ФЕДЕРАЛЬНОЕ АГЕНТСТВО

ПО ТЕХНИЧЕСКОМУ РЕГУЛИРОВАНИЮ И МЕТРОЛОГИИ

РЕКОМЕНДАЦИИ
ПО СТАНДАРТИЗАЦИИ

**Р 50.1.082—
2012**

Статистические методы

ПРИМЕРЫ ПРИМЕНЕНИЯ

Часть 4

Простые статистические приемы анализа данных

Издание официальное

Москва
Стандартинформ
2013

gost.e-ecolog.ru

Предисловие

1 РАЗРАБОТАНЫ Автономной некоммерческой организацией «Научно-исследовательский центр контроля и диагностики технических систем» (АНО «НИЦ КД»)

2 ВНЕСЕНЫ Техническим комитетом по стандартизации ТК 125 «Статистические методы в управлении качеством продукции»

3 УТВЕРЖДЕНЫ И ВВЕДЕНЫ В ДЕЙСТВИЕ Приказом Федерального агентства по техническому регулированию и метрологии от 29 ноября 2012 г. № 1283-ст

4 ВВЕДЕНЫ ВПЕРВЫЕ

Информация об изменениях к настоящим рекомендациям публикуется в ежегодном указателе «Руководящие документы, рекомендации и правила», а текст изменений и поправок — в ежемесячном информационном указателе «Национальные стандарты». В случае пересмотра (замены) или отмены настоящих рекомендаций соответствующее уведомление будет опубликовано в ежемесячном информационном указателе «Национальные стандарты». Соответствующая информация, уведомление и тексты размещаются также в информационной системе общего пользования — на официальном сайте Федерального агентства по техническому регулированию и метрологии в сети Интернет

© Стандартинформ, 2013

Настоящие рекомендации не могут быть полностью или частично воспроизведены, тиражированы и распространены в качестве официального издания без разрешения Федерального агентства по техническому регулированию и метрологии

Содержание

1 Область применения 1
2 Нормативные ссылки 1
3 Пояснения к применяемым статистическим терминам 1
4 Графическое представление данных 3
Библиография 19

Введение

Серия рекомендаций по стандартизации «Статистические методы. Примеры применения» включает разъяснения применения статистических методов в простой и доступной форме.

В серии рекомендаций представлены способы применения простых статистических методов и приемов, показано, как знание процесса производства может способствовать его улучшению, повышению эффективности, производительности и повышению качества изготавливаемой продукции.

В настоящих рекомендациях показаны роль статистических методов при анализе данных, а также простые методы графического представления данных, обеспечивающие получение предварительной информации об исследуемом объекте.

РЕКОМЕНДАЦИИ ПО СТАНДАРТИЗАЦИИ

Статистические методы

ПРИМЕРЫ ПРИМЕНЕНИЯ

Часть 4

Простые статистические приемы анализа данных

Statistical methods. Examples of application. Part 4. Simple examples of the data analysis

Дата введения — 2013—12—01

1 Область применения

В настоящих рекомендациях приведены наиболее часто применяемые на практике методы графического представления данных. На основе конкретных примеров показаны применение этих методов и приемы предварительного анализа данных.

Наблюдаемыми характеристиками могут быть температура, давление, усилие сжатия, скорость, массовая концентрация вещества в материале, наработка до отказа, уровень звукового давления и др. Измеренные значения характеристик являются данными наблюдений.

2 Нормативные ссылки

ГОСТ Р 50779.10—2000 Статистические методы. Вероятность и основы статистики. Термины и определения

ГОСТ Р 50779.11—2000 Статистические методы. Статистическое управление качеством. Термины и определения

Р 50.1.072—2010 Статистические методы. Примеры применения. Часть 1. Группировка данных

Примечание — При пользовании настоящими рекомендациями целесообразно проверить действие ссылочных документов в информационной системе общего пользования — на официальном сайте Федерального агентства по техническому регулированию и метрологии в сети Интернет или по ежегодному информационному указателю «Национальные стандарты», который опубликован по состоянию на 1 января текущего года, и по выпускам ежемесячного информационного указателя «Национальные стандарты» за текущий год. Если заменен ссылочный документ, на который дана недатированная ссылка, то рекомендуется использовать действующую версию этого документа с учетом всех внесенных в данную версию изменений. Если заменен ссылочный документ, на который дана датированная ссылка, то рекомендуется использовать версию этого документа с указанным выше годом утверждения (принятия). Если после утверждения настоящих рекомендаций в ссылочный документ, на который дана датированная ссылка, внесено изменение, затрагивающее положение, на которое дана ссылка, то это положение рекомендуется применять без учета данного изменения. Если ссылочный документ отменен без замены, то положение, в котором дана ссылка на него, рекомендуется применять в части, не затрагивающей эту ссылку.

3 Пояснения к применяемым статистическим терминам

Группу единиц или значений величины, отобранных из совокупности большего объема, в статистике называют выборкой. Единственное наблюдение можно рассматривать как выборку объема 1. Таким образом, выборка может состоять из 1, 2, 3.... *n* единиц или наблюдений.

Издание официальное

Выборку, отобранную случайным образом, называют случайной выборкой. Множество объектов (единиц), из которых отбирают выборку (например, всех студентов в колледже), называют совокупностью или генеральной совокупностью. Совокупность ограничена перечнем формирующих ее единиц, объектов (например, колледж, реестр, женщины старше 50 лет и т. п.). Существуют статистические показатели, которые могут быть применены и к группе единиц, и к совокупности.

Для группы объектов наблюдений важное значение имеют:

а) параметр положения;

б) параметр изменчивости;

в) модель изменчивости.

Существуют различные методы оценки параметров положения и изменчивости в пределах группы наблюдений. При этом важно учитывать модель изменчивости. Например, при исследовании возможностей процесса часто используют предположение о нормальности распределения, при формировании требований к надежности оборудования — о постоянной интенсивности отказов.

Обычно положение центра наблюдаемых данных характеризуют с помощью:

1) среднего арифметического (или выборочного среднего, представляющего собой сумму наблюдаемых значений, деленную на их количество);

2) медианы (значения, расположенного в центре выборки, если данные расположены в порядке неубывания или невозрастания);

3) моды (наиболее часто появляющегося в наблюдениях значения).

Два наиболее часто используемых показателя изменчивости:

- размах (разность между наибольшим и наименьшим значениями в выборке);

- стандартное отклонение (мера разброса данных вокруг среднего). Чем меньше изменчивость, тем меньше значение стандартного отклонения. Выборочная оценка стандартного отклонения имеет вид:

$$s = \sqrt{\frac{\sum_{i=1}^{n} X_i^2 - n\bar{X}^2}{n-1}} = \sqrt{\frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n-1}},$$

где X_i — i-е наблюдаемое значение;

$\bar{X}$ — выборочное среднее (среднее арифметическое);

n — количество наблюдений;

s — выборочное стандартное отклонение.

Преимущества и недостатки этих показателей приведены в таблице 1.

Таблица 1 — Преимущества и недостатки используемых статистических показателей

Показатель	Преимущества	Недостатки
Среднее	Легкость понимания, частое применение	Зависит от очень высоких или очень низких значений
Медиана	Не зависит от наибольшего и наименьшего значений	Необходимо использовать для вычисления все данные
Мода	Не зависит от наибольшего и наименьшего значений	Распределение может быть многомодальным
Размах	Легкость вычисления	Использует экстремальные значения
Стандартное отклонение	Более эффективно, чем размах	В простых случаях может быть вычислено вручную

Для выборки, состоящей из значений 7, 5, 10, 7 и 6.

Среднее арифметическое: (7 + 5 + 10 + 7 + 6)/5 = 7.

Медиана (центральное значение упорядоченной выборки 5, 6, 7, 7, 10): 7.

Мода (наиболее часто появляющееся значение в данных наблюдений): 7.

Размах (разность максимального и минимального значений): 10 – 5 = 5.

Стандартное отклонение (расчет приведен в таблице 2): 1,87.

Таблица 2 — Расчет стандартного отклонения

Выборочное значение	$(X-\bar{X})$	$(X-\bar{X})^2$
7	0	0
5	–2	4
10	3	9
7	0	0
6	–1	1

$$\sum_{i=1}^{5}(X_i-\bar{X}^2)=14.$$

Таким образом,

$$s=\sqrt{\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{n-1}}=\sqrt{\frac{14}{5-1}}=1{,}87.$$

Выборочное стандартное отклонение может быть вычислено с помощью компьютера или калькулятора.

4 Графическое представление данных

4.1 Точечная и штриховая диаграммы

На практике, когда доступно лишь несколько наблюдений, точечная или штриховая диаграммы (см. рисунки 1, 3 [1]) позволяют получить полезную предварительную картину ситуации. В некоторых случаях рассмотрения такой диаграммы бывает достаточно. Пример точечной диаграммы по данным таблиц приведен на рисунке 1 (на основе данных таблицы 3).

