В.В. Нешитой

СТАТИСТИЧЕСКИЙ АНАЛИЗ И РЕГУЛИРОВАНИЕ ТЕХНОЛОГИЧЕСКИХ ПРОЦЕССОВ НА БАЗЕ ОБОБЩЕННЫХ РАСПРЕДЕЛЕНИЙ С ПАРАМЕТРОМ СДВИГА

Методические рекомендации

Минск
БелГИСС
2001

УДК 658.562.012.7(083.131)
ББК 30.607
Н 59

Нешитой В.В.

Статистический анализ и регулирование технологических процессов на базе обобщенных распределений с параметром сдвига: Метод. рекомендации / В.В. Нешитой.– Мн.: БелГИСС, 2001. – 40 с.

Н 59

ISBN 985-6646-09-X

Настоящие методические рекомендации предназначены для всех специалистов, использующих статистические методы в управлении качеством продукции.

Существенное отличие их от других подобных рекомендаций заключается в том, что в данном документе для описания статистических распределений производственных погрешностей контролируемых параметров используются не отдельные распределения (нормальный закон, Вейбулла, Релея, Максвелла и др.), а обобщенные четырехпараметрические распределения автора, включающие как частные случаи большое количество известных распределений, образующих единую систему.

Этим достигается новый, значительно более высокий уровень точности решения прикладных задач, связанных с выравниванием статистических распределений.

Задачи по установлению наилучшей выравнивающей кривой распределения, нахождению оценок параметров, вычислению различных показателей, характеризующих состояние технологического процесса, доведены автором до программной реализации (в виде серии программ под общим названием SNR – системы непрерывных распределений).

УДК 658.562.012.7(083.131)
ББК 30.607

ISBN 985-6646-09-X

ВВЕДЕНИЕ

Настоящие методические рекомендации предназначены для статистического анализа точности и стабильности технологических процессов с целью их подналадки и последующего регулирования.

Точность технологического процесса характеризуется степенью близости действительных и нормативных значений параметров производимой продукции.

Стабильность технологического процесса – это его свойство сохранять неизменными значения основных параметров в течение некоторого интервала времени без вмешательства извне.

Статистическое регулирование технологического процесса – это его корректировка по результатам выборочного контроля параметров производимой продукции (объем выборки, как правило, не превышает 30 единиц продукции).

Статистический анализ производится на основе результатов измерения контролируемых параметров.

Значения этих параметров имеют некоторый разброс (рассеяние) относительно среднего. Чем меньше рассеяние, тем точнее технологический процесс.

Для описания статистических закономерностей разброса значений однородных случайных величин в настоящее время используются отдельные вероятностные модели, так называемые **законы распределения вероятностей случайных величин**, например нормальный закон, Вейбулла, Релея и некоторые другие. Но известные законы не могут с достаточной точностью описать все многообразие статистических распределений, встречающихся на практике. В то же время **решающим условием успешного анализа и регулирования технологического процесса является нахождение такого теоретического выравнивающего закона распределения, который наиболее точно описывает статистическое распределение**. Спрашивается, зачем нужна максимально высокая точность выравнивания статистического распределения? Ответ заключается в том, что **главный и наиболее общий показатель качества – ожидаемый процент брака – сосредоточен на концах распределения,** причем предельно допустимый его уровень в машиностроении составляет малую величину, равную 0,0027, или 0,27 %. Поэтому **грубое выравнивание просто не имеет смысла.**

Чтобы увеличить вероятность правильного установления типа выравнивающего распределения, в настоящих методических рекомендациях используются обобщенные четырехпараметрические распределения автора, которые включают как частные случаи большинство известных распределений, в том числе семейство кривых К.Пирсона, и образуют три системы непрерывных распределений (сокращенно SNR). Последние доведены автором до программной реализации.

Использование обобщенных распределений и соответствующих программ значительно упрощает и ускоряет процедуру нахождения выравнивающего распределения, поскольку оно **вычисляется** по статистическому ряду распределения за один прием без перебора отдельных частных случаев и проверки каждого из них по критериям согласия.

Основные задачи статистического анализа технологического процесса заключаются в том, чтобы [1]:

– дать оценку его качества по результатам измерения контролируемых параметров;

– при необходимости привести технологический процесс в **статистически управляемое состояние;**

– выявить основные факторы, влияющие на величину случайных и систематических погрешностей;

– установить закономерности изменения во времени показателей качества технологического процесса.

Статистически управляемый процесс – это технологический процесс, обеспечивающий в пределах нормы показатели рассеяния и уровня настройки контролируемого параметра. **Такой технологический процесс поддается статистическому регулированию.**

Методические рекомендации устанавливают порядок расчета на ПЭВМ показателей качества технологических процессов на базе обобщенных распределений, содержащих параметр сдвига.

Для расчетов используется программа SNR11M97 (в дальнейшем – Программа).

Методические рекомендации могут быть использованы для решения различных задач, в том числе [6, 11]:

– статистической оценки технологической точности производственного оборудования во время эксплуатации, перед сдачей в ремонт, после ремонта, при подготовке к внедрению статистического регулирования с помощью контрольных карт и в других необходимых случаях;

– статистической оценки технологической точности нового оборудования;

– корректировки конструкторской документации на основе статанализа результатов испытаний новых образцов и серийных изделий;

– устранения несоответствия между заданной точностью и возможностями реальных технологических процессов;

– выявления резервов производства и технологии;

– анализа контрольных карт и статистического регулирования технологических процессов;

– контроля качества механических свойств металлов;

– установления необходимости ремонта оборудования;

– определения качества выполненного ремонта оборудования;

– сравнительной оценки точности вариантов технологического процесса, оборудования и оснастки;

– сравнительной оценки точности режимов обработки;

– оценки эффективности управляющих воздействий.

