COMPARATIVE STATISTICAL ESTIMATION OF PARAMETERS OF THE RESPONSE FUNCTION AT PLANNING EXPERIMENT

B.N. Poliakov

Comparison of statistical estimations of parameters of the response functions, received for active multifactorial experiment and at its planning shows, and their insufficient reliability for the last substantiates.

In modern scientific methodology mathematical methods of planning, including, optimum experiments, take rather worthy place in fundamental and applied scientific researches. The methods for the first time offered by English mathematician R.Fisher are informative, productive and economic and widely enough are applied to the decision of the big circle of problems, including, for studying mechanisms of the phenomena, for the organization and active management of experimental researches in various scientific directions and branches of industry, for creation of mathematical models (of the response functions) technological processes of objects of automation and search of their optimum regimes.

Mathematical methods of planning of experiments is «the theory of common sense» for the first time have been applied and have received the further development at statement and carrying out of experimental researches exactly in metallurgy in 50th years of the last century, owing to world famous scientific proceedings and practical activities in not easy conditions of not free the bright Russian scientist - the mathematician Prof. V.V. Nalimov [1,2].

These popular mathematical methods have received a wide spreading and in mechanical engineering, for example, at carrying out of numerical computer experiments for the decision of problems of optimum design, from positions of various purposeful functions (criterions quality) forged, cast, welded and thermo-stressed designs and details of complex configurations of metallurgical, mountain and oil-boring equipment [3,4,5].

At all unconditional efficiency of methods of planning of experiments we consider necessary to pay attention to some mathematical features of formed of the response functions– of the equations regression, in the sense of exactness and reliability of statistical estimations of their parameters. But it is necessary to note, that, essentially, represented article does not in the least claim at some new theoretical (the more so mathematical) generalizations, but only on a concrete example shows, usually hidden from the beginner – the user, these features, on the basis of comparison of statistical characteristics of parameters of the equations regression received at active multifactorial experiment (AME) and at its planning, represent in the form of a Latin square (LS).

The basis for the executed comparative analysis was made with the experimental information received at research at the stand of a level of power losses in various designs of knot of compaction of a rod of the drilling pump 13Gr at a various combination of influencing factors [6]. The volume of initial sample at AME has composed 787 experiences, that considerably surpasses the size of sample of full factorial experiment. Then the sample corresponding to a Latin square was formed of initial sample.

The experimental information both samples has been processed on the computer under programs multiple regression and correlation analyses [7]. The equations regression have been constructing in the form of a linear polynomial in true coordinates - $y = a_0 + \sum_{1}^{4} a_i x_i$ and in logarithmic (posinomial)– $\ell n y = a_0 + \sum_{1}^{4} a_i \ell n x_i$,

where y - force of friction in compaction in rod cavity at forcing, $\times$ 10H;

x_1 - pressure of a pumped over liquid in the cylinder of the pump, $\times$ 10Па;

x_2 - the design of compaction (Open Society " Uralmash " and other concerns [6]);

$x_3 \cdot 10^2$ - speed of moving of a rod, m/s;

$x_4 \cdot 10^{-2}$ - effort preliminary draw up of compaction, $\times$ 10H.

The results of the statistical analysis samples – statistical characteristics of the response functions – the equations of regression are presented in the table1, and comparison of statistical estimations of their parameters – in table 2. The analysis of the information table 1 and 2 allows to note the following.

All investigated factors (independent variables) initial sample have normal functions of distribution, in this case arithmetic average values and dispersions (in table 1 - variation coefficients) substantially do not differ from similar parameters of sample in the form of a Latin square, and distinctions make in all ≈ 0,3÷2,6 %, i.e. within the limits of mistakes of measurements, that proves lawful of application regression and correlation analyses.

The accepted scheme of a Latin square has turned out correct as sample LS forms of the response function adequate to sample AME. The response functions also adequately reflects theoretical (or physical) a picture of interaction of factors and their influence on the dependent variable. The multiple correlation coefficients of the equations regression differ inconsequentially in the statistical plan, though for LS their absolute values are overestimated obviously (0,778 > 0,632). The indices significance – the partial correlation coefficients qualitatively correctly reflects interrelations and also confirms inconsequentiality of factors in sample LS, but on absolute value they are above essentially in comparison with sample AME. Besides the interrelation between factors in both samples practically is absent, and available cross – correlation coefficients are essential less of the partial correlation coefficients, that raises significance of the lasts. Therefore ranging of factors on force of their influence on a dependent variable (by way of lowering) is established on absolute quantities of values of the partial correlation coefficients, which form their arrangement, identical for both samples, in following sequence – x_1, x_4, x_3 and x_2. The factor x_2 turned out inconsequential for posinomials both samples (see the table1), after rejection which correlations and regressions coefficients and them 95 %-es confidential intervals are changing inconsequentially – in all within the limits of 0,5 %.

