

Robustness of Higher Levels Rotatable Designs for Two Factors against Missing Data

Nyakundi Omwando Cornelious¹ & Evans Mbuti Kilonzo²

^{1,2} Department of Mathematics, Physics, and Computing

Moi University, Eldoret

Kenya

ABSTRACT

Experimenters should be aware of the possibility that some of their observations may be unavailable for analysis. This paper considers a criterion which assesses the robustness for missing data when running four and five levels designs in estimating a full second order polynomial model. The criterion gives the maximum number of runs that can be missing and still allow the remaining runs to estimate a second order model for four and five levels.

Key words: Robustness Criterion, Missing Data, Four and Five Level Designs, Second Order Rotatability.

1.0 INTRODUCTION

Response surface methodology (RSM) is a collection of statistical and mathematical techniques used for the purpose of setting up a series of experiments for adequate predictions of responses. In a majority of experiments which utilize response surface methodology, there is a possibility that some observations may be unavailable or missing for analysis. Missing data is a problem that occurs in many experiments and can have substantial consequences on study quality [1]. Strategies to limit the impact of missing data on the analysis and interpretation of experiments are supported by the natural academy of sciences report [2]. The report recommends that a more principled approach to design and analysis in the presence of missing data is both needed and that careful design and conduct limit the amount and impact of missing data [1]. Experiments which utilize higher level models can become lengthy and as such missing data is commonly seen through subject dropout [2]. First introduced by [6], and also defined on page 32 of [5], data are said to be missing at random, conditional on the observed data. The problem of missing observations has been common lately in most experiments due to its existence in practice [8]. The causes of missing data include; loss of experimental units, miscoded data, and the use of experiments that take too long to complete, resulting in the cancellation of runs [7].

Therefore, the designs chosen must be robust in situations where some information is missing. The inclusion of sensitivity to outliers as one important property of an experimental design was done by [3]. Several other authors have investigated the robustness of designed experiments against missing data [4]. However, most of this research has focused on factorial designs and complete and incomplete block designs. The aim of this article is to study the robustness for missing data for different two factor designs in four and five dimensions for estimating a full second order polynomial model.

2.0 METHODS

Response surface methodology explores the relationships between several explanatory variables and one or more response variables with the aim of optimizing the response variables. It is assumed that the form of the functional relationship is unknown but that within the range of interest, the function could be represented by a Taylor series expansion of moderately low order. For example in two independent variables and to terms of second order model, the series has the form

$$y = B_1x_1 + B_2x_2 + B_{11}x_1^2 + B_{22}x_2^2 + B_{12}x_1x_2 \quad (1)$$

Where; y is the expected value of the dependent variable, x_1 and x_2 are the independent variables and B 's are unknown coefficient in the series. Normally, the specific problem considered is that of the choice of the combinations of the independent variables so as to achieve certain desirable properties in the final estimates.

Expanding equation (1) in a Taylor series gives an infinite series which might be expressed as

$$y = B_0 + \sum_{i=1}^k B_i x_i + \sum_{i < j=1}^k B_{ij} x_i x_j \tag{2}$$

2.1 Moment Conditions for SORD

A Second order rotatable design is possible if the following moment conditions and non-singularity conditions for second order rotatability are satisfied.

$$\sum_{u=1}^N x_{iu}^2 = N\lambda_2, \quad i = 1, 2, \dots, k, \tag{3}$$

$$\sum_{u=1}^N x_{iu}^4 = 3 \sum_{u=1}^N x_{iu}^2 x_{ju}^2 = 3N\lambda_4 \quad i \neq j = 1, 2, \dots, k, \tag{4}$$

2.2 Non-Singularity Conditions for TORD

The non- singularity conditions for second order rotatability can be given by,

$$\frac{\lambda_4}{\lambda_2^2} > \frac{k}{k+2} \tag{5}$$

for $k = 3, 4, \dots, k$

Where λ_2 and λ_4 in (5) are as obtained in (3), and (4) respectively.

2.3 Construction of SORD in Four and Five Levels

2.3.1 Thirty two points second order rotatable design in four dimensions

The design is represented by the set of thirty two points denoted by s_1 given by,

$$s_1 = s(a, a, a) + s(c_1, 0, 0) + s(c_2, 0, 0), \tag{6}$$

The set of points given in (6) yields the moment conditions given by;

$$\sum_{iu}^{32} x_{iu}^2 = 16a^2 + 2c_1^2 + 2c_2^2 = 32\lambda_2 \tag{7}$$

$$\sum_{iu}^{32} x_{iu}^4 = 16a^4 + 2c_1^4 + 2c_2^4 = 96\lambda_4 \tag{8}$$

$$\sum_{u=1}^{32} x_{iu}^4 - 3 \sum_{u=1}^{32} x_{iu}^2 x_{ju}^2 = c_1^4 + c_2^4 - 16a^4 = 0 \tag{9}$$

