Kimberly Bastien
Yash Jain
Rahul Sivaram
Linxi Sun

# LEGO DOE Report

## Executive Summary

In this project, we constructed a Lego car in order to analyze the combinations of 4 factors in which we believed would maximize the distance traveled. Using the principles of DOE, such as ANOVA and factorial analysis, we analyzed the results of several experimental runs using statistical methods in order to determine which combination of factors would maximize the distance traveled by the car. Amongst other metrics, statistical analysis resulted in our model having an R-Squared value of 77%. This not only reinforced our selection of factors we believed would be important in maximizing the distance travelled, but also gave credence to the heuristic observations we arrived at during the data collection phase. Overall, the experiment emphasized the significance of accurate and consistent measurements, as well as the importance of factor selection and the selection of the number and values of factors.

## Project Objectives and Goals

The purpose of this project was to maximize the distance that a Lego-constructed car can travel down a pre-constructed ramp based on an optimal combination of design components. Constrained by a limited number of Lego blocks, our team underwent several experimental processes, where the car was assembled based on several factors, in order to compare the distance of each design in order to conclude the optimal model.

## Experiment Description

All observations were conducted on the same-day in order to control process variables. Our team underwent a series of 48 trials, where 3 replications of each design was conducted, based on four factors and two levels: a low level and a high level. During each trial, the car was released at the top of a ramp that measured 34 cm (L) x  22.6 cm (H), with an angle that was set to 30°. The distance traveled was measured with a measuring tape and recorded. This allowed us to gather data on each combination of factors in order to detect the best car design model.

**Factors**

| Factor | Factor Name | Low Level | High Level |
|---|---|---|---|
| A | Weight | 28 | 40 |
| B | Length | 4.5 | 8 |
| C | Ground Clearance | 0.8 | 2 |
| D | Rear Wheel Size | Small | Big |

The factors that we used for this project were ***weight, length, ground clearance*** and ***rear wheel size***. The first factor that we used was the weight of the car. To see the impact of this factor, we checked by considering different weights of 28 gm and 40 gm. Our second factor length defines the base length of our car, i.e the length between the front wheel and rear wheel of the car. In analyzing the influence of the length, we experimented with different base lengths of 4.5 cm and 8 cm. Then, for our third factor we took the ground clearance into consideration; where ground clearance is the height of the base from the ground. To check how this matters, we changed the ground clearance from 0.8 cm to 2 cm. The final factor that we took into account was the size of the rear wheel. To see how this factor affects our result we experimented with different sizes of rear wheel denoted with "small" and "big".

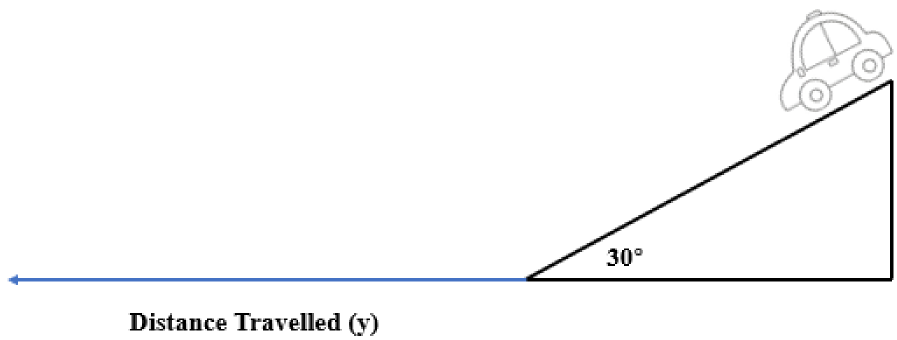
**Noise Factors**

The following noise factors were determined to be significant in our experimental setup and needed to be controlled:

- Material of the ramp:
  - Higher friction materials reduce the distance travelled.
  - Controlled using the same, smooth surfaced ramp for each experiment.
- Environment in which experiments were conducted:
  - Factors such as wind can affect the distance travelled.
  - Experiment was conducted in a closed, wind-controlled environment.
- Calibration of measurement equipment:
  - Minute changes in measurement method may change the inputs going into the analysis.
  - A measurement tape was fixed along the experiment track of the car in order to minimize the recalibration required for measuring the distance travelled.