The equations of regression in the form of posinomial is describe better properties samples both for AME, and for LS: the relative value of a standard error estimation - $S/\bar{y}$,% is less than ≈ in 4,5 times in comparison with linear polynomials in true coordinates (see the table 2), and exactness of approximation of the samples practically is identical (compare – 10,2 % and 12 %), i.e. takes place inconsequential of difference of absolute quantities of values $S/\bar{y}$, % for AME and LS, that is explained by practically absence of distinctions in ranges of variations of the observation data.

So, for of the response functions the following equations of regression are received:

- for sample AME:

$$y = 19{,}89 x_1^{0,23} x_2^{0,056} \left(x_3 \cdot 10^2\right)^{-0,11} \left(x_4 \cdot 10^{-2}\right)^{0,576},$$

- and for LS:

$$y = 23{,}48 x_1^{0,292} x_2^{-0,03} \left(x_3 \cdot 10^2\right)^{-0,29} \left(x_4 \cdot 10^{-2}\right)^{0,65}.$$

The main difference of the response function at planning experiment – in insufficient exactness and reliability of statistical estimations of parameters of the equation of regression are the broadest values of confidential intervals for absolute all statistical parameters and exactness the approximation, repeatedly exceeding values of similar parameters for sample AME, namely (see the table 2):

- for of the multiple correlation coefficient the relative value (in %) of width its of 95 % - a confidential interval - $\frac{\Delta R = R_2 - R_1}{R}$ or the relative value of its root-mean-square deviation - S_R / R,% ≈ in 3,6 times;

- for all regression coefficients relative value of width their of 95 % - es confidential intervals - $\Delta a_i / a_i$,% - in a range ≈ 2,0 ÷ 12,0 times (without taking into account of the inconsequential factor- x_2);

- for relative value of a standard error assessment (exactness of approximation of sample) - $S / \bar{y}$,% ≈ in 4,5 times.

So, for the sample planned in the form of a Latin square, exactness and reliability of the statistical parameters of the response function and their estimations come down sharply.

Thus, at all known undoubtedly positive qualities of methods of planning of experiments (including, significant reduction of quantity of experiences) is available essential shortage – low exactness and reliability of all statistical parameters of the response function and their estimations, that is objective property of small sample.

Presented statistically estimated glance at application of planning of experiment allows to make following conclusions:

- planning of experiment displays of the response function adequately AME, and statistical parameters and estimations of last are close on value to analogous parameters at AME, but their exactness and reliability are absolutely insufficient;

- the executed analysis specifically proves justice known in mathematical statistics of the proposition about, that reliability of a regression coefficients essentially depends from quantity of observations (size of sample) and from a standard error estimation, or from exactness, with which the equation regression describes of sample, and besides, that, it is impossible to reach an optimum of two criteria of quality simultaneously: in our case - considerably to reduce quantity of measurements (observations, experiences) and at the same time to receive high exactness and reliability of results;

- as at planning experiment the wide confidential intervals are formed, in particular, for of the regression coefficients, which values can make 50 and more percent from absolute value of the last, and in this case to be inessential (insignificant), that methodological more correctly to establish importance (the statistical significance), and also randomization of factors on absolute values of the partial correlation coefficients;

- shown insufficient (even low) reliability of the correlation and regression coefficients substantiates expediency to apply methods of scheduling only at the initial stage of experimental researches, after analysis of results which is follow to use, at least, the full factorial experiment, providing reception of more reliable and exact statistical assessments of parameters of the response function;

- full multifactorial and active experiment is the most reliable method of reception of authentic, exact and reliable of the response function.

REFERENCES

1. Nalimov V.V., Chernova N.A. Statistical methods of planning extreme experiments.-M.: Fizmatgiz, 1965, 340p.

2. Nalimov V.V. The Theory of Experiment. - M.: Science, 1971, 208p.

3. Shenk X.N. The theory of engineering Experiment.-Translation with engl.-M.: World, 1972, 381p.

4. Adler J.P., Markova E.V., Granovskii J.V. Planning of experiment at search of optimum conditions. - M.: Science.-1976, 279p.

5. Poliakov B.N. Increase quality of the technologies, load-carrying capacity of designs, durability of the equipment and efficiency of the automatic systems of rolling mills. ISBN 5-98947-023 –1. - SPb.: Renome, 2006, - 528p.

6. Kygelev V.S., Poliakov B.N. Definition of power losses in compaction of a rod of the drilling pump. Calculation and design of the oil-field equipment. Jobs VNIINEFTEMASHa.– M.: Mechanical engineering, 1975, 296 p.

