ADVANCES IN CAR DESIGN METHODS FOR DISTANCE OPTIMISATION: A TOY APPROACH

by

Sean Francis | Nicholas Gecks-Preston | Allan Huang

N11038276 | N10763104 | N10796118

Group 9’s Report

Submitted to Regression and Design 2023 Semester 1 (MXB242)

Queensland University of Technology

In Fulfilments of the Requirements

For Dr Leah South

May 28^{h}

Actual word count for analysis: 3050 (total 5969)

Includes: Abstract, Exploratory analysis, Results & Conclusion

Table of Contents

1.0 Abstract 1
2.0 Introduction 2
- **1.1 Importance of Study** 2
- **1.2 Aims and Variables of Study** 2
  - **1.2.1 Questions to be Answered** 2
- **1.3 Measurement of Variables** 3
  - **1.3.1 Additional weight (Categorical)** 3
  - **1.3.2 Slope of the Angle (Categorical)** 3
  - **1.3.3 Surface (Categorical)** 3
  - **1.3.4 Displacement (Continuous)** 3
- **1.4 Assumptions and Limitations** 4
  - **1.4.1 Assumptions** 4
  - **1.4.2 Limitations** 4

3.0 Literature Review 5
4.0 Methods 6
5.0 Exploratory Analysis 8
- **5.1 Reliability and Accuracy of the Data** 11

6.0 Results 12
- **6.1 Statistical Methods** 12
  - **6.1.2 Linear Regression** 12
  - **6.1.2 ANOVA (Analysis of Variance)** 13
- **6.2 The Regression Method** 14
  - **6.2.1 Initial Model** 14
  - **6.2.2 Log Transformed Model** 16
  - **6.2.2 Final Model: Nested Log Transformation** 17
  - **6.2.3 Model Metric Comparison** 19
- **6.3 ANOVA Analysis** 20
  - **6.3.3 Optimal Treatment** 22

7.0 Conclusion 23
8.0 Reference List 24
9.0 Appendices 25

1.0 Abstract

This report presents the investigative study conducted by ©Frequent Tests, commissioned by ©Matcha-Blocku to optimize their new line of matchbox vehicles. The study is motivated by the company's intention on capitalizing on the global toy industry. The three variables involved in this optimization research are weight, angle of the slope, and surface material. Investigating the variables was done through a factorial, complete, balanced, and randomised experimental design, where the order of trials was also randomised and replicated thrice.

The findings of the report match that of the literature review, where prior research indicates that weight has no bearing on total displacement, whilst surface material and slope angle do. The ideal combination we discovered involves no additional weight attached to the matchbox car, a slope angle of 30 degrees, with the slope covered by a layer of marble.

A new matchbox vehicle set developed based on these findings has the potential to disrupt and redefine the standards for toy vehicles, appealing to a wider audience by positioning ©Matcha-Blocku as a leader in the production of innovative and educational toys.

2.0 Introduction

1.1 Importance of Study

As of March 7th, 2023, the global toy industry is worth approximately 104.2 billion US dollars (Tighe, 2023). This means that any company that can successfully capitalise on this, can make a significant amount of revenue. This is precisely why the toy company ©Matcha-Blocku has hired our team of statisticians (©Frequent Tests). They intend to launch a new product line of match-block cars that come with an included ramp. Though they are not sure about the exact design, know they want their cars to travel as far as possible and function consistently between varying households. Thus, our company (©Frequent Tests) will investigate a range of variables that could affect the distance travelled by a matchbox car released from a ramp.

1.2 Aims and Variables of Study

Matchbox cars are toys that resemble real motor vehicles but are significantly reduced in size and lack the all-important 'motor' faculties. This deficiency leads to the movement being entirely affected by its design and the mechanical energy supplied.

This study aims to inform the toy manufacturer ©Matcha-Blocku about the ideal combination of ramp angle, weight, and surface material that would result in the optimal displacement of a matchbox car. Additionally, whether the matchbox car performs consistently across different surfaces.

1.2.1 Questions to be Answered

Below is a summary of the questions intended to be answered in this report:

1.) Does the weight of a vehicle have a significant effect on the distance travelled by a matchbox car?

2.) Does the angle of the slope have a significant effect on the distance travelled by a matchbox car?

3.) Does the type of surface the matchbox car is travelling on have a significant effect on the distance travelled by a matchbox car?

4.) What is the ideal combination of these variables in maximizing displacement?

1.3 Measurement of Variables

The following section will describe the variables being measured in this experiment. Each was chosen due to its reasonability to affect the displacement of the car.

1.3.1 Additional weight (Categorical)

The weight of the vehicle was to be adjusted incrementally by attaching variously sized weights to the matchbox car. This will be achieved by attaching otherwise identical objects that only differ in terms of weight, to minimize nuisance variables such as the shape and design of the toy – which would otherwise impact the aerodynamics of the vehicle albeit at a negligible level. Additionally, a precision scale (weight scale) will be used to determine the weight of each vehicle after adjustments, to ensure the weight will be increased consistently.

1.3.2 Slope of the Angle (Categorical)

The slope of the angle will be adjusted in increments of 15 degrees, from 30 to 60 degrees. (30, 45 & 60). Possible error in this variable could occur from the measurement using a protractor parallel to the table and thus it was decided that to minimize this error, a protractor and weight measurement method will be used instead. Under the assumption this is set up correctly, the effects of it will mitigate problems such as parallax error. A rudimentary diagram of this method can be found in appendix 1.

1.3.3 Surface (Categorical)

The two surfaces used in this experiment are concrete and wood, by which we hope to investigate the effects of surface friction and the matchbox car's wheel traction. The tests will be conducted in the same location in both instances to avoid any surface variations.

1.3.4 Displacement (Continuous)

To measure the displacement of the matchbox car, a tape measure will be used, recording the distance between the front end of the matchbox car to the base of the slope.

1.4 Assumptions and Limitations

The following section will outline the current assumptions and limitations of this experiment.

1.4.1 Assumptions

The following assumptions have been made in the context of this experiment:

- The matchbox car chosen is a reasonable representation of a typical matchbox car.
- External environmental factors will have a negligible impact on the results. (e.g., wind …)
- Measurements taken are accurate and precise.
- All the potential energy will be converted into kinetic energy with no loss. (Friction, wind resistance)
- The domain of the slope angle will be restricted to between zero and ninety degrees as an inclined plane must lie between these values to exist. (non-inclusive of zero and ninety)

1.4.2 Limitations

- Only two surfaces were tested, real households would have greater surface options.
- Only the variables slope angle, additional weight and surface material are considered.
- Due to time constraints, each treatment only had two replicates.
- Matchbox tyre tread is a nuisance variable that was not controlled for in this experiment.

3.0 Literature Review

As aforementioned, the experiment involves a matchbox car first rolling down a ramp and then traversing a flat surface. After the car has come to a stop, the total distance travelled will be measured for each treatment. Surface type, weight and slope angle will be manipulated to measure their effect on the distance travelled on a flat surface.

Three levels were chosen for slope angle: 30°, 45° and 60° respectively. As a continuous variable, equally spaced values were chosen for ease of calculation. Given the standardised slope length, varying the angle increases the height of the matchbox car, thus increasing gravitational potential energy (Lucas, 2014). This potential energy is converted to kinetic energy when the matchbox car is released and allowed to roll down the slope (Hamm, 2020). Thus, we predict that increasing the angle of the slope will increase the distance that matchbox car travels on the flat plane.

Friction is a force that acts against a body's motion. It occurs when two surfaces' grooves catch on each other when sliding past. This varies with the surface type. As such, two levels were chosen for the surface type: carpet and laminate. Given that carpet compresses by nature, its grooves will be deeper than laminate, and will slow the matchbox car at a higher rate than smooth laminate. Thus, it is expected that the matchbox car will travel further on the laminate surface than the carpet (Persson et al., 2003).

Three levels were chosen for weight: original matchbox car weight, plus 10 and 20 grams respectively. Increasing the mass of an object also increases its potential, and therefore, kinetic energy. Concurrently, increasing mass also increases its friction. Literature suggests that mass does not affect an object's acceleration down an inclined plane (Walding, 2019). Thus, this study seeks to inquire if varying its mass does indeed affect the matchbox car's distance.

Combining all these variables into a single formula using work and energy formulae (working in Appendix 2), yields:

$$d = \frac{l \sin(\theta)}{\mu_k}$$

Where d is the displacement travelled by the matchbox car along the flat surface, *l* is the length of the ramp in metres (m), theta (θ) is the angle of the slope in degrees ($^{\circ}$), and μ_k is the frictional coefficient of the surface (unitless).

