STATS4846/9864 Final Project
Ball Projection Experiment

Kecheng Xiao, Yunjie Xu, Nicholas Gregory DaFonseca Finn

Professor: Wenqing He

December 2020

Contents

1 Experiment Introduction

The main purpose of this experiment is to study the factors that will affect the drop distance of a ball-shaped object from a certain height level. To realize it, we designed a ball projection experiment and the analysis of the result is done by a 2^k Factorial Design. Also, while doing this experiment, we found out the processes of making paper-balls vary from person to person, which may result to differences in terms of the roundness or other features. To further reduce the concern of handmade paper-balls, the Nested Model Design is applied.[1]

2 PPDA Procedure

2.1 Problem

1. What factor will affect the drop distance of a ball-shaped object?
2. Is the maker of paper-balls a significant factor to the drop distance?

2.2 Plan

2.2.1 Experiment Unit

Two different types of balls are used:
1. Baseball
2. Paper-ball (Balled-up paper, and shape is fixed by some tape)

Note: The baseball have two levels: a standard size version ($radius = 3.4cm$), a smaller size version ($radius = 2.4cm$). The smaller version of the baseball is a dog toy, that has similar texture of a regular baseball. The paper-balls are made in the same dimensions. Within each levels, all ball samples are randomly selected from the shop or made by group members.

2.2.2 Procedure

Experiment 1
Push balls with a mechanical spring (the upper arm of the hole puncher) from a certain distance to the edge of the table with height, observe the distance from the table edge projected on the floor to the first drop point. Record the measured distance and repeat one more times. For the tables we used, the surface texture is either smooth wood or rough wood.

Experiment 2
All group members are provided with four paper to make two small size paper-balls, two paper per ball. The first ball's experiment result is considered as the first batch, so we have two batches in total. When the paper-balls are pressed to the standard dimensions, then they can be balled-up by tape. The method

3

of making paper-balls is not restricted. Finally, we use all balls to run the first experiment with all other levels are controlled as low.

2.2.3 Response

Distance from the table edge projected on the floor to the first drop point

2.2.4 Factors

1. Ball Type (2 levels: Paper-ball, Baseball; leveling by their weight)
2. Force of pushing (2 levels: Spring with half pressed or full pressed)
3. Distance rolled on the table (2 levels: 15cm or 45cm)
4. Table Height (2 levels: 0.5m or 1m)
5. Table Surface (2 levels: Smooth or Rough)
6. Maker (3 suppliers; studied by the second experiment)

2.3 Data

Experiment 1
Sample size of the first experiment is determined by the leveling of the factors(exclude the Maker) and the 2 replicate requirement, since we still want to analyze how the interactions of factors affect the final distance, so the sample size calculation is: $2 \times 2 \times 2 \times 2 \times 2 \times 2 = 64$. The detail of the data is in the Data1 file.

Experiment 2
The experiment of maker's effect on paper-balls includes 18 observations, which is determined by the data structure including three makers, two batches and three test runs, so the sample size calculation is: $3 \times 2 \times 3 = 18$. The detail of the data is in the "Nested Model Design" section.

2.4 Analysis

1. 2^k Factorial Design: Study all the potential two-level factors with one model.
2. Nested Model Design: Study the maker's effect on the paper-ball result.

3 Exploratory Data Analysis

3.1 Exploring Data Using Plots

In order to explore how each factor and its levels impact the Data, we will be using Boxplots. Boxplots work well with factored variables as it allows us to see the mean value as well as the range (IQR specifically) that output distances are corresponding to each factor level. This allows us to really visualize the relations between factor levels and the distance output.

