Sam Thomas

4/25/22

Textural properties of mango soy fortified probiotic yoghurt: optimisation of inoculum level of yoghurt and probiotic culture

Data Set

Run	ST (%)	LB (%)	BB (%)	LA (%)	Hardness	Cohesiveness	Adhesiveness	Springiness	Gumminess
1	2	2	2	2	71.38	0.42	−65.58	0.87	30.28
2	2	2	2	1	74.84	0.44	−88.26	0.93	32.82
3	2	2	1	2	78.04	0.42	−75.88	0.81	32.6
4	2	2	1	1	92.35	0.42	−95.40	0.99	38.67
5	2	1	2	2	80.56	0.44	−124.17	0.79	35.26
6	2	1	2	1	85.07	0.42	−127.97	1	35.56
7	2	1	1	2	84.18	0.44	−119.81	0.97	36.77
8	2	1	1	1	99.12	0.43	−145.69	1.07	42.79
9	1	2	2	2	91.8	0.46	−213.01	0.98	42.28
10	1	2	2	1	92.53	0.45	−220.60	0.98	41.38
11	1	2	1	2	90.59	0.43	−172.67	0.97	38.71
12	1	2	1	1	89.37	0.42	−148.83	0.69	37.41
13	1	1	2	2	95.78	0.43	−169.30	0.51	40.8
14	1	1	2	1	109.14	0.41	−206.20	0.88	44.85
15	1	1	1	2	102.43	0.42	−164.33	0.75	42.87
16	1	1	1	1	100.64	0.41	−167.17	0.74	41.4
17	2.5	1.5	1.5	1.5	88.88	0.42	−84.63	0.97	37.03
18	0.5	1.5	1.5	1.5	90.32	0.4	−103.34	0.88	36.15
19	1.5	2.5	1.5	1.5	78.97	0.41	−101.28	0.96	32.53
20	1.5	0.5	1.5	1.5	86.52	0.4	−119.30	1	34.65
21	1.5	1.5	2.5	1.5	87.45	0.43	−136.24	1.03	37.26
22	1.5	1.5	0.5	1.5	108.95	0.42	−130.42	0.88	46.02
23	1.5	1.5	1.5	2.5	83.62	0.43	−129.33	1	35.71
24	1.5	1.5	1.5	0.5	107.11	0.42	−136.67	1.02	44.93
25	1.5	1.5	1.5	1.5	73.65	0.46	−118.72	0.82	33.68

26	1.5	1.5	1.5	1.5	73.65	0.46	−118.70	0.82	33.68
27	1.5	1.5	1.5	1.5	72.78	0.45	−117.67	0.81	33.56
28	1.5	1.5	1.5	1.5	73.65	0.46	−118.45	0.82	33.68
29	1.5	1.5	1.5	1.5	73.66	0.46	−118.46	0.82	33.66
30	1.5	1.5	1.5	1.5	72.89	0.46	−116.89	0.8	32.89
31	1.5	1.5	1.5	1.5	72.65	0.46	−116.67	0.81	33.34

Objective

Yogurt is a product that is created with blends of two yogurt bacteria: *Streptococcus thermophilus* (ST) and *Lactobacillus bulgaricus* (LB). Additionally, probiotic yogurt contains two probiotic cultures named *Bifidobacterium bifidus* (BB) and *Lactobacillus acidophilus (LA)*. These probiotic cultures have many positive effects on health. Further, soy-based dairy products provide other benefits such as lowering the cholesterol level of yogurt.

In this study, instrumental texture profile analysis was used to select the correct level of bacteria culture in mango soy fortified probiotic yogurt (MSFPY) such that optimal texture is reached. To do so, 5 different response variables (Hardness, Cohesiveness, Adhesiveness, Springiness, Gumminess) were observed and compared to the control mango soy fortified yogurt (MSFY). I will assess the cohesiveness response. Cohesiveness is defined as, "the extent to which a material can be deformed before its rupture" (Kaur et al., 2009: 420).

Data collection

Data was collected by creating probiotic yogurt samples using: 78.3% toned milk, 14.5% soy milk, and 7.2% mango pulp. First, the soy milk and toned milk was pasteurized at 95°C for 5 min. Then, this mixture was cooled to 45°C, and the mango pulp was added via high-speed stirring. After that, the mixture was inoculated with freeze-dried cultures of ST, LB, BB, and LA at different levels. These levels were between 0.5% - 2.5% for each culture. Then, runs were set up using a central composite rotatable design, meaning there were 31 runs.

