

A Soft-Drink Experiment using Replicated Full Factorial (RFF) Design

Arfa Maqsood
Department of Statistics
University of Karachi

Rafia Shafi
Department of Statistics
University of Karachi

ABSTRACT

An experiment using factorial design allows one to examine simultaneously the effects of multi-independent factors and their degree of interactions. In this paper, a replicated full-factorial (RFF) design is run to determine the factors that have significant impact on the response of soft drink experiment. We consider the four factors each with two levels and observe the impact of these factors on the volume of foam of soft drink when pour into a glass. Our investigation finds that the significant main effects are soft drink type (A), amount of soft drink (C), and diameter of glass (D), whereas the significant two-factor interactions B (temperature) with C, and C with D. Furthermore, to support our analysis we do modeling using regression approach based on significant factors and interactions. From the analysis of model adequacy, it is observed that the assumptions underlying the estimated model are appropriate.

Keywords

Replicated full-factorial design, Soft drink, Interactions, Regression analysis, Model adequacy

1. INTRODUCTION

The most commonly used type of multifactor experiment is the factorial experiment. In a factorial experiment the treatments consist of combinations of two or more factors each at two or more levels. The number of treatment combinations in such experiment is the product of number of levels of all factors (Petersen [1]). A two-level factorial design in which each factor is studied at two levels (high and low; present and absent). These designs are widely used in factor-screening experiment where aim is to examine the effect of each factor and the joint effects of some specified factors on the response (Hedayat and Pesotan [2]). The effect of a factor is defined to be the change in response produced by a change in the level of the other factor. Similarly the joint effect of two factors is defined as the change in mean response produced by the effect of one factor at different levels of other factor (see Montgomery [3]).

A large number of empirical studies are available on the use of factorial design and its family which includes unreplicated full-factorial design, replicated full-factorial design, fractional factorial design and split-plot design. The applications of these designs can be found in quite diversified fields particularly in biological sciences. The reader might get benefit from the huge literature available on the topic. For instance a two-level factorial design has been employed to investigate the effect of temperature and altitude on current flow in an integrated circuit by Hicks [4]. Similarly, other valuable contributions towards empirical studies include Pais et al. [5], Yasmeen and Siddiqui [6], Kumar et al. [7], Ahmed and Suboohi [8], Chen et al. [9], and Collin [10].

The paper is organised as follows. Following this introduction a brief foreword of replicated full-factorial design is given in

section 2. Some important preliminaries regarding soft drink experiment are discussed in section 3.1. Section 3.2 discusses the results of soft drink experiment in detail. A regression model is also estimated to validate model adequacy in the same section. The last section of the paper recapitulate the conclusion.

2. REPLICATED FULL FACTORIAL (RFF) EXPERIMENT

A full-factorial (FF) design with k factors each at two levels needs 2^k runs. When the number of factors increase, the size of the experiment ultimately become very large. For example an experiment with 10 factors and two levels each produces 1024 runs. At some point this may becomes infeasible most probably due to high cost or insufficient resources. In such cases, fractional factorial design may be used by running to run any significant fraction of full design. See Chen et al. [9], Addleman [11], and Montgomery [3] for detailed description on fractional factorial design.

A single replicate of 2^k design is called an unreplicated factorial (URF) design in which each treatment combinations appears only once. The main drawback of this design is that it does not provide the estimate of error due to unavailability of degree of freedom for error. To analyze such designs, one approach is based on principle of sparsity of effects; that is, most processes are dominated by some main effects and low-order interactions. While high-order interactions are negligible and thus their mean squares are combined to estimate the error. On the contrary, a replicated full-factorial (RFF) design with $n2^k$ observations where each treatment is replicated n times. All the k main effects and $2^k - (k+1)$ interactions are estimated and tested for their significance with no restrictions. The estimate of error is easily obtained in RFF designs with $(n-1)2^k$ degrees of freedom without any abandonment of high-order interactions.

We perform an experiment of soft drink using RFF design with four factors soft drink type, temperature, amount of soft drink, and diameter of glass. More details about the experiment are elaborated in the following section.

3. A SOFT DRINK EXPERIMENT

3.1 Preliminaries

We investigate the effect of different factors on the volume of soft drink foam. The amount of foam can be affected by several factors. In the current study, first we consider some factors as fixed through out the experiment. These factors are listed in table 1. On the other hand we consider the four factors as main factors seemed to be more influential in producing the amount of foam. The factors each with two levels are given in table 2. These factors are represented by the uppercase letters A, B, C, and D.