4

Table Factors: Surface, Height, Length(Decelerate Distance)

Plot of Distance for Surface

Plot of Distance for Height

Plot of Distance for Table Length
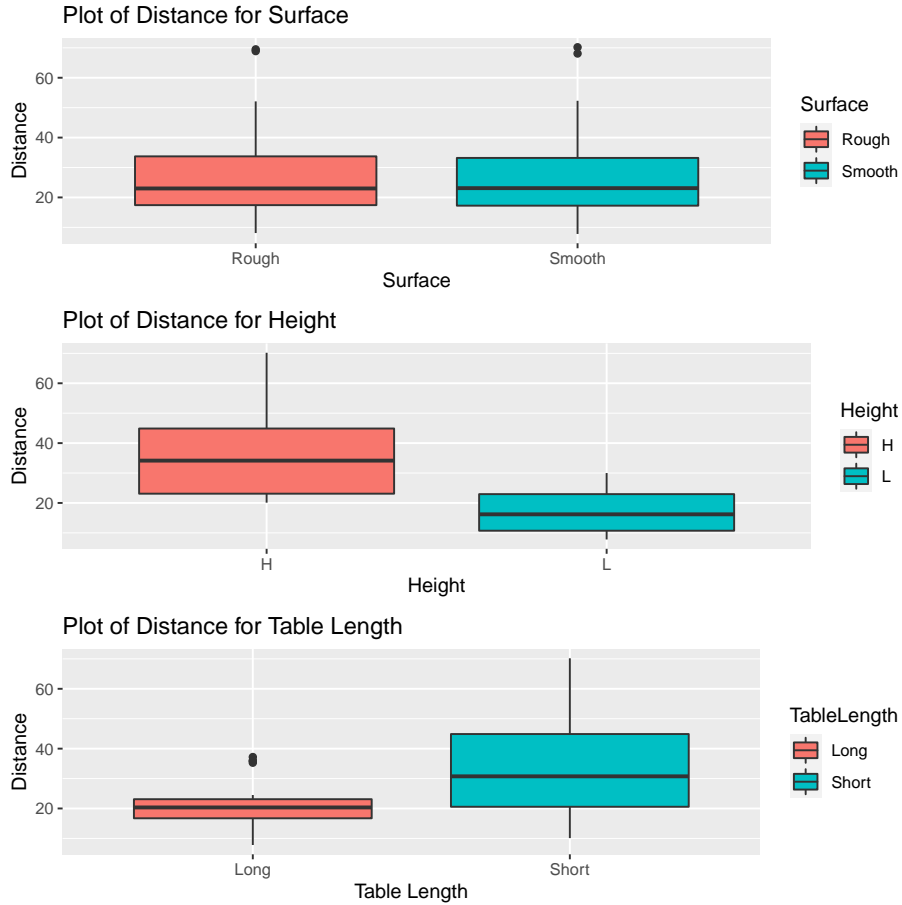
When looking out our first plot of Rough vs Smooth surface tables, we can almost see that it seemingly makes no impact on distance, as the boxplots are almost identical for each factor level. This could be foreshadowing of a factor that could be omitted from future use, or possibly used as a nuisance factor in other models. Next we can see that for the boxplot of High vs Low table height, we can see that higher Height increases distance a significant amount, which would make a lot of sense in regards to how the data was gathered. As it was a drop ball test, and the amount of time it would have to fall would impact it's total distance. Now for our plot of Table Lengths, we can see that there seems to be a much greater variability in the distances from the shorter decelerate distance in comparison to that of the longer decelerate distance. While also the values of the shorter decelerate distance were consistently larger than that of the longer decelerate distance. This could possibly be due to a longer decelerate distance providing less spontaneous results.

