IE 6309- Response Surface Methodology Project

Submitted by: *Poojan Patel, Preetam Kulkarni and Jaivardhan Sood*

Problem Statement

Everybody enjoys a cool lemonade during summer season. However, dissolution of salt in water to make lemonade takes long time if the temperature of solvent is very low (chilled water). On the other hand, it takes some time to cool the lemonade if the temperature of the solvent is high. Hence it is interesting to study the factors that affect the dissolution time of salt in water.

Response variable:

Dissolution time for salt (seconds)

Factor variables:

- Amount of solvent - water (ml)
- Stirring rate (stirs/sec)
- Temperature of solvent – water (Degree Fahrenheit)

Goal of the study:

The goal of this study is to determine the optimum temperature of water, stirring rate and amount of water required to minimize the time taken to dissolve a fixed amount of salt using Response Surface Methodology.

Data Collection Process

Half a tea spoon (2.84 grams) of salt in the form of crystals was dissolved in varying amount of water at different temperatures and at different levels of stirring rates to measure the time it took to dissolve the salt completely. To ensure that the salt is completely dissolved, we used a container made of glass to conduct the experiment. A one third measuring cup was used to measure the quantity of water which was then poured into the glass container and heated in a microwave. A temperature measuring device as shown in the figure below was used to take the temperature readings in degrees Fahrenheit. If the microwave overheated the water, we waited for a while to ensure the water is at the desired temperature. Then the predefined amount of salt (2.84 grams) was measured using a half teaspoon measuring spoon and was dropped into the glass of water. An ice-cream stick was used to stir the salt and water mixture by hand. As soon as the stirring began, a timer on a mobile phone was used to measure the time it took for the salt to disappear in the glass. Poojan Patel practiced stirring at a particular rate and he was the only one who was stirring the water to ensure the consistency in stirring rate.

Table 1. global design range, local design range and initial center for first-order model

	Amount of Water (ml)	Stirring Rate (Stirs/sec)	Temp. of water (Deg. Fahrenheit)
Global Range	150 – 300	2-9	75 – 160
Local Range	157-237	2-4	75-100
Initial Midpoint	197.16	3	87.5

Figures below show the experimental set up and the instruments used to conduct the experiment.

Figure 1. Temperature measuring device

Figure 2. Measuring cup used to measure water

Figure 3. Measuring spoon used to measure amount of salt

Figure 4. Ice-cream stick used to stir

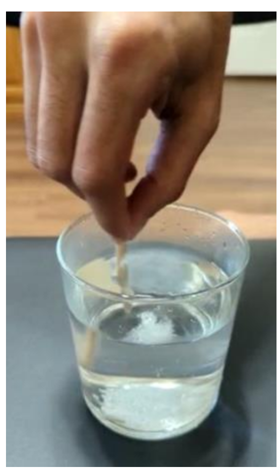

Figure 4. Experiment being conducted

First-Order RSM Iteration

The first-order local design range of the three variables varies as: 157.73 ml (2/3 cup) to 236.59 ml (1 cup) for amount of water, 75 F to 100 F for temperature, and 2 stirs/sec to 4 stirs/sec for stirring rate. The center point for the first order design is at: 197.16 ml for amount of water, 87.5 F for temperature, and 3 stirs/sec for stirring rate. The levels of the variables in coded units are:

$$z_1 = \frac{amount\ of\ water\ in\ ml - 197.16}{39.43} \qquad z_2 = \frac{Temperature\ in\ F - 87.5}{12.5} \qquad z_3 = \frac{stirring\ rate\ in\ sec - 3}{1}$$

The design employed was a 2^3 factorial with four center points (half range) as shown in Table 2 below.

Table 2. Results of first factorial design with four center points

	Variables in original units			**Variable in coded units**			**Response yield (secs)**
run*	**Amount of water (ml)**	**Temperature (F)**	**Stirring rate (stirs/sec)**	**z_1**	**z_2**	**z_3**	**y**
1	236.59	75	2	+1	-1	-1	73.85
2	236.59	75	4	+1	-1	+1	51.98
3	236.59	100	2	+1	+1	-1	60.26
4	236.59	100	4	+1	+1	+1	40.49
5	157.73	75	2	-1	-1	-1	76.87
6	157.73	75	4	-1	-1	+1	58.19
7	157.73	100	2	-1	+1	-1	67.26
8	157.73	100	4	-1	+1	+1	42.12
9	197.16	87.5	3	0	0	0	54.81
10	197.16	87.5	3	0	0	0	54.4
11	197.16	87.5	3	0	0	0	51.54
12	197.16	87.5	3	0	0	0	51.67

***The random order in which the experiment runs were performed was 2,5,9,4,7,3,1,10,6,12,8, and 11**

The outputs of the model fit of the first-order are as follows:

Table 3. SAS output for first-order model (coded form)

Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	3	1274.871100	424.957033	30.87	<.0001
Error	8	110.124967	13.765621		
Corrected Total	11	1384.996067			

R-Square	**Coeff Var**	**Root MSE**	**yield Mean**
0.920487	6.514465	3.710205	56.95333

Source	**DF**	**Type I SS**	**Mean Square**	**F Value**	**Pr > F**
z_1	1	39.8724500	39.8724500	2.90	0.1272
z_2	1	322.0722000	322.0722000	23.40	0.0013
z_3	1	912.9264500	912.9264500	66.32	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
z_1	1	39.8724500	39.8724500	2.90	0.1272
z_2	1	322.0722000	322.0722000	23.40	0.0013
z_3	1	912.9264500	912.9264500	66.32	<.0001

