M A T H E M A T I C A L   E C O N O M I C S

No. 8(15) 2012

ON THE DESIGN OF EXPERIMENTS CONSIDERING THE DIVISION OF THE EXPERIMENTAL AREA

Małgorzata Szerszunowicz

Abstract. The design of experiments is an important tool to improve production processes that use statistical methods. Designing experiments empowers not only setting properly the parameters of the production process and describing the influence of factors on the results, but also leads to improving the economic results of the analyzed process. The aim of this article is the issue of choosing the appropriate 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 suggested method takes into account the division of the experimental area and uses measures of spatial autocorrelation to determine the design points to carry out an experiment. The implementation of the mentioned method will be presented for selected factorial designs with particular reference to plans used to estimate the non-linear response surface.

Keywords: design of experiments, factorial design, measures of spatial autocorrelation.

JEL Classification: C99.

1. Introduction

The design of experiments is one of the methods of statistical quality control. This method is applied before beginning the manufacturing process in order to limit possible irregularities or deficiencies of the process, which is possible thanks to probabilistic and statistical methods. The design of experiments allows to improve the manufacturing process as well as influence the improvement of its economic aspects.

The aim of this article is to present the construction of a factorial design of experiment which uses spatial autocorrelation measures. The proposed design takes into account the division of the experimental area, which allows to reduce the number of experimental trials and determine the

Małgorzata Szerszunowicz
Department of Statistics, University of Economics in Katowice, 1 Maja Street 50, 40-287 Katowice, Poland.
E-mail: malgorzata.szerszunowicz@ue.katowice.pl

following points of the experimental area for which the variance of the response surface estimator has the lowest value. Due to the fact that the basic model considered in the design of experiment theory is a multivariate linear model, the proposed algorithm of design construction will be used for linear and nonlinear forms of response surface.

2. The basics of experimental design

The basis of 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):

– 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, the ranges over which these factors will be varied as well as the determination of the possibility to take them into account in the experiment;

– selection of a response variable which really provides useful information about the process under study;

– choice of a proper experimental design to determine the number of experiments and possible randomization restrictions;

– performance of the experiment;

– statistical analysis of received results;

– conclusions and recommendations for the described process following the data analysis.

The experiment is a sequence of *n* experimental trials, where the 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 set (Wawrzynek, 2009):

$$D_E = \left\{ \mathrm{x} = (x_1, x_2, \ldots, x_m) : x_i \in \mathrm{X}_i, i = 1, 2, \ldots, m \right\} \tag{1}$$

is an experimental area. A set of pairs (Wawrzynek, 2009)

$$P_n = \{\mathbf{x}_j, p_j\}_{j=1}^{n} \tag{2}$$

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:

$$\sum_{j=1}^{n} n_j = n \text{ and } \sum_{j=1}^{n} p_j = 1 \text{ for } j = 1, 2, \ldots, n.$$

Usually experimental research involves analyzing the influence of some number of non-random factors $X_1, X_2, \ldots, X_m$ on the result of variable *Y* but random factors may also have an impact on the starting variable *Y*. This correlation may be presented with the following statistical model (Miszczak, Ostasiewicz, Wawrzynek, 2008):

$$Y(X_1, X_2, \ldots, X_m) = y(X_1, X_2, \ldots, X_m) + \varepsilon, \tag{3}$$

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 of 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 (3) can be presented as a general linear model $\mathbf{Y} = \mathbf{F}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, where

$$\mathbf{Y}^{\mathbf{T}} = \left(Y_1 Y_2 \ldots Y_n\right), \tag{4}$$

$$\boldsymbol{\varepsilon}^{\mathbf{T}} = \left(\varepsilon_1 \varepsilon_2 \ldots \varepsilon_n\right), \tag{5}$$

$$\boldsymbol{\beta}^{\mathbf{T}} = \left[\beta_1 \beta_2 \ldots \beta_k\right], \tag{6}$$

$$\mathbf{f}^{\mathbf{T}}(\mathbf{x}) = \left(f_1(\mathbf{x}) f_2(\mathbf{x}) \ldots f_k(\mathbf{x})\right), \tag{7}$$

$$\mathbf{F} = \begin{bmatrix} f_1(\mathbf{x}_1) & \ldots & f_k(\mathbf{x}_1) \\ \vdots & \ddots & \vdots \\ f_1(\mathbf{x}_n) & \ldots & f_k(\mathbf{x}_n) \end{bmatrix} = \left[\mathbf{f}(\mathbf{x}_1)\ \mathbf{f}(\mathbf{x}_2) \ldots \mathbf{f}(\mathbf{x}_n)\right]^{\mathbf{T}} \tag{8}$$

