APPLICATION OF RESPONSE SURFACE METHODOLOGY FOR CAPTURING OPTIMUM RESPONSE IN A LONGITUDINAL SURVEY

[1]**O.M. Olayiwola**   [2]**G.N. Amahia ,**   [3]**A.A. Adewara** [4]**A. U. Chukwu**

1. Department of Mathematical Sciences, Redeemer's University, Mowe, Ogun State, Nigeria
2. Department of Statistics, University of Ibadan, Ibadan. Nigeria.
3. Department of Statistics, University of Ilorin, Ilorin. Nigeria.

Abstract

Non-response rates in surveys have been recognized as important indicators of data quality since they introduce bias in the estimates which increases the mean square error. This study was designed to apply Response Surface methodology in a Longitudinal survey to reduce non response and capture optimum response. Seven hundred and fifty (750) households in Oyo town were randomly selected. House-heads were interviewed in five waves. An interviewer-administered questionnaire was used to collect data on demographic characteristics and response predictors. Demographic characteristics were analyzed using summary statistics. Multi-way contingency tables were constructed to establish relationships and dependence structures among the variables under investigation. A log-linear model was fitted to constructed contingency tables to capture significant predictors of response. Using demographic characteristics, a Response Surface Model (RSM) was constructed and subjected to canonical analysis for the characterization of its turning point and to capture the combination of levels of response predictors that produced optimum response. Log-linear model showed that family size (x_1), duration of interview (x_2), and their interaction (x_1, x_2) significantly ($p < 0.05$) determined response rate. The RSM has an adjusted $R^2 = 0.722$. Canonical analysis of the RSM gave eigen-values -0.007 and -0.002. The turning point of the RSM was a maximum implying the point for optimum response. The response was optimum when the family size was three and duration of interview was twelve minutes.

Keywords: Longitudinal survey, Response predictors, Non-response rate.

Introduction

With the increased focus on experimental design, response surface methods have received considerable attention in recent decades. This interest has grown from the need for quality and precision in industry. Statistically designed experiments are one of the most powerful tools in statistical analysis as they can greatly increase the efficiency of experiments. The aim of such experimentation is to find out how a number of experimental variables affect a response, and to find the combination of conditions that provides the highest response, as well as to understand the relationship over a region of interest, Box and Liu (1999), Box (1999). Response surface methodology (RSM) originated with the work of Box and Wilson (1951), who were at the time involved in industrial research with ICI in the United Kingdom. There are many situations for which RSM has proved to be a very useful tool. Hill and Hunter (1966) illustrated chemical and processing applications of canonical analysis and use of multiple responses. Mead and Pike (1975) investigated the extent to which RSM had been used in applied research and gave examples from biological applications. Myers et al (1989) summarized the developments in RSM that had occurred since the review of Hill and Hunter (1966), while a more recent summary of the current status of RSM and some indication of possible developments was given by Myers (1999).

The exploration of an experimental region using response surface methods revolves around the assumption that the expected response, $E\{y)$, is a function of controllable variables $x_1, x_2,... ,x_k$; where the x_j's are suitably scaled and centred linear transformations of the independent variables.

According to Hill and Hunter (1996), *RSM* method was introduced by Box and Wilson (1951). Box and Wilson (1951) suggested to use a first-degree polynomial model to approximate the response variable. They acknowledged that this model is only an approximation, not accurate, but such a model is easy to estimate and apply, even when little is known about the process Wikipedia (2006). Moreover, Mead and Pike (1975) stated that the origin of RSM starts in the 1930s with use of Response Curves, Myers et al (1989).

According to research conducted Myers et al (1989), the orthogonal design was motivated by Box and Wilson (1951) in the case of the first-order model. For the second-order models, many subject- matter scientists and engineers have a working knowledge of the central composite designs (CCDs) and three-level designs by Box and Behnken (1960). Also, the same research states that another important contribution came from Hartley (1959), who made an effort to create a more economical or small composite design.

There exist many papers in the literatures about the response surface models.

According to Myers et al (1989), the important development of optimal design theory in the field of experimental design emerged following Word World 4 II. Kiefer (1958, 1959, 1961, 1962) was author who published his work on optimality. One of the important facts is whether the system contains a maximum or a minimum or a saddle point, which has a wide interest in industry. Therefore, RSM is being increasingly used in the industry. Also, in recent years, more emphasis has been placed by the chemical and processing field for finding regions where there is an improvement in response instead of finding the optimum response (Myers et al 1989). Optimization in simulation has been attempted by many methods; Fu (2002), Tekin and Sabuncuoglu (2004), and Kleijnen (2008a). These methods can be classified as either white-box or black-box methods. Examples of white-box methods are perturbation analysis (Ho and Cao, 1991; Glasserman, 1991) and the likelihood ratio score function (Rubinstein and Shapiro, 1993), which estimate gradients.

