Mxb242

MODELING OF READING SPEEDS BY OPTIMIZING PASSAGE CHARACTERISTICS

Actual word count for Final Report:

Table of Contents

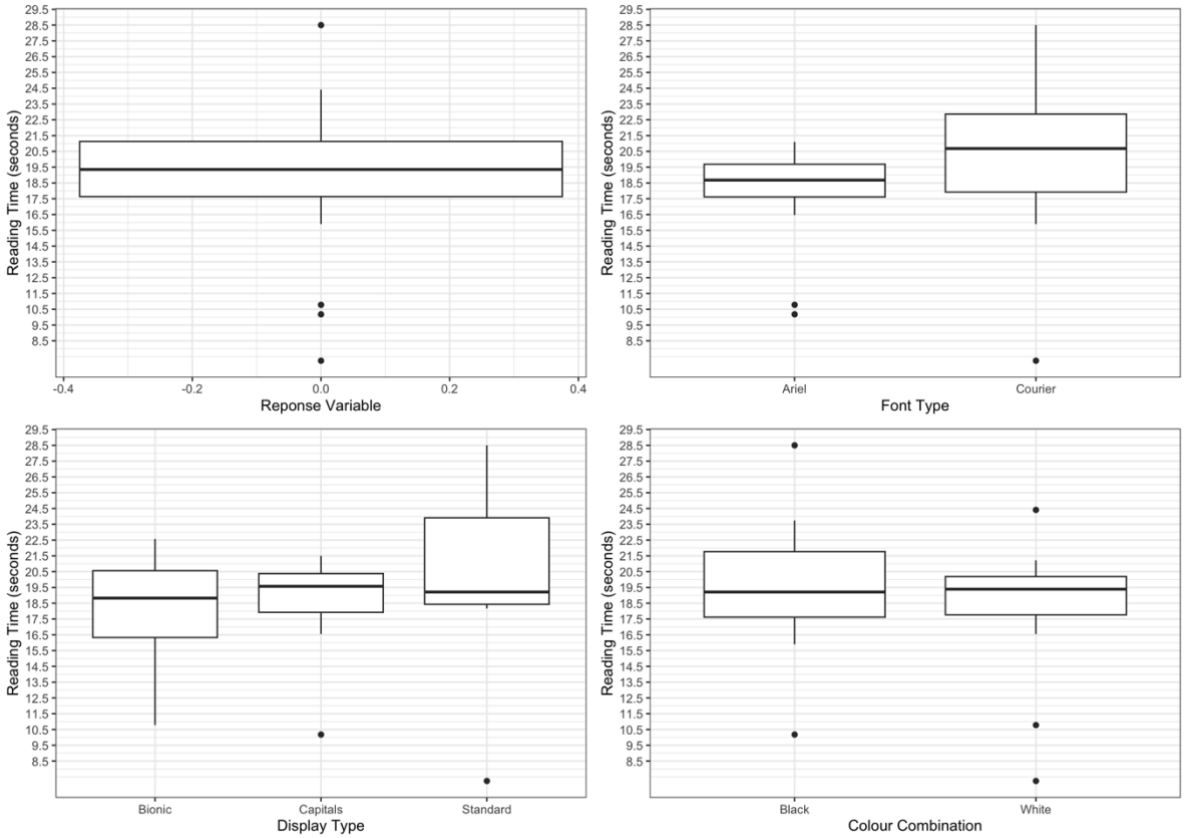
Abstract:

A study lead by TKK Digital Publishing Ltd is investigating typeface characteristics and their effects on general reading speeds. 24 participants were asked to read a passage with differing font types (Arial and Courier), display types (Bionic, Standard and Capital Letters) and screen colour (Black and white).

A revised linear model accounting for interaction of these variables revealed there was no significant statistical differences on reading speed. The outcomes of this study are attributed to an occurrence of a type-II error due to limited sample sizes and or the inadequacy of a linear model to fit this relationship. While no significant relationship was able to be determined, the study still highlights the strong dependencies within the variables.

4.0 Exploratory Analysis:

This section provides a surface analysis of the collected data. It aims to identify key trends and relationships between explanatory variables & the response variable.