4.2 Карта количества наблюдений

Карту количества наблюдений применяют к результатам измерений и к классифицированным данным. Ее используют для визуального представления частоты появления конкретных значений или определенного типа событий в серии. При этом для отражения количества данных конкретного значения или события определенной группы используют обозначения |, ||, |||, ||||, 𝍸.

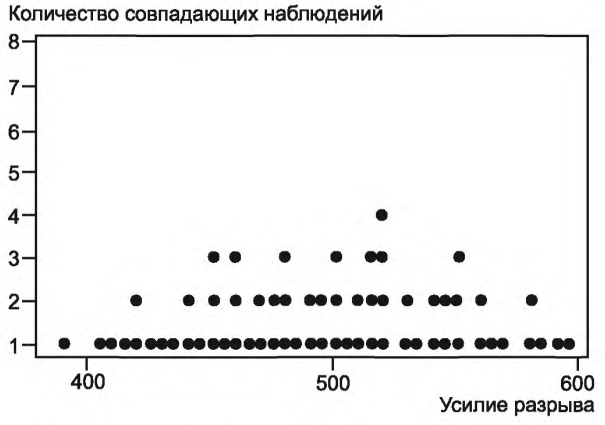


Рисунок 1 — Точечная диаграмма для значений усилия разрыва металлической проволоки (по данным таблицы 3)

Т а б л и ц а 3* — Результаты испытаний металлической проволоки на разрыв в порядке возрастания с округлением до 5 единиц

390	435	460	480	500	515	540	560
400	440	460	480	500	520	540	565
405	440	460	480	500	520	545	570
410	445	465	485	505	520	545	575
415	450	470	490	510	520	550	575
415	450	470	490	510	530	550	580
420	450	475	495	515	530	550	585
430	455	475	495	515	535	560	590
Среднее арифметическое: 495 Минимальное значение: 390 Максимальное значение: 590							

На рисунке 2 показаны примеры записей результатов измерений и классифицированных событий.

Карта количества результатов измерений		Карта количества событий/явлений	
21	𝍸	разорванная цепь	𝍸
22	𝍸	короткое замыкание	𝍸
23	𝍸 𝍸	непропаянное соединение	𝍸, 𝍸
24	𝍸 𝍸	разбрызгивание припоя	𝍸
25	𝍸 𝍸	наличие несоответствующего компонента	𝍸, 𝍸
26	𝍸	излом проводника	𝍸

Рисунок 2 — Типовая карта количества наблюдений

4.3 Карта «ствол и листья»

Карта «ствол и листья» показывает данные и их изменчивость. Это расширенная форма карты количества наблюдений или гистограммы. Карта состоит из двух частей:

а) в первой колонке указывают основные цифры в значениях наблюдений (ствол);

б) во второй колонке указывают следующие цифры в значениях наблюдений (листья).

На рисунке 3 показан пример карты «ствол и листья» для следующих упорядоченных данных: 29, 28, 41, 36, 36, 59, 50, 61, 44, 48, 35, 42, 53, 33, 31 после их расположения в порядке неубывания.

Ствол	Листья
2	8 9
3	1 3 5 6 6
4	1 2 4 8
5	0 3 9
6	1

Рисунок 3 — Карта «ствол и листья»

* Таблица заимствована из Р 50.1.072.

4.4 Диаграмма «ящик с усами»

Диаграмма «ящик с усами» очень полезна при анализе данных. Ее просто построить и легко интерпретировать. Так же, как точечную и штриховую диаграммы, эту диаграмму используют для отражения общих свойств или различий в группах данных. Основным элементом диаграммы является прямоугольник, высота которого охватывает область, где лежит 50 % значений наблюдаемой величины, с линией, указывающей медиану, и усами, показывающими максимальное и минимальное значения в выборке. На рисунке 4 представлены следующие элементы диаграммы:

M — значение медианы;

Q_1 — первый квартиль (значение, ниже которого лежит 1/4 значений наблюдаемой переменной);

Q_3 — третий квартиль (значение, выше которого лежит 1/4 значений наблюдаемой переменной);

min — минимальное наблюдаемое значение;

max — максимальное наблюдаемое значение;

На диаграмме «ящик с усами» знаком «*» могут быть указаны выбросы.

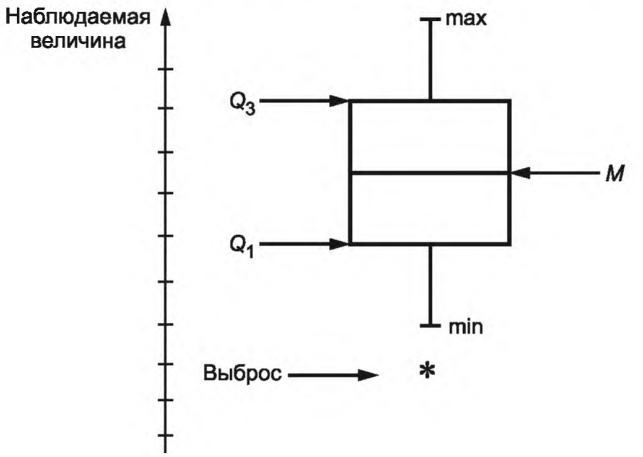


Рисунок 4 — Основной элемент диаграммы «ящик с усами»

На диаграмме можно указать доверительные границы для медианы. Если доверительные интервалы медианы для разных групп данных не пересекаются, это указывает на наличие существенных различий между значениями медиан в этих группах. Ширина «ящика» также может быть различной, что указывает на различия в объеме групп данных. Выбросы, показанные звездочкой, не используют при обработке выборки.

Диаграмма «ящик с усами» может быть дополнена более формальными статистическими методами, такими как дисперсионный анализ (ANOVA).

Пример 1 — Пример диаграммы «ящик с усами» показан на рисунке 5. Различные оттенки ткани конкретного способа окраски, полученной от трех различных поставщиков, были сопоставлены по установленным элементам одежды. Результаты представлены в форме диаграммы «ящик с усами» (см. рисунок 5).

На диаграмме указаны границы контролируемой переменной для трех поставщиков в зависимости от особенностей процесса окраски (характеризуемых положением медианы) и отклонения переменной, изображаемых в виде усов. Знак «*» указывает наличие двух выбросов, что свидетельствует о потере управляемости процессом окраски.

Диаграмма позволяет сделать следующие выводы:

а) поставщик 3 имеет процесс окраски с низким значением медианы и незначительными отклонениями от него;

б) поставщик 2 имеет процесс окраски с большим значением медианы контролируемой переменной и с более существенными отклонениями от него. Кроме того, имеется два значения из 30, явно не удовлетворяющие заказчика и подрывающие репутацию поставщика (продукция с таким значением контролируемой переменной может быть отозвана);

в) процесс поставщика 1 немного лучше, чем у поставщика 2, и имеет меньшую изменчивость.

Результат текущего состояния процесса окраски или лучшего метода принимают за точку отсчета или как образец сравнения для всех поставщиков.

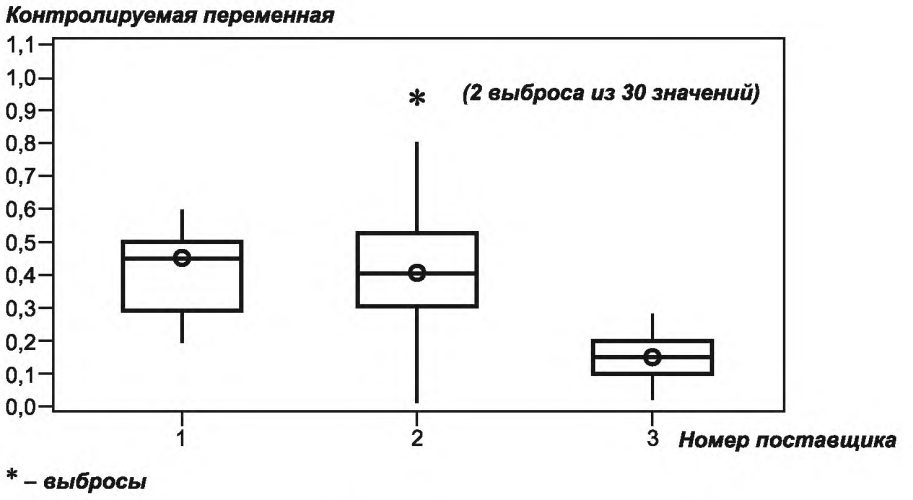


Рисунок 5 — Диаграмма «ящик с усами» для групп вариаций цвета одежды трех поставщиков[1)]

4.5 Многомерная карта

Многомерная карта является простым наглядным методом выявления и сравнения изменчивости данных из различных источников. Метод полезен для диагностики и исследований, но применим также для управления процессом. Карта состоит из вертикальных линий, соединяющих максимальные и минимальные значения характеристики (карта максимум-минимум). Точка на нулевом уровне соответствует идеальному значению характеристики. Чем длиннее линия, тем больше изменчивость.