1 ОБЩИЕ ПОЛОЖЕНИЯ

1.1 В соответствии с настоящими методическими рекомендациями вычисляются следующие основные показатели технологических процессов:

– технологический допуск;

– коэффициент точности;

– коэффициент уровня настройки;

– ожидаемый процент брака.

1.2 На базе статистического ряда распределения, полученного в результате замеров контролируемого параметра, с помощью Программы на ПЭВМ вычисляется выравнивающее распределение и точечные оценки его параметров.

1.3 По найденному закону распределения вычисляются показатели, характеризующие качество технологического процесса.

1.4 Программа вычисляет:

– показатели статистического распределения контролируемого параметра, включая среднее, дисперсию, среднее квадратическое отклонение, коэффициент вариации, центральные моменты 3 и 4-го порядков, показатели асимметрии и островершинности и некоторые другие;

– тип выравнивающей кривой;

– оценки параметров и нормирующего множителя выравнивающей кривой;

– значения плотности и функции распределения;

– квантили;

– процентили;

– доверительные вероятности;

– доверительные интервалы;

– координаты характерных точек – моды и точек перегиба;

– границы технологического допуска;

– показатель точности процесса (коэффициент рассеяния);

– показатель уровня настройки (коэффициент смещения);

– ожидаемый процент брака (общий, в том числе на нижней и верхней границах конструкторского допуска).

1.5 Программа строит статистическую и выравнивающую кривые распределения с указанием границ конструкторского и технологического допусков.

1.6 Программа выводит на экран все результаты расчетов.

1.7 Работа пользователей программы производится в несколько этапов.

На первом этапе осуществляется сбор статистических данных о контролируемом параметре, которые характеризуют состояние технологического процесса.

На втором этапе с помощью Программы осуществляется статистическая обработка данных, вычисляется выравнивающее распределение и на его основе производятся необходимые расчеты.

На третьем этапе по результатам расчетов принимаются соответствующие решения, т. е. либо о необходимости подналадки технологического процесса, либо продолжении работы.

Для подналадки технологического процесса необходимо выяснить причины его разладки и по возможности их устранить.

После устранения причин разладки необходимо повторить статистический анализ технологического процесса, чтобы оценить эффективность управляющих воздействий и в итоге добиться снижения брака ниже допустимого уровня.

Работу по первому этапу выполняет специалист-метролог.

Работу по второму этапу выполняет специалист статистической лаборатории, хорошо владеющий методами статистической обработки данных и в частности статанализа технологических процессов.

Работу по анализу причин разладки технологического процесса и их устранению проводит служба главного технолога.

2 ПОДГОТОВКА ИСХОДНЫХ ДАННЫХ

2.1 Статистический анализ технологического процесса заключается в выявлении закона распределения производственных погрешностей при производстве продукции, нахождении оценок его параметров и вычислении необходимых показателей, характеризующих состояние технологического процесса. При неудовлетворительных значениях показателей технологический процесс приводится в **статистически управляемое состояние.**

2.2 Для установления закона распределения контролируемого параметра необходимо отобрать не меньше 100 единиц продукции (по ряду мгновенных выборок при неизменной настройке технологического процесса) и измерить значения контролируемого параметра.

2.3 Средства измерения должны обеспечивать точность в пределах 1/10 – 1/20 ширины конструкторского допуска.

2.4 Результаты замеров заносятся в **контрольный листок (протокол измерения)**, где указывается код детали, номинальный размер контролируемого параметра, конструкторский допуск, точность средства измерения, время и исполнитель.

2.5 Контрольный листок за подписями контролера и технолога цеха передается последним в статистическую лабораторию на обработку.

2.6 По результатам статистической обработки контрольных листков с помощью Программы служба главного технолога совместно со статистической лабораторией и ОТК проводит анализ причин разладки технологического процесса и разрабатывает мероприятия по его отладке.

2.7 После приведения технологического процесса в статистически управляемое состояние для дальнейшего его регулирования по количественному признаку с помощью контрольных карт достаточно через определенные интервалы времени отбирать выборки из 25 – 30 замеров контролируемых параметров. Все последующие расчеты выполняются на ПЭВМ по той же Программе при несгруппированных данных.

2.8 На контрольной карте горизонтальными линиями отмечаются среднее значение контролируемого параметра, нижняя и верхняя границы регулирования.

2.9 Через определенные интервалы времени на контрольную карту наносятся выборочные значения контролируемых параметров.

Процесс считается налаженным, если выборочные значения контролируемых параметров находятся внутри границ регулирования.

2.10 Для регулирования уровня настройки (наладки) процесса используется контрольная карта среднего арифметического значения контролируемого параметра.

Для регулирования рассеяния – контрольная карта среднего квадратического отклонения или размаха.

Использование обобщенных распределений и соответствующих программ позволяет строить **контрольную карту ожидаемого уровня брака.**

3 СТАТИСТИЧЕСКАЯ ОБРАБОТКА ПРОИЗВОДСТВЕННЫХ ПОГРЕШНОСТЕЙ

3.1 Результаты замеров контролируемого параметра, записанные в контрольном листке, представляют собой **простой статистический ряд.**

3.2 Они могут быть обработаны на ПЭВМ с помощью Программы непосредственно либо после предварительной группировки в интервалы заданной ширины.

3.3 Число интервалов группирования можно оценить по формуле

$$k = 1 + 3{,}32 \cdot lg\, n,$$

округлив полученный результат до целых, где n – объем выборки (количество замеров).

3.4 Ширина интервала рассчитывается по формуле

$$h = \frac{x_{max} - x_{min}}{k}.$$

Величину h округляют в большую сторону.

