Project: Analysis of 2^5 Factorial Design using R

Submitted by:

Wasif Hameed

MSc Econometrics & Statistics
*Summer Semester **2011***
Pakistan Institute of Development Economics (PIDE)

Problem Statement:

Data of High Performance Ceramic Experiment:

This data set was taken from an experiment that was performed a few years ago at NIST by Said Jahanmir of the Ceramics Division in the Material Science and Engineering Laboratory. This example shows an independent analysis of a modified portion of the original data set. The original data set was part of a high performance ceramics experiment with the goal of characterizing the effect of grinding parameters on sintered reaction-bonded silicon nitride, reaction bonded silicone nitride, and sintered silicon nitride.
Only modified data from the first of the three ceramic types (sintered reaction-bonded silicon nitride) will be discussed in this illustrative example of a full factorial data analysis. The data and command structure is taken from this website: www.itl.**nist**.gov/

Solution:

Goal:

The experimenter is interested to know that during the production process how other important factors may affect the ceramic strength. In short **"Determine the effect of machining factors on ceramic strength".**

Defining Response variable and factors

Response variable Y = Mean of the ceramic strength
Number of observations = 32 (a complete 2^5 factorial design)

There are 5 factors each have two levels, for Factor 1 to Factor 3 which are quantitative data type we define their levels as: Low and High and for Factor 4 & 5 which are qualitative data type with there two levels each. Thus this according the situation this problem is suitable for the 2^5 Full Factorial Experiment. The details of these factors are given in the following table.

Table 1: Factors with their level

Factor 1	Table Speed (2 levels: Low (.025 m/s) and High (.125 m/s))
Factor 2	Down Feed Rate (2 levels: Low (.05 mm) and High (.125 mm))
Factor 3	Wheel Grit (2 levels: Low 140/170 and High 80/100)
Factor 4	Direction (2 levels: Longitudinal and Transverse)
Factor 5	Batch (2 levels: 1 and 2)

Creation of Data Matrix in R:

Commands used in R for this data matrix are given in **Appendix A.**

The data download from internet is in the .txt format. After saving it in the relevant working directory with R-Command the data is displayed as:

Two types of matrix for data display:

Type I:

Just import the .txt file and show in the R-output i.e.

```
    V1 V2 V3 V4 V5      V6 V7
1   X1 X2 X3 X4 X5       Y OD 2
   -1 -1 -1 -1 -1 680.45 17
3    1 -1 -1 -1 -1 722.48 30
4   -1  1 -1 -1 -1 702.14 14
5    1  1 -1 -1 -1 666.93  8
6   -1 -1  1 -1 -1 703.67 32
7    1 -1  1 -1 -1 642.14 20
8   -1  1  1 -1 -1 692.98 26
9    1  1  1 -1 -1 669.26 24
10  -1 -1 -1  1 -1 491.58 10
11   1 -1 -1  1 -1 475.52 16
12  -1  1 -1  1 -1 478.76 27
13   1  1 -1  1 -1 568.23 18
14  -1 -1  1  1 -1 444.72  3
15   1 -1  1  1 -1 410.37 19
16  -1  1  1  1 -1 428.51 31
17   1  1  1  1 -1 491.47 15
18  -1 -1 -1 -1  1 607.34 12
19   1 -1 -1 -1  1  620.8  1
20  -1  1 -1 -1  1 610.55  4
21   1  1 -1 -1  1 638.04 23
22  -1 -1  1 -1  1 585.19  2
23   1 -1  1 -1  1 586.17 28
24  -1  1  1 -1  1 601.67 11
25   1  1  1 -1  1 608.31  9
26  -1 -1 -1  1  1  442.9 25
```

```
 27   1 -1 -1  1  1 434.41 21
28  -1  1 -1  1  1 417.66  6
 29  1  1 -1  1  1 510.84  7
30  -1 -1  1  1  1 392.11  5
 31  1 -1  1  1  1 343.22 13
32  -1  1  1  1  1 385.52 22
 33  1  1  1  1  1 446.73 29
```

This above data matrix showed the strength of ceramic (Y) with actual randomized run order (OD) shown in the last column.

Type –II:

It is an alternate method to show the data matrix in R, command is in Appendix A. The out put is:

```
      [,1] [,2] [,3] [,4] [,5]   [,6] [,7]
 [1,]   -1   -1   -1   -1   -1 680.45   17
 [2,]    1   -1   -1   -1   -1 722.48   30
 [3,]   -1    1   -1   -1   -1 702.14   14
 [4,]    1    1   -1   -1   -1 666.93    8
 [5,]   -1   -1    1   -1   -1 703.67   32
 [6,]    1   -1    1   -1   -1 642.14   20
 [7,]   -1    1    1   -1   -1 692.98   26
 [8,]    1    1    1   -1   -1 669.26   24
 [9,]   -1   -1   -1    1   -1 491.58   10
[10,]    1   -1   -1    1   -1 475.52   16
[11,]   -1    1   -1    1   -1 478.76   27
[12,]    1    1   -1    1   -1 568.23   18
[13,]   -1   -1    1    1   -1 444.72    3
[14,]    1   -1    1    1   -1 410.37   19
[15,]   -1    1    1    1   -1 428.51   31
[16,]    1    1    1    1   -1 491.47   15
[17,]   -1   -1   -1   -1    1 607.34   12
[18,]    1   -1   -1   -1    1 620.80    1
[19,]   -1    1   -1   -1    1 610.55    4
[20,]    1    1   -1   -1    1 638.04   23
[21,]   -1   -1    1   -1    1 585.19    2
[22,]    1   -1    1   -1    1 586.17   28
[23,]   -1    1    1   -1    1 601.67   11
[24,]    1    1    1   -1    1 608.31    9
[25,]   -1   -1   -1    1    1 442.90   25
[26,]    1   -1   -1    1    1 434.41   21
[27,]   -1    1   -1    1    1 417.66    6
[28,]    1    1   -1    1    1 510.84    7
[29,]   -1   -1    1    1    1 392.11    5
[30,]    1   -1    1    1    1 343.22   13
[31,]   -1    1    1    1    1 385.52   22
[32,]    1    1    1    1    1 446.73   29
```

Saving the response, order variable and factors

After displaying the data we save the response (Y) and randomized run order (OD) variable. Also the factors i.e. speed, rate, grit, direction and batch number are saved as factors (see Appendix A for commands).

Declaring factors with abbreviation and interactions (2- factors interaction only)

The list below shows the abbreviation of factors:

Main Effects:

Speed ➦➦➦ s
Rate ➦➦➦ r
Grit ➦➦➦ g
Direction ➦➦➦ d
Batch ➦➦➦ b

Interaction term:

Direction*Batch ➦➦➦ db
Grit*Direction ➦➦➦ gd
Grit*Batch ➦➦➦ gb
Rate*Grit ➦➦➦ rg
Rate*Direction ➦➦➦ rd
Rate*Batch ➦➦➦ rb
Speed*Rate ➦➦➦ sr
Speed*Grit ➦➦➦ sg Speed*Direction ➦➦➦ sd
Speed*Batch ➦➦➦ sb

Data matrix with labeled variable and sign table of two factor interactions

Data Matrix:

	speed	rate	grit	direction	batch	strength	order
1	-1	-1	-1	-1	-1	680.45	17
2	1	-1	-1	-1	-1	722.48	30
3	-1	1	-1	-1	-1	702.14	14
4	1	1	-1	-1	-1	666.93	8
5	-1	-1	1	-1	-1	703.67	32
6	1	-1	1	-1	-1	642.14	20
7	-1	1	1	-1	-1	692.98	26
8	1	1	1	-1	-1	669.26	24
9	-1	-1	-1	1	-1	491.58	10
10	1	-1	-1	1	-1	475.52	16
11	-1	1	-1	1	-1	478.76	27
12	1	1	-1	1	-1	568.23	18
13	-1	-1	1	1	-1	444.72	3
14	1	-1	1	1	-1	410.37	19
15	-1	1	1	1	-1	428.51	31
16	1	1	1	1	-1	491.47	15
17	-1	-1	-1	-1	1	607.34	12
18	1	-1	-1	-1	1	620.80	1

19	-1	1	-1	-1	1	610.55	4
20	1	1	-1	-1	1	638.04	23
21	-1	-1	1	-1	1	585.19	2
22	1	-1	1	-1	1	586.17	28
23	-1	1	1	-1	1	601.67	11
24	1	1	1	-1	1	608.31	9
25	-1	-1	-1	1	1	442.90	25
26	1	-1	-1	1	1	434.41	21
27	-1	1	-1	1	1	417.66	6
28	1	1	-1	1	1	510.84	7
29	-1	-1	1	1	1	392.11	5
30	1	-1	1	1	1	343.22	13
31	-1	1	1	1	1	385.52	22
32	1	1	1	1	1	446.73	29