After that, texture profile analysis was conducted using TA.XT2 Texture Analyzer equipped with a 50kg load cell. This measured the hardness, cohesiveness, adhesiveness, springiness, and gumminess of each yogurt sample.

Model and Assumptions

The model used for this analysis is the Response Surface Model. The model used was a second-degree polynomial with the objective to optimize the response. Pictured below is the expected value of the polynomial with its surface derivatives.

$$E\{Y_i\} = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4}$$
$$+ \beta_{12} X_{i1} X_{i2} + \beta_{13} X_{i1} X_{i3} + \beta_{23} X_{i2} X_{i3} + \beta_{14} X_{i1} X_{i4} + \beta_{24} X_{i2} X_{i4} + \beta_{34} X_{i3} X_{i4}$$
$$+ \beta_{11} X_{i1}^2 + \beta_{22} X_{i2}^2 + \beta_{33} X_{i3}^2 + \beta_{44} X_{i4}^2 \quad i = 1, \ldots, n$$

Surface Derivatives wrt X_1, X_2, X_3 :

$$\frac{\partial E\{Y\}}{\partial X_1} = \beta_1 + \beta_{12} X_2 + \beta_{13} X_3 + 2\beta_{11} X_1 + \beta_{14} X_4$$

$$\frac{\partial E\{Y\}}{\partial X_2} = \beta_2 + \beta_{12} X_1 + \beta_{23} X_3 + 2\beta_{22} X_2 + \beta_{24} X_4 \qquad \frac{\delta E\{Y\}}{\delta X_4} = \beta_4 + \beta_{14} X_1 + \beta_{24} X_2 + \beta_{34} X_3 + 2\beta_{44} X_4$$

$$\frac{\partial E\{Y\}}{\partial X_3} = \beta_3 + \beta_{13} X_1 + \beta_{23} X_2 + 2\beta_{33} X_3 + \beta_{34} X_4$$

Goal: Choose X_1^*, X_2^*, X_3^* that optimize (maximize or minimize) response $(and\ X_4^*)$

Predictors can be "coded" or centered, but not needed in rsm package in R

Below is the model written in matrix form:

Second Order Polynomial With Interaction (k Factors)

$$\hat{Y} = \hat{\beta}_0 + \sum_{j=1}^{k} \hat{\beta}_j X_j + \sum_{j=1}^{k-1} \sum_{j'=j+1}^{k} \hat{\beta}_{jj'} X_j X_{j'} + \sum_{j=1}^{k} \hat{\beta}_{jj} X_j^2 = \hat{\beta}_0 + \mathbf{x}'\boldsymbol{\beta}_1 + \mathbf{x}'\boldsymbol{\beta}_2\, \mathbf{x}$$

where: $\mathbf{x} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_k \end{bmatrix}$ $\quad \hat{\boldsymbol{\beta}}_1 = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_k \end{bmatrix}$ $\quad \hat{\boldsymbol{\beta}}_2 = \begin{bmatrix} \hat{\beta}_{11} & \frac{1}{2}\hat{\beta}_{12} & \cdots & \frac{1}{2}\hat{\beta}_{1k} \\ \frac{1}{2}\hat{\beta}_{12} & \hat{\beta}_{22} & \cdots & \frac{1}{2}\hat{\beta}_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{2}\hat{\beta}_{1k} & \frac{1}{2}\hat{\beta}_{2k} & \cdots & \hat{\beta}_{kk} \end{bmatrix}$

Obtaining the optimal setting $\mathbf{x}^*$: $\dfrac{\partial \hat{Y}}{\partial \mathbf{x}} = \dfrac{\partial}{\partial \mathbf{x}}\left(\hat{\beta}_0 + \mathbf{x}'\hat{\boldsymbol{\beta}}_1 + \mathbf{x}'\hat{\boldsymbol{\beta}}_2\, \mathbf{x} \right) = \hat{\boldsymbol{\beta}}_1 + 2\hat{\boldsymbol{\beta}}_2\, \mathbf{x}$

setting to 0 $\quad \Rightarrow \quad 2\hat{\boldsymbol{\beta}}_2\, \mathbf{x}^* = -\hat{\boldsymbol{\beta}}_1 \quad \Rightarrow \quad \mathbf{x}^* = -\dfrac{1}{2}\hat{\boldsymbol{\beta}}_2^{-1}\hat{\boldsymbol{\beta}}_1$