Тогда произведение hk будет несколько больше размаха $R = x_{max} - x_{min}$ на величину ε, т.е.

$$\varepsilon = hk - R \quad (\varepsilon > 0).$$

3.5 Первый интервал имеет границы

$$x_{min} - \frac{\varepsilon}{2}; \quad x_{min} - \frac{\varepsilon}{2} + h.$$

Границы каждого следующего интервала увеличиваются на ширину интервала h.

3.6 Середина первого интервала равна

$$x_{min} - \frac{\varepsilon}{2} + \frac{h}{2}.$$

Середина каждого следующего интервала увеличивается на ширину интервала h.

3.7 Частота i-го интервала n_i определяется подсчетом числа замеров на i-м интервале.

3.8 Эмпирическая плотность распределения i-го интервала вычисляется по формуле

$$p_i = \frac{n_i}{nh}.$$

3.9 На базе полученного таким образом **интервального статистического ряда** строят **гистограмму.** По горизонтальной оси откладывают интервалы, над которыми строят прямоугольники высотой, равной эмпирической плотности i-го интервала p_i.

Гистограмма является одной из форм представления статистического закона распределения случайной величины.

Вместо гистограммы можно построить **полигон**, если соединить отрезками прямых середины верхних сторон прямоугольников.

3.10 Для сглаживания гистограммы и полигона используют **график подходящей плотности распределения**, который называется **выравнивающей кривой распределения.**

3.11 Выравнивающая кривая распределения вычисляется по Программе и используется в дальнейшем для **расчета показателей качества технологического процесса.**

4 ПОКАЗАТЕЛИ ТОЧНОСТИ И СТАБИЛЬНОСТИ ТЕХНОЛОГИЧЕСКОГО ПРОЦЕССА

4.1 В качестве меры точности технологического процесса используется **коэффициент точности (рассеяния)** [4]

$$K_T = \frac{\omega}{\delta} = \frac{X_в - X_н}{T_в - T_н},$$

где $\omega = X_в - X_н$ – ширина **поля рассеяния** (технологнческий допуск);

$X_в$, $X_н$ – верхняя и нижняя границы поля рассеяния;

$\delta = T_в - T_н$ – ширина **поля конструкторского допуска**;

$T_в$, $T_н$ – верхняя и нижняя границы поля конструкторского допуска.

В зарубежной литературе вместо показателя K_T используется коэффициент $C_p = 1/K_T$.

Ширина поля рассеяния вычисляется при заданной доверительной вероятности, например P = 0,9973 (или P = 0,9545), т. е. при условии, что 99,73 % (или 95,45 %) значений контролируемого параметра находятся внутри границ поля рассеяния. Для нормального закона $\omega = 6S$, где S – **выборочное среднее квадратическое отклонение**. Эта величина зависит от вида закона распределения и значений его параметров.

При использовании обобщенных распределений ширину поля рассеяния целесообразно определять из того же условия, т. е. $P = F(X_в) - F(X_н) = 0{,}9973$, при этом значения функции распределения равны: $F(X_н) = 0{,}00135$, $F(X_в) = 1 - 0{,}00135 = 0{,}99865$.

Точность технологического процесса (по крайней мере для нормального закона распределения погрешностей) считается достаточной при условии $K_T \le 0{,}75$. Как будет показано ниже, эта величина зависит от объема выборки.

4.2 Для характеристики наладки технологического процесса применяют коэффициент точности настройки, или показатель уровня настройки [4]

$$K_н = \frac{E}{\delta} = \frac{\left|\overline{X} - T_0\right|}{T_в - T_н},$$

где $\overline{X}$ – среднее выборочное значение контролируемого параметра;
$T_0 = (T_в + T_н)/2$ – середина поля конструкторского допуска.

Коэффициент точности настройки характеризует систематическую погрешность и может быть устранен.

4.3 Нестабильность технологического процесса по уровню наладки характеризуют коэффициентом смещения настройки [4]

$$K_{см} = \frac{\overline{X}(t) - \overline{X_0}}{\delta},$$

где $\overline{X_0}$, $\overline{X}(t)$ – начальное и конечное (на момент времени t) значения центра распределения контролируемого параметра.

4.4 Нестабильность технологического процесса по рассеянию характеризуют коэффициентом межнастроечной стабильности [4]

$$K_{м.с.} = \frac{S(t)}{S_0},$$

где S_0, $S(t)$ – начальное и конечное (на момент времени t) значения среднего квадратического отклонения контролируемого параметра.

Для вычисления двух последних коэффициентов необходимо иметь две выборки, полученные в разные моменты времени.

4.5 В заключение статанализа технологического процесса вычисляют ожидаемый уровень брака q в процентах. Он находится по формуле

$$q = 100\ \% - \left[F(T_в) - F(T_H)\right] \cdot 100\ \%.$$

При этом брак на нижней границе поля конструкторского допуска равен значению функции распределения $F(T_н)$, умноженной на 100 %

$$q_н = F(T_н) \cdot 100\ \%,$$

а на верхней границе поля конструкторского допуска

$$q_в = \left[1 - F(T_в)\right] \cdot 100\ \%.$$

Ожидаемый процент брака является наиболее общей и важной характеристикой качества технологического процесса.

При условии, когда границы конструкторского и технологического допусков совпадают, коэффициент точности $K_T = 1$, а ожидаемый уровень брака равен 0,27 % при любом законе распределения производственных погрешностей. Это предельно допустимый уровень брака.

4.6 Если $K_T < 1$, а $0 \le \left|K_H\right| \le \frac{1}{2}(1 - K_T)$, ожидаемый уровень брака не превышает предельно допустимого уровня при любом законе распределения производственных погрешностей.

В случае обобщенных распределений коэффициент уровня настройки целесообразно вычислять по формуле

$$K_н = \frac{\left|X_0 - T_0\right|}{T_в - T_н},$$

где $X_0 = \frac{X_в + X_н}{2}$ – середина поля рассеяния.

