A C T A   U N I V E R S I T A T I S   L O D Z I E N S I S
FOLIA OECONOMICA 285, 2013

Małgorzata Szerszunowicz *

SOME CONSTRUCTION OF FACTORIAL DESIGNS

Abstract. Design of experiments methods are an important tool to improve production processes using statistical methods. Designing experiments allows to set factors of the production process as well as describe the influence of factors on the results. Furthermore, the design methods of experiments help to improve economic results of the analyzed process.

The aim of this paper is the issue of choosing the optimal layout of experiments when the experimenter, because of the cost or conditions, has no possibility to implement the completion of the design of experiments. The method to determine successively points of a design will be suggested to carry out an experiment. Finally, implementation of the mentioned method will be presented for selected factorial designs.

Key words: Design of experiments, Factorial design, Sequence of experiments.

I. INTRODUCTION

Experimental design methods play a major role in of the manufacturing process. Thanks to probability and statistical design methods of experiment, it is possible to improve a manufacturing process.

The aim of this article is to present the construction of a factorial design of the experiment carried out in cost-limited restricted overall conditions and to determine the following points of the experimental area, which the variance of the response surface estimator has the lowest value.

II. THE BASICS OF EXPERIMENTAL DESIGN

The base of an experimental design is to use suitable rules to realize following experiments. The course of the preceding stage of the manufacturing process may be presented in the following scheme (Montgomery, 1997):

* M.Sc., Department of Statistics, University of Economics in Katowice.

– recognition and statement of a problem to determine all the aspects, circumstances and potential objectives of an experiment
– choice of factors and description of their levels, ranges over which these factors will be varied as well as determination of a possibility to take them into account in the experiment
– selection of response variable which really provides useful information about the process under study
– choice of a proper experimental design to determine a number of experiments and possible randomization restrictions
– performance of the experiment
– data analysis using statistical methods
– conclusions and recommendations for a described process following the data analysis

The experiment is a sequence of *n* experimental trials, where an experimental trial is a single act of obtaining variable values of a described Y, when each of the factors $X_1, X_2, \ldots, X_m$ is fixed. May the $\mathbf{X_1}, \mathbf{X_2}, \ldots, \mathbf{X_m}$ represent sets of all possible values of factors $X_1, X_2, \ldots, X_m$, then the experimental area is a set of points $\mathbf{x} = (x_1, x_2, \ldots, x_m)$, where $x_i \in \mathbf{X}_i, i = 1, 2, \ldots, m$. A set of pairs (Wawrzynek,1993) $P_n = \{\mathbf{x}_j, p_j\}_{j=1}^n$ defines the design of the experiment with *n* experimental trials, where $\mathbf{x}_j = (x_{1j}, x_{2j}, \ldots, x_{mj})$ and $p_j = \frac{n_j}{n}$, where n_j is a number of experimental trials in the point $\mathbf{x}_j$ of the experimental area, moreover $\sum_{j=1}^{n} n_j = n$ and $\sum_{j=1}^{n} p_j = 1$ for $j = 1, 2, \ldots, n$.

Usually an experimental research involves analyzing the influence of some number of non-random factors $X_1, X_2, \ldots, X_m$ on the result variable *Y*, but on the random factors may also have an impact on the starting variable *Y*. This correlation may be presented with the following statistical model (Wawrzynek ,2009)

$$Y(X_1, X_2, \ldots, X_m) = y(X_1, X_2, \ldots, X_m) + \varepsilon \tag{1}$$

where $EY(X_1, X_2, \ldots, X_m) = y(X_1, X_2, \ldots, X_m)$, $E\varepsilon = 0$ and $V\varepsilon = \sigma^2$, where σ^2 is a constant value independent from particular factor values. The object of the statistical research will be a function $y(x_1, x_2, \ldots, x_m)$ called a response surface. Arguments of the response surface are *m* realization of the non-random factors $X_1, X_2, \ldots, X_m$.