Figures 1A: Box Plots of Response & Explanatory Variables

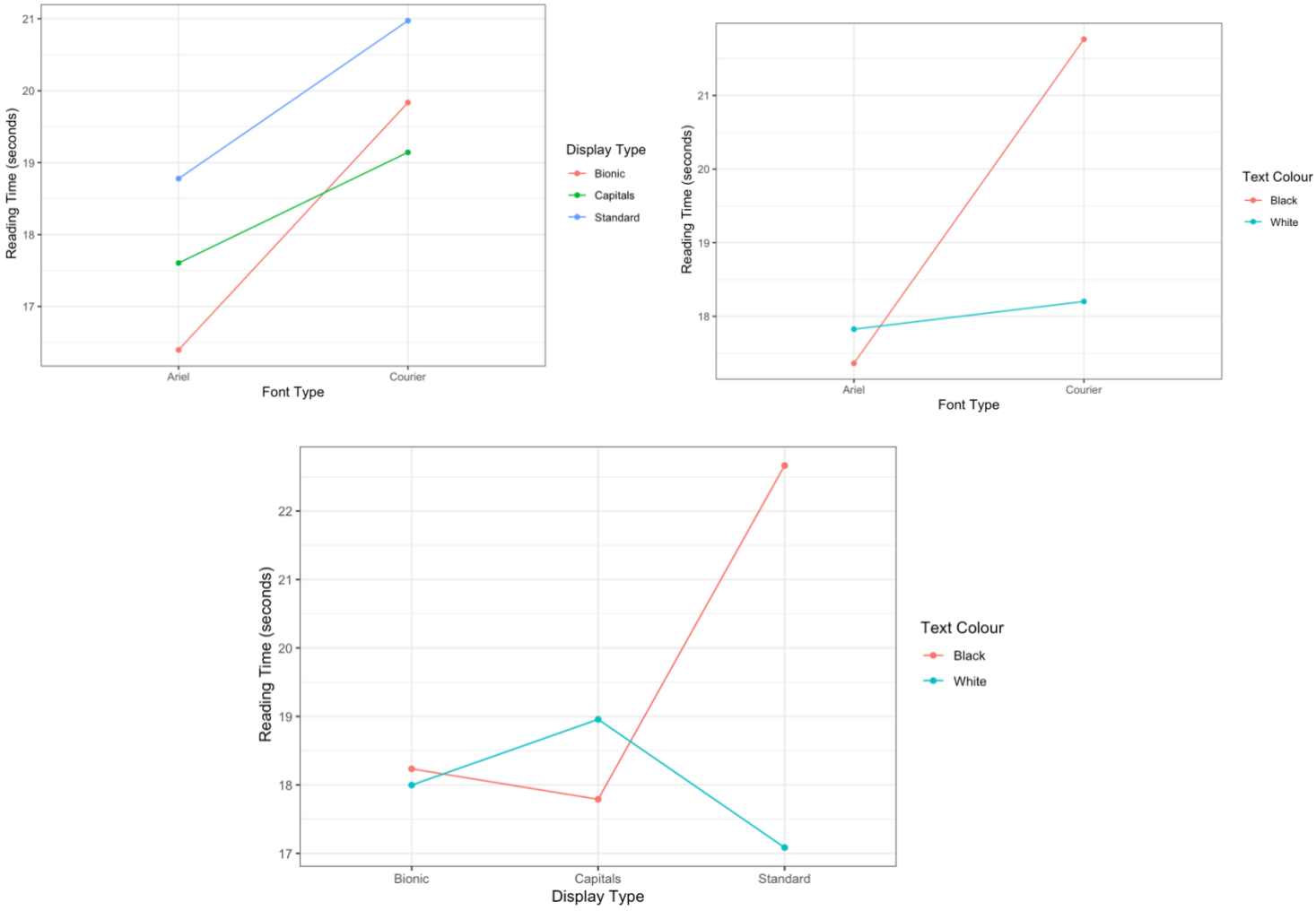
From Figure:1A, viewing the box plot of the response variable (reading time), there is a greater frequency of classified outliers within the lower portion (fast reading times) as

opposed to the latter, this indicates that from the sample there’s some responses that are abnormally fast when compared to the rest.

The distribution of the reading time against font type shows similar medians, with courier reading slightly slower suggesting ariel font is a more effective font for reading speed. Furthermore, the IQR for the ariel font has a considerably smaller spread compared to courier, however both IQR are within each other indicating that finding a significant difference between font types is difficult.

The distribution of reading time against the display type reveals similar trends, with only slight differences between median reading time scores. Display type capitals contains the lowest IQR spread with bionic and standard having similar larger IQR ranges. All IQR’s within display types also overlap suggesting finding a significant difference to reading time being difficult.

The distribution of reading time against colour combination again shows the median differences being minimal, with the black colour having a slightly faster median. Both colours have an IQR within each other, with the black level having a greater IQR spread implying less variation in reading times for the white colour condition.



Figures 1B: Interaction Plots Between Explanatory Variables

The interaction plots reveal trends implying that an interaction between covariates is present. Trends of font type with respect to display type demonstrate almost parrel lines, therefore an independent relationship among the variables is suggested. The trends also indicate that under all tested display types, the ariel font has a faster reading time than its alternative, ‘courier’ font.

This is contrary to the trends with font types with respect to colour as the clear cross over suggests a two-way interaction and therefore dependency. The data shows that under both colour conditions ariel has a fast-reading time, but when changing to courier under the black colour the reading time significantly increases. This interaction therefore suggests that effect of font type on reading time is dependent on the colour the text is displayed in. From these results it implies that the final model will include an interaction term between font type & colour.

Evidence of another interaction between covariates is present within the display text against text colour interaction plot. Clear cross-over occurs is the display types: bionic & capitals

under both colour conditions, thus implying the effect of display type on reading time is dependent on text colour. It's implied that the final model will also include a two-way interaction between display type & colour combination.

Although the interaction plots show potential dependency between multiple covariates, due to the low sample size this must be considered as it increases the variability within the results & introduces bias. However, from these explanatory trends, it suggests interaction terms would be present within the final model, with likely interactions between font type & text colour, as well as font type & text colour.

Flaws in Data Collection

The results suggest an issue with variability in the data. Reflecting on the data collection process provides insight into what improvements could be made if we were to investigate further. One issue the data collectors faced is availability of similar participants. In most instances, data collectors were unable to perform the test on similar individuals, testing with a variety of different demographics. Factors like age, cultural background and physical attributes varied between participants and may were not recorded during data collection and could not be accounted for in our model.

Another flaw in the data collection process was that it was assumed the time taken to speak a passage was an accurate reflection of the time taken to read a passage. This, coupled with the fact participants did not know they were being timed, led to a variance in how participants spoke the passage. Some participants happened to read quickly while others took time with pronunciation. In future, the best approach would be to let the participant read the passage and tell the experimenter when they finished.