Full Model

Cohesiveness was the response variable that was analyzed. All analysis was conducted in R using the rsm package. Note that an alpha level of 0.05 was used for all analysis. A summary of the full model with all first order, two-way interaction, and quadratic terms is pictured below:

```
rsm(formula = Cohesiveness ~ SO(ST, LB, BB, LA))

              Estimate  Std. Error t value  Pr(>|t|)
(Intercept)  2.6786e-03  6.1819e-02  0.0433  0.965975
ST           2.0655e-01  3.2077e-02  6.4391 8.184e-06 ***
LB           1.7988e-01  3.2077e-02  5.6078 3.927e-05 ***
BB           7.6548e-02  3.2077e-02  2.3864  0.029714 *
LA           1.1988e-01  3.2077e-02  3.7373  0.001795 **
ST:LB       -3.0000e-02  9.2783e-03 -3.2334  0.005200 **
ST:BB       -1.5000e-02  9.2783e-03 -1.6167  0.125490
ST:LA       -1.0000e-02  9.2783e-03 -1.0778  0.297111
LB:BB        2.0000e-02  9.2783e-03  2.1556  0.046688 *
LB:LA       -1.5000e-02  9.2783e-03 -1.6167  0.125490
BB:LA        1.0453e-15  9.2783e-03  0.0000  1.000000
ST^2        -4.0238e-02  6.9403e-03 -5.7978 2.722e-05 ***
LB^2        -4.5238e-02  6.9403e-03 -6.5182 7.084e-06 ***
BB^2        -2.5238e-02  6.9403e-03 -3.6365  0.002221 **
LA^2        -2.5238e-02  6.9403e-03 -3.6365  0.002221 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Multiple R-squared:  0.8764,   Adjusted R-squared:  0.7682
F-statistic: 8.102 on 14 and 16 DF,  p-value: 8.171e-05
```

Here, we see that all first order and quadratic terms are significant. Additionally, the ST and LB two-way interaction is significant as well as the LB and BB two-way interaction. Another thing to note is an Adjusted R-squared of 0.7682.

Additionally, the ANOVA is pictured below:

```
Analysis of Variance Table

Response: Cohesiveness
                     Df    Sum Sq     Mean Sq F value    Pr(>F)
FO(ST, LB, BB, LA)    4 0.0010167 0.00025417  2.9525  0.052791
TWI(ST, LB, BB, LA)   6 0.0018500 0.00030833  3.5817  0.019167
PQ(ST, LB, BB, LA)    4 0.0068979 0.00172447 20.0319 4.517e-06
Residuals            16 0.0013774 0.00008609
Lack of fit          10 0.0012917 0.00012917  9.0417  0.006974
Pure error            6 0.0000857 0.00001429

Stationary point of response surface:
      ST       LB       BB       LA
1.436188 1.629221 1.735258 1.606317
```

Similarly, we see significance of the quadratic terms and two-way interactions. Interestingly, the first order terms are just outside of our p-value of 0.05.

The stationary point of response surface in this example gives us our optimized response. In this case, the stationary point gives us the maximum cohesiveness.

Another interesting part of the ANOVA is that the Lack of fit is significant. This may point to a different model being better suited to this data. However, because of the nature of the experiment and measurements, there are small variances which create the significance in the Lack of fit. For this reason, we can continue using this type of model.

