Finding Optimal Popcorn-Popping Conditions

Isabel Arvelo, Molly Hellman, Kevin Molina, Shea van den Broek

2022-11-26

I. Introduction

What does the perfect bowl of popcorn look like? Is it the oil, salt, a combination of both? Since we are not professional tasters, we opted to focus on how efficient or productive batches of kernels are under different conditions by looking at the number of kernels popped each round—a result that is easy to measure quantitatively. Despite how simple it seems, there are many factors that go into the making of popcorn on a stove—the temperature of the stove, the size of a pot, the type of oil—among other variables. Different recipes for making popcorn suggest different combinations of these variables, so we were looking to isolate which produce the most popcorn and feed the maximum number of people with only one pot of kernels. Whether it's for a movie, perhaps a watch party with friends and family, knowing how to cook up a big bowl of popcorn can go a long way. We ultimately ran a randomized unreplicated 2^{7-4} experiment, which was $\frac{1}{16}$ of a full 2^7 factorial design. We included the following seven binary predictor variables representing the popcorn conditions: stove temperature (using a scale from 1 to 10), brand of kernel (Orville or American), the type of oil (either Canola or Olive oil), the amount of oil (either low or high), the number of kernels (either a low or high number), whether salt was added or not, the volume of the pot (either a small or large pot). We measured the number of kernels popped as our response variable. Our goal was to find the optimal combination of variables involved in the making of popcorn that would produce the most popped kernels.

Factors	Low Level (Coded -1)	High Level (Coded +1)
A = Stove Temperature	Low: 4	High: 8
B = Type of Kernel	Orville Redenbacher's	American's Best
C = Type of Oil	Canola Oil	Olive Oil
D = AB = Amount of Oil	Low: 2 TBSP	High: 4 TBSP
E = AC = Number of Kernels	Low: $\frac{1}{4}$ Cup	High: $\frac{3}{4}$ Cup
F = BC = Seasoning Salt	Salt	None
G = ABC = Volume of Pot	Small Pot	Large Pot

II. Design

We chose to implement a randomized, unreplicated 2^{7-4} design with 16 total runs. We found this design to be the best option due to the large number of variables we were examining, and the fact that the design allows us to work toward conclusions as to find which effects of all seven variables were significant and which contributed to higher numbers of popped kernels with the lowest number of total runs required. This made it the most efficient design in time and resources. We began with eight preliminary runs. We used the following design matrix using Yates's standard order, shown below:

A	B	C	AB	AC	BC	D
-1	-1	-1	1	1	1	-1
1	-1	-1	-1	1	1	1
-1	1	-1	1	1	-1	1
1	1	-1	-1	1	-1	-1
-1	-1	1	-1	1	-1	1
1	-1	1	1	1	-1	-1
-1	1	1	-1	1	1	-1
1	1	1	1	1	1	1

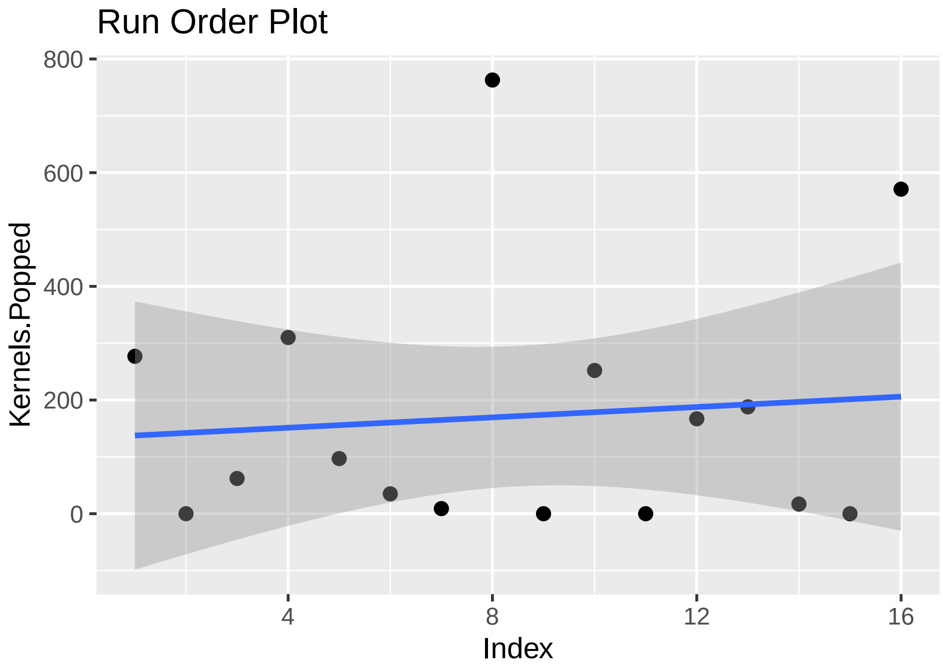
We utilized the generating equation ABC = -D, which also led to the following other interactions between effects (See Appendix 1).

After this first set of runs, we then conducted a fold-over design, switching the sign of our D effect, the amount of oil, in order to decipher further between confounding effects we found in the initial runs. This gave us the generating equation D = -AB for this next set of runs. Lastly, after analyzing the data from the two sets of runs combined, we found an optimal combination of the seven factors. We finished with a confirmatory run with these optimized conditions, to check against our first run with the same combination, and four runs with center point values, to be able to estimate the pure error in our model and to allow for a linearity check in our analysis.

III. Implementation

Regarding our implementation, there were a few possible issues in our experiment that could have compromised the independence between trials. For example, as we were completing trials with the same pots and different types of oil, there was a chance for residual oil to be in the pot before the next trial. In order to try to combat this, we cleaned each pot thoroughly between trials. Another possible problem could have been inconsistent heating from the stove, as we were changing the heat setting of the stovetop from our low level of heat (4 out of 10) to our high level of heat (8 out of 10). Finally, the differing methods of counting kernels also presents an issue. Namely, it is likely that different people have differing definitions of what is considered a popped kernel, especially when the state of the kernel is not easily distinguishable. In addition, because there are many kernels used in the experiment, and since we had to manually count each popped kernel ourselves, the possibility of human error in the calculation is present. In addition to the potentiality of non-independence, we also had to take into account the possibility of a nuisance variable present. Specifically, as we performed our experiment over the span of three days, we were left with three different batches of popcorn that form related populations, thus we needed to verify whether or not the blocking variable, in this case the batch corresponding to each day, was statistically significant. To do this, we conducted a one-way Analysis of Variance to compare the mean number of kernels popped across the three different batches (See Appendix 1). Ultimately, we failed to reject the null hypothesis and did not find evidence that the mean number of kernels popped varied between the three blocks as the p-value was far greater than our aα level of 0.05, therefore we opted to exclude the blocking variable from future analyses. Furthermore, we were also concerned about the run order of the experiment. As we conducted our experiment in the order determined by Yates' standard order, it is possible that this affected the results of our experiment by introducing statistical noise into our model as well as lurking variables, which would not have occurred had we randomized the sequence of the runs.

```
## `geom_smooth()` using formula 'y ~ x'
```



The run order plot does not indicate any strong temporal trends so it is fair to assume independence.