Sign Table up to 2-Factor interactions only

	s	**r**	**g**	**d**	**b**	**db**	**gd**	**gb**	**rg**	**rd**	**rb**	**sr**	**sg**	**ds**	**sb**
1	-1	-1	-1	-1	-1	1	1	1	1	1	1	1	1	1	1
2	1	-1	-1	-1	-1	1	1	1	1	1	1	-1	-1	-1	-1
3	-1	1	-1	-1	-1	1	1	1	-1	-1	-1	-1	1	1	1
4	1	1	-1	-1	-1	1	1	1	-1	-1	-1	1	-1	-1	-1
5	-1	-1	1	-1	-1	1	-1	-1	-1	1	1	1	-1	1	1
6	1	-1	1	-1	-1	1	-1	-1	-1	1	1	-1	1	-1	-1
7	-1	1	1	-1	-1	1	-1	-1	1	-1	-1	-1	-1	1	1
8	1	1	1	-1	-1	1	-1	-1	1	-1	-1	1	1	-1	-1
9	-1	-1	-1	1	-1	-1	-1	1	1	-1	1	1	1	-1	1
10	1	-1	-1	1	-1	-1	-1	1	1	-1	1	-1	-1	1	-1
11	-1	1	-1	1	-1	-1	-1	1	-1	1	-1	-1	1	-1	1
12	1	1	-1	1	-1	-1	-1	1	-1	1	-1	1	-1	1	-1
13	-1	-1	1	1	-1	-1	1	-1	-1	-1	1	1	-1	-1	1
14	1	-1	1	1	-1	-1	1	-1	-1	-1	1	-1	1	1	-1
15	-1	1	1	1	-1	-1	1	-1	1	1	-1	-1	-1	-1	1
16	1	1	1	1	-1	-1	1	-1	1	1	-1	1	1	1	-1
17	-1	-1	-1	-1	1	-1	1	-1	1	1	-1	1	1	1	-1
18	1	-1	-1	-1	1	-1	1	-1	1	1	-1	-1	-1	-1	1
19	-1	1	-1	-1	1	-1	1	-1	-1	-1	1	-1	1	1	-1
20	1	1	-1	-1	1	-1	1	-1	-1	-1	1	1	-1	-1	1
21	-1	-1	1	-1	1	-1	-1	1	-1	1	-1	1	-1	1	-1
22	1	-1	1	-1	1	-1	-1	1	-1	1	-1	-1	1	-1	1
23	-1	1	1	-1	1	-1	-1	1	1	-1	1	-1	-1	1	-1
24	1	1	1	-1	1	-1	-1	1	1	-1	1	1	1	-1	1
25	-1	-1	-1	1	1	1	-1	-1	1	-1	-1	1	1	-1	-1
26	1	-1	-1	1	1	1	-1	-1	1	-1	-1	-1	-1	1	1
27	-1	1	-1	1	1	1	-1	-1	-1	1	1	-1	1	-1	-1
28	1	1	-1	1	1	1	-1	-1	-1	1	1	1	-1	1	1
29	-1	-1	1	1	1	1	1	1	-1	-1	-1	1	-1	-1	-1
30	1	-1	1	1	1	1	1	1	-1	-1	-1	-1	1	1	1
31	-1	1	1	1	1	1	1	1	1	1	1	-1	-1	-1	-1
32	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

The sign table is given for 2 factor interaction only it is noted that we can generate the full sign table for 2^5 but it is unnecessary to build it because we can find all the interaction term (i.e. 3rd, 4th and 5th order) when we run ANOVA in the later section.

Analysis of Experiment

Analysis of experiment includes 5 basic steps and these are as follows:

1. **Graphical view of the data to have some idea about it**
2. **Create the theoretical model we want to run e.g. 2^3, 2^4, 2^5 etc**
3. **Fitting of model with data**
4. **Test the model assumptions using residual analysis and made the necessary transform if needed and fit the model with modified data.** 5. **Conclusion about the experimental objectives**

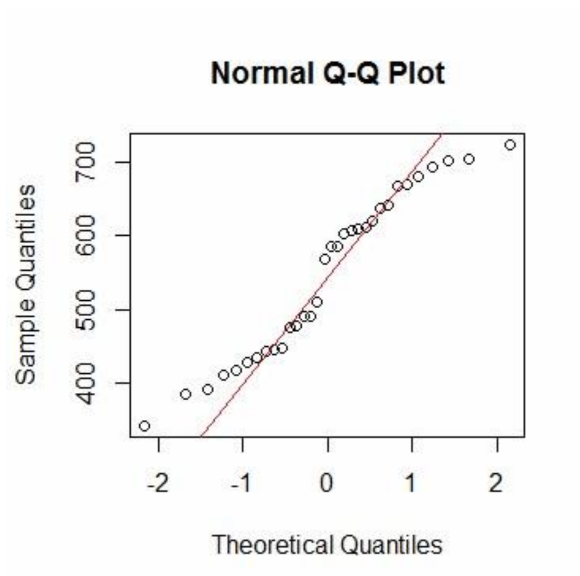
Step1:

The idea to plot the response variable in different ways to find out any trend if present or not in the data, for example the quadratic behavior of the data. It is also important that it may possible that our theoretical model could not detect these trends. Also it is helpful for further analysis to consider these trends for improvement of model.

Graphical View of the data for response variable(Y=Strength)

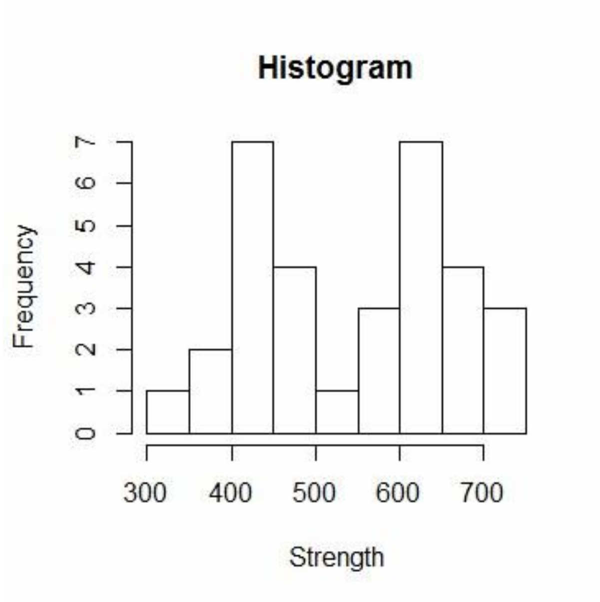
It includes four types of plots: Normal probability plot, Histogram, Box plot and Scatter plot.

Normal Probability Plot (strength)



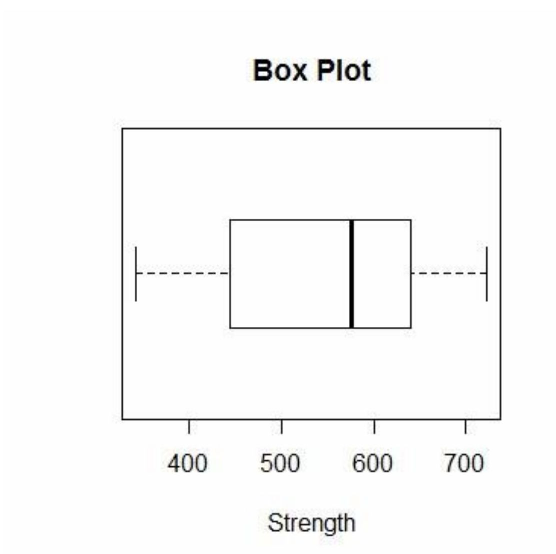
Histogram (Strength)

Minimum Strength = 343.22
Maximum Strength = 722.48

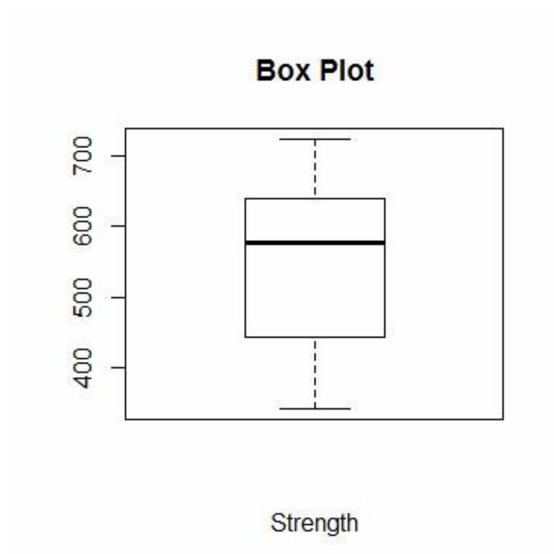


Box plot of response (strength)

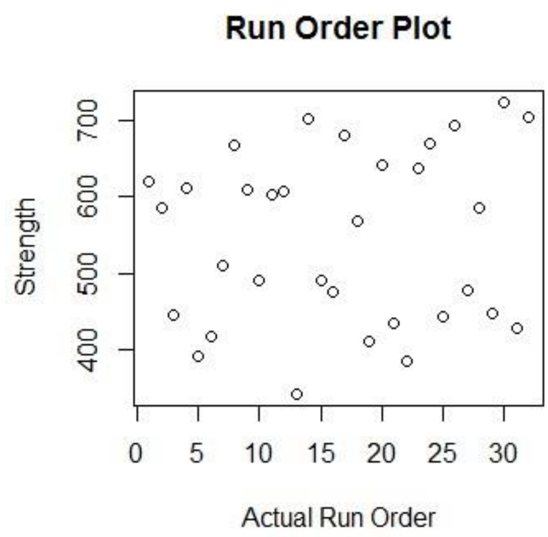
Horizontal view



Vertical view



Scatter Plot Run Order (OD) VS Strength (Y)



The normal probability plot shows the red line of theoretical normal distribution. From this plot it is clear that when we fit the model the residuals may be showed some trend and non-stationary of data. The histogram shows two curves which clearly indicate the separation of data (sub populations or Non-homogenous sample). The first curve is centered approximately at 450 while the other is at 650. The box plot mean line did not center at the central point which show non-normality. The actual run order did not show any trend and confirm the randomness of strength (Y) over time.

Relationship between response and factors

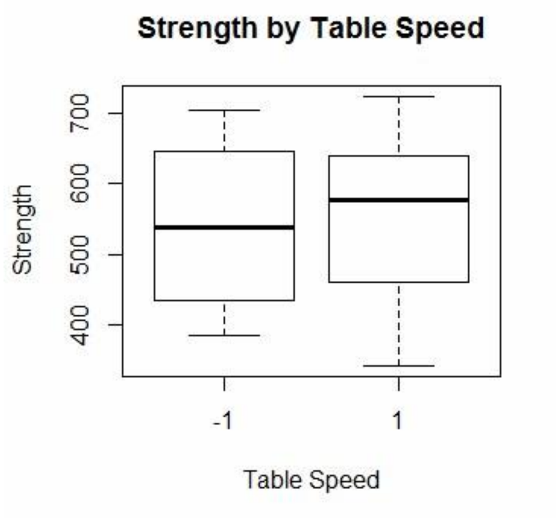
We are interested to see the behavior of response(Y) with factors: speed, rate, grit, direction and batch. The purpose of the graphical view is to check that whether these factors change the effect of average response level. The box plot is used to for this analysis. Five box plots are made for response versus each factor.