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 $\tilde{\mathbf{y}} = \mathbf{F}\tilde{\boldsymbol{\beta}}$, whose variance is described with the following equation:

$$\mathbf{V}\tilde{\mathbf{y}}(\mathbf{x}) = \sigma^2 \mathbf{f}^{\mathbf{T}}(\mathbf{x})(\mathbf{F}^{\mathbf{T}}\mathbf{F})^{-1}\mathbf{f}(\mathbf{x}). \tag{9}$$

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

Usually in the literature (Miszczak, Ostasiewicz, Wawrzynek, 2008; Kończak, 2007) on classical designs of experiments, the response surface functions are described in the following way:

$$y(\mathrm{x}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_m x_m + \beta_{12} x_1 x_2 + \ldots + \beta_{12\ldots m} x_1 x_2 \ldots x_m. \tag{10}$$

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.

In practice, the most commonly used 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 of overall costs. One of the methods to limit the number of the completed 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. Furthermore, the classical design of experiments usually does not allow to estimate the non-linear response surface, which often characterizes the process in the best way.

3. Measures of spatial autocorrelation

Spatial autocorrelation refers to the systematic spatial changes that are observed as clusters of similar value or spatial pattern. Positive spatial autocorrelation means that the objects observed in the considered region are more similar to objects in neighboring regions, while negative autocorrelation means that the objects are varied.

In order to determine the spatial autocorrelation of the considered objects, one uses global and local measures. The global measures are single-valued indicators of the spatial autocorrelation or general similarity between areas. Local measures are calculated for each area and inform about similarities or differences between the neighbouring areas. The most commonly used spatial autocorrelation measures are global and local Moran's statistics.

Global Moran's statistics is defined as follows (Kopczewska, 2011):

$$\mathrm{I} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{S^2 \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}}, \tag{11}$$

where $S^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$, x_i is a variable in areal unit i, $\bar{x}$ is the average of all researched areas, when w_{ij} is an element of a spatial matrix for a weight matrix; where a weight matrix ought to be row-standardized to one. Tests of significance of global Moran's statistics can be based on the theoretical moments or permutation approach. A positive and significant value of global Moran's statistics indicates a positive spatial autocorrelation, which states similarity between considered areas, whereas negative values of global Moran's statistic inform about the diversity of those areas.

Local Moran's statistic is indicated with the following formula (Kopczewska, 2011):

$$\mathrm{I}_i = \frac{(x_i - \bar{x})\sum_{i=1}^{n} w_{ij}(x_j - \bar{x})}{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}}, \tag{12}$$

where the designations of individual elements are the same as in global Moran's statistics. The tests of significance of local Moran's statistic are based on distribution indicated by a conditional randomization or permutation and inference based on the pseudo-significance levels (Anselin, 1995). Significantly positive values of local Moran's statistics mean that an area is surrounded by similar neighbouring areas. A negative value of local statistics determines different values of a researched variable in neighbouring areas.

The abovementioned global and local Moran's statistics will be used in the suggested construction of a factorial design of experiment.

4. Construction algorithm of a factorial design of an experiment

In the present section, the construction of factorial design of experiment considering a division of an experimental area will be presented. The purpose of the presented design of experiment will be to limit the number of the experimental trials carried out, which leads to a decrease of the costs of implementing the experiment.

Consider the following design of an experiment:

$$P_n = \begin{Bmatrix} (x_{11},x_{21}) & (x_{11},x_{22}) & \dots & (x_{11},x_{2k}) & (x_{12},x_{21}) & (x_{12},x_{22}) & \dots & (x_{12},x_{2k}) & \dots & (x_{1l},x_{2k}) \\ \frac{1}{n} & \frac{1}{n} & \dots & \frac{1}{n} & \frac{1}{n} & \frac{1}{n} & \dots & \frac{1}{n} & \dots & \frac{1}{n} \end{Bmatrix}$$

The design P_n considers two factors X_1 i X_2 occurring on levels *l* and *k*, respectively, and takes into account $n = lk$ experimental trials with the same weights. Supposing the costs of the studied process allow the experimenter to fulfil the limited number of experimental trials, which allow to estimate the parameters of response surface. It is worth noting that the use of the classical designs of experiments is pointless. Therefore, we should consider in what way we are to construct the design with a limited number of experimental trials providing the most accurate possible assessment of response surface.