The data collection did not consider the highly variable differences in reading speed that are not explained by changes in the three chosen factors. More practiced readers are likely to naturally read faster than those who do not frequently read. To rectify this, it is recommended that future data collection involves reading one standardised passage to gauge base reading speed and record this time for use as a continuous nuisance variable in an ANCOVA.

5.0 Results:

Statistical Methods:

This report will use a significance level of 0.05 as its industry standard, (Investopedia, n.d.), therefore the null hypothesis will be rejected with a p-value <0.05. The rejection of the null hypothesis doesn't indicate the alternative hypothesis being factual, rather evidence to reject the null hypothesis. Regression analysis techniques used within the report is linear regression through the stepwise approach. ANOVA will also be used and if applicable, the post-hoc analysis of Tukey's HSD incorporated.

Linear Regression:

The principle of linear regression is, "a technique that uses two or more independent variables to predict the outcome of a dependent variable, (Corporate Finance Institute, n.d.).

In the report's context, linear regression aims to model the relationship between the continuous variable (reading time) and the categorical predictor variables (font type, display type, text colour). This relationship is achieved through the utilisation of dummy variables that represent each level of the categorical predictors, as the predictor variables aren't continuous.

The assumptions that must hold for linear regression to be valid include:

Linearity: The relationship between the dependent variables and each independent variable should be linear, thus the change on independence should result in proportional changes in the dependent variable.
Independence: The observations from the regression analysis should be independent from one another.
Constant Variance: The variance of the error terms should be constant across the levels of the independent variables.
Normality: The residuals should follow a normal distribution as departures from normality can indicate that the model is not capturing all relevant predictors.
Multicollinearity: The independent variables should not be highly correlated with each other, this can cause unstable coefficient estimates and misinterpretation of the effects of individual predictors on the response variable.

These assumptions must hold to produce a valid multiple linear regression analysis and therefore interpretation of its results. Violations of these assumptions can lead to biased estimates and in-correct conclusions.

Analysis of Variance (ANOVA)

Analysis of variance (ANOVA) is a method to examine whether there are statistically significant differences in the means of the continuous response variable (reading time) across different levels of the categorical predictors (font type, display type, text colour).

ANOVA will determine the variability within the groups (each level of categorical predictors) and between the groups (between levels of the categorical predictors), along with a test-statistic (F-statistic) and its associated p-value. ANOVA will produce a statistically significant result if any of the factor level varies significantly from the overall population mean.

ANOV's assumptions are alike to linear regression assuming: independence, constant variance & normality. If the ANOVA results indicate a significant difference, a post-hoc analysis of Tukey's HSD will be used to assess the significance of differences between pairs of group means.

A full table of all tested hypothesis tests can be referred in APPENDIX A, with a sample of a single and two-way hypothesis test displayed below.

Null & Alternative Hypothesis for Font Type Covariate
$H_0: \mu_{A_1} = \mu_{A_2} = \cdots = \mu_{A_\alpha}$ *after accounting for B & C, i.e.* $\alpha_1 = \alpha_2 = \cdots \alpha_\alpha = 0$ H_a: *At least two means differ* *where* α = *effect of ith level of Font*
Null & Alternative Hypothesis for Font & Display Type Interaction
H_0: *Factors A & B do not interact to affect the mean response ,* *i.e* $(\alpha\beta)_{jk} = 0$ *for all* j, k H_a: *Factors A & B interact to affect the mean response* *where* α = *effect of ith level of Font & where* β = *effect of ith level of Display,*

ANOVA analysis was conducted using the full model, which is defined as:

Informal Full Model:
$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk} + \epsilon_{ijk}$
Definitions of Coefficients
where μ = *population mean of Y* *where* α = *effect of ith level of Font ,* *where* β = *effect of ith level of Display,* *where* γ = *effect of ith level of TextColour,* *where* $(\alpha\beta)_{ij}$ = 2 − *way interaction effect obetween Font & Display,* *where* $(\alpha\gamma)_{ik}$ = 2 − *way interaction effect obetween Font & TextColour,* *where* $(\beta\gamma)_{jk}$ = 2 − *way interaction effect obetween Display & TextColour,* *where* $(\alpha\beta\gamma)_{ijk}$ = 3 − *way interaction effect obetween Font, Display & TextColour,* *where* ϵ_{ijk} = *error term &* $N \sim (0, \sigma^2)$

The results table of the full factorial ANOVA analysis from R software includes:

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Font	1	34.27	34.27	1.368	0.265
Display	2	14.44	7.22	0.288	0.755
TextColour	1	14.42	14.42	0.575	0.463
Font:Display	2	3.72	1.86	0.074	0.929
Font:TextColour	1	24.32	24.32	0.970	0.344
Display:TextColour	2	50.70	25.35	1.011	0.393
Font:Display:TextColour	2	39.54	19.77	0.789	0.477
Residuals	12	300.73	25.06		

ANOVA Analysis

A summary of the ANOVA results is listed below in **Table 1.** Detailed analysis of the ANOVA results is in **Appendix A**.

