The mathematical model of the multivariate regression analysis carrying out is presented. Its practical application for rolled production is examined. Analysis of the model adequacy is carried out.

А. Н. ЧИЧКО, БНТУ, Л. А. ФЕКЛИСТОВА, В. И. ЩЕРБАКОВ, А. В. ВЕДЕНЕЕВ, РУП «БМЗ»

УДК 669.21.

ФАКТОРНЫЙ АНАЛИЗ ТЕХНОЛОГИИ ПРОЦЕССА ПРОКАТКИ (Часть 2)

Рассмотренный множественный регрессионный анализ в части 1 (А. Н. Чичко, Л. А. Феклистова, В. И. Щербаков, А. В. Веденеев. Факторный анализ технологии процесса прокатки. Ч. 1 // Литье и металлургия. 2011. № 1. С. 90–97) показал возможность применения подобного метода при выборе специальных характеристик. Напомним, что *специальные характеристики* – это характеристики продукции и процессов, назначенные потребителем, описывающие безопасность и правительственные нормы и/или назначенные поставщиком благодаря знаниям о продукции и процессе, требующие мониторинга и внесения в планы управления и другие технологические документы. Цель проведенного многомерного регрессионного анализа по определению влияния химического состава стали на величину прочности, относительного сужения и относительного удлинения стали 65К, предназначенной для изготовления металлокорда, достигнута. Поставленная задача построения математической модели и выбор одного или нескольких оптимальных параметров из многих выходных параметров выполнена, что подтверждено проведенными расчетами по статистическим данным. Предположения о характере распределения и зависимости в исследуемом направлении подтверждают нормальность распределения характеристик и гауссову модель линейной регрессии.

Задача второй части статьи – проверка адекватности модели и определение погрешности проведенных расчетов.

Основные методологические трудности возникают в ходе проверки выполнения априорных предпосылок регрессионного анализа и последующей оценки адекватности полученного уравнения. Поэтому выполнение этой задачи является не менее важной частью в сравнении с непосредственным регрессионным анализом. При проверке правильности спецификации нулевая гипотеза всегда состоит в том, что модель специфицирована корректно, а альтернативная гипотеза – в том, что имеется ошибка соответствующей спецификации.

Для наглядности в настоящей статье приведена математическая модель (см. рисунок) и рассмотрен пошагово алгоритм действий по проверке адекватности этой модели.

Начиная с шага 8 проведем вычисления, позволяющие оценить полученную модель.

Шаг 8. Вычислим значимость коэффициентов регрессии по критерию Стьюдента, так же как в [1, 6]. Данные для вычисления критериев приведены в таблице:

$$t = \frac{b_j \sqrt{n}}{\sigma_{yy}},$$

где σ_{yy} – ошибка воспроизводимости данных:

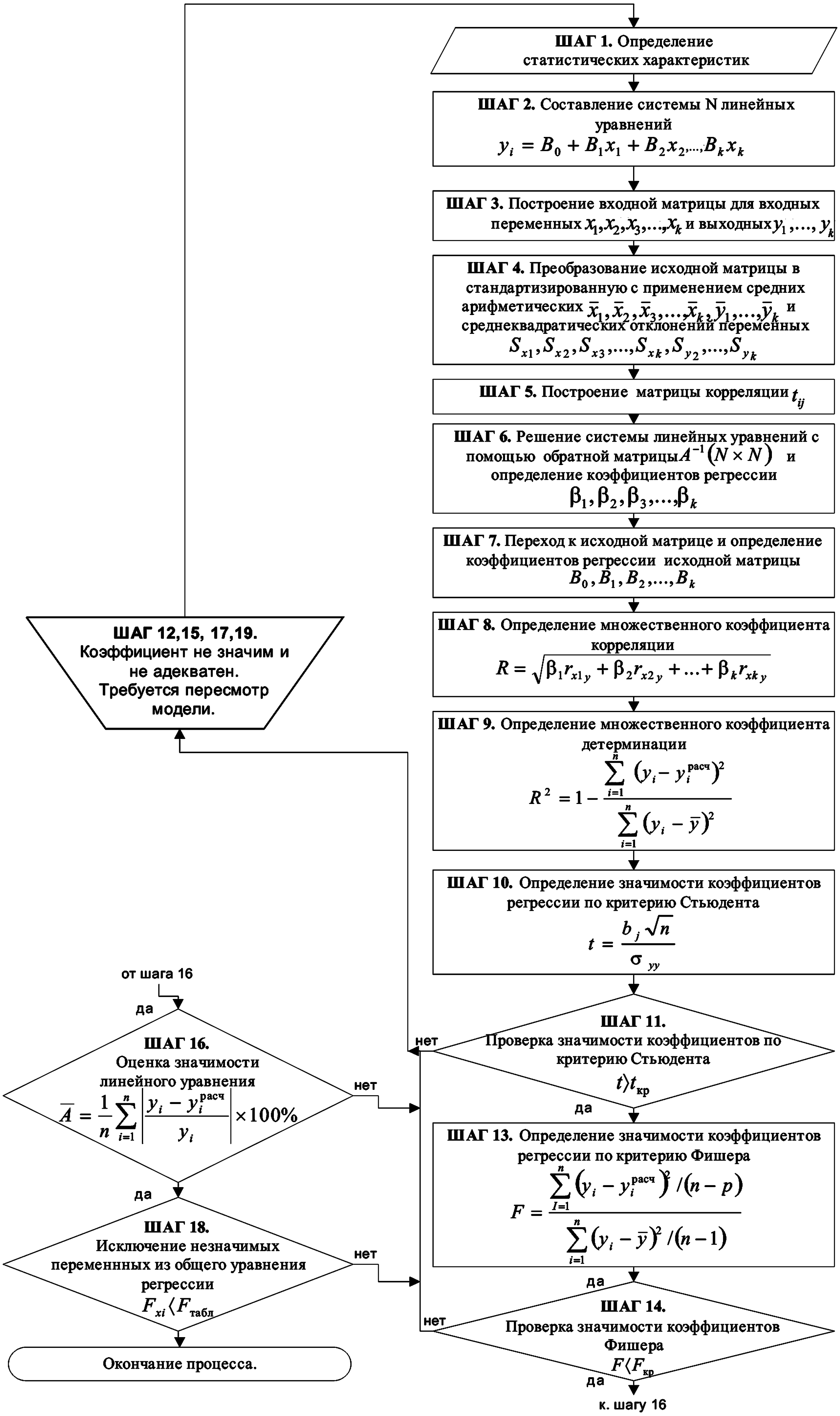
$$\sigma_{yy} = \sqrt{\frac{\sum_{i=1}^{n} y_i^2 - n \sum_{j=0}^{k} b_j^2}{n-k-1}}.$$

Здесь k – число регрессионных коэффициентов без нулевого; b_j – коэффициент регрессии; n – число наблюдений.

Проверка значимости коэффициентов, которая проводится по неравенству $|t| \geq t_{\text{кр}}$, показала, что

$$t_{0,05;88} \langle t \langle t_{0,05;88}, \; 2{,}06 \langle t \langle 2{,}06,$$

где $t_{0,05;88} = 2{,}06$ взято из таблицы [9, 10]. Так как условие $t_i \rangle t_{0,05;88}$ выполняется (табл. 5, часть 1), то все коэффициенты, влияющие на прочность, являются значимыми кроме химического элемента азота P; для относительного сужения все коэффициенты являются значимыми; для относительного удлинения – все рассмотренные кроме S, Ni, N_2.



Алгоритм определения специальных характеристик с помощью математической модели

Шаг 9. Вычислить значение критерия Фишера, используя формулу [1, 6, 9, 10]:

$$F = \frac{\sum_{I=1}^{n} (y_i - y_i^{\text{расч}})^2 / (n-k)}{\sum_{i=1}^{n} (y_i - \overline{y})^2 / (n-1)}.$$

Критерий Фишера при α = 0,05 для числа степеней свободы $k_1 = 79$ (число степеней свободы для числителя) и k_2 = 88 (число степеней свободы для знаменателя) находится из таблицы и равен $F_{\text{кр}}$ = 1,43. Наблюдаемые значения коэффициента Фишера $F_{\text{набл}}$ приведены в таблице. В этом случае для проверки адекватности используется неравенство $F_{\text{набл}} \langle F_{\text{кр}}$. При его выполнении модель считается значимой и адекватной.