4.7 Значение коэффициента точности зависит от объема выборки. Для статистически управляемого процесса ширина поля рассеяния должна быть меньше ширины конструкторского допуска как минимум на величину $6\sigma_x / \sqrt{n}$ (при нормальном законе распределения). Тогда коэффициент точности будет равен:

$$K_T = \frac{X_в - X_н}{T_в - T_н} = \frac{X_в - X_н}{X_в - X_н + \frac{6\sigma_x}{\sqrt{n}}}.$$

Поскольку ширина поля рассеяния равна $X_в - X_н = 6\sigma_x$, то

$$K_T = \frac{6\sigma_x}{6\sigma_x + \frac{6\sigma_x}{\sqrt{n}}} = \frac{\sqrt{n}}{1+\sqrt{n}}.$$

При $n = 10$ $K_T = 0{,}760$. При $n = 30$ $K_T = 0{,}846$. При $n = 100$ $K_T = 0{,}909$.

4.8 Показатель уровня настройки должен быть задан на интервале $0 \le |K_H| \le \frac{1}{2}(1 - K_T)$ или $0 \le |K_H| \le \frac{1}{2(1+\sqrt{n})}$. При $n = 10$ $0 \le |K_H| \le 0{,}120$. При $n = 30$ $0 \le |K_H| \le 0{,}077$. При $n = 100$ $0 \le |K_H| \le 0{,}045$.

При выполнении этих условий брак не превышает предельно допустимого уровня.

4.9 При заданных границах конструкторского допуска и объеме выборки n требуемое значение среднего квадратического отклонения в случае нормального закона распределения производственных погрешностей равно

$$\sigma_x = \frac{T_в - T_н}{6\left(1 + \frac{1}{\sqrt{n}}\right)}.$$

4.10 При заданных σ_x и n требуемое значение допуска равно

$$T_в - T_н = 6\sigma_x\left(1 + \frac{1}{\sqrt{n}}\right) = \frac{6\sigma_x}{K_T}.$$

Приложение 1
(справочное)

1 Основные понятия теории вероятностей и математической статистики

1.1 Случайные величины и их законы распределения

1.1.1 Случайной называют величину, которая при осуществлении испытаний может принять любое из возможных значений, причем заранее неизвестно, какое именно.

Случайные величины подразделяются на **дискретные и непрерывные**.

Дискретная случайная величина принимает возможные значения, которые заранее могут быть перечислены.

Для непрерывной случайной величины невозможно перечислить все возможные ее значения, т. к. они непрерывно заполняют некоторый промежуток.

1.1.2 Совокупность всех возможных значений случайной величины X образует **полную группу событий**, а сумма всех возможных вероятностей случайной величины X равна единице. **Эта суммарная вероятность распределена между всеми возможными значениями** x_i.

1.1.3 Всякое соотношение, устанавливающее взаимосвязь между возможными значениями случайной величины и их вероятностями, называется **законом распределения случайной величины.**

1.1.4 Закон распределения может быть задан различными способами: в виде таблицы, графика или аналитически.

При табличном задании закона распределения случайной величины X в верхней строке записываются возможные значения x_i, а в нижней – соответствующие им вероятности p_i.

При графическом представлении закона распределения строится многоугольник распределения: по оси абсцисс откладываются возможные значения, а по оси ординат – вероятности этих значений. Для наглядности через вершины построенных отрезков проводится ломаная линия.

Аналитически закон распределения случайной величины X задается в виде функции

$$p_i = f(x_i)\,.$$

1.1.5 Закон распределения дискретных и непрерывных случайных величин может быть задан **интегральной функцией распределения** или просто **функцией распределения *F(x)***, которая равна вероятности $P(X < x)$

$$F(x) = P(X < x) = P\,(-\infty < X < x)\,.$$

При этом $F(-\infty) = 0$, $F(+\infty) = 1$.

График функции распределения дискретной случайной величины имеет вид ломаной линии с горизонтальными ступеньками, а непрерывной случайной величины – монотонной кривой, возрастающей от 0 до 1.

1.1.6 Вероятность попадания случайной величины X на заданный интервал $x_1 < X < x_2$ выражается через функцию распределения

$$P(x_1 < X < x_2) = F(x_2) - F(x_1).$$

Для непрерывной случайной величины это равенство имеет вид

$$P(\alpha < X < \beta) = \int_{\alpha}^{\beta} f(x)dx = F(\beta) - F(\alpha),$$

где $f(x) = dF(x)/dx$.

1.1.7 Функция $f(x)$ называется **плотностью распределения** случайной величины X, или дифференциальным законом распределения.

При заданной плотности распределения непрерывной случайной величины X функция распределения вычисляется по формуле

$$F(x) = \int_{-\infty}^{x} f(x)dx.$$

Любая плотность распределения должна удовлетворять условию нормировки

$$\int_{-\infty}^{\infty} f(x)dx = 1.$$

График плотности распределения называется кривой распределения.

1.2 Числовые характеристики случайных величин

1.2.1 Закон распределения является исчерпывающей характеристикой случайной величины, хотя далеко не всегда может быть точно установлен его вид.

На практике во многих случаях бывает достаточно знать некоторые числовые характеристики исследуемой случайной величины, которые участвуют в различных расчетах. Наиболее часто используются математическое ожидание (М.о.), мода (Мо), медиана (Ме), квантиль, а также моменты – начальные и центральные и некоторые другие показатели.

1.2.2 Математическое ожидание дискретной случайной величины X равно

$$M(X) = \sum_{i=1}^{n} x_i p_i = m_x.$$

Для непрерывной случайной величины X

$$M(X)=\int_{-\infty}^{\infty} xf(x)dx=m_x .$$

Оценкой М.о. является среднее арифметическое (обозначается $\bar{x}$) возможных значений случайной величины X

$$\bar{x}=\frac{1}{n}\sum_{i=1}^{n} x_i .$$

1.2.3 Модой случайной величины X называется наиболее вероятное ее значение или значение случайной величины с наибольшей плотностью распределения.