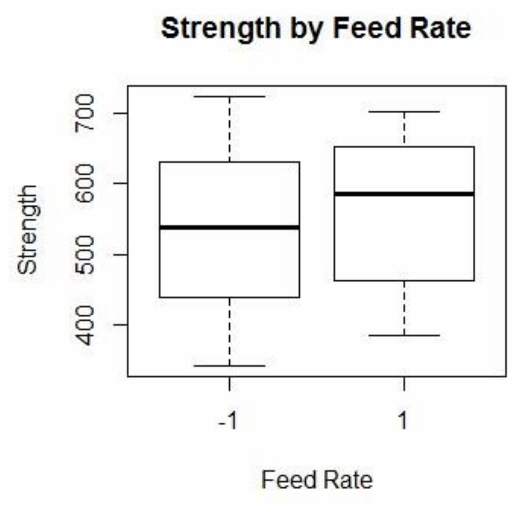
Box plot Strength VS Speed

(High and low value i.e. +1 and -1)

Box plot Strength VS Rate

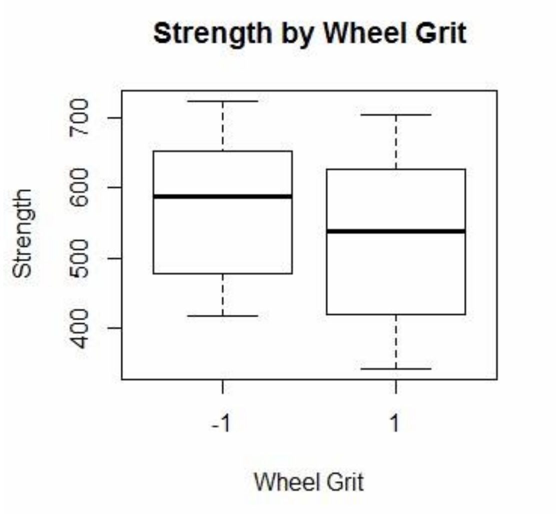
(High and low value i.e. +1 and -1)

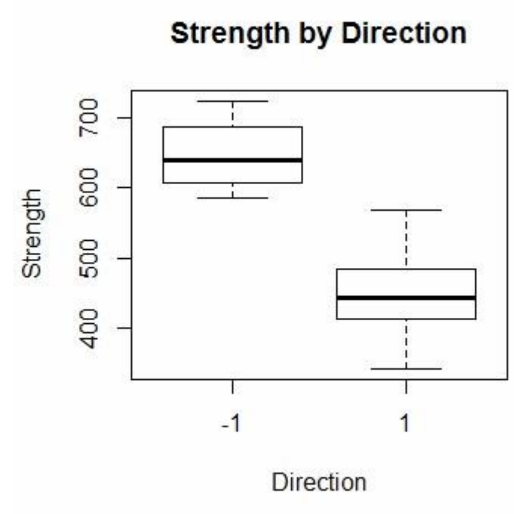




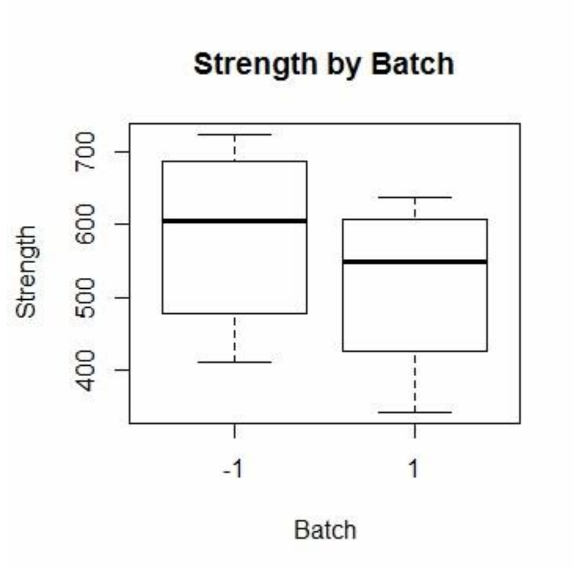
<u>Box plot Strength VS Grit</u>
(High and low value i.e. +1 and -1)

<u>Box plot Strength VS Direction</u>
(High and low value i.e. +1 and -1)





<u>Box plot for Strength VS Batch</u>
(High and low value i.e. +1 and -1)



It is clear from the above box plots for each factor with response that direction and the batch are changed rapidly with average response level while the grit showed some medium level change.

Step-II:

Creation of theoretical model

For a 2^5 full factorial design we can fit the model with mean term, main effect terms and all the interactions terms. The details are given in the following table:

Mean term	(1) = 1
Main effect terms	${}^5C_1$ = 5
Two-factor interaction terms	${}^5C_2$ = 10
Three-factor interaction terms	${}^5C_3$ = 10
Four-factor interaction terms	${}^5C_4$ = 5
Five-factor interaction term	${}^5C_5$ = 1

First we run the model with excluding the 4^{th} and 5^{th} order higher interactions. The obvious reason is the it is rare that these higher order interactions are significant. This assumption is also important because it will enable us to find out the sum of squares for error term. It should be noted that our experiment is 2^5 with single replicate.

Analysis is start with a theoretical model with 26 unknown constants, hoping the data will clarify which of these are the significant main effects and interactions we need for a final model.

Analysis of theoretical Model without 4^{th} and 5^{th} higher order interactions terms

The results are shown below with 26 terms include in the model i.e.

ANOVA TABLE

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
speed	1	894	894	2.817	0.144250	
rate	1	3497	3497	11.018	0.016019	*
grit	1	12664	12664	39.896	0.000735	***
direction	1	315133	315133	992.790	6.79e-08	***
batch	1	33654	33654	106.023	4.90e-05	***
speed:rate	1	4873	4873	15.350	0.007820	**
speed:grit	1	1839	1839	5.793	0.052802	.
speed:direction	1	1637	1637	5.158	0.063573	.
speed:batch	1	465	465	1.465	0.271637	
rate:grit	1	307	307	0.969	0.363033	
rate:direction	1	1973	1973	6.215	0.046974	*
rate:batch	1	199	199	0.627	0.458472	
grit:direction	1	3158	3158	9.950	0.019705	*
grit:batch	1	29	29	0.092	0.771312	
direction:batch	1	1329	1329	4.186	0.086715	.
speed:rate:grit	1	357	357	1.125	0.329695	
speed:rate:direction	1	5896	5896	18.573	0.005039	**
speed:rate:batch	1	145	145	0.456	0.524698	
speed:grit:direction	1	2	2	0.007	0.937573	
speed:grit:batch	1	30	30	0.096	0.767571	
speed:direction:batch	1	545	545	1.716	0.238168	
rate:grit:direction	1	44	44	0.140	0.721008	
rate:grit:batch	1	26	26	0.081	0.786049	
rate:direction:batch	1	167	167	0.527	0.495168	
grit:direction:batch	1	32	32	0.102	0.759969	
Residuals	6	1905	317			---

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

The values of R-Squared and Adjusted R-Squared are: 0.9951 and 0.9748 respectively. As a summary statement this model has high values of R-Squared and Adjusted RSquared but model showed that there are many unnecessary terms that have high p-value > 0.10. Also the main effects Rate, Grit, Direction, Batch and interactions Rate*Direction, Grit*Direction & Speed*Rate*Direction are significant at p-value > 0.05.

Analysis of theoretical Model without 5th order higher interaction term

The results are shown in the below ANOVA TABLE:

ANOVA TABLE

```
                           Df Sum Sq Mean Sq F value Pr(>F)
speed                       1    894     894   1.459 0.4403   rate
1   3497    3497   5.704 0.2524   grit                      1
12664   12664  20.655 0.1379   direction                 1
315133  315133 513.988 0.0281 * batch                     1
33654   33654  54.890 0.0854 . speed:rate                1
4873    4873   7.947 0.2170   speed:grit                1   1839
```

```
1839   2.999 0.3334   speed:direction            1   1637    1637
2.670 0.3496   speed:batch                  1    465     465
0.759 0.5439   rate:grit                    1    307     307
0.501 0.6077   rate:direction               1   1973    1973
3.218 0.3238   rate:batch                   1    199     199
0.325 0.6702   grit:direction               1   3158    3158
5.151 0.2642   grit:batch                   1     29      29
0.048 0.8629   direction:batch              1   1329    1329
2.167 0.3799   speed:rate:grit              1    357     357
0.582 0.5850   speed:rate:direction         1   5896    5896
9.616 0.1986   speed:rate:batch             1    145     145
0.236 0.7121   speed:grit:direction         1      2       2
0.003 0.9626   speed:grit:batch             1     30      30
0.050 0.8606   speed:direction:batch        1    545     545
0.888 0.5189   rate:grit:direction          1     44      44
0.073 0.8325   rate:grit:batch              1     26      26
0.042 0.8717   rate:direction:batch         1    167     167
0.273 0.6935   grit:direction:batch         1     32      32
0.053 0.8560   speed:rate:grit:direction    1    354     354
0.578 0.5862   speed:rate:grit:batch        1    356     356
0.581 0.5854   speed:rate:direction:batch   1     79      79
0.128 0.7810   speed:grit:direction:batch   1    233     233
0.380 0.6484   rate:grit:direction:batch    1    269     269
0.439 0.6273   Residuals                    1    613     613
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Here this model is run just to have an overview of the interactions terms with 4th order. It is clear from the ANOVA table only Direction is the significant effect and all other terms are insignificant which showed that this model not consider for further analysis.

Analysis of theoretical Model with all interaction terms included

This is the result of 2^5 full factorial models (32-terms) with single replicate and we only have the sum of squares and mean squares are available in the ANOVA table because no error degree of freedom is available for F-Test. These sum of squares can be useful for later analysis i.e. for ranking purpose.

ANOVA TABLE