***Пример 2** — Через каждый час измеряют диаметр трех деталей, изготовленных процессом. При этом фиксируют максимальный и минимальный диаметр для каждой детали. Результаты измерений приведены в таблице 4. На рисунке 6 представлена многомерная карта для трех вариантов анализа процесса и исследования доминирующих источников изменчивости:*

а) изменчивость в пределах единицы продукции;

б) изменчивость от единицы к единице;

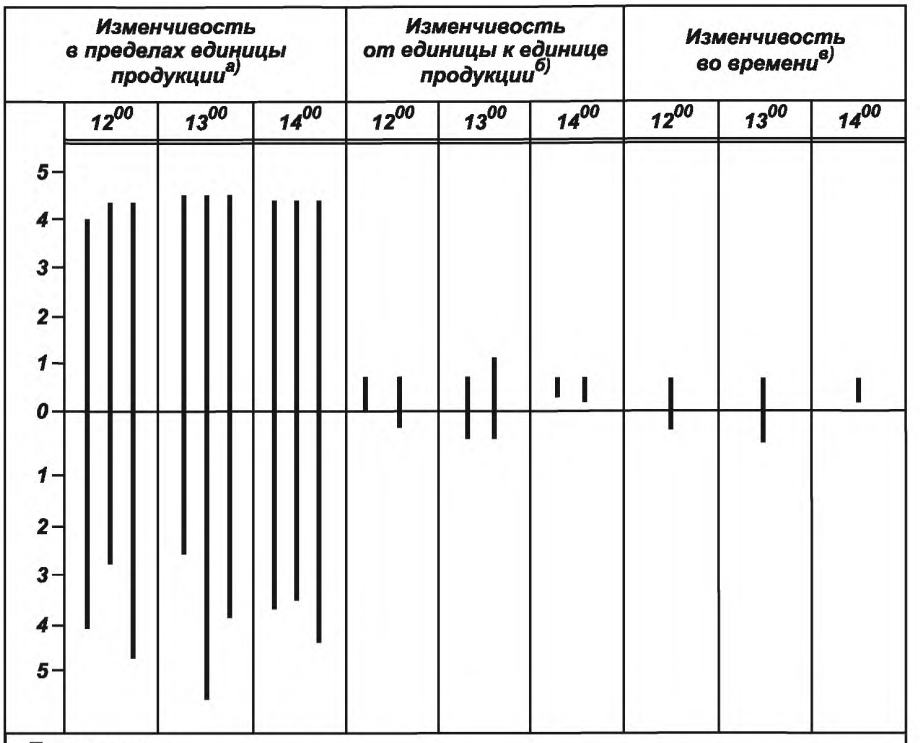
в) изменчивость во времени.

***Т а б л и ц а 4** — Результаты измерений диаметра деталей, выполненных последовательно через один час три раза по три детали*

Время измерений	№ детали	Наименьший диаметр вала, мм	Наибольший диаметр вала, мм	Средний диаметр вала, мм
12^{00}	1	36,0 (– 4,0)	44,0 (4,0)	40,0 (0,0)
	2	37,2 (– 2,8)	44,4 (4,4)	40,8 (0,8)
	3	35,2 (– 4,8)	44,2 (– 4,8)	39,7 (– 0,3)
13^{00}	1	37,2 (– 2,8)	44,4 (4,4)	40,8 (0,8)
	2	34,4 (– 5,6)	44,0 (4,0)	39,2 (– 0,8)
	3	37,2 (– 3,8)	44,0 (4,0)	40,1 (0,1)
14^{00}	1	36,2 (– 3,8)	44,4 (4,4)	40,3 (0,3)
	2	36,4 (– 3,6)	44,4 (4,4)	40,4 (0,4)
	3	36,0 (– 4,0)	44,4 (4,4)	40,2 (0,2)

П р и м е ч а н и е — В круглых скобках после результата измерений (х) указана разность (х – 40); 40 — номинальное значение диаметра детали.

[1)] В качестве контролируемой переменной использован показатель Delta E.



Примечания:

а) Приведенные отрезки показывают минимальное и максимальное отклонения результатов измерений диаметра от его номинального значения;

б) Приведенные отрезки показывают разности $(\bar{X}_{i+1} - \bar{X}_i)$ $(i = 1, 2)$, *где* $\bar{X}_i$ *– среднее отклонение диаметра i-й детали;*

в) Приведенные отрезки показывают изменчивость средних отклонений диаметров трех деталей, измеренных в i-й час $(i = 12^{00}, 13^{00}, 14^{00})$.

Рисунок 6 — Многомерная карта для анализа изменчивости процесса

4.6 P-D-диаграмма (положение-размер)

P-D-диаграмма является расширением многомерной карты на несколько характеристик. Горизонтальная ось P-D-диаграммы представляет собой идеальное значение (например, для овальности и конусности цилиндра). По вертикальной оси откладывают значения результатов измерений. Если эти значения совпадают с номинальным или условным значением общего среднего диаметра, это является идеальной ситуацией.

В примере показаны измерения внешнего диаметра цилиндра, проводимые для определения номинальных размеров его диаметра, овальности и конусности. Измерения проведены под прямым углом друг к другу на каждом основании цилиндра, как показано на рисунке 7а). Эти точки измерений обозначены A, B, C и D. Измерения проводят в каждом из четырех положений цилиндра.

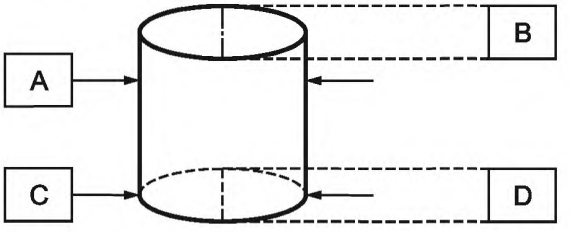


Рисунок 7а) — Измерения цилиндра для определения номинального значения овальности и конусности

На рисунке 7б) показаны допустимые отклонения диаметра и формы цилиндра.

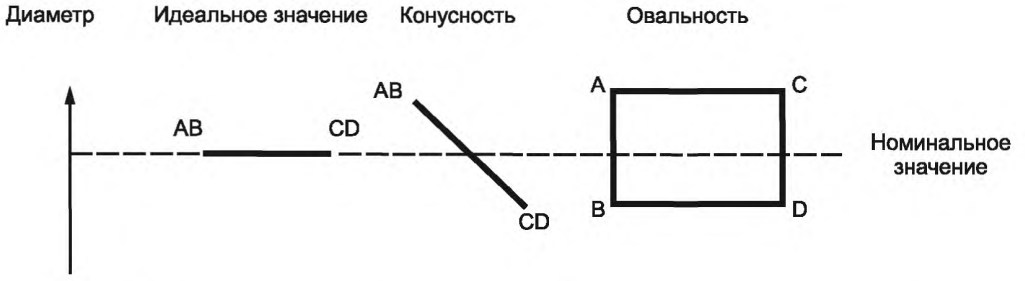


Рисунок 7б) — P-D-диаграмма, показывающая идеальное значение диаметра, овальность и конусность

Анализ P-D-диаграммы, представленной на рисунке 7в) с учетом P-D-диаграммы, представленной на рисунке 7б), показывает:

a) овальность:

- превышение диаметра в положении A диаметра в положении B указывает на наличие овальности на основании цилиндра AB, превышение диаметра в положении C диаметра в положении D указывает на наличие овальности на основании цилиндра CD;
- овальность постоянно увеличивается во времени до замены инструмента:

b) конусность:

- превышение средних значений диаметра в положении A и B, средних значений диаметра в положении C и D указывает на наличие конусности цилиндра;
- конусность постоянно увеличивается во времени до замены инструмента:

c) общий размер диаметра:

- среднее значение диаметра в положениях A, B и C, D позволяет получить оценку общего среднего диаметра;
- средний диаметр уменьшается во времени, все более отклоняясь от номинального значения до замены инструмента:

d) замена инструмента:

- достижение установленного значения общего диаметра и геометрической формы с некоторой остаточной овальностью на конце AB.