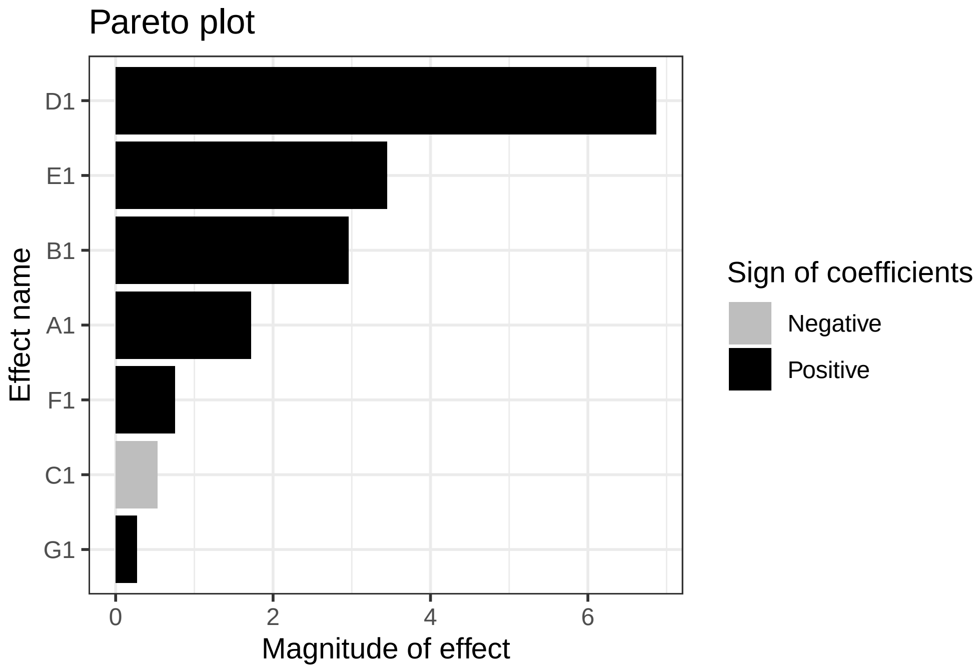
IV. Methodology

To evaluate the optimization of popcorn popping, we set seven different factors that were suspected to play significant roles in the process.Two different brands of kernels were used, Orvilles and America's Best, to distinguish whether the kernel brand was robust to the other six factors. A $\frac{3}{4}$ measuring cup and $\frac{1}{4}$ cup of kernels were the levels set for density, and temperature was differentiated by level four and eight on the stovetop's settings. For best results, popcorn should be popped in some kind of oil, so we tested both canola and olive oil at two tablespoons and four tablespoons. Finally, we tested the significance of relative pot size and presence of salt. Our first eight runs followed a 2^{7-4} factorial experimental design. The oil sat in the pot on the stove until it began to simmer, at which point the popcorn was added, and the pot closed with a lid. We allot each batch two minutes on the heat, then remove it to let it cool before manually counting how many had popped. Following the first eight runs we performed four midpoint runs and a replication of the highest scoring run from the first set to confirm it was the optimal point. The latter followed the same procedure and factor settings as our original eighth run. For the midpoints we set the qualitative factors and then took the middle value of each quantitative factor. We chose to use Orville kernels cooked in a large pot with no salt and canola oil, and set temperature to level 6 on the stove, using $\frac{1}{2}$ cup scoop of kernels, and 3 tablespoons of oil. These center point runs will serve as linearity tests in the below analysis.

V. Analysis

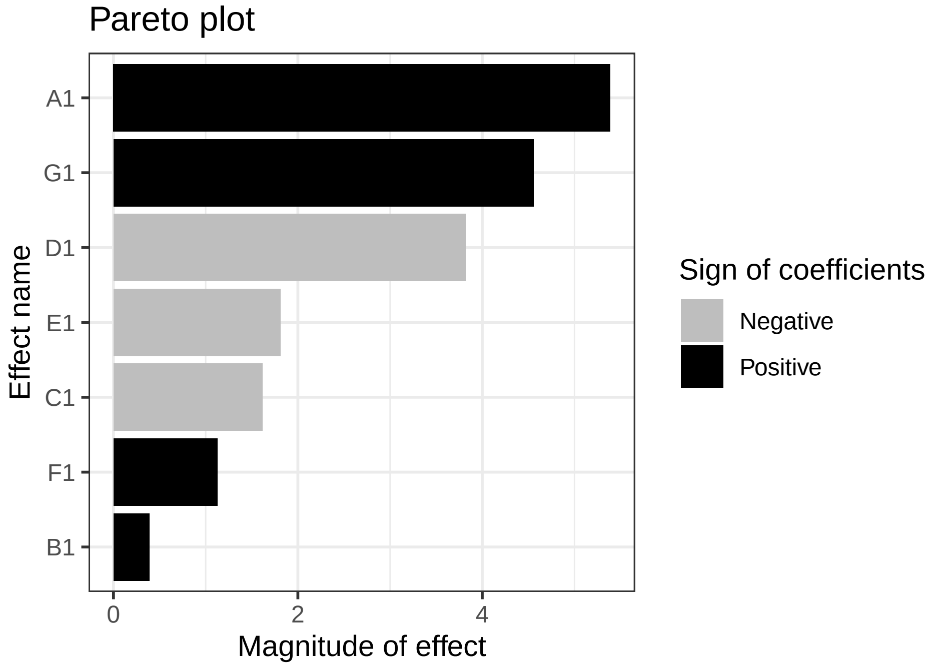
After visualizing the data, we immediately found that the box plots displayed inconsistent variance, making the model unfit, however a box-cox analysis suggested taking the square root of the output measure could fix this. This is a common transformation to apply to highly skewed count data. (See Appendix 1)

In doing so, variance across variables became roughly equal allowing us to move forwards with models that compare within group variance to between group variance. The boxplots of Stove Temperature (A), Amount of Oil (D), and Volume of Pot (G) suggested differences between the number of kernels popped for the two levels, but the other boxplots did not have a clear difference.



