Australian Journal of Basic and Applied Sciences, 4(8): 3418-3423, 2010
ISSN 1991-8178

Hodges Lehmann Estimator for Robust Design Using Dual Response Surface Approach

Amani Majid, Kassim Haron and Habshah Midi

Department of Mathematics, Faculty of Science,
Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor

Abstract: In this paper, we proposed the usage of the robust Hodges Lehmann estimator (HLE) as a location measure in the construction of optimal designs using the Dual Response Surface Approach (DRSA) where the location and scale measures were incorporated simultaneously into the model. The performance of six models with different pairs of location and scale estimators were compared. The results from a simulation study showed that for the non-normal and normal distribution with contaminated data, the model using HLE and inter-quartile range (IQR) performed better than the pair of the sample median and IQR as achieved by Park and Cho (2002).

Key words: Robust design, Taguchi method, Dual response, Hodges-Lehmann estimator, Expected bias

INTRODUCTION

Over the past ten years robust design has been identified as one of the most important designs technique used to improve industrial production. By minimizing the variability of the control factors, one can obtain the optimum operating conditions to control it. Hence robust design can achieve its objectives with keeping the process mean at its customer satisfactory target at the same time.

The pioneer Taguchi defined robust design as an "engineering methodology" for improving productivity during research and development so that high-quality product can be produced quickly and at low cost. However, in late 1980s and early 1990s much criticism have been drawn towards Taguchi's approach (Box, 1985; Vining and Myers, (1990), Myers, *et al*. (1992)) that led to much discussion to develop alternative approaches. An attractive alternative developed was the classical response surface approach, which was designed to construct an approximation model for the response *Y*. This model was then used to determine the best values of design parameters that optimize the response. In such problems the focus was on the mean of *Y* with the variance assumed to be small or constant. However, in this approach model inadequacy may occur which consequently may produced misleading optimization results. Another approach, the "Dual response surface approach (DRSA)" which consider two models; one for the mean and one for the variance at the same time was proposed and used to replace the classical response surface (Vining and Myers, 1990; Park and Cho (2002); Coetzer *et al*. (2008).

In this study we propose to incorporate the robust Hodges- Lehmann estimator (HLE) into the dual response model for generating optimal designs using the DRSA. The performance of the (location, scale) pairs consisting of HLE are then compared with those considered by Park and Cho, (2002). They found out that the model with (median, IQR) has the smallest expected bias of the mean response when a normality assumption is not met.

MATERIALS AND METHOD

The Location and Scale Estimators:

Let $\left(x_1, x_2, \ldots, x_k\right)$ be the *k* control factors vector and Y_{ij} represents the j^{th} response at the i^{th} design point where $i = 1,2,\ldots,n$. and $j = 1,2,\ldots,m$. The common estimators for the location and scale parameters are the sample mean and sample variance respectively.

Corresponding Author: Amani Majid, Department of Mathematics, Faculty of Science, Universiti Putra Malaysia,
43400 UPM, Serdang, Selangor
Email: amanimaj@hotmail.com

Tukey (1960), pointed out that these estimators can be heavily influenced by any single outlier especially the mean where if $y_i \to \pm\infty$ then $\overline{y}_i \to \pm\infty$. This prompted Park and Cho (2002) to also consider alternatives to the mean and variance in their study by using two models; sample median with sample median absolute deviation (MAD) and sample median with sample inter quartile range (IQR), apart from the model pairing sample mean with sample variance.