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Intercept	56.95333333	1.07104392	53.18	<.0001
z_1	-2.23250000	1.31175554	-1.70	0.1272
z_2	-6.34500000	1.31175554	-4.84	0.0013
z_3	-10.68250000	1.31175554	-8.14	<.0001

The *fitted RS model* is

$$\hat{y} = \mathbf{56.9533 - 2.2325}\ z_1 - \mathbf{6.345}\ z_2 - \mathbf{10.6825}\ z_3$$

Where,

$\hat{y}$ is the estimated mean dissolve time.

z_1 is the coded value of the amount of water.

z_2 is the coded value of temperature.

z_3 is the coded value of stirring rate.

From Table 3., the p-values for the three factors (predictor), amount of water, temperature, and stirring rate are 0.1272, 0.0013, and <.0001 respectively. This implies that the temperature and stirring rate are significant factors at α = 0.01 while the amount of water is only marginally significant even at α = 0.1.

The parameter estimates help in determining the gradient direction to perform steepest ascent/ descent. The estimates show that all three factors will impact the mean dissolve time negatively. Thus, increasing the amount of water, temperature, and stirring rate will reduce the mean time. The stirring rate is the most influential factor of all followed by the temperature. Increasing the stirring rate and temperature will significantly reduce the mean time for solute to dissolve in the solvent. Next step is to conduct lack-of-fit test to determine if the specified first-order model is sufficient or not.

Generic lack-of-fit test

Generic lack-of-fit test is conducted to assess the significance of the first-order model and verify if the first- order model is sufficient.

Hypothesis:

$H_0: Specified\ model\ is\ sufficient\ (first - order\ model)$

$H_1: Specified\ model\ is\ not\ sufficient$

Decision Rule:

Reject H_0 if ${F_{LOF}}^* > F_{Df_{LOF}, Df_{PE}, \alpha}$

Number of predictor factors, $p = 3$

Number of factorial points, $n_f = 8$

Number of center points, $n_0 = 4$

Number of points, $n = n_f + n_0 = 12$

Number of distinct treatment combinations z, $n_d = 9$

Number of observations at treatment combinations z, for z = 0, $n_z = 4$

Sample Variance, $(S_0)^2 = 3.031$

Sum of Squares for Pure Error, $SSPE = \sum(n_z - 1)(S_0)^2 = (4-1)(3.031) = 9.093$

Degree of Freedom for Pure Error, $Df_{PE} = n - n_d = 3$

Mean Squares for Pure Error, $MSPE = \frac{SSPE}{Df_{PE}} = \frac{9.093}{3} = 3.031$

Sum of Squares for Error, $SSE = 110.125$

Degree of Freedom for Error, $Df_E = 8$

Sum of Squares for lack-of-fit, $SSLOF = SSE - SSPE = 110.125 - 9.093 = 101.032$

Degree of Freedom for lack-of-fit, $Df_{LOF} = n_d - p - 1 = 9 - 3 - 1 = 5$

Mean Squares for lack-of-fit, $MSLOF = \frac{SSLOF}{Df_{LOF}} = \frac{101.032}{5} = 20.2064$

$$F_{LOF}{}^* = \frac{SSLOF}{Df_{LOF}} \Big/ \frac{SSPE}{Df_{PE}} = \frac{101.032}{5} \Big/ \frac{9.093}{3} = \frac{20.2064}{3.031} = 6.6667$$

$F_{Df_{LOF}, Df_{PE}, \alpha}; For\ \alpha = 0.5; F_{5,3,0.05} = 9.0135$

Thus, $F_{LOF}{}^* < F_{5,3,0.05}$

$p - value = 0.07466$

Table 4. ANOVA table for Generic lack-of-fit test for specified first-order model

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
LOF	5	101.032	20.2064	6.6667	0.07466
PE	3	9.093	3.031		
Error	8	110.125			

Thus, as per the test we Fail to Reject (FTR) H_0 and concluded that the specified first-order model is sufficient at $\alpha = 0.05$. This is technically a weak conclusion.

Residual Analysis of the Model

Residual Analysis is conducted to verify model assumptions made for the specified first-order model. By verifying the model assumptions, the validity of the model can be confirmed.

Model assumptions:

1. The residual error terms are normally distributed
2. The error terms have constant variance
3. No outliers present
4. The error terms are uncorrelated

Normality

Test for Normality:

Test for normality is conducted to determine if the error terms are normally distributed.

Hypothesis:

$H_0: Normality\ is\ OK$

$H_1: Normality\ is\ violated$

Decision rule:

Reject H_0 if $\hat{\rho} < c(\alpha, n)$

Table 5: SAS output of $\hat{\rho}$ value – specified first-order model

Pearson Correlation Coefficients, N = 12 Prob > |r| under H0: Rho=0		
	e	**enrm**
e Residuals	1.00000	0.95975 <.0001
enrm Normal Scores	0.95975 <.0001	1.00000

$\hat{\rho} = 0.98039$

$for\ \alpha = 0.1; c(\alpha, n) = 0.942$ (*Table B.6*)

$\hat{\rho} = 0.95975 > c(\alpha, n) = 0.942$

Thus, we FTR H_0 and concluded that the normality is OK. This is technically a weak conclusion.