Тогда полученные значения $F_{\text{набл}} \langle$ 1,43 с вероятностью $\gamma = 0,95$ могут свидетельствовать об адекватности построенной регрессии. Таким образом, регрессионные уравнения адекватны и имеют вид

$$y(\sigma_{\text{в}}) = 1177,5 - 116,1x_1 + 297,3x_2 - 355,8x_3 + 4723,5x_4 + 916,5x_5 + 204,8x_6 - 1237,2x_7 + 86,9x_8 + 146,5x_9, \quad (1)$$

$$y(\psi) = 109,7 - 84,8x_1 + 19,8x_2 - 7,8x_3 - 193,3x_4 + 45,8x_5 - 34,6x_6 + 101,9x_7 - 58,8x_8 - 437,4x_9, \quad (2)$$

$$y(\delta) = 36,1 - 13,1x_1 - 20,5x_2 - 14,2x_3 - 162,9x_4 + 105,5x_5 + 41,8x_6 + 82,2x_7 - 80,1x_8 + 137,2x_9. \quad (3)$$

Свободные коэффициенты при независимых переменных указывают на силу влияния факторов. Чем больше численная величина коэффициента, тем большее влияние оказывает фактор. Если коэффициент имеет знак плюс, то с увеличением значения фактора параметр оптимизации увеличивается, а если минус, то уменьшается. Исходя из этого, можно сделать следующее заключение: в уравнении (1) существенное влияние на изменение прочности $\sigma_{\text{в}}$ оказывает фактор x_4 (усиливает), которому соответствует химический элемент фосфор (P), и $x_7 - \text{Ni}$ (уменьшает); в уравнении (2) существенное влияние на изменение относительного сужения ψ оказывает фактор $x_7 - \text{Ni}$ (усиливает), а $x_9 - \text{N}_2$ (уменьшает); в уравнении (3) существенное влияние на изменение относительного удлинения δ оказывает фактор $x_9 - \text{N}_2$ (усиливает), а $x_4 - \text{P}$ (уменьшает).

Шаг 10. После того, как найдено уравнение линейной регрессии, проводится оценка значимости как уравнения в целом, так и отдельных его параметров. Чтобы иметь общее суждение о качестве модели из относительных отклонений по каждому наблюдению, определяют среднюю ошибку аппроксимации [11,12]:

$$\overline{A} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - y_i^{\text{расч}}}{y_i}\right| \cdot 100\%,$$

где $y_i^{\text{расч}}$ – теоретическое значение результативного признака, найденное исходя из уравнения регрессии. Средняя ошибка аппроксимации не должна превышать 8–10%.

Проверим уравнения регрессии $y = b_0 + b_1x_1 + b_2x_2 + ... + b_kx_k$, подставив в формулы средние зна-

Расчетная таблица для вычислений критериев Стьюдента и Фишера

Значения коэффициентов Стьюдента t											
t_1(C)	t_2(Si)	t_3(Mn)	t_4(P)	t_5(S)	t_6(Cr)	t_7(Ni)	t_8(Cu)	$t_9(\text{N}_2)$	$t_{0,05;88} = 2,06$		
–0,0651	0,2548	–0,553	7,6532	1,471	–0,2162	–1,4632	0,3267	4,2146	$t_{\text{кр}}(\sigma_b)$	–	–
0,1110	0,1675	–0,3079	–2,9940	1,1343	–0,29	1,1628	–0,59	–8,2492	–	$t_{\text{кр}}(\psi)$	–
–0,0970	–0,6658	–0,2921	–3,8370	2,5132	0,064	3,3498	–2,06	6,4932	–	–	$t_{\text{кр}}(\delta)$
Значения коэффициентов множественной корреляции R_i						Значения коэффициентов Фишера $F_{\text{кр}}$ = 1,43					
$R(\sigma_b)$		$R(\psi)$		$R(\delta)$		$F_{\text{набл}}(\sigma_b)$		$F_{\text{набл}}(\psi)$		$F_{\text{набл}}(\delta)$	
$R_1 = 0,43$		$R_2 = 0,28$		$R_3 = 0,33$		$F_1 = 0,9$		$F_2 = 1,08$		$F_3 = 0,89$	
Значения коэффициентов множественной детерминации R_i^2						–		–		–	
$R^2(\sigma_b)$		$R^2(\psi)$		$R^2(\delta)$		–		–		–	
$R_1^2 = 0,18$		$R_2^2 = 0,08$		$R_3^2 = 0,11$		–		–		–	