Table 1: Fixed factors in Soft drink experiment

| Fixed Factors | Levels |
|--------------------------------|------------------|
| Material of glass | Plastic |
| Soft drink | Bottle of 250 ml |
| Place of performing experiment | Room |
| Observer | Author |

A two replicate full-factorial design with four factors needs $2 \times 2^4 = 32$ runs which show 16 different treatment combinations each with two replicates. These treatment combinations are written in standard order as (1), a, b, ab, c, ac, bc, abc, d, ad, bd, abd, cd, acd, bcd, abcd. The plus (+) and minus (-) signs are associated with each combination to represent the high and low levels of the factors. For example, the treatment combination 'a' means the high level of factor A with all other factors at low level. Sometimes a different

notation is used to show the levels of factors as symbols of 0 and 1. The experiment is run in a completely random order and the data is presented in table 3. The table with the signs of each factor is called design matrix. The response variable is the volume of foam that produced when pouring soft drink into glass. For each run we measure the height of foam and calculate the volume using the formula as given below;

$$Volume = \pi r^2 h$$

Where h is the height of foam and $r = d / 2$, d is the diameter of glass.

The sixteen treatment combinations are well displayed geometrically as shown in figure 1. The two cubes are drawn one at low level of factor D and the other at high level of factor D. The total of two observations for each treatment combinations are shown at each corner of cubes.

Table 2: Factors and their levels in Soft-drink experiment

| Letter showing the factors | Factors | Levels | |
|----------------------------|----------------------|---------|----------|
| | | Low (-) | High (+) |
| A | Soft Drink | Pepsi | 7up |
| B | Temperature | Room | Cold |
| C | Amount of Soft drink | 40 ml | 60 ml |
| D | Diameter of glass | 3.6 cm | 4.8 cm |

Table 3: Response for RFF in Soft-drink experiment

| Run Number | Factors | | | | Treatment Combinations | Volume | |
|------------|---------|---|---|---|------------------------|-------------|--------------|
| | A | B | C | D | | Replicate I | Replicate II |
| 1 | - | - | - | - | (1) | 50.91 | 41.75 |
| 2 | + | - | - | - | a | 45.82 | 40.73 |
| 3 | - | + | - | - | b | 32.59 | 35.64 |
| 4 | + | + | - | - | ab | 33.60 | 20.37 |
| 5 | - | - | + | - | c | 66.19 | 61.10 |
| 6 | + | - | + | - | ac | 61.10 | 66.19 |
| 7 | - | + | + | - | bc | 88.59 | 92.66 |
| 8 | + | + | + | - | abc | 65.17 | 58.04 |
| 9 | - | - | - | + | d | 70.60 | 74.22 |
| 10 | + | - | - | + | ad | 63.36 | 48.88 |
| 11 | - | + | - | + | bd | 63.36 | 45.26 |
| 12 | + | + | - | + | abd | 54.31 | 47.07 |
| 13 | - | - | + | + | cd | 63.36 | 72.41 |
| 14 | + | - | + | + | acd | 65.17 | 61.55 |
| 15 | - | + | + | + | bcd | 57.93 | 57.93 |
| 16 | + | + | + | + | abcd | 68.79 | 59.74 |

