JMASM algorithms and code simple response surface methodology using RSREG (SAS)

Article · January 2016

6 authors, including:

Mohamad Shafiq
Universiti Sains Malaysia

Kasypi Mokhtar
Universiti Malaysia Terengganu

Hanafi A Rahim
Universiti Malaysia Terengganu

DigitalCommons@WayneState

Journal of Modern Applied Statistical Methods

Volume 15 | Issue 1 Article 44

5-1-2016

Simple Response Surface Methodology Using RSREG (SAS)

Wan Muhamad Amir
University Science Malaysia, wmamir@usm.my

Mohamad Shafiq
University Science Malaysia, shafiqmat786@gmail.com

Kasypi Mokhtar
University Malaysia Terengganu, kasypi@umt.edu.my

Nor Azlida Aleng
University Malaysia Terengganu, azlida_aleng@umt.edu.my

Hanafi A.Rahim
University Malaysia Terengganu, hanafi@umt.edu.my

See next page for additional authors

Recommended Citation

Amir, Wan Muhamad; Shafiq, Mohamad; Mokhtar, Kasypi; Aleng, Nor Azlida; A.Rahim, Hanafi; and Ali, Zalila (2016) "Simple Response Surface Methodology Using RSREG (SAS)," *Journal of Modern Applied Statistical Methods*: Vol. 15: Iss. 1, Article 44.
Available at: http://digitalcommons.wayne.edu/jmasm/vol15/iss1/44

Simple Response Surface Methodology Using RSREG (SAS)

Authors
Wan Muhamad Amir, Mohamad Shafiq, Kasypi Mokhtar, Nor Azlida Aleng, Hanafi A.Rahim, and Zalila Ali

Journal of Modern Applied Statistical Methods
May 2016, Vol. 15, No. 1, 855-867.

ISSN 1538 − 9472

JMASM Algorithms and Code

Simple Response Surface Methodology Using RSREG (SAS)

Wan Muhammad Amir
University of Science, Malaysia
Kelantan, Malaysia

Kasypi Mokhtar
University of Malaysia
Terengganu, Malaysia

Hanafi A. Rahim
University of Malaysia
Terengganu, Malaysia

Mohamad Shafiq
University of Science, Malaysia
Kelantan, Malaysia

Nor Azlida Aleng
University of Malaysia
Terengganu, Malaysia

Zalila Ali
University of Science, Malaysia
Penang, Malaysia

Response surface methodology (RSM) can be used when the response variable, *y*, is influenced by several variables, *x*'s. When treatments take the form of quantitative values, then the true relationship between response variables and independent variables might be known. Examples are given in SAS.

Keywords: Multiple linear regression, response surface methodology

Introduction

Response surface methodology (RSM) was introduced by Box and Wilson (1951). The RSM explores the relationships between several explanatory variables (*X*) and one or more response variables (*Y*). The main idea of RSM is to use a sequence of designed experiments to obtain an optimal response through linear models and second-degree polynomials. This model is only an approximation, but it is easy to apply even when little is known about the process. According to Montgomery (2005), RSM is a statistical technique that useful for modelling and analysis of problems in which a response of interest is influenced by some variables and the objective is to optimize the response variable, *Y*. The general form of RSM can be expressed as $y = f(x_1, x_2, x_3, \ldots, x_n) + \varepsilon$ or $y = \mathrm{f}'(\mathbf{x})\boldsymbol{\beta} + e$ where $\mathbf{x} = (x_1, x_2, \ldots, x_k)'$, $\mathrm{f}(\mathbf{x})$ is a vector function of p elements that consists of powers and cross-products of powers of $x_1, x_2, \ldots, x_k$ up to a certain degree, denoted by d (≥ 1), $\boldsymbol{\beta}$ is a vector of p unknown coefficients referred to as parameters, and e is an experimental error term.

Dr. Amir bin W Ahmad is an Associate Professor of Biostatistics. Email him at: wmamir@usm.my. Mohamad Shafiq Bin Mohd Ibrahim is a postgraduate student in the School of Dental Sciences. Email him at: shafiqmat786@gmail.com.

