

УДК 519.254:62-414

АЛГОРИТМ ПРОВЕДЕНИЯ ПЕРВИЧНОЙ СТАТИСТИЧЕСКОЙ ОБРАБОТКИ МАССИВОВ ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ

ЛАУХИН Д. В.¹, *д. т. н., проф.*,

БЕКЕТОВ А. В.², *к. т. н., доц.*,

РОТТ Н. А.³, *к. т. н., доц.*,

ЛАУХИН В. Д.⁴, *м. н. с.*

¹Кафедра материаловедения и обработки материалов, Государственное высшее учебное заведение «Приднепровская государственная академия строительства и архитектуры», ул. Чернышевского, 24-а, Днепро, 49600, Украина, тел. +38 (0562) 47-39-56, e-mail: d_laukhin@mail.ru, ORCID ID: [0000-0002-9842-499X](https://orcid.org/0000-0002-9842-499X)

²Кафедра материаловедения и обработки материалов, Государственное высшее учебное заведение «Приднепровская государственная академия строительства и архитектуры», ул. Чернышевского, 24-а, Днепро, 49600, Украина, тел. +38 (0562) 46-93-72, e-mail: lab120@mail.pgasa.dp.ua, ORCID ID: [0000-0002-1224-3355](https://orcid.org/0000-0002-1224-3355)

³Кафедра материаловедения и обработки материалов, Государственное высшее учебное заведение «Приднепровская государственная академия строительства и архитектуры», ул. Чернышевского, 24-а, Днепро, 49600, Украина, тел. +38 (0562) 46-93-72, e-mail: natalyrott@gmail.com, ORCID ID: [0000-0002-3839-6405](https://orcid.org/0000-0002-3839-6405)

⁴Лаборатория экспериментальных научных исследований, Государственное высшее учебное заведение «Приднепровская государственная академия строительства и архитектуры», ул. Чернышевского, 24-а, Днепро, 49600, Украина, тел. +38 (0562) 46-93-72, e-mail: laukhin.vladislav@gmail.com, ORCID ID: [0000-0002-6718-4639](https://orcid.org/0000-0002-6718-4639)

Аннотация. *Цель статьи.* Построение алгоритма предварительной (первичной) оценки массивов экспериментальных данных для дальнейшего получения математической модели исследуемого процесса. *Методика.* Использование основных закономерностей теории обработки массивов экспериментальных значений при первичном анализе данных. *Научная новизна.* Приведен алгоритм проведения первичного статистического анализа массивов экспериментальных данных. *Практическая значимость.* Разработка методов выявления статистически не достоверных значений в массивах экспериментальных данных с целью последующего их детального анализа и построения математической модели исследуемых процессов.

Ключевые слова: *алгоритм; первичная статистическая обработка; базовые статистические параметры; нормальное распределение*

АЛГОРИТМ ПРОВЕДЕННЯ ПЕРВИННОЇ СТАТИСТИЧНОЇ ОБРОБКИ МАСИВІВ ЕКСПЕРИМЕНТАЛЬНИХ ДАНИХ

ЛАУХІН Д. В.¹, *д. т. н., проф.*,

БЕКЕТОВ О. В.², *к. т. н., доц.*,

РОТТ Н. О.³, *к. т. н., доц.*,

ЛАУХІН В. Д.⁴, *м. н. с.*

¹Кафедра матеріалознавства та обробки матеріалів, Державний вищий навчальний заклад «Придніпровська державна академія будівництва та архітектури», вул. Чернишевського, 24-а, Дніпро, 49600, Україна, тел. +38 (0562) 47-39-56, e-mail: d_laukhin@mail.ru, ORCID ID: [0000-0002-9842-499X](https://orcid.org/0000-0002-9842-499X)

²Кафедра матеріалознавства та обробки матеріалів, Державний вищий навчальний заклад «Придніпровська державна академія будівництва та архітектури», вул. Чернишевського, 24-а, Дніпро, 49600, Україна, тел. +38 (0562) 46-93-72, e-mail: lab120@mail.pgasa.dp.ua, ORCID ID: [0000-0002-1224-3355](https://orcid.org/0000-0002-1224-3355)

³Кафедра матеріалознавства та обробки матеріалів, Державний вищий навчальний заклад «Придніпровська державна академія будівництва та архітектури», вул. Чернишевського, 24-а, Дніпро, 49600, Україна, тел. +38 (0562) 46-64-62, e-mail: natalyrott@gmail.com, ORCID ID: [0000-0002-3839-6405](https://orcid.org/0000-0002-3839-6405)

⁴Лабораторія експериментальних наукових досліджень, Державний вищий навчальний заклад «Придніпровська державна академія будівництва та архітектури», вул. Чернишевського, 24-а, Дніпро, 49600, Україна, тел. +38 (0562) 46-93-72, e-mail: laukhin.vladislav@gmail.com, ORCID ID: [0000-0002-6718-4639](https://orcid.org/0000-0002-6718-4639)

Анотація. *Мета статті.* Побудова алгоритму попередньої (первинної) оцінки масивів експериментальних даних для подальшого математичного моделювання процесу, що досліджується. *Методика.* Використання основних закономірностей теорії обробки масивів експериментальних значень для попереднього аналізу даних. *Наукова новизна.* Наведено алгоритм проведення первинного статистичного аналізу масивів експериментальних даних. *Практична значимість.* Розробка методів виявлення статистично не достовірних значень у масивах експериментальних даних із метою наступного детального аналізу та побудови математичних моделей процесів, що досліджуються.