Ball Factors: Size, Type

Plot of Distance for Ball size

Plot of Distance for Type of Ball
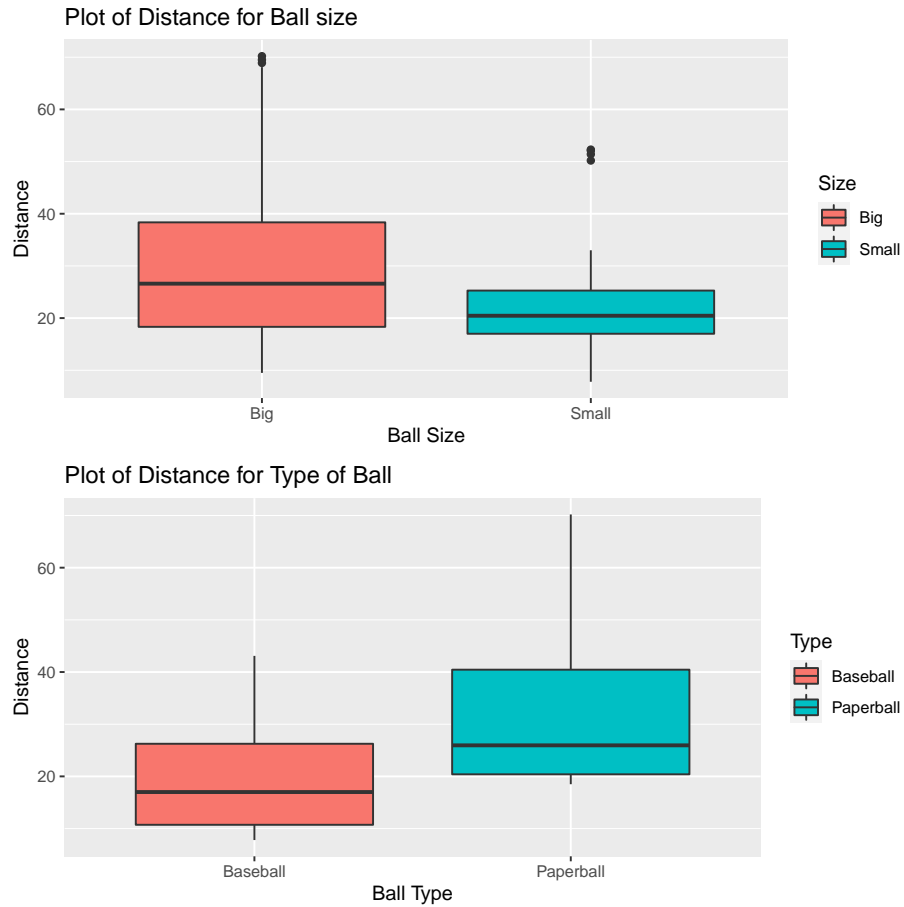
When comparing the distance travelled for each ball size, we can see that the average of each balls distance travelled is kind of similar. Although there seems to be a decently higher range for the larger balls. This could be due to physics possibly, and that larger balls fall faster with more force. Although the means aren't that far apart, it's just the larger sized balls had a higher upper IQR range. When finally analyzing the box plot of our two types of balls, we can see that there is a difference in distance outcomes between each ball, with paper-ball travelling farther confidently overall. The difference in means is not by quite a lot, and the plots are very similar in shape, just with paper-ball scaled up higher a bit more.

3.2 Analyzing the Effects in our Data

Given that we are utilizing a 2^5 factorial model, we should end up with a total of 32 factors:

5 Main Factors:

A = Surface Type (Rough, Smooth)

B = Table Height (Low, High)

C = Table Length (Short, Long)

D = Ball Size (Small, Big)

E = Ball Type (Baseball, Paperball)

And with this will also come 27 interaction factors

In order to fully analyze our factors, we must first calculate the estimates of each.

```
##           A           B           C           D           E          AB          AC
##   -0.003125  -20.109375  -13.840625   -7.284375   13.328125    0.078125   -0.115625
##          AD          BC          BD          CD          AE          BE          CE
##    0.040625    0.115625    9.478125    4.071875   -1.040625    1.815625   -5.096875
##          DE         ABC         ABD         ACD         BCD         ABE         ACE
##   -3.228125    0.040625   -0.265625   -0.115625   -0.009375   -0.109375    0.009375
##         BCE         BDE         CDE        ABCD        ABCE        ABDE        ACDE
##   -0.803125    3.609375   -0.390625   -0.015625   -0.140625    0.040625    0.096875
##        BCDE       ABCDE
##    1.178125    0.053125
```