```
                                Df Sum Sq Mean Sq
speed                            1    894     894
rate                             1   3497    3497
grit                             1  12664   12664
direction                        1 315133  315133
batch                            1  33654   33654
speed:rate                       1   4873    4873
speed:grit                       1   1839    1839
speed:direction                  1   1637    1637
speed:batch                      1    465     465
rate:grit                        1    307     307
```

```
rate:direction                   1   1973    1973
rate:batch                       1    199     199
grit:direction                   1   3158    3158
grit:batch                       1     29      29
direction:batch                  1   1329    1329
speed:rate:grit                  1    357     357
speed:rate:direction             1   5896    5896
speed:rate:batch                 1    145     145
speed:grit:direction             1      2       2
speed:grit:batch                 1     30      30

speed:direction:batch            1    545     545
rate:grit:direction              1     44      44
rate:grit:batch                  1     26      26
rate:direction:batch             1    167     167
grit:direction:batch             1     32      32
speed:rate:grit:direction        1    354     354
speed:rate:grit:batch            1    356     356
speed:rate:direction:batch       1     79      79
speed:grit:direction:batch       1    233     233
rate:grit:direction:batch        1    269     269
speed:rate:grit:direction:batch  1    613     613
```

Analysis continue with model that exclude the 4th and 5th order interaction terms

As at this step we did not know which subset of main effects and interaction terms is useful which constitutes a good model. The solution to this problem that we must control the entry or removal of independent variable (main effects and interaction terms) from the theoretical model. A suggested solution is to enter or removed the variables in the model "Step By Step" it is know as the ***Stepwise Regression***. In this problem we use the backward selection method which begins with all main effects and interaction terms and at each step removes the ***least useful predictor***. Generally at each step the coefficient of the entered variable is zero is tested using its F-Statistic and descion made about its removal based on generic function calculating ***Akaike's 'An Information Criterion' (AIC)*** for one or several fitted model objects for which a log-likelihood value can be obtained. When comparing models fitted by maximum likelihood to the same data, the model with smaller AIC is declared to be better the fit. We stop the stepping when the established criteria no longer exist. The results of stepwise regression with AIC are given below:

Summary of Result for Stepwise AIC

```
                      Df Sum Sq Mean Sq  F value   Pr(>F)
speed                  1    894     894    5.187 0.041862 *    rate
1   3497    3497   20.284 0.000722 *** grit                   1
12664   12664   73.454 1.84e-06 *** direction            1
315133  315133 1827.837 1.74e-14 *** batch                1
33654   33654  195.200 8.73e-09 *** speed:rate           1
4873    4873   28.262 0.000183 *** speed:grit           1   1839
1839   10.665 0.006756 **  speed:direction       1   1637    1637
9.496 0.009510 **  speed:batch           1    465     465
2.697 0.126440     rate:grit             1    307     307
```

```
1.783 0.206520     rate:direction         1   1973    1973
11.442 0.005442 **  rate:batch             1    199     199
1.155 0.303616     grit:direction         1   3158    3158
18.319 0.001069 **  direction:batch        1   1329    1329
7.707 0.016767 *   speed:rate:grit        1    357     357
2.071 0.175699     speed:rate:direction   1   5896    5896
34.196 7.87e-05 *** speed:rate:batch       1    145     145
0.839 0.377623     speed:direction:batch  1    545     545
3.159 0.100855     rate:direction:batch   1    167     167
0.970 0.344021     Residuals             12   2069     172
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

It is clear from the summarized result of stepwise regression with AIC and ignoring the interaction terms which have p-value greater then 0.05 we have left only the model which includes 12 terms. After running stepwise regression we got the total 19 main effects and interaction terms in the model of which 7 interaction terms are not significant at 5% level. The excluded interaction terms are two and three factors i.e. speed*batch, rate*grit, rate*batch, speed*rate*grit, speed*rate*batch, speed*direction*batch, rate*direction*batch.

Step-III:

Fitting of Model

Final Model

The final model we have for this problem is:

$$(Response\ strenght{:}Y) = Speed + Rate + Grit + Direction + Batch + (Speed \times Rate) + (Speed \times Grit) + (Speed \times Direction) + (Rate \times Direction) + (Grit \times Direction) + (Direction \times Batch) + (Speed \times Rate \times Direction) + \epsilon(Error)$$

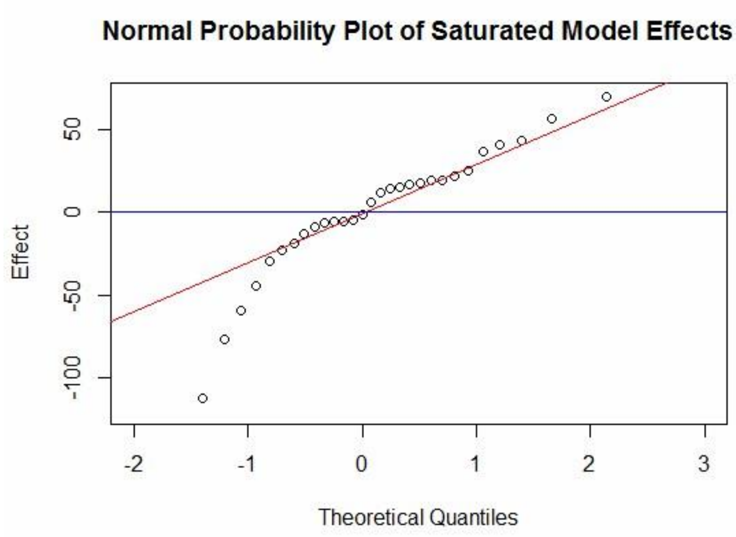
ANOVA Table

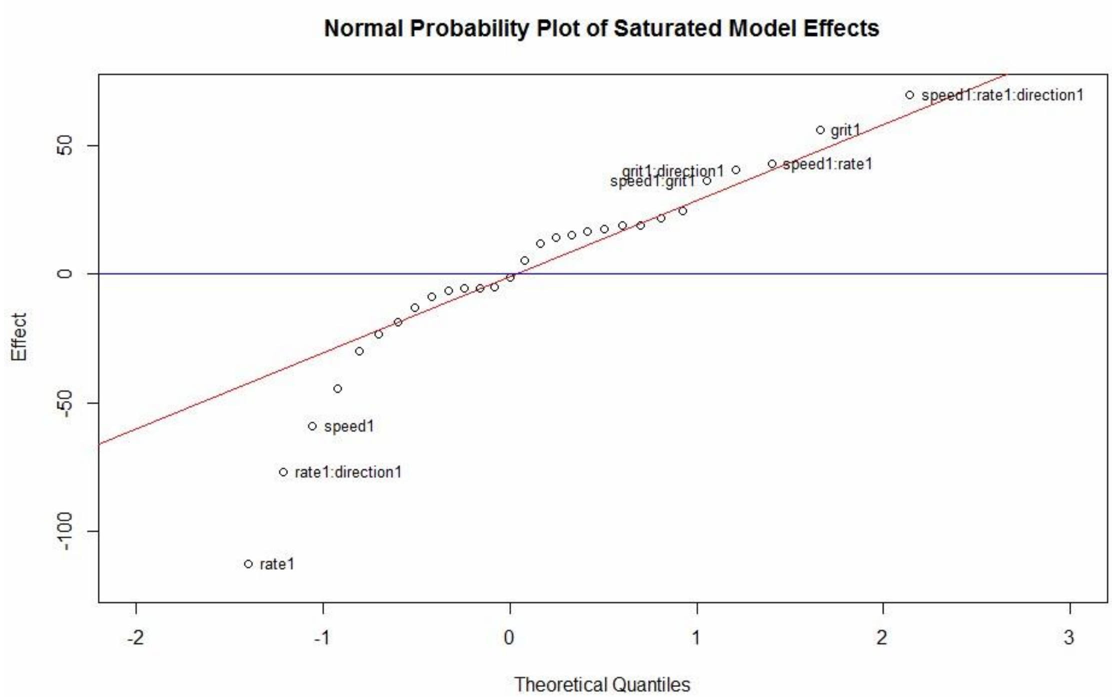
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
speed	1	894	894	3.994	0.060170	.
rate	1	3497	3497	15.619	0.000855	***
grit	1	12664	12664	56.560	4.14e-07	***
direction	1	315133	315133	1407.439	< 2e-16	***
batch	1	33654	33654	150.304	1.80e-10	***
speed:rate	1	4873	4873	21.762	0.000169	***
speed:grit	1	1839	1839	8.212	0.009896	**
speed:direction	1	1637	1637	7.312	0.014065	*
rate:direction	1	1973	1973	8.810	0.007897	**
grit:direction	1	3158	3158	14.106	0.001338	**
direction:batch	1	1329	1329	5.935	0.024859	*
speed:rate:direction	1	5896	5896	26.331	5.93e-05	***
Residuals	19	4254	224			-

--
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

The above table shows that all the main effects and interaction terms are significant for the final model with high values of R-Squared and Adjusted R-Squared i.e. 0.9891 & 0.9822.

It is helpful to use both numerical and graphical methods for descion related to which term to kept in the model. In order to reconfirm the results we obtained from Stepwise Regression we will make the Normal Probability plot to have an idea of all the effects along the theoretical regression line. For this consider the ANOVA Table results run for the full 2^5 factorial with 32 total combinations and we have 31 effects. The idea is that those effects that deviate significantly from the theoretical probability line are to be considered as significant.