Ключові слова: *алгоритм; первинна статистична обробка; базові статистичні параметри; нормальний розподіл*

ALGORITHM OF PRIMARY STATISTICAL ANALYSIS OF ARRAYS OF EXPERIMENTAL DATA

LAUKHIN D. V.¹, *Doctor of Technical Sciences, Prof.*,

BEKETOV A. V.², *Ph. D., Assos. prof.*,

ROTT N. A.³, *Ph. D., Assos. prof.*,

LAUKHIN V. D.⁴, *Junior scientist*

¹Department of Material Science and Treatment of Materials, SHEE «Pridneprovsk State Academy of Civil Engineering and Architecture», 24-a, Chernishevskogo str., Dnipro, 49600, Ukraine, +38 (0562) 46-93-72, e-mail: d_laukhin@mail.ru, ORCID ID: [0000-0002-9842-499X](https://orcid.org/0000-0002-9842-499X)

²Department of Material Science and Treatment of Materials, SHEE «Pridneprovsk State Academy of Civil Engineering and Architecture», 24-a, Chernishevskogo str., Dnipro, 49600, Ukraine, +38 (0562) 46-93-72, e-mail: lab120@mail.pgasa.dp.ua, ORCID ID: [0000-0002-1224-3355](https://orcid.org/0000-0002-1224-3355)

³Department of Material Science and Treatment of Materials, SHEE «Pridneprovsk State Academy of Civil Engineering and Architecture», 24-a, Chernishevskogo str., Dnipro, 49600, Ukraine, +38 (0562) 46-93-72, e-mail: natalyrott@gmail.com, ORCID ID: [0000-0002-3839-6405](https://orcid.org/0000-0002-3839-6405)

⁴Department of Material Science and Treatment of Materials, SHEE «Pridneprovsk State Academy of Civil Engineering and Architecture», 24-a, Chernishevskogo str., Dnipro, 49600, Ukraine, +38 (0562) 46-93-72, e-mail: laukhin.vladislav@gmail.com, ORCID ID: [0000-0002-6718-4639](https://orcid.org/0000-0002-6718-4639)

Annotation. Purpose. Construction of an algorithm for preliminary (primary) estimation of arrays of experimental data for further obtaining a mathematical model of the process under study. **Methodology.** The use of the main regularities of the theory of processing arrays of experimental values in the initial analysis of data. **Originality.** An algorithm for performing a primary statistical analysis of the arrays of experimental data is given. **Practical value.** Development of methods for revealing statistically unreliable values in arrays of experimental data for the purpose of their subsequent detailed analysis and construction of a mathematical model of the studied processes.

Keywords: *algorithm; primary statistical processing; basic statistical parameters; normal distribution*

Введение. Увеличение объемов производства металлопроката приводит к появлению больших массивов экспериментальных данных зависимостей комплекса эксплуатационных характеристик от параметров производства. Полученные в результате производственного контроля технологических режимов численные значения параметров процессов могут быть далее использованы (для получения математических моделей и/или количественной их оценки) лишь в том случае, если они достоверны, т. е. соответствуют определенным уровням и критериям математической статистики.

При этом в настоящее время обработка массивов экспериментальных данных не возможна без использования компьютеров и пакетов прикладных программ. Несмотря на впечатляющее разнообразие статистических пакетов, которым характеризуется современный мировой и отечественный рынок, важно правильно понимать область применения статистических методов решения того или иного класса задач. В этом аспекте большое значение имеет правильное формирование массивов экспериментальных данных для

последующей статистической обработки: классификация данных, выделение погрешностей, «выпадов», ошибок, оценка интервалов существования величин, соответствия распределений анализируемых величин критериям теории обработки экспериментальных данных и математической статистики.

В связи с этим, актуальной задачей является первичная обработка массивов экспериментальных данных с целью их последующего использования при построении адекватных математических моделей анализируемых процессов.

В общем случае возможно выделить три основных типа конечных прикладных целей статистических исследований [1]:

- **тип 1:** установление самого факта наличия (или отсутствия) статистически значимой связи между зависимыми и независимыми переменными. При такой постановке задачи статистический вывод имеет альтернативную природу «связь есть» или «связи нет» и может сопровождаться численной характеристикой степени тесноты исследуемой зависимости;
- **тип 2:** прогноз (восстановление неизвестных значений) исследуемых

результатирующих признаков (зависимых переменных). При такой постановке задачи статистический вывод включает в себя описание интервала (области) вероятных значений прогнозируемого показателя и сопровождается величиной доверительной вероятности, с которой гарантируется справедливость прогноза;

- тип 3: выявление причинных связей между независимыми и зависимыми переменными (частичное управление значениями зависимых переменных путем регулирования величин независимых переменных). Такая постановка задачи претендует на проникновение в физический смысл изучаемых статистических связей.