чения переменных $x_1, x_2, ..., x_k$ и получив при этом средние значения y_1, y_2, y_3:

$$y(\sigma_в) = 1177,5 - 116,1x_1 + 297,3x_2 - 355,8x_3 + 4723,5x_4 + 916,5x_5 + 204,8x_6 - 1237,2x_7 + 86,9x_8 + 146,5x_9 = 991,5 \text{ Н/мм}^2.$$

При этом расчетное значение прочности $y^{расч}$ = 991,46 Н/мм^2. Средняя ошибка аппроксимации составляет 0,001%. Следовательно, модель адекватна.

$$y(\psi) = 109,7 - 84,8x_1 + 19,8x_2 - 7,8x_3 - 193,3x_4 + 45,8x_5 - 34,6x_6 + 101,9x_7 - 58,8x_8 - 437,4x_9 = 48,24\%,$$

среднее значение относительного сужения $y^{расч}$ = $48,236\%$. Здесь средняя ошибка аппроксимации составляет 0,001% – модель адекватна.

$$y(\delta) = 36,2 - 13,1x_1 - 20,5x_2 - 14,2x_3 - 162,9x_4 + 105,5x_5 + 41,8x_6 + 82,2x_7 - 80,1x_8 + 137,2x_9 = 16,68\%,$$

среднее значение относительного удлинения $y^{расч}$ = $16,679\%$. Средняя ошибка аппроксимации составляет 0,001% – модель адекватна.

Таким образом, получены общие уравнения исследуемых характеристик. Задачей следующего этапа является нахождение специальных характеристик методом исключения незначимых из общего уравнения.

Шаг 11. Применяя *метод исключения* незначимых переменных из общего уравнения регрессии, можно выделить значимую и сконцентрировать внимание на ее управлении. Чтобы проверить переменные на этом шаге, необходимо определить частный F-критерий для каждой переменной $x_1, x_2, ..., x_k$ в предположении как будто данная переменная была включена в регрессионное уравнение последней.

В общем виде для фактора x_i частный F-критерий определится как [4]:

$$F_{xi} = \frac{R^2_{yx_1...x_i...x_k} - r^2_{yx_1...x_{i-1}x_{i+1}...x_k}}{1 - R^2_{yx_1...x_i...x_k}} \frac{n-k-1}{1},$$

где $R^2_{yx_1...x_i...x_k}$ – коэффициент множественной детерминации для модели с полным набором факторов; $R^2_{yx_1...x_{i-1i}x_{i+1}...x_k}$ – тот же показатель, но без включения в модель фактора x_i; n – число наблюдений; k – число параметров в модели (без свободного члена).

Фактическое значение частного F-критерия сравнивается с табличным при уровне значимости α и числе степеней свободы: 1 и $n-k-1$. Если фактическое значение F_{xi} превышает $F_{табл}(\alpha, k_1, k_2)$, то дополнительное включение фактора x_i в модель статистически оправдано и коэффициент чистой регрессии b_i при факторе x_i статистически значим. Если же фактическое значение F_{xi} меньше табличного, то дополнительное включение в модель фактора x_i не увеличивает существенно долю объясненной вариации признака y, следовательно, нецелесообразно его включение в модель; коэффициент регрессии при данном факторе в этом случае статистически незначим.