Two important models are commonly used in RSM. These are the first-degree model (d = 1),

$$y = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \varepsilon ,$$

and the second-degree model (d = 2),

$$y = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} \beta_{ij} x_i x_j + \sum_{i=1}^{k} \beta_{ii} x_i^2 + \varepsilon .$$

The relationship between **y** and $x_1, x_2, \ldots, x_k$ can be used to predict response values for given settings of the control variables. Also, the significance of the factor whose levels are presented by $x_1, x_2, \ldots, x_k$ can be determined and the optimum setting of $x_1, x_2, \ldots, x_k$ over a certain region of interest can be identified. RSMs are designs and models for searching for the optimum response through linear and second-degree polynomial models. When there is more than one response, then it is important to find the compromise optimum that does not optimize only one response (Myers, Khuri, & Carter; Oehlert, 2000; Ngo, 2012). In general, the response surface can be visualized graphically; this graph is very helpful in order to see the shape of a response surface.

As an example, consider the function of y (Triglycerides) plotted against the levels of x_1 (BMI) and x_2 (F7) as shown in Figure 1 (Ahmad, Shafiq, Halim, & Aleng, 2014).

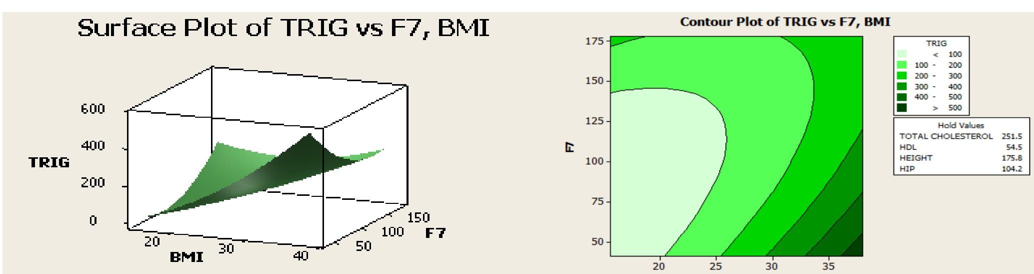

Figure 1. Surface plot of Trig vs F7, BMI, left; Contour plot of Trig vs F7, BMI, right

Materials and Methods

The relationship between response variable and independent variable is quite difficult to determine. The model parameters can be estimated most effectively if proper experimental design is used to collect the data. The relationship between response and independent variable is determined by a mathematical model called a regression model. There are two models involved, the first-order model and second-order model.

First-order model (method of least square)

In general, a first-order model takes the form

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

with two independent variables and, with n independent variables,

$$y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_q x_q + \varepsilon_i, \quad i = 1, 2, \ldots, n$$
$$y = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \varepsilon$$

where $(\beta_0, \beta_1, \beta_2, \ldots, \beta_q)$ are regression coefficients, $(x_0, x_1, x_2, \ldots, x_q)$ are independent or predictor variables, ε is random error, and y is a dependent, or response, variable.

Second-order model (method of least square)

With two independent variables:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{12} x_1 x_2 + \varepsilon$$

In general, a second-order model or second-degree polynomial (with n independent variables) expression takes the following form:

$$y = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} \beta_{ij} x_i x_j + \sum_{i=1}^{k} \beta_{ii} x_i^2 + \varepsilon \; .$$

The second degree polynomial is flexible because it can take a variety of function forms. The β_0, β_i, and β_{ij} are constant and ε is a term of error or residual between the observed and calculated value.

Case study I: First-order design

Suppose *A* and *B* denote the natural variables temperature (ζ_1), pressure ratio (ζ_2), and genuineness (*Y*). Then the transformation of these natural variables to coded variables is

$$C = \frac{X_1 + 220}{5}, \quad D = \frac{X_2 - 1.2}{0.1} \ .$$

The calculation of the coded variables is shown in Table 2.

The relationship between the response variable *y* and independent variable **x** is usually unknown. In general, the lower order polynomial model is used to describe the response surface, f. Results from Table 3 indicate that the linear models (p = 0.015) were statistically significant, suggesting that this model adequately fits the data.

