Małgorzata Złotoś
University of Economics in Katowice
Department of Statistics, Econometrics and Mathematics
e-mail: malgorzata.zlotos@ue.katowice.pl

On the use of permutation tests in the significance testing of the response surface function parameters

Abstract

The methods of experimental design were first used in agricultural experiments performed by R. A. Fisher. The development of experimental design methods took place along with their effective use in production companies. The most frequently used designs of experiments are the factorial designs of experiments. One of the stages of the factorial design of experiments is the estimation of the response surface function formula which describes the influence of factors on the response variable values. The task of the experimenter is to indicate the factors which have a significant influence on the response variable. In this case, in the classical approach, the *t*-test of the significance of particular parameters of the response surface function is used. The *t*-test requires the fulfilment of the assumptions about distribution and independence of the model errors. If the assumptions are not fulfilled, or the sample size is not sufficient, the use of the *t*-test is unjustified. Then the alternative approach to verify the significance of the response surface parameters is the use of a permutation test. Permutation tests use the simulation methods and do not entail the fulfilment of strict assumptions referring to the distribution of errors and the sample size of experimental data. The present paper deals with the use of a permutation test which allows us to assess the significance of response surface function parameters when the number of experimental data is small.

Keywords: design of experiments, permutation tests, response surface function
JEL Classification: C99, C12, C15

1. Introduction

Experimental design methods are among those statistical quality control tools which are used effectively in practice. Their implementation leads to some improvements in technological parameters of the manufacturing process, which enhances the quality of products and decreases financial losses related to the production process in question. The proper use of experimental design methods requires the adequate practical knowledge about the process and the knowledge of the statistical methods [Kończak 2007].

An experiment is a sequence of experimental trials. An individual experimental trial is an obtainment of the response variable Y with the fixed values of factors $X_1, X_2, \dots, X_m$. Then the design of the experiment is defined as the layout of factors levels in the further experimental trials. The dependence of the response variable Y and of the values of factors is defined as a statistical model [Wawrzynek 1993]

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

where $EY(X_1, X_2, \dots, X_m) = y(X_1, X_2, \dots, X_m)$, $E(\varepsilon) = 0$, $V(\varepsilon) = \sigma^2$ and σ^2 is a fixed value. The model (1) can be presented as the formula of the general linear model as follows [Wawrzynek 2009]:

$$\mathrm{Y}^{\mathrm{T}} = (Y_1\ Y_2\ \dots\ Y_m) \tag{2}$$

$$\varepsilon^{\mathrm{T}} = (\varepsilon_1\ \varepsilon_2\ \dots \varepsilon_n) \tag{3}$$

$$\beta^{T} = (\beta_1\ \beta_2\ \dots\ \beta_k) \tag{4}$$

$$\mathrm{f}^{T}(\mathrm{x}) = (f_1(\mathrm{x})\ f_2(\mathrm{x})\ \dots\ f_k(\mathrm{x})) \tag{5}$$

$$\mathrm{F} = \begin{bmatrix} f_1(\mathrm{x}_1) & \cdots & f_k(\mathrm{x}_1) \\ \vdots & \ddots & \vdots \\ f_1(\mathrm{x}_n) & \cdots & f_k(\mathrm{x}_n) \end{bmatrix} \tag{6}$$

where $f_i(\mathrm{x}_j) \equiv x_{ij}$, for $i = 1,2, \dots, k, j = 1, 2, \dots, n$. Then the response surface function is defined as $\mathrm{y} = \mathrm{F}\beta$.

Usually, the response surface functions which do not include any interaction between the factors are considered. Their formula is as follows [Montgomery 2001, Wawrzynek 2009]:

$$y(x_1, x_2, \dots, x_m) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_m x_m \,. \tag{7}$$

The response surface functions which take into account the interactions of the factors are also considered:

$$y(x_1, x_2, \dots, x_m) = \beta_0 + \beta_1 x_1 + \cdots + \beta_m x_m + \beta_{12} x_1 x_2 + \cdots + \beta_{m-1\,m} x_{m-1} x_m \,. \tag{8}$$

In the classical approach, in order to estimate the parameters of vector β of the response surface function, the least square method is used [Aczel 2000, Elandt 1964, Montgomery 1997, Wawrzynek 1993].

