STAT 425 Final Project

Wenxiao Yang*

*Department of Mathematics, University of Illinois at Urbana-Champaign

2021

1 Introduction

This project aims to identify the optimal operating conditions for Company XX's bubble wrap lines in order to increase production capacity. We investigate the results of a completely randomized design experiment. There are two factors: *line speed* (with three levels, $36, 37, 38 m/mm$), *percent loading of additives* (with three levels, $0, 2, 4\%$). And one response: *production rate* (lbs/hr). The experiment was replicated three times.
Our goal is to find the **optimal combination** of *line speed* and *percent load of additives* that results in the **highest production rate**.

2 Analysis

The goal of our regression analysis is to fit a two-way ANOVA model to determine the effects of various factors with different levels. We started with the full model, which included the effects of various factors as well as the interaction term.

2.1 Full Model: Significance

The full factor effects model is as follows:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}$$

Y_{ijk}: *Production rate* for *Line Speed* i, *Percent Loading of Additives* j, *replication* k.
α_i: effect of *Line Speed* i on *Production rate.*
β_j: effect of *Percent Loading of Additives* j on *Production rate.*
$(\alpha\beta)_{ij}$: interaction term.
The error terms satisfy the usual assumption $\varepsilon_{ijk} \sim \mathcal{N}\left(0, \sigma^2\right)$.
Sum Constraints: $\sum_i \alpha_i = 0, \sum_j \beta_j = 0, \sum_i (\alpha\beta)_{ij} = \sum_j (\alpha\beta)_{ij} = 0$.
$i = 36, 37, 38$.
$j = 0\%, 2\%, 4\%$.
$k = 1, 2, 3$

We first plot the Interaction Plots (Figure 1) and then use the partial F-test of ANOVA table (Figure 2) to test the interaction term's significance.
We find the interaction is presented in the plot. However, the p-value (0.68293) of the F-test is higher than 0.05, we conclude that the interaction terms are not statistically significant. So, we can remove it from the model.

2.2 Additive Model: Significance

Then we fit the regression of the additive model:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \varepsilon_{ijk}$$

Y_{ijk}: *Production rate* for *Line Speed* i, *Percent Loading of Additives* j, *replication* k.
α_i: effect of *Line Speed* i on *Production rate.*
β_j: effect of *Percent Loading of Additives* j on *Production rate.*
The error terms satisfy the usual assumption $\varepsilon_{ijk} \sim \mathcal{N}\left(0, \sigma^2\right)$.
Sum Constraints: $\sum_i \alpha_i = 0, \sum_j \beta_j = 0$.
$i = 36, 37, 38$.
$j = 0\%, 2\%, 4\%$.
$k = 1, 2, 3$

We use the partial F-test of the ANOVA table (Figure 3) to test the significance of both factors. Since the p-values of both factors are lower than 0.05 (*Line Speed*: 0.03777, *Percent Loading of Additives*: 0.02355), we can conclude that both factors are statistically significant.

2.3 Additive Model: Check Assumptions

Then we need to check the assumptions of the additive model.
To validate the assumptions, we perform the studentized Breusch-Pagan test (Figure 4), the Kolmogorov-Smirnov test (Figure 5), and the diagnostic plots (Figure 6) which include the 1. Residuals vs. Fitted plot and the 2. Normal Q-Q plot.
1. **Constancy of Variance (Homoscedasticity)**: By the Residuals vs. Fitted plot and studentized Breusch-Pagan test result ($p - value = 0.117 > 0.05$), we can conclude homoscedasticity assumption holds in the additive model.
2. **Normality**: By the Q-Q plot (obviously not a line) and Kolmogorov-Smirnov test result ($p - value = 1.156e - 07$), we can conclude Normality doesn't hold in the additive model.