Normality can be verified by analyzing the Normal Probability plot to check linearity of data points (residual points) against normal scores.

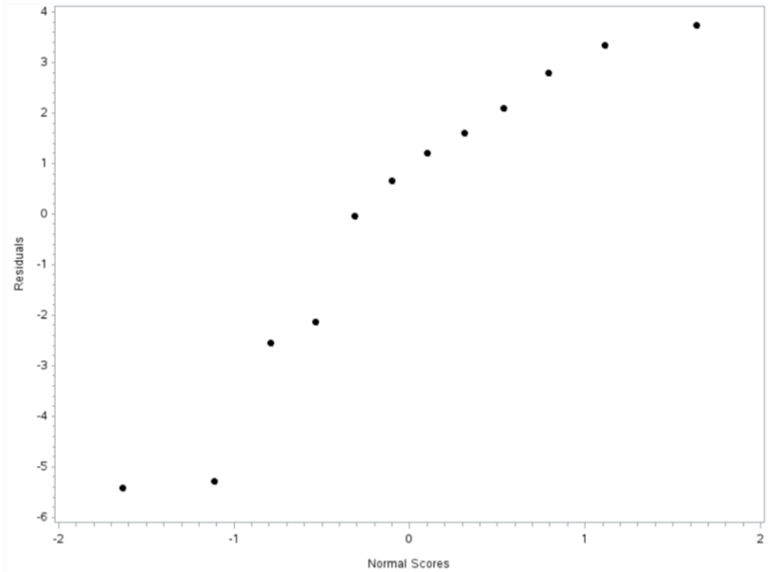

Figure 6. Normal Probability plot of the specified first-order model

It can be observed from Figure 6, that the distribution of points is pretty much linear with a longer lower tail. Thus, it can be concluded that the residual error distribution is close to normality.

Constant Variance

Constant Variance can be verified by analyzing the plot of residuals vs fitted values to check the spread of data points.

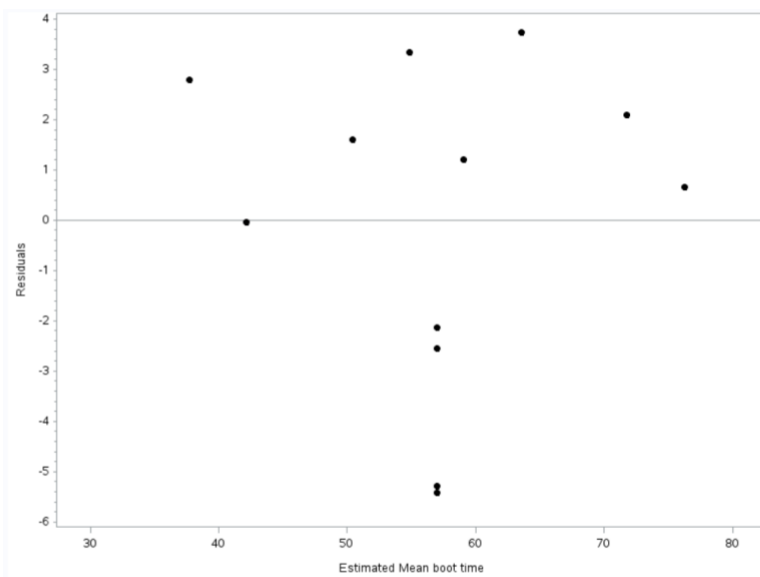

Figure 7. Residuals (e) vs fitted values ($\hat{y}$) plot of the specified first-order model

It can be observed from Figure 7, that there is no presence of a funnel shape. The points are also scattered randomly on both sides of the centerline. Thus, it can be concluded that the constant variance assumption is satisfied.

Outliers

Bonferroni Outlier Test:

Bonferroni outlier test is conducted to determine presence of outliers by employing a t-distribution to compare studentized residuals values (t_i) with most extreme studentized residual.

Decision rule:

If $|t_i| > t\ (1 - \left(\frac{\alpha}{2n}\right); n - p - 1)$ then observation i is an outlier

Table 6: SAS output for specified first-order model

Obs i	Water x_1	Temp x_2	Stirs x_3	Yield Y	z_1	z_2	z_3	Fitted value yhat	Residual e_i	R-student t_i
1	236.59	75	2	73.85	1	-1	-1	71.7483	2.10167	0.74819
2	236.59	75	4	51.98	1	-1	1	50.3833	1.59667	0.55903
3	236.59	100	2	60.26	1	1	-1	59.0583	1.20167	0.41672
4	236.59	100	4	40.49	1	1	1	37.6933	2.79667	1.02778
5	157.73	75	2	76.87	-1	-1	-1	76.2133	0.65667	0.22577
6	157.73	75	4	58.19	-1	-1	1	54.8483	3.34167	1.26973
7	157.73	100	2	67.26	-1	1	-1	63.5233	3.73667	1.46262
8	157.73	100	4	42.12	-1	1	1	42.1583	-0.03833	-0.01313
9	197.16	87.5	3	54.81	0	0	0	56.9533	-2.14333	-0.5777
10	197.16	87.5	3	54.4	0	0	0	56.9533	-2.55333	-0.69519
11	197.16	87.5	3	51.54	0	0	0	56.9533	-5.41333	-1.6921
12	197.16	87.5	3	51.67	0	0	0	56.9533	-5.28333	-1.63566

$$For\ \alpha = 0.1; t\left(1 - \left(\frac{\alpha}{2n}\right); n - p - 1\right) = 3.4788$$

From last column of Table 6, it can be observed that $|t_i| < t\ (1 - \left(\frac{\alpha}{2n}\right); n - p - 1)$ for all observations. Hence it can be concluded that there is no outlier.