2. The significance of response surface function

The response surface function is a mathematical description of the dependence of the factors on the response variable. The analysis of the response surface function particularly allows us to verify the model significance and the significance of the individual variable impact on the response variable. Then, in order to use proper parametric tests, is important to assume that the distribution of model residuals is a normal distribution with the expected value 0 and with a standard deviation σ^2 [Montgomery 2001].

In order to verify the response surface function model significance, the hypotheses were formulated as follows [Aczel 2000]:

$$\begin{gathered} H_0\colon \beta_1 = \beta_2 = \cdots = \beta_k = 0 \\ H_1\colon \beta_j \neq 0 \,,\ \text{dla pewnego } j. \end{gathered} \tag{9}$$

Assuming that the null hypothesis is true, the test statistic

$$F = \frac{(\hat{\beta}'X'y - n\bar{y}^2)/k}{(y'y - \hat{\beta}'X'y)/(n-k-1)} \tag{10}$$

has a F-Snedecor distribution with k and $n-k-1$ degrees of freedom. The null hypothesis should be rejected when $F > F_{\alpha,k,n-k-1}$.

The indication of factors which should be included or omitted in the considered model of response surface function is possible because of the verification of hypotheses formulated below:

$$\begin{gathered} H_0\colon \beta_j = 0 \\ H_1\colon \beta_j \neq 0 \,, \end{gathered} \tag{11}$$

where β_j is an established parameter of the response surface function. Then the value of the test statistic [Aczel 2000]

$$t = \frac{\hat{\beta}_j}{\sqrt{\sigma^2 C_{jj}}}, \tag{12}$$

is calculated, where C_{jj} is a diagonal element of matrix $(X'X)^{-1}$. If the value of the test statistic (12) satisfies the inequality $|t| > t_{\alpha/2,n-k-1}$, the null hypothesis should be rejected. Then the variable X_j should not be included in the considered model of response surface function.

3. Permutation tests

Permutation tests, like experimental design, were proposed by R. A. Fisher in the 1920s. However, due to computational difficulties, they did not have application as early as experimental design. It was not until the beginning of the 21st century that the permutation methods were developed [Kończak 2016].

A permutation test is described as a general method of estimating the probability of an event occurring. Permutation tests are an alternative to parametric tests which use only the sample data; do not require assumptions regarding the distribution form in the population; are resistant to the occurrence of outliers; and can be used for the sample with a small number of observations [Berry et al. 2014].

Good [2005] presents the procedure for permutation tests in the following stages:

1. Define the null-hypothesis and the alternative hypothesis.
2. Choose the formula of testing statistic T.
3. Determine the value of the test statistic (T_0) for the sample data.
4. Count the value of the test statistic T for a sufficiently large number (N) of permutations of a data set to obtain the set $\{T_1, T_2, \ldots, T_N\}$.
5. Determine the ASL value and make your decision.

If the alternative hypothesis is right-sided, then the value of ASL (*Achieving Significance Level*) is described as follows:

$$ASL = P(T_i \geq T_0). \tag{13}$$

Then the estimation of ASL value is determined on the basis of the following formula:

$$\widehat{ASL} = \frac{card\{i: T_i \geq T_0\}}{N}. \tag{14}$$

When the two-sided alternative hypothesis is considered, then the ASL value can be rewritten as follows:

$$ASL = P(\,|T_i| \geq |T_0|), \quad (15)$$

and the approximate value of ASL is calculated using the following formula:

$$\widehat{ASL} = \frac{card\,\{\,i: |T_i| \geq |T_0|\}}{N}. \quad (16)$$

The null-hypothesis should be rejected when the value $\widehat{ASL}$ is smaller than the assumed significance level.

The use of permutation tests for verifying ~~of~~ the significance of the model (7) or (8) and the significance of its parameters is connected with the description of the permutation rules of multidimensional data. O'Gorman [2012] gives four appropriate methods of this permutations: permutation of errors, permutation of residuals, permutation of independent variables and permutation of the response variable. Because of the fixed values of factors (independent variable) in particular experimental trials, the permutation of the response variable is taken into account for the considered data in experimental design. In order to verify the significance of the response surface model and the significance of parameters of the response surface function, in the procedure of the permutation test the test statistics (10) and (12) can be used accordingly [Kończak 2012, 2016].