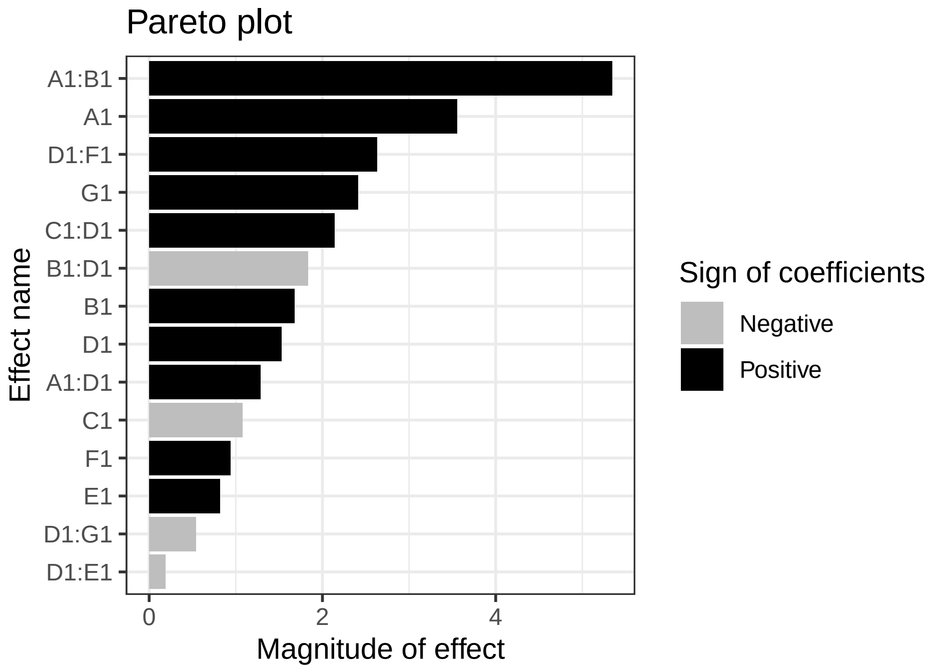
The first eight runs produced both a full normal and bayes plot (See Appendix 1) that showed none of the factors were active, however, this is likely indicative that multiple factors are significant, which confuses the plots. Due to this suspicion, we continued on with further testing, looking at a Lenth plot and Pareto plot, both of which suggested that factor D has an active effect. However, given our many aliases, we didn't know whether this was the main effect of D (Amount of Oil), or one of it's two way aliases: the interaction between A and B (Stove Temperature and Brand of Kernel), the interaction between C and G (Type of Oil and Size of the Pot), or the interaction between E and F (Number of Kernels and Salt). Using our domain knowledge, we ruled out the interaction with E and F, between seasoning of salt and number of kernels, but needed to perform more runs to decipher further between confounded effects.

After the first 8 runs, we wanted to "dislodge" the confounded effects from the first experiment. In order to do so, we flipped the sign of factor D and ran 8 more runs with the remaining factors in the same configuration. Since this was a 2^{7-4} design, we had 16 ways to choose generators but because factors 1, 4, and 2 were involved we decided to run the next 8 runs with $D = -AB$



We completed a second set of eight runs in order to dislodge the confounded effects, so we flipped the sign of factor D. We used the generating equation D = -AB. To analyze which effects were active, we again looked at a full normal plot and Bayes plot, but these were not informative due to what we believed to be the same issue as earlier (See Appendix 1). A Pareto plot and Lenth plot suggested that factor A, Stove Temperature, and factor G, Volume of Pot, were active factors. Looking at only this set of runs, we run into a similar confounding problem. We don't know whether the main effect of G (Volume of Pot) is significant or if its actually its two way alias B:D(Type of Kernel and Amount of Oil Interaction) or C:E (Type of Oil and Number of Kernels Interaction). The same goes for factor A which can be the main effect of stove temperature or the two way interactions B:D (type of kernel and amount of oil) or C:E(type of oil and number of kernels).

Next, we merged the data frames to try and estimate more of the two way interactions (See alias structure in Appendix 1).



From the first 16 runs, it appears that the two way interaction between A:B has the largest effect, followed by the main effect of A (See Appendix 1). Depending on where we draw the threshold of significance, the two way interaction between D:F would be the next largest and potentially significant effect. By folding the experiment, we were able to distinguish between the main effect of D from its alias A:B that appears to be what is actually driving the significant effect we observed in the first 8 runs. We feel fairly confident ruling out interactions that include factor F, seasoning of salt. That leaves A:B still confounded with two way interactions between C:G (Type of Oil and Volume of Pot Interaction) and A still confounded with the two way interactions between C:E (Type of Oil and Number of Kernels).