4.0 Methods

The basic protocol for the experiment was as follows:

1. The order for each test was randomised using R Studio.
2. The ramp was placed on the desired surface of either carpet or laminate.
3. The ramp was then secured at the desired, slope measured angle using the protractor and weight method.
4. The guard rails for the ramp and surface were set up.
5. The desired weight was securely attached to the matchbox vehicle.
6. The matchbox vehicle was placed at the top of the ramp.
7. A blocker was positioned in front of the matchbox vehicle to prevent it from moving.
8. The matchbox vehicle was released from the top of the ramp.
9. The distance travelled by the matchbox vehicle was measured and recorded in an Excel table. (Bottom of the ramp to the front of the final matchbox vehicle position).
10. Steps 2 to 8 were repeated for each replicate (in the order given by R studio in step 1).

The predictor variables of this experiment are as follows:

1. The angle of the slope, measured in degrees (˚).
2. The weight of the matchbox car, measured in kilograms (kg).
3. The surface the matchbox car will traverse (categorical).

Hence, we have chosen to use a factorial, complete, balanced, and randomised experimental design to ensure quality statistical analysis.

The three levels of slope angles delegated are 30, 45, and 60 degrees. They were chosen because they are equally spaced and are common angles used in trigonometry, providing easy measurement in the context of this experiment. The weight of the matchbox vehicle will be the unadjusted original weight of the vehicle (approximately 3.5×10^{-3}kg), and two additional increments of 1.0×10^{-3}kg that will be achieved using weights. The 1.0×10^{-3}kg increment was chosen as it is easily obtained and acts as a significant linear increase to the original weight of the vehicle. Regarding surfaces, carpet and tabletop laminate were chosen as they represent two common surfaces a child will experiment with matchbox vehicles. In addition to this, they have drastically different surface friction that will aid in answering a question of the experiment (Frictional coefficients).

In the experiment, each treatment will be a combination of the different factor levels: angle of the slope, weight, and surface material. This will amount to 18 unique treatments that will be replicated twice for a total of 36 trials, for the sake of finding a representative mean distance for each treatment.

There are constraints with attaching the weights to the matchbox car, which may cause variation in the data. This is due to possible changes in aerodynamics and the additional weight from the attachment method (tape, glue, etc). Furthermore, due to time constraints, it may not be feasible to measure more than two replicates of each treatment, which will limit the statistical power of our experiment.

One major limitation caused by this experimental design is the possibility of the matchbox car hitting the guard rails; expending kinetic energy that may have instead led to a greater distance travelled. In addition to this, other variables could have been considered such as sound, heat, friction, and the internal workings of the matchbox vehicle.

5.0 Exploratory Analysis

This section aims to identify key insights and possible relationships found within out data set. A copy of this data can be found in appendix 3 and the code used for this report found in appendix 4.

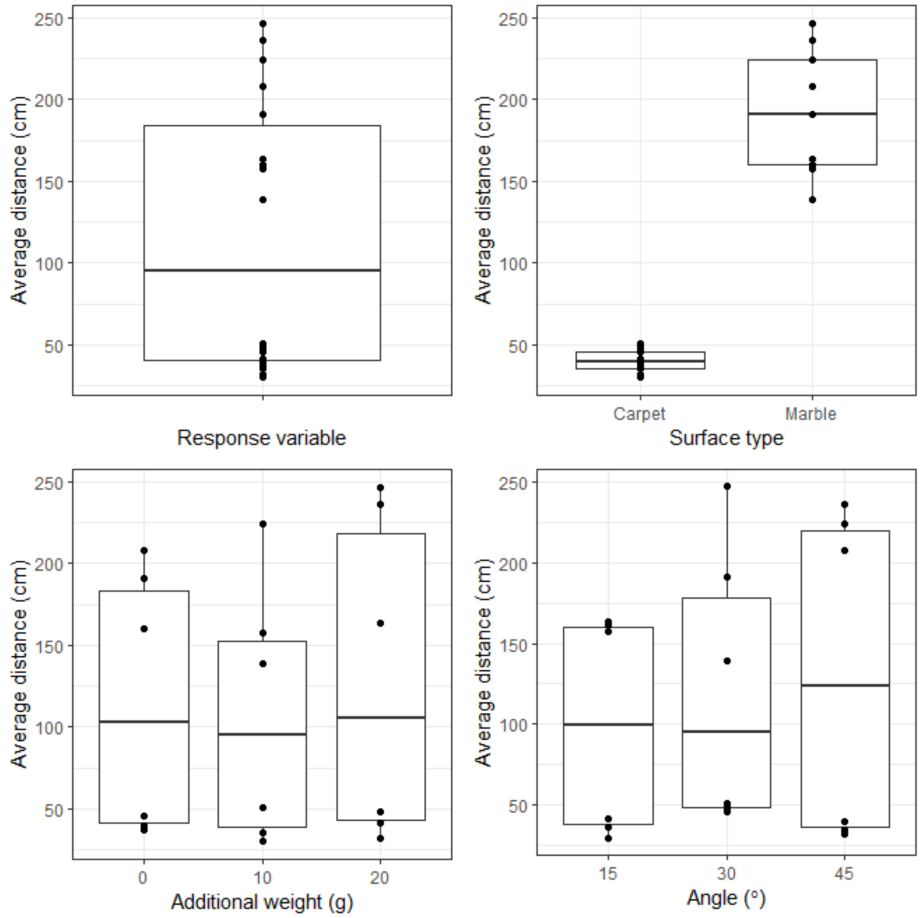


Fig. 1a: Distribution of response and explanatory variables

Within the distribution of the response variable, distance, plotted against the explanatory variable, there are two distinct groupings of points. This was determined to be caused by the drastic effect surface type had on the response material, as evidenced by the graph wherein the distance is now separated by surface type. Marble yields a much greater median distance than the results recorded for carpet and has a much greater interquartile range. Additional weight and angle appear to have little effect on displacement, though the highest levels of each have the greatest median distance travelled.

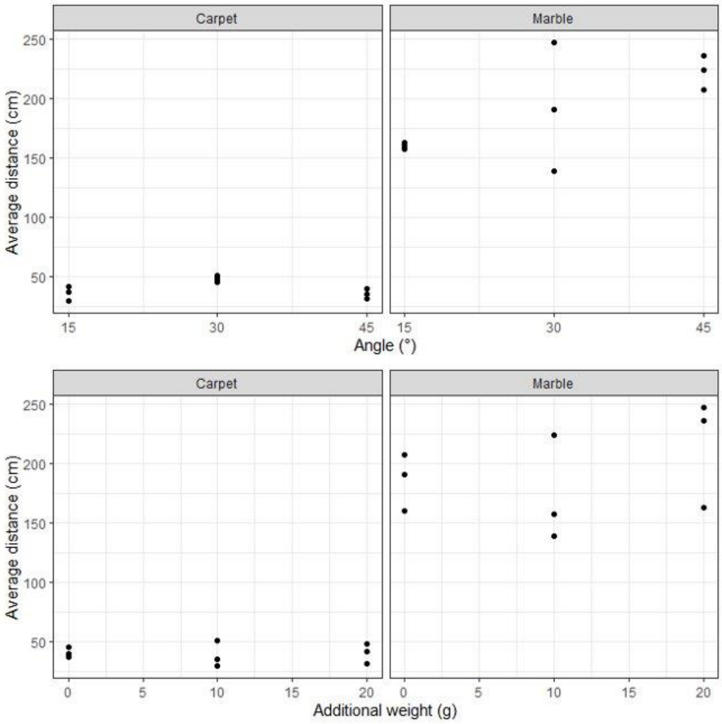


Fig. 1b: Displacement Separated by Surface Material

Upon separating the results based on the surface material, a subtle linear relationship is noticeable between the distance travelled by the vehicle and both the angle of the ramp and the additional weight of the vehicle. This suggests that the vehicle's distance is affected by these two factors. Conversely, results measured from when the experiment was completely on carpet indicate that there is no correlation between distance and either angle or additional weight. This would indicate the definite need to incorporate surface type into our model.

