

Машиностроение Mechanical engineering

Научная статья

Статья в открытом доступе

УДК 303.724.32

doi: 10.30987/2782-5957-2024-4-4-12

ПОВЫШЕНИЕ ТОЧНОСТИ РЕГРЕССИОННЫХ МОДЕЛЕЙ МЕТОДАМИ ПОЛИНОМИАЛЬНОЙ АППРОКСИМАЦИИ

Светлана Викторовна Степошина✉

Брянский государственный технический университет, Брянск, Россия
Sve-steposhina@yandex.ru, <https://orcid.org/0009-0000-7547-7557>

Аннотация

Цель исследования: оценить степень влияния метода аппроксимации на точность регрессионной модели.

Задача: повышение точности регрессионных моделей методами полиномиальной аппроксимации.

Методы исследования: метод наименьших квадратов, метод полиномиальной аппроксимации.

Новизна работы: разработаны зависимости для проведения регрессионного анализа многофакторного эксперимента с применением полиномиальной аппроксимации, проведен анализ влияния метода аппроксимации на точность регрессионной модели.

Результаты исследования: в результате применения разработанных зависимостей точность однофакторных уравнений регрессии повысилась в 2,4 раза, двухфакторных – 1,7 раза.

Выводы: применение полиномиальной аппроксимации позволяет повысить точность однофакторных и многофакторных регрессионных моделей, что увеличивает их адекватность, значимость коэффициентов и расширяет возможности их применения.

Ключевые слова: анализ, методы, линеаризация модели, аппроксимация, интерполяция, экстраполяция.

Ссылка для цитирования:

Степошина С.В. Повышение точности регрессионных моделей методами полиномиальной аппроксимации / С.В. Степошина // Транспортное машиностроение. – 2024. – № 4. – С.4-12. doi: 10.30987/2782-5957-2024-4-4-12.

Original article

Open Access Article

IMPROVING THE ACCURACY OF REGRESSION MODELS BY METHODS OF POLYNOMIAL APPROXIMATION

Svetlana Viktorovna Steposhina✉

Bryansk State Technical University, Bryansk, Russia
Sve-steposhina@yandex.ru, <https://orcid.org/0009-0000-7547-7557>

Abstract

The study objective is to assess the influence of the approximation method on the accuracy of the regression model.

The task is to increase the accuracy of regression models using polynomial approximation methods.

Research methods: the least square method, the polynomial approximation method.

The novelty of the work: dependencies for the regression analysis of a multifactor experiment using polynomial approximation is developed, the influence of the approximation method on the accuracy of the regression model is analyzed.

The study results: as a result of applying the developed dependencies, the accuracy of single-factor

regression equations increased by 2.4 times, two-factor equations - by 1.7 times.

Conclusions: the use of polynomial approximation makes it possible to increase the accuracy of single-factor and multifactor regression models, which

increases their adequacy, the significance of coefficients and expands the possibilities of their application.

Keywords: analysis, methods, model linearization, approximation, interpolation, extrapolation.

Reference for citing:

Steposhina SV. Improving the accuracy of regression models by methods of polynomial approximation. *Transport Engineering*. 2024;4:4-12. doi: 10.30987/2782-5957-2024-4-4-12.

Введение

В результате проведения эмпирических исследований в технологии машиностроения и других областях науки выявляется набор дискретных данных, количественно описывающий исследуемый процесс. Для дальнейшего применения количественных результатов эксперимента необходимо их привести к аналитической форме, провести регрессионный анализ и получить адекватное уравнение регрессии. Для приведения экспериментальных данных к аналитической форме применяют математические методы аппроксимации и интерполяции. Аппроксимацией (приближением) функции называется нахождение такой функции (аппроксимирующей функции), которая была бы близка заданной [1]. Интерполяцией называют такую разновидность аппроксимации, при которой кривая построенной функции определяется по дискретному набору точек и проходит точно через них [1]. После нахождения интерполяционного полинома, допустимо определить значение функции даже за пределами заданного интервала (провести экстраполяцию функции) [1].