Without more runs, we cannot know for sure which of these is truly important, but it seems that the main effect of A: Stove Temperature is definitely involved and we believe that the two way interaction between Stove Temperature and Kernel Brand is more likely than Type of Oil and Volume of Pot interaction. So using the results from our analysis and our own intuition, we decided that A and B seem to be the most important factors and built a reduced model with just Stove Temperature and Type of Kernel.

```
##             Df Sum Sq Mean Sq F value  Pr(>F)   
## A            1  202.2   202.2   5.236 0.04107 * 
## B            1   44.9    44.9   1.163 0.30214   
## A:B          1  456.5   456.5  11.818 0.00491 **
## Residuals   12  463.5    38.6                   
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

A reduced model including factors A, B and their interaction term helped to diffuse this discrepancy. In this model, both factor A and the interaction term gave significant p-values, 0.04107 and 0.00491 respectively. This gives us confidence in concluding that Stove Temperature and its interaction with Kernel Type are the active factors in this experiment, and any future work done should hone in on these effects. The root mean square error estimated from the insignificant factors we excluded was was 6.21, resulting in a coefficient standard error of 1.55.

Replication is the only way to obtain a pure error estimate. Therefore, we ran four center points with arbitrary, but consistent levels for the categorical variables and the cener value between the two levels of the quantitative variables. We can find the variances of these replicated runs to estimate the standard error.

Ideally, we expcted this to be roughly close to the standard error from reduced model, but we find that this is not the case. We ran four center points and we can use the variances of these fours runs to estimate the pure error which gives us an estimated RMSE of 1.04. Dividing this by the number of observations used to estimate the coefficients from the first 16 runs, we find a coefficient standard error of 0.261 which gave large t statistic values for every coefficient, indicating that all the effects are significant.

We explored this further with our linearity check (See Appendix 2), where we found two issues: the plots all showed a clear curve rather than a linear-looking relationship, and the spread for the middle points was significantly smaller than the spread of the points on both edges. We therefore attributed the extremely large t-statistics for every coefficient using the MSE from these center points to this smaller variance that does not accurately capture the true variation in the response.

Using the prediction profiler in JMP, we determined that the optimal combination of factors was the following: high stove temperature, American brand of kernels, Olive Oil, a high amount of oil, high number of kernels, no salt, and a large pot (See Appendix 3). This matches the profile of the run with the highest number of kernels popped that we ran in the first 16 runs. Our next step was to conduct a confirmatory run to see how well our model could actually predict at the optimum factor combination. We predicted 702 popped kernels for this point, and our confirmatory run resulted in 830 kernels popped. Using the standard error estimate from the center points, we calculated the difference between the observed and predicted divided by the standard error to be 8.854, which would mean that we did not predict the optimal combination well because our predicted value was significantly different from the observed value. However, when using the standard error estimate from our reduced model, we calculated that the observed value was only 1.488 standard errors away from the predicted value. Given the issues we found earlier with the pure standard error estimate from the center points being too small, we deduced that the reduced model's standard error estimate was better and therefore feel fairly confident that we are able to accurately predict the number of kernels popped at the optimal level.

VI. Conclusion

Through the twenty runs performed thus far, we have identified Stove Temperature, Brand of Kernel, and their interaction term as active effects in the number of kernels popped per run. So far, it does not appear that type of oil, seasoning, or number of kernels are important. In the next iteration of this experiment, flipping the sign of factor A, Stove Temperature, would serve to dislodge any confounding interaction aliases that may underlie the aforementioned results. With the results we have now, we are still not confident defining any one factor combination as the true optimal point. While the Yates standard order design allowed us to efficiently test every combination of factors, its limitation lies in the lack of data points produced. However, continuous replication will negate this as we continue to experiment. If we were to run the experiment again, we would reduce the number of kernels at each level to make the counting easier. This would likely reduce the variance of the response and allow us to complete more runs more quickly. One possible avenue for further research, as briefly mentioned in the introduction, is to actually look into taste as a response variable. Contextually, one can estimate that the taste of the popcorn may matter far more than number of kernels popped, of course, this would require official tasters, or at least consistent tasters, but if we were able to establish a system this could be another exploration to go along with the experiment we just completed. The question would then be: do the optimal conditions match when the response is taste versus when the response was the number of popped kernels, or are there clear differences?

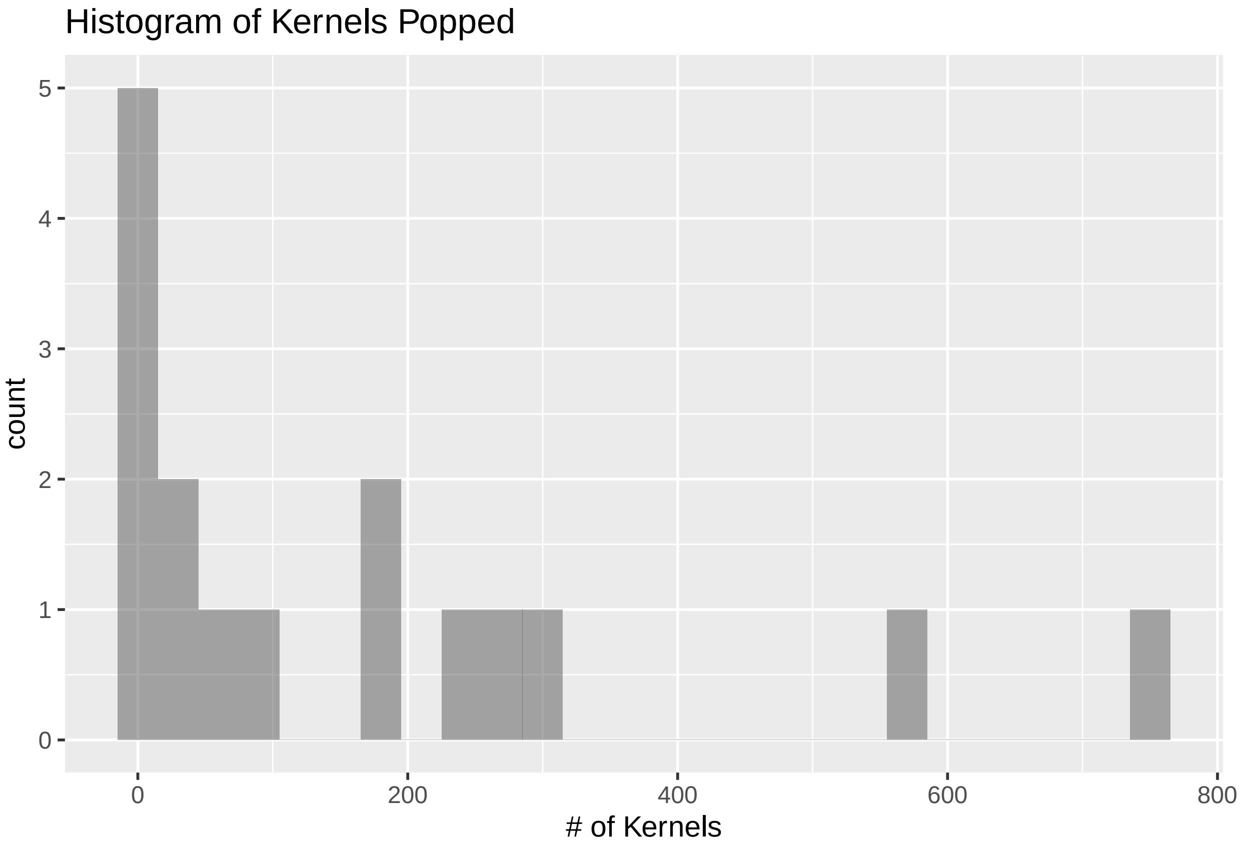
Appendix 1

