International Journal of Mathematics and Statistics Invention (IJMSI)
E-ISSN: 2321 – 4767 P-ISSN: 2321 - 4759
www.ijmsi.org Volume 3 Issue 7 || November. 2015 || PP-09-11

Reduction of Response Surface Design Model: Nested Approach

T. Deepthi and N.Ch. Bhatra Charyulu
Department of Statistics, University College of Science, Osmania University, Hyderbad-7.

ABSTRACT : *If the number of factors is more, only a few factors are important and underlying phenomena is of interest to study, moreover it is possible to eliminate the insignificant factors from the model, which are not affecting much the response, resulting to reduction of size of the model. It also reduces time, cost, effort and complexity of analysis of the model. Very few authors made attempts on reduction of response surface design. In this paper, an attempt is made to find the best choice model by selecting the factors by reducing the size of the first order response surface design model in nested approach and is illustrated with suitable examples.*

Keywords: *Nested approach, Model Reduction, First Order Response Surface Design Model.*

I. INTRODUCTION

Let Y be the vector of response corresponding a design matrix $X = ((x_{u1}, x_{u2}, \ldots, x_{uv}))$, where x_{ui} be the level of the i^{th} factor in the u^{th} treatment combination. Assume the functional form of the response surface design model can be expressed as

$$\underline{\mathbf{Y}} = X\underline{\boldsymbol{\beta}} + \underline{\boldsymbol{\varepsilon}} \tag{1.1}$$

where $\underline{\mathbf{Y}}_{N\times1} = (Y_1, Y_2, \ldots, Y_N)'$ is the vector of observations, $\mathbf{X}_{N\times p}$ be the Design matrix, $\underline{\boldsymbol{\beta}}_{p\times1}$ be the vector of parameters and $\underline{\boldsymbol{\varepsilon}}_{N\times1} = (\varepsilon_1, \varepsilon_2, \ldots \varepsilon_N)'$ be the vector of random errors and assume that $\underline{\boldsymbol{\varepsilon}} \sim N(0,\sigma^2 I)$. The factor-response relationship is given by $E(Y) = f(x_1, x_2, \ldots, x_v)$ is called the 'Response Surface'. Design used for fitting the response surface models are termed as 'Response Surface Design'. The least square estimate of β is $\hat{\beta} = (X'X)^{-1}X'Y$ and the variance–covariance is $V(\hat{\beta}) = (X'X)^{-1}\sigma^2$

If the number of factors is more, only a few factors are important and underlying phenomena of interest, and assume that it is possible to eliminate the insignificant factors from the model, which are not affecting much the response. As a result the time, cost, effort and data complexity can be minimized with the reduction of dimensionality of the model. Dimensionality reduction has enormous applications in various fields like, in agricultural, pharmaceutical, biological and computer sciences, mechanical and chemical engineering etc.

High-dimensional data sets/models make many mathematical challenges are bound to give rise to new theoretical developments. Very few articles can be found on the reduction of the dimensionality of response surface design model. But there is no significant work done on reduction of response surface design model.

II. REDUCTION OF FIRST ORDER RESPONSE SURFACE DESIGN MODEL

Consider the linear functional relationship between the responses and 'v'factors.

$$Y_u = \beta_0 + \beta_1 X_{1u} + \beta_2 X_{2u} + \ldots\ldots + \beta_v X_{vu} + \varepsilon_u \tag{2.1}$$

where $\mathbf{Y_u}$ be the u^{th} response at the design point X_u (u = 1, 2, …N),

$X_u = (1, X_{u1}, X_{u2}, \ldots X_{uv})$ be the u^{th} treatment combination of 'v' factors,

$\underline{\beta} = [\beta_0, \beta_1, \beta_2, \ldots \beta_v]'$ is the vector of parameters and

ε_u be the random error corresponding to u^{th} response Y_u. Assume $\underline{\boldsymbol{\varepsilon}} \sim N(0,\sigma^2 I)$.

