Matthew Dahle

Stat 424 – Dis 312

The Perfect Shrimp Rigatoni Pasta Dish

**Introduction:** My favorite thing in the world to eat is pasta. I cook it several times a week, and I particularly like to eat rigatoni pasta with cooked shrimp. I want to know how I can make the best shrimp rigatoni possible, so I have decided to study the factors that go in to making this dish as delicious as I can using statistical analysis. There are many ways to prepare this dish, but for this project I decided for the factors to be the type of rigatoni noodle, the type of pasta sauce, and the type of olive oil used to cook the shrimp with. Everything else in the cooking process such as time spent cooking, amount of sauce, number of shrimps, and seasoning of shrimp is used the same way in each dish. I collect and analyze my data by using a full factorial design, with two levels for each factor. The levels for pasta noodle are either Trader Joe’s rigatoni (+), or rigatoni from Fresh Madison market (-). The level for sauces is Prego Garlic and Herb sauce (+), and Prego Vodka Italian Sauce (-). Lastly, the level for olive oil is garlic infused olive oil (+), and regular virgin olive oil (-). Using these factors and levels I was able to create my experimental factorial design, and cooked each dish using various level settings. The table below shows clearly the factors and levels used in this experiment.

|  |  |  |  |
| --- | --- | --- | --- |
| Factor | Variable | Low Level (-) | High Level (+) |
| Pasta | A | Fresh Madison Pasta | Trader Joe’s Pasta |
| Sauce | B | Vodka | Garlic and Herb |
| Olive Oil | C | Virgin Olive Oil | Garlic Olive Oil |

**Data Collection Process:**

In order to collect my data for this experiment, I used different factor levels seen in the table above when cooking the pasta and organized it as a full factorial design. This means I cooked 8 dishes of pasta, and each dish had different factor settings making 8 unique pasta dishes. I replicated this experiment 3 times in a span of 3 weeks in order to have plenty of data. I prepared 1-2 pasta dishes and then served them to 8 different people who in return gave each dish a rating between 1.0-10. The people selected to judge the pasta dishes are all friends of mine, and they all ate 3 different types of pasta, giving me 24 ratings for data. I wanted to make sure I could randomize the order in which each person was served the 3 dishes of pasta and on which day they all got it. For the first trial, I assigned each person a number 1-8 and used a random number generator to see which person would eat which dish and repeated this process 2 more times for the replications. The random number generators made sure each person was getting a different version of the pasta each trial. I cooked the dishes in order of the factorial design where the 1st dish had levels (-,-,-), and the 8th dish had levels (+,+,+). This gave me great randomization because each person was having 3 of the 8 dishes and were each served on different days. A problem with this experiment is that people have subjective opinions on food, so some ingredients may be more appealing to others, and this is a limitation of the study. Another limitation is that I am just and amateur cook, and do not have the precision or knowledge of a real chef when creating this dish. But with enough data points and averaging of scores, we can obtain fair results on how to make this dish as tasty as we can. The planning matrix on the next page will show how this factorial design is set up, and Table 1 will show the real data.

|  |  |  |  |
| --- | --- | --- | --- |
| **Planning Matrix** | | | |
| Trial Number | Pasta Noodle | Sauce | Olive Oil |
| 1 | Fresh Madison (-) | Vodka (-) | Virgin Olive Oil (-) |
| 2 | Trader Joe’s (+) | Vodka (-) | Virgin Olive Oil (-) |
| 3 | Fresh Madison (-) | Garlic and Herb (+) | Virgin Olive Oil (-) |
| 4 | Trader Joe’s (+) | Garlic and Herb (+) | Virgin Olive Oil (-) |
| 5 | Fresh Madison (-) | Vodka (-) | Garlic Olive Oil (+) |
| 6 | Trader Joe’s (+) | Vodka (-) | Garlic Olive Oil (+) |
| 7 | Fresh Madison (-) | Garlic and Herb (+) | Garlic Olive Oil (+) |
| 8 | Trader Joe’s (+) | Garlic and Herb (+) | Garlic Olive Oil (+) |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Table 1** | | | | | | |
| Factors | | | Replication | | | Average |
| Pasta (A) | Sauce (B) | Oil (C) | 1 | 2 | 3 | Y-Bar |
| - | - | - | 7.8 | 7.5 | 8 | 7.76666667 |
| + | - | - | 7.6 | 7.6 | 8 | 7.73333333 |
| - | + | - | 7.3 | 7.6 | 7.4 | 7.43333333 |
| + | + | - | 7.9 | 7.6 | 7.4 | 7.63333333 |
| - | - | + | 8 | 8.3 | 7.9 | 8.06666667 |
| + | - | + | 7.8 | 8.1 | 8.4 | 8.1 |
| - | + | + | 8.8 | 9 | 8.9 | 8.9 |
| + | + | + | 8.9 | 9.2 | 9 | 9.03333333 |