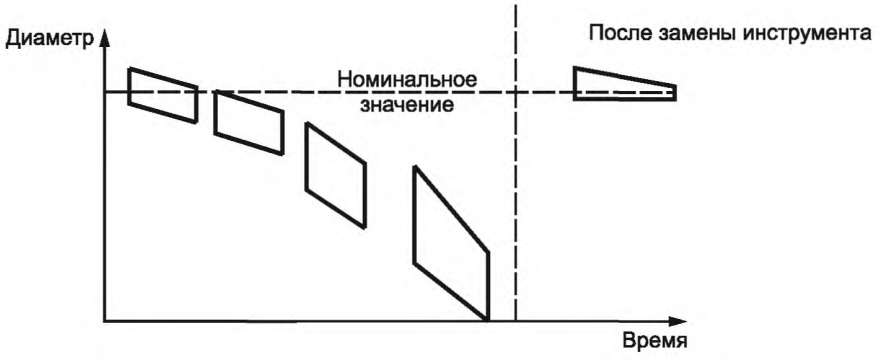


Рисунок 7в) — P-D-диаграмма, показывающая уменьшение среднего диаметра, увеличение отклонений от необходимой геометрической формы и благоприятные последствия замены инструмента

4.7 Графическое изображение гистограммы

Как только получено несколько наблюдений, могут быть построены линейная или точечная и штриховая диаграммы (см. рисунки 1 и 6) или карта «ствол и листья» (см. рисунок 3). Однако при наличии большого количества наблюдений обычно удобно располагать данные в порядке неубывания значений. Общая область изменения характеристики может быть разделена на равные интервалы и подсчитано количество наблюдений, попадающих в каждый интервал. Это количество называют частотой для дан-

ного интервала, а полученный в виде таблицы ряд чисел позволяет построить гистограмму. Существуют методы определения количества интервалов, исходя из общего количества наблюдений (например, правило Стерджеса). В таблице 5 приведено количество интервалов в соответствии с правилом Стерджеса.

Правило Стерджеса следует использовать в качестве грубой оценки количества интервалов. Количество интервалов, выбранных в конкретном случае, зависит от поставленной задачи.

Таблица 5 — Количество интервалов на основе общего количества наблюдений по правилу Стерджеса ($N = 1 + 3{,}3 \cdot \lg n$, где N — количество интервалов (классов); n — количество наблюдений)

Количество наблюдений	30—40	50 —90	100	200—300	400—700	800—1000	2000—3000	4000—6000	7000—10000
Количество интервалов	6	7	8	9	10	11	12	13	14

Данные наблюдений могут быть представлены несколькими способами, в том числе в виде:

a) таблицы частот;
b) диаграммы частот;
c) гистограммы;
d) таблицы накопленной частоты;
e) графика накопленной частоты.

Применение каждого из этих способов приведено в следующем примере.

Пример 3 — Проверка качества цинкового покрытия после гальванизации.

Отобранный образец в процессе испытаний должен выдержать воздействие кислоты при погружении его в ванну после гальванизации в течение четырех минут. В качестве данных исследований использовано 200 результатов. Измерения выполнены с точностью до 0,1 мин.

Результаты изменяются от 4,3 до 9,4 мин. В соответствии с правилом Стерджеса необходимо выбрать 9 интервалов для 200 результатов наблюдений. Это дало бы интервалы длиной (9,4 – 4,3)/9 = 0,57 мин. Результаты были расположены в порядке неубывания. Для простоты выбрано разбиение с интервалом 0,5 мин. Полученные результаты определения частоты показаны в таблице 6.

Таблица 6 — Частота времени до нарушения покрытия образца

Время до нарушения покрытия, мин	Количество наблюдений в интервале, частота, шт.	Количество наблюдений в интервале, %
Менее 4,1	0	0
от 4,1 до 4,5 вкл.	2	1
от 4,6 до 5,0 вкл.	5	2,5
от 5,1 до 5,5 вкл.	18	9
от 5,6 до 6,0 вкл.	27	13,5
от 6,1 до 6,5 вкл.	26	13
от 6,6 до 7,0 вкл.	39	19,8
от 7,1 до 7,5 вкл.	29	14,5
от 7,6 до 8,0 вкл.	25	12,5
от 8,1 до 8,5 вкл.	19	9,5
от 8,6 до 9,0 вкл.	6	3
от 9,1 до 9,5 вкл.	4	2
Свыше 9,5	0	0

В колонке 2 таблицы 6 приведено количество значений в каждом интервале, а в колонке 3 — та же величина в процентах от общего количества данных. Такая таблица позволяет увидеть распределение данных в целом.

На рисунках 8a)—e) представлены различные варианты представления данных таблицы 6. Они помогают оценить форму распределения данных и их соотношение с нижней границей поля допуска (4,0 мин).

На горизонтальной оси гистограммы указывают наблюдаемую характеристику. Количество наблюдений в данном интервале изображают в виде прямоугольника.

Иногда полезно указывать в гистограмме не фактическое количество данных в конкретном интервале, а их процент от общего количества данных (см. рисунок 8б)). Это является промежуточным этапом построения диаграммы частот.

Диаграмма кумулятивной относительной частоты показывает процент наблюдений, не превышающих указанного значения. Например, в соответствии с рисунком 8в) 58,5 % наблюдений не превосходят 7,0 мин. Следовательно, эта диаграмма полезна для определения расположения данных наблюдений относительно границы поля допуска.

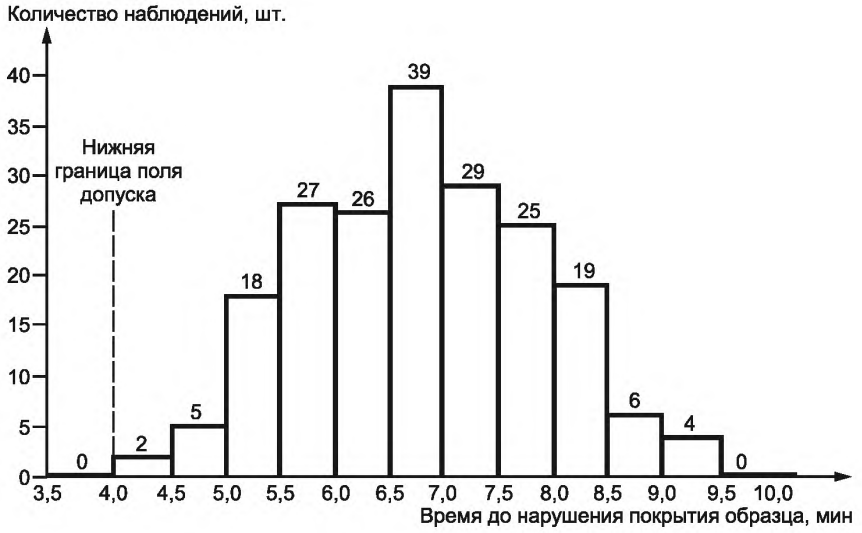


Рисунок 8а) — Количество наблюдений в конкретных интервалах

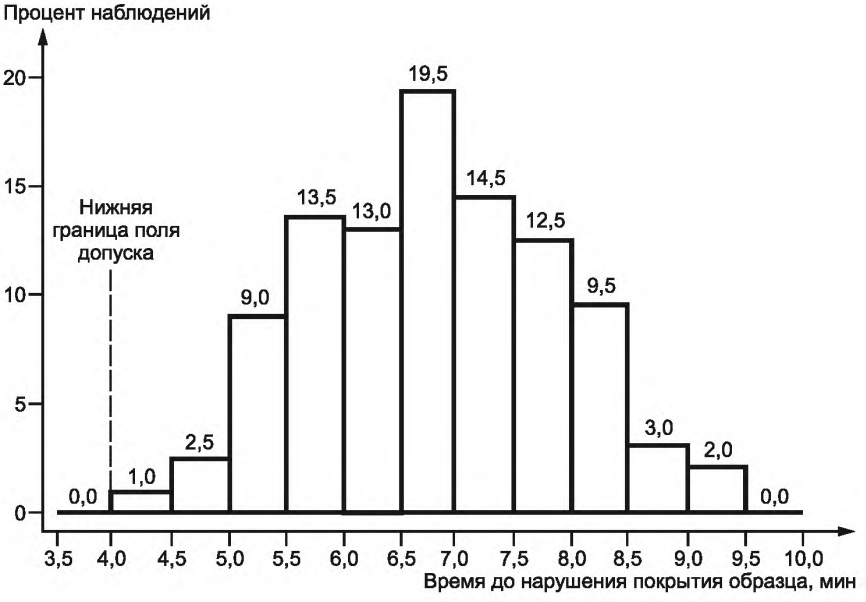


Рисунок 8б) — Гистограмма (в процентах)

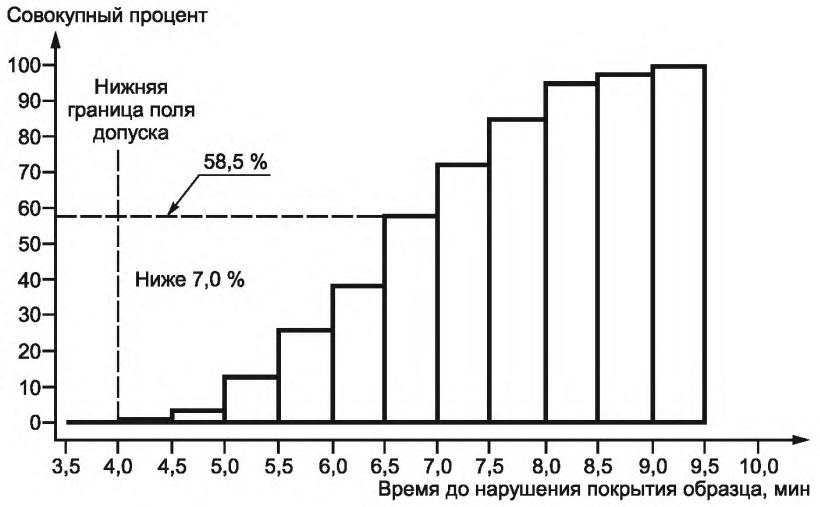


Рисунок 8в) — Гистограмма совокупного процента

Необходимо помнить, что совокупные значения процента относятся к верхней границе интервала, а не к его середине.