There are multiple interesting things to note about our output estimates. A big one being the very low estimated effect of Surface(A), which when you think about it would make some sense given our observed plot of Surface, where it seemed like changing between the two levels made no impact. In contrast, all other main effects have high estimations. Another to things to note is the highest estimated effects of BE, CD and DE, the two-way interaction effect.

3.3 Normal Q-Q Plot of Factors

To visualize the significance of effects, the Normal Q-Q plot is used. This allows us to easily compare our estimated factors that we previously calculated and the distances from the normal line almost give them a ranking. The single replicate assumption is applied with only using the first replicate value.

Normal Q–Q Plot
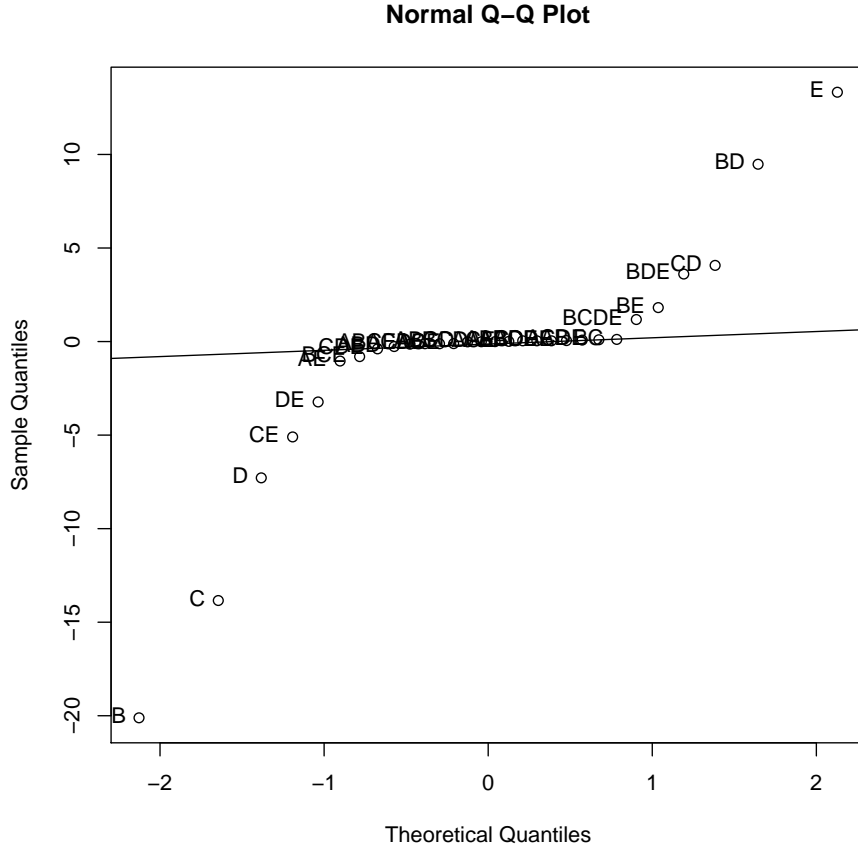
We can clearly see that all main factors are the most significant except the A. Many two-way interaction factors such as: BD, CD and CE are being close behind. Next up it's interesting to see that high-level interaction factors BDE and BCDE are the next closest. In fact, all factors include the A(Surface) have no significant effect, so we can conclude A has no interaction to other factors. Therefore, we can predict that the main factors other than A and the two-way interaction factors will be dominate factors in the 2^k Factorial Design.