Для проведения регрессионного анализа предварительно необходимо представить экспериментальные данные графически, чтобы определиться с видом регрессионной модели (линейный или нелиней-

ный). В классической теории регрессионного анализа широко применяются метод наименьших квадратов и линеаризация регрессионных моделей с помощью логарифмирования. Метод наименьших квадратов заключается в минимизации суммы квадратов отклонений экспериментальных и теоретических данных. В случае, когда распределение экспериментальных данных имеет нелинейный вид, широко применяют линеаризацию регрессионной модели методом логарифмирования. Линеаризация нелинейной модели заключается в том, что с помощью логарифмирования ее упрощают и аналитически приводят к линейному виду, что позволяет далее для обработки данных применять классические зависимости для линейной регрессии. Но данный подход вносит значительную погрешность в расчет выходного параметра по регрессионной зависимости, т.к. после линеаризации методом логарифмирования минимизируются не натуральные значения параметров процесса, а их логарифмы, которые имеют меньший порядок величин.

Так как большинство процессов описываются нелинейными зависимостями, возникает задача повышения точности регрессионных моделей, для решения которой применимы методы полиномиальной аппроксимации.

Материалы, модели, эксперименты и методы

В ходе проведения теоретического исследования были использованы экспериментальные данные из источника [2]. Для расчета уравнения однофакторной регрессии, описывающего зависимость качества поверхностного слоя детали (коэффициент упрочнения k , параметр шероховатости Ra) от силы обкатывания P , исследовались экспериментальные данные обра-

ботки деталей обкатыванием. В ходе натурального эксперимента в работе [2] обрабатывались наружные цилиндрические поверхности образцов из стали 40Х твердости 25HRCэ [2]. В качестве инструмента использовалась шариковая накатка радиусом с радиусом индентора 5мм. Результаты натурального эксперимента представлены в табл. 1.

Для расчета уравнения многофакторной регрессии, описывающего зависимость коэффициента упрочнения k от силы выглаживания P и продольной подачи S_0 , исследовались экспериментальные данные обработки деталей алмазным выглаживанием. В ходе натурального эксперимента об-

рабатывались внутренние цилиндрические поверхности образцов из стали 40Х, твердость 28HRCэ [2]. В качестве инструмента использовался алмазный выглаживатель с радиусом индентора 1,5 мм. Результаты натурального эксперимента представлены в табл. 2.

Таблица 1

Результаты обкатывания

Table 1

Running-in results

№ опыта	Сила, P , Н	Вид обработки	k	Ra , мкм	Rz , мкм
1	30	отделочная	1,09	0,471	1,1
2	80	отделочно-упрочняющая	1,3	0,214	1,34
3	300	упрочняющая	1,35	0,153	0,68
4	800	перенаклеп	1,44	0,131	0,73
5	1600	перенаклеп	1,1	0,195	4,02

Таблица 2

Результаты алмазного выглаживания

Table 2

Diamond smoothing results

№ опыта	Сила, P , Н	Подача, S_0 , мм/об	Вид обработки	k
1	20	0,08	отделочная	1,1
2	70	0,06	отделочно-упрочняющая	1,14
3	150	0,05	отделочно-упрочняющая	1,34
4	300	0,05	упрочняющая	1,73

Если уравнение регрессии рассчитывается в линейном виде, то выходной параметр $y(x)$ имеет вид [3, 4]:

$$y(x) = b_0 + b_1 x. \quad (1)$$

Результатом аппроксимации функции методом наименьших квадратов является расчет таких коэффициентов b_0 и b_1 , чтобы минимизировалась сумма квадратов отклонений экспериментальных данных y_i и значений функции по уравнению регрессии (1):

$$F = \sum_{i=1}^n (b_0 + b_1 x_i - y_i)^2 \rightarrow \min.$$

Для нахождения экстремумов функции находят частные производные функции, приравнивают их нулю и получают систему уравнений [Данилов]:

$$\begin{cases} \frac{\partial F}{\partial b_0} = 0, \\ \frac{\partial F}{\partial b_1} = 0. \end{cases}$$