Корректность используемых в статистическом анализе методов, а также интерпретация результатов в значительной степени зависят от некоторых основных свойств анализируемого процесса.

Для пояснения роли и места основных приемов статистического анализа и последующего математического моделирования удобно разложить эту схему на основные этапы исследования. Хотя, в общем случае, подобное разложение может носить условный характер [2].

Этап 1: исходный (предварительный) анализ исследуемой системы. В результате этого анализа определяются:

- основные цели исследования на неформализованном, содержательном уровне;
- совокупность единиц, представляющих предмет статистического исследования;
- перечень показателей, характеризующих состояние каждого из исследуемых объектов;

Этап 2: составление детального плана сбора исходной статистической информации. При составлении этого плана необходимо, по возможности, учитывать полную схему дальнейшего статистического анализа.

Этап 3: сбор исходных статистических данных и формирование их в виде таблиц данных (с целью дальнейшей обработки на ЭВМ). На данном этапе формируется

основная цель статистического анализа и определяется набор статистических инструментов, который необходим для достижения поставленной цели.

Этап 4: первичная статистическая обработка данных. В ходе первичной статистической обработки данных необходимо решить следующие задачи:

- преобразование (при необходимости) переменных (например, описанных текстом) в номинальную или порядковую шкалу;
- статистическое описание исходных совокупностей с определением пределов варьирования переменных;
- анализ резко выделяющихся переменных;
- восстановление пропущенных наблюдений;
- проверка статистической независимости последовательности наблюдений, составляющих массив экспериментальных данных;
- экспериментальный анализ закона распределения исследуемой генеральной совокупности.

Этап 5: вычислительная реализация выбранных методов статистической обработки массива данных и интерпретация полученных результатов. На этом этапе происходит выбор программного обеспечения, необходимого для реализации поставленной цели статистического исследования. В тех случаях, когда при интерпретации результатов проведенных исследований нельзя опереться на теоретическую базу, может оказаться полезным использование имитационного статистического моделирования [3].

При этом необходимо отметить, что одним из наиболее важных этапов статистической обработки является первичная обработка массивов данных. Основной целью данного этапа является оценивание основных свойств процесса и, как следствие, выбор основных вычислительных функций, которые возможно будет применить в дальнейшем для построения математических моделей исследуемых процессов.

К числу основных свойств (со статистической точки зрения) можно отнести стационарность и нормальность рассматриваемого процесса [4].

Стационарность процесса играет важную роль потому, что методы анализа нестационарных процессов существенно более громоздки, чем в стационарном случае. Если установлено, что в процессе содержатся периодические составляющие, это позволяет избежать в дальнейшем ошибок при интерпретации результатов анализа.

По-видимому, наиболее простой способ оценивания стационарности заключается в рассмотрении физической природы процесса. Если основные физические факторы, определяющие процесс, не зависят от времени, можно без дальнейшего исследования полагать изучаемый процесс стационарным.

Теоретически наличие периодических и/или почти периодических составляющих в случайном процессе проявляется в виде дельта-функций в его спектральной плотности [5]. На практике оказывается, что спектральная плотность содержит острые пики, которые можно ошибочно приписать узкополосному случайному шуму. Желательно поэтому установить наличие периодических составляющих, чтобы не путать их с узкополосным случайным шумом, спектральная плотность которого конечна.

Наиболее просто проверить, подчиняется ли исследуемый процесс нормальному закону, можно, измерив плотность вероятности значений процесса и сравнив ее с теоретически нормальным распределением. Если количество измерений достаточно велико, а ошибки измерений малы по сравнению с отклонениями функции от нормальной кривой, несоответствие ее нормальному распределению будет очевидно. Если выборочное распределение оценки плотности вероятности известно, можно применить критерии нормальности даже в том случае, когда статистические ошибки велики.

Технически реализация анализа свойств больших массивов экспериментальных данных затруднительна без использования пакетов специальных программ для статистической обработки [6]. Вместе с тем, необходимо отметить, что информация, получаемая с помощью пакетов прикладных программ, часто пересыщена результатами, которые не всегда будут использованы при дальнейшем статистическом анализе.

Цель статьи - построение алгоритма предварительной (первичной) оценки массивов экспериментальных данных для дальнейшего получения математической модели исследуемого процесса.