1. Определим частные F-критерии для уравнения (1):

$$y_{1.1}(\sigma_в) = 1104,5 - 238,5x_1 + 345,2x_2 + 14,4x_4 + 477,3x_5 - 222,8x_6 - 698,8x_7 - 26,9x_8 + 130,6x_9. \quad (1.1)$$

В этом уравнении исключена переменная $x_3(\text{Mn})$, так как значение F_{x3} оказалось меньше других значений в этом уравнении и составило 3,2, но больше $F_{табл}$ = 1,43. Следовательно, требуется дальнейший перерасчет уравнения.

$$y_{1.2}(\sigma_в) = 1180,7 - 387,3x_1 + 393,9x_2 + 1044,1x_5 - 310,8x_6 + 184,8x_7 - 330,4x_8 + 214,2x_9. \quad (1.2)$$

Здесь исключена переменная $x_4(\text{P})$, так как значение F_{x4} оказалось меньше других значений в этом уравнении и составило 1,8, но больше $F_{табл}$ = 1,43.

$$y_{1.3}(\sigma_в) = 1167,8 - 256,2x_1 + 1738,9x_5 - 448,1x_6 - 92,5x_7 - 212,5x_8 + 815,3x_9. \quad (1.3)$$

Здесь исключена переменная $x_2(\text{Si})$, так как значение F_{x2} оказалось меньше других значений в этом уравнении и составило 1,5, но больше $F_{табл}$ = 1,43. Следовательно, процедура исключения переменной из уравнения (1.3) продолжается.

$$y_{1.4}(\sigma_в) = 1161,7 - 230,1x_1 - 582,1x_6 + 137,5x_7 - 217,4x_8 + 1741,2x_9. \quad (1.4)$$

Исключение переменной $x_5(\text{S})$ привело к конечному решению системы уравнений, так как при перерасчете уравнения частный F-критерий переменной $x_6(\text{Cr}) = 0,6$, подлежащей исключению, стал меньше $F_{табл}$ = 1,43.

Отсюда следует, что специальными характеристиками, влияющими на прочность стали марки 65К, являются $x_1(\text{C})$, $x_6(\text{Cr})$, $x_7(\text{Ni})$, $x_8(\text{Cu})$, $x_9(\text{N}_2)$.

2. Определим частные F-критерии для уравнения (2):

$$y_{2.1}(\psi) = 55,9 + 6,1x_2 - 12,2x_3 - 113,6x_4 + 41,5x_5 - 20,9x_6 + 46,4x_7 - 33,9x_8 - 352,2x_9. \quad (2.1)$$

В этом уравнении исключена переменная x_1(C), так как значение F_{x1} оказалось меньше других значений в этом уравнении и составило 3,4, но больше $F_{\text{табл}}$ = 1,43. Следовательно, требуется дальнейший перерасчет уравнения.

$$y_{2.2}(\psi) = 49,7 + 7x_2 - 94,5x_4 + 25,9x_5 - 35,1x_6 + 62,3x_7 - 36,7x_8 - 348,5x_9. \quad (2.2)$$

Исключение переменной x_3(Mn) привело к конечному решению системы уравнений, так как при перерасчете уравнения частный F-критерий переменной $x_9(\mathrm{N}_2) = 0,9$, подлежащей исключению, стал меньше $F_{\text{табл}}$ = 1,43.

Следовательно, дальнейший перерасчет уравнения не требуется, все оставшиеся переменные можно отнести к специальным характеристикам, влияющим на относительное сужение стали марки 65К.

3. Определим частные F-критерии для уравнения (3):

$$y_{3.1}(\delta) = 27,6 + 1,8x_1 - 28,5x_2 - 16,6x_3 - 61,9x_4 + 111,3x_5 + 53,7x_6 - 35,9x_7 + 150,2x_9. \quad (3.1)$$

Переменная x_8(Cu), так как значение $F_{x8} = 10,9$ оказалось меньше других значений в этом уравнении, но больше $F_{\text{табл}}$ = 1,43.

$$y_{3.2}(\delta) = 27,4 - 7,1x_1 - 15,6x_3 - 68,9x_4 + 61,4x_5 + 63,4x_6 - 24,9x_7 + 107,7x_9. \quad (3.2)$$

Здесь переменная x_2(Si), так как значение $F_{x2} = 5,6$ оказалось меньше других значений в этом уравнении, но больше $F_{\text{табл}}$ = 1,43.

$$y_{3.3}(\delta) = 28,3 - 11,1x_1 - 11,1x_3 - 71,8x_4 + 24,9x_5 + 25,8x_7 + 174,6x_9. \quad (3.3)$$