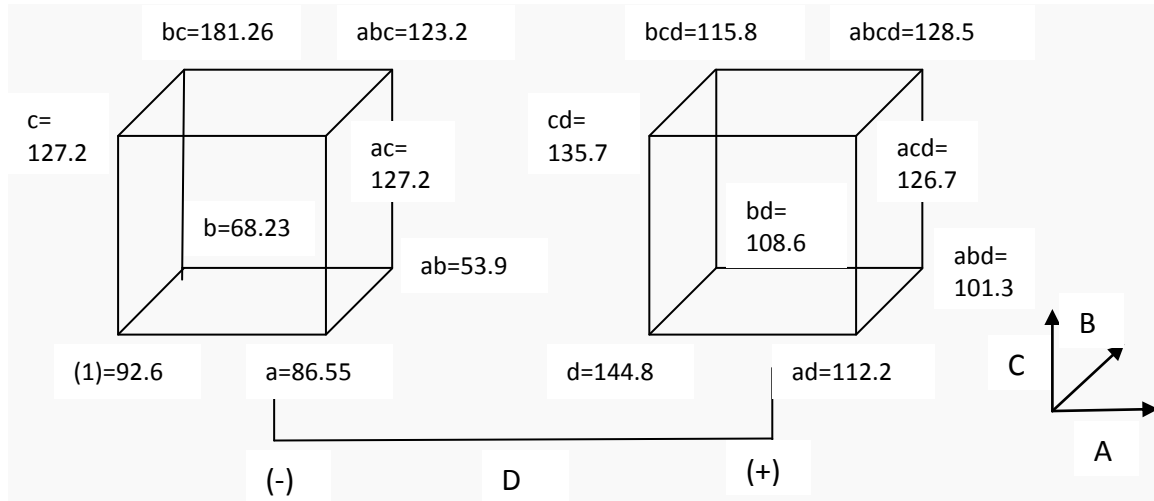


Fig 1: Geometric view of 2^4 design

3.2 Results and Discussion

The analysis of replicated full-factorial design for soft drink experiment is presented in this section. First, we estimate the effects of factors using the contrast formula as given below;

$$Effect_{(factor)} = \frac{Contrast}{n \cdot 2^{k-1}}$$

where contrast shows the total effect of a factor. The estimated standard error of the effect is given by;

$$SE(effect) = 2 \sqrt{\frac{MSE}{n \cdot 2^k}}$$

Where MSE is the mean square error obtained from the analysis of variance for factors. It is now possible to construct the confidence intervals for each effect using the formula given as $Estimate(effects) \pm t_{\alpha/2, N-p} \cdot SE(effects)$. This expression gives the $(1-\alpha)100\%$ confidence intervals of each factor effects. The standard errors of all factor effects are equal because the design matrix is orthogonal (see Montgomery [3]). The degrees of freedom on t are the number of degrees of freedom for error. The results on effect estimates along with their 95% confidence intervals are summarized in table 4. The effects corresponding to the factors A, C, D, BC, CD, ABD, ACD, and BCD have larger estimates showing more probably the important factors in soft drink experiment. This table indicates that for above significant factors, the approximate 95% confidence intervals do not include zero. The normal probability plot of these effects is shown in figure 2. All the effects that lie along the line are negligible, whereas the large effects are far from the line. The important factors from NPP include the main effects A, C, D, and the interaction effects BC, and CD. The main effects of A, B, C, and D are plotted in figure 3(a). This can be seen that the effects of C and D are positive while A and B have negative effects. Therefore, in order to maximize the volume of foam of soft drink one has to run the factors C, and D at high level. It is also necessary to examine the significant interactions. The factorial experiments are usually dominated by some main effects and low-order interactions. Most of the high-order interactions do not have significant contributions and thus are negligible. For this reason, we plotted all possible two-factor interactions in figure 3(b). The parallel lines in this

plot do not show the significant effect of one factor when the other factor moving from its low level to high level. However, a large effect can be seen in BC and CD interactions. The CD interaction indicates that amount of soft drink C has large effect at low level of diameter of glass D but a smaller effect at high level of D. The factor B is not important as a main effect but when combines with other factors, it has some remarkable significance. As we can observe from BC interaction which accounts for a 7.68 percent of the total variation (see table 4) in data. Based on sum of squares of each factor the percent contributions are calculated by dividing each sum of square to the total sum of squares. Last column of table 4 gives the percent contributions of each factor. 49.7 percent of the total variation are being explained due to main effects, 29.6 percent due to two-factor interactions, 11.7 percent due to three-factor interactions, and only 0.8 percent due to four-factor interactions.

The respective sum of square is defined by;

$$SS_{(factor)} = \frac{(Contrast)^2}{n \cdot 2^k}$$

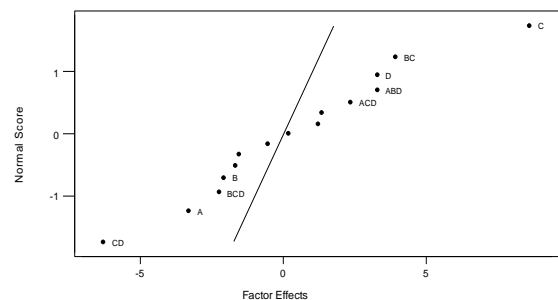


Fig 2: Normal Probability Plot (NPP) of the Effects

There are fifteen degrees of freedom distributed among sixteen treatment combinations in 2^4 designs. Four degrees of freedom are associated with the main effects of A, B, C, and D. six degrees of freedom are associated with two-factor interactions, four with three-factor interactions, and one with four factor interaction. The complete ANOVA is summarized in table 5. The ANOVA confirms our result drawn by the effect estimates. The effects corresponding to main effects A,C, D, and interactions BC, CD, and ABD all appear to be significant at 1% level.