Находим частные производные:

$$\begin{cases} 2 \sum_{i=1}^n (b_0 + b_1 x_i - y_i) = 0, \\ 2 \sum_{i=1}^n (b_0 + b_1 x_i - y_i) x_i = 0. \end{cases}$$

После преобразования получим:

$$\begin{cases} n b_0 + b_1 \sum_{i=1}^n x_i = \sum_{i=1}^n y_i, \\ b_0 \sum_{i=1}^n x_i + b_1 \sum_{i=1}^n x_i^2 = \sum_{i=1}^n (x_i y_i). \end{cases}$$

После решения данной системы уравнений методом подстановки получают классические зависимости для нахождения коэффициентов линейной регрессии [3, 4]:

$$\left\{ \begin{array}{l} b_1 = \frac{n \sum_{i=1}^n (x_i y_i) - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2}, \\ b_0 = \frac{\sum_{i=1}^n y_i - b_1 \sum_{i=1}^n x_i}{n}. \end{array} \right.$$

Если уравнение регрессии рассчитывается в степенном виде, то выходной параметр $y(x)$ имеет вид [3, 4]:

$$y(x) = b_0 x^{b_1} \quad (2)$$

Для линеаризации уравнения (2) его логарифмируют и далее для определения коэффициентов регрессии используют методику, приведенную выше.

В данной работе исследуется влияние на точность уравнения регрессии применение полиномиальной аппроксимации. Если графически экспериментальные данные распределены нелинейно, то уравнение регрессии допустимо искать в виде полинома второго порядка [3, 4]:

$$y(x) = b_2 x^2 + b_1 x + b_0 \quad (3)$$

Тогда функция минимизации квадратов отклонений будет иметь вид:

$$F = \sum_{i=1}^n (b_2 x_i^2 + b_1 x_i + b_0 - y_i)^2 \rightarrow \min.$$

$$\left\{ \begin{array}{l} \frac{\partial F}{\partial b_2} = 0, \\ \frac{\partial F}{\partial b_1} = 0, \\ \frac{\partial F}{\partial b_0} = 0; \end{array} \right.$$

Находим частные производные:

$$\left\{ \begin{array}{l} 2 \sum_{i=1}^n ((b_2 x_i^2 + b_1 x_i + b_0 - y_i) x_i^2) = 0, \\ 2 \sum_{i=1}^n ((b_2 x_i^2 + b_1 x_i + b_0 - y_i) x_i) = 0, \\ 2 \sum_{i=1}^n (b_2 x_i^2 + b_1 x_i + b_0 - y_i) = 0; \end{array} \right.$$

После преобразования получим:

$$\left\{ \begin{array}{l} b_2 \sum_{i=1}^n x_i^4 + b_1 \sum_{i=1}^n x_i^3 + b_0 \sum_{i=1}^n x_i^2 = \sum_{i=1}^n (y_i x_i^2), \\ b_2 \sum_{i=1}^n x_i^3 + b_1 \sum_{i=1}^n x_i^2 + b_0 \sum_{i=1}^n x_i = \sum_{i=1}^n (y_i x_i), \\ b_2 \sum_{i=1}^n x_i^2 + b_1 \sum_{i=1}^n x_i + b_0 n = \sum_{i=1}^n y_i. \end{array} \right.$$

Таким образом, получим систему трех линейных уравнений с тремя неизвестными, в результате решения которой найдем коэффициенты регрессии b_0 и b_1 и b_2 .

Для многофакторных экспериментов уравнение полиномиальной регрессии ищем в виде:

$$y(x) = b_1 x_1^2 + b_2 x_2^2 + b_3 x_1 + b_4 x_2 + b_0 \quad (4)$$

Тогда функция минимизации квадратов отклонений будет иметь вид:

$$F = \sum_{i=1}^n (b_1 x_i^2 + b_2 x_i^2 + b_3 x_i + b_4 x_i + b_0 - y_i)^2 \rightarrow \min$$

$$\left\{ \begin{array}{l} \frac{\partial F}{\partial b_1} = 0, \\ \frac{\partial F}{\partial b_2} = 0, \\ \frac{\partial F}{\partial b_3} = 0, \\ \frac{\partial F}{\partial b_4} = 0, \\ \frac{\partial F}{\partial b_0} = 0; \end{array} \right.$$