**Response Variable**

In this experiment, our response variable was the horizontal distance traveled measuring from the end of the makeshift ramp. We constructed a makeshift ramp and maintained the same height, length, and angle of incline (30°) for all factor combinations, resulting in 48 observations gathered throughout the experiment.




**Experimental Design**

The design of this experiment was based on carrying out trials on a Lego race-car by taking 4 factors and 3 replications into consideration. To decrease the number of variations obtained on experimenting with the model we made instant changes to above mentioned parameters. This allowed us to make comparisons between various sets of factors. We made sure that our car's performance was not affected by any external factors.

A few constraints were set for the experiment: The first constraint that we made sure of was that no extra Lego blocks were used in the project. The second constraint was adding the steering wheel and windshield to every model of the race-car that we designed. The third constraint was keeping the ramp angle to nearly 30°.

**Data Analysis**

The data comprised of:

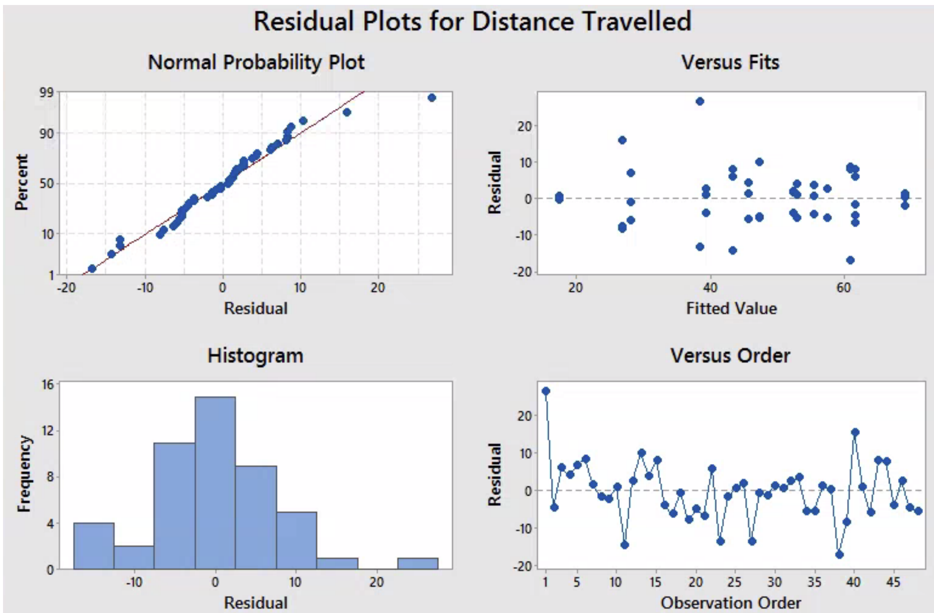
- 4 factors
- 2 levels per factor
- 3 experiment replications per factor combination

This resulted in a total of 48 records of data collected. A snapshot of the data collected is below:

| RunOrder | CenterPt | Blocks | Weight | Length | Ground Clearance | Rear Wheel Size | Distance Travelled |
|---|---|---|---|---|---|---|---|
| 6 | 1 | 1 | 28 | 4.5 | 0.8 | Big | 69.5 |
| 38 | 1 | 1 | 28 | 4.5 | 0.8 | Big | 44 |
| 43 | 1 | 1 | 28 | 4.5 | 0.8 | Big | 69 |
| 2 | 1 | 1 | 28 | 4.5 | 0.8 | Small | 51 |
| 25 | 1 | 1 | 28 | 4.5 | 0.8 | Small | 56 |
| 33 | 1 | 1 | 28 | 4.5 | 0.8 | Small | 59 |
| 8 | 1 | 1 | 28 | 4.5 | 2 | Big | 60 |
| 22 | 1 | 1 | 28 | 4.5 | 2 | Big | 67.5 |

Using this data, we moved on to performing data analysis of the design in Minitab.