Recent case studies of RSM optimization of stochastic simulation are presented by Irizarry et al (2001); a case study of RSM for deterministic simulation is presented by Ben-Gal and Bukchin (2002). Case studies of RSM applied to real, non-simulated systems are given in the standard RSM textbooks by Myers and Montgomery (2002) and Khuri and Cornell (1996). Originally, RSM was derived for problems with a single stochastic objective function and deterministic box constraints on the inputs. In practice, however, optimization problems may have constraints for the stochastic outputs. For example, inventory simulation may minimize the total holding and ordering cost under a service level constraint. In RSM, there are several approaches to solve constrained optimization problems. Khuri (1999) surveys most of these approaches, including the desirability function (Harrington, 1965; Derringer and Suich, 1980), the generalized distance (Khuri and Conlon, 1996), and the dual response, Myers and Carter (1973), Del Castillo and Montgomery(1993), Fan and Del (1999). Furthermore, Wei et al (1990) suggest so-called prediction-interval constrained goal programming. In all these approaches, the constrained optimization problem is reformulated by combining the constraints and the original objective function into a new, single objective function, using appropriate transformations. The resulting unconstrained optimization problem is solved through an ordinary nonlinear programming algorithm.

2. Methodology

Seven hundred and fifty (750) households in Oyo town were randomly selected. House heads were interviewed in five waves. An interviewer-administered questionnaire was used to collect data on demographic characteristics and response predictors. Demographic characteristics were analyzed using summary statistics. Multi-way contingency tables were constructed to establish relationships and dependence structures among the variables under investigation. A log-linear model was fitted to constructed contingency tables to capture significant predictors of response. Using demographic characteristics, a Response Surface Model (RSM) was constructed and subjected to canonical analysis for the characterization of its turning point and to capture the combination of levels of response predictors that produced optimum response.

3.0 RESULTS

The effect of the predictors of response is explained with respect to their odd ratios. Odd ratio greater than 1 means positive association, less than 1 means negative association and 1 means no association between the variables. Family size, duration of interview, education, number of visit, language of interview, familiarity, gender, house ownership, nationality and duration of residence in a community are positively related to the response rate. Age is negatively related to the response rate and there is no association between employment status and response rate (table 1).

Log-linear model showed that both duration of interview and family size together with their interaction significantly ($p < 0.05$) determined response rate. The first order RSM fit the data well (table 2), but the analysis of variance result (table 3) showed that there is no significant evidence of lack of fit (p-value = 0.990) at $\alpha = 5\%$. Therefore, we conducted a more appropriate model, a second-order Response Surface Model (table 4).

$$y = \beta_0 + \sum_{i=1}^{k} \beta_{ii} x_i^2 + \sum_{1 \leq i < j \leq k} \beta_{ij} x_i x_j + \in \dots\dots\dots\dots\dots\dots\dots\dots\dots 1.0$$

was fitted to the data. If the fitted second order model is written in matrix notation, then

$$\hat{y} = b_0 + X^1 b + X^1 B X \dots\dots\dots\dots\dots\dots\dots\dots\dots 2.0$$

Where

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_k \end{bmatrix}, \; b = \begin{bmatrix} b_1 \\ b_2 \\ \cdot \\ \cdot \\ \cdot \\ b_k \end{bmatrix}, \; B = \begin{bmatrix} b_{11} & \frac{1}{2}b_{12} & \cdot & \cdot & \frac{1}{2}b_{1k} \\ \frac{1}{2}b_{21} & b_{22} & \cdot & \cdot & \frac{1}{2}b_{2k} \\ \cdot & & \cdot & & \cdot \\ \cdot & & & \cdot & \cdot \\ \frac{1}{2}b_{k1} & \frac{1}{2}b_{k2} & \cdot & \cdot & b_{kk} \end{bmatrix} \quad\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 3.0$$

In equation (3.0), b is a (k x 1) vector of the first-order regression coefficients and B is a (k x k) symmetric matrix whose diagonal elements are the pure quadratic coefficients, and whose off-diagonal elements are one-half the mixed quadratic coefficients. Differentiating *y* in equation (2.0) with respect to the vector x and equating it to zero results in

$$\frac{dy}{dx} = b + 2BX = 0 \quad\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 4.0$$

Therefore, the stationary point is

$$X_s = -\frac{1}{2}B^{-1}b \quad\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 5.0$$

and the predicted response at the stationary point is

$$\hat{y}_s = \hat{\beta}_0 + \frac{1}{2}X_s^1 b \quad\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 6.0$$

The analysis of variance for second order response surface model (table 5) indicates that, there are significant interactions between the duration of interview and the family size and there is significant evidence of Lack of fit (p-value = 0.022) at α = 5% .