```
#aliases
aliases(lm(y~.^2,data=df3))
```

```
##
##  A = B:D = C:E = F:G
##  B = A:D = C:F = E:G
##  C = A:E = B:F = D:G
##  D = A:B = C:G = E:F
##  E = A:C = B:G = D:F
##  F = A:G = B:C = D:E
##  G = A:F = B:E = C:D
```

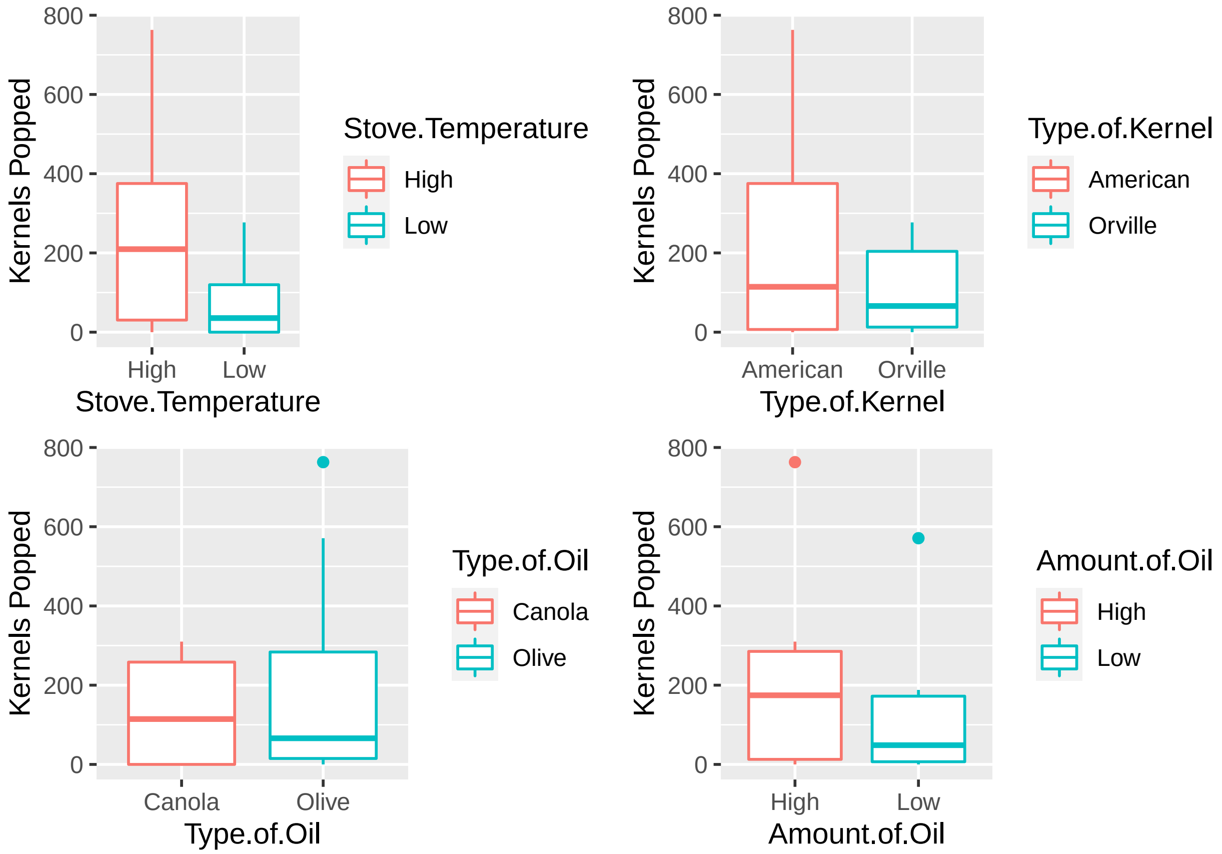
```
#Checking Block
summary(aov(KP.TF ~ Block, data = Kernels[c(1:16, 23), ]))
```

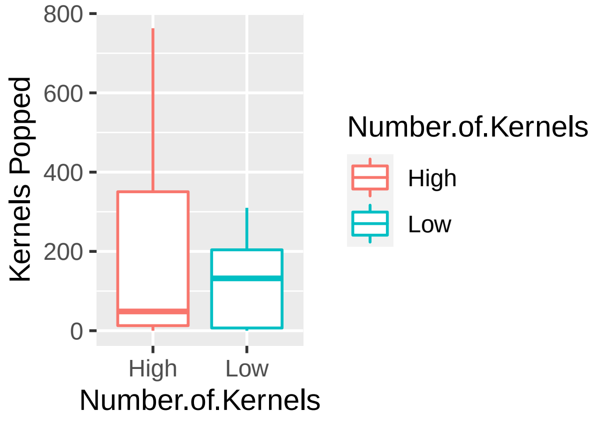
```
##             Df Sum Sq Mean Sq F value Pr(>F)
## Block        1   43.2   43.17   0.454   0.51
## Residuals   15 1424.9   94.99
```

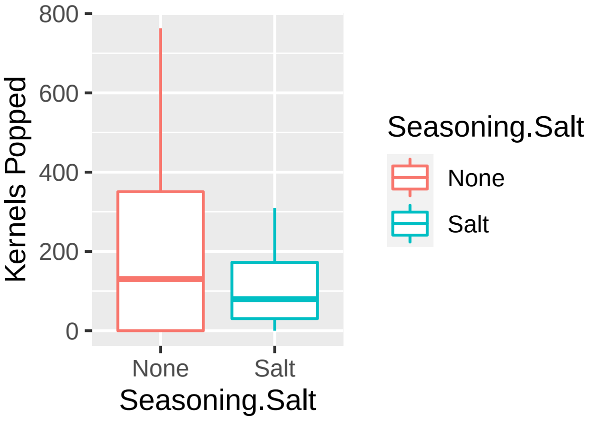


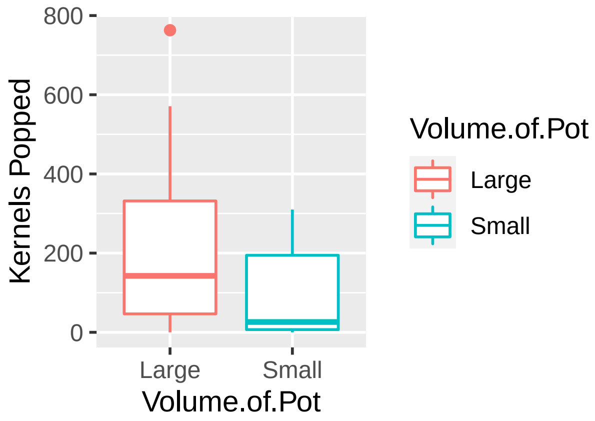
The distribution of kernels popped is highly skewed right

Initial Visualizations:

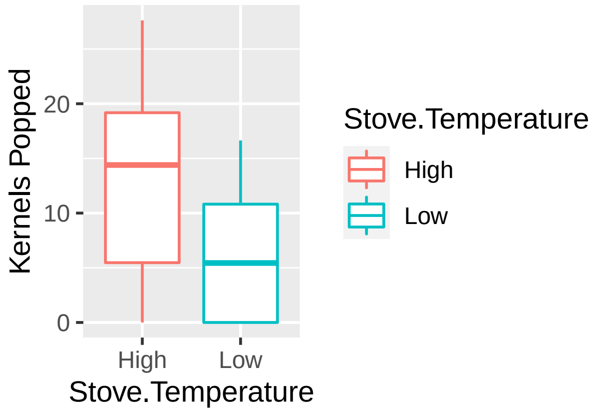


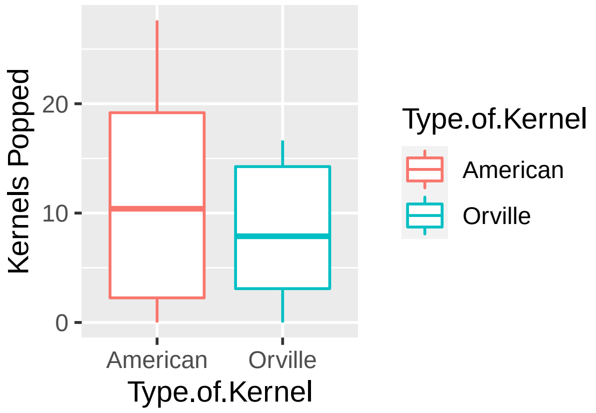

