STAT425_FinalProject_Summary

Giang Le

12/13/2021

Introduction

In this project, I analyze the data from a balanced two-factor factorial design. The two categorical factors are **line speed** and **loading of additives**. Both of them have three levels each. Line speed's levels are 36, 37, and 38. The loading of additives' levels are 0, 2, and 4. The response variable is the production rate, measured in lbs/hour.

The goal of our experiment and statistical analysis is to figure out the optimal combination of line speed and percent load of additives that results in the **highest production rate**. In this report, I present results of the data analysis, diagnostics, model fitting based on the collected data and draw final conclusion to recommend the best combination of line speed and percent load of additives.

Data Analysis and Model

The data and project description suggest that we have a factorial structure/design because the 3 levels of line_speed appear with the 3 levels of loading of additives. Because the factors are fully crossed, we have 3 x 3 treatments. The experiment is replicated three times for each treatment. We can decompose the model into $\mu_{ij} = \mu + \alpha_i + \beta_j + (\alpha\beta_{ij})$ where μ is the overall rate mean, α_i is the factor A, line_speed, fixed effect, β_j is the factor B, loading, fixed effect. and $(\alpha\beta_{ij})$ are the fixed interaction effects. The model also includes noise ϵ_{ijk} which is normal independently distributed with mean 0 and constant variance.

Model Fitting and Diagnostics

I first fit a general model with an interaction term. The interaction term's p-value is 0.68293 > 0.05 so the interaction effect is not statistically significant at 0.05 significance level. See Initial Model for the first fitted model with an interaction term. Next I plot the residuals vs. fitted, normal QQ plots to check for constant variance and normality assumption. The residuals vs. fitted plot shows larger variance values at smaller fitted values, so I suspected that the constant variance assumption has been violated. The normal QQ plot looks fine. See Diagnostics Plots 1 for all plots.

I conducted the studentized Breusch-Pagan test to test for constant variance of the residuals.

The p-value of this test is less than 0.05 so I conclude that we reject the null hypothesis of constant variance and that the constant variance assumption has been violated.

I considered a Box-Cox transformation in the next step.

Model Transformation and Rechecking the Assumptions

The Box-Cox plot suggests a lambda value between 1 and 2. I replotted the Box-Cox diagram to zoom into lambda values. See Box Cox Plot for the lambda values. The closest lambda value recommended to me is 1.6, therefore I decided to transform the response variable to (rate^1.6 - 1)/1.6. I rechecked the previous assumption and conducted the studentized Breusch-Pagan test to test for constant variance of the residual again. The p-value this time is 0.05529 so I concluded that the variance distribution is satisfactory after transformation and the variance is likely to be constant according to the null hypothesis.

Interaction Plot and Examining the Main Effects

The p-value of the interaction term of the transformed model is $0.70369 > 0.05$, so I reject the null hypothesis and conclude that the interaction effect of the transformed model is not statistically significant. I plotted the interaction plot to confirm that the interaction effect was not significant. The lines are roughly parallel (if noise is taken into account) so the previous conclusion is confirmed. See Interaction Plot 1 and Interaction Plot 2.

Fitting Main Effect Model

The main effect model includes only the loading and line_speed factors and both factors are statistically significant(See Main Effects Model). I also conducted a partial F-test to compare model 1 which is an additive model and model 2 which is an interaction model. The p-value is greater than 0.05, so I fail to reject model 1. I conclude that the additive model is satisfactory.

Recommend Optimal Combination of Levels

In order to recommend the best combination of levels of the loading and line_speed, I examined the summary of the main effects model (See Levels). Loading 4 produces an additive effect of 2117.4 lbs/hour and its p-value is 0.0107, so loading 4 is statistically significant in producing a positive additive effect on the production mean. The intercept is statistically significant and it represents the estimated combined effect of loading 0 and line speed 36. I noted that this is a positive effect on the production mean.

I used the TukeyHSD multiple comparisons of means with 95% family-wise confidence level to compare pairwise means more closely (See Tukey). In terms of the loading levels, loading 4 is statistically significant and the difference in its effect on the production mean is positive when compared with both loading 0 and loading 2. Therefore loading 4 is the best value among the three levels. In terms of the line speed, it can be seen that level 37 is significantly better than level 38, (p-value = 0.0230) however the difference between level 37 and 36 is not statistically significant. Combined with the intercept term (6539.4) in the previous summary, line speed 36 produces a statistically significant positive effect on the production mean.

In conclusion, I recommend loading 4 and line speed 36 to produce the most positive effect on the production rate (the highest production rate).