4 2^k Factorial Design

4.1 Model Assumption

All five factors are assumed to be fixed, and the treatment effects are defined as the deviations from the overall mean, so $\sum_{i=1}^{a} \tau = 0$, $\sum_{i=1}^{b} \beta = 0$. Similarly, all the interaction effects are fixed and are defined, such that $\sum_{i=1}^{a} \tau\beta = 0$, and random error component $\varepsilon \sim N(0, \sigma^2)$.

4.2 Model Screening: ANOVA

4.2.1 Full Model

```
## Analysis of Variance Table
##
## Response: Distance
##                 Df Sum Sq Mean Sq   F value    Pr(>F)
## x1               1    0.0     0.0    0.0001 0.9907210
## x2               1 6470.2  6470.2 5688.8618 < 2.2e-16 ***
## x3               1 3065.0  3065.0 2694.8813 < 2.2e-16 ***
## x4               1  849.0   849.0  746.4708 < 2.2e-16 ***
## x5               1 2842.2  2842.2 2499.0005 < 2.2e-16 ***
## x1:x2            1    0.1     0.1    0.0859 0.7713952
## x1:x3            1    0.2     0.2    0.1881 0.6674331
## x2:x3            1 1437.4  1437.4 1263.7847 < 2.2e-16 ***
## x1:x4            1    0.0     0.0    0.0232 0.8798496
## x2:x4            1  265.3   265.3  233.2476 2.993e-16 ***
## x3:x4            1   52.7    52.7   46.3746 1.064e-07 ***
## x1:x5            1    0.2     0.2    0.1881 0.6674331
## x2:x5            1   17.3    17.3   15.2341 0.0004595 ***
## x3:x5            1  415.7   415.7  365.4569 < 2.2e-16 ***
## x4:x5            1  166.7   166.7  146.5983 1.732e-13 ***
## x1:x2:x3         1    0.0     0.0    0.0232 0.8798496
## x1:x2:x4         1    1.1     1.1    0.9926 0.3265867
## x1:x3:x4         1    0.0     0.0    0.0012 0.9721682
## x2:x3:x4         1   10.3    10.3    9.0739 0.0050341 **
## x1:x2:x5         1    0.2     0.2    0.1881 0.6674331
## x1:x3:x5         1    0.2     0.2    0.1683 0.6843677
## x2:x3:x5         1  208.4   208.4  183.2704 8.578e-15 ***
## x1:x4:x5         1    0.0     0.0    0.0012 0.9721682
## x2:x4:x5         1   66.6    66.6   58.5807 1.009e-08 ***
## x3:x4:x5         1    2.4     2.4    2.1466 0.1526443
## x1:x2:x3:x4      1    0.0     0.0    0.0034 0.9536312
## x1:x2:x3:x5      1    0.3     0.3    0.2782 0.6015232
## x1:x2:x4:x5      1    0.0     0.0    0.0232 0.8798496
## x1:x3:x4:x5      1    0.2     0.2    0.1320 0.7187325
## x2:x3:x4:x5      1   22.2    22.2   19.5259 0.0001066 ***
## x1:x2:x3:x4:x5   1    0.0     0.0    0.0397 0.8433226
## Residuals       32   36.4     1.1
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

At first, we use the full model to fit. From the result of ANOVA table, we find result that the main factor, Surface, is insignificant and many of interaction effects including it are also insignificant, especially for higher level interactions.

4.2.2 Model with Significant Factors