Table 4: Estimated average effects and confidence intervals for RFF design in soft-drink experiment

| Factors | Effect Estimates | 95 % Confidence Intervals | Percent Contributions |
|---------|------------------|---------------------------|-----------------------|
| A | -7.15 | -11.73 to -2.57* | 5.43 |
| B | -4.51 | -9.09 to 0.08 | 2.16 |
| C | 18.60 | 14.02 to 23.19* | 36.78 |
| D | 7.11 | 2.52 to 11.69* | 5.37 |
| AB | -1.18 | -5.77 to 3.40 | 0.15 |
| AC | 0.37 | -4.21 to 4.96 | 0.01 |
| AD | 2.65 | -1.93 to 7.23 | 0.75 |
| BC | 8.50 | 3.92 to 13.09* | 7.68 |
| BD | -3.61 | -8.20 to 0.97 | 1.39 |
| CD | -13.60 | -18.18 to -9.02* | 19.66 |
| ABC | -3.33 | -7.91 to 1.25 | 1.18 |
| ABD | 7.09 | 2.51 to 11.67* | 5.34 |
| ACD | 5.08 | 0.50 to 9.66* | 2.74 |
| BCD | -4.86 | -9.44 to -0.28* | 2.51 |
| ABCD | 2.90 | -1.68 to 7.49 | 0.90 |

* Significant at 5% as these intervals do not include zero

Table 5: ANOVA table for RFF design in soft-drink experiment

| Sources of Variation | Degrees of Freedom | Sum of Squares | Mean Squares | F-Ratio | p-value |
|----------------------|--------------------|----------------|--------------|---------|---------|
| A | 1 | 409.09 | 409.09 | 10.94 | 0.004* |
| B | 1 | 162.43 | 162.43 | 4.34 | 0.054 |
| C | 1 | 2768.57 | 2768.57 | 74.05 | 0.000* |
| D | 1 | 403.96 | 403.96 | 10.80 | 0.005* |
| AB | 1 | 11.20 | 11.20 | 0.30 | 0.592 |
| AC | 1 | 1.11 | 1.11 | 0.03 | 0.865 |
| AD | 1 | 56.22 | 56.22 | 1.50 | 0.238 |
| BC | 1 | 578.41 | 578.41 | 15.47 | 0.001* |
| BD | 1 | 104.49 | 104.49 | 2.79 | 0.114 |
| CD | 1 | 1479.90 | 1479.90 | 39.58 | 0.000* |
| ABC | 1 | 88.82 | 88.82 | 2.38 | 0.143 |
| ABD | 1 | 402.26 | 402.26 | 10.76 | 0.005* |
| ACD | 1 | 206.61 | 206.61 | 5.53 | 0.032 |
| BCD | 1 | 188.80 | 188.80 | 5.05 | 0.039 |
| ABCD | 1 | 67.47 | 67.47 | 1.80 | 0.198 |
| Error | 16 | 598.21 | 37.39 | | |
| Total | 31 | 7527.55 | | | |