$$\begin{cases} 2 \sum_{i=1}^n ((b_1 x_1^2 + b_2 x_2^2 + b_3 x_1 + b_4 x_2 + b_0 - y_i) x_1^2) = 0, \\ 2 \sum_{i=1}^n ((b_1 x_1^2 + b_2 x_2^2 + b_3 x_1 + b_4 x_2 + b_0 - y_i) x_2^2) = 0, \\ 2 \sum_{i=1}^n ((b_1 x_1^2 + b_2 x_2^2 + b_3 x_1 + b_4 x_2 + b_0 - y_i) x_1) = 0, \\ 2 \sum_{i=1}^n ((b_1 x_1^2 + b_2 x_2^2 + b_3 x_1 + b_4 x_2 + b_0 - y_i) x_2) = 0, \\ 2 \sum_{i=1}^n (b_1 x_1^2 + b_2 x_2^2 + b_3 x_1 + b_4 x_2 + b_0 - y_i) = 0; \end{cases}$$

После преобразования получим:

$$\begin{cases} b_1 \sum_{i=1}^n x_1^4 + b_2 \sum_{i=1}^n (x_1^2 x_2^2) + b_3 \sum_{i=1}^n x_1^3 + b_4 \sum_{i=1}^n (x_1^2 x_2) + b_0 \sum_{i=1}^n x_1^2 = \sum_{i=1}^n (y_i x_1^2), \\ b_1 \sum_{i=1}^n (x_1^2 x_2^2) + b_2 \sum_{i=1}^n x_2^4 + b_3 \sum_{i=1}^n x_1 x_2^2 + b_4 \sum_{i=1}^n x_2^3 + b_0 \sum_{i=1}^n x_2^2 = \sum_{i=1}^n (y_i x_2^2), \\ b_1 \sum_{i=1}^n x_1^3 + b_2 \sum_{i=1}^n (x_1 x_2^2) + b_3 \sum_{i=1}^n x_1^2 + b_4 \sum_{i=1}^n (x_1 x_2) + b_0 \sum_{i=1}^n x_1 = \sum_{i=1}^n (y_i x_1), \\ b_1 \sum_{i=1}^n (x_1^2 x_2) + b_2 \sum_{i=1}^n x_2^3 + b_3 \sum_{i=1}^n (x_1 x_2) + b_4 \sum_{i=1}^n x_2^2 + b_0 \sum_{i=1}^n x_2 = \sum_{i=1}^n (y_i x_2), \\ b_1 \sum_{i=1}^n x_1^2 + b_2 \sum_{i=1}^n x_2^2 + b_3 \sum_{i=1}^n x_1 + b_4 \sum_{i=1}^n x_2 + n b_0 = \sum_{i=1}^n y_i. \end{cases}$$

Таким образом, получим систему из пяти линейных уравнений с пятью неизвестными, в результате решения которой

найдем коэффициенты регрессии b_0, b_1, b_2, b_3, b_4 .

Результаты

В первой части вычислительного эксперимента регрессионный анализ проводился с помощью приложения *MS Excel* и *SMath Solver*. С применением стандарт-

ных функций были получены уравнения степенной регрессии (2) на базе МНК и линеаризации логарифмированием (рис. 1).