The first stage of the model was fitted to the data by the method of least squares. The regression model is given by

$$Y = 84.1 + 0.850C + 0.250D \ .$$

Table 1. Process data for fitting, first-order model (natural variables) (Montgomery, 1984)

Temperature (ζ_1) X_1	Pressure Ratio (ζ_2) X_2	Genuineness (Y)
-225	1.1	82.8
-225	1.3	83.5
-215	1.1	84.7
-215	1.3	85.0
-220	1.2	84.1
-220	1.2	84.5
-220	1.2	83.9
-220	1.2	84.3

Table 2. Process data for fitting, first-order model (coded variables)

Temperature (ζ_1) *C*	**Pressure Ration (ζ_2) *D***	**Genuineness (*Y*)**
-1	-1	82.8
-1	1	83.5
1	-1	84.7
1	1	85.0
0	0	84.1
0	0	84.5
0	0	83.9
0	0	84.3

Table 3. Analysis of variance for genuineness (*Y*)

Source	**DF**	**Seq SS**	**Adj SS**	**Adj MS**	**F**	**P**
Regression	4	3.26	3.26	0.81500	12.23	0.034
Linear	2	3.14	3.14	1.57000	23.55	0.015
Square	1	0.08	0.08	0.08000	1.20	0.353
Interaction	1	0.04	0.04	0.04000	0.60	0.495
Residual Error	3	0.20	0.20	0.06667		
Pure Error	3	0.20	0.20	0.06667		
Total	7	3.46				

Table 4. Estimated regression coefficients for *Y*

Predictor	**Coef**	**SE Coef**	**T**	**P**
Constant	84.10	0.0894	940.27	0.000
C	0.85	0.1265	6.72	0.001
D	0.25	0.1265	1.98	0.105

Table 5. Analysis of variance (ANOVA)

Source	**DF**	**SS**	**MS**	**F**	**P**
Regression	2	3.14	1.570	24.53	0.003
Residual Error	5	0.32	0.064		
Total	7	3.46			

The residual plots do not indicate any problem with the model. A normal distribution with a mean of 1.246×10^{-14} and a standard deviation of 0.1690 appears to fit our sample data fairly well. The plotted points form a reasonably straight line. In our case, the residuals bounce randomly around the 0 line (residual vs. predicted value). This suggests that the assumption that the relationship is linear is reasonable. A higher *R*-squared value of 1 indicates how well the data fits the model and also indicates a better model.