Appendix

Initial Model

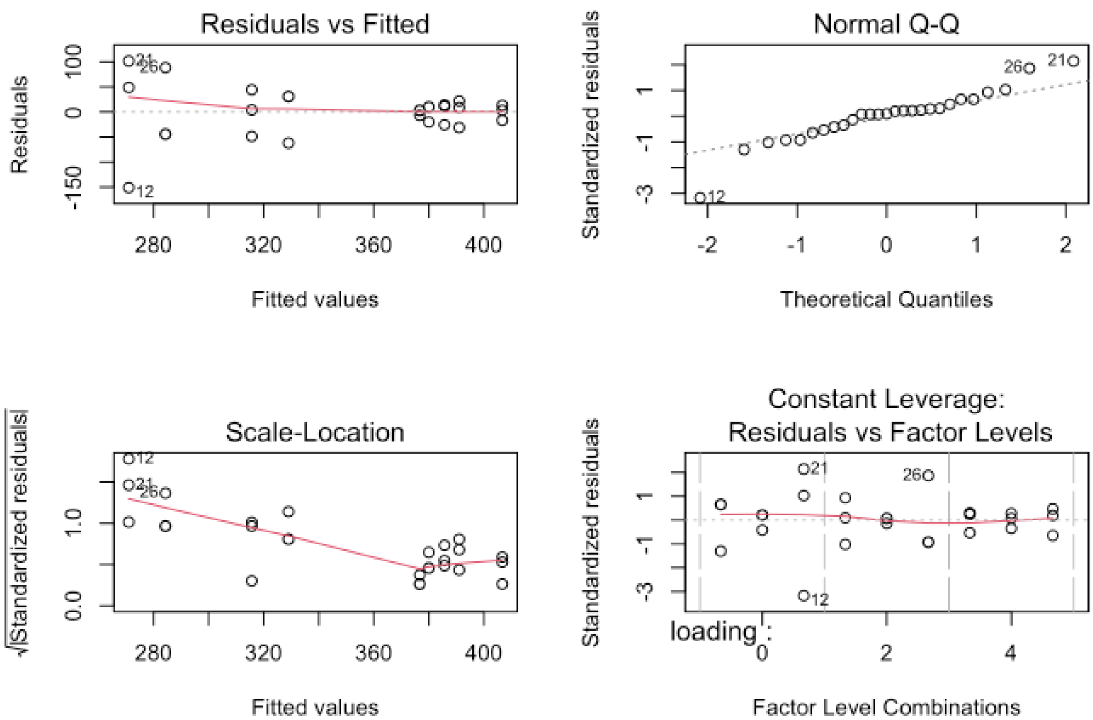
Fitting ANOVA model

```
bubblewrap <- read.csv("bubblewrap.csv")
bubblewrap$line_speed <- as.factor(bubblewrap$line_speed)
bubblewrap$loading <- as.factor(bubblewrap$loading)
lmod <- lm(rate ~ loading * line_speed, bubblewrap)
anova(lmod)
```

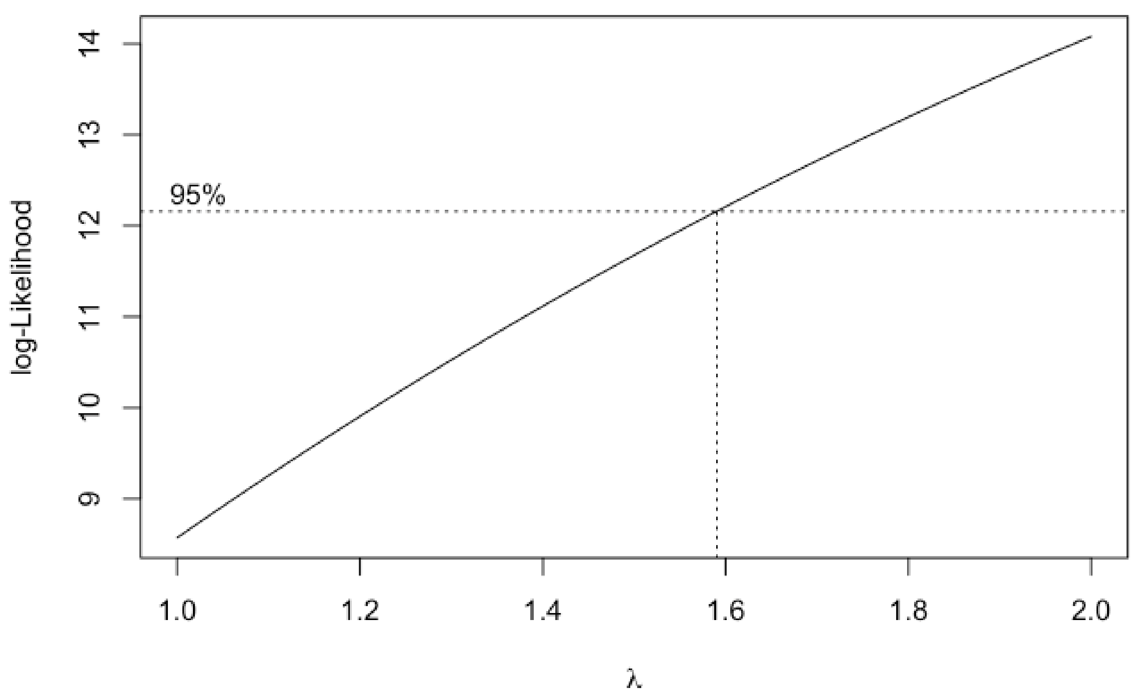
```
## Analysis of Variance Table
##
## Response: rate
##                    Df Sum Sq Mean Sq F value  Pr(>F)
## loading             2  28022 14011.1  4.1230 0.03357 *
## line_speed          2  23945 11972.3  3.5230 0.05114 .
## loading:line_speed  4   7844  1961.1  0.5771 0.68293
## Residuals     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