800
600
400
200
0
Kernels Popped
High
Low
Number.of.Kernels
Number.of.Kernels
High
Low

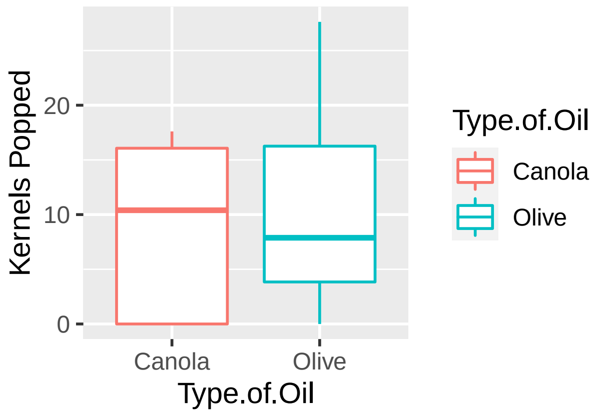

800
600
400
200
0
Kernels Popped
None
Salt
Seasoning.Salt
Seasoning.Salt
None
Salt

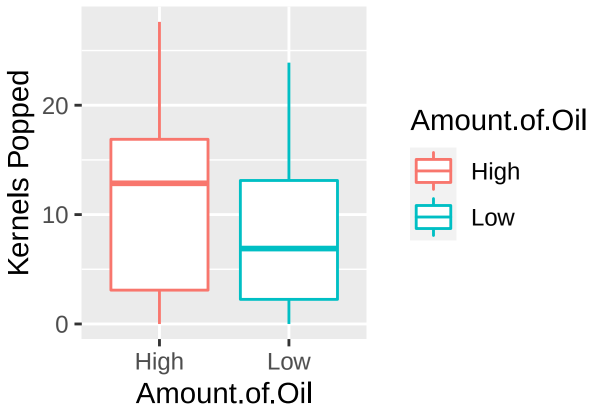

800
600
400
200
0
Kernels Popped
Large
Small
Volume.of.Pot
Volume.of.Pot
Large
Small

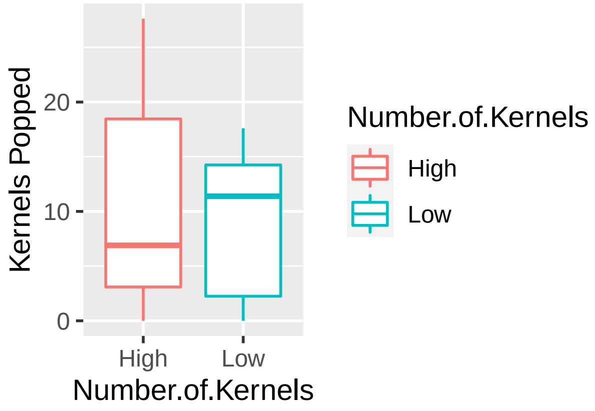
Transformed Visualizations:

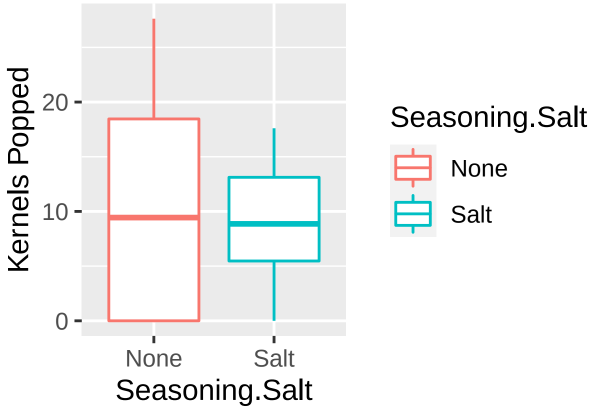


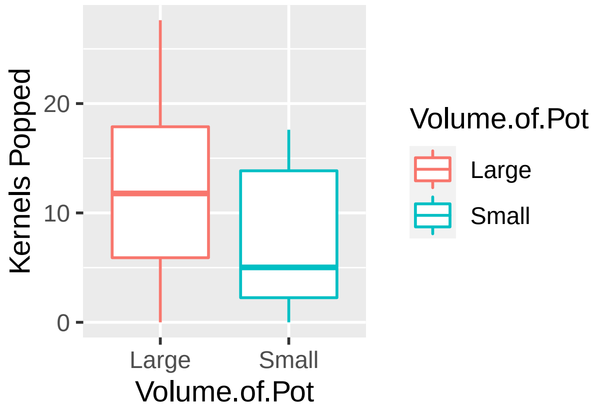




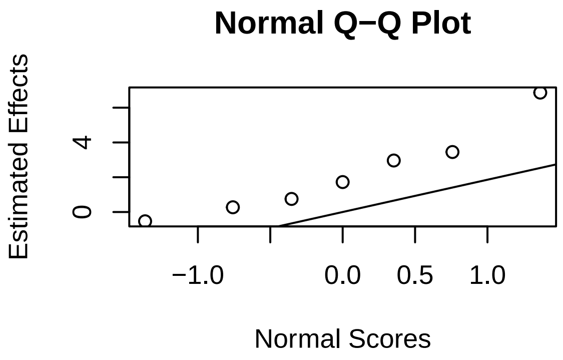


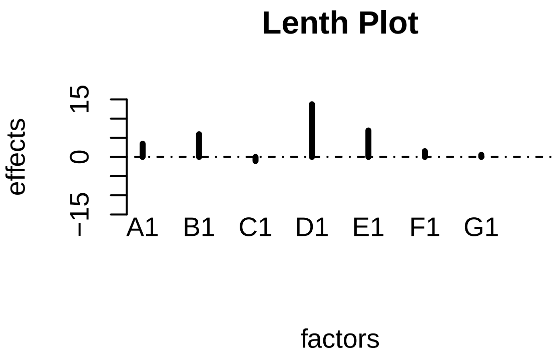

Kernels Popped
20
10
0
High
Low
Number.of.Kernels
Number.of.Kernels
High
Low

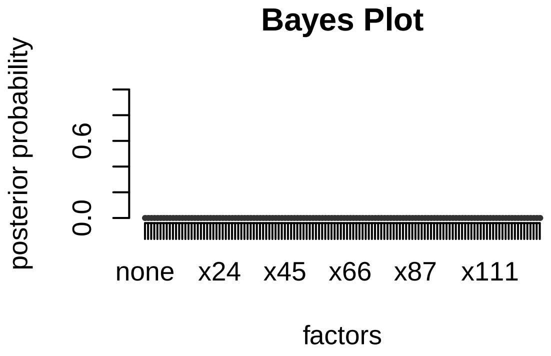

Kernels Popped
20
10
0
None
Salt
Seasoning.Salt