Гистограмма совокупного процента может быть представлена в форме гладкой кривой, как показано на рисунке 8г), или в виде прямой с помощью преобразования вертикального масштаба.

Диаграммы, представленные на рисунках 8а)—8 г), отражают фактическую ситуацию для объема выборки 200. Для прогнозирования качества гальванического покрытия исходя из имеющихся данных и предполагая, что процесс находится в устойчивом состоянии, часто прибегают к статистическому моделированию с использованием распределения вероятностей, соответствующего полученной гистограмме.

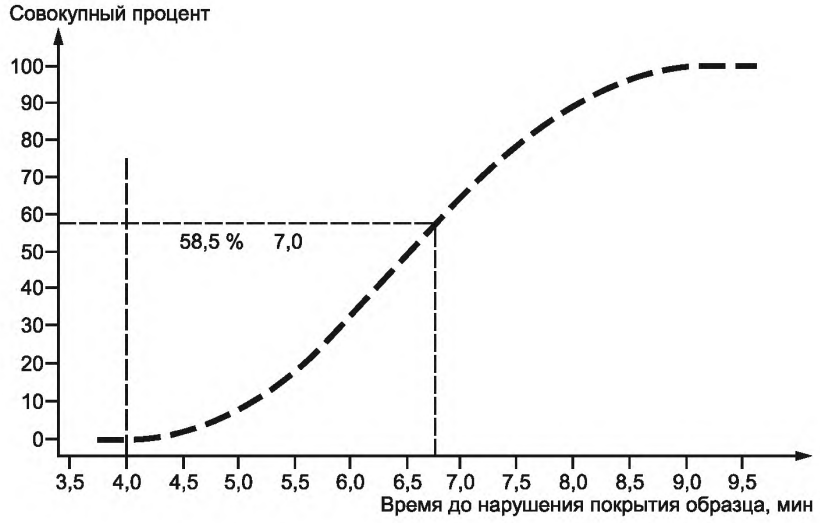


Рисунок 8г) — Диаграмма совокупного процента

В соответствии с рисунком 8а) можно предположить, что плотность распределения времени до нарушения цинкового покрытия под воздействием кислоты имеет колоколообразную форму и симметрична относительно среднего. Это соответствует нормальному распределению. Функция нормального распределения зависит от математического ожидания и стандартного отклонения. Характеристика,

которая может принимать нулевое значение, но является несимметричной относительно нуля, не может подчиняться нормальному распределению. Распределения с постоянной интенсивностью отказов, рассматриваемые при анализе надежности, также являются нормальными. Соответствие нормальному распределению наработки до отказа указывает на увеличивающуюся интенсивность отказов, т. е. на режим деградации.

Однако большое количество характеристик, рассматриваемых при анализе качества продукции или процессов, может быть описано нормальным распределением. Гистограмма, приведенная на рисунке 8д), показывает, что для описания данных рассматриваемого примера может быть использовано нормальное распределение.

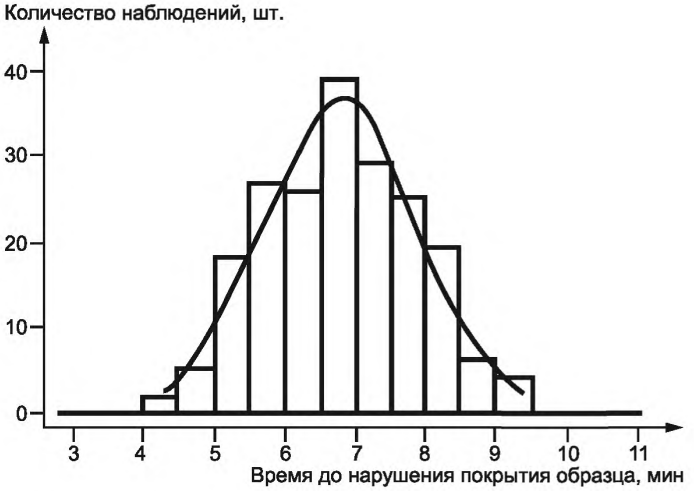


Рисунок 8д) — Нормальная кривая и гистограмма, построенная по данным примера 1 (среднее = 6,79, стандартное отклонение = 1,08)

Наложенная на гистограмму плотность нормального распределения построена с использованием в качестве параметров выборочного среднего 6,79 мин и стандартного отклонения 1,08 мин. Существует достаточно много статистических критериев для проверки отклонения распределения фактических данных от нормального распределения. Для этого могут быть использованы критерии Шапиро-Уилка, χ^2 и др.

Простой графический метод проверки соответствия фактических данных нормальному распределению состоит из составления графика совокупных частот в процентах на нормальной вероятностной бумаге. Если этот график представляет собой прямую линию, то существуют разумные основания полагать, что выборка соответствует нормальному распределению. Если график представляет собой другую кривую, для описания наблюдаемой характеристики следует использовать другое распределение. Этот метод используют также в статистическом управлении процессами для оценки пригодности и воспроизводимости процессов.

На рисунке 8е) приведен пример такого критерия для данных примера 3 (см. п. 4.7).

График совокупности частот на вероятностной бумаге:

- обеспечивает визуальное представление соответствия данных нормальному распределению;
- позволяет проводить простую экстраполяцию и прогнозирование значений вероятности за пределами области значений, полученных в процессе наблюдений;
- облегчает выявление ошибок;
- обеспечивает визуальную оценку соотношения наблюдений с границами поля допуска или установленными требованиями к изменчивости;
- обеспечивает оценку вероятности появления значений выше и/или ниже пределов поля допуска;
- служит для обнаружения отклонений от нормального распределения, например, гладкая вогнутая или выпуклая кривая на нормальной вероятностной бумаге указывает на несимметричность данных.

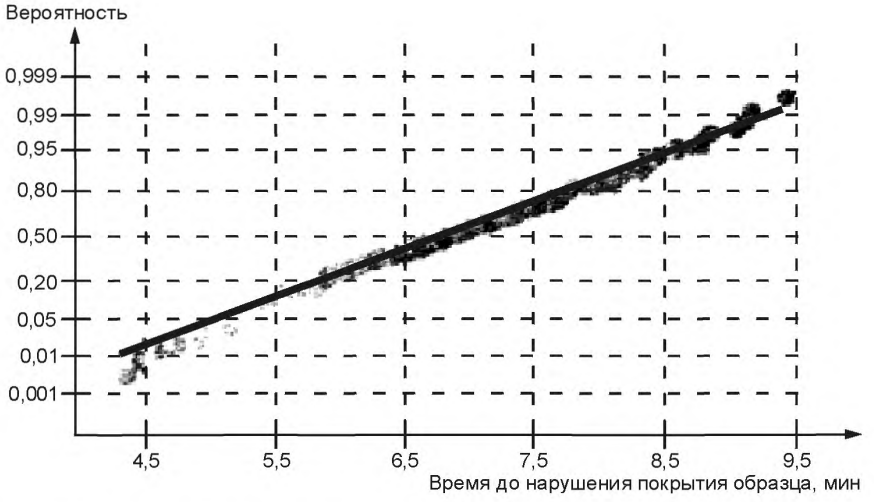


Рисунок 8е) — Прямая и данные наблюдений на нормальной вероятностной бумаге

Использование вероятностной бумаги для нормального или других распределений дает много преимуществ в интерпретации выборочных данных. Данные могут подчиняться лог-нормальному распределению, экспоненциальному распределению, распределению Вейбулла и др. Распределение Вейбулла часто используют при анализе надежности для моделирования различных режимов отказов: ранние отказы (убывающая интенсивность отказов), основная эксплуатация (постоянная интенсивность отказов) и деградация/старение (возрастающая интенсивность отказов). Вокруг прямой «наилучшей оценки» могут быть построены доверительные области. Границы доверительной области не являются прямыми линиями.