Factor	P-Value	Significant Effect (Yes or No)
Font (main effect)	0.265	No
Display (main effect)	0.755	No
Text Colour (main effect)	0.463	No
Font: Display (2-way interaction)	0.929	No
Font: Text Colour (2-way interaction)	0.344	No
Display: Text Colour (2-way interaction)	0.393	No
Font: Display: Text Colour (3-way interaction)	0.477	No

Table 1: A summary of the ANOVA results

According to the above **Table 1,** p-value for each factor (main effect); Font, Display and Text colour is greater than 0.05. Also, p-value for each interaction tested; Font: Display, Font: Text Colour, Display: Text Colour and Font: Display: Text Colour is greater than 0.05.

It has been selected earlier that a statistical significance (alpha) value of 0.05 for this experiment. A p-value larger than 0.05 declares that the null hypothesis is not statistically significant, so the null hypothesis is failed to be rejected. Therefore, it can conclude that there is strong evidence to suggest the followings:

- Text Font Type has no significant effect on mean reading time, after accounting for Display type and Text Colour.
- Text Display Type has no significant effect on mean reading time, after accounting for Font Type and Text Colour.
- Text Colour (black or white) has no significant effect on mean reading time, after accounting for Font Type and Display Type.
- There is no interaction effect between Font and Display Type average reading time.
- There is no interaction effect between Font and Text Colour on the average reading time.
- There is no interaction effect between Text Colour and Display on the average reading time.
- There is no interaction effect between Font, Text Colour and Display on the average reading time.

As the above ANOVA analysis did not reveal any significant difference between levels of any factor been tested, or any interactions been tested, no further analysis has been conducted. If it was discovered that a significant difference between any factor levels or interactions exists, then further analysis would be conducted such as Tukey's HSD. This post-hoc analysis would provide additional insight to understanding the exact differences between the comparing levels and their relative statistical significance.

Factorial ANOVA populates more reliable outcomes when there's a large range of data points. Limited variation of data, such in the investigation, can deteriorate the quality of analysis. Having various and diverse data points will improve the analysis and provide a more vigorous understanding of the factors of the experiment (font type, display type, text colour). To improve the Normality, larger sample size could have been used to ensure data will approximate a normal distribution.

Initial Regression Model

The initial regression model used is also the 'full model', this therefore includes all interactions between the explanatory variables. Below represents the formal equation of the "Full" linear regression model in this investigation:

$$Y = \beta_0 + \beta_1 F_c + \beta_2 D_c + \beta_3 D_s + \beta_4 C_w + \beta_5 (F_c D_c) + \beta_6 (F_c D_s) + \beta_7 (F_c C_w) + \beta_8 (D_c C_w) + \beta_9 (D_s C_w) + \beta_{10} (F_c D_c C_w) + \beta_{11} (F_c D_s T_w) + \epsilon$$

where:

$$F_c = Font\ Courier$$
$$D_c = Display\ Capital$$
$$D_s = Display\ Standard$$
$$T_c = Text\ Colour\ White$$

Interactions between these variables are denoted by the coefficient variables appearing in brackets, note that the model uses baseline coding using baseline variables:

Font: Ariel Display: Bionic Text Colour: Black

Summary Table of Coefficients & their Estimated Value:

Coefficient	Estimate Value
β_0	17.235
β_1	2.000
β_2	-1.595
β_3	1.975
β_4	-1.675
β_5	2.300
β_6	4.910
β_7	2.875
β_8	5.605
β_9	0.810
β_{10}	-8.400
β_{11}	-12.305

The table below displays the output using R software of the full linear regression model:

```
Call:
lm(formula = Time ~ Font * Display * TextColour, data = df)

Residuals:
   Min     1Q Median     3Q    Max
-8.585 -1.619  0.000  1.619  8.585

Coefficients:
                                          Estimate Std. Error t value Pr(>|t|)
(Intercept)                                17.235      3.540   4.869 0.000386 ***
FontCourier                                 2.000      5.006   0.400 0.696536
DisplayCapitals                            -1.595      5.006  -0.319 0.755499
DisplayStandard                             1.975      5.006   0.395 0.700121
TextColourWhite                            -1.675      5.006  -0.335 0.743709
FontCourier:DisplayCapitals                 2.300      7.080   0.325 0.750872
FontCourier:DisplayStandard                 4.910      7.080   0.694 0.501181
FontCourier:TextColourWhite                 2.875      7.080   0.406 0.691825
DisplayCapitals:TextColourWhite             5.605      7.080   0.792 0.443905
DisplayStandard:TextColourWhite             0.810      7.080   0.114 0.910804
FontCourier:DisplayCapitals:TextColourWhite -8.400     10.012  -0.839 0.417884
FontCourier:DisplayStandard:TextColourWhite -12.305    10.012  -1.229 0.242629
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5.006 on 12 degrees of freedom
Multiple R-squared:  0.3762,    Adjusted R-squared:  -0.1955
F-statistic: 0.658 on 11 and 12 DF,  p-value: 0.7519
```

As is to be expected following the results of the ANOVA analysis, the initial regression model indicates that none of the tested factors have a statistically significant effect on a person's reading speed.

As mentioned previously, the multiple linear regression model incorporated dummy variable encoding which sets a predictor variable’s level with a baseline.

From the analysis, the only statistically significant coefficient is the intercept (p-value: $3.86*10^{-4}$), which represents the estimated reading time when all dummy variables are set to the baseline (0 encoding). Therefore, the intercept coefficient interpretation is that there is evidence to reject that the intercept is equal to 0, implying that there is a non-zero baseline reading time (grand mean) when all predictor variables are set to their reference levels. The respective baseline reading time is 17.235 seconds.