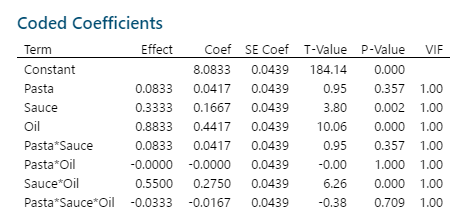
**Regression and Modeling:**

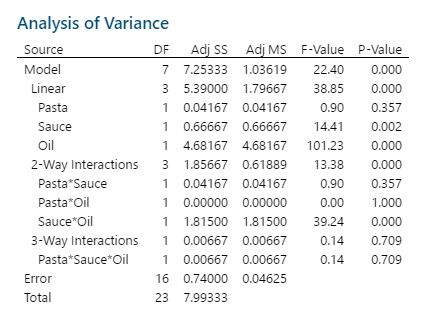
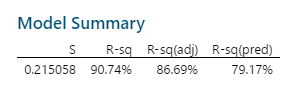
I followed a normal regression model where Y is the rating of the pasta dish, is the intercept, is the corresponding factor or interaction factor and is set to -1 if the level is (-) and 1 if the level is (+). β’s are the effect for each X main factor and interaction factors, and an error term . We divide the main effects and interaction effects by 2 to create the β coefficients. The general model is shown below:

After using some Minitab analysis, we come up with the regression equation of Y = 8.0833 + 0.0417 Pasta + 0.1667 Sauce + 0.4417 Oil + 0.0417 Pasta\*Sauce - 0.0000 Pasta\*Oil + 0.2750 Sauce\*Oil - 0.0167 Pasta\*Sauce\*Oil. Looking at the p-values of the coefficients, we see that sauce, olive oil, and the interaction effect of sauce and olive oil are significant. This makes sense according to the effect heredity principle, and since only 3 factors are significant this also confirms the effect sparsity principle. We can see from main effect plots that setting all the factors equal to the + level provides us with maximum results, with a maximum score of 9.033. I believe that all the factors are important in this regression since we need all the ingredients to create the dish, but if I wanted to follow the principle of Occam’s razor, I could use a regression model with only sauce, olive oil, and the interaction between the 2 as regressors.

|  |
| --- |
| Aliases |
| I |
| A |
| B |
| C |
| AB |
| AC |
| BC |
| ABC |

**Results:**

****

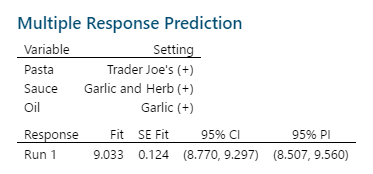












**Discussion of Results:**

There are countless ways to prepare a pasta dish, but by analyzing the data above, we learn a lot about the factors and what works. We can see from the half-normal plot and coded coefficients table that factors B (Sauce), C (Oil), and BC (Sauce\*Oil) are all significant factors in explaining the scores of the pasta dishes. The residual plots help us visualize the errors in the regression equation and judging on how there is no clear pattern among the residuals and the points are evenly dispersed around 0, we can assume this model is normally distributed. The histogram looks about normal, and the normal probability plot also has nothing very interesting that jumps out. Looking at the ANOVA table, we further see the importance factors B, C, and BC have on this model, with all being significant in explaining the variation within the data points. When we look at the interaction effect plots, we can see that our significant interaction BC has an antagonistic relationship. It makes sense that sauce and oil help create large effects, as they are the real factors that effect the taste of the dish, while type of noodle proved to be of lesser importance. Looking at the main effects plot also clearly explains this idea, with factor A having a very small positive slope, and factors B and C having very large positive slope, indicating the big effect they have on determining the score of the dish. More analysis can be done to better our understanding of this experiment, but all this analysis points that the best way to prepare shrimp rigatoni is with Trader Joe’s pasta, Garlic and Herb sauce, and garlic olive oil.