The model (1) can be presented as a general linear model $\mathbf{Y} = \mathbf{F}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ where

$$\mathbf{Y}^{\mathbf{T}} = \left(Y_1 Y_2 \ldots Y_n\right) \tag{2}$$

$$\boldsymbol{\varepsilon}^{\mathbf{T}} = \left(\varepsilon_1 \varepsilon_2 \ldots \varepsilon_n\right) \tag{3}$$

$$\boldsymbol{\beta}^{\mathbf{T}} = \left[\beta_1 \beta_2 \ldots \beta_k\right] \tag{4}$$

$$\mathbf{f}^{\mathbf{T}}(\mathbf{x}) = \left(f_1(\mathbf{x}) f_2(\mathbf{x}) \ldots f_k(\mathbf{x})\right) \tag{5}$$

$$\mathbf{F} = \begin{bmatrix} f_1(\mathbf{x}_1) & \ldots & f_k(\mathbf{x}_1) \\ \vdots & \ddots & \vdots \\ f_1(\mathbf{x}_n) & \ldots & f_k(\mathbf{x}_n) \end{bmatrix} = \left[\mathbf{f}(\mathbf{x}_1)\ \mathbf{f}(\mathbf{x}_2) \ldots \mathbf{f}(\mathbf{x}_n)\right]^{\mathbf{T}} \tag{6}$$

and $f_i(\mathbf{x}_j) \equiv x_{ij}$, for $i = 1,2,\ldots k, j = 1,2,\ldots,n$ (Wawrzynek, 2009). Then the response surface is defined with the equation $\mathbf{y} = \mathbf{F}\boldsymbol{\beta}$. In order to estimate the parameters of the response surface function one usually uses the method of the least squares (Aczel, 2000). In this way, we get the response surface estimator $\tilde{\mathbf{y}} = \mathbf{F}\tilde{\boldsymbol{\beta}}$ which variance is described with the equation

$$\mathbf{V}\tilde{\mathbf{y}}(\mathbf{x}) = \sigma^2 \mathbf{f}^{\mathbf{T}}(\mathbf{x})(\mathbf{F}^{\mathbf{T}}\mathbf{F})^{-1}\mathbf{f}(\mathbf{x}). \tag{7}$$

It is noted that value of the variance depends only on the choice of a suitable design of the experiment and exactly from the value of the square matrix $\mathbf{F}^{\mathbf{T}}\mathbf{F}$ elements.

Usually in the literature (Miszczak *et al.*,2008, Kończak, 2007) on classical designs of experiments the response surface functions, which consider the presented below interaction of factors, are described

$$y(\mathrm{x}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_m x_m + \beta_{12} x_1 x_2 + \ldots + \beta_{12\ldots m} x_1 x_2 \ldots x_m. \tag{8}$$

In justified cases the experimenter may omit the interaction of factors then the response surface function is stated with the following formula

$$y(\mathbf{x}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_m x_m \tag{9}$$

or

$$y(\mathbf{x}) = \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_m x_m . \tag{10}$$

Estimation of the above response surface function is done through fulfilling the experiment for m factors with n_i levels each, then the experiment involves $n = \prod_{i=1}^{m} n_i$ experimental trials.

III. PROPOSED CONSTRUCTION OF A FACTORIAL DESIGN OF THE EXPERIMENT

The most used in practice experimental designs are full and fractional factorial designs of experiments. The full-factorial designs of experiments involve carrying out all possible experimental trials, what with a great number of factors directly involves the extension of manufacturing process and the increase overall costs. One of the methods to limit the number of being fulfilled experimental trials in order to estimate the parameters of the response surface functions is to use fractional designs of factorial experiments which include the interaction of particular factors. Unfortunately not always costs or conditions of carrying the manufacturing process allow to carry out on experiment according to the full or fractional design.

Take into consideration the following design of the experiment

$$P_n = \begin{Bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \ldots & \mathbf{x}_n \\ \frac{1}{n} & \frac{1}{n} & \ldots & \frac{1}{n} \end{Bmatrix}. \tag{11}$$

The design of P_n considers n experimental trials with the same weights $\frac{1}{n}$. Supposing costs of a studied process allow the experimenter to fulfil the smallest number of experimental trials which allow to estimate the parameters of response surface functions omitting the interaction of factors. It is worth noting that the use the designs of experiments is pointless. We should consider in what way we are to construct the design with the smallest number of experimental trials as well as in what way next points of the experimental area ought to obtain the most precise results of the experiment.