The residuals from this model have a range of between -8.585 & 8.585, with a median of 0. The standard residual error of represents that on average, the actual reading time deviates from the predicted reading time by approximately 5.006 units.
The R^2 vale indicates how well the independent variables explains the variability of the dependent variable. From this model, approximately 37.62% of the variation in reading time is explained by the independent variables: font type, display type & text colour.

The adjusted R^2 value of -0.1955 indicates that the model has no predictive value and either removing predictors or a different model should be investigated further.

The F-statistic tests the overall significance of the regression model as it compares the variability explained by the model to the variability that remains unexplained, with a F-statistic of 0.658 & a p-value of 0.7519, this determines that the model is not statistically significant.

Based on these findings from the initial model, it can be concluded that there is no evidence to support a significant relationship between the measured predictor variables and reading time. The initial model does not adequately explain the variability and no predictor variables have a significant effect on reading time.

Typically, when the p-value is greater than the alpha threshold (>0.05), the coefficients wouldn't be interpreted, but to provide insight to what the main effect coefficients would imply if significant include:

1) A participant that uses font courier would increase the estimated reading time by 2 units if all other variables are held constant at their base line, therefore display would be bionic & text colour black.
2) A participant using the capital display type would reduce their estimate reading time by 1.595 units when all other variables are held constant, therefore ariel font & text colour black
3) A participant using text colour white would reduce their estimated reading time by 1.675 units if all other variables are held constant, therefore ariel font and bionic font type.

Analysis of Unusual Observations:

The analysis of the experimental data is paramount as the experimental data may include unusual observations that can have a major influence on the reliability and validity of the developed linear model. Unusual observations can derive from randomness or error within the investigation. Unusual observations cannot be disregarded immediately but must be investigated to test its influence and therefore reasoning to include/exclude within the data.

Unusual observations were analysed within the full linear model defined previously.

Outliers:

Outliers are classified as unusual in the y-direction, this is defined by using the standardised residuals: $r_i^{standard} = \frac{Y_i - \hat{Y}_i}{s\sqrt{1 - H_{ii}}}$, where $s = \sqrt{\frac{\Sigma(Y_i - \hat{Y}_i)^2}{n-(k+1)}}$ is the usual estimate of the error standard deviation σ. The bounds to define an observation as an outlier are $\left|r_i^{standard}\right| > 2$. To account for the standard error including the outlier observation in the standard error estimate, the studentised residuals is used which excludes the outlier observation in the standard error estimate. An observation to be defined as an outlier in the student residuals also has the same bounds defined as: $\left|r_i^{student}\right| > 2$.

Using R, both student & standardised tests were performed to identify outliers, which found two observations having an absolute value > 2 in both tests:

Studentised Test	
Observation	**Score**
6	-3.25191038
20	3.25191038
Standardised Test	
Observation	**Score**
6	-2.42523922
20	2.42523922

Leverage Points

Leverage points are classified as extreme observations within the x-direction that can have a disproportionate & significant influence on the estimated line of fit & thus regression model. Leverage points are constrained between: $\frac{1}{n} \leq H_{ii} < 1$. To be classified as a leverage point, it must be greater by a multiple of 2 of the average leverage, $avg(Hii) = \frac{k+1}{n}$, therefore can be formally represented as: $H_{ii} > \frac{2(k+1)}{n}$.

Using R, it was found that no observations were classified as leverage points within the observations.

Influential Observations:

Influential observations can be both or either outlier and leverage points, they are defined as observations that have a significant influence on the fitted line if included when compared to if excluded. To quantity this influence, the observations are measured through *cooks distance*, defined as:

$D_i = \sqrt{\frac{\Sigma_{j=1}^{N}(\hat{Y}_j - \hat{Y}_{j(-i)})^2}{(k+1)s^2}}$, with $(-i)$ representing the exclusion of ith sample from the estimate. The bounds to define an influential observation is $D_i > 1$. Using R, it was found that no observation was defined as an influential observation. The table below represents observation6 & 20 relative cooks distance:

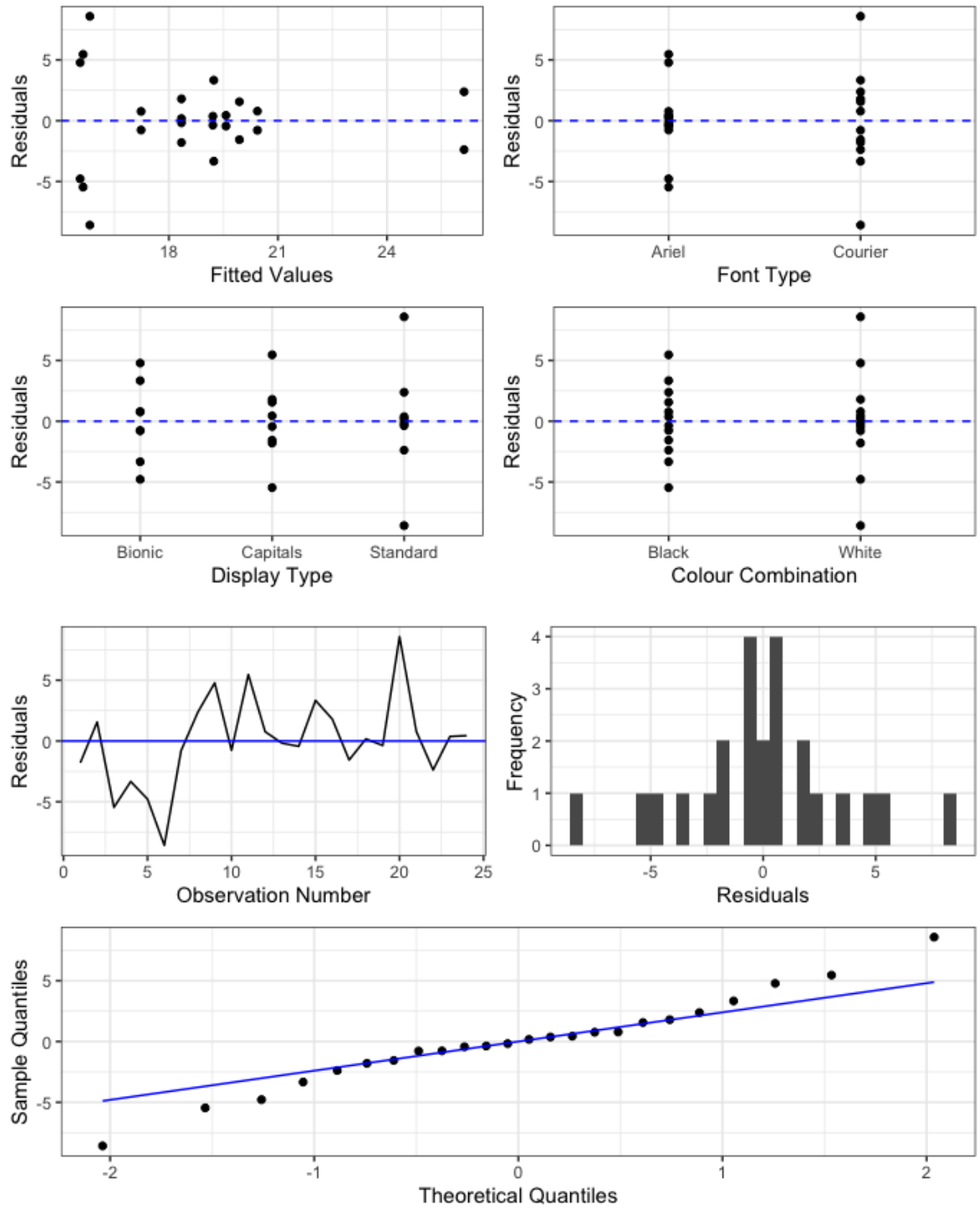
Observation	Cooks distance
6	0.4901487726
20	0.4901487726

Due to no observations being formally classified as leverage points & influential observations, no observations were removed from the investigation's data, however, with the classification of 2 outliers, this must be carefully considered when interpreting regression

results. This is because it can cause bias in the interpreted relationship between categorical predictor levels and the response variable.

Analysis of Initial Regression Model Assumptions

Below presents the assumption plots for the initial ‘full’ linear model.



Figures 2A: Linear Regression Assumption Plots

Some issues with the data are uncovered by the model assumption testing.

Firstly, there appears to be some tapering in for the fitted values vs residuals plot, suggesting that the data may not be accurately described by a linear model. The tapering of residuals against fitted values also indicates non constant variance.

Through the plots of residuals against all individual categorical variables, fanning is present within the graphs, further indicating non-constant variance trends.

Although tapering & fanning exists within the residual plots against the fitted & all explanatory variables, there isn't enough data to suggest that a significant non-linear relationship exists, but due to the small sample size, identifying significant trends is extremely difficult.

Within the residuals against observation order, the trend suggests that the first 6 observations have a lower trend than others, with the remaining residuals having relatively normal variance. This affect may be attributed to the variance between data collectors within the experiment.

The QQ plot also suggests breaches of normality within the tail ends (left & right) deviating from the QQ-line. This therefore diminishes the quality of the model, noting however that the ANOVA results are quite robust to issues with normality.

Based off the experimental design being complete & randomised, it's expected that this design hasn't formally introduced dependency between the variables.

Although the plots demonstrate clear violations of constant variance & normality, the sample size is considerably small and can increase the risk of bias within the model.

Final Regression Model

To determine a linear regression model which is more appropriate for the distribution, the stepwise regression approach was used. Although stepwise 'determines' the most effective linear model for the distribution, model effectiveness is dependent on the evaluation method.

There is no single evaluation criteria to indicate an optimal model, rather a selection of tools. AIC (Akaike Information Criterion) is a score that accounts the 'goodness' of fit & complexity of the model which penalises by the number of predictors. BIC (Bayesian Information Criterion) is alike but differs in the penalty given for the parameter quantity. Furthermore, AICc (correct AIC) adjusts for smaller sample sizes & implies a smaller bias factor, (REFF). Among these criteria also exists the familiar R and R^2 values which as mentioned refer to the effectiveness of the models ability to explain the data's variability.

In the context of this investigation, the AIC & R^2 values will be used to evaluate the appropriate of the model. Therefore, the stepwise approach will iterate through the linear models & eliminate predictor variables to reduce its relative AIC score, thus suggesting an optimal linear regression model.

The table below represents the initial linear regression model compared to the final iteration of the backward stepwise regression process.

Informal Defined Model	AIC SCORE	R^2
Time ~ Font + Display + TextColour + Font*Display + Font*TextColour + Display*TextColour + Font*Display*TextColour	84.68	0.3762
Time ~ 1	74.004	*NA*

The results from backward stepwise approach determined that the final linear regression model that had the lowest AIC score was a linear regression model that included no explanatory variables.
Stepwise regression using the forward direction was also incorporated to verify this final model.