* Significant at 1% level

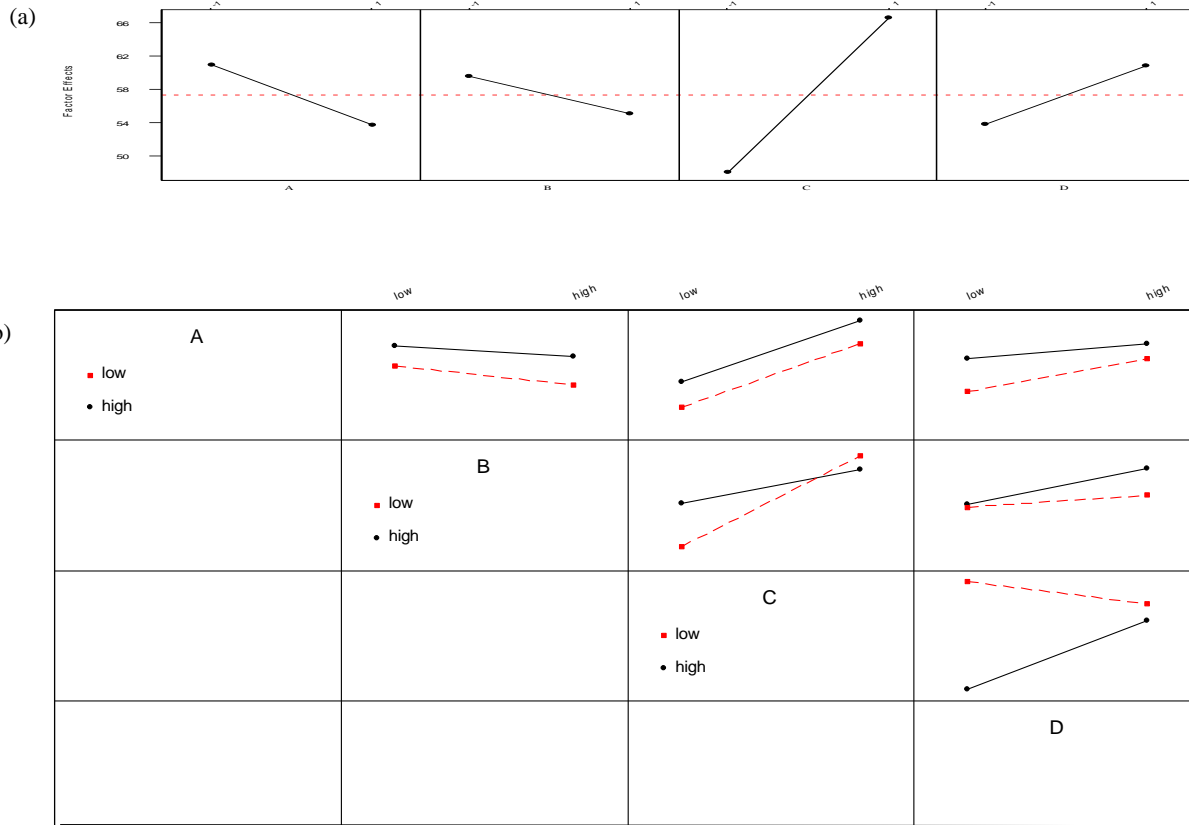


Fig 3: (a) Plot of Main Effects (b) Two-Factor Interaction plots

3.3 The Regression Model and Model Adequacy

In 2^k factorial designs, we use the regression approach to confirm the results of experiments. The regression model for soft drink experiment is based on the significant factors as given below;

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_{23} X_2 X_3 + \beta_{34} X_3 X_4 + \beta_{124} X_1 X_2 X_4 + e$$

Where $X_1, X_2, X_3,$ and X_4 are the coded variables for factors A, B, C, and D respectively. X_i can assume the values +1 or -1 for high or low level of i th factor respectively. β_0 is the grand average of all 32 observations. All other regression coefficients are one-half the corresponding factor effect estimates. Thus the fitted regression model for soft drink experiment is given by;

$$\hat{Y} = 57.32 - 3.57X_1 - 2.25X_2 + 9.3X_3 + 3.55X_4 + 4.25X_2X_3 - 6.8X_3X_4 + 3.54X_1X_2X_4$$

The fitted regression model can further be used to obtain residuals and hence helpful in model adequacy checking. There are some assumptions underlying the use of analysis of variance technique, which must be validated. For this purpose, the primary diagnostic tool is the residual analysis.

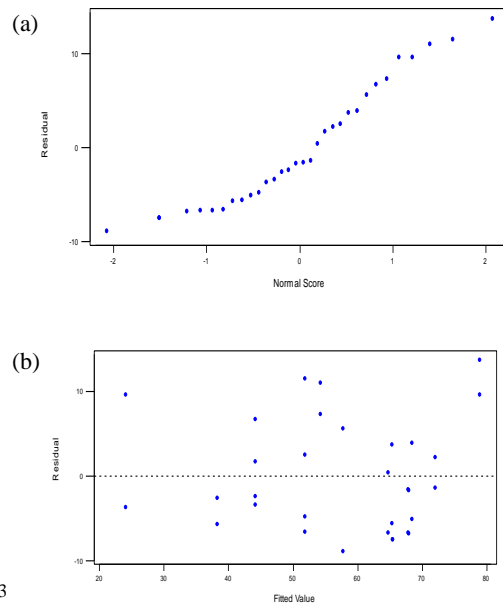


Fig 4: (a) Normal Probability Plot (b) Residuals versus Predicted values

The residuals are obtained as the difference between the observed and predicted volume of foam produced as $\hat{e} = Y - \hat{Y}$. A normal probability plot of residuals is shown in figure 4(a). All the points lie reasonably along a straight line, which support the assumption of normality. We plot residuals versus predicted volume in figure 4(b). This plot is also satisfactory to the assumption of equality of variance.

