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 3(302), 2014

*Małgorzata Szerszunowicz**

ON THE BOOTSTRAP METHOD OF ESTIMATION OF RESPONSE SURFACE FUNCTION

Abstract. Nowadays, in many fields of science it is necessary to carry out miscellaneous analyses using classical statistical methods, which usually have correct assumptions. These assumptions in the research realities cannot always be met, which makes it impossible to carry out analyses and leads to incorrect conclusions and recommendations.

The study of the production process largely consists in the use of tools of statistical quality control which are based on classical statistical methods. These methods result in some improvements in technological and economic results of the manufacturing process. One of the tools of statistical quality control is the design of experiments, whose important element is the estimation of response surface function.

The aim of this paper is to present the bootstrap method of estimation of response surface function and its use for empirical data.

Keywords: design of experiments, response surface function, bootstrap

I. INTRODUCTION

The methodology of design of experiments was introduced at the beginning of the 20th century by R. A. Fisher in agricultural experiments. The theory of design of experiment is now being developed in many areas of science, including biology, medicine, chemistry, engineering and statistical quality control.

Design of experiments as a tool used to improve the results of the production process has applications in companies and a beneficial influence on technological and economic results of the manufacturing process.

II. DESIGN OF EXPERIMENTS

The application of the methods of design of experiments in statistical quality control leads to determining the factors which most significantly affect the variable characterizing the investigated process as well as allows to specify the

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

values of factors for which result variable reaches the desired value or the smallest variability.

The effectiveness of the use of design of experiments while preparing the production process is connected, among other things, with a detailed formulation of the problem and the proper selection of factors and response variable characterizing the process under study. Moreover, one ought to determine the number of experimental trials and take into account the possible randomization restrictions. Then, after performing the experiment, one should analyze the results and formulate conclusions and recommendations for the process under study (Montgomery, 1997).

The dependence between result variable and the set of factors is an issue of the analysis of variance or a problem of the analysis of regression. This paper will look at the regression analysis which – in the framework of the theory of design of experiment – involves the introduction of the appropriate terminology and designation.

The experiment is a sequence of *n* successive experimental trials which are a single result of the value of response variable *Y*, with the fixed values of $X_1, X_2, \ldots, X_m$ specified on sets $\mathrm{X}_1, \mathrm{X}_2, \ldots, \mathrm{X}_\mathrm{m}$ respectively. The experimental area is a set of points $\mathrm{x} = (x_1, x_2, \ldots, x_m)$, where $x_i \in \mathrm{X}_i, i = 1, 2, \ldots, m$. Then the set of pairs $P_n = \{\mathrm{x}_j, p_j\}_{j=1}^{n}$ is a design of experiment with *n* experimental trials, where $\mathrm{x}_j = (x_{1j}, x_{2j}, \ldots, x_{mj})$ and $p_j = \frac{n_j}{n}$, where n_j is the number of experimental trials in point x_j of the experimental area, and also $\sum_{j=1}^{n} n_j = n$, $\sum_{j=1}^{n} p_j = 1$ for $j = 1, 2, \ldots, n$. The relationship between the set of factors and the response variable characterizing the process in the best way is presented in the form of the following statistical model:

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

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 of the values of the factors. This model can be presented as a general linear model (Wawrzynek, 2009) $\mathrm{Y} = \mathrm{F}\beta + \varepsilon$, where

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

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

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

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

$$F = \begin{bmatrix} f_1(x_1) & \ldots & f_k(x_1) \\ \vdots & \ddots & \vdots \\ f_1(x_n) & \ldots & f_k(x_n) \end{bmatrix} = \left[f(x_1)\ f(x_2) \ldots f(x_n)\right]^{T} \tag{6}$$

and $f_i(x_j) \equiv x_{ij}$, for $i = 1,2,\ldots k$, $j = 1,2,\ldots,n$. Then the response surface function is defined as a formula $y = F\beta$. The realization of the classical design of experiments leads to the estimation of parameters of response surface function defined as (Wawrzynek, 2009)

$$y(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_m x_m + \beta_{12} x_1 x_2 + \ldots + \beta_{12\ldots m} x_1 x_2 \ldots x_m . \tag{7}$$

The estimation of parameters of response surface function is carried out with the use of the least squares method. The use of this method depends on certain assumptions being fulfilled, which in the case of many different production processes cannot always be ensured (Montgomery, 2001). Also, the use of linear response surface function to describe the manufacturing process can lead to some simplification of the nature of the relationship between response variable and the set of factors. Then it may be successful to use nonlinear response surface functions.

III. BOOTSTRAP ESTIMATION

The procedures for designing manufacturing processes do not always allow to verify the assumptions of the applied methods, which results in the formulation of unfounded recommendations for the process under study. In the case of the estimation of the parameters of response surface function, one of the ways to avoid improper conduct is to use the nonparametric method of estimation of the parameters of response surface function. One of the nonparametric methods of estimation is the bootstrap estimation of the parameters of regression function.