Informal Defined Model	AIC SCORE
Time ~ 1	74.004
Time ~ Font	74.234
Time ~ TextColour	75.276
Time ~ Display	77.275

The results from the "forward" stepwise regression determine that any inclusion of a main effect independent variable (explanatory variable) would increase the models AIC score. Therefore, based on the findings from the stepwise approach, it has concluded that none of the predictors or their interactions significantly contribute to explaining the data & that a model with no predictors is the most appropriate and best fitting according to the AIC score.

The final linear model is defined as Y = 18.79 seconds, this represents the intercept and that the prediction of all response times being 18.79 is the most accurate compared to the addition of any predictor variables. Below represents the summary table of the described final model & the formal model:

$$\hat{Y} = \beta_0$$
$$\hat{Y} = 18.79$$

```
Call:
lm(formula = df$Time ~ 1)

Residuals:
     Min       1Q   Median       3Q      Max
-11.5483  -1.1508   0.5667   2.3417   9.7117

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  18.7883     0.9346    20.1 4.32e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.578 on 23 degrees of freedom
```

6.0 Conclusion:

The investigations' purpose was to aid TKK Digital Publishing Ltd in finding type face characteristics to optimise for reading speed. The initial questions proposed asked if the explored variables had a significant effect on reading time, with the measured variables being:

1) Font type: ariel & courier
2) Display type: bionic, standard & capital
3) Colour combination: black & white

It also proposed what combination of measured typeface variables determine the most optimised reading speed.

After statistical analysis, it found the measured font types, display types and text colours did not have a significant effect on reading speed. Even through a revised linear model with the initial inclusion of interactions, it indicated that no predictor variables had a significant effect on reading speed. This finding can be caused by 3 main reasons:

1) There is no relationship between reading speed and typeface characteristics measured within this investigation.
2) There is a relationship between the response variables and explanatory variables, but due to the small sample size a type-II error has occurred failing to detect a significant relationship between the predictors and dependent variable.
3) A relationship between the response variable & the explanatory variables exists but it cannot be significantly represented through a linear model and other model types such as logistic should be used.

Recommendations:

It should be noted that the sample size within this investigation was considerable small, with a 2 replicates per treatment. Furthermore, reading speed has extremely high variability between sampling units and therefore produces extremely 'noisy' data making determining significant relationships difficult. Although this investigation couldn't identify a significant relationship between the measured font type, display type or text colour variables, future iterations could include:

1) ANCOVA by measuring a participant's base-reading speed as a continuous nuisance variable could enable improved relationships.
2) Increase the sample size and replicates per treatment to enable a larger sample size improving the ability of the linear regression models to identify a significant relationship.

References

Investopedia. (n.d.). P-Value. In *Investopedia*. https://www.investopedia.com/terms/p/p-value.asp#:~:text=A%20p%2Dvalue%20less%20than,null%20hypothesis%20is%20not%20rejected.

Corporate Finance Institute. (n.d.). Multiple linear regression. In *Corporate Finance Institute*. , (https://corporatefinanceinstitute.com/resources/data-science/multiple-linear-regression/#:~:text=Multiple%20linear%20regression%20refers%20to%20a%20statistical%20technique%20that%20uses,variable%20in%20the%20total%20variance.) .

Appendix:

Appendix A:

Factor	**Null hypothesis**	**Alternative hypothesis**	**P value**	**Accept/ reject null hypothesis**	**Reasoning**
Font	Font has no effect on mean reading time, after accounting for display and text colour. Where are the population mean reading time for the font type Arial and Courier	Where are the population mean reading time for the font type Arial and Courier	0.265	accept	P value is higher than 0.05, thus accept the null hypothesis. Font type has no effect on mean reading time.
Display	display has no effect on mean reading time, after accounting for font and text colour. Where are the population mean reading time for the Display types, Capital, Sentence case and Biometric	or at least two differ. Where are the population mean reading time for the Display types, Capital, Sentence	0.755	accept	P value is higher than 0.05, thus accept the null hypothesis. Display type has no effect on mean reading time.

		case and Biometric			
Text Colour	Font has no effect on mean reading time, after accounting for display and Font. Where are the population mean reading time for the Text Colour Black and white	Where are the population mean reading time under the Text Colour Black and white	0.463	accept	P value is higher than 0.05, thus accept the null hypothesis. Text Colour has no effect on mean reading time.
Font : Display (Interaction)	There is no interaction effect between font and display on the average.	There is an interaction effect between font and display on the average.	0.929	accept	P value is high, thus accept the null hypothesis. There is no interaction effect between font and display average.
Font : Text Colour (Interaction)	There is no interaction effect between font and text colour on the average.	There is an interaction effect between font and text colour on the average	0.344	accept	P value is higher than 0.05, thus accept the null hypothesis. There is no interaction effect between font and text colour average.
Text Colour : Display (Interaction)	There is no interaction effect between text colour and display on the average.	There is an interaction effect between text colour and display on the average	0.393	accept	P value is higher than 0.05, thus accept the null hypothesis. There is no interaction effect between text colour and display average.
Font : Display : Text Colour (Interaction)	There is no interaction effect between font, text colour and display on the average.	There is an interaction effect between font, text colour and display on the average	0.477	accept	P value is higher than 0.05, thus accept the null hypothesis. There is no interaction effect between font, text colour and display average.