Методика и результаты исследования. Проблемой, которая рассматривается в данной работе, является получение количественной зависимости (построение математической модели) взаимосвязи параметров производства (в частности, температуры конца чистовой прокатки) на комплекс основных механических характеристик (σ_T , σ_B , δ_5) низкоуглеродистой микролегированной стали 09Г2С. Для решения поставленной задачи необходимым условием является проверка соответствия имеющегося массива данных требованиям построения математической модели, т. е. проверка возможности построения математической модели без нормировки (или другого типа преобразования) исходных данных. С этой целью авторами работы предложен алгоритм первичной статистической обработки массива данных, который включал следующие операции:

- формирование таблиц массивов данных, которые подлежат последующей вычислительной обработке (выявление явных «выпадов», пропущенных значений);
- вычисление и анализ базовых статистических показателей;
- проверка соответствия массива обрабатываемых данных нормальному распределению (выполняется в соответствии с требованиями [7]);
- построение гистограмм типа «коробки»;
- на основании полученных данных выявление интервалов значений, которые

могут быть далее использованы для построения математической модели (определение статистически достоверных интервалов значений анализируемых показателей).

Вычисления проводились с использованием программного комплекса StatSoft Statistica (v7.0) [8].

Формирование таблиц массивов данных, которые подлежат последующей вычислительной обработке. Массив значений, который подвергся статистической обработке, приведен в таблице 1.

Анализ приведенных данных показывает, что в массиве анализируемых величин отсутствуют явные «выпады» и пропущенные значения. Таким образом, данный массив может подвергаться дальнейшей статистической обработке.

Вычисление и анализ базовых статистических показателей. Результаты расчета базовых статистических показателей представлены в таблице 2.

В приведенной таблице рассчитаны следующие статистические показатели:

- *минимум* – минимальное значение исследуемого параметра;
- *максимум* – максимальное значение исследуемой величины;
- *среднее* – среднее арифметическое значение исследуемой величины;
- *размах* – разность между наибольшим и наименьшим значениями исследуемой величины;
- *мода* – представляет собой наиболее типичное, наиболее вероятное значение, своеобразный центр, вокруг которого разбросаны все значения исследуемой величины;
- *медиана* – представляет собой такое значение исследуемой величины, которое разделяет весь объем значений на две равные по численности части;
- *дисперсия* – характеризует меру изменчивости (вариации) исследуемой величины и представляет собой средний квадрат отклонений случаев от среднего значения исследуемой величины. В отличие от других показателей дисперсия

может быть разложена на составные части, что позволяет оценить влияние различных факторов на вариацию признака;

- *стандартное отклонение (среднее квадратичное отклонение)* – характеризует величину отклонений значений от среднего;
- *коэффициент асимметрии* – характеризует степень смещения значений относительного среднего по величине и направлению;
- *эксцесс* – характеризует степень концентрации случаев вокруг среднего значения и является мерой «крутости кривой». В кривой нормального распределения эксцесс равен нулю.

Вывод о реальном интервале, в котором находятся анализируемые параметры, делается на основании комплексного анализа рассчитанных базовых статистических показателей. При этом определение формы кривой является одной из самых важных задач, т. к. статистический материал в обычных условиях дает по определенному параметру характерную, типичную только для него кривую распределения. Всякое искажение формы кривой означает нарушение или изменение нормальных условий возникновения анализируемых параметров.

Искажение формы кривой распределения (по сравнению с кривой нормального распределения) можно определить, анализируя два статистических показателя: коэффициент асимметрии и эксцесс. При этом рассматриваются следующие условия:

- $A_s > 0$ – правосторонняя асимметрия распределения (кривая смещена вправо от центра);
- $A_s < 0$ – левосторонняя асимметрия распределения (кривая смещена влево от центра);
- $k > 0$ – «островершинное» распределение;
- $k < 0$ – «плосковершинное» распределение;

В качестве примера в таблице 3 приведен анализ коэффициента асимметрии и эксцесса для исследуемых величин.

Важной характеристикой, анализируемой на основании рассчитанных

статистических показателей, является надежность среднего значения. В случае близкого совпадения распределения исследуемого массива величин с нормальным законом выполняются соотношения [2]:

$$|Cp - Me| = 15...1; \quad (1)$$

$$Cp = Mo = Me. \quad (2)$$

Для рассматриваемых массивов данных анализ надежности среднего приведен в таблице 3.

Следующим критерием анализа является оценка однородности распределения. Для этих целей можно использовать два статистических показателя: дисперсия и среднее квадратичное отклонение. Оба показателя характеризуют отклонение значений

массива от среднего арифметического. При этом среднее квадратичное отклонение в большей степени характеризует точность проведения измерений (т. е. абсолютную и относительные ошибки). Дисперсия является вариативным признаком и позволяет судить о колебаниях значений массива вокруг среднего арифметического, т. е. анализ дисперсии распределения позволяет оценивать разброс значений (явные «выпады» и «провалы»).

Согласно теоретическим данным [см. например 1 – 5], незначительная флуктуация среднего значения должна соответствовать значению вариативного признака $\cong 40\%$. Оценка однородности исследуемых распределений приведена в таблице 3.