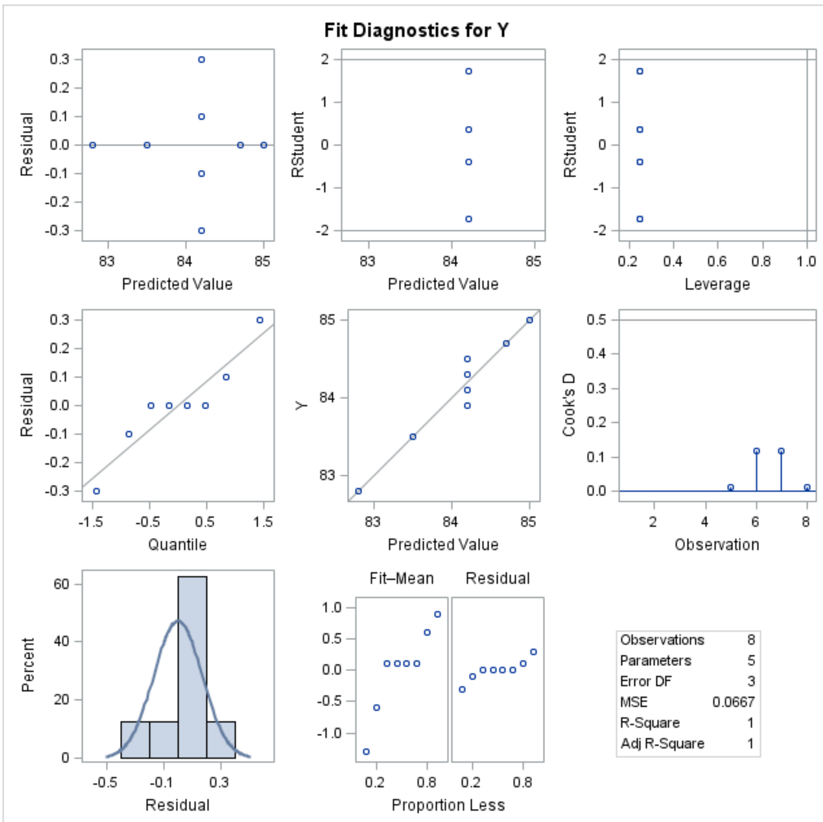

Figure 2. Residual plot for *Y*

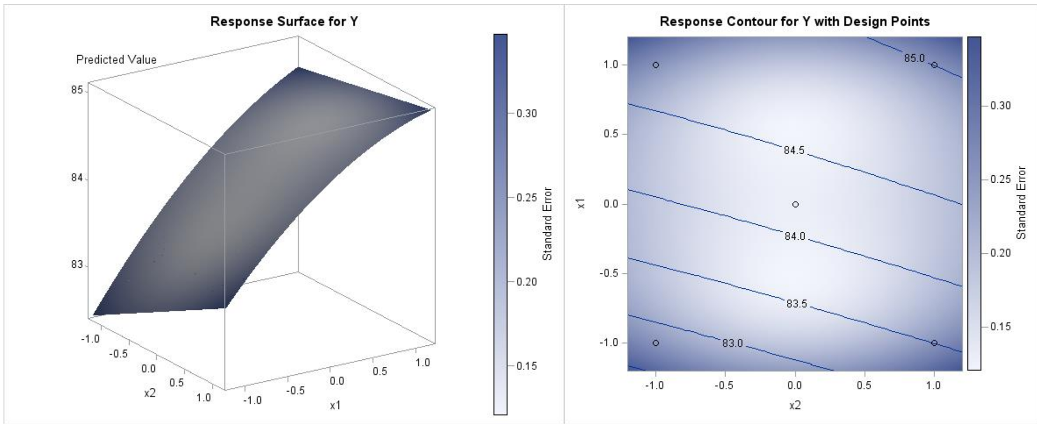

Figure 3. Contour and surface plot for genuineness Y vs. temperature X_1 and pressure ratio X_2.

The counter and surface plots indicate that the highest value of Genuineness is obtained when temperature is high and pressure ratio level is high. This area appears at the upper right corner of the plot.

Performing response surface analysis using the SAS RSREG procedure

```
Data predict ;
input x1 x2 Y;
cards;
-225   1.1   82.8
-225   1.3   83.5
-215   1.1   84.7
-220   1.2   84.1
-220   1.2   84.5
-220   1.2   83.9
-220   1.2   84.3
;
/*plots=(surface)*/
ods graphics on;
proc rsreg data=predict plots=(surface);
model y=x1 x2/lackfit;
run;
```

```
ods graphics off;

/*surface(3D)*/
ods graphics on;
proc rsreg data=predict plots=surface(3D);
model y=x1 x2/lackfit;
run;
ods graphics off;

/*Plot all*/
ods graphics on;
proc rsreg data=predict plots=all;
model y=x1 x2/lackfit;
run;
ods graphics off;
```

Case study II: Second-order design

Let A and B denote the natural variables reading (ζ_1) X_1, reading (ζ_2) X_2, and response (Y). Then the transformation of these natural variables to coded variables is

$$C = \frac{X_1 - 95}{5}, \quad D = \frac{X_2 - 165}{5}$$

The calculation of the coded variables is shown in Table 7.

The relationship between the response variable Y and independent X_1 and X_2 is usually unknown. In general, the lower order polynomial models are used to describe the response surface, f. Because the linear model suggested that a higher model is needed to adequately model the response surface, the full quadratic model is fitted. For the full quadratic model (Table 8), the p-value for lack of fit is 0.089, suggesting that this model adequately fits the data.

The second stage of the model was also fitted to the data by the method of least squares. We get the following model in coded variables by using the regression method. The fitted regression model is given by

$$Y = 81.00 + C + 0.5D - 1.5C^2$$

Table 6. Process data for fitting, second-order model (natural variables) (Montgomery, 1984)

Reading (ζ_1) X_1	Reading (ζ_2) X_2	Response (Y)
90	160	78
90	170	79
100	160	80
100	170	81
95	165	81
95	165	82
95	165	82
95	165	80
95	165	80

Table 7. Process data for fitting, second-order model (coded variables)

Reading (ζ_1) C	Reading (ζ_2) D	Response (Y)
-1	-1	78
-1	1	79
1	-1	80
1	1	81
0	0	81
0	0	82
0	0	82
0	0	80
0	0	80

Table 8. Analysis of variance for genuineness (Y)

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Regression	4	10	10	2.5	2.5	0.198
Linear	2	5	5	2.5	2.5	0.198
Square	1	5	5	5.0	5.0	0.089
Interaction	1	0	0	0.0	0.0	1.000
Residual Error	4	4	4	1.0		
Pure Error	4	4	4	1.0		
Total	8	14				