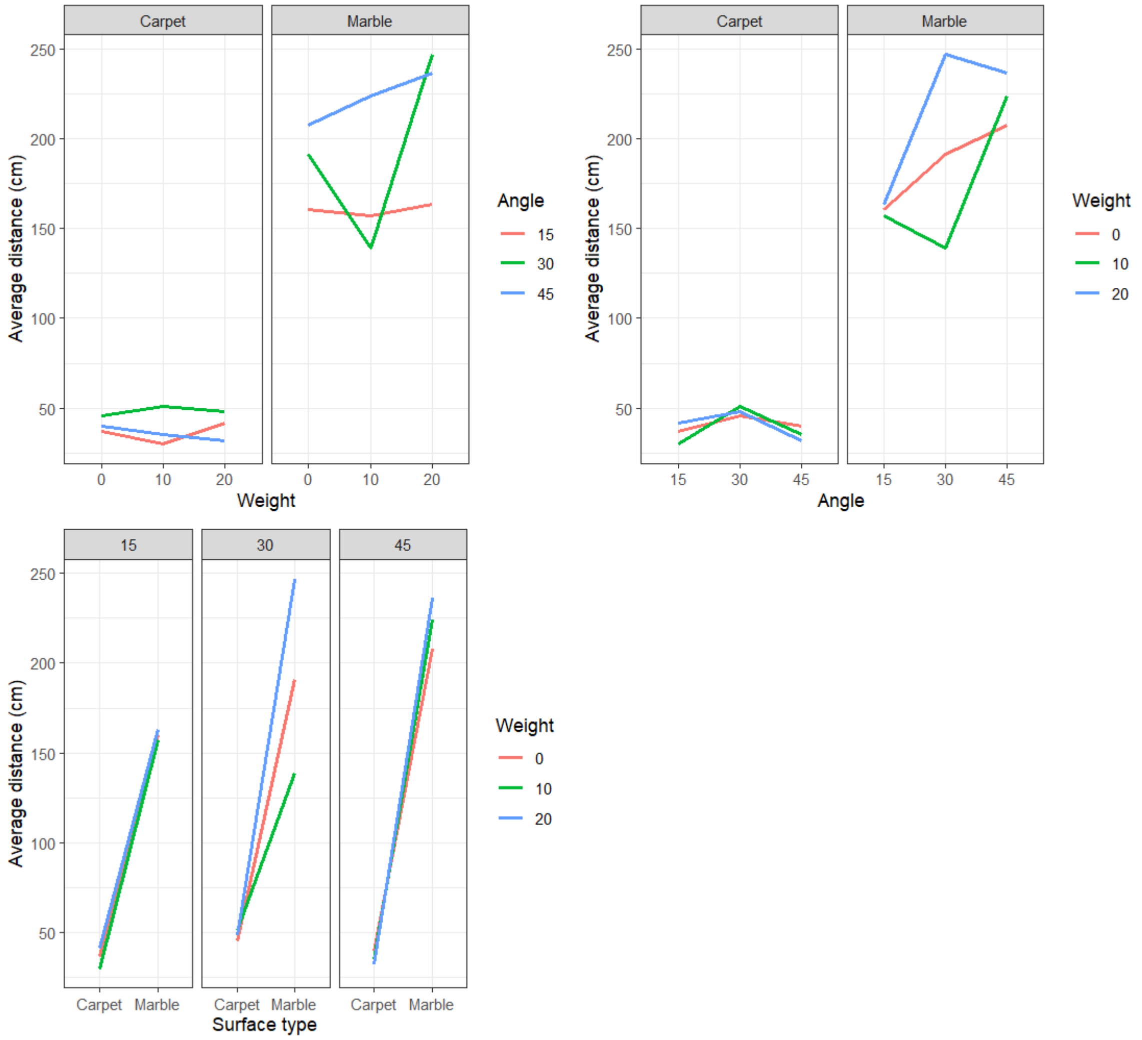


Fig. 1c: Interaction between variables

An advantage of a factorial design is its ability to analyse interaction between variables. In Fig. 1c, we can see that none of the Angle or Weight lines are parallel, most of which don't follow a consistent trend. The carpet facet of the Average distance vs Angle graph however, show a similar trend of peaking at 30^{o}, however all still crossover at some point. In contrast to this, the Surface type plot does show a trend, yet almost all facets either converge or diverge. It is difficult from this to tell whether interactivity exists. Collating information from all three plots suggests that there may be interaction between weight and angle, and possibly surface type and weight. A three-way interactivity may also exist between the variables, but it is difficult to tell graphically. This analysis confirms the need for interactions terms when building our model.

5.1 Reliability and Accuracy of the Data

The data was collected by strictly following the methodology outlined in section 4, resulting in a consistent dataset where errors were minimised. The results generally adhere to the predictions made by the literature review and thus are a representative sample. A standard IQR outlier test presented a few outliers as see in figure 2. This however seems unreliable to take the IQR test of the whole data since two distinct groupings can be seen. Thus, we decided to disregard the removal of outliers.

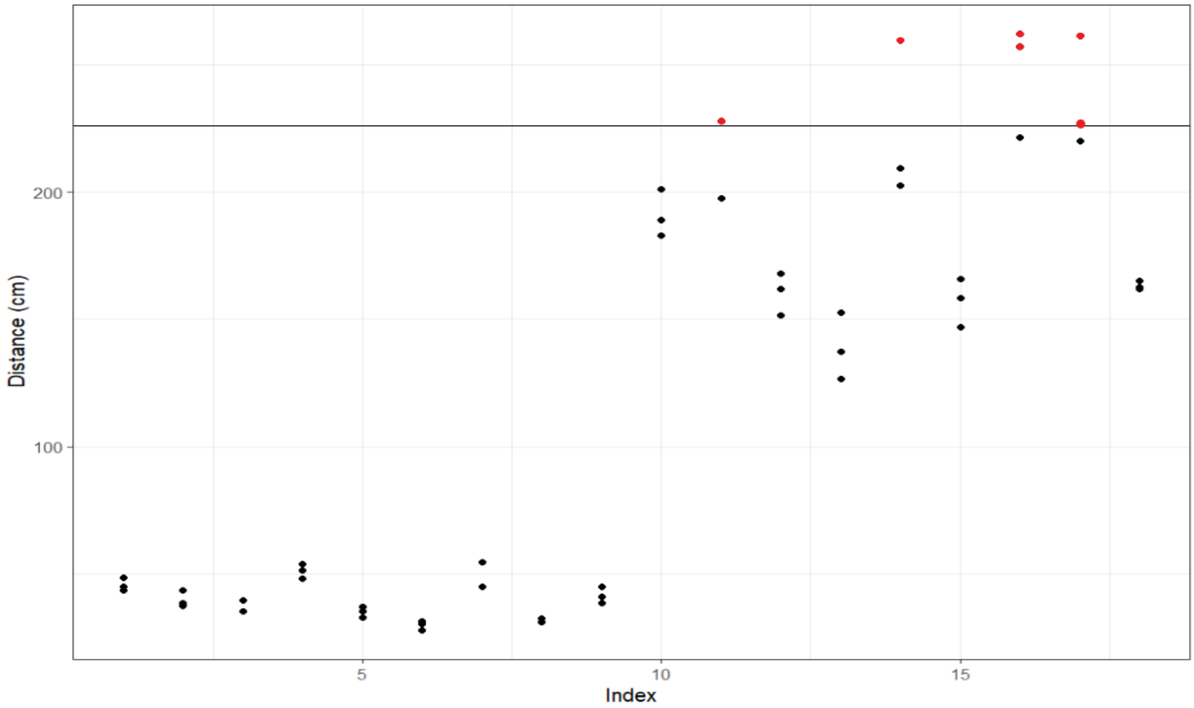


Fig. 2: IQR Test Identified Outliers

6.0 Results

6.1 Statistical Methods

We have chosen a significance (alpha) value of 0.05; we reject results with p values >0.05. This report will conduct two types of statistical analysis, a linear regression-based approach, and an ANOVA.

6.1.2 Linear Regression

Linear regression is a process in which the relationship between a response variable and explanatory variable/s are quantified. We investigate how the distance travelled by a toy car is affected by changing the surface type, weight, and angle of ramp that the car traverses. As we have used more than one explanatory variable, we will use multiple linear regression to quantify our results. The goal of this approach is to describe the response variable with a linear equation and estimate its parameters based on the data collected. We will use the least-squares method of parameter estimation, which aims to minimize the squared error of the model and data. A model based on this approach must adhere to some assumptions for the results to be considered reliable, these include:

1. Linearity: justification that there is indeed a linear relationship between the response and explanatory variables
2. Independence of residuals: That the previous residuals do not affect the result of the next residual.
3. Constant variance: the variance of the data is homoscedastic.
4. Normality: The Residuals follow a normal distribution.

A challenge of this method is the selection of covariates as this can drastically affect the fit and usefulness of the model. Therefore, a covariate selection technique known as stepwise regression was used. This is an algorithmic process that will systematically add or remove covariates (based on the direction) and use Akaike information criterion (AIC) values to determine whether this improved or not. AIC is a metric based on the log-likelihood ratio test which can be used compare models of differing parameter amounts.

6.1.2 ANOVA (Analysis of Variance)

Analysis of variance is a statistical tool used to determine whether the population mean of the response is the same across all levels our factor (South, 2023). The total sum of squares is the addition of the sum of square deviations between groups and the sum of square deviations within groups.

$$SST = SS_{\{between\}} + SS_{\{within\}}$$

Where SST is total sum of squares,
SS is sum of squares.