4.8 Нормальное распределение

В подразделе 4.7 приведен пример обоснования применения нормального распределения для описания типа изменчивости рассматриваемой характеристики.

Часто результаты наблюдений за процессом подчиняются нормальному распределению, поэтому его применяют при анализе устойчивости процесса. Для получения достоверных заключений о работе процесса в прошлом и прогнозирования его поведения необходимо иметь информацию о том, что в период наблюдений отсутствовали «специальные причины» изменчивости процесса.

Изменчивость нестабильного процесса непредсказуема. Поэтому основной задачей статистического управления процессом является проверка и обеспечение его стабильности.

При статистическом управлении процессами очень часто приходится иметь дело с нормальным распределением. Кроме того, распределение среднего арифметического выборок или подгрупп часто близко к нормальному, даже если объем выборки 4 или 5 в случаях, когда распределение индивидуальной характеристики не является нормальным.

Математическое ожидание (среднее) и дисперсия (квадрат стандартного отклонения) однозначно определяют вид нормального распределения. Следовательно, пользователь имеет возможность подобрать значения этих параметров в соответствии с имеющейся реальной ситуацией. Это легко продемонстрировать в графической форме и в форме таблицы (таблица 6). На рисунке 9 показана плотность нормированного нормального распределения (симметричная, имеющая колоколообразную форму). На графике отмечены наиболее часто применяемые процентные точки распределения. На рисунке 9 показан график плотности нормированного нормального распределения в интервале от – 4 до + 4. Областью определения нормального распределения является вся ось ($-\infty$, $+\infty$). Рисунок 9 относится к теоретическому распределению совокупности.

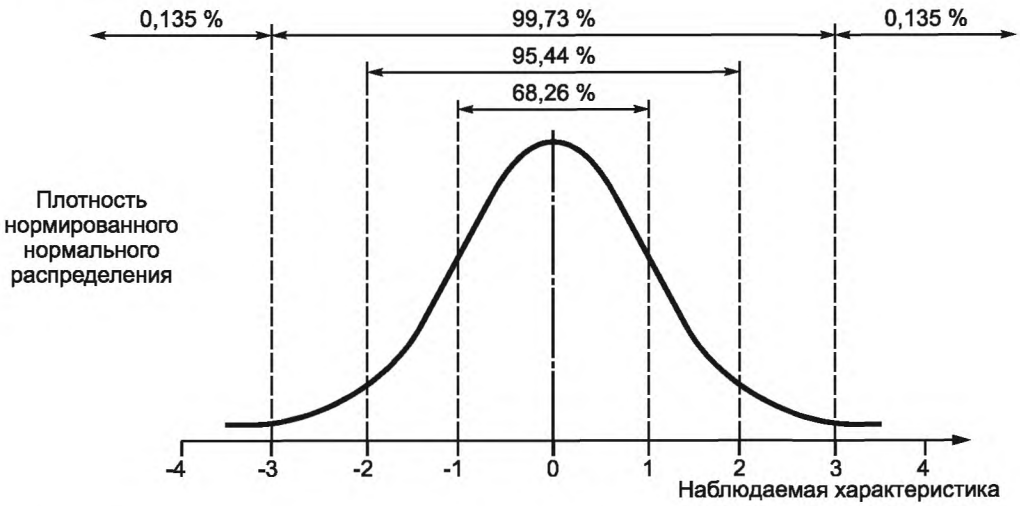


Рисунок 9 — Плотность и процентные точки нормированного нормального распределения

Из рисунка 9 видно, что:

а) 99,73 % значений переменной лежат в пределах значений, отстоящих от среднего на ± 3 стандартных отклонения;

б) из оставшихся 0,27 % значений 0,135 % лежат левее точки, отстоящей от среднего на 3 стандартных отклонения влево, и 0,135 % значений лежат правее точки, отстоящей от среднего на 3 стандартных отклонения вправо;

в) 95,44 % значений переменной лежат в пределах значений, отстоящих от среднего на ± 2 стандартных отклонения;

г) более чем две трети значений переменной (68,26 %) лежат в пределах значений, отстоящих от среднего на ± 1 стандартное отклонение.

Это показывает простые, но полезные свойства нормального распределения. Такие свойства распределения позволяют делать заключения о возможности описания данных с помощью нормального распределения по указанным процентным точкам.

Следующий пример показывает, как с помощью таблиц функции нормального распределения определить долю распределения выше или ниже заданного значения, если известны среднее и стандартное отклонения нормального распределения. Альтернативой использования такой таблицы является применение графика вероятностной прямой, показанного на рисунке 8е).

Пример 4 — Использование таблицы функции нормированного нормального распределения

Границы поля допуска: U = 42 + 4; L = 42 – 4.

Среднее μ ***= 40.***

Стандартное отклонение σ ***= 2,2.***

Контролируемый параметр устойчивого процесса подчиняется нормальному распределению.

Какой средний процент значений выходного параметра процесса находится вне границ поля допуска?

Для определения процента выхода процесса за верхнюю границу поля допуска необходимо определить z_U***:***

$$z_U = (U - \mu)/\sigma = (46 - 40)/2,2 = 2,73.$$

В соответствии с таблицей функции нормированного нормального распределения[1)] Ф(2,73) = 0,99653 (см. [1]). Таким образом, выше верхней границы поля допуска лежит доля распределения $1 - \Phi(z_U) = 1 - 0,99683 = 0,00317$ ***или 0,32 %.***

Для определения процента значений контролируемого параметра ниже нижней границы поля допуска необходимо определить z_L***:***

$z_L = (\mu - L)/\sigma = (40 - 38)/2,2 = 0,91.$

[1)] $\Phi(z) = 1/\sqrt{2\pi} \int_{-\infty}^{z} \exp\{-t^2/2\}\, dt.$

В соответствии с таблицей функции нормированного нормального распределения Ф(0,91) = 0,81861 (см. [1]). Таким образом, ниже нижней границы поля допуска лежит доля распределения (1 – 0,81861) · 100 % = 18,14 %. Следовательно, средний общий процент несоответствующих выходов составляет 18,14 %.

Пример 5 — Анализ размеров одежды

Анализ размеров одежды на основе репрезентативной выборки, представленной заказчиком, показал, что такая характеристика, как рост, подчиняется нормальному распределению со средним 75 см и стандартным отклонением 7 см. По результатам анализа было сделано предположение, что:

- 16 % (15,87 %) целевой совокупности показателей имеют рост 182 см (среднее + 1 стандартное отклонение);

- 25 % (25,14 %) целевой совокупности показателей имеют рост менее 170 см (среднее — 2/3 стандартного отклонения);

- 59 % (58,99 %) целевой совокупности показателей имеют рост от 170 до 182 см.

Такие оценки позволяют заказывать одежду соответствующего размера.

4.9 Распределение Вейбулла

Распределение Вейбулла имеет три параметра: формы, положения, масштаба. Это распределение часто применяют при решении задач надежности, диагностики состояния процесса и выделения элементов, влияющих на изменчивость качества продукции.

Параметр формы распределения Вейбулла β может принимать любое положительное действительное значение. Этот параметр определяет форму распределения Вейбулла. Например, распределение Вейбулла с:

1) $\beta < 1$ соответствует гиперэкспоненциальным распределениям;

2) $\beta = 1$ соответствует экспоненциальному распределению;

3) $1 < \beta < 3{,}5$ соответствует асимметричным распределениям, асимметрия которых уменьшается с ростом β (для $\beta = 3{,}5$ распределение становится близко к симметричному нормальному распределению);

4) $\beta > 3{,}5$ остается в значительной степени симметричным (небольшая асимметрия) и становится более остроконечным с ростом β.

Эта информация позволяет определить форму распределения Вейбулла, если известен его параметр формы. На рисунке 10 показан вид распределения Вейбулла в зависимости от параметра формы.

Любое распределение Вейбулла может быть изображено в виде прямой линии (независимо от значения β) на вероятностной бумаге распределения Вейбулла.

Распределение Вейбулла играет главную роль в задачах анализа надежности, поскольку во многих случаях позволяет описывать отказы на различных этапах жизненного цикла объекта. Обычно выделяют три этапа: приработка или ранние отказы (уменьшение интенсивности отказов), основной режим эксплуатации (постоянная интенсивность отказов) и деградация/старение (увеличение интенсивности отказов). Ниже приведены особенности распределения Вейбулла в зависимости от значения β:

- $\beta < 1$ соответствует уменьшению интенсивности отказов (режим приработки или ранних отказов);
- $\beta = 1$ соответствует постоянной интенсивности отказов (основной режим эксплуатации);
- $\beta > 1$ соответствует увеличению интенсивности отказов (режим деградации/старения).