```
## Analysis of Variance Table
##
## Response: Distance
##              Df Sum Sq Mean Sq  F value    Pr(>F)
## x2            1 6470.2  6470.2 7952.607 < 2.2e-16 ***
## x3            1 3065.0  3065.0 3767.244 < 2.2e-16 ***
## x4            1  849.0   849.0 1043.511 < 2.2e-16 ***
## x5            1 2842.2  2842.2 3493.418 < 2.2e-16 ***
## x2:x3         1 1437.4  1437.4 1766.677 < 2.2e-16 ***
## x2:x4         1  265.3   265.3  326.063 < 2.2e-16 ***
## x3:x4         1   52.7    52.7   64.828 1.840e-10 ***
## x2:x5         1   17.3    17.3   21.296 2.953e-05 ***
## x3:x5         1  415.7   415.7  510.882 < 2.2e-16 ***
## x4:x5         1  166.7   166.7  204.934 < 2.2e-16 ***
## x2:x3:x4      1   10.3    10.3   12.685 0.0008443 ***
## x2:x3:x5      1  208.4   208.4  256.198 < 2.2e-16 ***
## x2:x4:x5      1   66.6    66.6   81.891 6.000e-12 ***
## x2:x3:x4:x5   2   24.6    12.3   15.148 7.943e-06 ***
## Residuals    48   39.1     0.8
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

From the ANOVA table above, we remove all the insignificant factor based on $\alpha = 0.05$. Then, we fit the model with only significant factors. Here is the result. Since all the factors are significant right now, thus, we assume this model can be one of the final models.

4.2.3 Step-wise Model

```
##
## Call:
## lm(formula = Distance ~ x2 + x3 + x4 + x5 + x2:x3 + x2:x4 + x3:x4 +
##     x2:x5 + x3:x5 + x4:x5 + x2:x3:x4 + x2:x3:x5 + x2:x4:x5 +
##     x3:x4:x5 + x2:x3:x4:x5, data = df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.7500 -0.4500 -0.1000  0.5062  1.5500
##
## Coefficients:
##                                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)                      23.6000     0.4510  52.328  < 2e-16 ***
```

```
## x2L                              -13.2500    0.6378 -20.774  < 2e-16 ***
## x3Short                           18.8000    0.6378  29.476  < 2e-16 ***
## x4Small                           -1.9000    0.6378  -2.979  0.00453 **
## x5Paperball                       12.5000    0.6378  19.598  < 2e-16 ***
## x2L:x3Short                      -15.7000    0.9020 -17.406  < 2e-16 ***
## x2L:x4Small                        0.1000    0.9020   0.111  0.91219
## x3Short:x4Small                   -8.3750    0.9020  -9.285 2.71e-12 ***
## x2L:x5Paperball                   -1.3000    0.9020  -1.441  0.15601
## x3Short:x5Paperball               14.2750    0.9020  15.826  < 2e-16 ***
## x4Small:x5Paperball              -13.6750    0.9020 -15.161  < 2e-16 ***
## x2L:x3Short:x4Small                7.9250    1.2756   6.213 1.19e-07 ***
## x2L:x3Short:x5Paperball           -9.7250    1.2756  -7.624 8.19e-10 ***
## x2L:x4Small:x5Paperball           12.8750    1.2756  10.093 1.87e-13 ***
## x3Short:x4Small:x5Paperball        6.2750    1.2756   4.919 1.06e-05 ***
## x2L:x3Short:x4Small:x5Paperball   -9.4250    1.8040  -5.225 3.74e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.902 on 48 degrees of freedom
## Multiple R-squared:  0.9975,Adjusted R-squared:  0.9968
## F-statistic:  1302 on 15 and 48 DF,  p-value: < 2.2e-16
```

Before we check the model adequacy, we use the step-wise selection to verify that the model we selected above is the best model. After we find the row model selected by step-wise selection, we still see there are some effects are insignificant. We get rid of those factors. Then we obtain the final model from step-wise selection.