The algorithm of the construction of the design of experiment that considers the aforementioned circumstance of the experiment is presented in the following steps:

1. We determine the minimal number of experimental trials $n_{\min}$ in order to estimate all the parameters of the studied response surface function such as function (10).

2. Using local Moran's statistics, we analyze a spatial autocorrelation for the considered experimental area.

3. From all points of the experimental area we choose those for which the value of local Moran's statistics is not significant.

4. From the chosen design points we choose an appropriate number of points for which the inequality $\mathbf{det}\,(\mathbf{F}^{\mathbf{T}}\mathbf{F}) > 0$ works (Box, Draper, 1971; Wawrzynek, 1993) and the value of local Moran's statistics is close to 0. The established design is defined as:

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

5. Taking into account the chosen experimental trials, we determine the form of the variance of the response surface function estimator (9).

6. We calculate the value of variance of response surface estimator for the points of the experimental area which have been determined in step 3 and have not been included in $P_{n_{\min}}$ design.

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

8. The following experimental trial in the fulfilled experiment is carried out as defined in the previous step point of the experimental area.

We create a design of an experiment presented in (14):

$$P_{n_{\min}+1} = \left\{ \begin{matrix} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_{n_{\min}} & \mathbf{x}_{n_{\min}+1} \\ \dfrac{1}{n_{\min}+1} & \dfrac{1}{n_{\min}+1} & \cdots & \dfrac{1}{n_{\min}+1} & \dfrac{1}{n_{\min}+1} \end{matrix} \right\}. \tag{14}$$

The mentioned construction of the design of the experiment allows not only to carry out the experiment with a limited number of experimental trials, but also to determine the following points of the experimental area, ensuring the most accurate possible estimate response surface.

5. Application of the suggested construction of design of a factorial experiment

The application of the construction algorithm of design of an experiment will be described in the linear and non-linear form of response surface function and for the experimental area stated as:

$$D_E = \left\{ (x_1, x_2) : x_1 \in \{1,2,3,4,5\}, x_2 \in \{1,2,3,4\} \right\}. \tag{15}$$

Consider the response surface function in (16):

$$y(\mathrm{x}) = \beta_1 x_1 + \beta_2 x_2 . \tag{16}$$

The value of global Moran's statistics for response surface (16) and the considered experimental area equals 0.74, which indicates a strong positive spatial autocorrelation.

According to the algorithm of the design construction presented in the previous section, first we have to determine the smallest number of experimental trials and therefore we suppose that $n_{\min} = 2$. Then we have to determine the values of local Moran's statistics for particular points of the experimental area. The results are presented in Table 1.

Table 1. Values of local Moran's statistics for response surface (16)

No	Sign	x_1	x_2	$y(x)$	Local statistics
1	*aa*	1	1	2	2.6923
2	*ab*	1	2	3	1.7308
3	*ac*	1	3	4	**0.5769**
4	*ad*	1	4	5	**0.0769**
5	*ba*	2	1	3	1.4615
6	*bb*	2	2	4	0.5385
7	*bc*	2	3	5	**0.0769**
8	*bd*	2	4	6	**0.0385**
9	*ca*	3	1	4	**0.3462**
10	*cb*	3	2	5	**0.0769**
11	*cc*	3	3	6	**0.0256**
12	*cd*	3	4	7	**0.5769**
13	*da*	4	1	5	**0.0256**
14	*db*	4	2	6	**0.0256**
15	*dc*	4	3	7	0.5604
16	*dd*	4	4	8	1.7308
17	*ea*	5	1	6	**0.0769**
18	*eb*	5	2	7	**0.5769**
19	*ec*	5	3	8	1.7308
20	*ed*	5	4	9	2.6923

Source: own elaboration.

Then, from the experimental trials with highlighted and insignificant values of local statistics, we choose the smallest number of experimental trials in order to match $\mathbf{det}\,(\mathbf{F}^{\mathbf{T}}\mathbf{F}) > 0$ and the value of the local statistic is close to 0. We construct the initial plan"

$$P_2 = \begin{Bmatrix} (3,3) & (4,1) \\ \frac{1}{2} & \frac{1}{2} \end{Bmatrix}. \tag{17}$$

The variance of response surface function estimator is presented in the following formula:

$$V\tilde{y}(\mathrm{x}) = \frac{1}{81}\sigma^2\left(10\, x_1^2 - 26x_1x_2 + 25x_2^2\right). \tag{18}$$

Counting the value of the variance (18) for the points of the experimental area that are not included in the initial plan and for which local statistics are insignificant, we obtain the lowest value for the experiment, marked as *cb*. Therefore, the created design of the experiment is presented in (19):

$$P_3 = \begin{Bmatrix} (3,3) & (4,1) & (3,2) \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{Bmatrix}. \tag{19}$$

The design of the experiment allows us to estimate the parameters of the response surface (16) through carrying out two experimental trials and to determine the next points of design of experiment characterized with the best results.