2.4 Box-cox Transformation

In the additive model, the normality assumption is violated. Because our response variable only contains positive values, the Box-cox transformation may be used to address assumptions violations. The λ found by Box-cox transformation is 5.151515 (Figure 7). We use $\lambda = 5.5$ to transform the model. Then the new transformed model is as following:

$$\frac{Y_{ijk}^{5.5} - 1}{5.5} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}$$

Y_{ijk}: *Production rate* for *Line Speed* i, *Percent Loading of Additives* j, *replication* k.
α_i: effect of *Line Speed* i on *Production rate.*
β_j: effect of *Percent Loading of Additives* j on *Production rate.*
$(\alpha\beta)_{ij}$: interaction term.
The error terms satisfy the usual assumption $\varepsilon_{ijk} \sim \mathcal{N}\left(0, \sigma^2\right)$.
Sum Constraints: $\sum_i \alpha_i = 0, \sum_j \beta_j = 0, \sum_i (\alpha\beta)_{ij} = \sum_j (\alpha\beta)_{ij} = 0$.
$i = 36, 37, 38$.
$j = 0\%, 2\%, 4\%$.
$k = 1, 2, 3$

2.5 New Transformed Model: Significance

For the new transformed model, we also need to check whether the interaction is significant or not. We plot the Interaction Plots (Figure 8) and then use the partial F-test of ANOVA table (Figure 9) to test the interaction term's significance.
We find the interaction is presented in the plot. However, the p-value (0.9063497) of the F-test is higher than 0.05, we conclude that the interaction terms are not statistically significant. So, we can remove it from the model.

2.6 New Additive Model: Significance

Then we fit the regression of the additive model:

$$\frac{Y_{ijk}^{5.5} - 1}{5.5} = \mu + \alpha_i + \beta_j + \varepsilon_{ijk}$$

Y_{ijk}: *Production rate* for *Line Speed* i, *Percent Loading of Additives* j, *replication* k.
α_i: effect of *Line Speed* i on *Production rate.*
β_j: effect of *Percent Loading of Additives* j on *Production rate.*
The error terms satisfy the usual assumption $\varepsilon_{ijk} \sim \mathcal{N}\left(0, \sigma^2\right)$.
Sum Constraints: $\sum_i \alpha_i = 0, \sum_j \beta_j = 0$.
$i = 36, 37, 38$.
$j = 0\%, 2\%, 4\%$.
$k = 1, 2, 3$

We use the partial F-test of the ANOVA table (Figure 10) to test the significance of both factors. Since the p-values of both factors are lower than 0.05 (*Line Speed*: 0.0067840, *Percent Loading of Additives*: 0.0002654), we can conclude that both factors are statistically significant.

2.7 New Additive Model: Check Assumptions

Then we need to check the assumptions of the new additive model.
To validate the assumptions, we perform the studentized Breusch-Pagan test (Figure 11), the Kolmogorov-Smirnov test (Figure 12), and the diagnostic plots (Figure 13) which include the 1. Residuals vs. Fitted plot and the 2. Normal Q-Q plot.
1. **Constancy of Variance (Homoscedasticity)**: By the Residuals vs. Fitted plot and studentized Breusch-Pagan test result ($p - value = 0.117 > 0.05$), we can conclude homoscedasticity assumption holds in the additive model.
2. **Normality**: By the Q-Q plot (obviously not a line) and Kolmogorov-Smirnov test result ($p - value = 1.156e - 07$), we can conclude Normality doesn't hold in the additive model.
The Box-cox transformation is incapable of resolving the normality violation issue. The new additive model, on the other hand, has a higher Adjusted R-squared (Adjusted $R^2 = 0.5604$, Figure 14) than the previous additive model (Adjusted $R^2 = 0.3258$, Figure 15). As a result, we should draw conclusions using the new additive model.

2.8 Tukeys Paired Comparison

Then we can use Tukeys paired comparison to test the difference between different factor levels.
1. **Line Speed**: According to the Tukeys paired comparison results (Figure16, Figure 17), the 95% family-wise confidence level of
Effect of 37 Line Speed - Effect of 36 Line Speed is $[2.022656e + 12, 2.340860e + 13]$;

Effect of 38 Line Speed - *Effect of 36 Line Speed* is $[-1.144041e+13, 9.945539e+12]$;
Effect of 38 Line Speed - *Effect of 37 Line Speed* is $[-2.415603e+13, -2.770089e+12]$.
These means *Effect of 37 Line Speed* is **statistically larger** than *Effect of 36 Line Speed*; *Effect of 38 Line Speed* and *Effect of 36 Line Speed* are **statistically the same**; *Effect of 38 Line Speed* is **statistically smaller** than *Effect of 37 Line Speed*.
Then we can conclude *Effect of 37 Line Speed*>Effect of 38 Line Speed≈*Effect of 36 Line Speed*, **37 is the best factor level of line speed**.