Переменная x_6(Cr), так как значение $F_{x6} = 3,9$ оказалось меньше других значений в этом уравнении, но больше $F_{\text{табл}}$ = 1,43.

$$y_{3.4}(\delta) = 24,2 - 13,3x_1 - 50,8x_4 + 20,6x_5 + 26,4x_7 + 161,1x_9. \quad (3.4)$$

Исключение переменной $x_3(\mathrm{Mn}) = 1,1$ в уравнении (3.3) привело к конечному решению системы уравнений, так как при перерасчете уравнения (3.4) частный F-критерий переменной $x_9(\mathrm{N}_2) = 0,7$, подлежащей исключению, стал меньше $F_{\text{табл}}$ = 1,43.

Следовательно, дальнейший перерасчет уравнения не требуется, все оставшиеся переменные можно отнести к специальным характеристикам, влияющим на относительное сужение стали марки 65К.

Таким образом, в рассмотренном регрессионном анализе проведено исследование влияния химического состава катанки диаметром 5,5 мм стали 65К марки $(\mathrm{C, Si, Mn, P, S, Cr, Ni, Cu, N_2})$ на механические свойства (прочность (σ_b), относительное сужение (ψ), относительное удлинение (δ)). По результатам определены специальные характеристики $(\mathrm{C, Si, Mn, P, S, Cr, Ni, Cu})$ для данной марки стали, оказывающие влияние на механические свойства. Так, в случае определения прочности (σ_b) как специальной характеристики следует выделять следующие специальные параметры: $\mathrm{C, Cr, Ni, Cu, N_2}$; относительного сужения (ψ) – $\mathrm{Si, Cr, Ni, Cu, N_2, S, P}$; относительного удлинения (δ) – $\mathrm{C, Ni, N_2, S, P}$. Сила связи определена с помощью множественного коэффициента корреляции. Полученные значения показывают умеренную корреляцию между химическим составом, прочностью и относительным удлинением и слабую степень корреляции с относительным сужением (см. таблицу).

Суммарную меру общего качества уравнения регрессии (его соответствия статистическим данным) показывает коэффициент детерминации. Как показывают расчеты, приведенные в таблице, линейная связь между химическим составом и пределом прочности, относительным удлинением является слабой, а с относительным сужением практически отсутствует. Поэтому следует провести исследования влияния других факторов на зависимую переменную y.

Литература

1. К у м э Х. Статистические методы повышения качества / Пер. с англ. М.: Финансы и статистика, 1990.

2. Ш и ш к и н И. В., С т а н я к и н В. М. Квалиметрия и управление качеством. М.: Изд-во ВЗПИ, 1992.

3. А д л е р Ю. П., М а р к о в а Е. В., Г р а н о в с к и й Ю. В. Планирование эксперимента при поиске оптимальных условий. Изд. 2-е. М.: Наука, 1976.

4. Д р е й п е р Н., С м и т Г. Прикладной регрессионный анализ: В 2-х кн. Кн. 2 / Пер. с англ. 2-е изд., перераб. и доп. М.: Финансы и статистика, 1987.

5. Н е у й м и н Я. Г. Модели в науке и технике. История, теория, практика. Л.: Наука, 1984.

6. Ч и ч к о А. Н., С о б о л е в В. Ф., Ч и ч к о О. И. Статистические методы регулирования качества продукции в литейном производстве. Мн.: БНТУ, 2006.
7. Рекомендации. Прикладная статистика. Методы обработки данных. Основные требования и характеристики. М.: ВНИИС, 1987.
8. Теория статистики / Под ред. Р. А. Шмойловой. М.: Финансы и статистика, 1998.
9. И в а ш о в А. Линейная алгебра. Матрицы: Учеб. пособ. М.: ВНИИС, 2004.
10. О р л о в А. И. Эконометрика: Учеб. пособ. М.: Экзамен, 2002.
11. Е ф и м о в В. В. Улучшение качества проектов и процессов: Учеб. пособ. Ульяновск: УлГТУ, 2004.
12. Е ф и м о в В. В. Статистические методы в управлении качеством: Учеб. пособ. Ульяновск: УлГТУ, 2003.