Table 9. Estimated regression coefficients for *Y*

Predictor	Coef	SE Coef	T	P
Constant	81.0	0.4472	181.122	0.000
C	1.0	0.5000	2.000	0.116
D	0.5	0.5000	1.000	0.374
C^2	-1.5	0.6708	-2.236	0.089

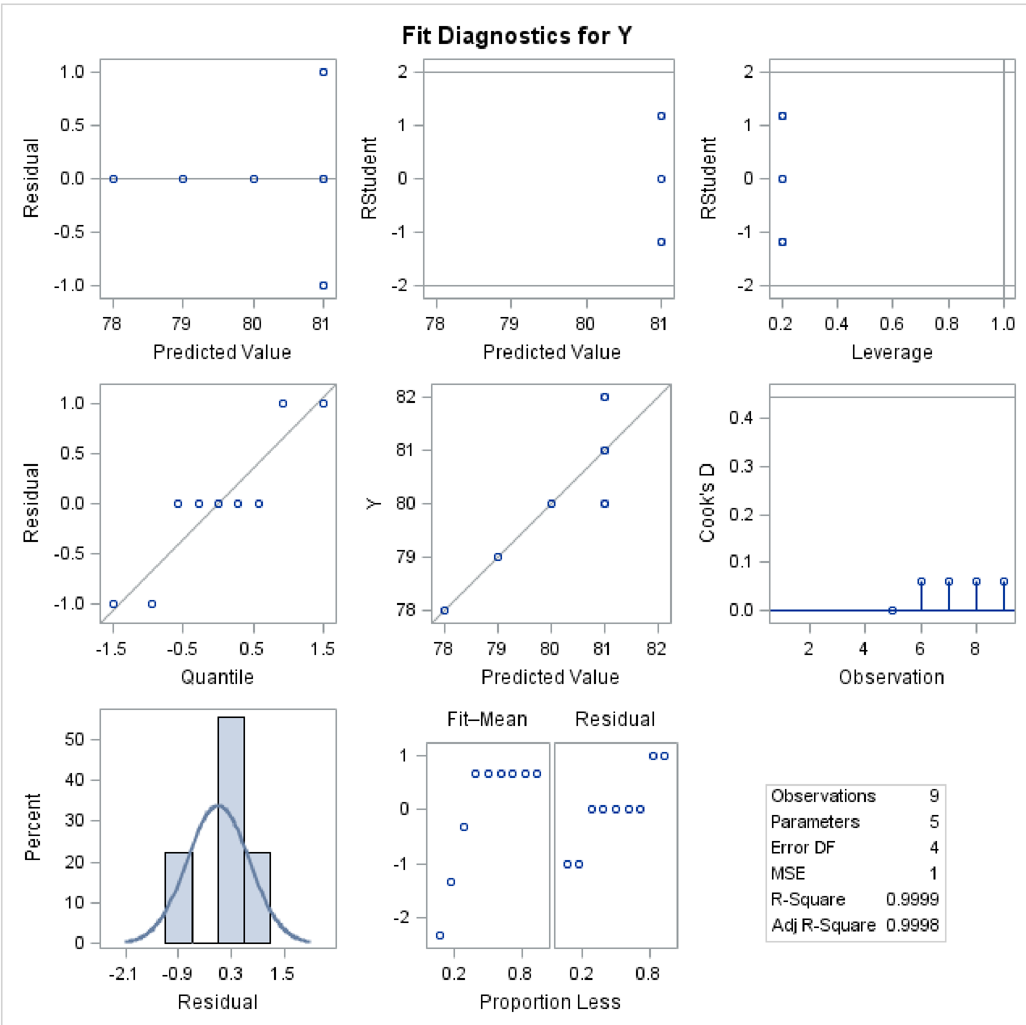

Figure 4. Residual plot for *Y*

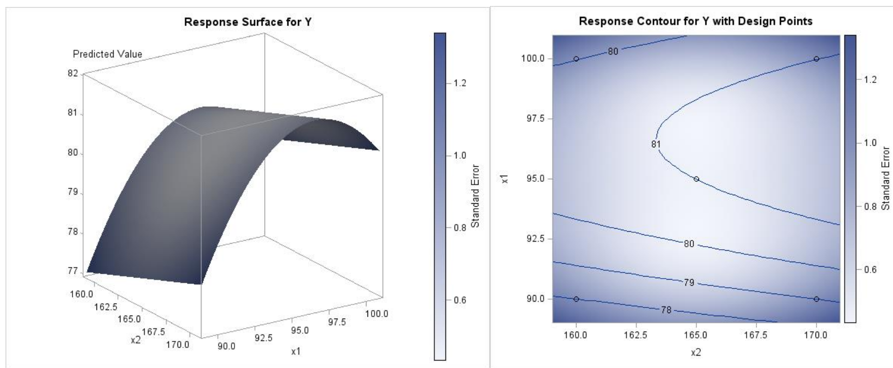

Figure 5. Contour and surface plot for response Y vs. reading (X_1) and reading (X_2)

The residual plots do not indicate any problem with the model. In our case, the residuals bounce randomly around the 0 line (residual vs predicted value). This suggests that the assumption that the relationship is linear is reasonable. How well the estimated model fits the data can be measured by the value of R^2. The R^2 lies in the interval [0, 1]. A higher R-squared value of 0.99 indicated how well the data fit the model and also indicates a better model.

The counter and surface plots indicate that the highest value of response Y is obtained when the reading X_2 is high and the reading X_1 is in the range of 92 to 98. This area appears at the right corner of the plot. In addition, we can see the shape of the of the response surface and get a general idea of the response Y at various settings of reading X_1 and reading X_2.

Response surface methodology are design and models for working with continuous treatments when finding the optima or describing the response is the goal (Oehlert, 2000).

Performing response surface analysis using the SAS RSREG procedure

```
Data predict ;
input x1 x2 Y;
cards;
90      160    78
90      170    79
```

```
100     160     80
100     170     81
95      165     81
95      165     82
95      165     80
95      165     80
;
/*plots=(surface)*/
ods graphics on;
proc rsreg data=predict plots=(surface);
model y=x1 x2/lackfit;
run;
ods graphics off;

/*surface(3D)*/
ods graphics on;
proc rsreg data=predict plots=surface(3D);
model y=x1 x2/lackfit;
run;
ods graphics off;

/*Plot all*/
ods graphics on;
proc rsreg data=predict plots=all;
model y=x1 x2/lackfit;
run;
ods graphics off;
```

Summary and Conclusion