7. Karlinskaia F.M., Makarov J.D., Poliakov B.N. Algorithms of statistical processing of the data of observations by means of the COMPUTER: Ref.inform. 15-70-5 / NIIinformtiazhmash. M., 1970, 46p.

Table 1

Statistical characteristics of parameters of the equations of regression at multifactorial experiment and at its planning

Experiment. Volume of sample	Kind the equations regressesion	y; x_i	x_{min}	x_{max}	$\bar{y}; \bar{x}_i$	V_{X_i} %	r_{1i}	R	S_R	R_1	R_2	S	a_i	$\alpha_{\pm 0,95}$
Multi factorial experiment	$y = a_0 + \sum_1^4 a_i x_i$	y	18	888,0	271,6	67,7	-						24,1	34,52÷13,7
		x_1	2,5	254	105	73,6	0,524						1,18	1,32÷1,05
		x_2	1,0	5,0	3,12	45,6	0,163	0,592	0,023	0,547	0,638	148,3	17,43	24,8÷10,03
		$x_3 \cdot 10^2$	5,0	85,0	45,0	62,8	-0,136						-0,72	-0,35÷-1,08
n = 787		$x_4 \cdot 10^{-2}$	8,0	48,0	23,8	52,2	0,332						4,25	5,1÷3,41
	$\ell n y = a_0 + \sum_1^4 a_i \ell n x_i$	y	2,9	6,79	5,32	15,5	-						2,99	3,03÷2,94
		x_1	0,9	5,54	3,96	40,3	0,500	0,632	0,021	0,590	0,674	0,638	0,23	0,26÷0,20
		x_2*	0	1,61	1,0	56,7	0,049						0,056	0,136÷-0,02
		$x_3 . 10^2$	1,6	4,44	3,45	29,2	-0,167						-0,11	-0,06÷-0,15
		$x_4 . 10^{-2}$	2,1	3,87	3,01	19,9	0,473						0,576	0,65÷0,50
Planning -Latin square n = 25	$y = a_0 + \sum_1^4 a_i x_i$	y	18	775	252,6	69,8	-						45,72	98,44÷-7,01
		x_1	2,5	202	96,06	71,8	0,641	0,731	0,104	0,527	0,935	120,3	1,457	2,22÷0,69
		x_2*	1,0	5,0	3,0	47,1	0,127						10,89	48,18÷-26,4
		$x_3 . 10^2$	5,0	85,0	45,0	62,8	-0,40						-1,88	-0,014÷-3,743
		$x_4 . 10^{-2}$	8,0	40,0	24,0	47,1	0,422						4,95	9,608÷0,284
	$\ell n y = a_0 + \sum_1^4 a_i \ell n x_i$	y	2,9	6,65	5,24	16,3	-						3,156	3,39÷2,29
		x_1	0,9	5,31	3,89	40,6	0,652	0,778	0,088	0,605	0,951	0,537	0,292	0,441÷0,143
		x_2*	0	1,61	0,96	59,3	-0,03						-0,03	0,382÷-0,445
		$x_3 . 10^2$	1,6	4,44	3,45	29,2	-0,48						-0,29	-0,058÷-0,525
		$x_4 . 10^{-2}$	2,1	3,69	3,04	18,7	0,568						0,65	1,065÷0,24

*- inconsequential factor

Conventional designations: n – volume of sample; x_{min}, x_{max} - ranges of change of factors; $\bar{y}, \bar{x}_i$ - arithmetic average values y and x_i; V_{X_i} - variation coefficient; r_{1i} - partial correlation coefficient; R - multiple correlation coefficient; S_R - root-mean-square deviation R; R_1 и R_2 - the bottom and top borders 95 %-s confidential intervals R; S - standard error estimation; a_i - regression coefficients; $\alpha_{\pm 0,95}$ - confidential intervals of regression coefficients at confidential probability α = 0,05.

Table 2

Comparative statistical estimations of parameters of the response functions
at multifactorial experiment and at planning

Statistical estimations		Active multifactorial experiment		Planning of experiment (A Latin square)	
		$y = a_0 + \sum_{1}^{4} a_i x_i$	$\ell n y = a + \sum_{1}^{4} a_i \ell n x_i$	$y = a_0 + \sum_{1}^{4} a_i x_i$	$\ell n y = a_0 + \sum_{1}^{4} a_i \ell n x_i$
S_R / R,%		3,9	3,3	14,2	11,3
$\Delta R / R$,%		15,4	13,3	55,8	44,5
$\Delta a_i / a_i$,%	a_0	86,3	3	230,7	34,9
	a_1	22,8	24,6	105	102
	a_2	84,9	284*	684,7*	2668*
	a_3	102,5	83,2	198,6	160
	a_4	39,8	26,04	188,5	127
$S / \bar{y}$,%		54,6	12	47,6	10,2

* - inconsequential factor