2. **Percent Loading of Additives**: According to the Tukeys paired comparison results (Figure18, Figure 19), the 95% family-wise confidence level of
Effect of 2% Percent Loading of Additives - *Effect of 0% Percent Loading of Additives* is $[-1.309924e+13, 8.286709e+12]$;
Effect of 4% Percent Loading of Additives - *Effect of 0% Percent Loading of Additives* is $[6.234150e+122.762009e+13]$;
Effect of 4% Percent Loading of Additives - *Effect of 2% Percent Loading of Additives* is $[8.640413e+123.002636e+13]$.
These mean *Effect of 2% Percent Loading of Additives* and *Effect of 0% Percent Loading of Additives* are **statistically the same**; *Effect of 4% Percent Loading of Additives* is **statistically larger** than *Effect of 0% Percent Loading of Additives*; *Effect of 4% Percent Loading of Additives* is **statistically larger** than *Effect of 2% Percent Loading of Additives*.
Then we can conclude *Effect of 4% Percent Loading of Additives*>Effect of 2% Percent Loading of Additives≈*Effect of 0% Percent Loading of Additives*, **4% is the best factor level of percent loading of additives**.

All in all, we conclude that 37 **Line Speed** and 4% **Percent Loading of Additives** is the **optimal combination** we want.

3 Summary

The goal of this project is to find the best combination of two variables. First, we construct the untransformed with interaction and test the significance of the interaction term. Since the interaction test is statistically insignificant in this model, we remove it. Second, we test the significance of the two factors in the remaining additive models, and we find that they are both statistically significant. Third, we examine the additive model's assumptions and discover that the homoscedasticity assumption is correct, but the normality assumption is violated. Fourth, to attempt to correct the normality violation, we apply the Box-cox transformation to the additive model and obtain the new transformed model. Fifth, we examine the interaction term of the newly transformed model. We discover that the interaction term is still statistically insignificant and eliminate it. Sixth, we examine the significance of the two factors in the new additive model. Both of these factors are statistically significant. Seventh, we examine the new additive model's assumptions. The homoscedasticity assumption is still valid, but the normality assumption is still violated, which cannot be fixed by the Box-cox transformation (Since we only have 27 data points, it is not surprised to violate the normality assumption.). Eighth, because the new additive model fit the data better, we use it to find the best factor combination. The best line speed is 37, and the best additive loading percentage is 4%. We conclude the optimal combination is 37 **Line Speed** and 4% **Percent Loading of Additives**.

4 Conclusion (layman)

According to the project's test results, there is no combination effect of line speed and percent loading of additives, so the optimal combination is the optimal line speed with the optimal percent loading of additives. In comparison to 36 and 38, we discover that 37 is the best line speed for increasing production rate the most. And, when compared to 0% and 2%, the optimal percent loading of additives is 4%. Overall, we can conclude that the best combination is 37 Line Speed and 4% Additives Loading.