Let,

$$c_1^2 = xa^2, \text{ and } c_2^2 = ya^2 \tag{10}$$

Then from (9),

$$x^2 + y^2 = 16 \tag{11}$$

Assuming that $x = 3$, then $y = 2.645751311$ and hence the values of x and y are real and positive as required for rotatability.

$$\text{Thus } c_1^4 = 9a^4, \text{ and } c_2^4 = 7a^4 \tag{12}$$

Substituting (12) in (8) and (7) gives;

$$\lambda_4 = 0.5a^4 \text{ and } \lambda_2 = 0.852859456a^2 \tag{13}$$

Substituting (13) to (5) gives,

$$\frac{\lambda_4}{\lambda_2^2} > \frac{k}{k+2} = 0.987408764 > 0.6666 \tag{14}$$

The conditions given in 3, 4 and 5 are satisfied, hence s_1 forms a four level rotatable design for two factors.

2.3.1 Eighty two points second order rotatable design in five dimensions

The design is represented by the set of thirty two points denoted by s_2 given by,

$$s_2 = s(f, f, 0, 0, 0) + s(a, a, a, a, a) + s(c, 0, 0, 0, 0), \tag{15}$$

The set of points given in (15) yields the moment conditions given by;

$$\sum_{u=1}^{82} x_{iu}^2 = 32a^2 + 16f^2 + 2c^2 = 82\lambda_2 \tag{16}$$

$$\sum_{u=1}^{82} x_{iu}^4 = 32a^4 + 16f^4 + 2c_2^4 = 246\lambda_4 \tag{17}$$

$$\sum_{u=1}^{82} x_{iu}^4 - 3 \sum_{u=1}^{82} x_{iu}^2 x_{ju}^2 = 2f^4 + c_2^4 - 32a^4 = 0 \tag{18}$$

$$c_2^2 = xa^2, \text{ and } f^2 = ya^2 \tag{19}$$

Then from (19),

$$2x^2 + y^2 = 32 \tag{20}$$

Assuming that $x = 2$, then $y = 4.898979486$ and hence the values of x and y are real and positive as required for rotatability.

$$\text{Thus } c_2^4 = 4a^4, \text{ and } f^4 = 24a^4 \tag{21}$$

Substituting (21) in (16) and (17) gives;

$$\lambda_4 = 1.723577236a^4 \text{ and } \lambda_2 = 1.394922827a^2 \tag{22}$$

Substituting (22) to (5) gives,

$$\frac{\lambda_4}{\lambda_2^2} > \frac{k}{k+2} = 0.885789212 > 0.6666 \tag{14}$$

The conditions given in 3, 4 and 5 are satisfied, hence s_2 forms a five level rotatable design for two factors.

2.3 Robustness for Missing Data for Rotatable Designs.

A design is said to be disconnected if it fails to satisfy both the moments and the non-singularity conditions for rotatability. The minimum number of observations that can result to a disconnected eventual design is referred to a break down number (BD). A useful measure when planning an experiment to reduce or even prevent the possibility of a disconnected eventual design is the concept of minimum number of observations that a planned design is required to lose for the corresponding eventual design to be not rotatable. Thus planned designs with high breakdown numbers are advantageous on grounds of robustness for missing data. In this study, the two designs; D_1 and D_2 which are second order rotatable designs in four and five dimensions respectively are evaluated on their robustness for missing data.

3.0 RESULTS AND DISCUSSION

Theorem 1: The thirty two points second order rotatable design in four dimensions can lose a maximum of 14 experimental runs and still attain both the moments and non-singularity conditions for second order rotatability.

Proof

The set of points given in (6) with 14 missing experimental runs yields the moment conditions given by;

$$\sum_{u=1}^{32} x_{iu}^2 = 2a^2 + 2c_1^2 + 2c_2^2 = 32\lambda_2 \tag{15}$$

$$\sum_{u=1}^{32} x_{iu}^4 = 2a^4 + 2c_1^4 + 2c_2^4 = 96\lambda_4 \tag{16}$$

$$\sum_{u=1}^{32} x_{iu}^4 - 3 \sum_{u=1}^{32} x_{iu}^2 x_{ju}^2 = c_1^4 + c_2^4 - 4a^4 = 0 \tag{17}$$

$$\text{Let } c_1^2 = xa^2, \text{ and } c_2^2 = ya^2 \tag{18}$$

Then from (18),

$$x^2 + y^2 = 4 \tag{19}$$

Assuming that $x = 1$, then $y = 1.732050808$ and hence the values of x and y are real and positive as required for rotatability.