The turning point or stationary point of the second order RSM (figure 1) was obtained as,

$$X^* = -\frac{1}{2}(B^{-1})b = \begin{Bmatrix} 11.6423 \\ 3.000 \end{Bmatrix}$$

at the turning point, Duration of interview (x_1) = 12 minutes and family size (x_2) = 3. The nature of the turning point was obtained by determining the eigen-values (λ_1, λ_2) of the characteristic equation below

$$|B - \lambda I| = \left| \left\{ \begin{pmatrix} -0.007 & 0.000 \\ 0.000 & -0.002 \end{pmatrix} - \lambda \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right\} \right| = 0$$

$\lambda_1 = -0.007$ and $\lambda_2 = -0.002$

Both λ_1 and λ_2 are negative. This implied that the turning point of the model was a maximum point. Solving the model gives duration of interview *(x_1)* to be twelve minutes and family size *(x_2)* to be three. Thus, we concluded that duration of interview of twelve minutes and family size of three resulted in optimum response in a longitudinal survey.

5. Conclusion

Females respond better to survey questions than males. The higher the educational qualification, the higher the response rate . The response rate from those that were living with their spouse was higher than those that were not living with their spouse. The response from those that were interviewed with English language was higher compared with those that were interviewed with Yoruba language. Respondents at the middle age (50-79 years) respond better to survey questions compared with youth and old age respondents The response rate from those that are familiar with the interviewer was higher than those that are not familiar with the interviewer. Response rate increased from first visit to fourth visits and at fifth visit, it declined. Response rate from tenants was higher than the owner occupiers. There was no significant difference in the response rate from unemployed respondents and employed respondents. The more the number of years a respondent has spent in his/her community, the more they response to survey questions. The response from Nigerians was higher than that of the non Nigerians.

6. References

1. Ben-Gal, I., and J. Bukchin. (2002). The ergonomic design of working environment via rapid prototyping tools and design of experiments. IIE Transactions 34 (4): 375 – 391.
2. Box, G.E.P.(1999), 'Statistics as a catalyst to learning by scientific method. Part II – a discussion'.Journal of Quality Technology 31, 16-29.
3. Box, G.E.P. and D.W. Behnken (1960), 'Some new three-level designs for the study of quantitative variables', Technometrics 2, 455-475.

4. Box, G.E.P. and K.B. Wilson (1951), 'On the experimental attainment of optimal conditions'. Journal of the Royal Statistical Society, Series B 13, 1-45.

5. Box, G.E.P. and P.Y.T. Liu (1999), 'Statistics as a catalyst to learning by scientific method. Part I - an example', Journal of Quality Technology 31, 1-15.

6. del CastiUo, E. and D. C. Montgomery (1993), 'A nonhnear programming solution to the dual response problem'. Journal of Quality Technology 25, 199-204.

7. Derringer, G. and R. Suich (1980), 'Simultaneous optimization of several response variables'. Journal of Quality Technology 12, 214-219.

8. Fan, S. K., and E. Del Castillo (1999): Calculation of an optimal region of operation for dual response system fitted from experimental data. Journal of the operational Research Society 50(8): 826 – 836.

9. Fu, M. C. 2002. Optimization for simulation: theory vs. practice. INFORMS Journal on Computing 14 (3): 192 - 215

10. Glasserman, P. 1991. Gradient Estimation via Perturbation Analysis. Kluwer Academic, Dordrecht.

11. Harrington, Jr., E. C. (1965), 'The desirability function". Industrial Quality Control 21, 494-498.

12. Hartley H.O (1959): Smallest Composite Designs for Quadratic Response Surfaces,". Biometrics, 15, 611–624

13. Hill, W.J. and W.G. Hunter (1966), 'A review of response surface methodology: A literature survey', Technometrics 8, 571-590.

14. Ho, Y. C., and X. R. Cao. 1991. Perturbation Analysis of Discrete Event Dynamic Systems. Kluwer Academic, Dordrecht

15. Irizarry, M., J. R. Wilson, and J. Trevino. 2001. A flexible simulation tool for manufacturing cell design, II: Response sorface analysis and case study. IIE Transactions 33: 837 – 846.