We began by first testing the assumptions required by a linear model by analysing the characteristics of the residuals.




From the 4-in-1 charts above, we can interpret the following information about the residuals:

- The normal probability shows almost all points lie on the centerline, indicating that our data is normally distributed.
- The data points in the versus fit chart are mostly evenly as well as randomly distributed above and below the centerline; indicating a linear relationship between the independent and response variables is appropriate.
- The histogram indicates a fairly normal distribution.
- The data points in the versus order chart do not follow a discernable pattern, and are randomly distributed along the center, another positive indicator of linear relations for the independent and response variables.

From the charts above, we can be confident our independent and response variables have an appropriate linear relationship. We can observe a few outlier points in the plots above, which we attribute as noise coming from possible errors in measurement of the response variable.

Moving forward, we perform an ANOVA test to determine the significance of our model, and its constituent factors.

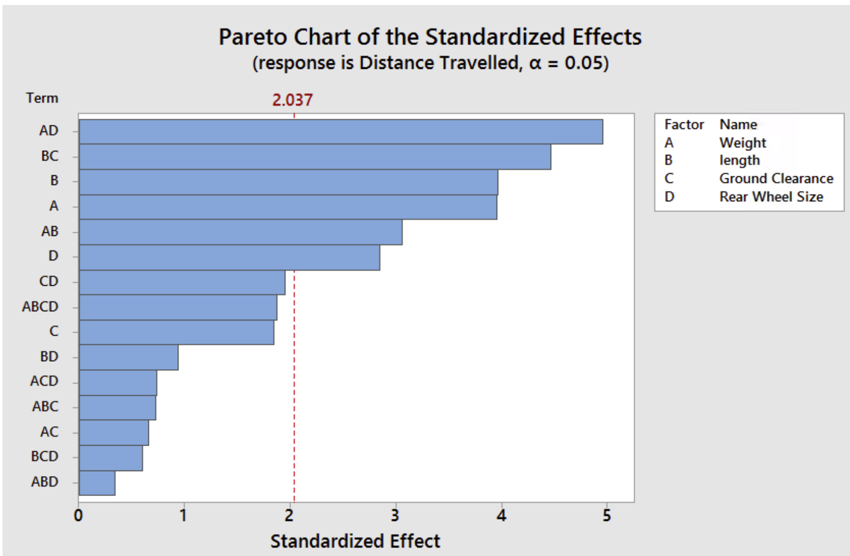
| Model Summary | | | |
|---|---|---|---|
| S | R-sq | R-sq(adj) | R-sq(pred) |
| 9.3775 | 76.98% | 66.19% | 48.21% |

**Overall Statistics of Regression Analysis**

Beginning with the overall model summary, we observe an R-Squared value of approximately 77%. This shows that the model explains around 77% of the variability in the data, positively reinforcing the selection of our factors for this experiment. However, the predicted R-Squared value is much lesser than the R-Squared value, meaning we cannot rely on this model to predict future/unseen values of the response variable. This may once again be attributed to errors on the measurement side of the response variable, as well as a need for refinement of the factor levels; The factors considered while designing the car were significant, but the levels of the factors would benefit from optimization.