Reduced Model

Next, we reduce the model and only keep the terms that were significant in the full model. Specifically, we keep all first order and quadratic terms. Additionally, we keep the ST:LB and LB:BB two-way interactions. The summary of the reduced model is pictured below:

```
rsm(formula = Cohesiveness ~ FO(ST, LB, BB, LA) + TWI(ST, LB) ·
    TWI(LB, BB) + PQ(ST, LB, BB, LA))

              Estimate Std. Error t value  Pr(>|t|)
(Intercept)  0.0926786  0.0482351  1.9214  0.069049 .
ST           0.1690476  0.0267989  6.3080 3.706e-06 ***
LB           0.1573810  0.0305779  5.1469 4.908e-05 ***
BB           0.0540476  0.0267989  2.0168  0.057344 .
LA           0.0823810  0.0223908  3.6792  0.001487 **
ST:LB       -0.0300000  0.0098168 -3.0560  0.006236 **
LB:BB        0.0200000  0.0098168  2.0373  0.055068 .
ST^2        -0.0402381  0.0073431 -5.4797 2.305e-05 ***
LB^2        -0.0452381  0.0073431 -6.1606 5.099e-06 ***
BB^2        -0.0252381  0.0073431 -3.4370  0.002609 **
LA^2        -0.0252381  0.0073431 -3.4370  0.002609 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Multiple R-squared:  0.827,     Adjusted R-squared:  0.7405
F-statistic: 9.562 on 10 and 20 DF,  p-value: 1.214e-05
```

Now, all terms are significant except for the LB:BB interaction and the BB first order term. That being said, their p-values are very close to being below the 0.05 threshold for alpha. Also, the Adjusted R-squared for the reduced model is 0.7405. This is not a large decrease compared to the Adjusted R-squared of the full model at 0.7682.

Next, is the ANOVA for the reduced model:

```
Response: Cohesiveness
                    Df    Sum Sq     Mean Sq F value      Pr(>F)
FO(ST, LB, BB, LA)   4 0.0010167 0.00025417  2.6374    0.064403
TWI(ST, LB)          1 0.0009000 0.00090000  9.3391    0.006236
TWI(LB, BB)          1 0.0004000 0.00040000  4.1507    0.055068
PQ(ST, LB, BB, LA)   4 0.0068979 0.00172447 17.8945 2.176e-06
Residuals           20 0.0019274 0.00009637
Lack of fit         14 0.0018417 0.00013155  9.2083    0.006035
Pure error           6 0.0000857 0.00001429

Stationary point of response surface:
      ST        LB        BB        LA
1.495841 1.622267 1.713540 1.632075
```

Once again, the quadratic terms are significant. Also, the ST:LB interaction is significant while the LB:BB interaction is not significant. Interestingly, the First Order terms are not significant according to the ANOVA.

Again, the stationary point of response surface gives us our maximum for the cohesiveness response.

Also, Lack of fit is significant once again. However, as explained in the full model, this is not a big deal due to the nature of the experiment.

Model Comparison

To compare the two models, the following hypotheses will be tested:

H_0: $\beta_{ST,LB} = \beta_{ST,LA} = \beta_{LB,LA} = \beta_{BB,LA} = 0$ vs. H_A: Not all $\beta_{ij} = 0$

Then, the Test Statistic is the following: $TS: F^* = \dfrac{[\frac{SSE(R)-SSE(F)}{df_{E(R)}-df_{E(F)}}]}{[\frac{SSE(F)}{df_{E(F)}}]}$

The rejection region is: $F^* > F(0.95,4,16) = 3.007$

Below is the ANOVA showing the comparison between the reduced model (Model 1) and the full model (Model 2):

```
Analysis of Variance Table

Model 1: Cohesiveness ~ FO(ST, LB, BB, LA) + TWI(ST, LB) + TWI(LB, BB) +
    PQ(ST, LB, BB, LA)
Model 2: Cohesiveness ~ FO(ST, LB, BB, LA) + TWI(ST, LB, BB, LA) + PQ(ST,
    LB, BB, LA)
  Res.Df       RSS Df Sum of Sq      F Pr(>F)
1     20 0.0019274
2     16 0.0013774  4   0.00055 1.5972 0.2233
```

We see that F* = 1.5972. Therefore, our test statistic does not fall into the rejection region, and we fail to reject H0. Hence, we continue to use the reduced model.

Our final equation for the reduced model appears as follows:

$$\hat{Y} = 0.092 + 0.169ST + 0.157LB + 0.054BB + 0.082LA - 0.030(ST * LB) + 0.020(LB * BB) - 0.040ST^2 - 0.045LB^2 - 0.025BB^2 - 0.025LA^2$$