In the above normal probability plot the largest effects direction and batch are not shown; the purpose is to reduce the scale on y-axis to have better idea of other effects. Also as we know that these two effects are significant so these are excluded form the computation of normal probability plot. The effects that are clustered near OR at the normal line (red color line) are considered to be non-significant and those which are away from it are significant. From above plot most of the effects are close to center line and follows the normal model. Further the significant effects are labeled with their names. After analyzing the probability plot it is concluded that the decision about the entrance OR removal of effects are the same as obtained from stepwise regression. At this stage model is accepted for further analysis.

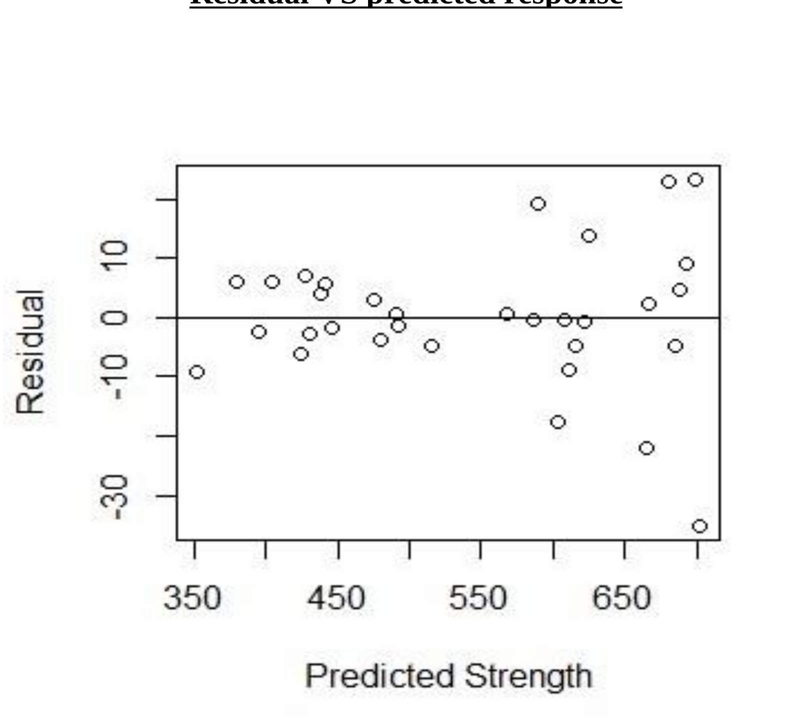
Step-IV:

Residual Analysis (Assumptions of Model)

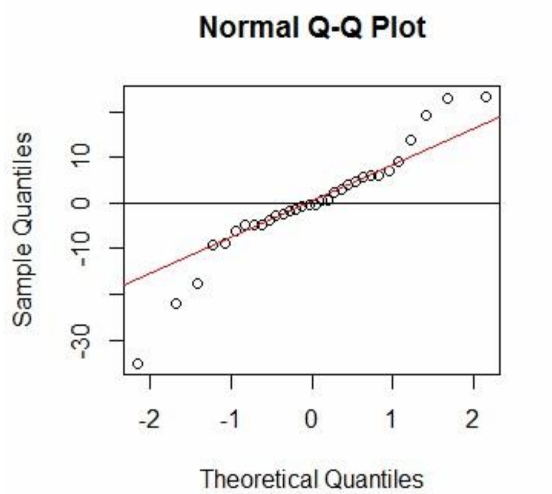
For diagnosing the assumptions of model the residual analysis is performed by viewing several graphs. It includes the following graphs:

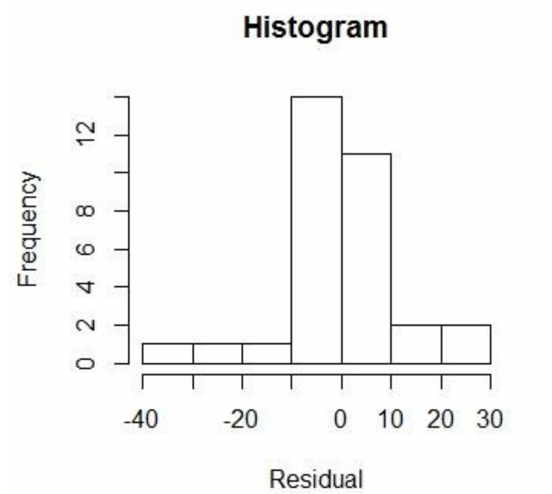
1. Residual VS predicted response
2. Normal probability plot
3. Histogram
4. Box plot
5. Scatter plot

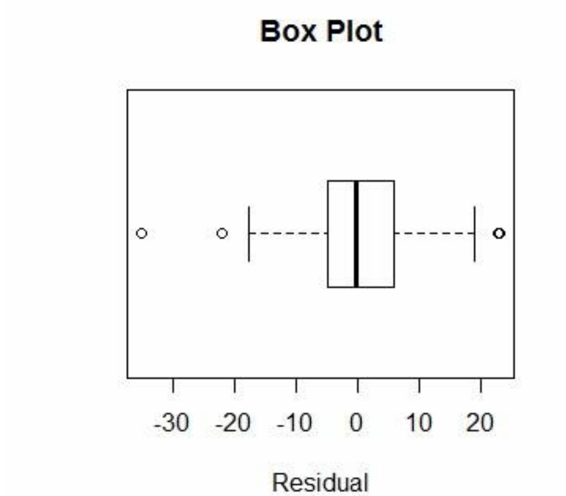
Residual VS predicted response

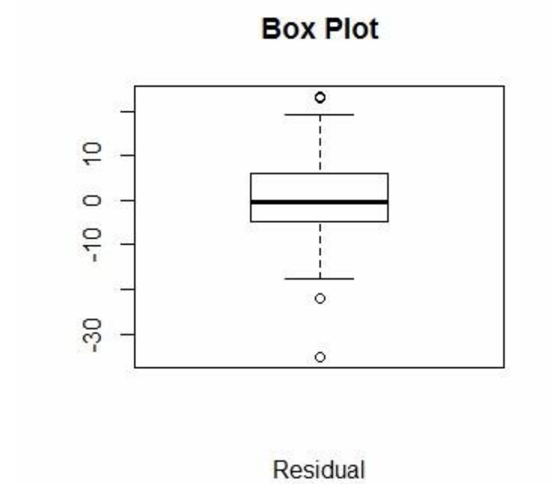


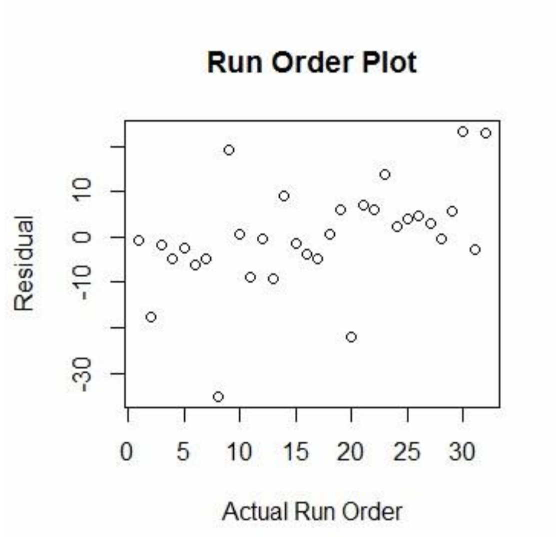
It is clear from the residual plot that the residual appear to spread out more (a funnel type shape) with higher values of predicted strength. It shows the non-normality of common variance. i.e $\epsilon \sim (N(0, \sigma^2))$.


Normal Q-Q Plot
Sample Quantiles
-30
-10 0 10
-2 -1 0 1 2
Theoretical Quantiles


Histogram
Frequency
0 2 4 6 8 12
-40 -20 0 10 20 30
Residual


Box Plot
-30 -20 -10 0 10 20
Residual


Box Plot
-30 -10 0 10
Residual



From the normal probability plot revealed that when the theoretical quantile is greater than one the residual tends to non-normality i.e. residual quantiles are getting away from the theoretical probability line. The Histogram also confirms the nonstationary behavior of the residuals and it frequency goes up to 13 for data range between -10 to 10 while for all other bins it remains at approximately 2. Similarly box plot shows variations from central value in both views i.e. horizontally and vertically with identifying the outliers. Also the scatter plot is obtained with actual randomized order versus residuals. It confirms the tendency of non-normality of residuals by showing some very high and low values with patterns.

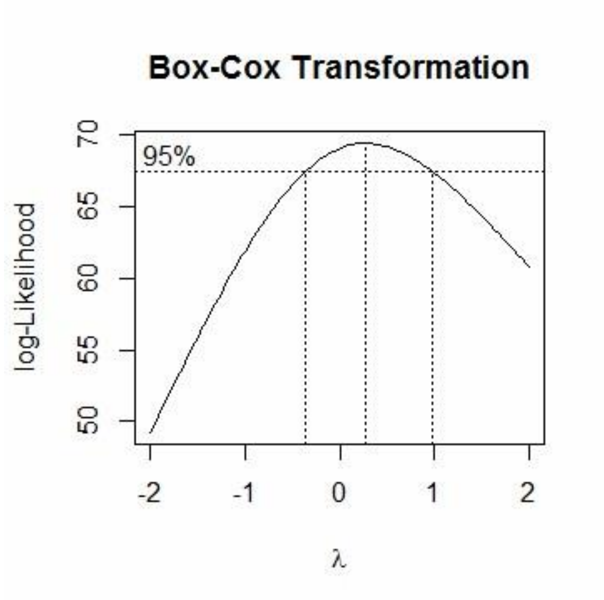
Hence there is no doubt about the non-normality of residuals. A remedy for situation is to make the transformation of the response (strength=Y) variable. There are many transformations that are available in the literature like the Log- Transformation, Square-Root Transformation etc. Among these the Box-Cox transformation is used for ‘Y’ variable.