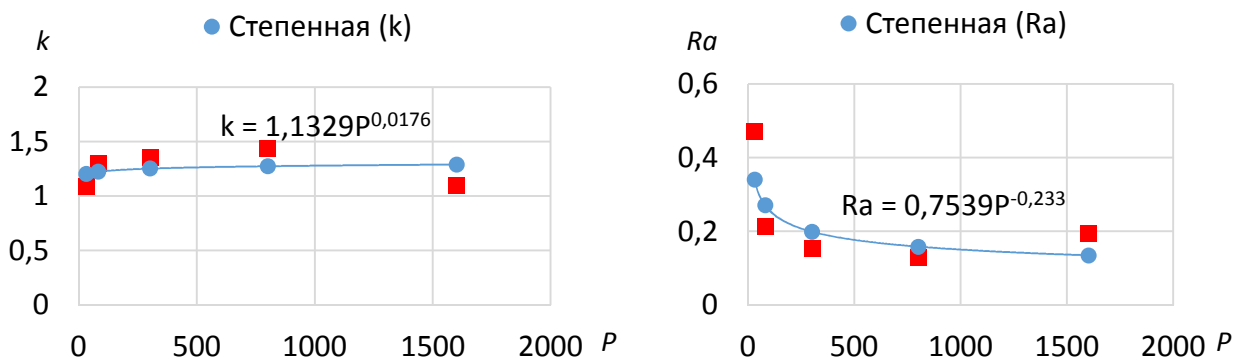


Рис. 1. Графики уравнений регрессии в степенной форме

■ – экспериментальные данные; ● – теоретические данные

Fig. 1. Graphs of regression equations in power form

■ – experimental data; ● – theoretical data

Во второй части вычислительного эксперимента регрессионный анализ про-

водился с помощью полиномиальной регрессии (3) (рис. 2).

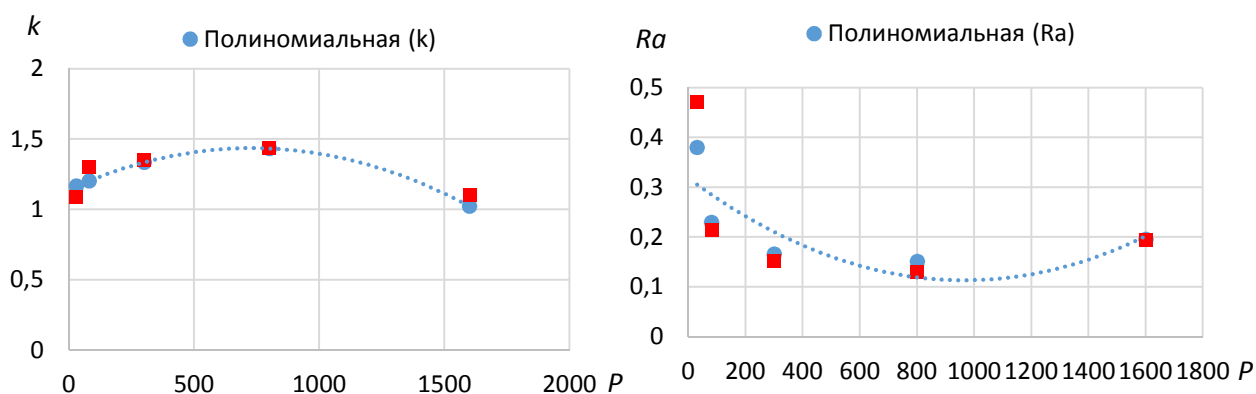


Рис. 2. Графики уравнений регрессии в полиномиальной форме

■ – экспериментальные данные; ● – теоретические данные

Fig. 2. Graphs of regression equations in polynomial form

■ – experimental data; ● – theoretical data

Результаты проведения вычислительного эксперимента представлены в таблицах 3-7.

Для обкатывания полученные уравнения степенной регрессии имеют вид:

$$k(P) = 1,1329P^{0,0177},$$

где $k(P)$ – коэффициент упрочнения, P , Н – сила обкатывания.

$$Ra(P) = 0,7539P^{-0,233},$$

где $Ra(P)$, мкм – среднеарифметическое отклонение профиля шероховатости.

Таблица 3

Результаты вычислительного эксперимента методом линейной аппроксимации (обкатывание)

Table 3

Results of a computational experiment using the linear approximation method (running-in)

№ опыта	Сила, P , Н	k	Ra , мкм	k_T	$\Delta k_{абс}$	$\Delta k_{отн}, \%$	Ra_T , мкм	$\Delta Ra_{абс}, \text{мкм}$	$\Delta Ra_{отн}, \%$
1	30	1,09	0,471	1,20	0,11	10,35	0,34	0,13	27,54
2	80	1,3	0,214	1,22	0,08	5,87	0,27	0,06	26,91
3	300	1,35	0,153	1,25	0,10	7,22	0,12	0,05	30,45
4	800	1,44	0,131	1,27	0,17	11,50	0,16	0,03	21,23
5	1600	1,1	0,195	1,29	0,19	17,27	0,14	0,06	30,70
Среднее значение					0,11	10,35		0,13	27,54