In this section, an attempt is made to fit a response surface design model in an iterative nested approach. The step by step procedure for finding the best model with selected factors is presented below.

Step 1: Let $(Y_1, Y_2, \ldots, Y_N)'$ be the vector of N observations, and X_{Nxv} be the design matrix, $F_1, F_2, \ldots, F_v$ are v factors. Assume initially $Y = \varepsilon_1$.

Step 2: Choose the maximum correlation coefficient factor as X_1 with Y and assume the nested model as $Y = \beta_{01} + \beta_1 X_1 + \varepsilon_2$.

Step 3: Evaluate the estimated responses and residuals ($\varepsilon_{i+1} = \varepsilon_i - \hat{Y}$ for i = 1,2,..v-1).

Step 4: Choose the maximum the correlated factor as X_i in the i^{th} step, between the residual and unselected factors, and assume the nested model as $\varepsilon_{i+1} = \beta_{0i+1} + \beta_{i+1}X_{i+1} + \varepsilon_{i+2}$.

Step 5: Estimate the nested model as $\hat{\varepsilon}_{i+1} = \hat{\beta}_{0i+1} + \hat{\beta}_{i+1}$ X_{i+1}. Test the significance of the model.

Step 6: Repeat the steps 3-5, if the model is significant and the resulting model is the best model with the selected factors can be expressed in the form

$$\hat{Y} = \hat{\beta}_0 + \sum_{m=1}^{k} \hat{\beta}_m X_m \text{ where } \hat{\beta}_0 = \hat{\beta}_{01} + \hat{\beta}_{02} + ... + \hat{\beta}_{0k} \text{ and } k \le v.$$

Note:

1. At each step the estimation of value of the parameter is uses least square method.
2. It reduces the size of the original model by selecting a subset of variables from the original set of variables iteratively in a forward approach.
3. The nested approach avoids the problem of multicollinearity by selecting single variable at each step.
4. The choice of selection of variables in the model is same when compared with Forward and Step wise regression approach.
5. It is a time consuming process due to estimation of parameters and testing its significance in each step of iteration.

The method for reducing the size of first order response surface design model (non-orthogonal design) is illustrated in the example 2.1 is presented below.

EXAMPLE 2.1: Consider the design and analysis strategy illustrated with a 27 run experiment (Taguchi (1987, P.423)) to study the problem of PVC insulation for electric wire, to understand the compounding method of Plasticizer, Stabilizer, and filler for avoiding embrittlement of PVC insulation, and finding the most suitable process conditions. Among the nine factors, two are about Plasticizer: DOA (X_1) and DOP (X_2); two about stabilizer: Tribase (X_3) and Dyphos (X_4); three about Filler: Clay (X_5), Titanium white (X_6), and Carbon (X_7); the remaining two about process condition: number of revolutions of screw (X_8) and cylinder temperature (X_9). All nine factors are continuous and their levels are chosen to be equally spaced. The measure is the embrittlement temperature (Y). The 27 runs of experimental values and the respective design points are presented below.
Y=[5,2,8,-15,-6,-10,-28,-19,-23,-13,-17,-7,-23,-31,-23,-34,-37,-29,-27,-27,-30,-35,-35,-38,-39,-40, -41]'
X=[(0,0,0,0,0,0,0,0,0),(0,0,0,0,1,1,1,1,1),(0,0,0,0,2,2,2,2,2),(0,1,1,1,0,0,0,2,2), (0,1,1,1,1,1,1,0,0), (0,1,1,1,2,2,2,1,1), (0,2,2,2,0,0,0,1,1), (0,2,2,2,1,1,1,2,2), (0,2,2,2,2,2,2,0,0), (1,0,1,2,0,1,2,0,1), (1,0,1,2,1,2,0,1,2), (1,0,1,2,2,0,1,2,0), (1,1,2,0,0,1,2,2,0), (1,1,2,0,1,2,0,0,1), (1,1,2,0,2,0,1,1,2), (1,2,0,1,0,1,2,1,2),(1,2,0,1,1,2,0,2,0), (1,2,0,1,2,0,1,0,1), (2,0,2,1,0,2,1,0,2), (2,0,2,1,1,0,2,1,0), (2,0,2,1,2,1,0,2,1), (2,1,0,2,0,2,1,2,1), (2,1,0,2,1,0,2,0,2), (2,1,0,2,2,1,0,1,0), (2,2,1,0,0,2,1,1,0), (2,2,1,0,1,0,2,2,1), (2,2,1,0,2,1,0,0,2)]′ respectively.