Seasoning.Salt
None
Salt


Kernels Popped
20
10
0
Large
Small
Volume.of.Pot
Volume.of.Pot
Large
Small

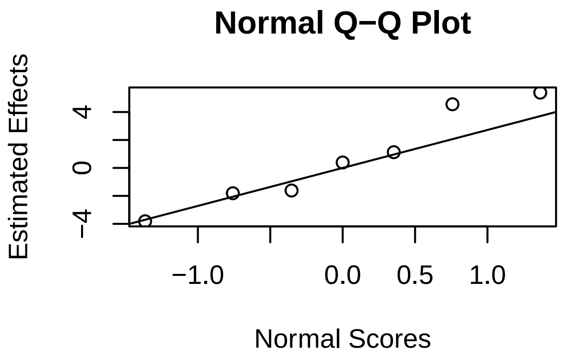
First 8 runs

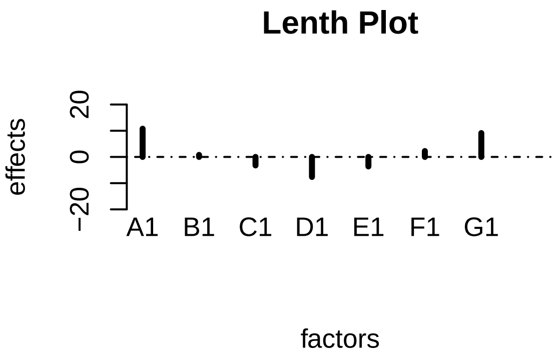

Normal Q−Q Plot
Estimated Effects
0
4
−1.0
0.0
0.5
1.0
Normal Scores

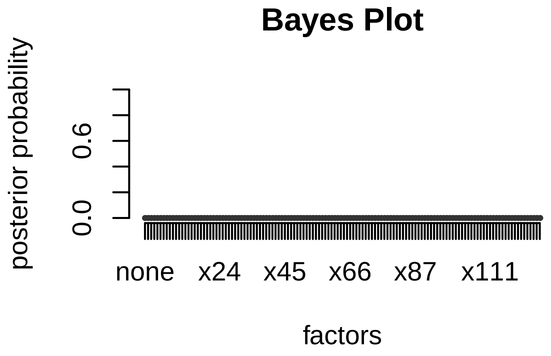

Lenth Plot
effects
15
0
−15
A1
B1
C1
D1
E1
F1
G1
factors


Bayes Plot
posterior probability
0.0
0.6
none
x24
x45
x66
x87
x111
factors

Second 8 runs



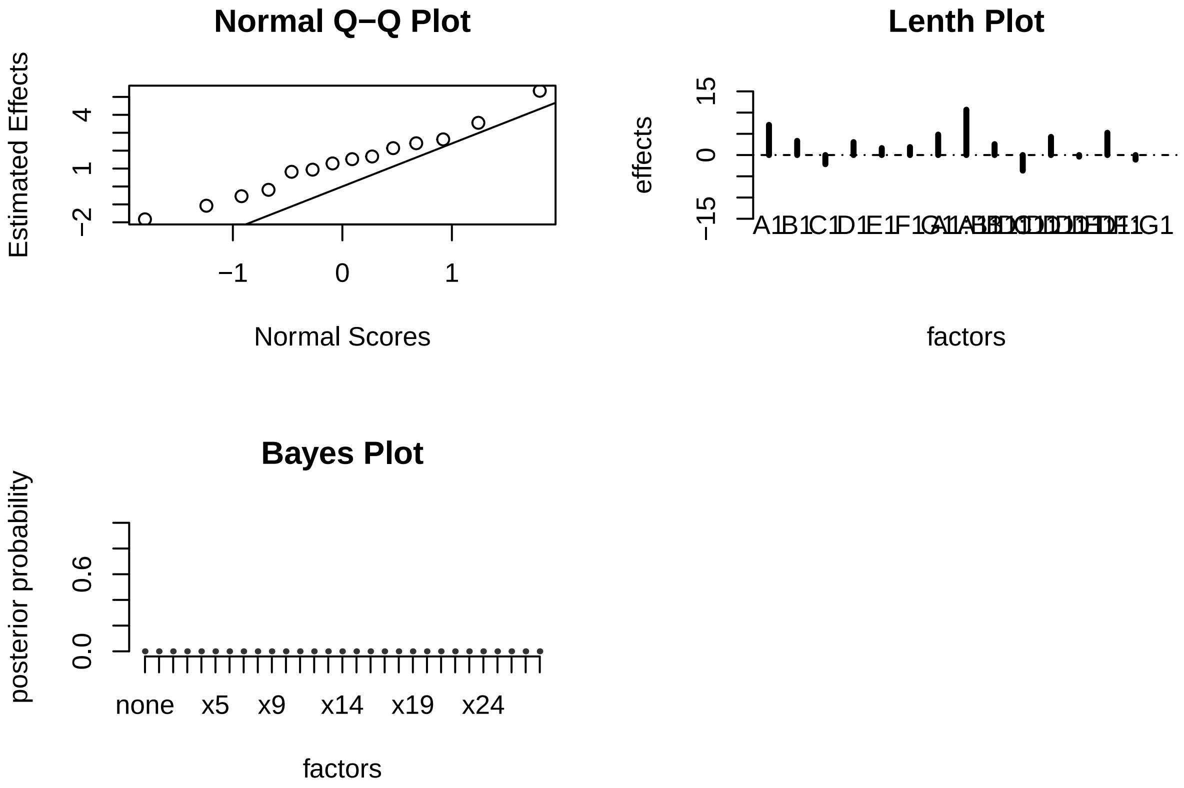




First 16 Runs

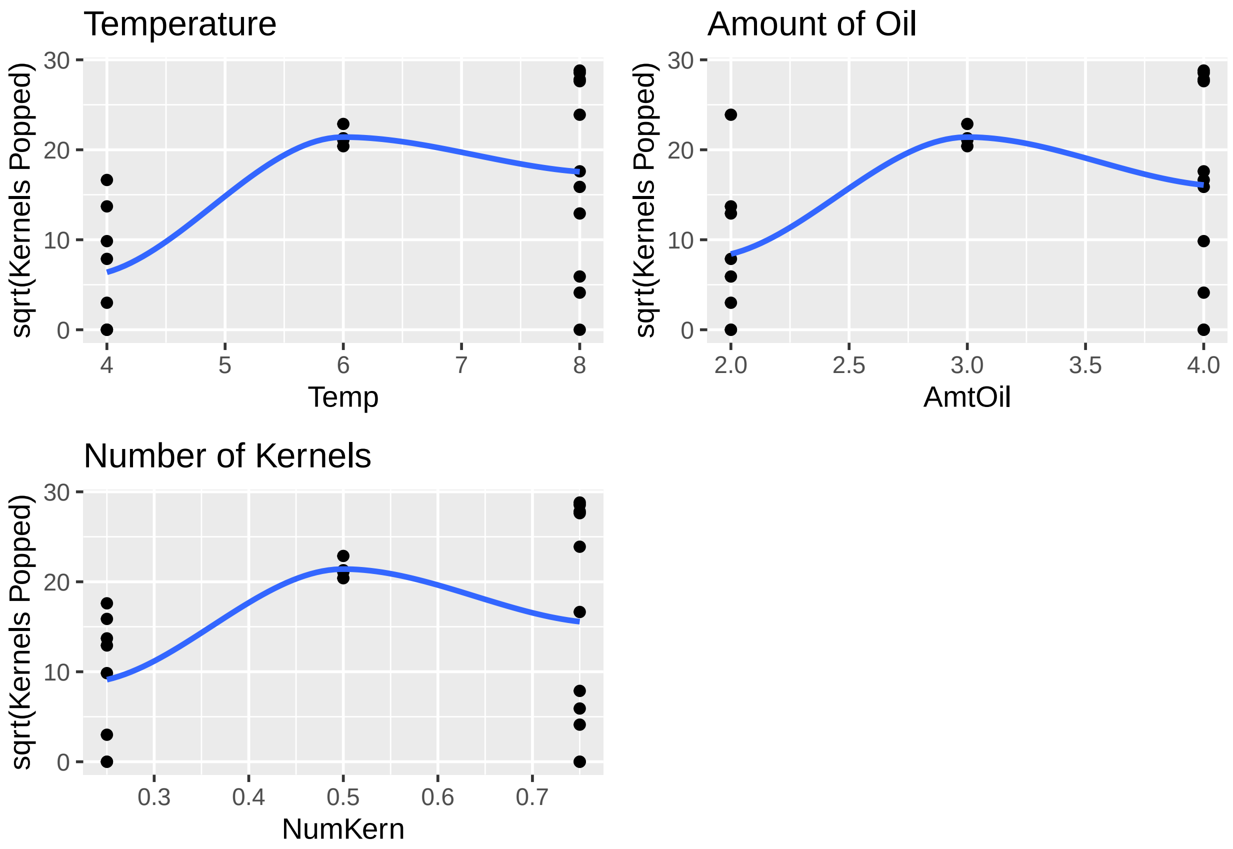
Alias Structure

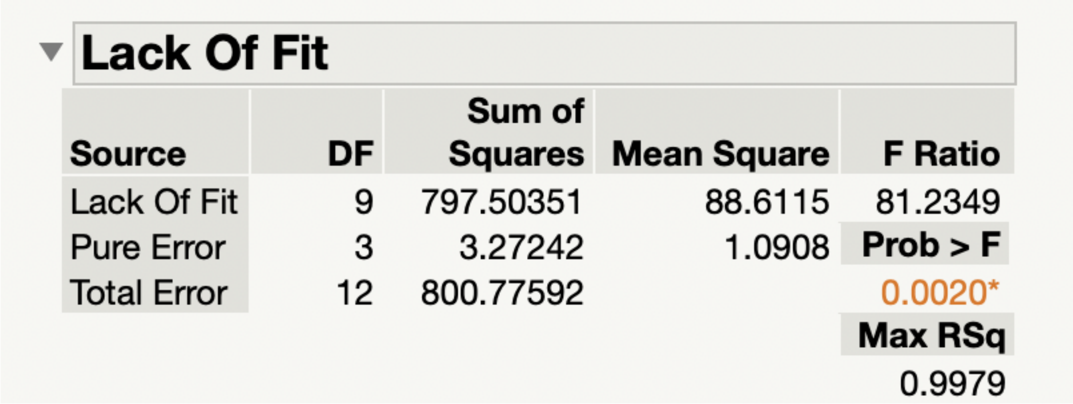
```
##
##  A = C:E = -F:G
##  B = C:F = -E:G
##  C = A:E = B:F
##  E = A:C = -B:G
##  F = -A:G = B:C
##  G = -A:F = -B:E
##  A:B = -C:G = E:F
```