1.2.4 Медиана случайной величины X определяется соотношением

$$P(X<Me)=P(X>Me).$$

Она делит площадь под кривой распределения пополам.

1.2.5 Точка x_p, в которой функция распределения равна вероятности P

$$F(x_p)=P,$$

называется **квантилью** с заданным уровнем вероятности P.

Квантиль с уровнем вероятности $P = 0{,}5$ совпадает с медианой.

1.2.6 Начальный момент *r*-го порядка задается формулами:

– для дискретной случайной величины X

$$\nu_r=M(X^r)=\sum_{i=1}^{n} x_i^r p_i ;$$

– для непрерывной случайной величины X

$$\nu_r=M(X^r)=\int_{-\infty}^{\infty} x^r f(x)dx .$$

Начальный момент первого порядка представляет собой математическое ожидание m_x

$$\nu_1=\sum_{i=1}^{n} x_i p_i=m_x .$$

Эмпирические (выборочные) начальные моменты порядка r вычисляются по формуле

$$\nu_r^*=\frac{1}{n}\sum_{i=1}^{n} x_i^r .$$

Если n возможных значений непрерывной случайной величины X сгруппированы в k интервалов шириной h, то

$$\nu_r^* = \sum_{i=1}^{k} x_i^r p_i^* h_i = \sum_{i=1}^{k} x_i^r \frac{n_i}{n},$$

где x_i – середина i-го интервала;

n_i – число событий в i-м интервале (частота интервала);

h_i – ширина i-го интервала;

$p_i^* = n_i/(nh_i)$ – эмпирическая плотность распределения;

$n = \sum_{i=1}^{k} n_i$ – объем выборки.

1.2.7 Центральный момент *r*-го порядка задается формулами:

– для дискретной случайной величины X

$$\mu_r = M\left[(X - m_x)^r\right] = \sum_{i=1}^{n} (x_i - m_x)^r p_i,$$

– для непрерывной случайной величины X

$$\mu_r = \int_{-\infty}^{\infty} (x - m_x)^r f(x)dx.$$

Эмпирические центральные моменты вычисляются по формуле

$$\mu_r^* = \frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})^r,$$

где $\bar{x}$ – среднее значение случайной величины X.

При небольших объемах выборки ($n < 30$) можно использовать исправленные (несмещенные) оценки для центральных моментов [5, с. 386]

$$M_2 = \frac{n}{n-1}\mu_2^*$$

$$M_3 = \frac{n^2}{(n-1)(n-2)}\mu_3^*$$

$$M_4 = \frac{n(n^2 - 2n + 3)}{(n-1)(n-2)(n-3)}\mu_4^* - \frac{3n(2n-3)}{(n-1)(n-2)(n-3)}\mu_2^{*2},$$

но более целесообразно принимать объем выборки $n \geq 30$ и не вводить указанных поправок.

Если данные сгруппированы в k интервалов шириной h_i, то

$$\mu_r^* = \sum_{i=1}^{k}\left(x_i - \bar{x}\right)^r p_i h_i = \sum_{i=1}^{k}\left(x_i - \bar{x}\right)^r \frac{n_i}{n},$$

где x_i – середина i-го интервала; p_i – эмпирическая плотность i-го интервала; n_i – частота i-го интервала.

1.2.8 Центральный момент второго порядка называется **дисперсией** и обозначается $D(X)$:

$$D(X) = \mu_2 = M\left[(X - m_x)^2\right] = M\left(X^2\right) - m_x^2 = \nu_2 - \nu_1^2.$$

1.2.9 Квадратный корень из дисперсии называется **средним квадратическим отклонением**

$$\sigma(x) = \sigma_x = \sqrt{D(X)}.$$

Эмпирическое среднее квадратическое отклонение обозначается $S(x)$.

Размерность величины σ_x совпадает с размерностью самой случайной величины X.

Величины $D(X)$ и σ_x характеризуют **рассеяние случайной величины X** около ее математического ожидания m_x – **центра распределения**.

При этом **среднее квадратическое отклонение среднего n одинаково распределенных взаимно независимых случайных величин в** $\sqrt{n}$ **раз меньше среднего квадратического отклонения отдельной случайной величины X**, т. е.

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}}.$$

1.2.10 Форма кривой распределения характеризуется двумя показателями: **асимметрии** (скошенности) – β_1 и **островершинности** (эксцесса) – β_2.

Показатель асимметрии вычисляется по формуле

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3}.$$

Для симметричных распределений центральный момент 3-го порядка $\mu_3 = 0$ и, следовательно, $\beta_1 = 0$. При $\mu_3 > 0$ распределение имеет положительную (правостороннюю) асимметрию – правая ветвь кривой распределения более пологая и длинная, а при $\mu_3 < 0$ – левостороннюю асимметрию.

Показатель островершинности рассчитывается по формуле

$$\beta_2 = \frac{\mu_4}{\mu_2^2}.$$

Показатели β_1, β_2 введены К. Пирсоном.

Достоинством показателей асимметрии и островершинности является то, что они зависят только от **параметров формы** (как правило, не более двух) и не зависят от **параметров масштаба и сдвига.** С их помощью может быть установлен **тип кривой распределения** и найдены **оценки параметров выравнивающей (аппроксимирующей) кривой**, т. е. «теоретической» кривой, которая наилучшим образом описывает (выравнивает) **статистическое распределение**, представленное, например, гистограммой.

Другим достоинством показателей асимметрии и островершинности, а также моментов является возможность и простота вычисления этих величин как по сгруппированным, так и несгруппированным статистическим данным.