Classical factorial designs of experiments are usually used to estimate the parameters for a linear response surface function (Kończak, 2007). In experiments where response surface has a nonlinear form, using classical design of experiments is not justified.

Therefore, consider the response surface function in (20):

$$y(\mathrm{x}) = \sin\left(\frac{\beta_1x_1 + \beta_2x_2}{2}\right)\cos\left(\frac{\beta_1x_1 + \beta_2x_2}{2}\right). \tag{20}$$

We will construct the design of the experiment according to the suggested algorithm taking into account the experimental area formed (15). For the considered experimental area and response surface function (20) the value of global Moran's statistics equals 0.42, which proves a positive spatial autocorrelation of experimental area points. Supposing that the expected costs of a planned experiment allow to carry out a maximum of five experimental trials. Then the smallest number of experimental trials to carry out allows us to estimate the parameters of response surface function: $n_{\min} = 4$. Consequently, on the basis of the carried out experimental trials, it will be possible to determine the fifth point of the design with the smallest dispersion. After determining the number of experimental trials, we indicate the insignificant values of local Moran's statistics. The results are presented in Table 2. Then, according to the approved criterion, we construct the initial plan:

$$P_4 = \begin{Bmatrix} (1,2) & (2,1) & (3,4) & (5,2) \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{Bmatrix}. \tag{21}$$

The variance of response surface function estimator is shown in (22):

$$V\tilde{y}(\mathrm{x}) = \frac{1}{23}\sigma^2\left(3\cos^2\alpha - 4\sin\alpha\cos\alpha + \frac{25}{13}\sin^2\alpha\right), \tag{22}$$

where $\alpha = \frac{x_1 + x_2}{2}$.

Table 2. Values of local Moran statistics for nonlinear response surface (20)

No	Sign	x_1	x_2	$y(x)$	Local statistics
1	*aa*	1	1	0.4546	**0.1217**
2	*ab*	1	2	0.0706	**– 0.0019**
3	*ac*	1	3	– 0.3784	**0.2738**
4	*ad*	1	4	– 0.4795	**0.2738**
5	*ba*	2	1	0.0706	**– 0.0076**
6	*bb*	2	2	– 0.3784	**0.3308**
7	*bc*	2	3	– 0.4795	**0.3042**
8	*bd*	2	4	– 0.1397	**0.2928**
9	*ca*	3	1	– 0.3784	**0.3593**
10	*cb*	3	2	– 0.4795	**0.1724**
11	*cc*	3	3	– 0.1397	**0.1597**
12	*cd*	3	4	0.3285	**0.0837**
13	*da*	4	1	– 0.4795	**0.3752**
14	*db*	4	2	– 0.1397	**0.1597**
15	*dc*	4	3	0.3285	**0.3824**
16	*dd*	4	4	0.4947	1.9563
17	*ea*	5	1	– 0.1397	**0.2484**
18	*eb*	5	2	0.3285	**0.0837**
19	*ec*	5	3	0.4947	1.9563
20	*ed*	5	4	0.2061	**0.9582**

Source: own elaboration.

The variance takes the smallest value for the point of experimental area with sign *aa*. Therefore, the constructed design of experiment for nonlinear response surface is (23):

$$P_5 = \begin{Bmatrix} (1,2) & (2,1) & (3,4) & (5,2) & (1,1) \\ \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} \end{Bmatrix}. \qquad (23)$$

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

6. 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 enables us to carry out an experiment with the minimal number of experimental trials, which has an influence on the overall costs and time of the manufacturing process. The presented algorithm allows us to determine the next point of the experimental area with the most precise results.

Literature

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

Anselin L. (1995). *Local indicators of spatial association-LISA*. Geographical Analysis. Vol. 27. Pp. 93-115.

Box M.J., Draper N.R. (1971). *Factorial designs, the |X'X| criterion, and some related matters*. Technometrics. Vol. 13. No. 4. Pp. 731-742.

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

Kopczewska K. (2011). *Ekonometria i statystyka przestrzenna z wykorzystaniem programu R CRAN*. CeDeWu. Warszawa.

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

Montgomery D.C. (1997). *Introduction to Statistical Quality Control*. John Wiley & Sons. 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.