Another robust location estimator adopted by research workers in the past is the Hodges-Lehmann estimator (HLE), which is based on the rank of the response values. For a random sample ($x_1, x_2, \ldots, x_n$), the HLE is given by:

$$\text{HLE} = \begin{Bmatrix} W_{(k+1)} & \textit{if } M \textit{ is odd} \\ W_{(k)+(k+1)/2} & \textit{if } M \textit{ is even} \end{Bmatrix} \tag{1}$$

That is, HLE is the median of the $\frac{n(n+1)}{2}$ Walsh averages W_r where

$$W_r = \frac{x_i + x_j}{2} \quad \begin{Bmatrix} i \leq j \\ r = 1, 2, \ldots, M \\ i, j = 1, 2, \ldots, n \end{Bmatrix} \tag{2}$$

(Abu-Shawiesh and Abdullah, 2000; Wei, 2007):

The median absolute value (MAD) and the inter quartile range (IQR) are the robust scale estimators that can be good alternatives to the variance. By denoting $y_{[k]}$ as the k^{th} order statistics then the MAD and IQR become:

$$MAD(Y_\downarrow 1, \ldots Y_\downarrow n) = MED_i \left\{ Y_i - MED_j (Y_J) \right\} \tag{3}$$

$$IQR(Y_1, \ldots Y_n) = Y_{[3n/4]} - Y_{[n/4]} \tag{4}$$

Where MED is the median.

Let the random variable $Y \sim N(\mu, \sigma^2)$, Z another random variable $\sim N(0,1)$ with the cumulative function F(.), so Y-μ = $Z\sigma$ have the same distribution.

Let $\hat{\sigma}$ = L.MAD, where L is a constant scale factor, which depend on the distribution.

$$\text{If } \mathrm{p}(|y - \mu| \leq MAD) = \mathrm{p}^{\left(\left|\frac{y-\mu}{\sigma}\right| \leq \frac{MAD}{\sigma}\right)}$$

$$= p\left(|Z| \leq \frac{MAD}{\sigma}\right) = \frac{1}{2}$$

Since MAD is the 75% percentile for a symmetric distribution about the mean then:

$$\frac{MAD}{\sigma} = F^{-1}(3/4)$$

Thus the scale factor to use the MAD is $\frac{3}{4}$ of the normal distribution with σ=1 then L becomes

$$L = 1/F^{-1}\left(\frac{3}{4}\right)$$

And

$$MED^{(|Z|)} = F^{-1}\left(\frac{3}{4}\right)$$

When n is large size,

$$MED_j(Y_j) \rightarrow \mu \;\; and \;\; Z \rightarrow F^{-1}(p) =$$

Therefore

$$MAD(Y_{\downarrow}1,...Y_{\downarrow}n) \rightarrow MED\{|Y-\mu|\} = MED(|Z|)\sigma^2$$

$$= F^{-1}\left(\frac{3}{4}\right)\sigma^2$$

And

$$IQR \rightarrow \left\{F^{-1}\left(\frac{3}{4}\right) - F^{-1}\left(\frac{1}{4}\right)\right\}\sigma^2$$

Hence, we have

$$D^2 = \left[MAD(Y_1,...Y_n) \Big/ F^{-1}\left(\frac{3}{4}\right) \right]^2 \qquad (5)$$

$$Q^2 = \left[MAD(Y_1,...Y_n) \Big/ \left\{F^{-1}\left(\frac{3}{4}\right) - F^{-1}\left(\frac{3}{4}\right)\right\} \right]^2 \qquad (6)$$

Dual Response Surface Approach (DRSA): The response surface methodology (RSM) techniques are useful for the modeling and analysis of problems in which one or more responses of interest are influenced by several variables. The method was introduced by G.E.P. Box and K.B. Wilson (1951). The main idea of RSM is to use a sequence of designed experiments to optimize the response *Y*. The classical response surface methods focus on the mean value of *Y* where the variance is assumed to be small constant or equal to zero (Unal and Yeniay, 2003). However, if the variance is not constant, RSM may not be adequate and the result can be misleading.

In such situations, the Dual Response Surface Approach (DRSA) may be used as an alternative to the classical RSM (Vining and Mayers, 1990). This approach utilizes response surface in modeling process relationship by separately estimating the response functions for the process mean and variance for the system under investigation. Then based on the optimization strategy chosen, these functions are optimized simultaneously over the region of interest to determine the system optimum operating conditions (Shaibu and Cho, 2009).