```
# Interaction term is not significant.
```

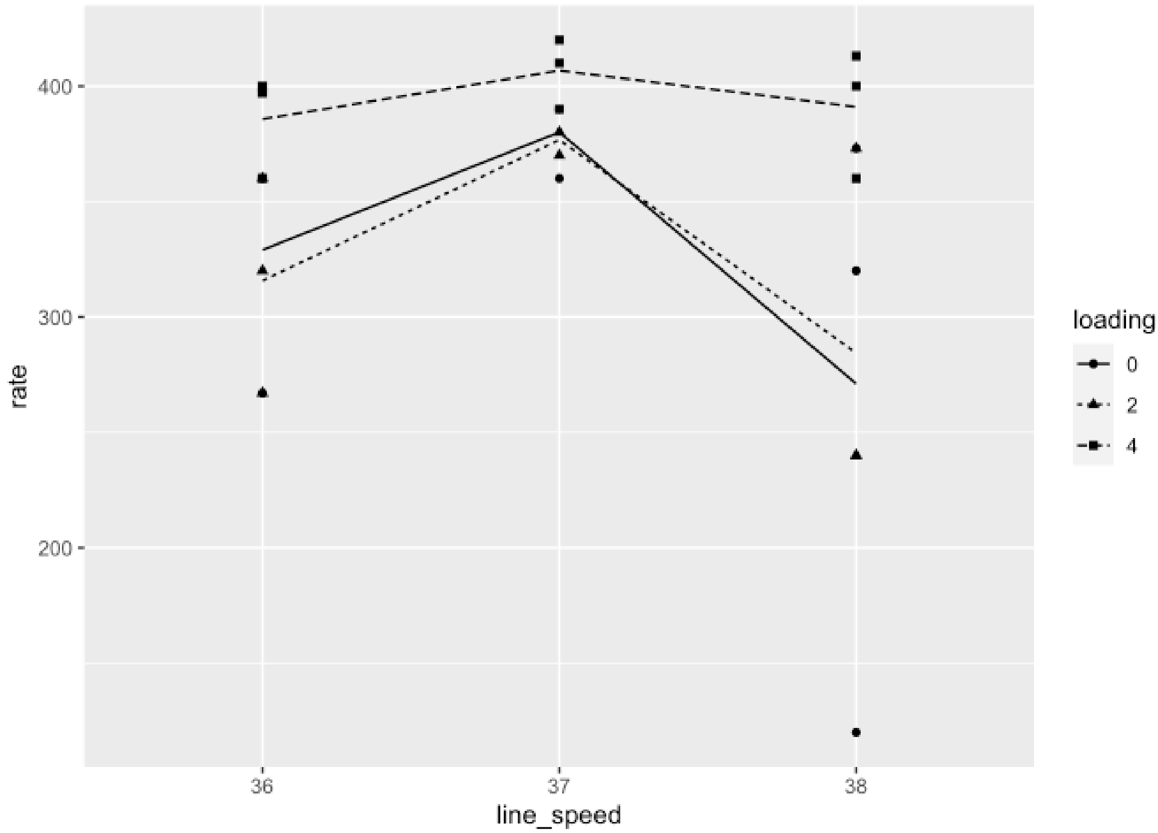
Diagnostics Plots 1



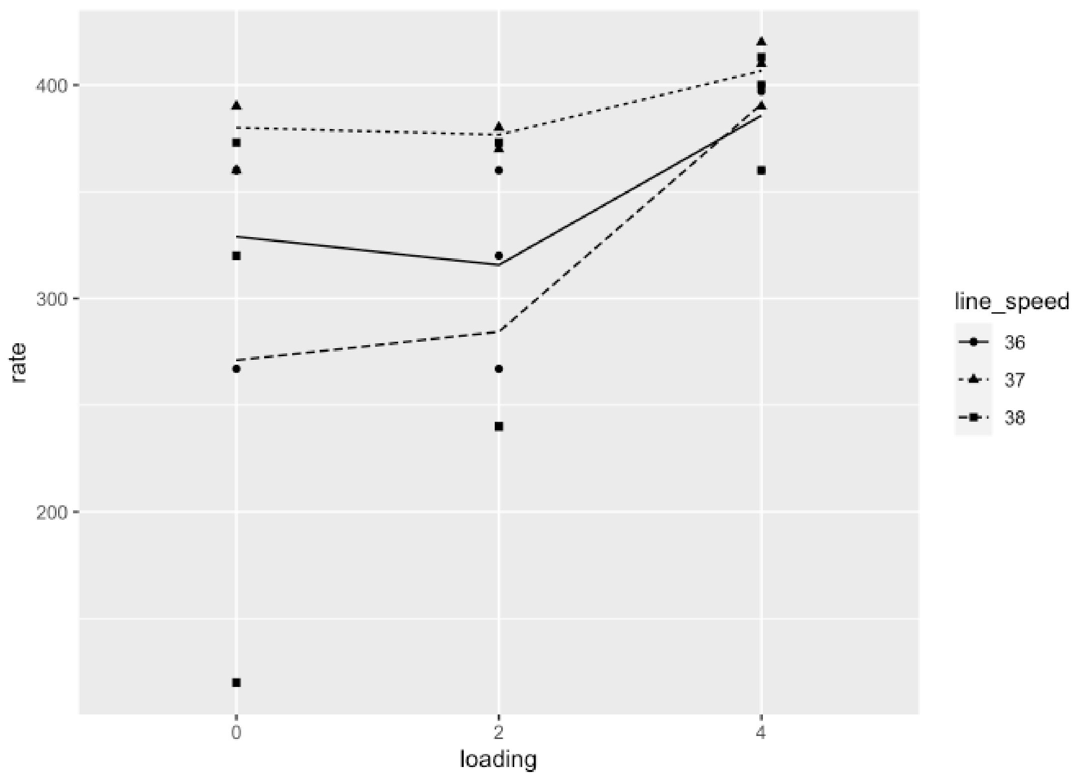
Box Cox Plot



Interaction Plot 1



Interaction Plot 2



Main Effects

Removing the interaction and test the main effects.

