assumptions are valid, we will be using diagnostic plots heavily involving fitted data and residuals. Residuals are the difference between the data points in the data set and the fitted values given by linear regression. To test for linearity we will be plotting the fitted values by the residuals, if there is no discernible pattern in the plot we know the assumption is valid. As for normality we will be plotting our residuals on a qq plot, if normal probability plot of residuals in the qq plot follow a straight line we know the assumption is valid. As for testing the Homogeneity of variance assumption we will be scale-location plot which plots fitted values by the square root of residuals. With this plot we will be able to identify if the residuals are spread equally along the predictors. By seeing a horizontal line we know that the variances are equal and the assumption is valid. Lastly, to test for independence we will be looking at the variance inflation factor (VIF) score of the model which measures the degree of colinearity. The lower the score the more independent the data is with a value lower than five indicating that the model does not break the assumption.

results

With the models finalized we can move onto the results of the study.

Primary research question results

Table 1: The trimmed mean is trimmed by 10 percent

Method	Min	Q1	Median	Q3	Max	Mean	Trimmean	Var	SD	Range
Phone	21.00	35.67	41.00	47.00	73.67	42.03	41.69	98	9.90	52.67
Keyboard	28.67	47.33	50.33	64.67	100.33	56.07	54.87	267	16.35	71.67

In Table_1 we can see the numerical and graphical summaries of the mean typing speeds for both keyboards and phones. We see that for phones, the min, q1, median, q3 and max values are lower than keyboards. In addition, the mean, trimmed mean and range values are lower for phones compared to keyboards. From this we can determine that keyboards typing speeds using the mean trials values were greater than phone typing speed values. In addition we notice that the variance and standard deviation values for keyboards are much higher than phones, so while keyboards did on average have higher typing scores it also contained a greater amount of spread in the data among the participants.

Secondary research question results

After determining which typing method produced faster typing speed we can move on and answer our secondary research question of understanding the quantitative and qualitative factors that influence typing. To do this we need to look which variables were selected by the step-wise model and the values of the coefficients. We will first begin with the factors that influence typing speeds on keyboards.

Table 2: Coefficient values in final model

category	vals	STD_ERROR	PVALS
intercept	165.103	50.336	0.00245
Age	-5.226	2.349	0.03304
Participant is gamer	14.159	5.144	0.00954
Participant is injured	-18.080	8.806	0.04806

After preforming the step-wise function, the variables that were found to most significant to the mean typing speed on keyboards across all three trails were age, if the participant is injured and if the participant is a gamer. Starting with age we see that the coefficient value is -5.226 meaning as the participant's age increases by one year then the mean keyboard typing score decreases at about 5 words per minute. In a similar vein, the participant is injured then their typing mean scores decrease by roughly 18 words per minute with a coefficient of -18.080. Lastly, if the participant self reported to being a gamer than their mean score increased by roughly 14 words with a coefficient of 14.159. In addition judging by the lowest p value being a gamer was also the most significant contribution to the mean keyboard scores.

Moving on, we will now discuss the findings for the typing test on phones.

Table 3: Coeffient values in final model

category	vals	STD_ERROR	PVALS
intercept	39.9180	16.4983	0.02181
Mean Phone Accuracy	0.2253	0.1373	0.11121
iPhone user	7.8717	3.6301	0.03818
size screen	-1.9689	0.6529	0.00518
Participant is gamer	4.6824	2.5850	0.08011
Participant is fluent in English	7.1447	3.0856	0.02761
Participant is professionally fluent in English	5.2323	2.8724	0.07850

From table_2 we can see the results of the coefficients for the phone model. We notice that in contrast to keyboard model the step-wise model resulted in a greater amount of variables in the model comparatively. Starting with mean phone accuracy we see that as the accuracy percentage increases by one percent that the typing score increases by 0.2 words. In addition if the participant uses an Iphone instead of an Andoid we see that the typing score increases by 7.8 words. Moving onto screen size. Each for one centimeter that the screen size is bigger, the typing score decreases by about 2 words. In addition, if the participant is a gamer than the typing score increases by 7 words. Lastly we see that being fully fluent and professionally fluent in English increase typing score by 7.1 and 5.2 words respectively. From the data we see that being a gamer and being fluent in English had the highest contributions to increasing typing scores on phones.

conclusion

To conclude, after asking 40 senior level university students the to take 3 trials of typing tests on phones and keyboards and answer a series of questions we were able to the data to answer the primary and secondary research questions of the study. From our data the main takeaways are that in the study typing on keyboards had higher typing speeds compared to phones, but typing scores on keyboards had a much higher variance compared to phones. In addition our secondary main take away is that the most significant factors that increase keyboard typing speeds are age, being a gamer and the absence of an injury. As for phones the main factors that lead to increasing typing scores are being an accurate in typing, using an Iphone, using a phone with a smaller screen size, being a gamer and being fluent in English.

After preforming our study it is important to understand some limitations that exist in our study. One major limitation has to do with the sampling group. Two issues with with the sample group is that only 40 people were included in the survey. With billions of typists the sampling group was comparatively small. In addition another major issue with the sampling group is that all the participants are very close in age and are all senior statistics student in UOFT. This greatly the sampling bias in the study. Lastly, one limitation that exists with the study is only one typing test was used, this greatly forces to study to lean on the goodness of the typing test, if the test is faulty or biased this will greatly effect results. With all the students being from a single forth year university class there is a large degree of sampling bias and colinearity. Since the subjects are all in the same university and around the same age they more more likely to have similar typing abilities. For future studies it is recommended that the study have a much larger subject count and that the subjects be as diverse as possible. This will help mitigate the issue of sampling bias and colinearity.