4. Example

It is assumed that the brake horsepower (response variable Y) developed by an automobile engine depends on the engine speed in revolutions per minute (factor X_1), the road octane number of the fuel (factor X_2) and the engine compression (factor X_3). The experimental data for $n = 12$ experimental trials are presented in Table 1.

Table 1. The experimental data.

No.	X_1	X_2	X_3	Y
1	2000	90	100	225
2	1800	94	95	212
3	2400	88	110	229
4	1900	91	96	222
5	1600	86	100	219

6	2500	96	110	278
7	3000	94	98	246
8	3200	90	100	237
9	2800	88	105	233
10	3400	86	97	224
11	1800	90	100	223
12	2500	89	104	230

Source: Montgomery D. C. [2001], *Design and Analysis of Experiments*, John Wiley & Sons, Inc., New York

The response surface function which does not include interactions between factors is taken into consideration and it is in the following form:

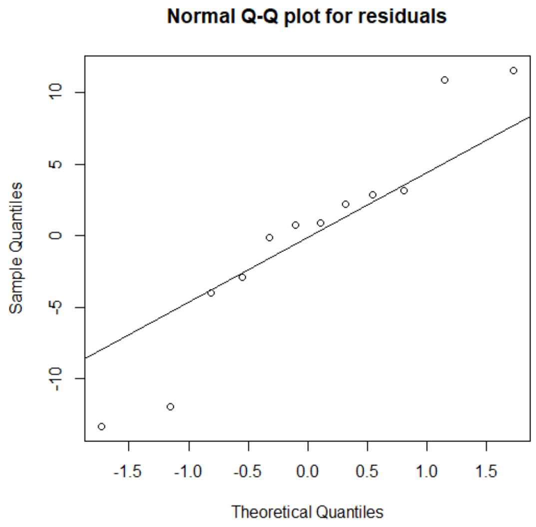
$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3. \tag{17}$$

Using the least mean square method, the values of response surface function parameters were estimated:

$$y = -266.031 + 0.011x_1 + 3.135x_2 + 1.867x_3. \tag{18}$$

For the estimated values of the model residuals, Shapiro-Wilk's test was prepared. The obtained *p-value* = 0,3706, so there are no arguments for rejecting the null-hypothesis saying that the distribution of residuals is the normal distribution (Figure 1).

Figure 1. Normal Q-Q plot for residuals of model (18)



Source: the author's own elaboration

The significance of the response surface model was verified and the value of statistic $F = 11.12$ was obtained. The calculated value of statistic F is bigger than the critical value $F_{0,05,3,8} = 4{,}07$, so the null-hypothesis should be rejected, which means that the response surface model (18) is significant. Moreover, the significance of the response surface parameters was investigated with parametric test t. The results are presented in Table 2.

Table 2. The results of testing significance of response surface function (18).

Parameter	Estimate	Standard Error	t - value	p - value
β_0	-266.03	92.674	-2.871	0.021
β_1	0.017	0.004	2.390	0.044
β_2	3.135	0.844	3.712	0.006
β_3	1.867	0.535	3.494	0.008

Source: the author's own elaboration

The considered experimental data include a small number of experimental trials, so because of the assumptions about the distribution of residuals the use of parametric test of model significance or the t test can be unfounded. Therefore, in order to confirm or deny the obtained results, the proper permutation tests were carried out.

The permutation test of model significance uses the test statistic (10). For $N = 1000$ permutations of the response variable, formulas of response surface functions were estimated, with the appropriate values of the test statistic F. According to (14), the value $\widehat{ASL} = 0.003$ was estimated. Therefore, on the significance level $\alpha = 0.05$ the null-hypothesis should be rejected, which confirms that the response surface function model (18) is significant.

The significance of response surface function parameters was verified with the test statistic (12). For every parameter of response surface function $N = 1000$ permutations of the response variable were performed and respond surface functions were estimated with t-statistic values. Then, according to (16), the values of $\widehat{ASL}$ were estimated. The results are presented in Table 3.