16. Khuri A.I, Cornell J.A. 1996. ResponsenSurfaces: Design and analysis. 2nd edition. Marcel Dekker, Monticello, NY.

17. Khuri, A.L (1999), 'Discussion', Journal of Quality Technology 31, 58-60.

18. Kiefer, J. (1958), 'On the non-randomised optimality and the randomised non-optimality of symmetrical designs'. The Annals of Mathematical Statistics 29, 675-

19. Kiefer, J. (1959), 'Optimum experimental designs'. Journal of the Royal Statistical Society, Series B 21, 272-319.

20. Kiefer, J. (1961a), 'Optimum designs in regression problems. 11". The Annah of Mathematical Statistics 32, 298-325.

21. Kiefer, J. (1961b), Optimum experimental designs V, with apphcations to systematic and rotatable designs, m 'Fourth Berkeley Symposium'. Vol. 1, pp. 381- 405.

22. Kiefer, J. (1962), 'Two more criteria equivalent to D-optimahty to designs". The Annals of Mathematical Statistics 33, 792-796.

23. Kleijnen, J.P.C., W. van Beers, and I. van Nieuwenhuyse. 2008. Constrained optimization in simulation: a novel approach. Working paper, Tilburg University, Tilburg, Netherlands.

24. Mead, R. and D.J. Pike (1975), 'A review of response surface methodology from a biometric viewpoint', Biometrics 31, 803-851.

25. Myers, R.H. (1999), 'Response surface methodology - current status and future directions'. Journal of Quality Technology 31, 30-44.

26. Myers, R.H. and W.H. Carter, Jr. (1973), 'Response surface techniques for dual response systems', Technometrics 15, 301-317.

27. Myers, R.H., and D.C. Montgomery. 2002. Response Surface Methodology: Process and Product Optimization Using Designed Experiments, second edition. Wiley, New York.

28. Myers, Raymond H., Khuri, Andre I. and Carter, Walter H., Jr. (1989). Response surface methodology: 1966-1988. Technometrics 31 (2): 137-153 http://www.jstor.org/ (accessed January 29, 2007).

29. Rubinstein, R.Y., and A. Shipiro. 1993. Discrete Event Systems: Sensitivity Analysis and Stochastic Optimization by the score Function Method. Wiley, Chichester.

30. Rubinstein, R.Y., and D. P. Kroese. 2004. The Cross- E ntropy Method: A Unified Approach to Combinatorial Optimization, Monte-Carlo Simulation and Machine Learning. Springer, New York.

31. Tekin, E., and I. Sabuncuoglu. 2004. Simulation Optimization: a comprehensive review on theory and applications. IIE Teansactions 36: 1067 – 1081.

32. Wei, C.J, D.L Olson, and E.M. White. 1990. Simultaneous optimization in process quality control via prediction-interval constrained programming. Journal of the Operational Research Society 41 (12): 1161 – 1167.

Table 1: Odd Ratios for Predictors of Response

Predictors of Response and Response Rate	Odd ratios
Response rate * tribe	1.266
Response rate * age	0.7596
Response rate * language of interview	1.1411
Response rate * familiarity with interviewer	1.4064
Response rate * education	2.7511
Response rate * number of visit	2.7899
Response rate * gender	1.1853
Response rate * house ownership	1.1219
Response rate * family size	1.7402
Response rate * duration of interview	1.1185
Response rate * spouse kind of settlement	1.3298
Response rate * employment status	1.007
Response rate * year of reciding	1.137

Table 2: Fitted First Order Surface Response Model

Model	Coefficient	P-value
Constant	0.418	0.011
Duration of interview (x_1)	0.058	0.049
family size (x_2)	-0.002	0.026

a Predictors: (Constant), Familysize, Duration
b Dependent Variable: Response Rate

Table 3: Analysis of Variance for lack of fit for first order Response Surface Model

Variation	Mean Square	p-value
Residual	29.56	
Lack of fit	28.41	0.990
Pure Error	0.96	

Table 4: The fitted Second order Surface Response Model

Model	Coefficient	p-value
Constant	0.154	.018
Duration	0.163	.002
familysize	0.011	.017
duration *duration	-0.007	.001
familysize *familysize	-0.002	.006
duration* familysize	0.000	.039

Table 5: Summary for the fitted Second Order Surface Response Model

Model	R-square	Adjusted R- square	P-value
Duration of interview, family size, duration of interview*family size	0.855	0.722	0.001

Fig 1: Graphical illustration of Predicted Percentage Response Rate