In this study, we investigate the performance of HLE as a location estimator to be paired with the scale parameters MAD, IQR as adopted by Park ad Cho (2002) into the dual response surface that estimate the overall shape of the curve with much precision as possible. By assuming a second order polynomial model to represent the response function so that the fitted response functions for the mean (measure of location) and the variance (measure of scale) of the response *Y* are given by $\hat{\mu}(x)$ and $\hat{\sigma}^2(x)$ respectively,

where

$$E(y) = \hat{\mu}(x) = \hat{\beta}_0 + \sum_i \hat{\beta}_i x_i + \sum_i \sum_j \hat{\beta}_{ij} x_i x_j \qquad (7)$$

And

$$V(y) = \hat{\sigma}^2(x) = \hat{\alpha}_0 + \sum_i \hat{\alpha}_i x_i + \sum_i \sum_j \hat{\alpha}_{ij} x_i x_j \quad (8)$$

The main objective of the robust design, that is finding the optimum operating conditions of the control factors $\left(x_1^*, x_2^*, \ldots, x_k^*\right)$, is carried out by minimizing the function $L = \left[\hat{\mu}(x) - T_0\right]^2 + \hat{\sigma}^2(x)$, where T_o is the target response value.

RESULTS AND DISCUSSION

Six different models formed with different combination of the location and scale estimators were considered as shown in Table 1.

The R language were used to generate random responses (Y) from Normal, 2% contaminated Normal to check how the contaminated data affects estimators.

The responses are also generated from another two distributions to check the lack of normal assumption such as Double exponential and Logistic respectively whose PDF's as follows:

$$f(x) = \frac{1}{2\beta} e^{\left((-|x-\mu|)/\beta\right)} \quad (9)$$

$$f(x) = \frac{1}{2\beta\left[1 + \cosh\left\{(x-\mu)/\beta\right\}\right]} \quad (10)$$

Using a 3^3 factorial design with three control factors $x_i = (x_{i1}, x_{i2}, x_{i3})$, ($i$ =1, 2… 27), occurring at each five replicates of y_{ij}, (j =1, 2, … 5), $y \sim \phi\left(\mu(x_i), \sigma^2(x_i)\right)$ where

$$\mu(x) = 50 + 5(x_1^2 + x_2^2 + x_3^2) \quad \text{and} \quad \sigma^2(x) = 100 + 5\left\{\left(x_1 - 0.5\right)^2 + x_2^2 + x_3^2\right\}$$

Table 1: Six models with different (location, scale) pairs

Model	$\hat{\mu}(x)$	$\hat{\sigma}^2(x)$
A	Mean	Variance
B	Median	MAD
C	Median	IQR
D	HLE	Variance
E	HLE	MAD
F	HLE	IQR

Table 2: Estimates of coefficients of $\hat{\mu}(x)$ for contaminated normal distribution.

	Models		
Coefficients	A (Mean)	B (Median)	C (HLE)
β_0	56.780	54.26	53.571
β_1	3.7000	0.3889	0.2500
β_2	-3.1333	0.1667	-1.2778
β_3	-1.8000	-1.4444	-0.5556
β^2_1	7.9889	9.3889	8.6389
β^2_2	-3.4444	-1.6111	0.8889
β^2_3	1.0889	7.5556	7.5556
β_{12}	-3.8667	-3.4167	-1.2917
β_{13}	-0.2667	1.25500	-0.1667
β_{23}	1.6333	2.8333	1.8750

Table 3: Optimal settings of six models for contaminated normal distribution.