When incorporating the degrees of freedom for each (SS_A & SSE) a F-test statistic can be constructed, under the assumptions of:

- Independence of residuals
- Constant variance of residuals
- Normality of residuals

This means we can test the hypotheses:

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_n$$

$$H_a : \text{at least two differ.}$$

This result rationale can generalise to experiments with more than one factor by partitioning the total sum of squares (SST) between factors, possibly including affirming the presence of interactions between factors. Thus, allows us to statistically test for difference among a factor while accounting for additional factors. If in fact there not enough evidence to reject the null hypothesis, then no further statistical analysis (for this purpose) is needed. Conversely, if there is enough evidence to reject the null hypothesis further statistical analysis can be conducted in our case Tukey's HSD or Tukey's Honest Significant Difference. Tukey's HSD is a post-hoc statistical test which compares the mean response of every possible pair of treatments, including interactions between treatments. Though in this report Tukey's HSD is done directly after the ANOVA, it is not necessary as it corrects for family-wise error.

6.2 The Regression Method

6.2.1 Initial Model

Testing all the variations of stepwise regression including two-way interactions produced several models which were compared using Bayesian Information Criterion (BIC), the results were tabulated.

Model	*BIC*
Forward-One-Way	214.7385
Backward-One-Way	178.2938
Both-One-Way	178.2938
Both-Two-Way	180.2729

Table 1: BIC scores of Initial Stepwise Regression based models.

Backward-One-Way and Both-One-Way are both the lowest score as they are equal, this probably means that they are the same model, after inspection this was found to be true. Thus Both-One-Way was selected to be our initial model. The formula can be found below. Please note that the forward and backwards directions for two-way models were not included as forward often produced worst results than backward and backward usually produces the same result as both directions.

$$\text{Distance} \mid t_i, x_i = 22.37 + 151.65 \ \times S_1(t_i) + \ A_1(x_i) + \epsilon_i$$

Given

$$S_1(t) = \begin{cases} 0 : t = \text{Carpet} \\ 1 : t = \text{Marble} \end{cases} : t \in \{\text{Carpet, Marble}\} \ ,$$

$$A_1(x) = \begin{cases} 0 : \ x = 15 \\ 22.08 : \ x = 30 \\ 30.92 : \ x = 45 \end{cases} : \ x \in \{15, 30, 45\}$$

Upon analysis the only statistically significant covariate was surface type ($p = 6.29 \times 10^{-9}$) see appendix 5.

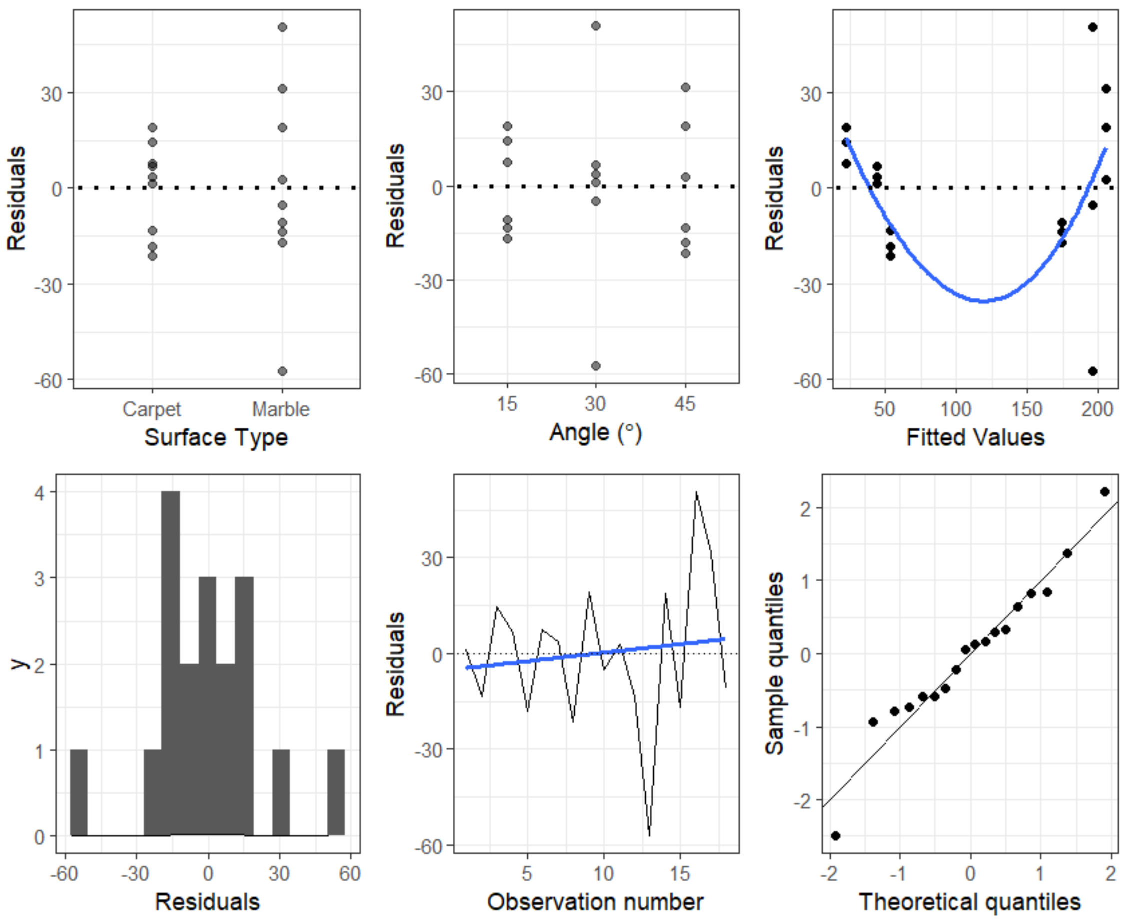


Fig. 3: Linearity assumptions verification for initial model

Testing the model assumptions of linearity and independence, normality and homoscedasticity of residuals was completed.

The Residuals vs Observation number plot shows the dependence between observations. Ideally, the line of best fit is y = 0, as this shows that the residuals have no trend, and therefore are independent of previous treatments. In this model, the line of best fit has a gradient much greater than 0, suggesting there is some dependence within the residuals. As such, the assumption of independent residuals has been violated.

The histogram shows that the residuals are somewhat normal, with a mean of approximately 1. However, the variance is far greater than 1 (as seen from the x-axis, where x has a range of 120), which makes the Gaussian overlay impossible to see. This normality is further seen in the Q-Q Plot, which shows from -0.5 onwards the residuals are normally distributed. Below this value, there is some deviance from the y = x line, indicating some positive skewness in the normal plot. While the residuals are normally distributed, their variance is not equal to 1, and as such, the assumption that the residuals are normally distributed with a mean = 0 and a variance = 1 has been violated.

In Fig. 3, the Residuals vs Surface type plot shows fanning, from Carpet to Marble, indicating non-constant variance. This is also seen in ‘Residuals vs Angle’ plot where the 30º level has much greater variance than the other levels. Furthermore, this is confirmed in the Residuals vs Fitted Values plot, which show a great fanning, accompanied by a loess line of best fit (which appears to be of polynomial or power form). As such, the assumption of homoscedasticity within the residuals has been violated. It can be seen above many violations of assumptions can be noted this motivated the use of a log-transformation of our response variable 'distance'.

6.2.2 Log Transformed Model

The process of a log-transformation is applying the natural log function to every data point of the distance variable.

$$[y_1, y_2, \dots, y_n] \rightarrow [\log y_1, \log y_2, \dots, \log y_n]$$

Since the model is now predicting transformed data, it is best practice to run stepwise regression again rather than just adding the log response term our initial model.

Model	*BIC*
Forward-One-Way (Log)	48.935380
Backward-One-Way (Log)	-1.228313
Both-One-Way (Log)	-1.228313
Both-Two-Way (Log)	-3.809952

Table 2: BIC scores of Log-Transformed Stepwise Regression based models.