	Correlations	Nested model	Mean Squares & R^2 values
1	-0.762, -0.601, -0.124, -0.108, 0.052, -0.039, 0.114, 0.007, -0.026 .	$Y(=\varepsilon_1)= -8.830 - 13.343X_1 +\varepsilon_2$	MSR=3019.919, MSE= 87.283, R^2 =0.581 Significant
2	(-0.861, -0.124, -0.099, 0.148, 0.007, 0.244, -0.057, -0.108.	$\varepsilon_2 = 9.481 + -9.481X_2+ \varepsilon_3$	MSR=1617.990, MSE=22.564, R^2 = 0.741, Significant
3	-0.245, -0.195, 0.291, 0.013, 0.480, -0.113, -0.212.	$\varepsilon_3 = - 2.685 + 2.686X_7+ \varepsilon_4$	MSR= 129.836 MSE = 17.370, R^2 =0.230, Significant
4	-0.279, -0.222, 0.332, 0.015, -0.128, -0.241.	$\varepsilon_4 = - 1.631 + 1.630X_5+ \varepsilon_5$	MSR= 47.834 MSE = 15.457, R^2 =0.110, Insignificant

The nested reduced model is Y = - 2.034 - 13.343 X_1 - 9.481 X_2 +2.686 X_7, with error sum of squares 431.513 with 23 degrees of freedom and with an R^2 value is 0.917. The general least square fitted model is Y = 0.256 – 12.996 X_1 – 9.5 X_2 – 1.389 X_3 – 1.111 X_4 + 1.611 X_5 + 0.055 X_6 + 2.666 X_7 – 0.611 X_8 – 1.166 X_9 with error sum of squares 296.629 with 17 degrees of freedom and with an R^2 value is 0.943 . The stepwise, Forward and backward equations obtained are Y = - 2.408 – 12.940 X_1 - 9.503 X_2 + 2.663 X_7 with error sum of squares 431.513 with 23 degrees of freedom with an R^2 value is 0.917. It can be noted that the values of parameters β_3, β_4, β_5, β_6, β_8, β_9 are insignificant in nested and other regression approaches. The mean square error values for full and nested reduced models are 17.449 and 18.761.

The method for reducing the size of orthogonal first order response surface design model is illustrated in the example 2.1 is presented below.

EXAMPLE 2.2: Consider the 2^4 factorial experimental design illustrated in Khuri and Cornel (1996, page 79), in the study of the hydrogenolysis of Canadian lignite using carbon-monoxide and hydrogen mixtures as reducing agents, the input variables studied were X_1 = Temperature; X_2 = CO (H_2 ratio); X_3 = Pressure; and X_4 = Contact time. One of the response variables under investigation was Y = Percentage lignite conversion. The levels of four factors are with, X_1: Reaction temperature: 380 ^{o}c ,460 ^{o}c; X_2: Initial CO / H_2 ratio (molar ratio) is ¼: ¾ ; X_3: Initial Pressure (MPa) 7.10, 11.10; X_4: Contact time at reaction temperature (min) 10, 50.