Serial Correlation

Serial correlation of error terms can be checked by analyzing the residuals (e) vs observation number plot.

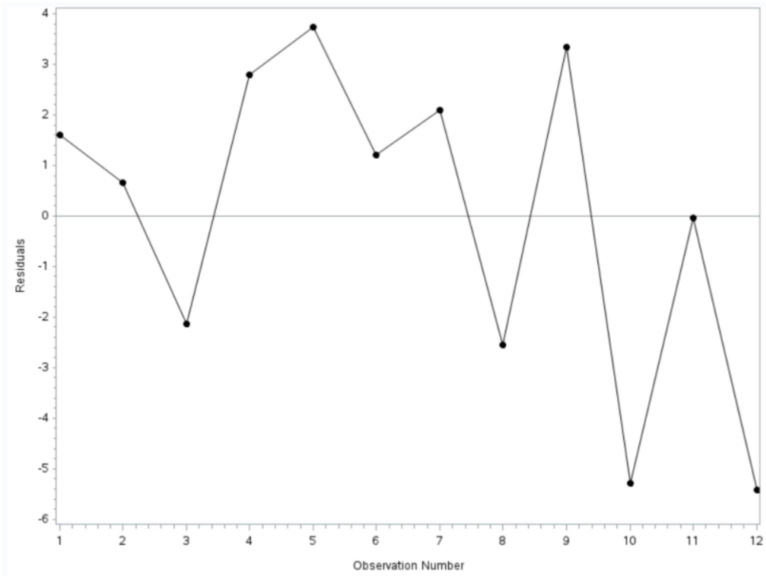

Figure 8. Residuals (e) vs observation number plot for specified first-order model

It can be observed from Figure 8, that the plot has random jaggedness. A slight downward trend can be observed as the experimenter got better at stirring at a constant rate with each run. However, the number of runs is less, and it can be concluded that the errors are not serially correlated (uncorrelated).

Thus, all the model assumptions are satisfied.

Steepest Descent

First-Order analysis indicated that the specified first-order model is sufficient. The next step is to follow the path of steepest descent to reach the maximum of the response surface using gradient direction.

The steepest descent gradient can be obtained from the parameter estimates of the specified first-order model.

The gradient is given by: $\left(\frac{\partial\,\hat{y}}{\partial\,z_1}, \frac{\partial\,\hat{y}}{\partial\,z_2}, \frac{\partial\,\hat{y}}{\partial\,z_3}\right)^T$

For the specified first order model: $\hat{y} = 56.9533 - 2.2325\;z1 - 6.345\;z2 - 10.6825\;z3$

$$Steepest\ descent\ gradient = (2.2325, 6.345, 10.6825)^T$$

Considering step size = 0.1, because we wanted the stirring rate to be an integer as it becomes difficult to gauge stirring rate in fractions. Small step size also means a small jump due to high differences in the significance level of the three variables. Based on the step size the coded gradient direction is given by

$$Gradient\ Direction\ (coded) = (0.22325, 0.6345, 1.06825)^T$$

So, the appropriate direction to move from the center in steps proportion to direction gradient.

The amount of water increases in steps of 8.25 ml, temperature increases by 7.5 F, and the stirring rate increases by 1 to find the maximum response where the mean dissolve time is minimum.

Table 7. Results of points on the path of steepest descent

	Variables in original units			Variable in coded units			Response yield (secs)
run*	Amount of water (ml)	Temperature (F)	Stirring rate (stirs/sec)	z_1	z_2	z_3	y
9, 10, 11, 12	197.16	87.5	3	0	0	0	53.1 (average)
13	205	95	4	0.209	0.594	1	49.01
14	214	102.5	5	0.418	1.188	2	32
15	221	110	6	0.627	1.782	3	24.2
16	238	117.5	7	0.836	2.376	4	24.5

Table 7 represents the data collection for points on the path of steepest descent. The observation that gave the quickest mean dissolve time was at amount of water of 221 ml, temperature of 110 F and stirring rate of 6 stirs/sec. Technically the salt should dissolve faster as temperature and stirring rate increases however it is practically impossible to maintain constant and faster stirring rate. Thus, we see a higher dissolve time value at a faster stirring rate. But at this point the values that are obtained can be considered satisfactory and a new first-order model can be developed at the point of maximum response yield.

Second First-Order RSM Iteration

The second first-order local design range of the three variables varies as: 200 ml (2/3 cup) to 240 ml (1 cup) for amount of water, 100 F to 120 F for temperature, and 5 stirs/sec to 7 stirs/sec for stirring rate. The center point for the first order design is at: 220 ml for amount of water, 110 F for temperature, and 6 stirs/sec for stirring rate. The levels of the variables in coded units are:

$$z_1 = \frac{amount\ of\ water\ in\ ml - 220}{20} \quad z_2 = \frac{Temperature\ in\ F - 110}{10} \quad z_3 = \frac{stirring\ rate\ in\ sec - 6}{1}$$

The second design employed was a 2^3 factorial with four center points as shown in Table 7 below.