Both-One-Way was selected to be our log-transformed model as it has the lowest BIC score. The formula can be found below.

$$\log Y_i | x_i, t_i, z_i = 3.584710 + 1.499070 S_1(x_i) + A_2(t_i) + W_1(z) + G(x_i, t_1) + \epsilon_i$$

Given

$$S_1(t) = \begin{cases} 0: & \mathrm{t = Carpet} \\ 1: & \mathrm{t = Marble} \end{cases} : t \in \{\text{Carpet, Marble}\},$$

$$A_2(x) = \begin{cases} 0: & x = 15 \\ 0.299376: & x = 30 \quad x \in \{15, 30, 45\} \\ -0.007551: & x = 45 \end{cases}$$

$$W_1(z) = \begin{cases} 0: & x = 0 \\ -0.080147: & x = 10 \; x \in \{0,10,20\} \\ 0.059220: & x = 20 \end{cases}$$

$$A_3(x) = \begin{cases} 0: & x = 15 \\ -0.144773: & x = 30 \; x \in \{15, 30, 45\} \\ 0.334543: & x = 45 \end{cases}$$

$$G(x, t) = A_3(x) \times S_1(t)$$

If predicting the non-log response this would be the approximation.

$$y_i | x_i, t_i, z_i \approx e^{3.584710 + 1.499070 S_1(x_i) + A_2(t_i) + W_1(z) + G(x_i, t_1)}$$

Though if you refer to appendix 6 the summary for this model indicated that none of the levels of weight were in fact statistically significant, this lead to a partial-f test being conducted with a nested model that removed weight. The results from this test concluded that the extra parameter was not needed. The test can be found in appendix 7. Thus, our final model can be seen below.

6.2.2 Final Model: Nested Log Transformation

$$\log Y_i | x_i, t_i, z_i = 3.577735 + 1.499070 S_1(x_i) + A_4(t_i) + G(x_i, t_1) + \epsilon_i$$

Given

$$S_1(t) = \begin{cases} 0 : \ \mathrm{t} = \mathrm{Carpet} \\ 1 : \ \mathrm{t} = \mathrm{Marble} \end{cases} : t \in \{\mathrm{Carpet,\ Marble}\},$$

$$A_4(x) = \begin{cases} 0 : \ x = 15 \\ 0.299376 : \ x = 30 \\ -0.007551 : \ x = 45 \end{cases} x \in \{15, 30, 45\}$$

$$A_5(x) = \begin{cases} 0 : \ x = 15 \\ -0.144773 : \ x = 30 \\ 0.334543 : \ x = 45 \end{cases} x \in \{15, 30, 45\}$$

$$G(x, t) = A_5(x) \times S_1(t)$$

If predicting the non-log response this would be the approximation.

$$y_i | x_i, t_i, z_i \approx e^{3.577735 + 1.499070 S_1(x_i) + A_4(t_i) + G(x_i, t_1)}$$

The summary for this model can be seen in appendix 8.

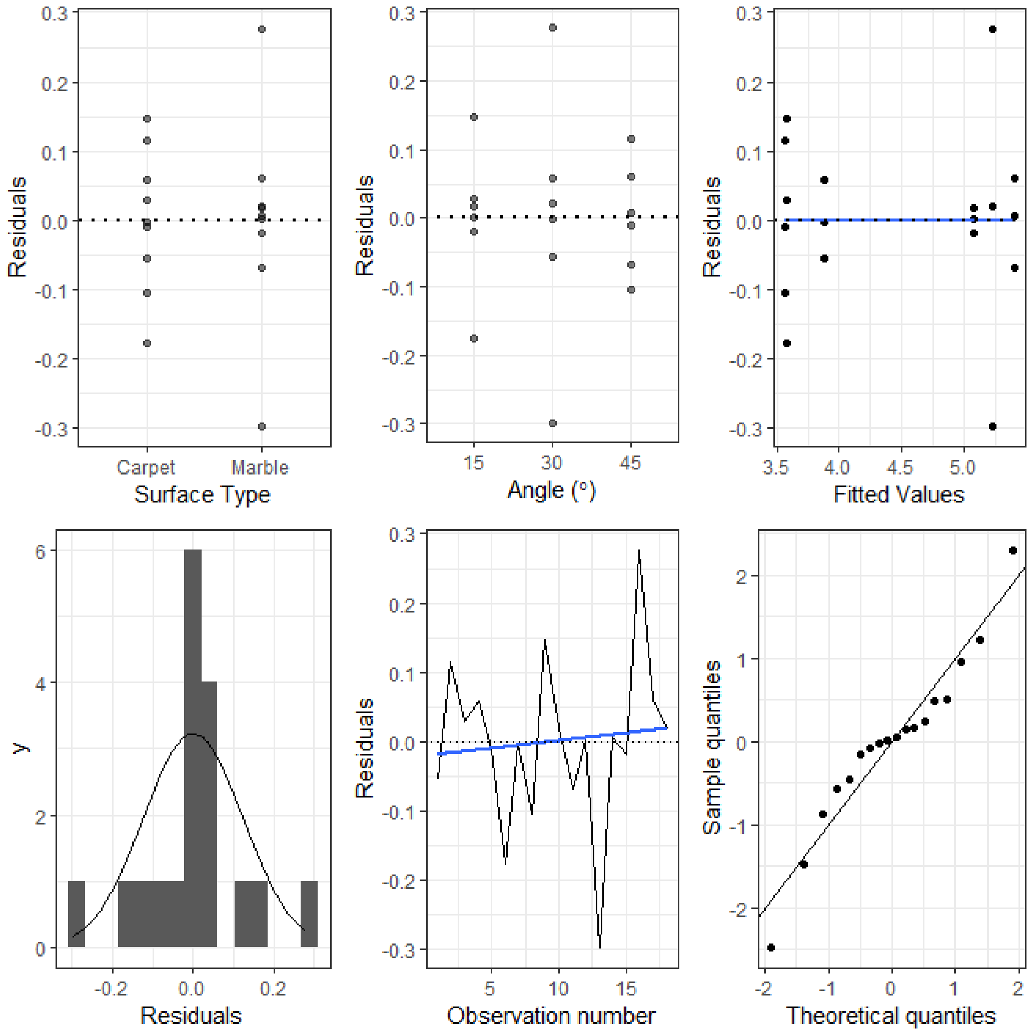


Fig. 4: Linearity assumptions verification for final model

Surface_type: Angle. As can be seen in the top row of graphs in Fig. 4, fanning is still present in the residuals, showing heteroscedasticity. As such, the assumption of homoscedasticity has been violated.

In the Residuals vs Observation number plot, while the line of best fit does have a non-zero gradient, it is greatly due to the residual value decreasing heavily. Further investigations and/or data are required to confirm the residuals' independence.

The histogram in Fig. 4 has an approximate mean of 0, and a much lower variance than 1. Looking at the Q-Q plot to confirm normality, there is a deviance from the y = x line at around x = -0.5, with some of the points lying on the line at x = 0. This graph suggests that the residuals aren't quite normally distributed. As such, the assumption of normality has been violated.

6.2.3 Model Metric Comparison

models <chr>	r.squared <dbl>	adj.r.squared <dbl>	sigma <dbl>	statistic <dbl>	p.value <dbl>	df <dbl>	logLik <dbl>	AIC <dbl>	BIC <dbl>
Inital: Step One Way	0.9184415	0.9009647	25.9944357	52.55198	7.268099e-08	3	-81.92095	173.84190	178.293756
Intermediate: Step Two Way (log)	0.9826529	0.9705099	0.1417812	80.92359	5.330484e-08	7	14.91165	-11.82330	-3.809952
Final : Nested (log)	0.9775867	0.9682478	0.1471185	104.67930	1.805610e-09	5	12.60559	-11.21118	-4.978579

Table 3: A table of the model metrics for all major models

In addition to improving the assumptions a log model has also improved many metrics as seen in table 3. Notably the log models have better adjusted r-squared value. The Intermediate model does have a higher r-squared value than the final model. This is completely fine as from the partial-f test conducted earlier in the report believe this is an overfitted model. The BIC which accounts penalises models with more variables agrees with our conclusion as the BIC is lower for the Final model. One notable of the log- transformation drawback is now predictions made by the model, which are not the log of the response are only approximations. This now cements our final model is ready for an ANOVA.

6.3 ANOVA Analysis

As previously mentioned, an ANOVA is a statistical tool used to in determining whether the population mean of the response is the same across all levels of each factor. This is done using R on our final model. A summary table of the results can be found below. The raw summary can also be seen in appendix 9.

Factor	*P-Value*	*Significant (Yes or No)*
Surface Type	3.46×10^{-11}	Yes
Angle	0.05347	No
Surface Type: Angle (Interaction)	0.0417	Yes

Table 4: A summary of the ANOVA results.

It can be seen in the table above that both surface type and surface type: angle (interaction) are significant. Therefore, we have strong evidence to suggest that the population mean distance is not the same for a marble and carpet surface types. In addition to this we have strong evidence to suggest an interaction. The effect of the surface type depends on the angle used. (Visa versa) It can be seen that angle is in fact almost significant ($\alpha = 0.05$), an argument could be made for including this as significant but, since we are faithful followers of Jacob Bernoulli, we must absolve from such a notion as we preselected a significance level of 0.05. Since we have significant evidence of a difference between mean response of some factors, we can use Tukey's HSD to give us a more granular understanding of the differences between groups. Undergoing Tukey's HSD, a summary table can be found below, and the original can be found in appendix 10.