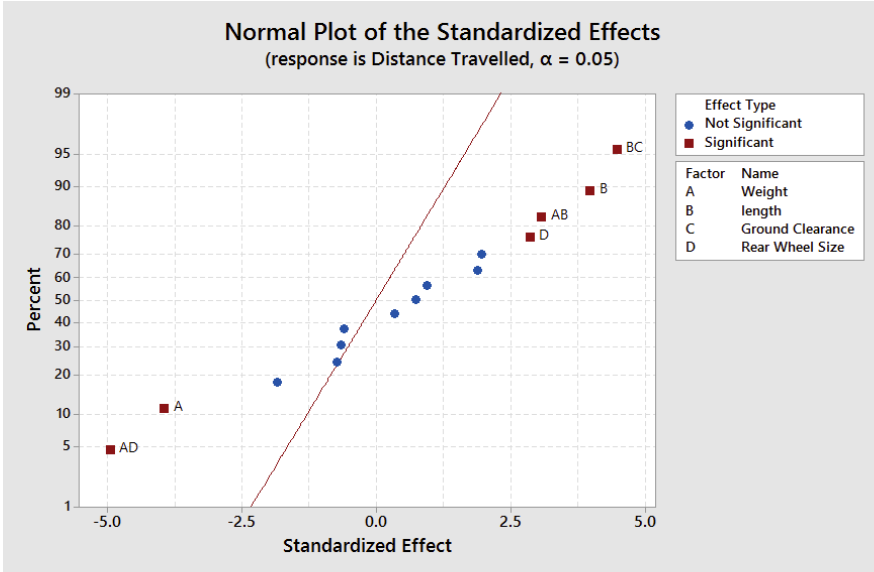
| Source | DF | Adj SS | Adj MS | F-Value | P-Value |
|---|---|---|---|---|---|
| **Model** | 15 | 9410.9 | 627.39 | 7.13 | 0 |
| Linear | 4 | 3775.8 | 943.95 | 10.73 | 0 |
| Weight | 1 | 1376 | 1376.02 | 15.65 | 0 |
| length | 1 | 1386.8 | 1386.75 | 15.77 | 0 |
| Ground Clearance | 1 | 300 | 300 | 3.41 | 0.074 |
| Rear Wheel Size | 1 | 713 | 713.02 | 8.11 | 0.008 |
| **2-Way Interactions** | 6 | 5189.3 | 864.88 | 9.84 | 0 |
| Weight*length | 1 | 825 | 825.02 | 9.38 | 0.004 |
| Weight*Ground Clearance | 1 | 38.5 | 38.52 | 0.44 | 0.513 |
| Weight*Rear Wheel Size | 1 | 2160.1 | 2160.08 | 24.56 | 0 |
| length*Ground Clearance | 1 | 1752.1 | 1752.08 | 19.92 | 0 |
| length*Rear Wheel Size | 1 | 77.5 | 77.52 | 0.88 | 0.355 |
| Ground Clearance*Rear Wheel Size | 1 | 336 | 336.02 | 3.82 | 0.059 |
| **3-Way Interactions** | 4 | 135.8 | 33.95 | 0.39 | 0.817 |
| Weight*length*Ground Clearance | 1 | 46 | 46.02 | 0.52 | 0.475 |
| Weight*length*Rear Wheel Size | 1 | 10.1 | 10.08 | 0.11 | 0.737 |
| Weight*Ground Clearance*Rear Wheel Size | 1 | 48 | 48 | 0.55 | 0.465 |
| length*Ground Clearance*Rear Wheel Size | 1 | 31.7 | 31.69 | 0.36 | 0.553 |
| 4-Way Interactions | 1 | 310.1 | 310.08 | 3.53 | 0.07 |
| Weight*length*Ground Clearance*Rear Wheel Size | 1 | 310.1 | 310.08 | 3.53 | 0.07 |
| **Error** | **32** | **2814** | **87.94** | | |
| **Total** | **47** | **12224.9** | | | |

From the table above, the highlighted rows show the factors (including interaction terms) which the model has found not to be statistically significant at a 95% CI. This conclusion is drawn by looking at the p-values for the factor/interaction. We can further confirm this from the DOE charts below. It is to be noted that the ground clearance does not have a statistically significant p-value according to the results of the analysis. However, it has statistical significance in the interaction with it and weight. To test this observation, we re-ran the model without ground clearance as a factor, and the R-Squared value of our model dropped to ~55%. This shows that although ground clearance did not have a significance as a base factor, it was important to retain it for its significant interaction effect on the response variable in concert with length.