Table 3. The values of $\widehat{ASL}$.

Parameter	$\widehat{ASL}$ - value
β_0	0.078

β_1	0.030
β_2	0.005
β_3	0.008

Source: own elaboration

On the basis of the performed calculations, it can be seen that conclusions for parametric tests and for appropriate permutation tests are similar. It should be emphasized that the permutation tests did not require fulfilment of the assumptions regarding the distribution of residuals of the considered model.

5. Conclusions

The methods of experimental design are used primarily in statistical quality control procedures. One of the fundamental stages of experimental design is to estimate the response surface function model and its analysis, which allows formulating proper recommendations for the production process in question. In particular, the analysis of the response surface function relies on the assessment of the significance of the estimated model and its parameters, where classical parametric tests are used. The present study analyses the response surface function for an experiment which involves a small number of experimental trials, which may lead to incorrect conclusions in the case of parametric tests. Then, it was proposed to use permutation tests, which do not require fulfilment of the restrictive assumptions regarding the distribution of model residuals and can be used for a data set with a small number of observations.

References

Aczel A. [2000], *Statystyka w zarządzaniu*, Wydawnictwo Naukowe PWN, Warszawa

Berry K. J., Johnston J. E., Mielke Jr. P. W. [2014], *A Chronicle of Permutation Statistical Methods*, Springer International Publishing, New York

Elandt R. [1964], *Statystyka matematyczna w zastosowaniu do doświadczalnictwa rolniczego*, Państwowe Wydawnictwo Naukowe, Warszawa

Good P. [2005], *Permutation, Parametric and Bootstrap Tests of Hypotheses,* Springer Science Business Media, Inc., New York

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

Kończak G. [2012], *On Testing the Significance of the Coefficients in the Multiple Regression Analysis*, Acta Universitatis Lodziensis, Folia Oeconomica, nr 269, s. 61-71

Kończak G. [2016], *Testy permutacyjne. Teoria i zastosowania.*, 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

O'Gorman T.W. [2012], *Adaptive Tests of Significance Using Permutations of Residuals with R and SAS*, John Wiley and Sons, New Jersey

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

O wykorzystaniu testów permutacyjnych w ocenie istotności parametrów funkcji powierzchni odpowiedzi

Streszczenie

Metody planowania eksperymentów po raz pierwszy wykorzystane zostały w doświadczeniach rolniczych przeprowadzonych przez R.A. Fishera. Rozwój metod planowania eksperymentów nastąpił wraz z ich efektywnym wykorzystaniem w praktyce przedsiębiorstw produkcyjnych. Do najczęściej wykorzystywanych planów eksperymentów należą plany eksperymentów czynnikowych. Jednym z etapów eksperymentu czynnikowego jest oszacowanie postaci funkcji powierzchni odpowiedzi, która opisuje wpływ czynników na wartości zmiennej wynikowej. Zadaniem eksperymentatora jest wskazanie tych czynników, które istotnie oddziałują na zmienną wynikową. W tym celu, w podejściu klasycznym, wykorzystuje się test *t* istotności poszczególnych parametrów funkcji powierzchni odpowiedzi. Test *t* wymaga spełnienia założeń dotyczących postaci rozkładu i niezależności reszt modelu. Jeżeli założenia te nie są spełnione lub liczebność próby nie jest dostatecznie duża wykorzystanie testu *t* jest nieuzasadnione. Wówczas alternatywnym podejściem w weryfikacji istotności parametrów funkcji powierzchni odpowiedzi jest wykorzystanie testu permutacyjnego. Testy permutacyjne wykorzystują metody symulacyjne oraz nie wymagają spełnienia restrykcyjnych założeń dotyczących postaci rozkładu lub liczebności próby danych eksperymentalnych. Przedmiotem referatu jest zastosowanie testu permutacyjnego, który

pozwoli na ocenę istotności parametrów funkcji powierzchni odpowiedzi w przypadku, gdy liczebność danych eksperymentalnych jest niewielka.

Słowa kluczowe: funkcja powierzchni odpowiedzi, planowanie eksperymentów, testy permutacyjne