Groups being compared	P-value
Marble - Carpet	0
30 - 15	0.0496042
Marble:15 - Carpet:15	0.0000004
Carpet:30 - Carpet:15	0.0000001
Marble:45 - Carpet:15	0.0000001
Carpet:30 - Marble:15	0.0000043
Carpet:45 - Marble:15	0.0000004
Marble:30 - Carpet:30	0.0000011
Marble:45 - Carpet:30	0.0000003
Carpet:45 - Marble:30	0.0000001
Marble:45 - Carpet:45	0.00000001

Table 5: A summary of significant outputs from Tukey's HSD

Table 4: A Table of significant outputs from Tukey's HSD

The table above reveals the mean of the response is significant between surface types and several different interaction terms. Terms of importance can be seen in yellow; these indicate that when angle is kept constant between surface types, the mean response is honestly significantly different. This evidence will help inform optimal treatment.

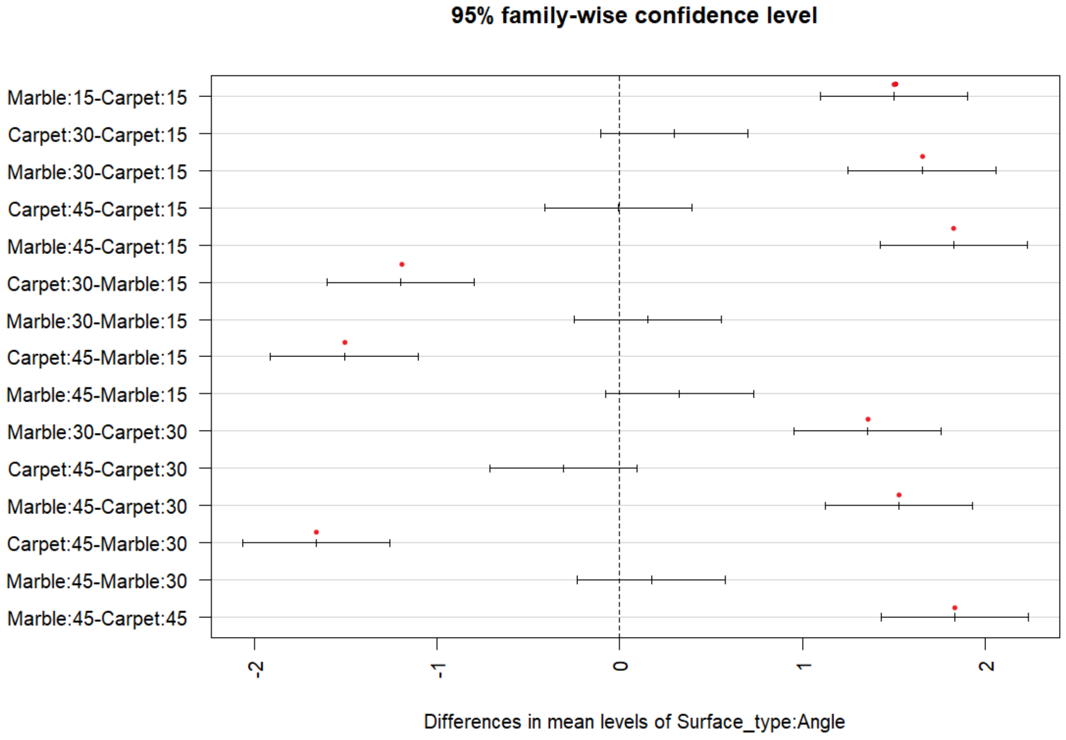


Fig. 5: Tukey's HSD Plot for Interactions. (Red dots are significant)

Looking at Figure. 5 (and the table in Appendix 10), Tukey's HSD shows that for both surfaces changing angles from 15^{o} to 30^{o} yields a statistically significant difference (as seen in the p value of $0.496 < 0.05$). However, going to 45^{o} from either 15^{o} or 30^{o} does not yield a statistically significant difference (as the p values are >0.05). This is further seen in the Surface_type:Angle section, where every instance of 45^{o} (keeping the surface type constant) yields a p value of > 0.05. As such, the data suggests that there is an increase in average distance going from 15^{o} to 30^{o}, however, further experiments are required to confirm if there is a statistically significant difference going to 45^{o}.

6.3.3 Optimal Treatment

In this report we were requested by ©Matcha-Blocku to find the optimal combination of ramp angle, added weight and surface type to optimise for distance in their new line of match block car. It was found in our model that weight played no significant effect and thus not adding extra weight would be our recommendation to save on manufacturing costs. With specifications on ramp angle and surface type, if you examine the plot below.

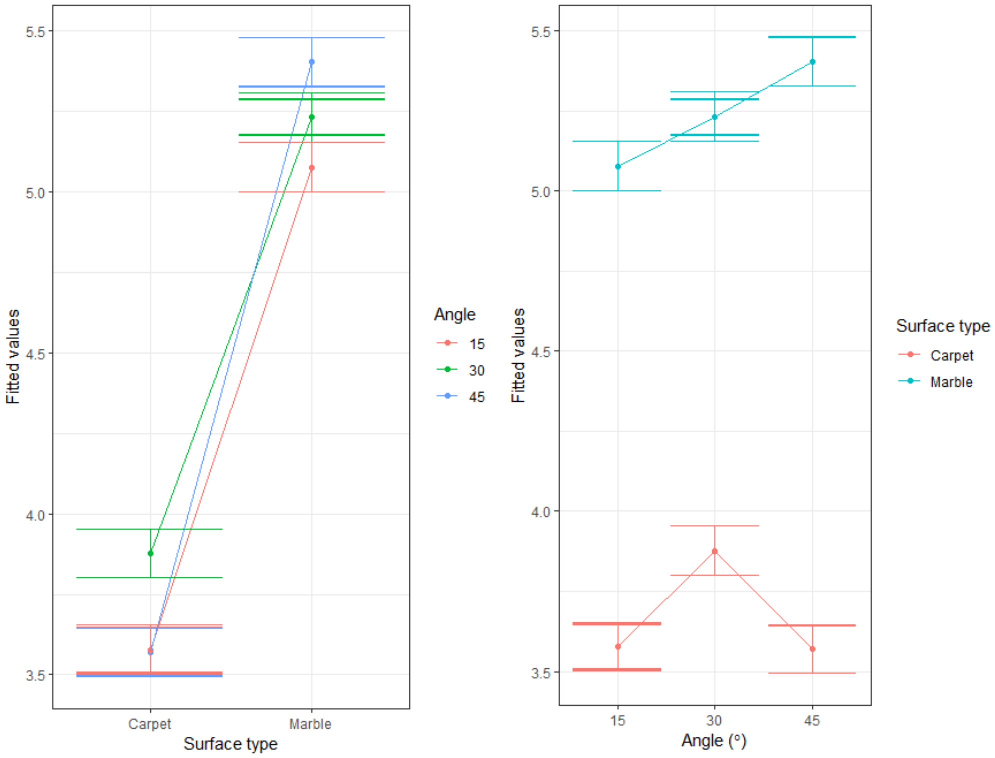


Fig. 6: Final model visualisation

That marble outperforms carpet and thus is recommended by us to be the surface type suggested in marketing material. In addition to this Tukey's HSD indicates 30 degrees as the safest choice and thus will be our recommendation for ramp angle. A table has been made to sum up these recommendations.

Factor	*Recommendation*	*Reason*
Surface type	Marble	ANOVA + Tukey HSD
Additional weight	0	Cost + Model Choice
Angle	30°	Tukey's HSD

Table 6: Summary of product recommendations

7.0 Conclusion

In conclusion the purpose of this report was to find the optimal combination of ramp angle, added weight and surface type that allows a matchbox car to travel its furthest distance. This information will help inform ©Matcha-Blocku on its new line of matchbox car and ramp. To do this a factorial, complete, balanced, and randomised experiment was conducted to gather data on this topic. This data was then statistically modelled using linear regression and further investigated with ANOVA and Tukey's HSD. Results from this motivated our recommendations that zero grams of additional weight should be added to the matchbox car, and an angle of 30 degrees should be used for the ramp, However, there was an indication that 45 degrees could be a better pick but for our sample, there was not enough support to be sure. Finally, the surface type of marble should be the suggested play surface in advertising material.

8.0 Reference List

Hamm, K. (2020). *Biomechanics of Human Movement*. Pressbooks. Retrieved March 27, 2023.