Для обкатывания полученные уравнения полиномиальной регрессии имеют вид:

$$k(P) = -5,4672 \cdot 10^{-4} P^2 + 0,0008P + 1,1429,$$

$$Ra(P) = 6,4626 \cdot 10^{-8} P^2 - 0,0001P + 0,1902,$$

где $k(P)$ – коэффициент упрочнения;
 $Ra(P)$, мкм – среднеарифметическое от-

клонение профиля шероховатости; P , Н – сила обкатывания.

Таблица 4

Результаты вычислительного эксперимента методом полиномиальной аппроксимации
(обкатывание)

Table 4

*Results of a computational experiment using the polynomial approximation method
(running-in)*

№ опыта	Сила, P, Н	k	Ra, мкм	k _т	Δk _{абс}	Δk _{отн} , %	Ra _т , мкм	ΔRa _{абс} , мкм	ΔRa _{отн} , %
1.	30	1,09	0,471	1,17	0,08	7,01	0,38	0,28	60,24
2.	80	1,3	0,214	1,20	0,10	7,43	0,23	0,03	14,67
3.	300	1,35	0,153	1,33	0,02	1,21	0,17	0,01	8,51
4.	800	1,44	0,131	1,43	0,01	0,49	0,15	0,02	15,70
5.	1600	1,1	0,195	1,02	0,08	6,97	0,20	0,001	0,33
Среднее значение					0,05	4,62		0,03	10,26

Для алмазного выглаживания отверстий полученное уравнение степенной регрессии имеет вид:

$$k(P, S_0) = 11,02P^{0,42} \cdot S_0^{1,42},$$

где $k(P, S_0)$ – коэффициент упрочнения; P, Н – сила выглаживания; S₀, мм/об – продольная подача.

Таблица 5

Результаты вычислительного эксперимента методом линейной аппроксимации
(алмазное выглаживание отверстий)

Table 5

*Results of a computational experiment using the linear approximation method
(diamond hole smoothing)*

№ опыта	Сила, P, Н	Подача, S ₀ , мм/об	k	k _т	Δk _{абс}	Δk _{отн} , %
1	20	0,08	1,1	1,09	0,01	0,88
2	70	0,06	1,14	1,19	0,05	4,47
3	150	0,05	1,34	1,24	0,10	7,57
4	300	0,05	1,73	1,79	0,06	3,21
Среднее значение					0,05	4,03

Для алмазного выглаживания отверстий полученное уравнение полиномиальной регрессии имеет вид:

$$k(P, S_0) = 7,485 \cdot 10^{-6} P^2 - 0,151 S_0^2 + 0,0001P + 0,3742 S_0 + 1,067.$$

Таблица 6

Результаты вычислительного эксперимента методом полиномиальной аппроксимации
(алмазное выглаживание отверстий)

Table 6

*Results of a computational experiment using the polynomial approximation method
(diamond hole smoothing)*

№ опыта	Сила, P, Н	Подача, S ₀ , мм/об	k	k _т	Δk _{абс}	Δk _{отн} , %
1	20	0,08	1,1	1,10	0,001	0,088
2	70	0,06	1,14	1,13	0,007	0,650
3	150	0,05	1,34	1,27	0,071	5,318
4	300	0,05	1,73	1,79	0,059	3,409
Среднее значение					0,03	2,37

Таблица 7

Сравнение погрешностей регрессии, полученных методами линейной и полиномиальной аппроксимации

Table 7

Comparison of regression errors obtained by linear and polynomial approximation methods