The pareto chart above shows which terms are significant:

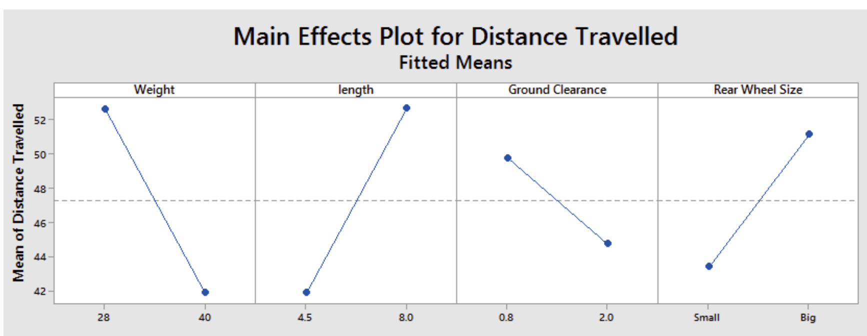
- **Base Terms:** A, B, D
- **Interaction Terms:** BC, AB, AD




The normal plot of standardized effects above further shows the polarity of effect of significant terms on the response variable i.e, are the effects negative or positive. Along with this, it too shows the significant vs insignificant effect factors and interactions. We can summarize this as:

- **Positive effects:** BC, B, AB, D
- **Negative effects:** A, AD

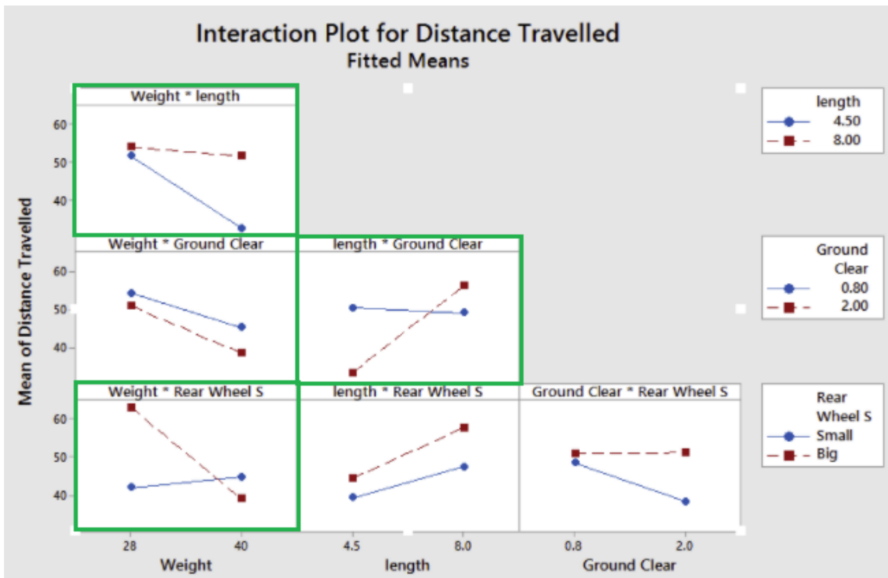
The interaction between length and ground clearance has the strongest positive effect, and the interaction between weight and rear wheel size has the strongest negative effect.




From the Main Effects Plot for the independent variables, we can understand the response variable is maximised when the terms are set at the following levels in combination:

| Factor | Optimal Level | Level Value |
|---|---|---|
| Weight | Low | 28 |
| Length | High | 8 |
| Ground Clearance | Low | 0.8 |
| Rear Wheel Size | High | Big |

However, keeping in mind the significance plots above have shown us the significance of certain interaction terms, we must also analyse the effect plots of the interactions to refine our understanding of the ideal combination of factors for a maximised response variable.