Lucas, J. (2014, June 12). What is kinetic energy? Retrieved March 28, 2023, from https://www.livescience.com/46278-kinetic-energy.html

Persson, B., Albohr, O., Mancosu, F., Peveri, V., Samoilov, V., & Sivebaek, I. (2003). On the nature of the static friction, kinetic friction, and creep. *Wear, 254*(9), 835-851. doi:10.1016/s0043-1648(03)00234-5

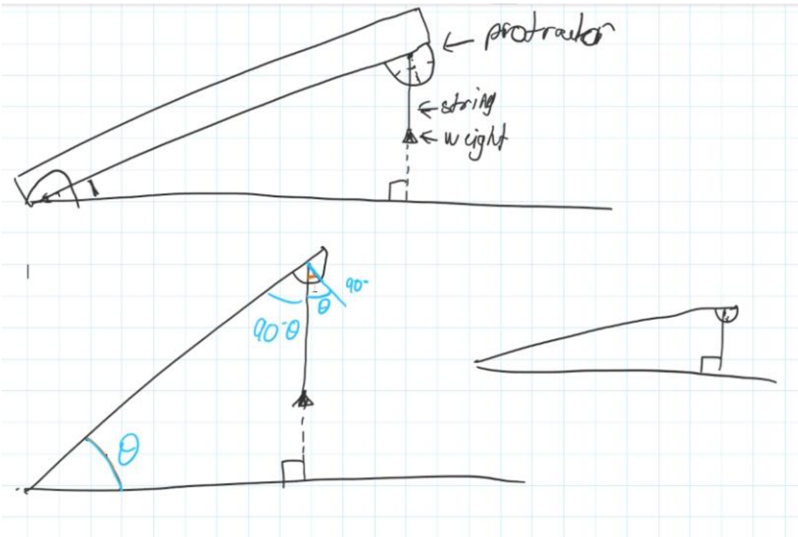
South, L. (2023). MXB242 Lecture Content. Retrieved May 24, 2023, from https://www.qut.edu.au/study/unit?unitCode=MXB242

Tighe, D. (2023, March 7). Toy industry - statistics & facts. Retrieved March 27, 2023, from https://www.statista.com/topics/1108/toy-industry/#topicOverview

Walding, R. (2019). New Century Physics for Queensland. Docklands, Victoria: Oxford University Press. Retrieved March 26, 2023

9.0 Appendices

Appendix 1: (Protractor and weight measuring method)



Appendix 2: (Calculations for the estimated car distance)

At the top of the ramp, the car has potential energy (PE) in Joules (J):

$$\mathrm{PE} = \mathrm{mgh} = mg \times l \times \sin(\theta)$$

Where m is mass (kg), g is acceleration due to gravity (9.8 m/s^2), and *l* is length of ramp (m).

The matchbox car is then released; once it has reached the base of the ramp, all of its potential energy has been converted into kinetic energy. The change in potential energy is the work done by gravity. We can calculate work:

$$W = F \times d$$

Where W is work in joules (J), F is force in newtons (N) and d is displacement of car on flat surface in metres (m).

Work done by gravity is equal to the kinetic energy, thus the work done by friction can be equated to PE using work formula:

$$mg \times l \times \sin(\theta) = F_{friction} \times d$$

Where **d** is displacement in metres (m), and:

$$F_{friction} = \mu_k \times mg$$

Where μ_k is frictional coefficient (unitless).

Combining this gives:

$$mg \times l\sin(\theta) = \mu_k \times mg \times d$$

Mass and gravity cancels:

$$l\sin(\theta) = \mu_k d$$

Thus, distance can be given by:

$$d = \frac{l\sin(\theta)}{\mu_k}$$

Appendix 3: (Raw Data)

Treatment	Surface_type	Weight	Angle	Distance_1	Distance_2	Distance_3	Distance_avg
1	Carpet	0	30	48.5	45	43.5	45.67
2	Carpet	0	45	37.5	38.5	43.5	39.83
3	Carpet	0	15	35.5	39.5	35.5	36.83
4	Carpet	10	30	51.5	54	48	51.17
5	Carpet	10	45	35.5	33	37	35.17
6	Carpet	10	15	28	30.5	31.5	30
7	Carpet	20	30	45	54.5	45	48.17
8	Carpet	20	45	32.5	31	32.5	32
9	Carpet	20	15	41	45	38.5	41.5
10	Marble	0	30	201	183	189	191
11	Marble	0	45	197.5	197.5	228	207.67
12	Marble	0	15	162	168	151.5	160.5
13	Marble	10	30	137.5	152.5	126.5	138.83
14	Marble	10	45	209.5	202.5	259.5	223.83
15	Marble	10	15	158.5	166	147	157.17
16	Marble	20	30	257	221.5	262	246.83
17	Marble	20	45	220	261.5	227	236.17
18	Marble	20	15	165	162	162.5	163.17

Appendix 4: (Raw Code)

https://github.com/ngecksp/ToyCarDesign.git

Appendix 5: (Initial model summary)

```
Call:
lm(formula = Distance_avg ~ Surface_type + Angle, data = df)

Residuals:
    Min      1Q  Median      3Q     Max
-57.272 -13.504   1.975  12.751  50.728

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)           22.37      12.25   1.826   0.0893
Surface_typeMarble   151.65      12.25  12.375 6.29e-09
Angle30               22.08      15.01   1.471   0.1633
Angle45               30.92      15.01   2.060   0.0585

(Intercept)        .
Surface_typeMarble ***
Angle30
Angle45            .
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 25.99 on 14 degrees of freedom
Multiple R-squared:  0.9184,    Adjusted R-squared:  0.901
F-statistic: 52.55 on 3 and 14 DF,  p-value: 7.268e-08
```

Appendix 6: (Log-transformed model summary)

```
Call:
lm(formula = log(Distance_avg) ~ Surface_type + Weight + Angle +
    Surface_type:Angle, data = df)

Residuals:
      Min        1Q    Median        3Q       Max
-0.224986 -0.067088  0.004225  0.075992  0.211097

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)               3.584710   0.094521  37.925 3.87e-12 ***
Surface_typeMarble        1.499070   0.115764  12.949 1.42e-07 ***
Weight10                 -0.080147   0.081857  -0.979   0.3506
Weight20                  0.059220   0.081857   0.723   0.4860
Angle30                   0.299376   0.115764   2.586   0.0271 *
Angle45                  -0.007551   0.115764  -0.065   0.9493
Surface_typeMarble:Angle30 -0.144773  0.163715  -0.884   0.3973
Surface_typeMarble:Angle45  0.334543  0.163715   2.043   0.0683 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1418 on 10 degrees of freedom
Multiple R-squared:  0.9827,    Adjusted R-squared:  0.9705
F-statistic: 80.92 on 7 and 10 DF,  p-value: 5.33e-08
```

Appendix 7: (Partial F test)

```
Analysis of Variance Table

Model 1: log(Distance_avg) ~ Surface_type + Angle + Surface_type:Angle
Model 2: log(Distance_avg) ~ Surface_type + Weight + Angle + Surface_type:Angle
  Res.Df     RSS Df Sum of Sq      F
1     12 0.25973
2     10 0.20102  2  0.058707 1.4602
  Pr(>F)
1
2 0.2777
```

Appendix 8: (Log nested model – final model)

```
Call:
lm(formula = log(Distance_avg) ~ Surface_type + Angle + Surface_type:Angle,
    data = df)

Residuals:
     Min       1Q   Median       3Q      Max
-0.29816 -0.04662  0.00429  0.05068  0.27729

Coefficients:
                           Estimate Std. Error t value Pr(>|t|)
(Intercept)                3.577735   0.084939  42.121 2.08e-14 ***
Surface_typeMarble         1.499070   0.120122  12.480 3.12e-08 ***
Angle30                    0.299376   0.120122   2.492   0.0283 *
Angle45                   -0.007551   0.120122  -0.063   0.9509
Surface_typeMarble:Angle30 -0.144773   0.169878  -0.852   0.4108
Surface_typeMarble:Angle45  0.334543   0.169878   1.969   0.0724 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1471 on 12 degrees of freedom
Multiple R-squared:  0.9776,	Adjusted R-squared:  0.9682
F-statistic: 104.7 on 5 and 12 DF,  p-value: 1.806e-09
```

Appendix 9: (ANOVA summary of final model)

```
Analysis of Variance Table

Response: log(Distance_avg)
                   Df  Sum Sq Mean Sq  F value    Pr(>F)
Surface_type        1 10.9839 10.9839 507.4830 3.462e-11 ***
Angle               2  0.1631  0.0816   3.7683   0.05370 .
Surface_type:Angle  2  0.1813  0.0907   4.1885   0.04171 *
Residuals          12  0.2597  0.0216
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
>
```