The algorithm below shows the construction of the experiment design that considers the mentioned circumstance of the experiment.

1. We determine the minimal number of experimental trails $n_{\min}$ in order to estimate all the parameters of the studied response surface function.

2. From the experimental area we choose $n_{\min}$ points for which the inequality $\det(\mathrm{F}^{\mathrm{T}}\mathrm{F}) > 0$ works. The established design for the minimal number of experimental trials is defined as

$$P_{n_{\min}} = \begin{Bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \dots & \mathbf{x}_{n_{\min}} \\ \dfrac{1}{n_{\min}} & \dfrac{1}{n_{\min}} & \dots & \dfrac{1}{n_{\min}} \end{Bmatrix}. \tag{12}$$

3. Taking into account the chosen experimental trails, we determine the form of the variance of the response surface function estimator (7).

4. We calculate the value of variance of response surface estimator for these points of the experimental area which have not been included in $P_{n_{\min}}$ design. In justified case a value of the variance of response surface is calculated for all the points of the experimental area.

5. We find a point of the experimental area in which a value of the variance of response surface is the lowest.

6. The following experimental trial in the fulfilled experiment is carried out in defined, in the previous step, a point of the experimental area.

We create a design of an experiment presented in the form below

$$P_{n_{\min}+1} = \begin{Bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \dots & \mathbf{x}_{n_{\min}} & \mathbf{x}_{n_{\min}+1} \\ \dfrac{1}{n_{\min}+1} & \dfrac{1}{n_{\min}+1} & \dots & \dfrac{1}{n_{\min}+1} & \dfrac{1}{n_{\min}+1} \end{Bmatrix}. \tag{13}$$

The mentioned construction of the experiment design allows not only to carry out the experiment with the smallest number of experimental trials, but also allows to define the next point of the design in which dispersion values of the response surface function are the lowest.

IV. APPLICATION OF THE SUGGESTED CONSTRUCTION OF DESIGN OF FACTORIAL EXPERIMENT

In this part the application of the suggested construction of the experiment design will be described to estimate the response surface function stated as:

$$y(x) = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 . \tag{14}$$

The classical method of experimental designs allows to estimate the response surface function (14) to use the full factorial design 2^4. This design contains four factors on two levels: the upper and the lower ones. The design of the experiment is presented in the table 1.

Table 1. Full factorial design of experiment 2^4

Nr	X_1	X_2	X_3	X_4
1	+	+	+	+
2	+	+	+	–
3	+	+	–	–
4	+	–	–	–
5	–	–	–	–
6	+	+	–	+
7	+	–	+	+
8	–	+	+	+

Nr	X_1	X_2	X_3	X_4
9	–	+	–	+
10	+	–	+	–
11	–	+	+	–
12	+	–	–	+
13	–	+	–	–
14	–	–	+	–
15	–	–	–	+
16	+	–	–	+

Source: own elaboration.

For the studied response surface function and experiment area which may be presented as vertices of the four-dimensional square:

$$K_4 = \{\, \mathbf{x} = (x_1, \ldots, x_4) : -1 \le x_i \le 1,\ i = 1, \ldots, 4 \,\} , \tag{15}$$

We should construct a design of the experiment with the smallest number of experimental trials and define the next point of the experimental design.

According to the algorithm of the design construction, presented in the previous chapter, in the first step we have to determine the minimal number of experimental trials which allows to estimate the parameters of the response surface function. For the analyzed function (14) we should carry out $n_{\min} = 4$ experimental trials. Then from the experimental area points we choose the next experimental trials in order to match $\mathbf{det}(\mathbf{F}^{\mathrm{T}}\mathbf{F}) > 0$. The matching points are, among the ones identified with the experimental trials number 9, 11, 13 and 14 from the table 1. In this way the experimental design with the minimal number of experimental trials has a form of:

$$P_4 = \begin{Bmatrix} (-1,1,-1,1) & (-1,1,1,-1) & (-1,1,-1,-1) & (-1,-1,1,-1) \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{Bmatrix}. \tag{16}$$

The matrix of this design is

$$\mathbf{F} = \begin{bmatrix} -1 & 1 & -1 & 1 \\ -1 & 1 & 1 & -1 \\ -1 & 1 & -1 & -1 \\ -1 & -1 & 1 & -1 \end{bmatrix}. \tag{17}$$