```
maineffects <- lm((rate^1.6 - 1)/1.6 ~ loading+line_speed, bubblewrap)
anova(maineffects)
```

```
## Analysis of Variance Table
##
## Response: (rate^1.6 - 1)/1.6
##             Df   Sum Sq  Mean Sq F value  Pr(>F)
## loading      2 28930917 14465458  5.5741 0.01101 *
## line_speed   2 22317352 11158676  4.2998 0.02653 *
## Residuals   22 57093032  2595138
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anova(maineffects, transformed_mod)
```

```
## Analysis of Variance Table
##
## Model 1: (rate^1.6 - 1)/1.6 ~ loading + line_speed
## Model 2: (rate^1.6 - 1)/1.6 ~ loading * line_speed
##   Res.Df      RSS Df Sum of Sq      F Pr(>F)
## 1     22 57093032
## 2     18 50907715  4   6185317 0.5468 0.7037
```

Levels

```
summary(maineffects)
```

```
##
## Call:
## lm(formula = (rate^1.6 - 1)/1.6 ~ loading + line_speed, data = bubblewrap)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4505.6  -814.4   336.4   938.9  2457.4
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    6539.4      693.2   9.433 3.45e-09 ***
## loading2       -149.3      759.4  -0.197   0.8459
## loading4       2117.4      759.4   2.788   0.0107 *
## line_speed37   1474.3      759.4   1.941   0.0651 .
## line_speed38   -708.3      759.4  -0.933   0.3611
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1611 on 22 degrees of freedom
## Multiple R-squared:  0.473,  Adjusted R-squared:  0.3772
## F-statistic: 4.937 on 4 and 22 DF,  p-value: 0.005399
```

Tukey

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = (rate^1.6 - 1)/1.6 ~ loading + line_speed, data = bubblewrap)
##
## $loading
##          diff        lwr      upr     p adj
## 2-0 -149.3292 -2057.0047 1758.346 0.9789294
## 4-0 2117.3881   209.7126 4025.064 0.0277944
## 4-2 2266.7173   359.0417 4174.393 0.0180082
##
## $line_speed
##             diff        lwr       upr     p adj
## 37-36  1474.3184  -433.3572 3381.9939 0.1508898
## 38-36  -708.2984 -2615.9739 1199.3772 0.6258881
## 38-37 -2182.6168 -4090.2923 -274.9412 0.0230281
```