Формула для определения интенсивности отказов имеет вид: $\lambda = \frac{\beta}{\alpha} t^{\beta-1}$, где α — параметр масштаба, t — время эксплуатации. Таким образом, если $\beta = 1$, что соответствует постоянной интенсивности отказов, то $T = \frac{1}{\lambda} = \alpha$, средняя наработка равна α.

Если $\beta = 2$, интенсивность отказов линейно возрастает во времени, если $\beta = 3$, интенсивность отказов возрастает пропорционально квадрату времени.

В тех случаях, когда исследуемая характеристика не принимает нулевого значения, играет роль третий параметр распределения Вейбулла γ. Параметр γ является параметром положения, который в задачах надежности часто равен нулю. В задачах надежности объект может отказать до начала функционирования, например, при хранении или поставке. Если $\gamma \neq 0$, такая ситуация проявляется на вероятностной бумаге распределения Вейбулла тем, что результаты наблюдений располагаются на гладкой кривой, а не на прямой линии. В таком случае, вычитая значение оценки γ из результатов каждого наблюдения и нанося полученные значения на график, можно получить прямую, демонстрирующую согласованность данных с распределением Вейбулла. Может потребоваться несколько повторений этой процедуры, чтобы достигнуть наилучшей оценки γ и, следовательно, получить распределение, хорошо описывающее данные наблюдений.

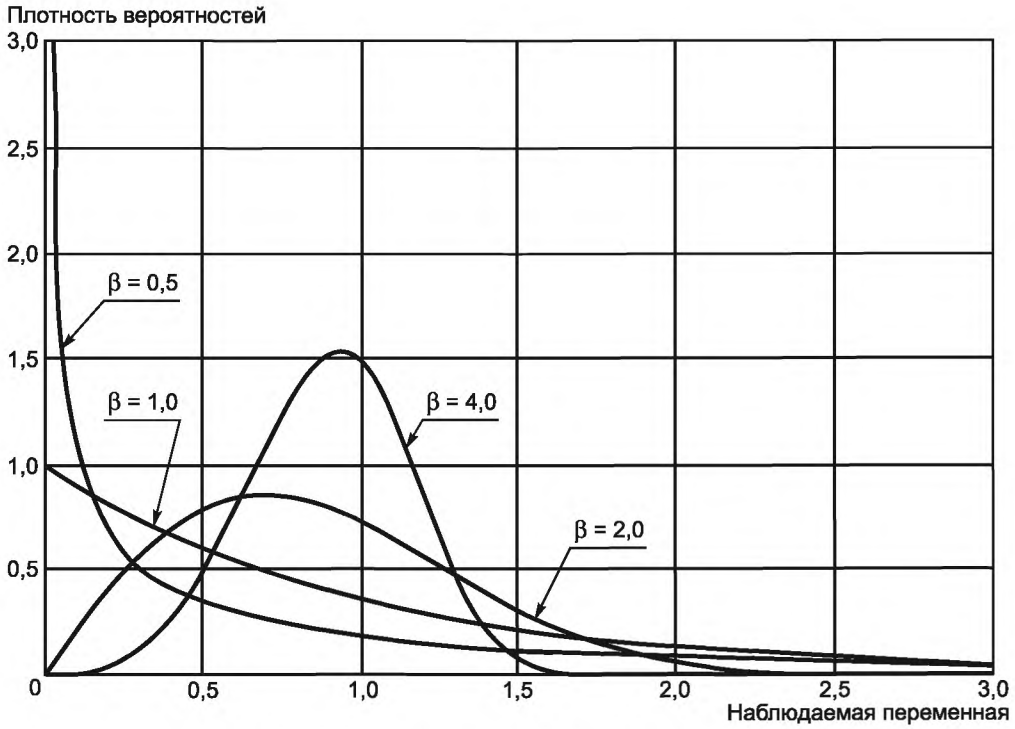


Рисунок 10 — Вид плотности распределения Вейбулла с α = 1

Пример 6 — Наработки объектов до отказа (в часах), проверенных в аналогичных условиях, составили: 179, 507, 949, 1454, 2317, 3345, 4302, 5687, 7674, 12315.

Необходимо определить оценку:

I) интенсивности отказов;

II) вероятности безотказной работы за 1000 час.

На рисунке 11 показана функция распределения Вейбулла с параметрами α = 3875, γ = 0 на вероятностной бумаге. Интенсивность отказов является постоянной. По рисунку можно определить, что за 1000 часов отказало 23 % объектов. Следовательно, вероятность безотказной работы за 1000 часов составляет 77 %.

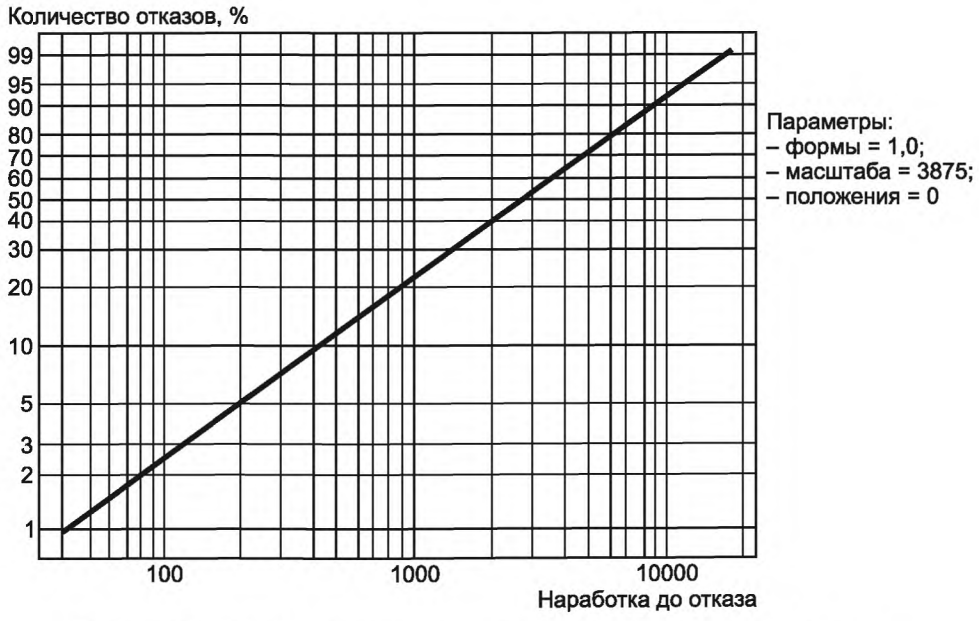


Рисунок 11 — График распределения Вейбулла на вероятностной бумаге

Вероятность безотказной работы за время t определяют по формуле

$$R(t) = \exp\left\{-\frac{t^{\beta}}{\alpha}\right\}.$$

Следовательно, $R(t) = e^{-\frac{1000^{1,0}}{3875}} = 77\ \%$.

4.10 График

График является способом представления данных в виде непрерывной кривой. График обеспечивает наглядность представления информации.

Существует несколько видов графиков:

а) линейный, когда горизонтальная и вертикальная оси имеют линейный масштаб (см. рисунок 13);

б) лог-линейный, когда горизонтальная ось имеет линейный масштаб, а вертикальная ось — логарифмический масштаб;

в) логарифмический, когда обе оси имеют логарифмический масштаб;

г) номограмма, когда график обеспечивает расчет в соответствии с формулой.

4.11 Линия регрессии

Линию регрессии используют для проверки наличия корреляции между переменными. Взаимосвязь переменных характеризуют направлением и формой линии регрессии. При определении корреляции между переменными следует помнить о различиях между корреляцией и причинно-следственной связью.

Пример линии регрессии показан на рисунке 12.

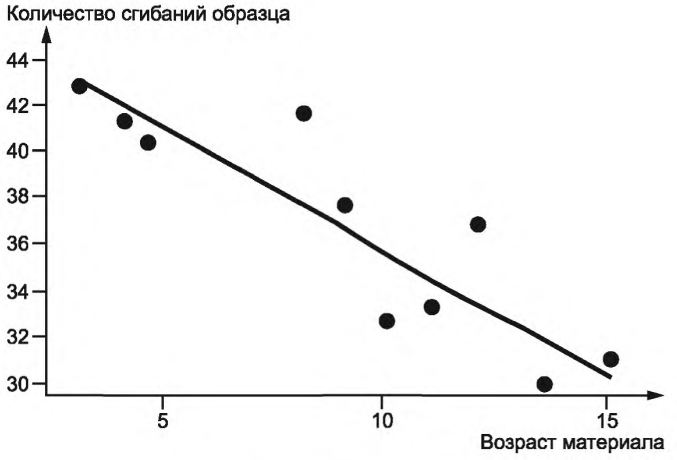


Рисунок 12 — Линия регрессии количества сгибаний образца каучука в зависимости от возраста материала

Близость к линейной зависимости характеризует коэффициент корреляции r. Чем ближе $|r|$ к 1, тем ближе зависимость переменных к линейной. В данном случае $r = 0{,}9$.