Then the variance of the response surface estimator is presented with the following formula

$$\begin{aligned} &\mathrm{V}\tilde{\mathrm{y}}(\mathrm{x}) = \sigma^2 \mathrm{f}^{\mathrm{T}}(\mathrm{x})(\mathrm{F}^{\mathrm{T}}\mathrm{F})^{-1}\mathrm{f}(\mathrm{x}) = \\ &= \frac{1}{2}\sigma^2\left(x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_1x_2 - x_1x_4 + x_2x_3 + x_3x_4\right) \end{aligned}. \tag{18}$$

Counting the value of a variance (15) for the experimental area points considered in P_4 design of the lowest value, we obtain for the points identified with the experimental trials with number 6, 7, 10 and 16 from the table 1. In this case the experimenter chooses subjectively from the given experimental area points in the following experimental trial of the fulfilled experiment. An exemplary design of the experiment shown in the form:

$$P_5 = \begin{Bmatrix} (-1,1,-1,1) & (-1,1,1,-1) & (-1,1,-1,-1) & (-1,-1,1,-1) & (1,-1,1,1) \\ \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} \end{Bmatrix}. \tag{19}$$

The presented construction of the experimental design allowed to estimate the parameters of the response surface through carrying out 4 accordingly chosen experimental trials and to determine the experimental area point with the lowest variance. As presented, the fulfilled design of the experiment has a significant impact on a number of the experimental trials.

V. SUMMARY

The experimental design methods are an important tool to improve the manufacturing process. The right application of the experimental designs leads to improving the quality of the carried out process and its products. The most commonly used designs of experiments are full and fractional factorial designs. Unfortunately the use of these designs in practice is not always justified because of the circumstance or number of the experimental trials. The construction of the factorial experimental design presented in this article, allows to carry out an experiment with the minimal number of experimental trials what has an influence on overall costs and time of the manufacturing process. The presented algorithm allows tor determine the next point of the experimental area with the most precise results alike.

REFERENCES

Aczel A. D. (2000), *Statystyka w zarządzaniu*, Wydawnictwo Naukowe PWN, Warszawa

Kończak G. (2007), *Metody statystyczne w sterowaniu jakością produkcji*, Wydawnictwo Akademii Ekonomicznej, Katowice

Miszczak W., Ostasiewicz W., Wawrzynek J. (2008), *Projektowanie próby z elementami planowania eksperymentów*, Wydawnictwo Uniwersytetu Ekonomicznego, Wrocław

Montgomery D. C. (1997), *Introduction to statistical quality control*, John Wiley & Sons, Inc., New York

Wawrzynek J. (2009), *Planowanie eksperymentów zorientowane na doskonalenie jakości produktu*, Wydawnictwo Uniwersytetu Ekonomicznego, Wrocław

Wawrzynek J. (1993), *Statystyczne planowanie eksperymentów w zagadnieniach regresji w warunkach małej próby*, Wydawnictwo Akademii Ekonomicznej, Wrocław

Małgorzata Szerszunowicz

O PEWNEJ KONSTRUKCJI PLANÓW EKSPERYMENTÓW CZYNNIKOWYCH

Metody planowania eksperymentów są istotnym narzędziem doskonalenia procesów produkcyjnych wykorzystującym metody probabilisyczne i statystyczne. Planowanie eksperymentów pozwala na odpowiednie ustawienie parametrów procesu produkcyjnego oraz określenie wpływu czynników na jego wyniki. Ponadto metody planowania eksperymentów umożliwiają poprawę ekonomicznych rezultatów badanego procesu.

Przedmiotem niniejszego artykułu jest zagadnienie wyboru optymalnego układu doświadczeń wówczas, gdy eksperymentator, ze względu na koszty bądź warunki, nie ma możliwości realizacji kompletnego planu doświadczeń. Zaproponowana zostanie metoda wyznaczania kolejnych punktów planu, w których należy przeprowadzić doświadczenie. Zastosowania prezentowanej metody przedstawione zostaną dla wybranych planów eksperymentów czynnikowych. Własności metody zostały sprawdzone w analizach symulacyjnych.