5 Appendix

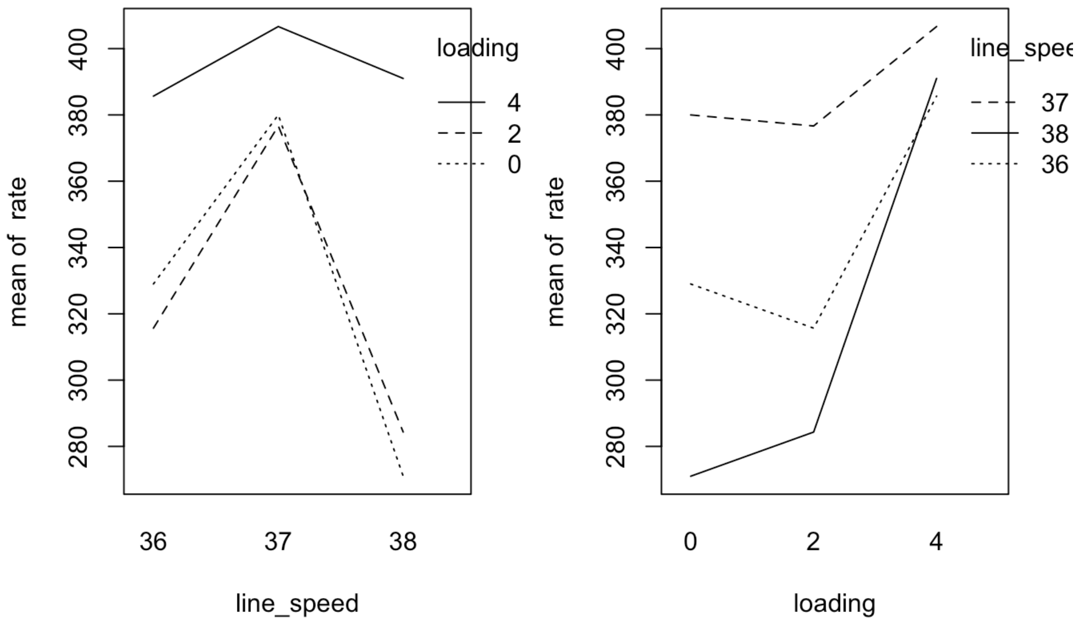

Figure 1: Interaction Plots: Full Model

```
Analysis of Variance Table

Response: rate
                  Df Sum Sq Mean Sq F value  Pr(>F)
line_speed         2  23945 11972.3  3.5230 0.05114 .
loading            2  28022 14011.1  4.1230 0.03357 *
line_speed:loading 4   7844  1961.1  0.5771 0.68293
Residuals         18  61169  3398.3
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 2: ANOVA Table: Full model

```
Analysis of Variance Table

Response: rate
           Df Sum Sq Mean Sq F value  Pr(>F)
line_speed  2  23945   11972  3.8165 0.03777 *
loading     2  28022   14011  4.4664 0.02355 *
Residuals  22  69014    3137
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 3: ANOVA Table: Additive Model

```
	studentized Breusch-Pagan test

data:  model2
BP = 7.3817, df = 4, p-value = 0.117
```

Figure 4: Studentized Breusch-Pagan Test Result: Additive Model

```
	One-sample Kolmogorov-Smirnov test

data:  residuals(model2)
D = 0.55556, p-value = 1.156e-07
alternative hypothesis: two-sided
```