Y = [53.3,78, 62.4, 78.9, 75.9, 75.4, 71.3, 84.4, 64.5, 67.5, 72.8, 85.3, 71.4, 83.3, 82.9, 81.7] ′ be the vector of responses at the design points [(-1,-1,-1,-1), (1,-1,-1,-1), (-1,1,-1,-1), (1,1,-1,-1), (-1,-1,1,-1), (1,-1,1,-1), (-1,1,1,-1), (1,1,1,-1), (-1,-1,-1,1), (1,-1,-1,1), (-1,1,-1,1), (1,1,-1,1), (-1,-1,1,1), (1,-1,1,1), (-1,1,1,1), (1,1,1,1)]′ respectively.

	Correlations	Nested Model	Mean Squares & R^2 values
1	(0.574, 0.361, 0.456, 0.214)	$Y(=\varepsilon_1) = 74.313+5.0\ X_1+\varepsilon_2$	MSR=400, MSE= 58.278 R^2 =0.329, Significant
2	(0.441, 0.557, 0.261)	$\varepsilon_2= -0.001 + 3.975\ X_3 + \varepsilon_3$	MSR=252.810,MSE=40.221 R^2 =0.310, Significant
3	(0.531, 0.314)	$\varepsilon_3= 0.0000+ 3.15X_2+ \varepsilon_4$	MSR=158.76, MSE =28.881 R^2 =0.282, Significant
4	(0.371)	$\varepsilon_4= 0.001.+ 1.862\ X_4+ \varepsilon_5$	MSR=55.502, MSE =24.916 R^2 =0.137, Insignificant

The nested reduced model is Y = 74.312 + 5 X_1 + 3.15 X_2 + 3.975 X_3, with error sum of squares 404.328 with 12 degrees of freedom and with an R^2 value is 0.667. The backward elimination procedure also resulting to the same as that of nested approach. The full model is Y = 74.313 + 5 X_1 + 3.15 X_2 + 3.975 X_3 + 1.862 X_4 with error sum of squares 348.825, 11 degrees of freedom and with an R^2 value is 0.713 . The stepwise and forward approaches gives Y = 74.313 + 5 X_1 + 3.975 X_3 with error sum of squares 563.088, 13 degrees of freedom and with an R^2 value is 0.537. It can be noted that the values of parameters β_4 is insignificant in nested and β_2, β_4 are insignificant in other regression approaches. The mean square error values for full and nested reduced models are 33.694 and 31.711. The percentage of loss of error due to the elimination of six components with respect to full model is 15.911 % .

Acknowledgments: The authors are grateful to the UGC for providing financial assistance for completing this work under the UGC-MRP and also thankful the referee for suggestions to improve the version.

References

[1] Bhatra Charyulu N.Ch. and Ameen Saheb SK. (2015): " A Note on Reduction of Dimensionality of First Order Response Surface Design Model"; International Journal of Statistika and Mathematika, Vol 12 (3).

[2] Christian Gogu, Haftka., R. T. Satish, K .B. and Bhavani (2009): "Dimensionality Reduction Approach for Response Surface Approximation: Application to Thermal Design", AIAA Journal, Vol.47, No.7, pp 1700-1708.

[3] Christine M. Anderson-Cook, Connie M. Borror, Douglas C. Montgomery (2009):" Response surface design evaluation and comparison", Journal of Statistical Planning and Inference.

[4] Feng-Jenq Lin, (2008):"Solving Multicollinearity in the process of Fitting Regression Model using the Nested Estimate Procedure", Quality and Quantity, Vol. 42 (3), pp 417-426.

[5] Khuri A.I. and Cornell J.A. (1996): " Response Surface. Design and Analysis 2nd edition Marcel Dekker, Newyork.

[6] Shao-wei Cheng and C. F. J. Wu (2001): " Factor Screening and Response Surface Exploration", Statistica Sinica, 11, 553-604.

[7] Taguchi G. (1987): " System of Experimental Design". Unipub / Kraus International Publications, White Plains, New York.

[8] Tengfei Long and Weili Jiao (2012): "An Automatic Selection and Solving Method for Rational Polynomial Coefficients based on Nested Regression", proceedings of the 33rd Asian Conference on Remote Sensing held on November 26-30, 2012. Ambassador City, Pattaya, Thailand.