Factorial designs with the RSMs provided a preliminary idea of the dependent variables with two independent variables by plotting a contour and surface response function. The factorial designs are widely used in experiments when the curvature in the response surface is concerned. Actually, it is easier to understand the behavior of the data by using graphing and canonical analysis. This response surface method reveals the finding with more explicitly due to the surface plot performance. It provides the comprehensive information and also give the general idea of dependent variables at the various setting of two independent variables. After a proper design is conducted, the response surface analysis can be made by any statistical computer software as such SAS and then statistical analyses can be applied to draw the appropriate conclusions from the study.

References

Ahmad, W. M. A. W., Shafiq, M., Halim, N., & Aleng, N. A. (2014). A study of triglycerides level in three distinct phases of human blood pressure: A case study from previous projects. *Applied Mathematical Sciences, (8)*46, 2289-2305. doi: 10.12988/ams.2014.42145

Box, G. E. P. & Wilson, K. B. (1951). On the experimental attainment of optimum conditions. *Journal of the Royal Statistical Society*. *Series B (methodological) 13*(1), 1-45. Available from http://www.jstor.org/stable/2983966

Montgomery, D. C. (1984). *Design and analysis of experiments* (2nd ed). New York: Wiley.

Montgomery, D. C. (2005). *Design and analysis of experiments: Response surface method and designs* (6th ed.). Hoboken, NJ: John Wiley & Sons, Inc.

Myers, R. H., Khuri, A. I., & Carter, W. H., Jr. (1989). Response surface methodology: 1966-1988. *Technometrics 31*(2), 137-153. doi: 10.1080/00401706.1989.10488509

Ngo, T. H. D. (2012). The steps to follow in a multiple regression analysis. *Proceedings of the SAS Global Forum 2012 Conference (paper 333-2012)*. Cary, NC: SAS Institute Inc. Retrieved from http://support.sas.com/resources/papers/proceedings12/333-2012.pdf

Oehlert, G. W. (2000). *A first course in design and analysis of experiments*. New York, NY: W. H. Freeman.