Figure 5: Kolmogorov-Smirnov Test Result: Additive Model

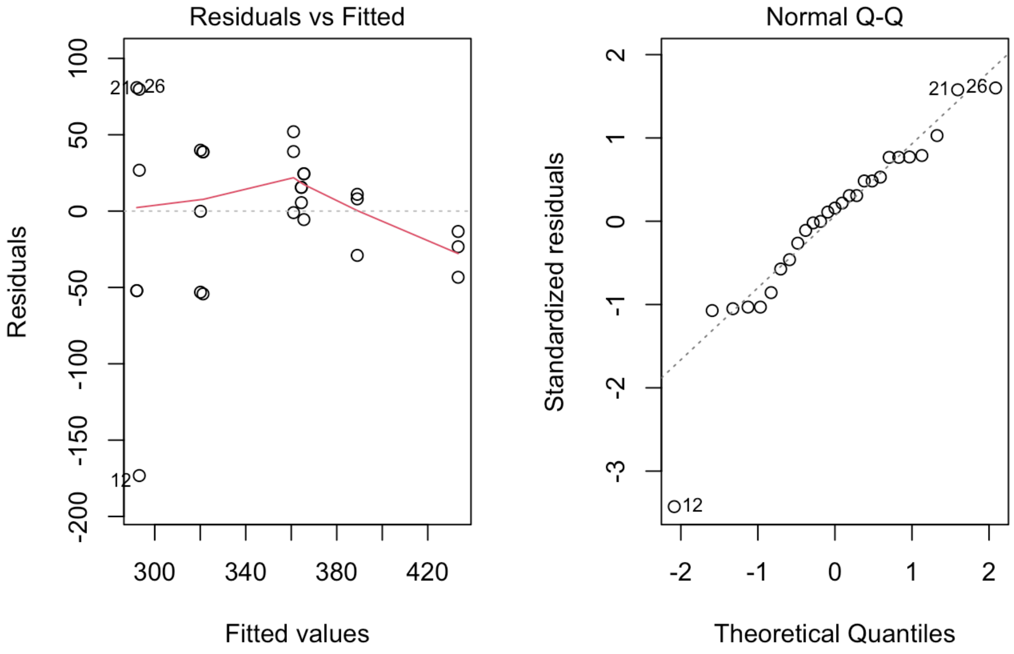

Figure 6: Diagnostic Plots: Additive Model

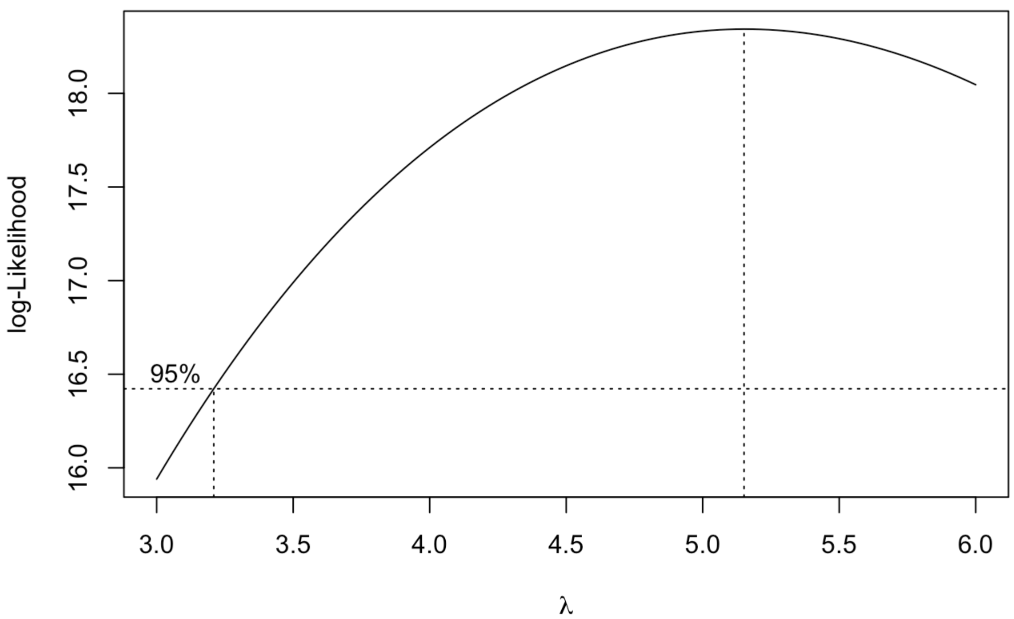

Figure 7: Box-cox Transformation: Find Optimal λ

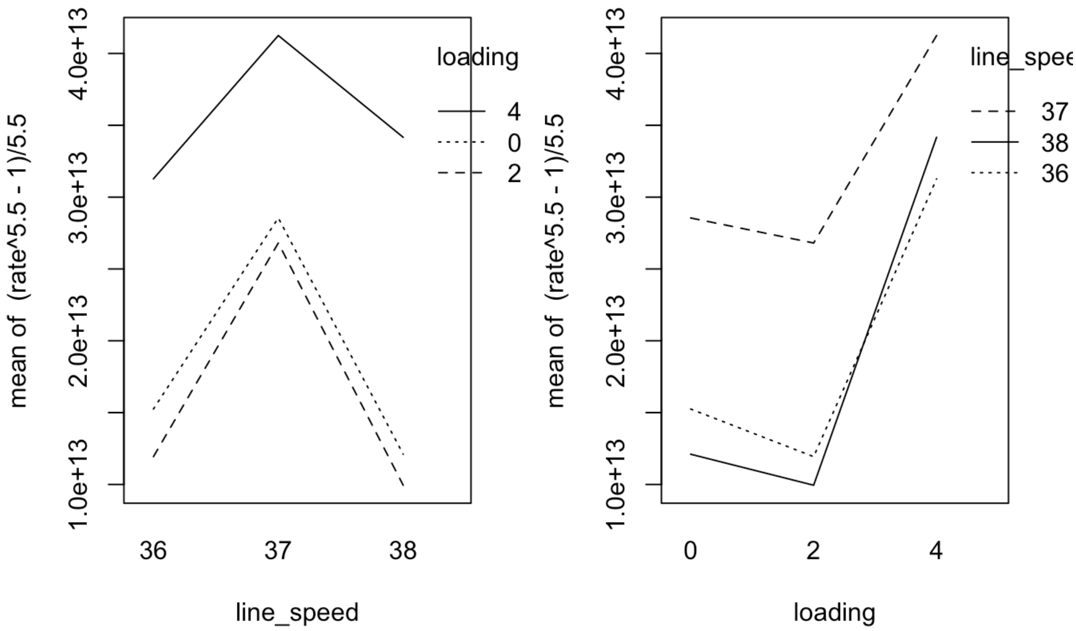

Figure 8: Interaction Plots: New Transformed Model

```
Analysis of Variance Table

Response: (rate^5.5 - 1)/5.5
                   Df     Sum Sq    Mean Sq F value    Pr(>F)
line_speed          2 1.0305e+27 5.1525e+26  5.4567 0.0140461 *
loading             2 1.9983e+27 9.9915e+26 10.5814 0.0009154 ***
line_speed:loading  4 9.4140e+25 2.3535e+25  0.2492 0.9063497
Residuals          18 1.6996e+27 9.4425e+25
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 9: ANOVA Table: New Transformed Model

```
Analysis of Variance Table

Response: (rate^5.5 - 1)/5.5
           Df     Sum Sq    Mean Sq F value    Pr(>F)
line_speed  2 1.0305e+27 5.1525e+26  6.3193 0.0067840 **
loading     2 1.9983e+27 9.9915e+26 12.2541 0.0002654 ***
Residuals  22 1.7938e+27 8.1536e+25
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 10: ANOVA Table: New Additive Model

```
        studentized Breusch-Pagan test

data:  model3
BP = 9.3792, df = 4, p-value = 0.05229
```

Figure 11: Studentized Breusch-Pagan Test Result: New Additive Model

```
        One-sample Kolmogorov-Smirnov test

data:  residuals(model3)
D = 0.51852, p-value = 9.902e-07
alternative hypothesis: two-sided
```