Contour Plots

To visualize the findings, contour plots were created in R. Since there were 4 predictor variables, there are 6 contour plots as viewed below:
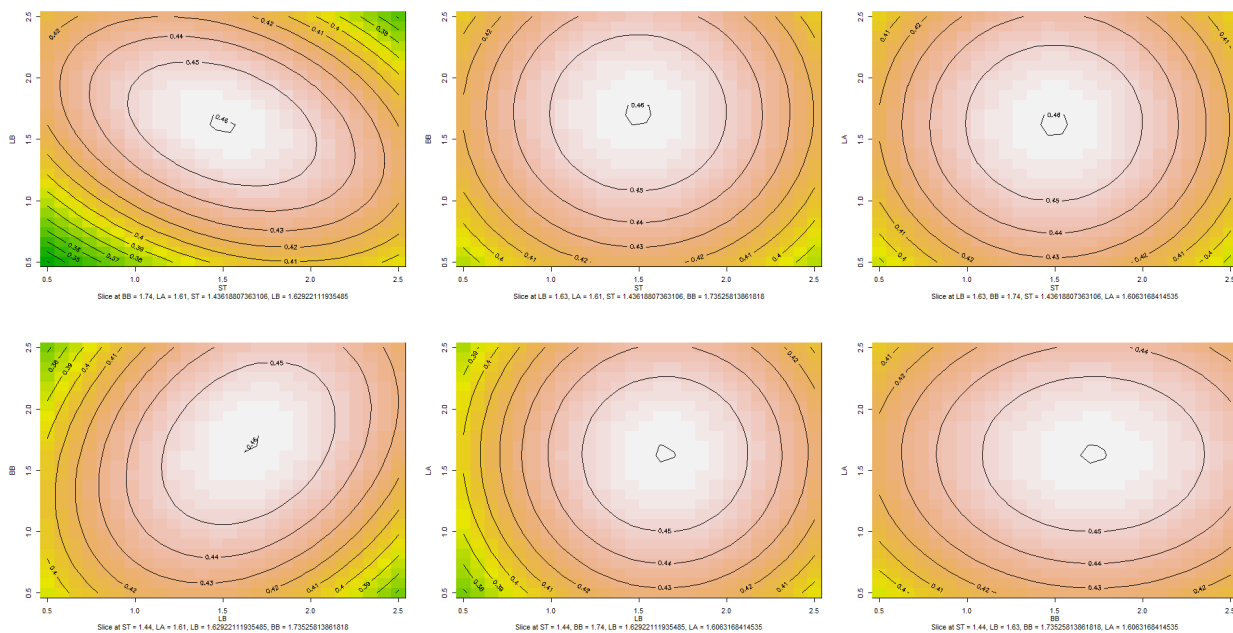
Each contour plot is very similar. The general trend is that cohesiveness is maximized whenever the level of each bacterium is either at or slightly above 1.5%. Then, cohesiveness decreases smoothly and evenly as bacteria levels go further from that 1.5% mark.