Таблица 1

Массив данных для первичной статистической обработки / An array of data for the primary statistical processing

Анализируемые параметры							
Предел текучести, σ_T , МПа	Предел прочности, σ_B , МПа	Относительное удлинение, δ_5 , %	Температура конца прокатки, $T_{кон}$ °С	Предел текучести, σ_T , МПа	Предел прочности, σ_B , МПа	Относительное удлинение, δ_5 , %	Температура конца прокатки, $T_{кон}$ °С
370	495	30	753	360	510	34	754
365	495	25	746	375	510	30	748
380	520	30	746	370	510	30	748
370	510	29	749	357	500	31	758
380	520	27	748	370	510	29	754
370	510	33	750	360	500	33	747
370	510	30	748	375	500	33	751
365	500	30	753	375	500	30	750
380	500	36	754	380	510	28	753
375	520	32	759	370	510	27	746
370	510	32	756	380	495	28	752
375	500	34	749	365	495	32	754
365	520	29	752	365	500	32	746
380	510	26	757	375	510	28	756
375	520	29	746	375	510	32	750
369	495	27	751	370	500	30	755
375	510	36	746	370	500	34	759

Таблица 2

Значения базовых статистических показателей / The values of the basic statistical indicators

Анализируемый параметр	Количество измерений	Минимум, МПа	Максимум, МПа	Среднее (Cp), МПа	Размах, МПа	Мода (Mo), МПа	Медиана (Me), МПа	Дисперсия	Стандартное отклонение	A_s , Коэффициент асимметрии	κ , Эксцесс
σ_T , МПа	35	357	380	372	23	370	370	39,129	6,255	-0,4113	-0,3663
σ_B , МПа	35	495	520	507	25	510	510	69,034	8,309	0,1905	-1,0525
δ_5 , %	35	25	36	30	11	30	30	7,252	2,693	0,2032	-0,3681
$T_{кон}$ °С	35	745	759	751	14	746	751	16,987	4,121	0,3012	-0,9640

Анализ базовых статистических показателей / Analysis of basic statistical indicators

Анализируемый параметр	Симметрия распределения		Надежность среднего значения		Однородность распределения	
	A_s	k	$ C_p - Me $	Равенство	Станд. откл	Дисперсия
σ_T , МПа	<0	<0	2	$C_p \neq Mo = Me$	< 10%	< 40%
σ_B , МПа	>0	<0	3	$C_p \neq Mo = Me$	< 10%	> 40%
δ_s , %	>0	<0	0	$C_p = Mo = Me$	< 10%	< 40%
$T_{кон}$ °C	>0	<0	5	$C_p = Me \neq Me$	< 10%	< 40%

Таким образом, на основании полученных данных можно сделать следующую количественную оценку исследуемых распределений:

- *предел текучести* – распределение имеет плоскую вершину и левостороннюю асимметрию. Наблюдаются незначительные отклонения среднего значения от медианы, при этом показатели вариативных признаков находятся в пределах допустимых значений. Это свидетельствует о незначительных отклонениях значений массива от среднего и, как следствие, об отсутствии явных «выпадов» и «провалов» анализируемых значений;
- *предел прочности* – распределение имеет плоскую вершину и правостороннюю асимметрию (относительно среднего значения). Наблюдаются незначительные отклонения среднего значения от медианы, при этом значения дисперсии превышают допустимый уровень. Это свидетельствует о рассеянии значений массива возле среднего. Возможной причиной установленного факта может являться наличие явных «выпадов» и «провалов» анализируемых значений;
- *относительное удлинение* – распределение имеет плоскую вершину и правостороннюю (относительно среднего значения) асимметрию. Наблюдаются равенство показателей среднего (среднего арифметического, моды, медианы), что свидетельствует о близости формы распределения к нормальному закону. Показатели вариативных признаков находятся в пределах допустимых значений;
- *температура конца прокатки* – распределение имеет плоскую вершину и

правостороннюю асимметрию (относительно среднего значения). Наблюдаются незначительные отклонения среднего значения от моды, при этом показатели вариативных признаков находятся в пределах допустимых значений. Это свидетельствует о незначительных отклонениях значений массива от среднего и, как следствие, об отсутствии явных «выпадов» и «провалов» анализируемых значений.

Проверка соответствия массива обрабатываемых данных нормальному распределению

Процедура проверки соответствия внешнего вида исследуемых распределений нормальному закону состоит в построении для массива значений гистограммы в координатах:

$\text{число случаев (попаданий в заданный диапазон)} = f(\text{диапазон значений})$. (3)

Количество интервалов значений задается на основании соответствующего расчета [6], в котором учитывается количество измерений. Построенные для исследуемых массивов гистограммы приведены на рисунке 1.

Характерное свойство нормального распределения состоит в том, что 68 % всех его значений лежит в диапазоне ± 1 стандартное отклонение от среднего, а 95 % – в диапазоне ± 2 стандартных отклонения.