Figure 12: Kolmogorov-Smirnov Test Result: New Additive Model

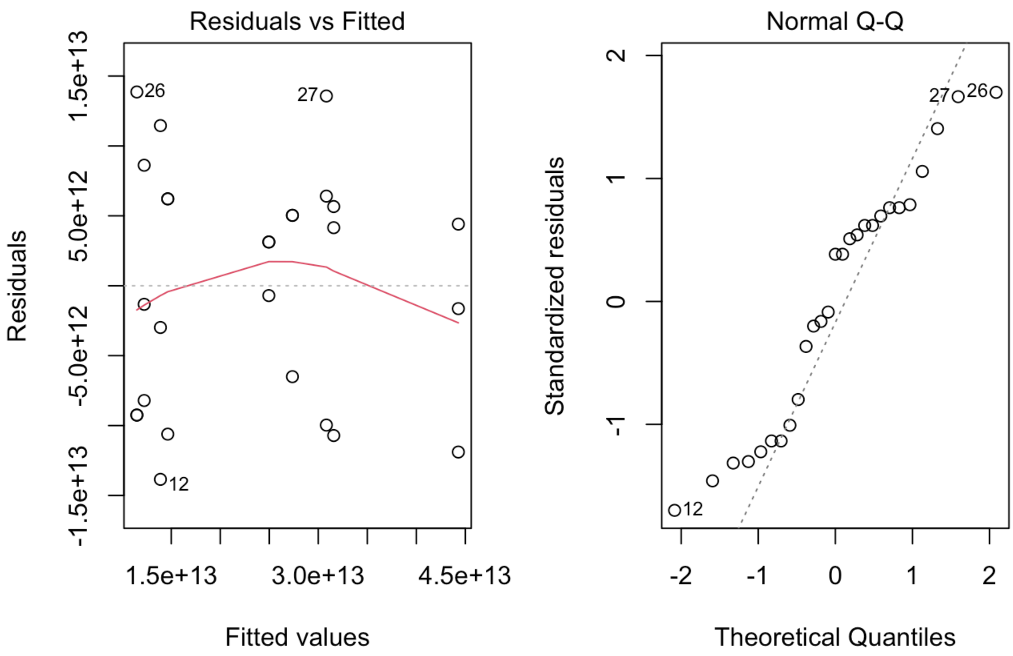

Figure 13: Diagnostic Plots: New Additive Model

```
Call:
lm(formula = rate ~ line_speed + loading, data = new.data)

Residuals:
    Min      1Q  Median      3Q     Max
-173.22  -26.17    8.00   32.78   80.89

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   321.222     24.102  13.327 5.17e-12 ***
line_speed37   44.333     26.403   1.679   0.1073
line_speed38  -28.000     26.403  -1.060   0.3004
loading2       -1.111     26.403  -0.042   0.9668
loading4       67.778     26.403   2.567   0.0176 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 56.01 on 22 degrees of freedom
Multiple R-squared:  0.4295,    Adjusted R-squared:  0.3258
F-statistic: 4.141 on 4 and 22 DF,  p-value: 0.01192
```

Figure 14: Regression Summary: Additive Model

```
Call:
lm(formula = (rate^5.5 - 1)/5.5 ~ line_speed + loading, data = new.data)

Residuals:
       Min         1Q     Median         3Q        Max
-1.385e+13 -8.729e+12  3.128e+12  5.938e+12  1.386e+13

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   1.465e+13  3.886e+12   3.769 0.001057 **
line_speed37  1.272e+13  4.257e+12   2.987 0.006792 **
line_speed38 -7.474e+11  4.257e+12  -0.176 0.862221
loading2     -2.406e+12  4.257e+12  -0.565 0.577593
loading4      1.693e+13  4.257e+12   3.977 0.000638 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 9.03e+12 on 22 degrees of freedom
Multiple R-squared:  0.628,     Adjusted R-squared:  0.5604
F-statistic: 9.287 on 4 and 22 DF,  p-value: 0.0001491
```

Figure 15: Regression Summary: New Additive Model

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

Fit: aov(formula = (rate^5.5 - 1)/5.5 ~ line_speed + loading, data = new.data)