The bootstrap method for the specified statistical model in the form of

$$y_i = f(\mathrm{x}_i, \beta) + \varepsilon_i \tag{8}$$

can be presented in the following phases (Domański, Pruska, 2000):

1. Determination of the values of $\hat{\beta}$ estimator with the use of the least squares method.
2. Estimation of the vector of residuals according to the following formula: $\hat{\varepsilon}_i = y_i - f(\mathrm{x}_i, \hat{\beta})$.
3. Generation of a bootstrap sample $(\varepsilon_1*, \varepsilon_2*, \ldots \varepsilon_n*)$ according to the specified probability distribution $P(Z = \hat{\varepsilon}_i) = \dfrac{1}{n}$.
4. Formation of a sample $(Y_1*, Y_2*, \ldots, Y_n*)$ with realizations $y_i* = f(\mathrm{x}_i, \hat{\beta}) + \varepsilon_i*$.
5. Estimation of the vector of parameters for the specified model with the use of the least squares method on the basis of the obtained values $(y_1*, y_2*, \ldots, y_n*)$.
6. Generation of a vector of the values $(\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_N)$ with N-time realization of points 3, 4 and 5 of the present algorithm.

The determined vector describes the bootstrap distribution of the estimator of the parameters of response surface function. This vector allows one to define the appropriate confidence interval for the parameters in question (Kończak, 2013).

IV. THE USE OF THE BOOTSTRAP METHOD FOR ESTIMATION OF RESPONSE SURFACE FUNCTION

Let us use a data experiment consisting in testing the injection molding process (Hedayat *et al.*, 1999). The response variable which characterizes the process under study in the best way is the percentage of the shrinkage of the molded product, which depends on these five factors considered in the experiment: cycle time (X_1), form temperature (X_2), holding down pressure (X_3), cavity thickness (X_4) and injection speed (X_5). Each of the factors takes two values: -1 and 1. The results of the performed experiment are shown in Table 1.

Table 1. The results of the injection molding experiment

No.	X_1	X_2	X_3	X_4	X_5	Y	No.	X_1	X_2	X_3	X_4	X_5	Y
1	–1	–1	1	–1	1	2.5	**17**	–1	–1	–1	–1	–1	2.2
2	–1	–1	1	–1	1	2.7	**18**	–1	–1	–1	–1	–1	2.3
3	–1	–1	1	–1	1	0.3	**19**	–1	–1	–1	–1	–1	2.1
4	–1	–1	1	–1	1	0.3	**20**	–1	–1	–1	–1	–1	2.3
5	–1	1	–1	1	–1	0.5	**21**	1	1	1	–1	–1	2.0
6	–1	1	–1	1	–1	0.4	**22**	1	1	1	–1	–1	1.9
7	–1	1	–1	1	–1	3.1	**23**	1	1	1	–1	–1	1.9
8	–1	1	–1	1	–1	2.8	**24**	1	1	1	–1	–1	1.8
9	–1	1	1	1	1	2.0	**25**	1	1	–1	–1	1	4.0
10	–1	1	1	1	1	1.8	**26**	1	1	–1	–1	1	4.6
11	–1	1	1	1	1	1.9	**27**	1	1	–1	–1	1	1.9
12	–1	1	1	1	1	2.0	**28**	1	1	–1	–1	1	2.2
13	1	–1	–1	1	1	3.0	**29**	1	–1	1	1	–1	2.1
14	1	–1	–1	1	1	3.0	**30**	1	–1	1	1	–1	1.0
15	1	–1	–1	1	1	3.1	**31**	1	–1	1	1	–1	4.2
16	1	–1	–1	1	1	3.1	**32**	1	–1	1	1	–1	3.1

According to the classical methods of design of experiments, on the basis of the results of the experiment one ought to estimate the parameters of linear response surface function presented as formula (7). This procedure may lead to a significant simplification of the relationship between the response variable and the set of factors. Now, on the basis of the experimenter's knowledge, let the response surface function be defined as follows:

$$y(\mathrm{x}) = \beta_1 \exp(x_1) + \beta_2 \exp(x_2) + \exp(\beta_3 x_3) + \beta_4 x_4^2 x_5^2 + \beta_5 \frac{1}{x_5} \tag{9}$$

The parameters of response surface function (9) have been estimated according to the bootstrap method of estimation for different number of replications. The results of this estimation with specified confidence intervals are shown in Table 2.

Table 2. The results of estimation of response surface function (9) according to the bootstrap method

	$N = 100$			$N = 1000$			$N = 5000$		
	Estimate	2.50%	97.50%	Estimate	2.50%	97.50%	Estimate	2.50%	97.50%
$\hat{\beta}_1$	–0.29	–0.93	0.30	0.26	–1.73	2.84	7.61	0.41	14.16
$\hat{\beta}_2$	–0.07	–1.00	0.45	–0.68	–3.02	2.29	–3.15	–10.08	4.30
$\hat{\beta}_3$	–1.07	–1.55	–0.52	–0.32	–2.39	1.86	–2.75	–3.72	1.10
$\hat{\beta}_4$	2.81	1.52	5.17	5.80	1.52	9.83	–19.82	–34.96	–0.15
$\hat{\beta}_5$	–1.01	–2.16	0.19	2.94	–0.17	5.70	6.34	–2.48	19.75

Source: own work.

