STA305H1S Sec L0201 , W2023

Factorial Design of Hotdogs Toppings

Yiliu Cao
1006664334
April 5, 2023

Introduction

Hotdogs are one of the most popular fast foods in Western countries. People love hotdogs as it is convenient and have a unique taste. I am also like hotdogs. However, I love the hotdogs because they have a variety of toppings, and you can choose your unique combination according to your reference.
At the University of Toronto, the hotdog is one of the most well-known foods, and the most popular one is "MAMA's Best," which is in front of Sidney Smith Hall. You can choose different toppings based on your flavor, including Mushrooms, Sweet Corn, Onions, etc. However, "Mushrooms, Onions, and Pickles" are the three most-consumed toppings. Therefore, it is worth detecting whether these three toppings indeed can make the hotdog more delicious or if it is simply a coincidence. In addition, there are two most popular types of hotdogs: All-Beef and Jumbo chicken.
In this project, I will conduct a blocking factorial design with the three most-consumed toppings as factors and “Taste Rating" as the outcome to find whether the toppings and their combinations and different types of hotdogs will affect the flavor of the hotdogs. Hence this experiment can help the students to make a better choice of toppings and improve the tasty of their hotdogs.

Materials and Method

As I mentioned, this experiment will be the 2^3 factorial design with two complete blocks to see whether the three toppings of hotdogs and the two kinds of hotdogs will influence the taste. The data used in this experiment will be generated by R instead of collected from an actual survey.

Setup

The three factors (toppings) in this experiment will be Mushrooms(M), Onions(O), and Pickles(P), and each of them only have two levels: adding or not adding, represented by 1 and -1, respectively. For instance, {M, O, P}={1, 0, 0} means only the mushrooms to the hotdogs. {M, O, P}={0, 1, 1} means the onions and pickles are added to the hotdogs.
In addition, this experiment contains two complete blocks, each corresponding to a type of hotdog: All-Beef(1) and Jumbo chicken(2). Since these are complete blocks, thus each block contains all the unique combinations of factors. In other words, every possible combination of toppings is considered for both All-beef and Jumbo Chicken hotdogs.
Moreover, I set the replication to be 50, meaning there will be 50 experimental units for each unique combination of factors in each block. This implies that the block size is $50 * 8 = 400$ for each block. Hence, the total number of experimental units will be $400 * 2 = 800$.

Data

Following the procedures described in the setup, we can construct the design matrix for the experiment by first creating the design matrix for the single replication and then replicating for

50 times. The probability of units being assigned to each factor combination is completely randomized. I will set the first half of the data in block 1 and the second half in block 2. Therefore, we have different combinations of factors and assigned the blocks. However, we still need the outcome “Taste Rating” of the hotdogs under various combinations of toppings. The “Taste Rating (Rating)” is the outcome of this experiment and is a quantitative variable.
As I mentioned, the data of outcomes used in this experiment is generated by R. To generate the data of ratings of hotdogs; I will assume that:
1) The initial rating is 50, meaning that the flavor of hotdogs without any toppings is 50.
2) The rating of a hotdog is cumulative by all the added toppings. This means that a hotdog’s final rating is calculated by the initial rating of 50 plus the effect of each topping.
3) The effect of each topping can increase, decrease or maintain the flavor of the hotdog. In other words, the impact of each topping can be positive, negative, or 0. However, the effect of each topping is highly to have positive effects on the ratings of the hotdogs in my settings. The exact mechanism is shown below.
4) The maximum flavor of hotdogs is 100, and the minimum is 0.
Based on the above three assumptions, we can generate the data of ratings by the following steps:
Suppose M, O, P are random variables representing the change in ratings of the hotdog for Mushrooms, Onions, and Pickles, respectively. Each random variable follows a Normal distribution: $M \sim N(15, 5^2)$, $O \sim N(15, 10^2)$ and $P \sim N(10, 5^2)$
With the above settings, we can calculate the expected value and variance of effects for any combinations of toppings. For example, the expected rating of the hotdog with mushrooms and pickles should follow the normal distribution with a mean of 50+15+10=75 and a variance of 25+25=50.
Following the above procedure, we can generate the ratings in our experiment. Combined with the design matrix in the setup we had earlier, we can get the data for this experiment.

Statistical Analysis

This experiment will consist of step-by-step statistical analysis to find whether the individual topping or their combinations affect the ratings of the hotdog.