Figure 1 - Experimental Design

	Arial Font	Courier Font	
Sentence Case Display (default)	2	2	Black Font/White Background
	2	2	White Font/Black Background
Upper case Display			Black Font/White Background
			White Font/Black Background
Bionic Display			Black Font/White Background
			White Font/Black Background

Appendix B

Null & Alternative Hypothesis For Font Type Covariate
$H_0: \mu_{A_1} = \mu_{A_2} = \cdots = \mu_{A_a}\ after\ accounting\ for\ B\ \&\ C, i.e. \alpha_1 = \alpha_2 = \cdots\ \alpha_\alpha = 0$ $H_a: At\ least\ two\ means\ differ$
Null & Alternative Hypothesis For Display Type Covariate
$H_0: \mu_{B_1} = \mu_{B_2} = \cdots = \mu_{B_b}\ after\ accounting\ for\ A\ \&\ C, i.e\ \beta_1 = \beta_2 = \cdots \beta_\beta = 0$ $H_a: At\ least\ two\ means\ differ$
Null & Alternative Hypothesis For Colour Type Covariate
$H_0: \mu_{C_1} = \mu_{C_2} = \cdots = \mu_{C_c}\ after\ accounting\ for\ A\ \&\ B, i.e\ \gamma_1 = \gamma_2 = \cdots \gamma_\gamma = 0$ $H_a: At\ least\ two\ means\ differ$
Null & Alternative Hypothesis For Font & Display Type Interaction
$H_0: Factors\ A\ \&\ B\ do\ not\ interact\ to\ affect\ the\ mean\ response$, $i.e\ (\alpha\beta)_{jk} = 0\ for\ all\ j, k$ $H_a: Factors\ A\ \&\ B\ interact\ to\ affect\ the\ mean\ response$
Null & Alternative Hypothesis Display & Colour Type Interaction
$H_0: Factors\ B\ \&\ C\ do\ not\ interact\ to\ affect\ the\ mean\ response$, $i.e\ (\beta\gamma)_{jk} = 0\ for\ all\ j, k$ $H_a: Factors\ B\ \&\ C\ interact\ to\ affect\ the\ mean\ response$
Null & Alternative Hypothesis For Font & Colour Type Interaction
$H_0: Factors\ A\ \&\ C\ do\ not\ interact\ to\ affect\ the\ mean\ response$, $i.e\ (\alpha\gamma)_{jk} = 0\ for\ all\ j, k$ $H_a: Factors\ A\ \&\ C\ interact\ to\ affect\ the\ mean\ response$
Null & Alternative Hypothesis For Font, Display & Colour Type Interaction
$H_0: Factors\ A, B\ \&\ C\ do\ not\ interact\ to\ affect\ the\ mean\ response$, $i.e\ (\alpha\beta\gamma)_{jkl} = 0\ for\ all\ j, k\ \&\ l$ $H_a: Factors\ A, B\ \&\ C\ interact\ to\ affect\ the\ mean\ response$

Appendix C – Meeting Minutes

Meeting Summaries	
Time and Location	**Summary**

09/04/2024, 16:00-18:00
 	
Week 6 Workshop
	• Marks for data collection plan released
	  ○ Reviewed feedback as a group
	  ○ Team overall happy with result
	• Todo for future report.
	  ○ Agree on changes to experiment as per the feedback.
	  ○ Print and sign the document confirming every team member has reviewed and understands the data collection plan.

14/04/2024, 16:00-18:00
 	
Week 7 Workshop
	• Refinement of experimental plan
	  ○ Narrowing of experimental scope.
	  ○ Consultation with Sara about our project and how to avoid pitfalls
	• Formalisation of final experimental data collection plan
	  ○ R script developed for 24 random experimental trials

	○ Trials divided up ○ Agreement to collect data within the next 2 weeks
23/04/2024, 16:00-18:00 Week 8 Workshop	• Some data has been collected. By all members. ○ Agreement all data to be collected next week • Temporary "Dummy data" generated in R to work on preliminary data analysis script. • "Dummy Data" used to script graphs to make them Readable for reporting purposes.
30/04/2024, 16:00-18:00 Week 9 Workshop	• All data collected for 24 random trials allocated in week 7 workshop • Creation of PowerPoint shared document and rough outline of presentation structure • Agreement that data interpretation and analysis in written report should be completing before summarising on PowerPoint
07/05/2024, 16:00-18:00 Week 10 Workshop	• Significant progress made on data analysis final report ○ Much progress made at home • Data analysis and graphs created • Graphs Ready for Presentation • Organisation of PowerPoint Presentation slide structures.
14/05/2024, 16:00-18:00 Week 11 Workshop	• Work started on bringing important graphs into presentation • Agreement on slide designs • Slides were made and a script was written well before the due date.
17/05/2024, 16:00-18:00 Week 11 PowerPoint Presentations	• 14:00-17:00 Group 20 Presentation • All went well
17/05/2024, 16:00-18:00 Week 12	• Working on final report. • Progress is made in online calls • Many team members balancing end of semester assignment rush.
17/05/2024, 16:00-18:00 Week 13	• Final touches on report • Formalising final interpretations of the study • Fixing up of report formatting

Appendix D: Paragraph Consent to Participants

Participants were given the below statement and were not tested without prior verbal consent:

“For a unit in statistics at QUT, we have been asked to choose a topic of interest to us and collect data to investigate it. We are interested in studying different typefaces and would be grateful for your assistance. These data are for educational purposes only and no individual will be able to be identified in any writing or discussion about the data.”