Table 8. Results of second factorial design with four center points

	Variables in original units			Variable in coded units			Response yield (secs)
run*	Amount of water (ml)	Temperature (F)	Stirring rate (stirs/sec)	z_1	z_2	z_3	y
17	240	100	2	+1	-1	-1	42.04
18	240	100	4	+1	-1	+1	27.16
19	240	120	2	+1	+1	-1	26.28
20	240	120	4	+1	+1	+1	23.66
21	200	100	2	-1	-1	-1	45.23
22	200	100	4	-1	-1	+1	24.77
23	200	120	2	-1	+1	-1	28.48
24	200	120	4	-1	+1	+1	24.71
25	220	110	6	0	0	0	31.8
26	220	110	6	0	0	0	31.4
27	220	110	6	0	0	0	30.01

28	220	110	6	0	0	0	31.26

***The random order in which the experiment runs were performed was 17,23,20,25,18,21,26,24,27,22,19, and 28**

The outputs of the model fit of the second first-order are as follows:

Table 9. SAS output for the second first-order model (coded form)

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	382.3550375	127.4516792	8.70	0.0067
Error	8	117.2408292	14.6551036		
Corrected Total	11	499.5958667			

R-Square	Coeff Var	Root MSE	yield Mean
0.765329	12.52410	3.828198	30.56667

Source	DF	Type I SS	Mean Square	F Value	Pr > F
z_1	1	2.0503125	2.0503125	0.14	0.7181
z_2	1	162.6306125	162.6306125	11.10	0.0104
z_3	1	217.6741125	217.6741125	14.85	0.0049

Source	DF	Type III SS	Mean Square	F Value	Pr > F
z_1	1	2.0503125	2.0503125	0.14	0.7181
z_2	1	162.6306125	162.6306125	11.10	0.0104
z_3	1	217.6741125	217.6741125	14.85	0.0049

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Intercept	30.56666667	1.10510571	27.66	<.0001
z_1	-0.50625000	1.35347255	-0.37	0.7181
z_2	-4.50875000	1.35347255	-3.33	0.0104
z_3	-5.21625000	1.35347255	-3.85	0.0049

The fitted RS model is

$$\hat{y} = 30.5667 - 0.50625\ z_1 - 4.50875\ z_2 - 5.21625\ z_3$$

Where,

$\hat{y}$ is the estimated mean dissolve time.

z_1 is the coded value of the amount of water.

z_2 is the coded value of temperature.

z_3 is the coded value of stirring rate.

From Table 9, the p-values for the three factors (predictors), amount of water, temperature, and stirring rate are 0.7181, 0.0104, and 0.0049 respectively. This implies that the stirring rate is a significant factor and temperature is a marginally significant factor at α = 0.01, while the amount of water is not significant even at α = 0.1.

The parameter estimates of the amount of water is insignificant while temperature is only marginally significant which indicates that a second order model may be required. However, the next step is to conduct a lack-of-fit test to determine if the specified first-order model is sufficient or not.

Generic lack-of-fit test

Generic lack-of-fit test is conducted to assess the significance of the second first-order model and verify if the model is sufficient.

Hypothesis:

$H_0: Specified\ model\ is\ sufficient\ (second\ first-order\ model)$

$H_1: Specified\ model\ is\ not\ sufficient$

Decision Rule:

Reject H_0 if ${F_{LOF}}^* > F_{Df_{LOF},Df_{PE},\alpha}$

Number of predictor factors, $p = 3$

Number of factorial points, $n_f = 8$

Number of center points, $n_0 = 4$

Number of points, $n = n_f + n_0 = 12$

Number of distinct treatment combinations z, $n_d = 9$

Number of observations at treatment combinations z, for z = 0, $n_z = 4$

Sample Variance, $(S_0)^2 = 0.5975$

Sum of Squares for Pure Error, $SSPE = \sum(n_z - 1)(S_0)^2 = (4-1)(0.5975) = 1.7925$

Degree of Freedom for Pure Error, $Df_{PE} = n - n_d = 3$

Mean Squares for Pure Error, $MSPE = \frac{SSPE}{Df_{PE}} = \frac{1.7925}{3} = 0.5975$

Sum of Squares for Error, $SSE = 117.24085$

Degree of Freedom for Error, $Df_E = 8$

Sum of Squares for lack-of-fit, $SSLOF = SSE - SSPE = 117.24085 - 1.7925 = 115.448$

Degree of Freedom for lack-of-fit, $Df_{LOF} = n_d - p - 1 = 9 - 3 - 1 = 5$

Mean Squares for lack-of-fit, $MSLOF = \frac{SSLOF}{Df_{LOF}} = \frac{115.448}{5} = 23.09$

$$F_{LOF}{}^{*} = \frac{SSLOF}{Df_{LOF}} \Big/ \frac{SSPE}{Df_{PE}} = \frac{101.032}{5} \Big/ \frac{9.093}{3} = \frac{23.09}{0.5975} = 38.6438$$

$F_{Df_{LOF},Df_{PE},\alpha}; For\ \alpha = 0.01; F_{5,3,0.05} = 28.237$

Thus, $F_{LOF}{}^{*} > F_{5,3,0.01}$

$p - value\ =\ 0.006331$

Table 10. ANOVA table for Generic lack-of-fit test for second first-order model

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
LOF	5	115.448	23.09	38.6438	0.006331
PE	3	1.7925	0.5975		
Error	8	117.24085			

Thus, as per the test we Reject H_0 and concluded that the specified second first-order model is not sufficient at $\alpha = 0.01$.