$line_speed
               diff           lwr           upr     p adj
37-36  1.271563e+13  2.022656e+12  2.340860e+13 0.0179125
38-36 -7.474332e+11 -1.144041e+13  9.945539e+12 0.9831589
38-37 -1.346306e+13 -2.415603e+13 -2.770089e+12 0.0120449
```

Figure 16: Confidence Interval: Line Speed

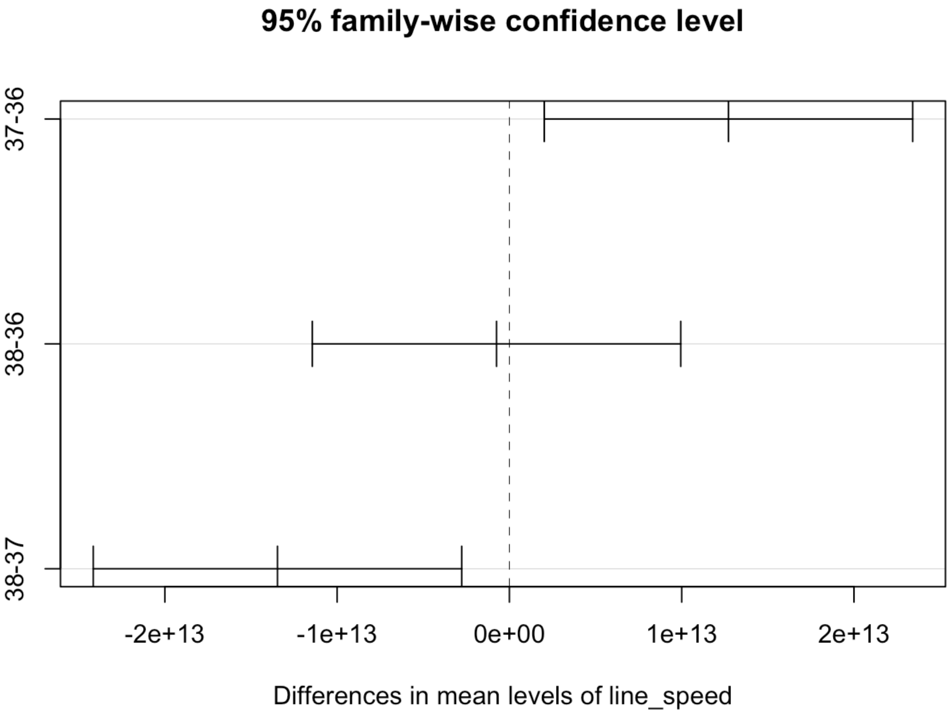

Figure 17: Confidence Interval Plot: Line Speed

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

Fit: aov(formula = (rate^5.5 - 1)/5.5 ~ line_speed + loading, data = new.data)

$loading
             diff           lwr          upr     p adj
2-0 -2.406263e+12 -1.309924e+13 8.286709e+12 0.8397419
4-0  1.692712e+13  6.234150e+12 2.762009e+13 0.0017759
4-2  1.933339e+13  8.640413e+12 3.002636e+13 0.0004541
```

Figure 18: Confidence Interval: Percent Loading of Additives

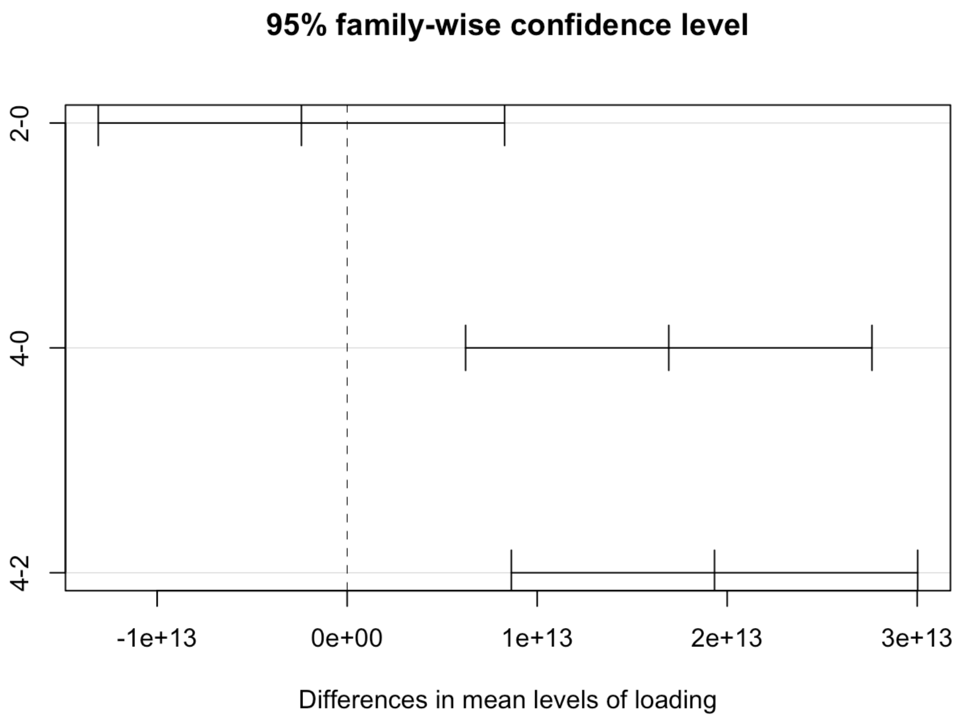

Figure 19: Confidence Interval Plot: Percent Loading of Additives