4. CONCLUSION

In this paper, we applied a replicated full factorial (RFF) design to the soft drink experiment. We considered the four factors soft drink type (A), temperature (B), amount of soft drink (C), and diameter of the glass (D). With two levels of each factor, we decided to run a 2^4 design with two replicates. The volume of foam produced was taken as the response variable and we investigated for the significant factors that can maximize the volume.

Our analysis showed that the main effects of factor soft drink type (A), amount of soft drink (C), and diameter of glass (D) are the significant. We found the negative effect of factor A, i.e., the type of soft drink “pepsi” makes larger amount of foam as compare to the type of soft drink “7up”. To maximize the foam of soft drink, one has to run the factors C, and D at high levels. Their positive effects indicate that pouring larger amount of soft drink (C at high level) into a glass of bigger diameter (D at high level) increases the amount of foam, which in turns increases the volume. However, the main effect temperature (B) is appeared to be insignificant.

On the other hand, the interaction between temperature (B) and amount of soft drink (C), and the interaction between amount of soft drink (C) and diameter of glass (D) are the significant two-factor interactions. There is a large effect in average response of factor C when taken at low level of factor D, and a little effect of factor C at high level of D. With these observations, our analysis suggests that the experimenter has to use the low levels of soft drink type (pepsi), with high level of soft drink amount (60 ml), and high level of glass diameter (4.8 cm) to maximize the foam volume.

5. ACKNOWLEDGEMENT

This research is partially supported by Dean Grant from University of Karachi, Pakistan. The authors wish to thank the Department of Statistics, University of Karachi for providing the computing and research facilities.

6. REFERENCES

[1] Petersen, R. G. 1985, Design and Analysis of Experiments, Marcel Dekker, Inc., New York.

[2] Hedayat, A. S., and Pesotan, H. 1992, Two-Level Factorial Designs for Main Effects and Selected Two-Factor Interactions, *Statistica Sinica*, 2: 453-464.

[3] Montgomery, D. C. 2009, Design and Analysis of Experiments, 7th Ed., John Wiley & Sons, Inc.

[4] Hicks, C. R. 1999, Fundamental Concepts in the Design of Experiments, 5th Ed., Oxford University Press.

[5] Pais, M. C., Peretta, I. S., Yamanaka, K., and Pinto, E. R. 2014, Factorial Design Analysis Applied to the Performance of Parallel Evolutionary Algorithms, *Journal of the Brazilian Computer Society*, 20(6): 1-17.

[6] Yasmeen, F., and Siddiqui, A. J. 2014, A Cake Baking Experiment Using Fractional Factorial Split Plot (FFSP) Design, *International Journal of Recent Development in Engineering and Technology*, 2(2): 7-12.

[7] Kumar, L., Reddy, M. S., Managuli, R. S., and Girish, P. K. 2015, Full Factorial Design For Optimization, Development and Validation of HPLC Method to Determine Valsartan in Nanoparticles, *Saudi Pharmaceutical Journal*, 23: 549-555.

[8] Ahmed, E., and Suboohi, A. 1993, On Determining Sparse Factors for Cake Baking Using Unreplicated Fractional Factorial Experiments, *Quality Engineering*, 5(4): 571-581.

[9] Chen, J., Sun, D. X., and Wu, C. F. J. 1993, A Catalog of Two-level and Three-level Fractional Factorial Designs with Small Runs, *International Statistical Review*, 61: 39-56.

[10] Collin, L. M., Dziak, J. J., and Li, R. 2009, Design of Experiments with Multiple Independent Variables: A Resource Management Perspective on Complete and Reduced Factorial Designs, *Psychol Methods*, 14(3): 202-224.

[11] Addelman, S. 1963, Techniques for Constructing Fractional Replicate Plans, *Journal of American Statistical Association*, 58: 45-71.