$$\text{Thus } c_1^4 = a^4, \text{ and } c_2^4 = 3a^4 \tag{20}$$

Substituting (12) in (15) and (16) gives;

$$\lambda_4 = 0.104166666a^4 \text{ and } \lambda_2 = 0.233253175a^2 \tag{21}$$

Substituting (21) to (5) gives,

$$\frac{\lambda_4}{\lambda_2^2} > \frac{k}{k+2} = 1.914580521 > 0.6666 \tag{22}$$

The conditions given in 3, 4 and 5 are satisfied, hence a four level rotatable design for two factors is realized from the thirty two design with fourteen missing runs. It is clear that the thirty two points SORD in four dimensions has a breakdown number of fourteen (BD = 14), thus if the design is compared with other SORDs in four dimensions with $BD < 14$, the thirty two points SORD is considered to be more robust for missing data.

Theorem 2: The eighty two points second order rotatable design in five dimensions can lose a maximum of 30 experimental runs and still attain both the moments and non-singularity conditions for second order rotatability.

Proof

The set of points given in (15) with 30 missing experimental runs yields the moment conditions given by;

$$\sum_{u=1}^{82} x_{iu}^2 = 2a^2 + 16f^2 + 2c^2 = 82\lambda_2 \tag{23}$$

$$\sum_{u=1}^{82} x_{iu}^4 = 2a^4 + 16f^4 + 2c^4 = 246\lambda_4 \tag{24}$$

$$\sum_{u=1}^{82} x_{iu}^4 - 3 \sum_{u=1}^{32} x_{iu}^2 x_{ju}^2 = 4f^4 + 2c^4 - 4a^4 = 0 \tag{25}$$

$$\text{Let } f^2 = xa^2, \text{ and } c^2 = ya^2 \tag{26}$$

Substituting (26) on (25) gives,

$$2x^2 + y^2 = 2 \tag{27}$$

Assuming that $x = 1$, then $y = 1$ and hence the values of x and y are real and positive as required for rotatability.

$$\text{Thus } f^4 = a^4, \text{ and } c^4 = a^4 \tag{28}$$

Substituting (28) in (23) and (24) gives;

$$\lambda_4 = 0.081300813a^4 \text{ and } \lambda_2 = 0.243902439a^2 \tag{29}$$

Substituting (29) to (5) gives,

$$\frac{\lambda_4}{\lambda_2^2} > \frac{k}{k+2} = 1.36666667 > 0.6666 \tag{30}$$

The conditions given in 3, 4 and 5 are satisfied, hence a five level rotatable design for two factors is realized from the thirty two design with thirty missing runs. It is clear that the eighty two points SORD in five dimensions has a breakdown number of thirty (BD = 30), thus if the design is compared with other SORDs in five dimensions with $BD < 30$, the eighty points SORD is considered to be more robust for missing data.

4.0 CONCLUSION

In this paper, higher level rotatable designs were considered for two factors; the thirty two points SORD in four dimensions, and eighty two points SORD in five dimensions were considered for robustness for missing data. The thirty two points SORD was found to lose a maximum of fourteen points and still attain both the moments and non-singularity conditions for SORD while the eighty two points lost a maximum of thirty points and still attained the conditions for rotatability. The breakdown numbers form the basis for missing data robustness where designs with higher BDs are considered to be more robust for missing data.

REFERENCES

- [1] Godolphin, P., & Godolphin, E. (2019). Robustness of crossover trials against subject drop-out- examples of perpetually connected designs. *Statistical Methods in Medical Research*, 28(3), 788-800.
- [2] Godolphin, J. (2019). Connectivity of Incomplete Block Designs. *Quality and reliability Engineering International*, 35(5), 1299-1287.
- [3] Box, G. E., & Draper, N. (1975). Robust designs. *Biometrika*, 62(2), 347-352.
- [4] Herzberg, A., & Andrews, D. (1976). Some considerations in the optimal design of experiments in non-optimal solutions. *journal of royal statistical society: series B (Methodological)*, 38(3), 284-289.
- [5] Molenbergs, G., & Kenward, G. (2007). *Missing data in clinical studies*. Chichester, UK: Wiley.
- [6] Rubin, D. (1976). Inference and missing data. *Biometrika*, 63(3), 581-592.
- [7] Tanco, M., & Elizabeth, V. (2013). Robustness of three-level response surface designs against missing data. *IIE Transactions*, 45(5), 544-553.
- [8] Whittinghill, D. (1998). A note on the robustness of Box and Benken designs to the unavailability of data. *Metrika*, 48(1), 49-52.