Since the generic lack-of-fit test showed that the first-order model is not sufficient, a second-order lack-of-fit test needs to conducted to check if a second order design would be suitable.

Interactions and quadratic terms are added to the model for two-factor interactions and combined quadratic effects (a standard second-order model does not enable separate estimation of quadratic effects of all factors).

A second order lack of fit test is conducted to check if the second-order model is sufficient.

Second-order lack-of-fit test

Number of predictor factors, $p = 3$

Number of factorial points, $n_f = 8$

Number of center points, $n_0 = 4$

Number of points, $n = n_f + n_0 = 12$

Number of distinct treatment combinations z, $n_d = 9$

Number of observations at treatment combinations z, for z = 0, $n_z = 4$

z_1 and z_2 Interaction estimate, $\gamma_{12} = \frac{1}{n_f} * Z_{12} * y = \frac{1}{8}(-42.04 - 27.16 + 26.28 + 23.66 + 45.23 + 24.77 - 28.48 - 24.71) = -0.30625$

z_1 and z_3 Interaction estimate, $\gamma_{13} = \frac{1}{n_f} * Z_{13} * y = \frac{1}{8}(-42.04 + 27.16 - 26.28 + 23.66 + 45.23 - 24.77 + 28.48 - 24.71) = 0.84125$

z_2 and z_3 Interaction estimate, $\gamma_{23} = \frac{1}{n_f} * Z_{23} * y = \frac{1}{8}(42.04 - 27.16 - 26.28 + 23.66 + 45.23 - 24.77 - 28.48 + 24.71) = 3.61875$

$SSR(z_1 z_2) = n_f(\gamma_{12})^2 = 8(-0.30625)^2 = 0.75$

$SSR(z_1 z_3) = n_f(\gamma_{12})^2 = 8(0.84125)^2 = 5.6616$

$SSR(z_2 z_3) = n_f(\gamma_{12})^2 = 8(3.61875)^2 = 104.763$

Sum of Squares for the two-factor interaction effects, $SSI = sum\ of\ all\ interaction\ SS = SSR(z_1 z_2) + SSR(z_1 z_3) + SSR(z_2 z_3) = 0.75 + 5.6616 + 104.763 = 111.1747$

Degree of Freedom for two-factor interaction effects, $Df_I = 3$ (1 for each term)

Mean Squares for the two-factor interaction effects, $MSI = \frac{SSI}{Df_I} = \frac{111.1747}{3} = 37.058$

$\bar{Y}_{fac} = 30.29125$

$\bar{Y}_{ctr} = 31.1175$

Sum of Squares for the combined quadratic effects, $SSQ = \left(\frac{n_f n_o}{n}\right)\left(\bar{Y}_{fac} - \bar{Y}_{ctr}\right)^2 = \left(\frac{8*4}{12}\right)(30.29125 - 31.1175)^2 = 1.8205$

Degree of Freedom for combined quadratic effects, $Df_Q = 1$

Mean Squares for the combined quadratic effects, $MSQ = \frac{SSQ}{Df_Q} = \frac{1.8205}{1} = 1.8205$

Sum of Squares due to higher-order effects, $SSH = SSLOF - SSI - SSQ = 115.448 - 111.1747 - 1.8205 = 2.4528$

Degree of Freedom for higher-order effects, $Df_H = Df_{LOF} - Df_I - Df_Q = 5 - 3 - 1 = 1$

F-test for Interaction:

$H_0{}^I: all\ estimable\ \gamma_{ij} = 0\ (two - factor\ interaction\ effects\ are\ insignificant)$

$H_1{}^I: not\ all\ \gamma_{ij} = 0(two - factor\ interaction\ effects\ are\ significant)$

Decision Rule:

Reject $H_0{}^I$ if $F_I{}^* > F_{Df_I, Df_{PE}, \alpha}$

$$F_I{}^* = \frac{MSI}{MSPE} = \frac{37.058}{0.5975} = 62.02$$

$F_{Df_I, Df_{PE}, \alpha} = F_{3,3,0.01} = 29.457$

$F_I{}^* > F_{Df_I, Df_{PE}, \alpha}$

$p - value\ =\ 0.00337$

Thus, as per the test we Reject H_0 and concluded that the two-factor interaction effects are significant.

F-test for Quadratic:

$H_0{}^Q: \sum \gamma_{ij} = 0$ $(Cummulative\ curvature\ esitmate\ is\ insignificant)$

$H_1{}^Q: \sum \gamma_{ij} \neq 0$ $(Cummulative\ curvature\ esitmate\ is\ significant)$

Decision Rule:

Reject $H_0{}^Q$ if $F_Q{}^* > F_{Df_Q, Df_{PE}, \alpha}$

$$F_Q{}^* = \frac{MSQ}{MSPE} = \frac{1.8205}{0.5975} = 3.047$$

$F_{Df_Q, Df_{PE}, \alpha} = F_{1,3,0.1} = 5.54$

$F_Q{}^* < F_{Df_Q, Df_{PE}, \alpha}$

$p - value\ =\ 0.17922$

Thus, as per the test we FTR H_0 and concluded that the quadratic effects are not significant.