Transformation of Response Variable (Y) The Box-Cox Transformation is defined as:

$$\frac{(Y_i)^{\lambda} - 1}{\lambda\left[\left(\prod_{i=1}^{n} Y_i\right)^{\frac{1}{n}}\right]^{\lambda - 1}}$$

where Y_i is the ith response and λ is parameter

In Box-Cox transformation an optimum value of λ is calculated that maximize the negative log-likelihood. The result is shown as:




```
The optimum is found at lambda =0.26 but Use lambda = 0.2
to match output in the page. lambda = 0.2
```

After running the Box-Cox transformation package in R we obtain the value of λ= 0.26 but it is recommended that use λ=0.2 for better results. The values of transformed variable i.e. ***Y-Transformed = Y****

Thus Response (Strength) ⇨ Y   Transformed (newstrength) ⇨⇨⇨ Y*

```
 [1] 1779.446 1813.888 1797.410 1768.034 1798.662
 [6] 1746.658 1789.874 1770.013 1601.773 1584.461
[11] 1587.988 1679.086 1550.015 1509.458 1531.175
[16] 1601.655 1715.599 1727.764 1718.519 1743.065
[21] 1695.141 1696.058 1710.415 1716.483 1547.925
[26] 1538.092 1518.269 1621.992 1486.870 1422.313
[31] 1478.527 1552.316
```

<u>Analysis after transformation</u>

The results with transformed variable are shown in the following table:

<u>ANOVA Table</u>

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
speed	1	1055	1055	5.368	0.031824 *
rate	1	4294	4294	21.854	0.000165 ***
grit	1	14778	14778	75.208	4.95e-08 ***
direction	1	315042	315042	1603.291	< 2e-16 ***
batch	1	32528	32528	165.541	7.91e-11 ***

```
speed:rate            1   6532    6532   33.244   1.48e-05 ***
speed:grit            1   1728    1728    8.796   0.007940 **
speed:direction       1   1652    1652    8.408   0.009185 **
rate:direction        1   2619    2619   13.329   0.001700 **
grit:direction        1   5172    5172   26.323   5.94e-05 ***
direction:batch       1    119     119    0.604   0.446556
speed:rate:direction  1   7366    7366   37.485   6.93e-06 ***
Residuals            19   3733     196
 ---
 Signif. codes:
 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

It is clear from the above ANOVA table that after transformation the term (direction*batch) becomes insignificant with high p-value i.e. 0.446. Except this interaction all other terms are significant at 5% level of significant.

Now reconsidered the model without (direction*batch) interaction term. There are total 11 terms remained in the model and the ***Linear Model*** is:

$$\left(\textit{Transformed Strength } Y^*\right) = \textit{Speed} + \textit{Rate} + (\textit{Speed} \times \textit{Rate}) + \textit{Grit} + (\textit{Speed} \times \textit{Grit}) + \textit{Direction} + (\textit{Direction} \times \textit{Speed}) + (\textit{Rate} \times \textit{Direction}) + (\textit{Speed} \times \textit{Rate} \times \textit{Direction}) + (\textit{Grit} \times \textit{Direction}) + (\textit{Batc})$$

Descriptive Statistics for Residuals

Min	1Q	Median	3Q	Max
-29.373	-3.978	-1.298	6.800	20.994

Results of T-test

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 1649.967      2.453 672.532  < 2e-16 *** s
5.741      2.453   2.340 0.029750 *   r
11.584      2.453   4.722 0.000131 *** sr
14.288      2.453   5.824 1.07e-05 *** g           -
21.490      2.453  -8.759 2.79e-08 *** sg           -
7.349      2.453  -2.996 0.007146 **  d           -
99.222      2.453 -40.443  < 2e-16 *** ds
7.185      2.453   2.929 0.008302 **  rd
9.047      2.453   3.688 0.001459 **  srd
15.172      2.453   6.184 4.85e-06 *** gd          -
12.714      2.453  -5.182 4.53e-05 *** b           -
31.883      2.453 -12.996 3.28e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

t-test shows that all the estimates are significant at 5% level of significance. The value of Multiple-Squared and Adjusted R-Squared with F-test value are given below:

Summary

Residual standard error:	13.88
Multiple R-squared:	0.9903
Adjusted R-squared:	0.9849
F-statistic: 185.4 and p-value: < 2.2e-16	

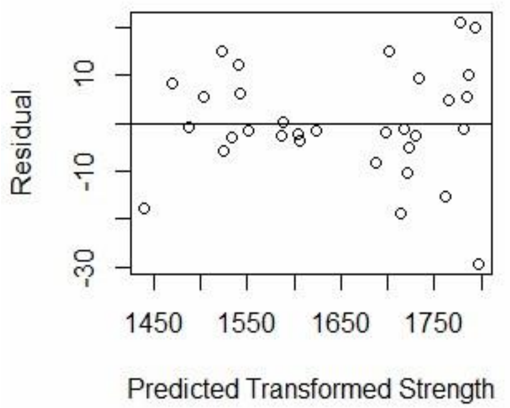
Multiple R-Squared and Adjusted R-Squared show the high values and the F-test for the equality of means revealed enough high value for the rejection of null hypothesis.

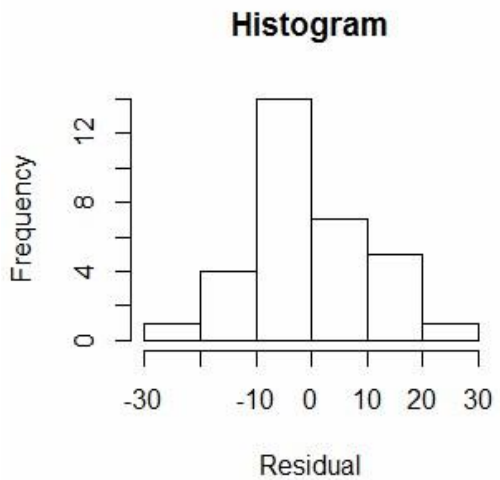
Checking the model Assumptions

Residual analysis with transformed response variable

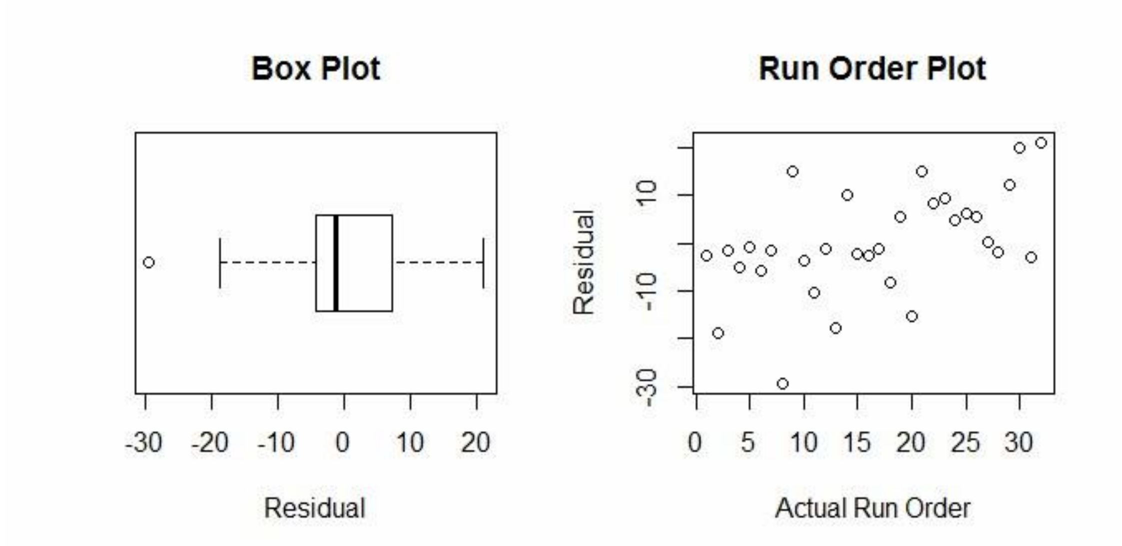
The residual analysis includes these types of plots: Residual versus predicted response, Histogram, Box plot and Scatter plot.

Residual Plot





It is obvious from the plot of residuals that they showed better randomness behavior as earlier. There is one or two outliers presented in the plot but other than these values the residuals show stationarity of values. Thus the transformation resolved the issue of increased variance with respect to increased strength. The histogram shows quite good normal curvature behavior also the existence of sub-population did not represent in the histogram after transformation.



The Box plot manifests only single outlier in the residual when compared with un-transformed residuals and shows the reduction of outliers. Finally the scatter plot gives clear indication of stationary residuals and did not confirm any time dependent patterns. Over all after the analysis it is concluded that we accept our model with transformed response variable. Also it is summarize that model contain not any unnecessary term(s) and did not violate any of the basic assumption of model.

Step-V:

Conclusion about the experimental objectives

From this analysis it is clear by the estimates of effects that the "Direction" is the most important factor in the analysis. Secondly the factor "Batch" is more crucial which is followed by the "Wheel Grit". On other hand the "Feed Rate" and "Table Speed" plays most important role nearly every significant interaction term but both proves their self as least important main effects through out the analysis. It is also clear that the higher interaction effects are very difficult to explain thus we restricted our analysis for 3^{rd} order interaction terms only. The model is initially accepted with 12 terms including the 3^{rd} order interaction which is confirmed by the *stepwise regression* and normal probability plot with labeling. But at final stage we accept model with transformed response (strength) with 11 terms included in the model.

Recommendations:

Since our model includes the two qualitative variables i.e. Direction and Batch so it is recommended that the optimal analysis may be perform in order to find the optimal solution. Further for transformation purpose analysis may be performed with log transformation i.e. $Y^{*}=logY$ which may give improved model

APENDIX – A

Working Directory

```
## To check the working directory getwd()
 ………………………………
 setwd("E:/Stats/Mphil_First_Semester/Experimental_Design") getwd()
 [1] "E:/Stats/Mphil_First_Semester/Experimental_Design"
```

Data Matrix Visualization