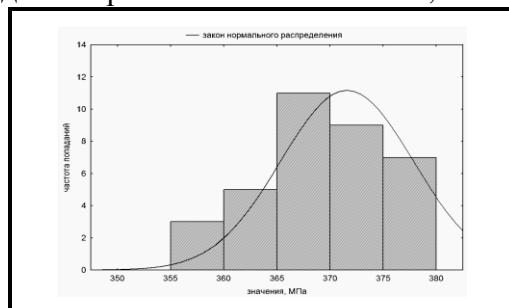
Таким образом, можно сделать следующие предварительные выводы:

- *предел текучести* – область наиболее достоверных значений – 365...377 МПа, область статистически не достоверных значений – < 360 МПа;
- *предел прочности* – область наиболее достоверных значений – 497...514 МПа,

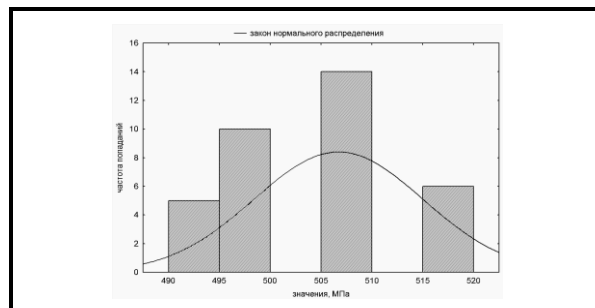
область статистически не достоверных значений – < 491 МПа;

- *относительное удлинение* – область наиболее достоверных значений – 28...33 %, область статистически не достоверных значений – < 25%, > 35 %;

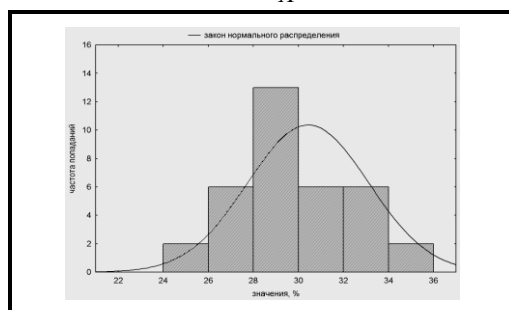
- *температура конца прокатки* – область наиболее достоверных значений – 747...755 °С, область статистически не достоверных значений – > 759 °С;



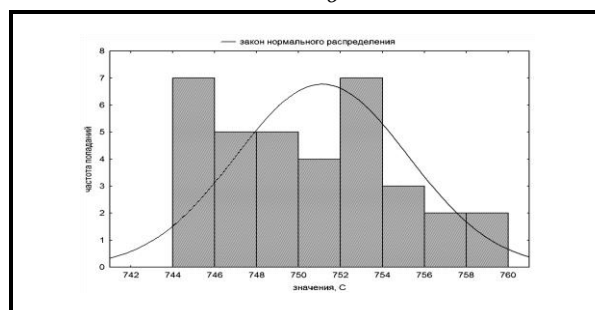
А



б

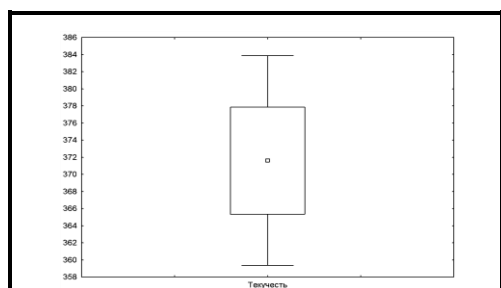


В

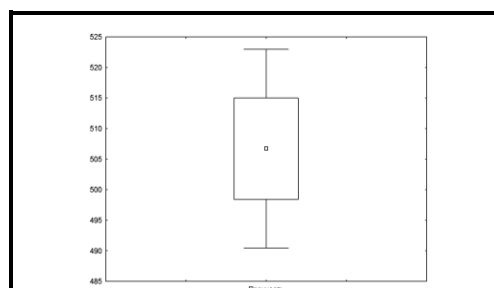


г

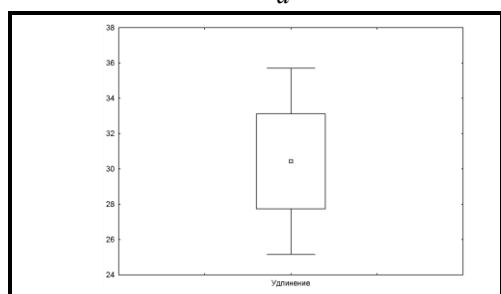
Рис. 1. Гистограммы распределения исследуемых величин: а – предела текучести; б – предела прочности; в – относительного удлинения; г – температура конца прокатки