3D Model Plots

The 3D Model plots give another visualization of the results. In 3-dimensional form, we see that cohesiveness is maximized whenever the level of each bacterium is either at or slightly above 1.5%. Then, cohesiveness smoothly decreases as bacteria levels go further from the 1.5% mark. This trend allows all the 3D plots to appear as a concave-down hill.
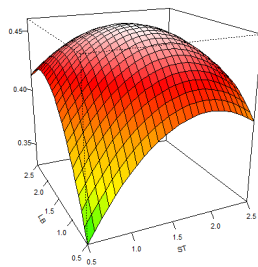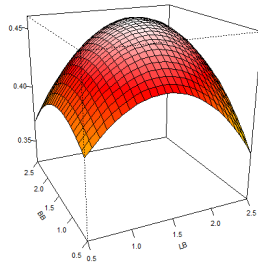
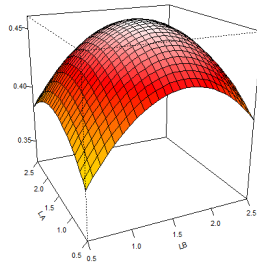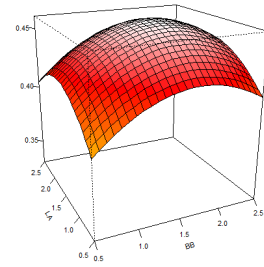Slice at BB = 1.74, LA = 1.61, ST = 1.43618807363106, LB = 1.62922111935485   Slice at LB = 1.63, LA = 1.61, ST = 1.43618807363106, BB = 1.73525813861818   Slice at LB = 1.63, BB = 1.74, ST = 1.43618807363106, LA = 1.6063168414535
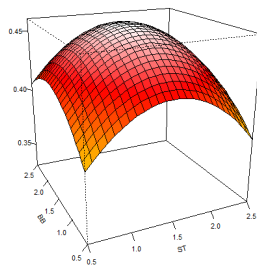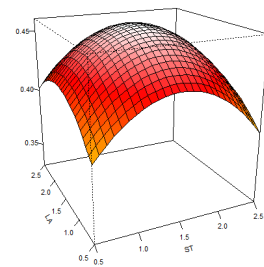
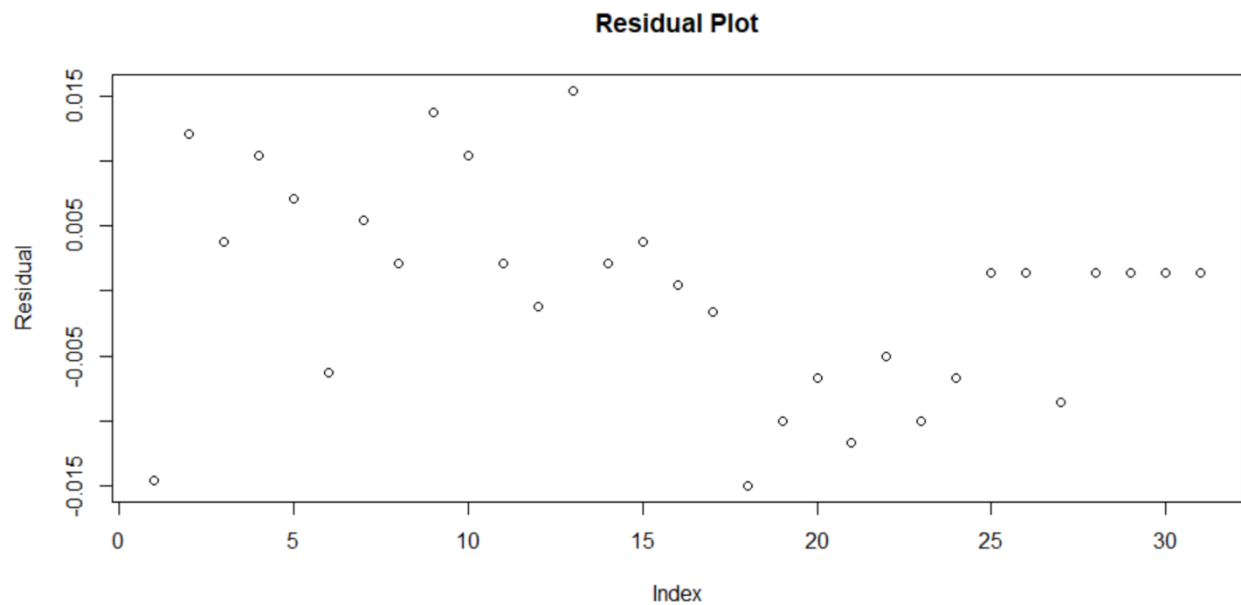Slice at ST = 1.44, LA = 1.61, LB = 1.62922111935485, BB = 1.73525813861818   Slice at ST = 1.44, BB = 1.74, LB = 1.62922111935485, LA = 1.6063168414535   Slice at ST = 1.44, LB = 1.63, BB = 1.73525813861818, LA = 1.6063168414535

Diagnostics

Diagnostics were conducted to make sure that assumptions were satisfied for the model. First, the residuals from our model were plotted:


Residual Plot

We see that the residuals generally fall into an even band between -0.015 and 0.015. Also, there does not seem to be any large trend between the residuals and run number (index). Note that 6 of the last 7 runs have very similar residuals. This is because the last 7 runs used the same levels of predictor

variables (ST = 1.5, LB = 1.5, BB = 1.5, LA = 1.5). Because of this, the independence and homoscedasticity assumptions are satisfied.