```
## First method to visualize the data matrix
> mydata=read.table("2^5_Fact.txt")
> mydata

## Second method to show the data matrix
> m = matrix(scan("2^5_Fact.txt",skip=1),ncol=7,byrow=T)
Read 224 items
> m
```

Saving the Response (Y) and Order variable (OD)

```
> strength = m[,6]
> order = m[,7]
```

Declaring the factors and saving

```
> speed = as.factor(m[,1])
> rate = as.factor(m[,2])
> grit = as.factor(m[,3])
> direction = as.factor(m[,4]) >
batch = as.factor(m[,5])
```

Declaring numeric variables and interactions with two factors only

```
> s = m[,1]
> r = m[,2]
> g = m[,3]
> d = m[,4]
> b = m[,5]
> db = d*b
> gd = g*d
> gb = g*b
> rg = r*g
> rd = r*d
> rb = r*b
> sr = s*r
```

> sg = s*g
> ds = s*d
> sb = s*b

Data matrix with labeled variable and sign table of two factor interactions

>df = data.frame(speed,rate,grit,direction,batch,strength,order, s,r,g,d,b,db,gd,gb,rg,rd,rb,sr,sg,ds,sb)
Data matrix
> df[,1:7]
Two Factor interactions sign table
> df[,8:22]

Step 1:

Normal Probability Plot for response(strength)

> qqnorm(strength)
> qqline(strength, col = 2)

Histogram (Strength)

First we find the min and max values
> min(strength)
[1] 343.22
> max(strength)
[1] 722.48
>hist(strength, main="Histogram", xlab="Strength")

Box plot of response (strength)

##Horizontal view
> boxplot(strength, horizontal=TRUE, main="Box Plot", xlab="Strength") ##Vertical view
> boxplot(strength, horizontal=FALSE, main="Box Plot", xlab="Strength")

Scatter Plot Run Order VS Strength

>plot(order, strength, xlab="Actual Run Order", ylab="Strength", main="Run Order Plot")

Box plot for Strength VS Speed

>boxplot(strength~speed, data=df, main="Strength by Table Speed", xlab="Table Speed",ylab="Strength")

Box plot for Strength VS Rate

>boxplot(strength~rate, data=df, main="Strength by Feed Rate", xlab="Feed Rate",ylab="Strength")

Box plot for Strength VS Grit

>boxplot(strength~grit, data=df, main="Strength by Wheel Grit", xlab="Wheel Grit",ylab="Strength")

Box plot Strength VS Direction

>boxplot(strength~direction, data=df, main="Strength by Direction", xlab="Direction",ylab="Strength")

Box plot for Strength VS Batch

>boxplot(strength~batch, data=df, main="Strength by Batch", xlab="Batch",ylab="Strength")

Step-II:

Theoretical Model:

Theoretical Model (Excluding the 4th and 5th order higher interactions)

> model = aov(strength~(speed+rate+grit+direction+batch)^3,data=df)
> summary(model)

Theoretical Model (Excluding the 5th order higher interactions)

> model2 = aov(strength~(speed+rate+grit+direction+batch)^4,data=df)
> summary(model2)

Theoretical Model (Considering All Interactions Terms)

> model3 = aov(strength~(speed+rate+grit+direction+batch)^5,data=df)
> summary(model3)

Continue with the model that excluded the 4th and 5th order higher interaction terms

Fitting the model:

Stepwise regression based on AIC

> sreg = step(model,direction="backward")

```
Start:  AIC=182.76
strength ~ (speed + rate + grit + direction + batch)^3
                       Df Sum of Sq    RSS    AIC
- speed:grit:direction  1       2.1 1906.6 180.80
- rate:grit:batch       1      25.6 1930.1 181.19
- speed:grit:batch      1      30.4 1934.9 181.27
- grit:direction:batch  1      32.5 1937.0 181.30
- rate:grit:direction   1      44.5 1949.0 181.50
<none>                              1904.5 182.76
- speed:rate:batch      1     144.7 2049.2 183.10
- rate:direction:batch  1     167.3 2071.8 183.46
- speed:rate:grit       1     357.0 2261.6 186.26
- speed:direction:batch 1     544.6 2449.1 188.81
- speed:rate:direction  1    5895.6 7800.1 225.88
```

```
Step:  AIC=180.8
strength ~ speed + rate + grit + direction + batch + speed:rate +
speed:grit + speed:direction + speed:batch + rate:grit +      rate:direction
+ rate:batch + grit:direction + grit:batch +      direction:batch +
speed:rate:grit + speed:rate:direction +      speed:rate:batch +
speed:grit:batch + speed:direction:batch +      rate:grit:direction +
rate:grit:batch + rate:direction:batch +      grit:direction:batch

                        Df Sum of Sq    RSS    AIC
- rate:grit:batch        1      25.6 1932.2 179.22
- speed:grit:batch       1      30.4 1937.0 179.30
- grit:direction:batch   1      32.5 1939.1 179.34
- rate:grit:direction    1      44.5 1951.1 179.53
<none>                               1906.6 180.80
- speed:rate:batch       1     144.7 2051.4 181.14
- rate:direction:batch   1     167.3 2074.0 181.49
- speed:rate:grit        1     357.0 2263.7 184.29
- speed:direction:batch  1     544.6 2451.2 186.84
- speed:rate:direction   1    5895.6 7802.3 223.89

Step:  AIC=179.22
strength ~ speed + rate + grit + direction + batch + speed:rate +
speed:grit + speed:direction + speed:batch + rate:grit +      rate:direction
+ rate:batch + grit:direction + grit:batch +      direction:batch +
speed:rate:grit + speed:rate:direction +      speed:rate:batch +
speed:grit:batch + speed:direction:batch +      rate:grit:direction +
rate:direction:batch + grit:direction:batch

                        Df Sum of Sq    RSS    AIC
- speed:grit:batch       1      30.4 1962.6 177.72
- grit:direction:batch   1      32.5 1964.7 177.75
- rate:grit:direction    1      44.5 1976.7 177.95
<none>                               1932.2 179.22
- speed:rate:batch       1     144.7 2076.9 179.53
- rate:direction:batch   1     167.3 2099.5 179.88
- speed:rate:grit        1     357.0 2289.3 182.65
- speed:direction:batch  1     544.6 2476.8 185.17
- speed:rate:direction   1    5895.6 7827.8 221.99

Step:  AIC=177.72
strength ~ speed + rate + grit + direction + batch + speed:rate +
speed:grit + speed:direction + speed:batch + rate:grit +      rate:direction
+ rate:batch + grit:direction + grit:batch +      direction:batch +
speed:rate:grit + speed:rate:direction +      speed:rate:batch +
speed:direction:batch + rate:grit:direction +      rate:direction:batch +
grit:direction:batch

                        Df Sum of Sq    RSS    AIC
- grit:direction:batch   1      32.5 1995.0 176.25
- rate:grit:direction    1      44.5 2007.1 176.44
<none>                               1962.6 177.72
- speed:rate:batch       1     144.7 2107.3 178.00
```

```
- rate:direction:batch   1     167.3 2129.9 178.34
- speed:rate:grit        1     357.0 2319.6 181.07
- speed:direction:batch  1     544.6 2507.2 183.56
- speed:rate:direction   1    5895.6 7858.2 220.11

Step:  AIC=176.25
strength ~ speed + rate + grit + direction + batch + speed:rate +
speed:grit + speed:direction + speed:batch + rate:grit +      rate:direction
+ rate:batch + grit:direction + grit:batch +      direction:batch +
speed:rate:grit + speed:rate:direction +      speed:rate:batch +
speed:direction:batch + rate:grit:direction +      rate:direction:batch

                        Df Sum of Sq    RSS    AIC
- grit:batch             1      29.4 2024.4 174.71
- rate:grit:direction    1      44.5 2039.5 174.95
<none>                               1995.0 176.25
- speed:rate:batch       1     144.7 2139.8 176.49
- rate:direction:batch   1     167.3 2162.4 176.82
- speed:rate:grit        1     357.0 2352.1 179.51
- speed:direction:batch  1     544.6 2539.6 181.97
- speed:rate:direction   1    5895.6 7890.7 218.25

Step:  AIC=174.71
strength ~ speed + rate + grit + direction + batch + speed:rate +
speed:grit + speed:direction + speed:batch + rate:grit +      rate:direction
+ rate:batch + grit:direction + direction:batch +      speed:rate:grit +
speed:rate:direction + speed:rate:batch +      speed:direction:batch +
rate:grit:direction + rate:direction:batch

                        Df Sum of Sq    RSS    AIC
- rate:grit:direction    1      44.5 2068.9 173.41
<none>                               2024.4 174.71
- speed:rate:batch       1     144.7 2169.1 174.92
- rate:direction:batch   1     167.3 2191.7 175.25
- speed:rate:grit        1     357.0 2381.4 177.91
- speed:direction:batch  1     544.6 2569.0 180.34
- speed:rate:direction   1    5895.6 7920.0 216.37

Step:  AIC=173.41
strength ~ speed + rate + grit + direction + batch + speed:rate +
speed:grit + speed:direction + speed:batch + rate:grit +      rate:direction
+ rate:batch + grit:direction + direction:batch +      speed:rate:grit +
speed:rate:direction + speed:rate:batch +
    speed:direction:batch + rate:direction:batch

                        Df Sum of Sq    RSS    AIC
<none>                               2068.9 173.41
- speed:rate:batch       1     144.7 2213.6 173.57
- rate:direction:batch   1     167.3 2236.2 173.90
- speed:rate:grit        1     357.0 2425.9 176.50
- speed:direction:batch  1     544.6 2613.5 178.89
- grit:direction         1    3158.3 5227.2 201.07
- speed:rate:direction   1    5895.6 7964.5 214.54
```

Summary of Result for AIC
> summary(sreg)

Model after stepwise regression and removing non-significant terms from the model (It is our final model)
>newmodel = aov(formula = strength ~ speed + rate + grit + direction + batch + speed:rate
+ speed:grit + speed:direction + rate:direction + grit:direction + direction:batch
+ speed:rate:direction, data = df)
>summary(newmodel)

Normal probability plot (consider model-3 with all treatment effects)

NOTE: **There is an alternate method to find the rank i.e. Just create the Sum of Squares column into vetor and divide each entry of this vector by the Total of Sum of Squares.**
> model3 = aov(strength~(speed+rate+grit+direction+batch)^5,data=df)
> summary(model3)