4.2.4   Final Model

```
## Analysis of Variance Table
##
## Response: Distance
##              Df Sum Sq Mean Sq  F value     Pr(>F)
## x2            1 6470.2  6470.2 7952.607 < 2.2e-16 ***
## x3            1 3065.0  3065.0 3767.244 < 2.2e-16 ***
## x4            1  849.0   849.0 1043.511 < 2.2e-16 ***
## x5            1 2842.2  2842.2 3493.418 < 2.2e-16 ***
## x2:x3         1 1437.4  1437.4 1766.677 < 2.2e-16 ***
## x2:x4         1  265.3   265.3  326.063 < 2.2e-16 ***
## x3:x4         1   52.7    52.7   64.828 1.840e-10 ***
## x2:x5         1   17.3    17.3   21.296 2.953e-05 ***
## x3:x5         1  415.7   415.7  510.882 < 2.2e-16 ***
## x4:x5         1  166.7   166.7  204.934 < 2.2e-16 ***
## x2:x3:x4      1   10.3    10.3   12.685 0.0008443 ***
```

```
## x2:x3:x5     1  208.4   208.4  256.198 < 2.2e-16 ***
## x2:x4:x5     1   66.6    66.6   81.891 6.000e-12 ***
## x2:x3:x4:x5  2   24.6    12.3   15.148 7.943e-06 ***
## Residuals   48   39.1     0.8
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

   Since the result from the step-wise selection is the same from the first model. We can determine that the first model is the final model.

4.3   Model Adequacy
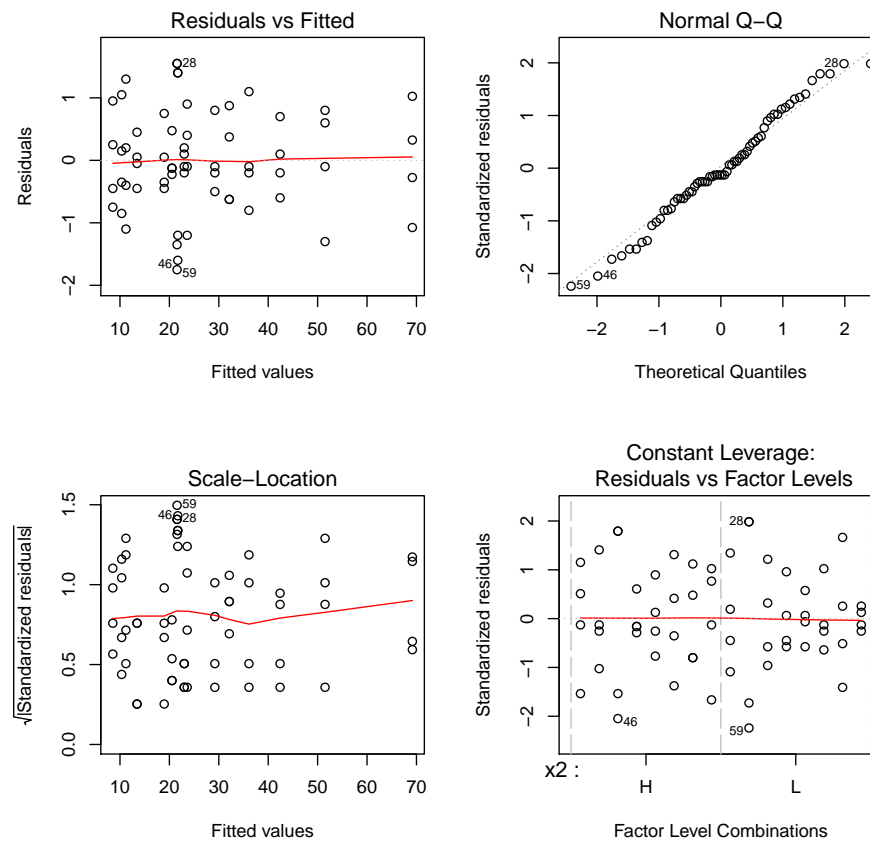
4.3.1   Conclusion

Residual vs. Fitted Plot Residuals "bounce randomly" around the 0 line, which indicates the linearity assumption holds. Residuals roughly form a "hori-

zontal band" around the 0 line, indicating the equal variance assumption holds.
Normal Q-Q plot The Q-Q plot gives a linear pattern with no obvious fat tails, thus the error terms should come from a normal distribution.
Scale-Location The scale of residuals change to the standardized residual. The assumption holds in first plot, so will also hold in here.
Residuals vs. Leverage There are three influential points. But the influence will not be significant by observing the plot.

5 Nested Model Design

5.1 Table

Since the paper-balls are handmade, to reduce the concern from the effect of different makers, the Nested Model Design is used, data are listed as follows:

	Maker 1		Maker 2		Maker 3	
Batch	1	2	1	2	1	2
run 1	21.5	22.1	19.5	21.1	20.1	22.4
run 2	20.0	21.6	20.6	20.5	21.6	21.0
run 3	20.6	23.5	20.5	21.5	21.5	21.2

5.2 Calculation

We will take the supplier and the batch are random factors, due to group members are randomly assigned and balls are made with random methods, so:

$$SST = \sum^a \sum^b \sum^n y_{ijk}^2 - \frac{y_{...}^2}{abn} = \sum(21.5^2 + 20.0^2 + ... + 21.2^2) - \frac{380.8^2}{3 \times 2 \times 3} = 8071.42 - 8056.04 = 15.38$$