Normal Q−Q Plot
Estimated Effects
−2
1
4
−1
0
1
Normal Scores
Lenth Plot
effects
−15
0
15
A1B1C1D1E1F1GA1A1B1D1D1D1D1DF1G1
factors
Bayes Plot
posterior probability
0.0
0.6
none
x5
x9
x14
x19
x24
factors

Appendix 2

Linearity Check:

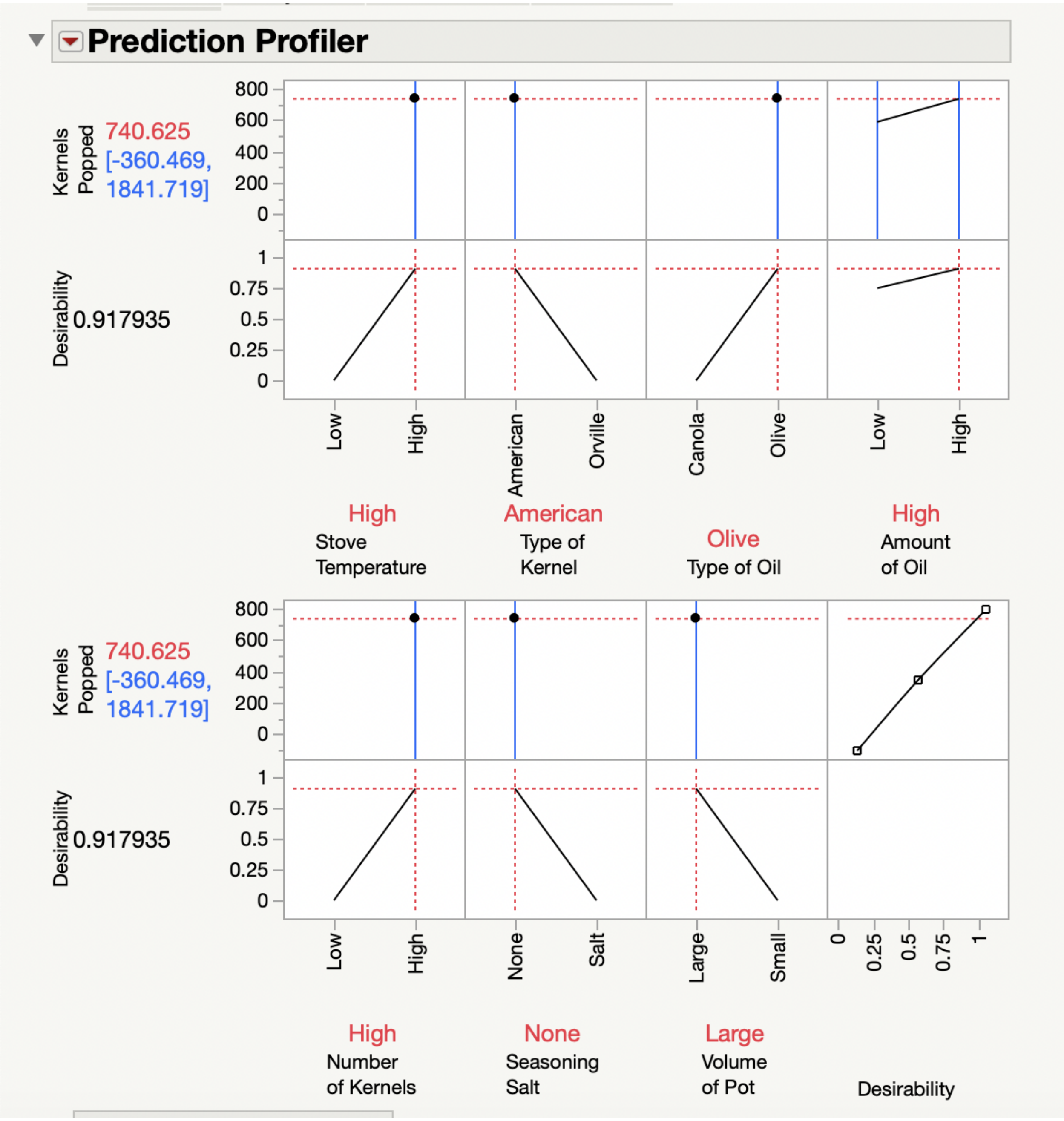




Lack Of Fit

Source	DF	Sum of Squares	Mean Square	F Ratio
Lack Of Fit	9	797.50351	88.6115	81.2349
Pure Error	3	3.27242	1.0908	**Prob > F**
Total Error	12	800.77592		0.0020*
				Max RSq
				0.9979

Appendix 3



Prediction Profiler

Kernels Popped 740.625 [-360.469, 1841.719]

Desirability 0.917935

High Stove Temperature

American Type of Kernel

Olive Type of Oil

High Amount of Oil

High Number of Kernels

None Seasoning Salt

Large Volume of Pot

Figure 1: JMP Prediction Profiler