Наиболее успешно эти показатели могут быть использованы в случае распределений, заданных на ограниченном с обеих сторон интервале.

1.3 Сравнение средних и дисперсий двух случайных величин

1.3.1 Сравнение средних двух случайных величин.

На практике часто возникает необходимость проверки равенства средних значений двух случайных величин.

Рассмотрим два случая [3].

Случай 1. Две случайные величины X, Y подчиняются нормальному закону распределения с **неизвестными математическими ожиданиями** a_1, a_2 и **известными дисперсиями** σ_1^2 и σ_2^2.

Требуется проверить нулевую гипотезу $H_0 : a_1 = a_2$ против альтернативной гипотезы $H_a : a_1 \neq a_2$ на основании двух независимых выборок объемами n_1 и n_2, извлеченных из исследуемых генеральных совокупностей.

Проверка осуществляется в следующем порядке.

1. Вычисляем средние арифметические значения

$$\bar{x} = \frac{1}{n}\sum x_i; \quad \bar{y} = \frac{1}{n}\sum y_i .$$

2. Вычисляем дисперсию разности $\bar{x} - \bar{y}$:

$$D\left(\bar{x} - \bar{y}\right) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}.$$

3. Вычисляем критерий u

$$u = \frac{\bar{x} - \bar{y}}{\sqrt{D\left(\bar{x} - \bar{y}\right)}}.$$

4. Задаем уровень значимости $\alpha \, (\alpha \le 0{,}1)$.

5. По таблице нормального распределения при заданном уровне значимости α находим критическое значение $u_{\alpha/2}$, удовлетворяющее условию

$$P(|u| \ge u_{\alpha/2}) = \alpha .$$

6. При $|u_{набл.}| < u_{\alpha/2}$ нет оснований отклонять нулевую гипотезу. В противном случае нулевая гипотеза отклоняется в пользу альтернативной.

Случай 2. Две случайные величины *X*, *Y* подчиняются нормальному закону распределения с **неизвестными математическими ожиданиями** a_1, a_2 **и неизвестными, но равными между собой дисперсиями** $\sigma_1^2 = \sigma_2^2$.

Проверка нулевой гипотезы $H_0 : a_1 = a_2$ против альтернативной гипотезы $H_a : a_1 \ne a_2$ осуществляется в следующем порядке.

1. Вычисляем выборочную статистику

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2 - 2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}},$$

где $\bar{x}$ и $\bar{y}$ – средние арифметические;

$$S_1^2 = \frac{1}{n_1}\sum (x_i - \bar{x})^2; \quad S_2^2 = \frac{1}{n_2}\sum (y_i - \bar{y})^2$$ – оценки дисперсий.

2. Задаем уровень значимости α.

3. По таблице распределения Стьюдента при заданной вероятности α и числе степеней свободы $n_1 + n_2 - 2$ находим критическую точку $t_{\alpha/2; n_1+n_2-2}$.

4. Если $|t_{набл.}| < t_{\alpha/2; n_1+n_2-2}$, то нет оснований для отклонения нулевой гипотезы. В противном случае нулевая гипотеза отклоняется в пользу альтернативной.

1.3.2 Сравнение дисперсий двух случайных величин.

Из двух нормальных генеральных совокупностей извлечены две выборки объемами n_1 и n_2. **Параметры нормальных законов** $a_1, a_2, \sigma_1, \sigma_2$ **неизвестны.**

Требуется проверить нулевую гипотезу $H_0 : \sigma_1^2 = \sigma_2^2$ против альтернативной гипотезы $H_a : \sigma_1^2 \ne \sigma_2^2$.

Проверка осуществляется в следующем порядке [3].

1. Вычисляем средние значения и оценки дисперсий для обеих выборок

$$\bar{x}_1 = \frac{1}{n_1}\sum x_{i1}; \quad \bar{x}_2 = \frac{1}{n_2}\sum x_{i2};$$

$$S_1^2 = \frac{1}{n_1 - 1}\sum (x_{i1} - \bar{x}_1)^2;$$

$$S_2^2 = \frac{1}{n_2 - 1}\sum (x_{i2} - \bar{x}_2)^2.$$

2. Вычисляем критерий Фишера

$$F = \frac{S_1^2}{S_2^2},$$

где S_1^2 и S_2^2 – соответственно наибольшая и наименьшая оценки дисперсий.

3. Задаем уровень значимости α критерия F.

4. По таблице квантилей F – распределения по уровню значимости α и числу степеней свободы $\nu_1 = n_1 - 1$ и $\nu_2 = n_2 - 1$ находим критическую точку $F_{\alpha/2;\ \nu_1;\ \nu_2}$.

5. Если $F_{набл.} < F_{\alpha/2;\ \nu_1;\ \nu_2}$, то нет оснований для отклонения нулевой гипотезы. В противном случае нулевая гипотеза отклоняется в пользу альтернативной.

1.4 Оценка коэффициента корреляции и индекса корреляции

1.4.1 Оценка коэффициента корреляции.

Для оценки степени тесноты **линейной связи между случайными величинами *X* и *Y*** используется коэффициент корреляции $r_{y/x}$ [2, с.319].

Если между случайными величинами X и Y существует функциональная зависимость, то $r_{y/x} = \pm 1$; если X и Y независимы, то $r_{y/x} = 0$.

Коэффициент корреляции вычисляется по формуле

$$r_{y/x} = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n\sum x_i^2 - \left(\sum x_i\right)^2}\sqrt{n\sum y_i^2 - \left(\sum y_i\right)^2}}.$$

1.4.2 Оценка индекса корреляции.

Индекс корреляции $R_{y/x}$ используется в качестве меры тесноты связи **в случае криволинейной зависимости между признаками *X* и *Y*.**

Он представляет собой **корреляционное отношение**, вычисленное на основании результатов выравнивания y по x по некоторой линии [2, с. 329].