1. Pre-check the interactions. First, we will pre-check the interactions visually for all pairs of factors by plotting the interaction plots.
2. Fit the full model. The full model is the model that contains all the factors and their interactions (i.e., factorial effects)
3. Summary the full model and implement the t-test. From the t-test, we can know whether the factors or their combinations are significant in predicting the ratings of the hotdogs.
4. Construct the ANOVA table for the full model and implement the F-test. From F-test, it can tell us whether there is a difference in the means of each effect.
5. If all the effects are significant by the t-test, go to step 7)
6. If some effects are not significant by the t-test. Then we need to construct a reduced model including only the significant effects and blocks. Then repeat steps 2) ~ 6).
7. Check the assumptions of ANOVA for the latest model. Now all the effects in the model should be significant, we need to check the assumptions of ANOVA. Violations of ANOVA may result in the bias of prediction.
8. Find the estimated main and interaction effects (i.e., factorial effects), variance and build the 95% confidence interval. Using the summary table, we can find the factorial effect estimates by doubling the coefficient estimates. Similarly, the estimated variance of factorial effects is

double the estimated standard error of coefficients and square it. Therefore we calculate the confidence interval as well.

9. Interpretations and conclusions.

Results and Discussion

Pre-check the interactions of all paired factor visually by the interaction plot.

Figure 1: The interaction between Mushrooms and Onions

Figure 2: The interaction between Mushrooms and Pickles

Figure 3: The interaction between Onions and Pickles

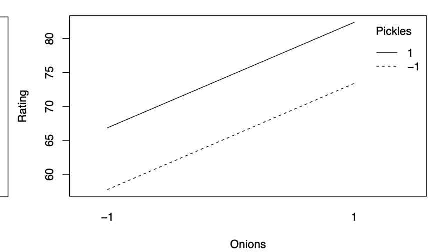


Figure 1&2&3 shows the interaction plots of M&O, M&P, and O&P. It seems that only M&O (the left one) shows an interaction effect between them. Then build the full model, which contains all the factors and their interactions, into the summary table and ANOVA table.

Table 1: The summary of the full model

Source	Estimate	Standard Error	T value	P values
Intercept	70.7200	0.4199	168.412	0.00000
Mushrooms	7.0675	0.2969	23.802	0.00000
Onions	7.7950	0.2969	26.252	0.00000
Pickles	4.5250	0.2969	15.239	0.00000
Mushrooms:Onions	-0.8000	0.2969	-2.694	0.00720
Mushrooms:Pickles	-0.1450	0.2969	-0.488	0.62545
Onions:Pickles	-0.0275	0.2969	-0.093	0.92623
Mushrooms:Onions:Pickles	-0.2625	0.2969	-0.884	0.37694
Blocks	-1.2450	0.5939	-2.096	0.03636

Table 2: The ANOVA table of the full model

Source	Degree of Freedom	Sum of Squares	Mean Squares	F value	P value
Mushrooms	1	39960	39960	566.5335	0.00000
Onions	1	48610	48610	689.1697	0.00000
Pickles	1	16381	16381	232.2368	0.00000
Mushrooms:Onions	1	512	512	7.2590	0.00720
Mushrooms:Pickles	1	17	17	0.2385	0.62545
Onions:Pickles	1	1	1	0.0086	0.92623
Mushrooms:Onions:Pickles	1	55	55	0.7815	0.37694
Blocks	1	310	310	4.3951	0.03636
Residuals	791	55792	71	NA	NA

Table 1 shows the coefficients' estimates, standard error, and P values in the full model. Using the significance level 0.05, we can observe that only the effects of M, P, O, and M:O have a p-value less than 0.05, meaning that these effects are significant in predicting the flavor of hotdogs, and hence we should keep them.

Table 2 shows that the significant effects both have a p-value less than 0.05, meaning there is a significant difference between the means of different levels of these effects.

Hence, from the summary and ANOVA table of the full model, it is reasonable only to keep the factorial effects of M, P, O, and M:P. According to the statistical analysis we designed, we need to construct a reduced model.