$$SSA = \frac{1}{bn} \sum^a y_{i..}^2 - \frac{y_{...}^2}{abn} = \frac{129.3^2 + 123.7^2 + 127.8^2}{6} - \frac{380.8^2}{18} = 8058.84 - 8056.04 = 2.80$$

$$SSB_{(A)} = \frac{1}{n} \sum^a \sum^b y_{ij.}^2 - \frac{1}{bn} \sum^a y_{i..}^2 = \frac{\sum(62.1^2 + ... + 64.6^2)}{3} - \frac{129.3^2 + 123.7^2 + 127.8^2}{6} = 5.70$$

$$SSE = \sum^a \sum^b \sum^n y_{ijk}^2 - \frac{1}{n} \sum^a \sum^b y_{ij.}^2 = 8071.42 - 8064.54 = 6.88$$

5.3 ANOVA

SS		DF	MS	$All - Random$	F^*	$F_{0.05,v_1,v_2}$	H_0
SSA	2.80	2	1.40	$MSA/MSB_{(A)}$	0.737	9.55	Not Reject
$SSB_{(A)}$	5.70	3	1.90	$MSB_{(A)}/MSE$	3.316	3.49	Not Reject
SSE	6.88	12	0.573				
SST	15.38	17					

5.4 Conclusion

From the ANOVA result, we can see there is no different between makers and batches under $\alpha = 0.05$, but the batch effect will be significant under $\alpha = 0.01$. This indicates that paper-balls performances in different batches are different in a slight level, but the differences caused by the maker effect are very small, and will not be statistically significant.

6 Discussion

From the physics point of view, the experiment of this project is similar to a flat parabolic motion study. The two core factors are the initial velocity and height of falling. According to the physics law, the factors: Surface, Table Length(Decelerate Distance), Ball Size and Ball Type should influence on initial velocity, so the statistics conclusions are correct in the majority. One exception is that the table surface has no significant effect. After group members' discussion, the most likely cause is the rough table texture did not produce enough fractions to decelerate the ball motion.

7 Conclusion

With the Exploratory Data Analysis, the 2^k Factorial Design and the Nested Model Design, we complete the experiment with credible results. Firstly, we come up some initial thought from the two level comparison with the Boxplots. Then, we discover the significant factors that will affect the drop distance of a ball-shaped object from a certain height level, which are the Ball Type, Ball Size, Table Height, and Table Length(Decelerate Distance Length on Table). The Table Surface will not have effect, because the ball motion are not sensitive to the slightly rougher texture. Finally, we excluded the effect of makers by the Nested Model results.

References

[1] Douglas C. Montgomery. *Design and Analysis of Experiments*. John Wiley Sons, 2012.