Индекс корреляции вычисляется по формуле

$$R_{y/x} = \sqrt{1 - \frac{\sigma^2_{y/x}}{\sigma^2_y}},$$

где $\sigma^2_y = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \overline{y}\right)^2$; $\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$; $\sigma^2_{y/x} = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \overline{y_x}\right)^2$;

$\overline{y}$ – среднее значение y;

$\overline{y}_x$ – вычисленное значение y при заданном x по некоторой выравнивающей кривой.

В случае линейной зависимости между случайными величинами *Y* и *X* индекс корреляции численно равен линейному коэффициенту корреляции.

Индекс корреляции является показателем степени близости линии регрессии к фактическим данным. Он может принимать значения на интервале $0 \le R_{y/x} \le 1$.

1.5 Анализ временных рядов

Статистические данные за ряд периодов (дней, часов), например, средние значения или средние квадратические отклонения контролируемого показателя производственного процесса представляют собой **временной (динамический) ряд.**

Он содержит как детерминированную **(неслучайную) составляющую,** которую называют **трендом,** так и **случайную составляющую**.

Для анализа технологического процесса важным является выделение тренда. Решение этой задачи дает возможность осуществлять прогноз и, следовательно, предсказывать момент наступления разладки технологического процесса.

Один из методов выделения тренда заключается в подборе наиболее подходящей выравнивающей кривой. При этом в качестве математических моделей используются различные функции: уравнения прямой, экспоненты, параболы и т. д.

Для описания динамических рядов целесообразно использовать уравнение автора

$$y = y_0\left(1 + \alpha u t\right)^{\frac{1}{u}} = \left(A + Bt\right)^{\frac{1}{u}}, \tag{1.1}$$

где $A = y_0^u$; $B = \alpha u y_0^u$.

Последнее уравнение включает как частные случаи множество функций, в том числе прямую при $u = 1$ и экспоненту при $u \to 0$.

Действительно, при $u \to 0$ из (1.1) имеем

$$y = y_0 e^{\alpha t} = e^{\ln y_0 + \alpha t} . \qquad (1.2)$$

Для нахождения оценок параметров уравнения (1.1) его следует привести к виду

$$y^u = A + Bt .$$

Обозначив $y^u = Y$, можем записать

$$Y = A + BT.$$

Это уравнение прямой. Оценки параметров A, B (при заданном u) находятся **по методу наименьших квадратов**:

$$B = \frac{n\sum T_i Y_i - \sum T_i \sum Y_i}{n\sum T_i^2 - \left(\sum T_i\right)^2}; \quad A = \frac{1}{n}\left(\sum Y_i - B\sum T_i\right) .$$

Значения параметра u задаются с некоторым шагом, например, $\Delta u = 0{,}1$ или $\Delta u = 0{,}01$ или другим шагом. При меньшем шаге достигается более высокая точность аппроксимации.

При каждом заданном значении параметра u, а также вычисленных оценках параметров A, B необходимо вычислять коэффициент корреляции $r_{Y/T}$ и индекс корреляции. При достижении наибольшего значения коэффициента корреляции (или индекса корреляции) вычислительный процесс прекращается.

Эти расчеты осуществляются на ПЭВМ по заранее разработанной программе, например программе автора GRAF25F.

Приложение 2
(справочное)

2 Система непрерывных распределений, содержащих параметр сдвига

2.1 Для выравнивания статистических распределений контролируемых показателей технологического процесса используется одна из систем непрерывных распределений, заданная обобщенными плотностями В. Нешитого [7-10]:

$$p(x) = N(x - l)^{k\beta - 1}\,[1 - \alpha u(x - l)^{\beta}]^{\frac{1}{u} - 1}\;, \qquad (2.1)$$

$$p(x) = Ne^{k\beta x}(1 - \alpha u e^{\beta x})^{\frac{1}{u} - 1}\;, \qquad (2.2)$$

где α, β, k, u, l – параметры распределений;

N – нормирующий множитель, который находится из условия нормировки:

$$\int_{-\infty}^{\infty} p(x)\,dx = 1.$$

Параметр l в первой плотности и α – во второй являются **параметрами сдвига**; k, u – **параметры формы** (для обеих плотностей).

В формуле (2.1) в общем случае параметр $\beta = 1$, а в случае симметричных распределений $\beta = 2$, $k = 1/2$.

При изменении параметра сдвига кривая распределения перемещается вдоль горизонтальной оси, не меняя своей формы.

Система распределений, заданная обобщенной плотностью (2.1), относится к первой дополнительной системе непрерывных распределений автора и включает почти все семейство кривых К.Пирсона.

2.2 В зависимости от значений параметров α, u, а также знака параметра β все распределения, заданные плотностями (2.1) и (2.2), разделяются на типы (см. рис. 2.1).

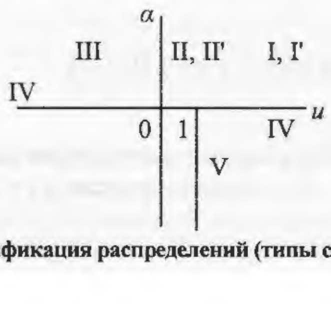

Рис. 2.1 – **Классификация распределений (типы со штрихом – при $\beta < 0$).**

2.3 На параметры распределений разных типов наложены ограничения (см. табл. 2.1).

Таблица 2.1 – **Значения параметров распределений разных типов**

Тип кривой	Параметры кривой		
	u	α	k
I, I′	$0 < u < \infty$	$\alpha > 0$	$0 < k < \infty$
II, II′	$u \to 0$		
III	$-\infty < u < 0$		$0 < k < 1 - \dfrac{1}{u}$
IV	$u \to \pm\infty$	$\alpha u < 0$	
V	$1 < u < \infty$	$\alpha < 0$	

2.4 Ниже записаны все типы распределений, заданных обобщенной плотностью (2.1).