Table 11. ANOVA table for second-order lack-of-fit test

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Interaction	3	111.1747	37.058	62.02	0.00337
Quadratic	1	1.8205	1.8205	3.047	0.17922
Higher-Order	1	2.4528			
PE	3	1.7925	0.5975		
Error	8				

Since the second-order lack-of-fit test indicated the presence of significant second order interactions, design an experiment to fit the second-order model.

Second-Order Design

Since the first-order model was not sufficient a second-order model design as per the new local design range is required. The purpose of second-order model design is to ensure efficient RS estimation, efficient parameter estimation, and allow a test for model's lack-of-fit.

A second-order Center-Composite Design (CCD) can be generated by adding data points to the first-order design.

The conditions to make the CCD rotatable and orthogonal are:

1) $\alpha = (n_f)^{1/4}\ ; where\ \alpha\ should\ be\ an\ integer$

2) $n_0 = 4\sqrt{n_f} + 4 - 2p\,; where\ n_0\ should\ be\ an\ integer$

Where,

$n_f = factorial\ points$

$n_0 = center\ points$

$\alpha = axial\ points$

$p = Number\ of\ factors$

Considering $n_f = 16$

$\alpha = (n_f)^{1/4} = (16)^{1/4} = 2\ in\ each\ direction\ on\ each\ axis\ (for\ each\ factor)$

$n_0 = 4\sqrt{n_f} + 4 - 2p = 4\sqrt{16} + 4 - 2 * 3 = 14$

The CCD design employed was a 2*2^3 complete factorial design with Resolution V, 14 center points, and 2 axial points as shown in Table 12 below.

Table 12. Results of second-order CCD data points

		Variables in original units			**Variable in coded units**			**Response yield (secs)**
	run*	**Amount of water (ml)**	**Temperature (F)**	**Stirring rate (stirs/sec)**	**z_1**	**z_2**	**z_3**	**Y**
Second first-order design	17	240	100	5	1	-1	-1	42.04
	18	240	100	7	1	-1	1	27.16
	19	240	120	5	1	1	-1	26.28
	20	240	120	7	1	1	1	23.66
	21	200	100	5	-1	-1	-1	45.23
	22	200	100	7	-1	-1	1	24.77
	23	200	120	5	-1	1	-1	28.48
	24	200	120	7	-1	1	1	24.71
	25	220	110	6	0	0	0	31.8
	26	220	110	6	0	0	0	31.4
	27	220	110	6	0	0	0	30.01
	28	220	110	6	0	0	0	31.26
Additional runs to form Central-Composite Design	29	240	100	5	1	-1	-1	39.06
	30	240	100	7	1	-1	1	28.11
	31	240	120	5	1	1	-1	26.73
	32	240	120	7	1	1	1	24.55
	33	200	100	5	-1	-1	-1	43.77
	34	200	100	7	-1	-1	1	23.81
	35	200	120	5	-1	1	-1	28.92
	36	200	120	7	-1	1	1	26.91
	37	180	110	6	-2	0	0	35.25
	38	260	110	6	2	0	0	28.11

	39	220	90	6	0	-2	0	38.71
	40	220	130	6	0	2	0	21.52
	41	220	110	4	0	0	-2	41.04
	42	220	110	8	0	0	2	21.25
	43	220	110	6	0	0	0	33.25
	44	220	110	6	0	0	0	28.51
	45	220	110	6	0	0	0	34.33
	46	220	110	6	0	0	0	31.72
	47	220	110	6	0	0	0	29.92
	48	220	110	6	0	0	0	28.89
	49	220	110	6	0	0	0	30.87
	50	220	110	6	0	0	0	30.14
	51	220	110	6	0	0	0	32.55
	52	220	110	6	0	0	0	32.37

***The random order in which the additional runs for CCD performed was 51, 41, 46, 32, 52, 40, 47, 39, 31, 49, 35, 33, 44, 38, 42, 30, 48, 43, 34, 37, 29, 45, 50, and 36**

Second-Order Analysis

The results of the second-order analysis are as shown below:

Table 13. SAS output of RSREG procedure for the second-order CCD model

Coding Coefficients for the Independent Variables		
Factor	**Subtracted off**	**Divided by**
Water	220.000000	40.000000
Temp	110.000000	20.000000
Stirs	6.000000	2.000000

Response Surface for Variable yield	
Response Mean	30.752500
Root MSE	1.618329
R-Square	0.9465
Coefficient of Variation	5.2624

Regression	DF	Type I Sum of Squares	R-Square	F Value	Pr > F
Linear	3	988.140012	0.7758	125.77	<.0001
Quadratic	3	5.939734	0.0047	0.76	0.5289
Cross product	3	211.562619	0.1661	26.93	<.0001
Total Model	9	1205.642366	0.9465	51.15	<.0001

Residual	DF	Sum of Squares	Mean Square	F Value	Pr > F
Lack of Fit	5	23.941017	4.788203	2.28	0.0840
Pure Error	21	44.152693	2.102509		
Total Error	26	68.093709	2.618989		

Parameter	DF	Estimate	Standard Error	t Value	Pr > |t|	Parameter Estimate from Coded Data
Intercept	1	304.000625	79.532177	3.82	0.0007	31.129375
Water	1	-0.099146	0.404414	-0.25	0.8083	-1.940833
Temp	1	-1.152958	0.808827	-1.43	0.1659	-8.174167
Stirs	1	-52.061667	7.176822	-7.25	<.0001	-9.700833
water*water	1	-0.0000336	0.000715	-0.05	0.9629	-0.053750
temp*water	1	-0.002059	0.002023	-1.02	0.3180	-1.647500
temp*temp	1	-0.004047	0.002861	-1.41	0.1691	-1.618750
stirs*water	1	0.048656	0.020229	2.41	0.0236	3.892500
stirs*temp	1	0.347938	0.040458	8.60	<.0001	13.917500
stirs*stirs	1	-0.147188	0.286083	-0.51	0.6113	-0.588750