№ п/п	Вид обработки	Вид аппроксимации	$\overline{\Delta k_{abc}}$	$\overline{\Delta k_{отн}}, \%$	$\overline{\Delta Ra_{abc}}, \text{мкм}$	$\overline{\Delta Ra_{отн}}, \%$
1.	Обкатывание	Линейная	0,11	10,35	0,13	27,54
2.		Полиномиальная	0,05	4,62	0,03	10,26
3.	Алмазное выглаживание от- верстий	Линейная	0,05	4,03	-	-
4.		Полиномиальная	0,03	2,37	-	-

Заключение

В результате проведения вычислительного эксперимента доказано, что применение полиномиальных уравнений регрессии повышает точность расчетов (табл. 7). Для классического (однофакторного эксперимента) относительная погрешность уравнения регрессии, полученного виде полинома 2 рода, в 2,4 раза (с 27,54 % до 10,26 %) меньше погрешности уравнения, полученного с помощью линеаризации. Для многофакторного эксперимента применение полиномиального уравнения регрессии позволяет снизить относительную погрешность отклонения теоретических данных от экс-

периментальных в 1,7 раза (с 4,03 % до 2,37 %). Таким образом, можно сделать вывод, что для обработки классических (однофакторных) экспериментов, целесообразно применять полиномиальные уравнения регрессии. Для многофакторных экспериментов применение полиномиальных уравнений регрессии существенно повышает вычислительную сложность алгоритма, поэтому вопрос о целесообразности их применения исследователь должен принимать с учетом конкретных требований к точности регрессионной модели.

СПИСОК ИСТОЧНИКОВ

1. Алгоритмизация в инженерных задачах : учебное пособие /С. В. Степошина, О. Н. Федонин, С. Ю. Съянов. – Москва ; Вологда : Инфра-Инженерия, 2023. – 116 с.
2. Степошина, С.В. Выбор рациональных условий обкатывания и алмазного выглаживания внутренних цилиндрических поверхностей деталей из конструкционных легированных сталей. [Текст]: дис. ... канд. техн. наук: 05.02.07: защищена 20.06.13: утв. 09.12.13 /

Степошина Светлана Викторовна. - М., 2013. – 149 с.

3. Данилов А.М. Интерполяция, аппроксимация, оптимизация: анализ и синтез сложных систем: моногр. / А.М. Данилов, И.А. Гарькина. – Пенза: ПГУАС, 2014. – 168 с.
4. Львовский, Е.Н. Статистические методы построения эмпирических формул/ Е.Н. Львовский.- М.:Выш.школа, 1982. – 224 с.

REFERENCES

1. Steposhina SV, Fedonin ON, Syanov SYu. Algorithmization in engineering problems: textbook. Moscow; Vologda: Infra-Enzheneria; 2023.
2. Steposhina SV. Choice of rational conditions for rolling and diamond smoothing of internal cylindrical surfaces of parts made of structural alloy steels [dissertation]. [Moscow (RF)]; 2013.

3. Danilov AM, Garkina IA. Interpolation, approximation, optimization: analysis and synthesis of complex systems: monograph. Penza: PGUAS; 2014.
4. Lvovsky EN. Statistical methods for constructing empirical formulas. Moscow: Visshaya Skola; 1982.

Информация об авторе:

Степошина Светлана Викторовна – кандидат технических наук, тел. 8-953-273-25-14, доцент кафедры «Автоматизированные технологические

системы» Брянского государственного технического университета.

Steposhina Svetlana Viktorovna – Candidate of Technical Sciences, phone: 8-953-273-25-14, Associate Professor of the Department of Automated

Technological Systems at Bryansk State Technical University.

**Статья опубликована в режиме Open Access.
Article published in Open Access mode.**

Статья поступила в редакцию 01.03.2024; одобрена после рецензирования 13.03.2024; принята к публикации 26.03.2024. Рецензент – Сьянов С.Ю., кандидат технических наук, доцент Брянского государственного технического университета, заместитель главного редактора журнала «Транспортное машиностроение».

The article was submitted to the editorial office on 01.03.2024; approved after review on 13.03.2024; accepted for publication on 26.03.2024. The reviewer is Syanov S.Yu., Candidate of Technical Sciences, Associate Professor of Bryansk State Technical University, Deputy Editor-in-Chief of the journal *Transport Engineering*.