```
                                 Df  Sum Sq Mean Sq
speed                            1    894     894 rate
1   3497    3497 grit                             1
12664   12664 direction                        1
315133  315133 batch                            1
33654   33654 speed:rate                       1
4873    4873 speed:grit                      1   1839
1839 speed:direction                  1   1637    1637
speed:batch                      1    465     465
rate:grit                        1    307     307
rate:direction                   1   1973    1973
rate:batch                       1    199     199
grit:direction                   1   3158    3158
grit:batch                       1     29      29
direction:batch                  1   1329    1329
speed:rate:grit                  1    357     357
speed:rate:direction             1   5896    5896
speed:rate:batch                 1    145     145
speed:grit:direction             1      2       2
speed:grit:batch                 1     30      30
speed:direction:batch            1    545     545
rate:grit:direction              1     44      44
rate:grit:batch                  1     26      26
rate:direction:batch             1    167     167
grit:direction:batch             1     32      32
speed:rate:grit:direction        1    354     354
speed:rate:grit:batch            1    356     356
speed:rate:direction:batch       1     79      79
speed:grit:direction:batch       1    233     233
rate:grit:direction:batch        1    269     269
speed:rate:grit:direction:batch  1    613     613
```

Saving effects (treatment effects) in a vector but removing the intercept

> model3ef=model3$effects
> model3ef=model3ef[-1]

Sorting effects and their labels **Sort:**
> sort_ef = model3ef[order(model3ef)]
>model3lab=names(sort_ef)

```
direction1                                batch1
                       -561.366770                               -183.450016
                             grit1                  speed1:rate1:direction1
                       -112.534276                                -76.782958
                             rate1                           rate1:direction1
                        -59.137108                                -44.415145
                            speed1                speed1:direction1:batch1
                        -29.905314                                -23.336292
          speed1:rate1:grit1:batch1                rate1:direction1:batch1
                        -18.874448                                -12.934751
     speed1:rate1:direction1:batch1                  rate1:grit1:direction1
                         -8.872422                                 -6.669785
            grit1:direction1:batch1                            grit1:batch1
                         -5.697513                                 -5.418206
                 rate1:grit1:batch1                 speed1:grit1:direction1
                         -5.057581                                 -1.454872
                speed1:grit1:batch1                      speed1:rate1:batch1
                          5.510130                                 12.029654
                      rate1:batch1       speed1:grit1:direction1:batch1
                         14.112084                                 15.257597
      rate1:grit1:direction1:batch1                              rate1:grit1
                         16.413716                                 17.534480
      speed1:rate1:grit1:direction1                       speed1:rate1:grit1
                         18.821415                                 18.895661
                     speed1:batch1 speed1:rate1:grit1:direction1:batch1
                         21.564989                                 24.761112
                  direction1:batch1                        speed1:direction1
                         36.453122                                 40.462418
                       speed1:grit1                         grit1:direction1
                         42.880723                                 56.199079
                       speed1:rate1
                         69.803814
```

Lable:
> model3lab=names(sort_model3ef)
> large=c(1,2
> model3lab

```
[1] "direction1"                          "batch1"
 [3] "grit1"                               "speed1:rate1:direction1"
 [5] "rate1"                               "rate1:direction1"
 [7] "speed1"                              "speed1:direction1:batch1"
 [9] "speed1:rate1:grit1:batch1"           "rate1:direction1:batch1"
[11] "speed1:rate1:direction1:batch1"      "rate1:grit1:direction1"
[13] "grit1:direction1:batch1"             "grit1:batch1"
```

```
[15] "rate1:grit1:batch1"                     "speed1:grit1:direction1"
[17] "speed1:grit1:batch1"                    "speed1:rate1:batch1"
[19] "rate1:batch1"
"speed1:grit1:direction1:batch1"
[21] "rate1:grit1:direction1:batch1"          "rate1:grit1"
[23] "speed1:rate1:grit1:direction1"          "speed1:rate1:grit1"
[25] "speed1:batch1"
"speed1:rate1:grit1:direction1:batch1"
[27] "direction1:batch1"                      "speed1:direction1"
[29] "speed1:grit1"                           "grit1:direction1"
[31] "speed1:rate1"
```

Ignoring the two largest effects i.e. Direction and Batch

```
> model1ef=model1ef[-large]
> model1lab=model1lab[-large]
```

Generating the theoretical normal probability plots

Ranking:

```
>ip = ppoints(length(sort_ef))
> ip
```

[1] 0.01612903 0.04838710 0.08064516 0.11290323 0.14516129
[6] 0.17741935 0.20967742 0.24193548 0.27419355 0.30645161
[11] 0.33870968 0.37096774 0.40322581 0.43548387 0.46774194
[16] 0.50000000 0.53225806 0.56451613 0.59677419 0.62903226
[21] 0.66129032 0.69354839 0.72580645 0.75806452 0.79032258
[26] 0.82258065 0.85483871 0.88709677 0.91935484 0.95161290
[31] 0.98387097

Normal Porbability

\> zp = qnorm(ip)
\> zp
[1] -2.14119812 -1.66069761 -1.40074506 -1.21123213
[5] -1.05741423 -0.92524456 -0.80754104 -0.70009021
[9] -0.60017878 -0.50593365 -0.41598722 -0.32929135
[13] -0.24500622 -0.16242937 -0.08094729 0.00000000
[17] 0.08094729 0.16242937 0.24500622 0.32929135
[21] 0.41598722 0.50593365 0.60017878 0.70009021
[25] 0.80754104 0.92524456 1.05741423 1.21123213
[29] 1.40074506 1.66069761 2.14119812

Plot of Ranking VS Normal Probability

Note: After ignoring the Direction and Batch we get this range (min= -118.345, max= 69.803)

```
>plot(zp,sort_ef, ylim=c(-120,70), xlim=c(-2,3),ylab="Effect",
xlab="Theoretical Quantiles", main="Normal Probability Plot of
Saturated Model Effects")
> qqline(sort_ef, col=2)
> abline(h=0, col=4)
```

Adding labels for 10 largest effects in the Normal Probability Plot

```
> text(-2,90,"Direction and Batch not shown",pos=4)
> small = c(6:(length(sort_ef)-3))
> small2 = c((length(sort_ef)-4):(length(sort_ef)-3))
> text(zp[-small],sort_ef[-small],label=model1lab[-small],pos=4,cex=0.8)
> text(zp[small2],sort_ef[small2],label=model1lab[small2],pos=2,cex=0.8)
```

Step – IV: Checking the model assumptions

Plot residuals versus predicted response

```
> plot(predict(newmodel),newmodel$residuals,ylab="Residual", xlab="Predicted
Strength")   >
abline(h=0)
```

1. Normal Probability Plot

```
> qqnorm(newmodel$residuals)
```

2. Histogram hist(newmodel$residuals, main="Histogram",
xlab="Residual")

3. Box Plot

```
> boxplot(newmodel$residuals, horizontal=FALSE, main="Box Plot", xlab="Residual") >
boxplot(newmodel$residuals, horizontal=TRUE, main="Box Plot", xlab="Residual")
```

4. Scatter Plot

```
> plot(order, newmodel$residuals, xlab="Actual Run Order", ylab="Residual", main="Run
Order Plot")
```

Transformation of the Data (response variable) for Normalization

Mass package used for transformation of data when residuals are not normal
Using BOX-COX transformation: > bc = boxcox(newmodel)

```
> title("Box-Cox Transformation")
> lambda = bc$x[which.max(bc$y)]
>  lambda  [1]
0.2626263
```

Calculation of Geometric Mean

Note: After adding the psych package to library which is used for the calculation of Geometric Mean.

$$\frac{(Y_i)^{\lambda} - 1}{\lambda\left[\left(\prod_{i=1}^{n} Y_i\right)^{\frac{1}{n}}\right]^{\lambda - 1}}$$

> newstrength = (strength^lambda - 1)/ (lambda*(geometric.mean(strength)^(lambda-1)))

> newstrength

Adding "newstrength" in the data frame i.e. df

> df = data.frame(df,newstrength)

Run model1 with transformed response variable i.e.
Response(Strength) Y Transformed (newstrength) Y*

> summary(aov(formula = newstrength ~ speed + rate + grit + direction + batch + speed:rate + speed:grit + speed:direction + rate:direction + grit:direction + direction:batch + speed:rate:direction, data = df))

Adding the significant interaction (speed*rate*direction) in the linear model

> srd = s*r*d

> df = data.frame(df,srd)

Running model without interaction direction*batch

> newmodel = lm(formula = newstrength ~ s + r + sr + g + sg + d + ds + rd + srd + gd + b, data=df)

> summary.lm(newmodel) `Call: lm(formula = newstrength ~ s + r + sr + g + sg + d + ds + rd + srd + gd + b, data = df)`

Residual Analysis with transformed strength (response variable)

Plot of Residual VS Predicted

> plot(predict(newmodel),newmodel$residuals,ylab="Residual", xlab="Predicted Transformed Strength") >
abline(h=0)

Histogram

> hist(newmodel$residuals, main="Histogram", xlab="Residual")

Box Plot

```
> boxplot(newmodel$residuals, horizontal=TRUE, main="Box Plot", xlab="Residual")
```

Scatter Plot

```
> plot(order, newmodel$residuals, xlab="Actual Run Order", ylab="Residual", main="Run Order Plot")
```