Table 3: The summary of the reduced model

Source	Estimate	Standard Error	T value	P values
Intercept	70.7200	0.4194	168.622	0.00000
Mushrooms	7.0675	0.2966	23.832	0.00000
Onions	7.7950	0.2966	26.285	0.00000
Pickles	4.5250	0.2966	15.258	0.00000
Mushrooms:Onions	-0.8000	0.2966	-2.698	0.00713
Blocks	-1.2450	0.5931	-2.099	0.03613

Table 4: The ANOVA table of the reduced model

Source	Degree of Freedom	Sum of Squares	Mean Squares	F value	P value
Mushrooms	1	39959.645	39959.64500	567.943634	0.0000000
Onions	1	48609.620	48609.62000	690.885122	0.0000000
Pickles	1	16380.500	16380.50000	232.814898	0.0000000
Mushrooms:Onions	1	512.000	512.00000	7.277020	0.0071324
Blocks	1	310.005	310.00500	4.406079	0.0361263
Residuals	794	55864.625	70.35847	NA	NA

Table 3 shows the summary of the reduced model, which only contains the significant effects and the block. Compared to the full model, we can see that all the effects in the reduced model are significant in predicting the flavor of hotdogs. Table 4 shows that the p-values of all the factorial effects are less than 0.05, meaning there is a significant difference in the means of different factors and their interactions.

In addition, we can also observe that the block effect has a p value less than 0.05 in the summary table of both the full and reduced model. This indicates that there is a statistically significant difference in the flavor of the different hotdogs (All-Beef and Jumbo Chicken).
Meanwhile, we also need to checking the assumptions of constancy of error variance for the ANOVA of the reduced model.

Figure 4: The distribution of residuals across all fitted values

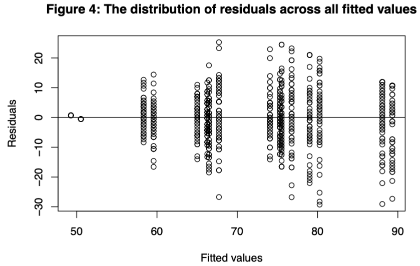
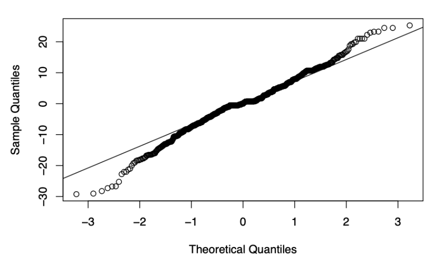


Figure 5: The normal QQ plot of residuals



From the Figure 4 and Figure 5, we can see that the constancy of variance is basically satisfied except the two obvious outliers. In addition, it seems that our data is light tailed and these may influence the precision of our model.
Furthermore, we can estimate the factorial effects, their estimated variance and 95% confidence interval.

Table 5: The summary of the reduced model

Effects	Factorial Effects	Estimated Variance	95%Confidence Interval
Mushrooms	14.135	0.3518862	(12.971, 15.299)
Onions	15.590	0.3518862	(14.426, 16.754)
Pickles	9.050	0.3518862	(7.886, 10.214)
Mushrooms:Onions	-1.600	0.3518862	(-2.764, -0.436)

From table 5, we can find the estimates of main and interaction effects in our reduced model. The main effects are close to our setting, which meets the expectations

Using the results we have above, we can build the linear model:
$\hat{y} = 70.72 + \frac{14.135}{2}M + \frac{15.59}{2}O + \frac{9.05}{2}P - \frac{1.6}{2}M * O - 1.245Blk$, where {M, O, P} represent add(1) or not add(-1) the corresponding topping. In other words, the increase in the flavor of hotdogs for M, O and P is 14.135/2, 15.590/2 and 9.050/2 respectively. However, if you only choose to add or not add both of mushrooms and onion, then the flavor will both decrease extra 0.8. The flavor will only increase if you only add one of them. Besides, *Blk* represent the types of the hotdog, 0 mens All-beef and 1 means Jumbo-chicken. For example, if you buy a Jumbo-chicken and want to add mushrooms and pickles, then flavor of this hotdog is
$70.72 + \frac{14.135}{2} + \frac{9.05}{2} - \frac{1.6}{2}(1)(-1) - \frac{2.49}{2} = 81.8675.$

Conclusion

To sum up, we found that not all the factorial effects are significant. The only significant effects are M, O, P and M:O. In addition, the main effect meet our initial settings. However, it is surprisingly that the interaction effect for mushrooms and onions is negative, meaning that the the hotdog will be more tasty if you only add one of them. With this results, students can know that the three most-consumed toppings can indeed increase the tase of the hotdogs. They can also know how to have a more delicious hotdog.
However, there are still some limitations about this experiment. The replication may not be large enough and assumptions of ANOVA seem not satisfy perfectly. The accuracy may be influenced.