Factor	DF	Sum of Squares	Mean Square	F Value	Pr > F
water	4	40.472595	10.118149	3.86	0.0136
temp	4	602.553770	150.638442	57.52	<.0001
stirs	4	774.178620	193.544655	73.90	<.0001

The fitted second order model is

$$\boldsymbol{\hat{y} = 304 - 0.0991\, z_1 - 1.153\, z_2 - 52.0616\, z_3 - 0.0000336\, {z_1}^2 - 0.00206\, z_1 z_2 - 0.00405\, {z_2}^2 + 0.0486\, z_1 z_3 + 0.348\, z_2 z_3 - 0.1472\, {z_3}^2}$$

Where,

$\hat{y}$ is the estimated mean dissolve time.

z_1 is the coded value of the amount of water.

z_2 is the coded value of temperature.

z_3 is the coded value of stirring rate.

$z_1 z_2, z_1 z_3, z_2 z_3$ are the interaction terms

${z_1}^2, {z_2}^2, {z_3}^2$ are the quadratic terms

Table 14. ANOVA table for second-order model

Residual	DF	Sum of Squares	Mean Square	F Value	Pr > F
Linear	3	988.140012	329.38	125.77	<.0001
Quadratic	3	5.939734	1.979911	0.76	0.5289
Interaction	3	211.562619	70.52087	26.93	<.0001
Total Error	26	68.093709	2.618989		
Total	35	1273.736074			

Second-order generic lack-of-fit test

$H_0: Specified\ model\ is\ sufficient\ (second\ -order\ model)$

$H_1: Specified\ model\ is\ not\ sufficient$

Table 15. ANOVA table for second-order generic lack-of-fit test

Residual	DF	Sum of Squares	Mean Square	F Value	Pr > F
Lack of Fit	5	23.941017	4.788203	2.28	0.0840
Pure Error	21	44.152693	2.102509		
Total Error	26	68.093709	2.618989		

The F value is 2.28 which is less than $F_{(5,21,0.05)}$ = 2.6847. Hence, we fail to reject H_0.

The second order model is sufficient. From the ANOVA the linear and the interaction terms are the only ones that are significant at an α of 0.05. However, the quadratic terms are insignificant. We now conduct Canonical analysis to find out if it is possible to find a minimum.

Canonical Analysis

As shown in Table 16, the uncoded stationary point is (235.77, 123.06, 7.6) i.e., when one teaspoon of salt is added to 235.77 ml of water which is at a temperature of 123.06 degrees Fahrenheit and the stirring rate is maintained at 7.6, the predicted mean time to dissolve the salt is 24.26 seconds.

Table 16. Canonical Analysis of Response Surface Based on Coded Data

Factor	Critical Value	
	Coded	Uncoded
water	0.394432	235.777291
temp	0.653205	123.064108
stirs	0.785971	7.571941
Predicted value at stationary point: **24.264621**		

Eigenvalues	Eigenvectors		
	water	Temp	stirs
6.003342	0.147733	0.658087	0.738307
0.258566	0.963501	-0.264269	0.042762
-8.523157	0.223252	0.705042	-0.673108
Stationary point is a saddle point.			

Because one of the eigen values is less than zero and the other two are greater than zero, the stationary point is a saddle point. To minimize the time taken by the salt to dissolve, the negative eigen values should be maximized in the global range which is 9 stirs per second. Hence, when a teaspoon of salt is dissolved in 235.77 ml of water which is at a temperature of 123.06 degrees Fahrenheit and the stirring rate is maintained at 9 stirs per second, the minimum mean time to dissolve can be observed.

Final Discussion

In this study, the intent was to identify the combination of factors that will result in the least amount of time taken to dissolve a fixed amount of salt in water. The Factors for this project were the volume of water, temperature of the water and the stirring rate per second. We first identified a global range of values and tried to identify a first order model that would help explain the variation in the time taken to dissolve the salt in water. Then we conducted a generic lack of fit test on this first order model and found out that it is sufficient. Hence, we are still in a region where the response is linear.

Model assumptions were verified at this point to ensure that they are satisfied. Now, a gradient was generated to follow the path of steepest descent and move to another point where the response surface could be nonlinear. A point was selected and again a first order model created. This time the generic lack of fit test indicated that the model was not sufficient.

At this point, we used Center-Composite-Design to generate a second order model and conducted a lack of fit test. The test indicated that the second order model is sufficient indicating that there are significant interactions.

With canonical analysis, we identified the existence of a saddle point. Hence, further analysis is required to find the factor value combination that will lead to the minimum mean time to dissolve salt in water at a given temperature and stirring rate. We expect to see a minimum time to dissolve salt when the factor variable values were 235.77 ml of water, temperature of 123.06 degrees Fahrenheit and a stirring rate of 9 stirs per second.

Future Work

During the experiment, stirring rate was hard to control as we were not using any equipment to do this. For future analysis, using stirring equipment which is controlled by a motor will help improve accuracy of the experiment.