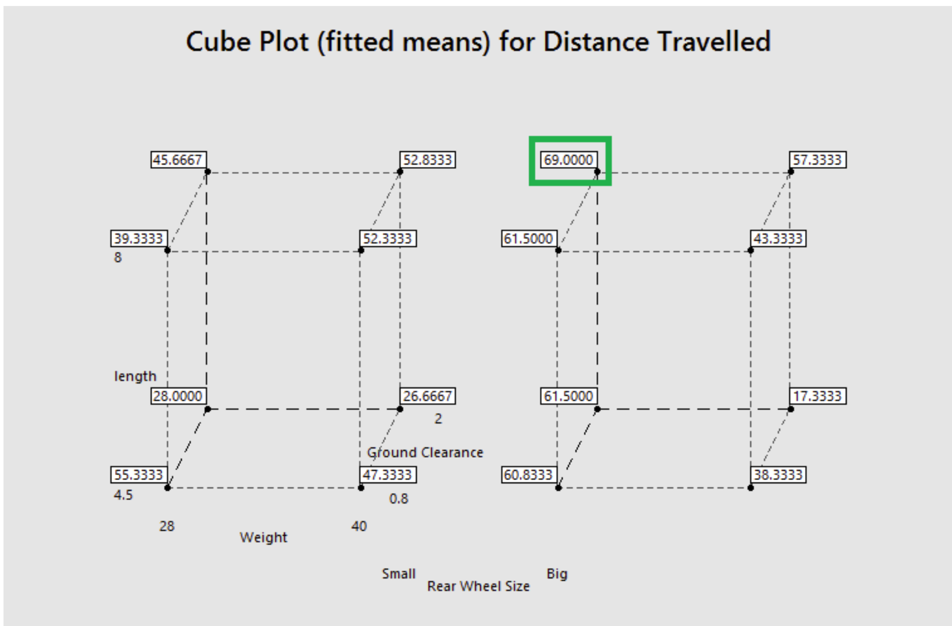



Analysing the interaction plots above, the following conclusions can be drawn:

- Based on the results of the normal plot of standardized effects, we analyse the following factor interactions: AD, BC, AB
    - AD (Weight * Rear Wheel Size, Bottom Left Corner):
        - A low weight in combination with a high rear wheel size results in the highest mean of the response variable.
    - BC (Length * Ground Clearance, Center):
        - A high length with a high ground clearance results in the highest mean of the response variable.
    - AB ( Weight * Length, Top Left Corner):
        - A low weight with a high length results in the highest mean of the response variable.

It can be observed that the interaction chart has refined our understanding of the ideal levels for the factors. It shows a low ground clearance maximises the response variable, however when viewed in concert with its interactions with the factor of length, we can conclude that a high ground clearance is the better choice.

From the cube plot below, we determine the optimal levels for each factor in order to maximise the response variable, "Distance Travelled."




The maximum mean value of distance travelled can be observed at the corner point of the right cube, highlighted in the green square. The cube chart tells us the ideal combination of factor-levels is:

| Factor | Optimal Level | Level Value |
|---|---|---|
| Weight | Low | 28 |
| Length | High | 8 |
| Ground Clearance | High | 2 |
| Rear Wheel Size | High | Big |

**Financial Analysis**

The best model of our car is the model with less weight, higher length, higher ground clearance and larger rear wheels. This model covered the maximum average distance compared to others. The minimum cost that would be required to build this model would be $10,600. To build this model cost-effective we had to compromise with the number of lego blocks. We had to use the blocks having lower cost price.

Our recommendations would be:

- Using the parts which are cheaper and more durable to build the car.
- The raw material used to manufacture wheels could be improved so that they have a better lifetime.

**Conclusion and Recommendations**

In conclusion, we found that the most important factors were weight and length in our experiment. Both have a great effect on the distance travelled. By comparison, the ground clearance has the least effect on our response variable as a base factor, but was important as an interaction factor. Besides, a low weight in combination with a high rear wheel size result in the highest mean distance travelled. The optimal combination of factors of the Lego car we chose includes low level weight (28 g), high level ground clearance (2 cm), high level length (8cm), and big rear wheel size, because we found that this car achieved the maximum average distance travelled compared to the other models.

Based on the results, we recommend that we could complete this experiment with more factors, because we found that R Squared explained 77% of data variation and other factors may contribute to decreasing the error in the model, such as center of gravity, spoiler, and wheelbase. The second recommendation is that we could introduce more levels for factors, and this, in turn, optimizes the value of a significant factor which would maximize the distance travelled. Finally, improving the measurement accuracy and reducing bias is a good way to help us obtain more accurate experimental results, for example, introducing an automated measurement system to keep measurements consistent and free of human error.