Appendix 10: (Tukey's HSD Output of the ANOVA of the final model)

```
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = nested.log, data = df)

$Surface_type
                  diff      lwr      upr p adj
Marble-Carpet 1.562327 1.411221 1.713433     0

$Angle
             diff           lwr       upr     p adj
30-15  0.22698947  0.0003839722 0.4535950 0.0496042
45-15  0.15972011 -0.0668853869 0.3863256 0.1865975
45-30 -0.06726936 -0.2938748529 0.1593361 0.7148133

$`Surface_type:Angle`
                            diff         lwr         upr     p adj
Marble:15-Carpet:15  1.499070342  1.09559075  1.90254994 0.0000004
Carpet:30-Carpet:15  0.299375977 -0.10410362  0.70285557 0.2007767
Marble:30-Carpet:15  1.653673297  1.25019370  2.05715289 0.0000001
Carpet:45-Carpet:15 -0.007551264 -0.41103086  0.39592833 0.9999998
Marble:45-Carpet:15  1.826061821  1.42258223  2.22954141 0.0000000
Carpet:30-Marble:15 -1.199694365 -1.60317396 -0.79621477 0.0000043
Marble:30-Marble:15  0.154602955 -0.24887664  0.55808255 0.7865086
Carpet:45-Marble:15 -1.506621607 -1.91010120 -1.10314201 0.0000004
Marble:45-Marble:15  0.326991478 -0.07648812  0.73047107 0.1412311
Marble:30-Carpet:30  1.354297320  0.95081773  1.75777691 0.0000011
Carpet:45-Carpet:30 -0.306927241 -0.71040684  0.09655235 0.1826830
Marble:45-Carpet:30  1.526685844  1.12320625  1.93016544 0.0000003
Carpet:45-Marble:30 -1.661224562 -2.06470416 -1.25774497 0.0000001
Marble:45-Marble:30  0.172388523 -0.23109107  0.57586812 0.7073175
Marble:45-Carpet:45  1.833613085  1.43013349  2.23709268 0.0000000
```

Appendix 11: (Randomised Order of treatments)

The rest can be found in the submission folder.

	Surface_Type	Angle	Weight	Run Order
1	Carpet	15	0	2
2	Marble	15	0	16
3	Carpet	30	0	10
4	Marble	30	0	3
5	Carpet	45	0	24
6	Marble	45	0	30
7	Carpet	15	10	51
8	Marble	15	10	6
9	Carpet	30	10	21
10	Marble	30	10	39
11	Carpet	45	10	49
12	Marble	45	10	45
13	Carpet	15	20	32
14	Marble	15	20	17
15	Carpet	30	20	19
16	Marble	30	20	1
17	Carpet	45	20	37
18	Marble	45	20	50
19	Carpet	15	0	22
20	Marble	15	0	5
21	Carpet	30	0	4
22	Marble	30	0	13
23	Carpet	45	0	40
24	Marble	45	0	23
25	Carpet	15	10	35
26	Marble	15	10	48
27	Carpet	30	10	28
28	Marble	30	10	47
29	Carpet	45	10	8
30	Marble	45	10	11

Appendix 12: (Run order script)

```r
# Creating three factors (or blocking factors)
# and listing their levels
Factor1 ← c("Carpet","Marble")
Factor2 ← c("15","30","45")
Factor3 ← c("0","10","20")

# expand.grid is an easy way to get all combinations of factor levels
Design ← expand.grid(Factor1,Factor2,Factor3)
names(Design) ← c("Surface_Type","Angle","A.Weight")

# We can create replicates by using rbind
Design2 ← rbind(Design,Design)
Design3 ← rbind(Design2,Design)

# Now we need to decide how/when to do each of these treatments.

# Typically we'll want to randomise the run order, which can
# be done like this:
Design3$RunOrder ← sample(1:nrow(Design3),replace=FALSE)

# You can randomise other things by duplicating and adjusting
# the line above.

# The line below shows a preview of the design
head(Design3)

```

Appendix 13: (Planning checklist)

MXB242: Regression and Design

Planning Checklist

Part 1

Please fill in all entries in the following tables and tick boxes to state agreement. It is important as part of your future careers that you take project planning seriously – especially sections relating to ethics (if applicable).

1. We have consulted the unit coordinator and teaching team to discuss this project. ☒

2. We have all completed the "Respect and Safety at QUT" training module. ☒

3. This project will not have any physical, emotional or psychological effects beyond those associated with everyday living. ☒

4. We commit to reporting any changes to our plan after any pilot study or at any subsequent stage. ☒

5. This project will not involve issues that are offensive or sensitive. ☒

6. If this project involves peoples' cooperation or assistance, this cooperation will be requested in person. ☒

7. If this project involves a survey or a non-intrusive experiment/study involving peoples' cooperation or assistance, the following statement will be provided to each participant: ☒

   "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 .... 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."

Part 2

It is important that each member of the group reviews this document before submission to verify that they have understood the plan before data collection and assessment commences.

Signed by each member of the group:

1.	Nicholas Gecks-Pearson	N. Gecks
	(Name)	*(Signature)*
2.	Sean Francis	S. Francis
	(Name)	*(Signature)*
3.	Allan Huang	A. H
	(Name)	*(Signature)*
4.		
	(Name)	*(Signature)*

If you have four group members, describe the planned additional work in consideration of the fourth member.

N. A.

Appendix 14: (Work distribution table)

Work Distribution Table	
Nick	Introduction + Formatting
Allan	Methods + Referencing
Sean	Literature Review + Appendix

This is a rough plan of what sections were focused on by each person, but everyone individually contributed, reviewed and worked on each section.

Appendix 15: (Meeting summaries)

Group Meetings Summary	
Friday 3rd of March	- Team was formed and skillsets were discussed.
Friday 10th of March	- Group communication method was made. - Typesetting application was decided. (Word)
Friday 17th of March	- Finalised first experiment idea. - First experiment idea was rejected. - Brainstormed second experiment idea. - Second experiment idea was sent off to Leah.
Monday 20th of March	- Second experiment was given the green light from Leah.
Friday 24th of March	- The work distribution table was made (see appendix 5.)
Saturday 25th of March	- Report formatting and Title Page was done (Nick) - Research articles were found for the literature review (Sean) - Brainstormed methods and completed planning checklist (Allan)

Wednesday 29th of March	- Literature review draft was done (Sean) still needs to be checked by other group members. - Methods draft was done (Allan) still needs to be checked by other group members. - A diagram of the protractor and weight method (Sean) - Introduction draft was done (Nick) still needs to be checked by other group members.
Sunday 2nd of April	- Assumptions and limitation were tweaked + added finished pieces to appendix (Allan) - Calculations for the appendix were done (Sean) - Rewrote motivations of study (Nick)
Tuesday 4th of April	- Bibliographies and Intext referencing (Allan) - Protocol in methods section was edited + Fixed up grammar in Lit Review (Nick) - Finalised methods section (Sean)
Wednesday 5th of April	- Report was submitted (Allan & Nick) - Final Edits and tweaks to all parts of document (Sean + Nick + Allan) this included grammar and proofreads. - Checked document with exemplar (Sean)
Friday 28th of April	- Meeting to discuss required apparatus for data collection (Sean + Nick + Allan) - Edits to Data collection plan (Sean + Nick + Allan)

Tuesday 2nd of May	- Conducted data collection plan (Sean + Nick + Allan) - Results recorded and shared (Sean)
Friday 5th of May	- Meeting regarding presentation slides (Sean + Nick + Allan) - Initial template for power point created and shared with group members (Allan) - Shared R Studio file created and shared with group members (Nick)
Monday 8th of May	- Exploratory Analysis of data conducted (Sean + Nick) - Meeting to discuss enquiries for our tutor during practical class this week to prepare for presentation (Sean + Nick + Allan)
Wednesday 10th of May	- Data analysis sprint session (Sean + Nick + Allan) - Required graphs for presentation discussed and finalised (Sean + Nick + Allan)
Friday 12th of May	- Power point finalised using required graphs (Allan) - Power point checked with group members to be finalised (Sean + Nick + Allan)
Monday 15th of May	- Meeting to discuss slide distribution & topics to each cover (Sean + Nick + Allan) - Begin practice for presentation (Sean + Nick + Allan)
Tuesday 16th of May	- Practice run of presentation over voice call (Sean + Nick + Allan)
Wednesday 17th of May	- Practice run in person 5 hours prior to presentation (Sean + Nick + Allan)

Friday 19th of May	- Meeting to discuss feedback from data collection plan and edit planning and distribution (Sean + Nick + Allan) - Edited references, and updated method (Allan + Sean) - Edited hypothesised best configuration (Nick)
Monday 22nd of May	- Exploratory analysis completed (Allan) - Abstract and conclusion begun (Allan + Nick)
Tuesday 23rd of May	- Discussion regarding results section of report (Nick + Sean) - Appendix updated (Nick + Sean)
Wednesday 24th of May	- Abstract and conclusion completed (Nick + Allan)
Thursday 27th of May	- Finalised report, collective proof reading and updated table of contents (Sean + Nick + Allan)