The bootstrap distribution of parameter estimators of response surface function (9) is graphically presented in Figure 1 (for $N = 1000$).

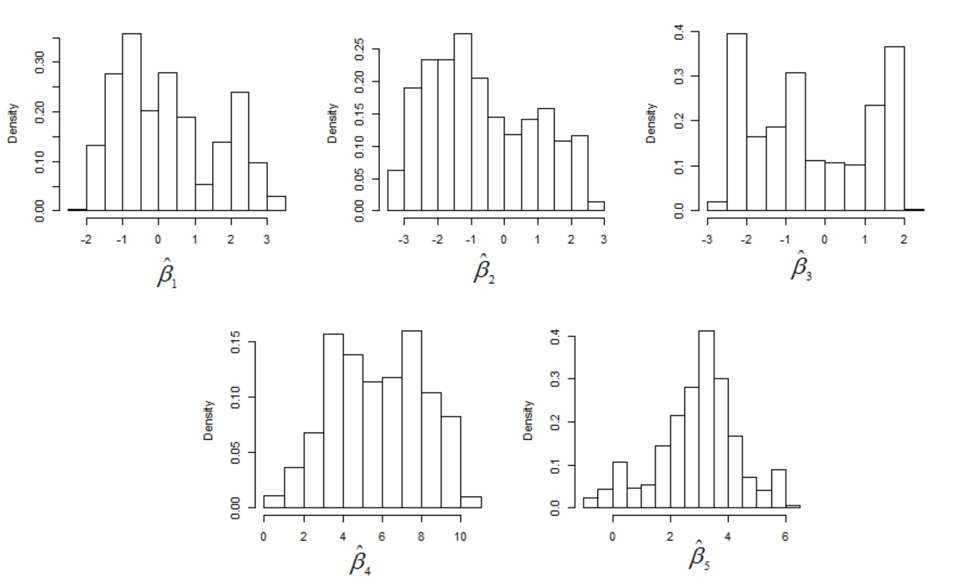

Figure 1. The bootstrap distribution of parameters

The presented distribution of the parameters of nonlinear response surface function allows one to describe the interdependence of the studied variables in greater detail. Also, one can define the confidence intervals for the values of the estimated response surface function.

V. CONCLUSIONS

Design of experiments as one of the methods for statistical quality control allows one to describe the influence of factors on the response variable on the basis of the estimated response surface function. Classical procedures for designing experiments usually use linear response surface functions, which can lead to some generalization of the relationship between the results of the considered process and the value of the factors.

In this paper, an alternative method of estimation of linear and nonlinear response surface functions has been presented. The method uses the bootstrap estimation of parameters of regression function. The proposed non-classical method has been used for the estimation of the parameters of a nonlinear function to describe the injection molding process.

ACKNOWLEDGEMENTS

The research was supported by the Polish Science Center grant DEC-2011/03/B/HS4/05630.

REFERENCES

Domański C., Pruska K. (2000), *Nieklasyczne metody statystyczne*, Polskie Wydawnictwo Ekonomiczne, Warszawa.

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

Kończak G. (2012), *Wprowadzenie do symulacji komputerowych,* Wydawnictwo uniwersytetu Ekonomicznego w Katowicach, Katowice.

Montgomery, D. C. (2001), *Design and Analysis of Experiments*, John Wiley & Sons, Inc., New York.

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.

Małgorzata Szerszunowicz

O BOOTSTRAPOWEJ METODZIE ESTYMACJI PARAMETRÓW FUNKCJI POWIERZCHNI ODPOWIEDZI

Obecnie w wielu dziedzinach nauki niezbędnym jest przeprowadzanie różnorodnych analiz wykorzystujących zazwyczaj klasyczne metody statystyczne, które wymagają spełnienia określonych założeń. Założenia te w realiach badanych zjawisk nie zawsze mogą być spełnione, co uniemożliwia przeprowadzenie analiz lub prowadzi do niewłaściwych wniosków i zaleceń.

Badanie przebiegu procesu produkcyjnego polega przede wszystkim na wykorzystaniu narzędzi statystycznej kontroli jakości, których podstawą są klasyczne metody statystyczne. Metody te prowadzą do polepszenia rezultatów technologicznych i ekonomicznych procesu produkcyjnego. Jednym z narzędzi kontroli jakości jest planowanie eksperymentów, którego ważnym etapem jest estymacja funkcji powierzchni odpowiedzi.

Przedmiotem niniejszego artykułu jest przedstawienie bootstrapowej metody estymacji funkcji powierzchni odpowiedzi oraz jej wykorzystanie dla pewnych danych empirycznych.