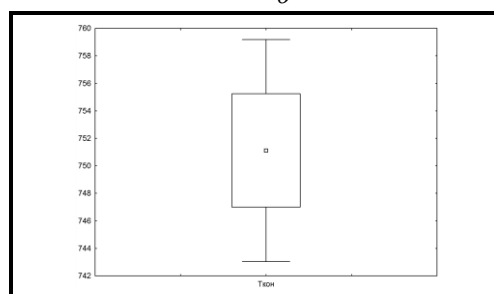
а



б



в



г

Рис. 2. Гистограммы типа «ящик»: а – предела текучести; б – предела прочности; в – относительного удлинения; г – температура конца прокатки

Построение гистограмм типа «коробки»

Построение гистограмм типа «коробки» используется для визуализации областей

наиболее достоверных значений, выделенных на основании анализа диаграмм распределения. Существует несколько вариантов построения данного типа

диаграмм. В данной работе использовались диаграмма вида:

$$\frac{\text{среднее значение}}{\text{стандартное отклонение}} / \frac{\text{среднее значение} + 1,96 \times \text{стандартное отклонение}}{\text{среднее значение}} \quad (4)$$

Внешний вид построенных гистограмм для анализируемых массивов приведен на рисунке 2. При этом при построении гистограмм использованы следующие условные изображения:

- Среднее;
- ▭ Среднее+стандартное отклонение;
- ▭ Среднее+1,96 стандартного отклонения

Анализ гистограмм позволяет выделить две области:

- область наиболее достоверных значений (выделена прямоугольником);
- область статистически значимых значений (выделена вертикальными и горизонтальными линиями).

Определение статистически достоверных интервалов значений анализируемых показателей

Выявление статистически достоверных интервалов значений осуществляется на основании совместного анализа данных таблиц 2, 3 и рисунков 1, 2.

Для исследуемых в работе массивов статистически достоверные интервалы обобщены и представлены в таблице 4.

Во всем массиве значений можно выделить несколько интервалов:

- *интервал наиболее достоверных значений* – область данных, которые распределяются в пределах 40 % отклонений от средних показателей (среднего арифметического, моды, медианы). Значения, которые попадают в эту область, могут быть использованы для дальнейшей математической обработки без какого-либо дополнительного анализа;
- *интервал статистически значимых значений* – данные, которые попадают в доверительный интервал среднего ($Cp + 0,95 \times Cp$). Эти значения могут быть использованы для дальнейшей математической обработки только после дополнительного анализа (выявление причин, почему значения значительно отклонились от показателей среднего);
- *интервал статистически не достоверных значений* – данные, которые попадают за границы доверительного интервала среднего («выпады» и «провалы»). Эти величины не могут быть приняты во внимание при построении математической модели описываемого процесса. При этом необходимо провести детальный анализ причин появления таких значений, т. к. их возникновение может быть связано с еще не установленными и/или не учтенными составляющими описываемого процесса.

Таблица 4

Статистически достоверные интервалы

Анализируемый параметр	Интервал наиболее достоверных значений	Интервал статистически значимых значений	Интервал статистически не достоверных значений
σ_{Γ} , МПа	$365 < x < 377$	$377 < x < 384$ $360 < x < 365$	$x < 360$
$\sigma_{\text{в}}$, МПа	$497 < x < 514$	$514 < x < 522$ $491 < x < 497$	$x < 491$
δ_5 , %	$28 < x < 33$	$33 < x < 35$ $25 < x < 28$	$x < 25\%$; $x > 35\%$
$T_{\text{кон}}$ °С	$747 < x < 755$	$755 < x < 759$ $743 < x < 747$	$x > 759$

Выводы. 1. В работе показано построение алгоритма предварительной (первичной) оценки массивов экспериментальных данных для дальнейшего получения математической модели исследуемого процесса.

2. Проанализирован метод формирования таблиц массивов данных для первичной статистической обработки.

3. На примере массивов экспериментальных данных показан расчет и последующий анализ базовых статистических показателей.

4. Проанализированы основные этапы проверки соответствия массива обрабатываемых данных нормальному распределению.

5. На основании проведенного комплексного анализа статистической

информации показан способ определения значений анализируемых показателей.
статистически достоверных интервалов