Зависимость среднего количества сгибаний y от возраста материала t можно описать уравнением $y = 46{,}4—1{,}08t$.

4.12 Диаграмма «Парето»

Диаграмма «Парето» представляет собой простой графический метод изображения ранжирования свойств, проблем или причин проблем по их значимости.

Диаграмма Парето показывает в порядке убывания относительный вклад каждого элемента (или причины) в проблему. Относительный вклад каждого элемента может быть оценен на основе относительной частоты, относительной стоимости или другого показателя. Вклады показывают в форме диаграммы. Иногда одновременно показывают линию накопленного вклада. Пример показан на рисунке 14.

Из рисунка 13 видно, что 65 % от общего количества несоответствий процесса окраски составляют отслаивание, потеки и наплывы. Эти недостатки были отобраны для работ по улучшению качества.

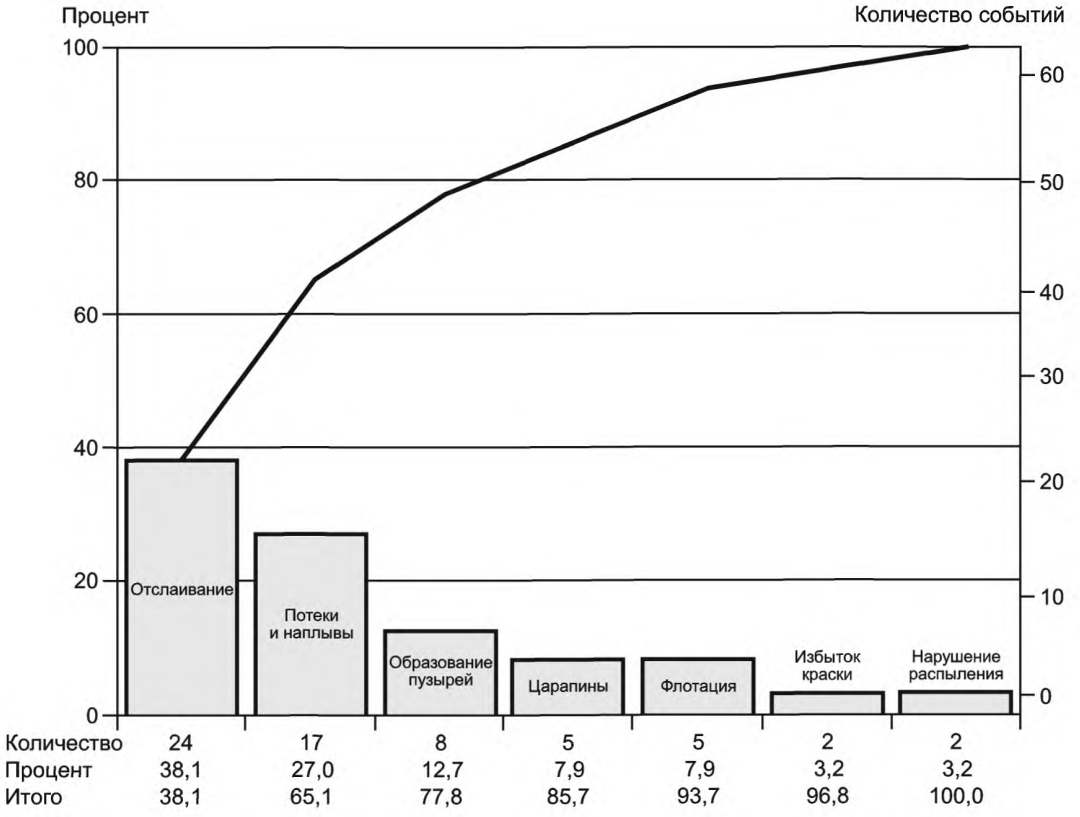


Рисунок 13 — Относительный вклад различных видов несоответствий процесса окраски

4.13 Диаграмма причин и последствий

Диаграмму причин и последствий часто называют диаграммой рыбьего скелета (из-за ее формы) или диаграммой Исикавы (по имени ее создателя). Эту диаграмму применяют в тех случаях, когда необходимо показать соотношение между причинами и следствиями. Существует несколько типов диаграммы, основанных на принципах формирования главных ветвей диаграммы, включая:

a) общий метод 4M (персонал, машины, материалы, технологии);
b) общий метод 4P (персонал, процедуры, производство (агрегат, цех), процесс);
c) анализ процесса (по этапам процесса в соответствии с их последовательностью);
d) метод объединения (по подгруппам);
e) метод анализа (на основе технического анализа).

На рисунке 14 показана диаграмма причин и последствий для литейного производства.

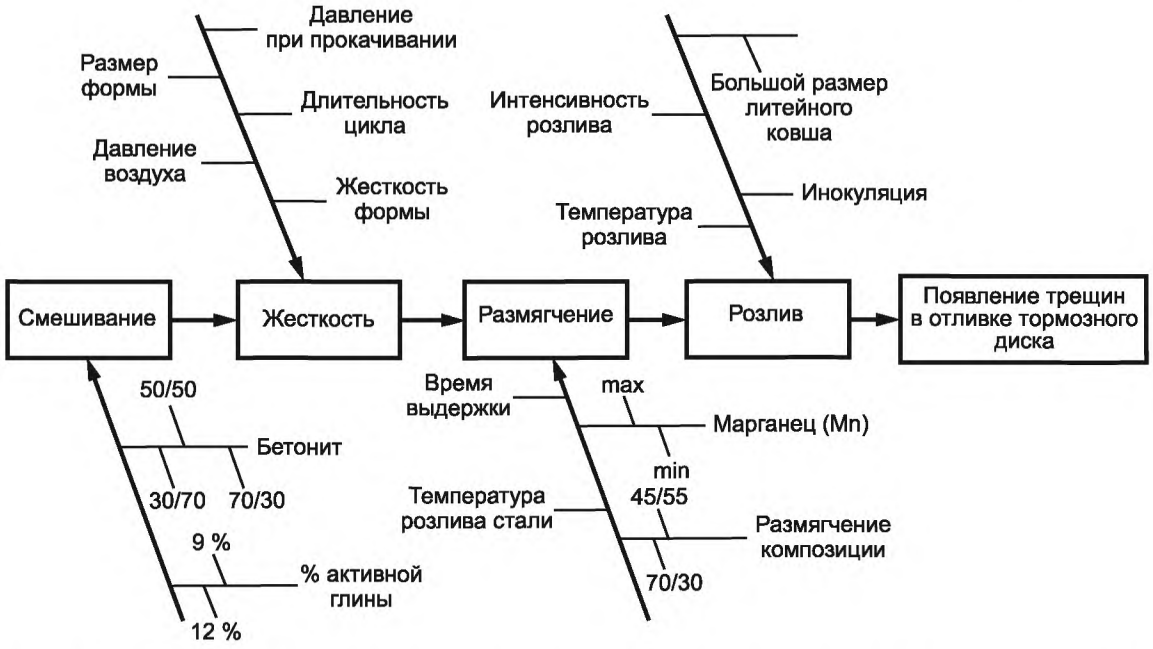


Рисунок 14 — Диаграмма причин и последствий появления трещин в отливке для литейного производства

Библиография

[1] Большев Л.Н., Смирнов Н.В. Таблицы математической статистики, М.: Наука, 1983

УДК 658.562.012.7:65.012.122:006.354 ОКС 03.120.30 Т59

Ключевые слова: статистические методы, среднее арифметическое, медиана, изменчивость данных, причины изменчивости данных

Редактор *С.Д. Золотова*
Технический редактор *В.Н. Прусакова*
Корректор *Е.Д. Дульнева*
Компьютерная верстка *Ю.В. Демениной*

Сдано в набор 03.10.2013. Подписано в печать 24.10.2013. Формат 60 × 84⅛. Гарнитура Ариал.
Усл. печ. л. 2,79. Уч.-изд. л. 1,95. Тираж 124 экз. Зак. 1224.

ФГУП «СТАНДАРТИНФОРМ», 123995 Москва, Гранатный пер., 4.
gost.e-ecolog.ru gost.e-ecolog.ru
Набрано во ФГУП «СТАНДАРТИНФОРМ» на ПЭВМ.
Отпечатано в филиале ФГУП «СТАНДАРТИНФОРМ» — тип. «Московский печатник», 105062 Москва, Лялин пер., 6.

gost.e-ecolog.ru