Model	x^*	$\hat{\mu}(x^*)$

A	(-0.03074, 0.00544, -0.01964)	56.85642
B	(0.11010, -0.05576, -0.00250)	54.53166
C	(0.08972, 0.00264, -0.03242)	54.10326
D	(-0.01666, -0.00886, -0.02720)	59.26712
E	(0.12928, -0.04664, 0.00142)	54.53612
F	(0.08920, 0.00722, -0.02540)	53.66976

Table 4: Estimated bias of $\hat{\mu}\left(x^*\right)$ for the six models

	Model					
Data	A	B	C	D	E	F
Normal	3.40	4.2	3.64	3.40	4.19	3.70
Normal. (contaminated)	6.86	4.53	4.10	9.27	4.54	3.67
Logistic	6.54	6.59	5.98	6.60	6.79	6.01
Double exponential	6.05	5.75	5.11	6.14	5.72	4.90

By assuming T_0 = 50, a total of 500 iterations were run with each having 27 design points and 135 responses. The expected bias measure of the mean response was used as the criterion in comparing the performance of the estimators.Table 2 and Table 3 show the resulting estimate of the coefficients of the mean response and the optimal settings respectively for the normal distribution with contamination.

From the optimal responses shown in Table 3, the expected bias of the optimal response for each model with respect to the targeted value T_0 = 50 is computed and displayed in Table 4. Clearly, model F with the pair (HLE, IQR) is the best choice among the six models since it produced optimal settings with the smallest expected bias of $\hat{\mu}(x^*)$ for the double exponential and normal distribution with contaminated data and the second smallest expected bias (with minimal difference) for the logistic distribution.

Conclusions:

In this study, we have proposed to construct optimal designs using the Dual Response Surface Approach (DRSA) by incorporating the robust HLE estimator together with the scale estimators considered by Park and Cho[4] into the model. The results from a series of simulation revealed that the model with the pair (HLE, IQR) performed better than the model using sample median and IQR, concluded to be the best model by Park and Cho (2002). The model (HLE, IQR) yielded optimal designs with the smallest expected bias of $\hat{\mu}(x^*)$ for the double exponential and normal distribution with contaminated data and second smallest expected bias for the logistic distribution.

REFERENCES

Abu-Shawiesh, M.O. and M.B. Abdullah, 2000. New Robust Statistical Process Control Chart For Location, Quality Engineering, 12(2): 149-159.

Box, G.E.P. and K.B. Wilson, 1951. On the Experimental Attainment of Optimum Conditions (with discussion). Journal of the Royal Statistical Society Series, B 13(1): 1-45.

Box, G.E.P., 1985. Discussion of O_-line Quality Control, Parameter Design and theTaguchi methods. Journal of Quality Technology, 17: 198-206.

Coetzer, R.L.J., R.F. Rossouw and D.K.J. Lim, 2008. DUAL Response Surface O ptimization with hard-to-control variables for sustainable Gazifier Performance. Journal of the Royal Statistical Society, 57(5): 567-587.

Myers, R.H., A.I. Khuri and G.G. Vining, 1992. Response Surface Alternatives to the Taguchi Robust Design Problem. American Statistician, 46: 131-139.

Park, C. and B.R. Cho, 2002. Development of Robust Design Under Contaminated and Non-Normal Data.

Shaibu, A.B. and B.R. Cho, 2009. Another view of dual response surface modeling and optimization in robust parameter design, Int Adv Manuf Technol., 14: 631-641.

Tukey, J.W., 1960. A Survey of Sampling from Contaminated Distributions. InI. Olkin, S. Ghurye, W. Hoe_ding, W. Madow and H. Mann, editors, Contributions to Probability and Statistics; University Press: Stanford, pp: 448-485.

Unal, R. and O. Yeniay, 2003. Reducing Deign Risk Using Robust Design Methods: A Dual Response Surface Approach, Final report, ODU project No:113091,NASAGrant No:NAG-1-01086.

Vining, G.G. and R.H. Myers, 1990. Combining Taguchi and Response Surface Philosophies:a Dual Response Approach. Journal of Quality Technology, 22: 38-45.

Wei, G.C., 2007. Comparisn between a Robust control chart based on Hodges Lehmann estimator and classical hotelling's T square chart. pp: 8-11.