Additionally, a QQ-plot for the residuals is shown below:

Normal Q-Q Plot
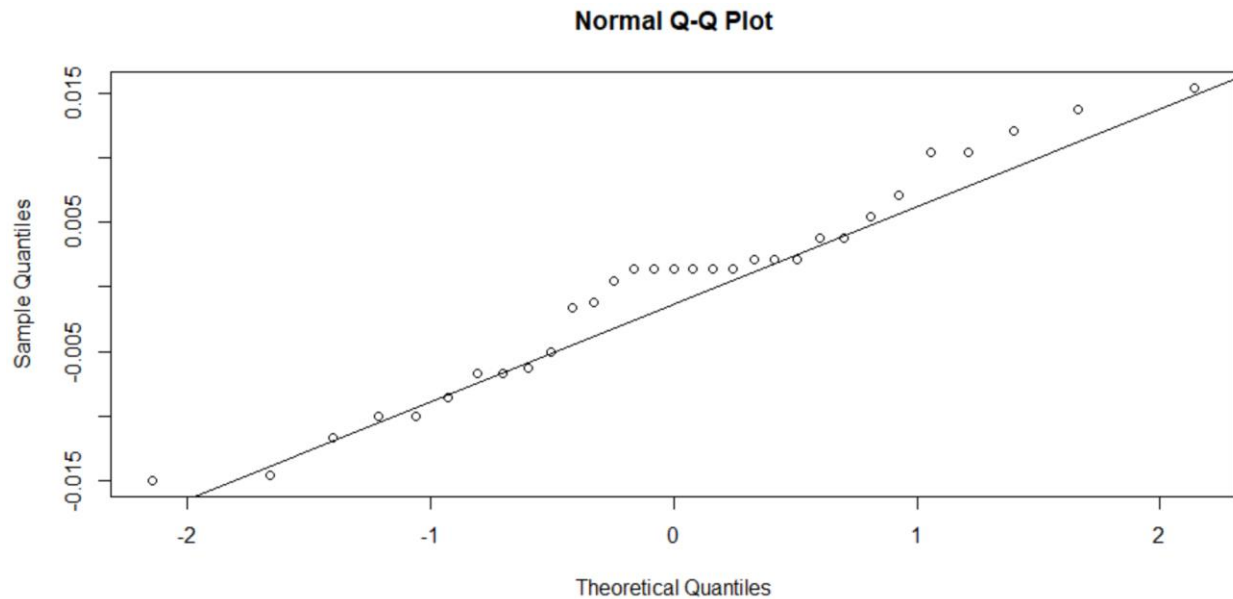
The residuals are very close to following a normal distribution. Even the tails almost follow a normal distribution. There is a slight abnormality towards the center of the distribution due to the multiple repeated tests with the same predictor values. However, this data is close enough to a normal distribution to satisfy the normality condition.

Conclusion

Overall, we have concluded that a model to predict cohesiveness in terms of ST, LB, BB, and LA bacteria is the following:

$$\hat{Y} = 0.092 + 0.169ST + 0.157LB + 0.054BB + 0.082LA - 0.030(ST*LB) + 0.020(LB*BB) \\ - 0.040ST^2 - 0.045LB^2 - 0.025BB^2 - 0.025LA^2$$

Additionally, the value that maximizes cohesiveness is with the following predictor levels:

ST = 1.500, LB = 1.622, BB = 1.714, LA = 1.632

Interestingly, my model is slightly different from the model that was created according to the original paper. In the original paper, their model does not include any first order terms. It only includes quadratic terms, the ST:LB interaction, and the LB:BB interaction.

Further, since the goal of the experiment was to create a MSFPY with similar properties as the control MSFY yogurt, the predictor levels that give the maximized cohesiveness is not the "best" conclusion. According to the study, optimum levels of cohesiveness are between 0.39 and 0.40 while the stationary point from my model gives a cohesiveness of 0.46.

Hence, in order to optimize the five responses so as to produce acceptable ranges for each, the author super-imposed plots for the five responses in relation to the four predictors. This is beyond the scope of my analysis. However, the author states that the optimal bacterial levels are as follows.

ST = 0.79, LB = 0.75, BB = 0.79, LA = 0.95

When inputting these values into my model, we get a cohesiveness of 0.37. This value is very close to the optimal value of 0.39, especially when the predictor levels have been selected with four other responses in mind.

Citation

Kaur, Harbinder, et al. "Textural Properties of Mango Soy Fortified Probiotic Yoghurt: Optimisation of

Inoculum Level of Yoghurt and Probiotic Culture." *International Journal of Food Science & Technology*, vol. 44, no. 2, 2009, pp. 415–424., https://doi.org/10.1111/j.1365-2621.2008.01789.x.