ИСПОЛЬЗОВАННАЯ ЛИТЕРАТУРА

1. Айвазян С. А. Прикладная статистика. Исследование зависимостей / С. А. Айвазян, И. С. Енюков, Л. Д. Мешалкин ; под ред. С. А. Айвазяна. – Москва : Финансы и статистика, 1985. – 487 с. – Режим доступа: <http://www.twirpx.com/file/46852/>.
2. Айвазян С. А. Прикладная статистика. Основы моделирования и первичная обработка данных / С. А. Айвазян, И. С. Енюков, Л. Д. Мешалкин. – Москва : Финансы и статистика, 1983. – 471 с. – Режим доступа: <http://www.twirpx.com/file/46842/>.
3. Бендат Дж. Прикладной анализ случайных данных [пер. с англ.] / Бендат Дж., Пирсон А. – Москва : Мир, 1989. – 540 с. Режим доступа: <http://mexalib.com/view/3368>
4. Спирин Н. А. Методы планирования и обработки результатов инженерного эксперимента : консп. лекц. / Спирин Н. А., Лавров В. В. ; под общ. ред. Н. А. Спирина. – Екатеринбург : ГОУ ВПО УГТУ-УПИ, 2004. – 257 с. – Режим доступа: <http://window.edu.ru/resource/524/28524/files/ustu131.pdf>
5. Фаддеев М. А. Элементарная обработка результатов эксперимента : учеб. пособие / Фаддеев М. А. – Нижний Новгород : Изд-во Нижегород. госуниверситета, 2010. – 122 с. – Режим доступа: http://www.unn.ru/books/met_files/ElemTreat.pdf
6. Использование пакета Statistica 5.0 для статистической обработки опытных данных : метод. указ. для диплом. проект. для студ. лесного фак-та спец. 260400 "Лесное хозяйство" и 260500 "Садово-парковое и ландшафтное строительство" / Сарат. гос. агр. ун-т ; сост. Кобанов С. В. – Саратов : Саратов. гос. ун-т, 2001. – 42 с. – Режим доступа: <http://www.ict.edu.ru/ft/004911/mu.pdf>
7. Статистические методы. Проверка отклонения распределения вероятностей от нормального распределения : ГОСТ Р ИСО 5479-2002. – Введ. 2002-07-01. – Москва : Госстандарт России, 2002. – 31 с. – Режим доступа: <http://docs.cntd.ru/document/gost-r-iso-5479-2002>
8. Академия Анализа Данных StatSoft. Курсы в области промышленности: управление качеством, контроль качества. – Режим доступа: <http://statsoft.ru/academy/courses/industry/list.php>

REFERENCES

1. Ajvazyan S.A., Enyukov I.S. and Meshalkin L.D. Prikladnaya statistika. Issledovanie zavisimostej. [Applied statistics. Dependency research]. Moskva: Finansy i statistika, 1985, 487 p. Available at: <http://www.twirpx.com/file/46852/>.
2. Ajvazyan S.A., Enyukov I.S. and Meshalkin L.D. Prikladnaya statistika. Osnovy modelirovaniya i pervichnaya obrabotka dannyx [Applied statistics. Fundamentals of modeling and primary data processing]. Moskva: Finansyi i statistika, 1983, 471 p. Available at: <http://www.twirpx.com/file/46842/>. (in Russian).
3. Bendat Dzh. and Pirson A. Prikladnoj analiz sluchajnyx dannyx [Applied analysis of random data]. Moskva: Mir, 1989, 540 p. Available at: <http://mexalib.com/view/3368>. (in Russian).
4. Sparin N.A. and Lavrov V.D. Metodyi planirovaniya i obrabotki rezultatov inzhenerenogo eksperimenta [Planning methods and results processing of an engineering experiment]. Ekaterinburg: GOU VPO UGTU-UPI, 2004, 257 p. Available at: <http://window.edu.ru/resource/524/28524/files/ustu131.pdf>. (in Russian).
5. Faddeev M.A. Elementarnaya obrabotka rezultatov eksperimenta [Elementary processing of experimental results]. Nizhnij Novgorod: Izd-vo Nizhnegorod. gosuniversiteta, 2010, 122 p. Available at: http://www.unn.ru/books/met_files/ElemTreat.pdf. (in Russian).
6. Kobanov S.V., ed. Ispol'zovanie paketa Statistica 5.0 dlya statisticheskoy obrabotki opytnyx dannyx: metod. ukaz. dlya diplom. projekt. dlya stud. lesnogo fak-ta spec. 260400 "Lesnoe hozyajstvo" i 260500 "Sadovo-parkovoe i landshaftnoe stroitel'stvo" [The Statistica 5.0 package using for statistical processing of experimental data: a methodology instructions for a diploma project for students of the forest department, 260400 speciality of "Forestry" and 260500 speciality of "Garden and park and landscape construction"]. Saratov: Saratov. gos. un-t, 2000, 42 p. Available at: <http://www.ict.edu.ru/ft/004911/mu.pdf>. (in Russian).
7. Statisticheskie metody. Proverka otkloneniya raspredeleniya veroyatnostej ot normalnogo raspredeleniya, GOST R ISO 5479-2002, vved. 2002-07-01 [Statistical methods. The deviation checking of the probability distribution from the normal distribution: the State Standards of Russia ISO 5479-2002, dated on 2002-07-01]. Moskva: Gosstandart Rossii, 2002, 31 p. Available at: <http://docs.cntd.ru/document/gost-r-iso-5479-2002>. (in Russian).
8. Akademiya Analiza Danyx StatSoft. Kursy v oblasti promyshlennosti: upravleniya kachestvom, kontrol' kachestva [Academy of Data Analysis StatSoft. Courses in the field of industry: quality management, quality control]. Available at: <http://www.statsoft.ru>. (in Russian).

Рецензент: Большаков В. И., д-р т. н., проф.

Надійшла до редколегії: 11. 03.2017 р. Прийнята до друку: 14.03.2017 р