Тип I: $$p(x) = \frac{\beta(\alpha u)^k \Gamma\left(k + \dfrac{1}{u}\right)}{\Gamma(k)\Gamma\left(\dfrac{1}{u}\right)} (x-l)^{k\beta-1}\left[1 - \alpha u (x-l)^{\beta}\right]^{\frac{1}{u}-1},$$

$$l < x < (1/\alpha u)^{1/\beta} + l .$$

Распределения этого типа при $\beta = 1$, $ku = 1$ являются симметричными.

Тип I′: $$p(x) = \frac{\beta(\alpha u)^k \Gamma\left(k + \dfrac{1}{u}\right)}{\Gamma(k)\Gamma\left(\dfrac{1}{u}\right)} \frac{1}{(x-l)^{k\beta+1}}\left[1 - \frac{\alpha u}{(x-l)^{\beta}}\right]^{\frac{1}{u}-1},$$

$$(\alpha u)^{1/\beta} + l < x < \infty .$$

Тип II: $$p(x) = \frac{\beta\alpha^k}{\Gamma(k)}(x-l)^{k\beta-1} e^{-\alpha(x-l)^{\beta}}, \quad l < x < \infty .$$

Тип II′: $$p(x) = \frac{\beta\alpha^k}{\Gamma(k)} \frac{1}{(x-l)^{k\beta+1} e^{\alpha/(x-l)^{\beta}}}, \quad l < x < \infty .$$

Типы III – V: $$p(x) = \frac{\beta(-\alpha u)^k \Gamma\left(1 - \dfrac{1}{u}\right)}{\Gamma(k)\Gamma\left(1 - \dfrac{1}{u} - k\right)} \frac{(x-l)^{k\beta-1}}{\left[1 - \alpha u (x-l)^{\beta}\right]^{1-\frac{1}{u}}},$$

$$l < x < \infty .$$

Здесь $\Gamma(k) = \int_0^\infty x^{k-1} e^{-x} dx$ – гамма-функция. При целых k

$$\Gamma(k) = (k-1)(k-2)\dots 1 = (k-1)\,!$$

Для обобщенной плотности (2.2) нормирующий множитель такой же, как и для плотности (2.1.).

2.5 Обобщенная плотность (2.1) при $\beta = 2$, $k = 1/2$ дает группу симметричных распределений Ic – Vc типов с центром распределения $\bar{x} = l$.

Из этой группы симметричных распределений наиболее часто используется распределение типа IIc (нормальный закон).

Тип IIc: $$p(x) = \sqrt{\frac{\alpha}{\pi}} e^{-\alpha(x-l)^2} , \quad -\infty < x < \infty .$$

2.6 Оценивание параметров распределений, заданных обобщенными плотностями (2.1) и (2.2), осуществляется по методу центральных моментов. При этом абсолютное значение параметра β плотности (2.1) принимается равным единице: $|\beta| = 1$.

2.7 Для вычисления типа выравнивающей кривой используются критерии В. Нешитого *L*, *u*, а также показатели асимметрии β_1 и островершинности β_2, введенные К. Пирсоном.

Критерий *L* (в Программе обозначен через *L*1) рассчитывается по формуле [7, с. 85]:

$$L = \frac{4\beta_2 - 3\beta_1}{4 + \beta_1},$$

где $\beta_1 = \dfrac{\mu_3^2}{\mu_2^3}$, $\quad \beta_2 = \dfrac{\mu_4}{\mu_2^2}$.

Критерий u (он же – параметр формы обобщенной плотности (2.1)) вычисляется по формуле [7, с. 82]:

$$u = -\frac{A(C - a^2 A)}{6CE} = -\frac{AD}{6CE},$$

где величина $a \approx -\bar{x}$ (при $l = 0$) и является одним из корней квадратного уравнения

$$Aa^2 + Ba + C = 0 \ .$$

Величины $A, \ldots, E$ вычисляются по формулам:

$$A = 2\beta_2 - 3\beta_1 - 6$$

$$B = \frac{\mu_3}{\mu_2}(3 + \beta_2)$$

$$C = \mu_2(4\beta_2 - 3\beta_1)$$

$$D = C - a^2 A$$

$$E = \beta_2 - \beta_1 - 1 .$$

2.8 На рис. 2.2 дана классификация распределений, заданных обобщенной плотностью (2.1) при $|\beta| = 1$, по критериям u, L.

Здесь нормальный закон распределения (IIc тип), а также все распределения II типа представлены точкой ($u \to 0$; $L = 3$).

Распределения II′ типа лежат на оси ординат $L > 3$.

Симметричные распределения Ic – IIIc типов лежат на кривой

$$L = 3\frac{2 + u}{2 + 3u} .$$

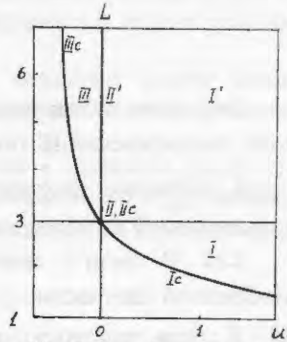

Рис. 2.2 – **Классификация распределений по критериям *u*, *L*.**

Распределения I, I′ и III типов занимают определенные области.

2.9 На рис. 2.3 дана классификация распределений, заданных обобщенной плотностью (2.1) при $|\beta| = 1$, по критериям β_1, β_2 .

На рис. 2.3 нормальный закон распределения (тип IIc) представлен точкой ($\beta_1 = 0$; $\beta_2 = 3$).

Распределения I типа лежат ниже прямой $\beta_2 = 3 + 1{,}5\ \beta_1$, т. е. распределений II типа (при $\mu_3 > 0$).

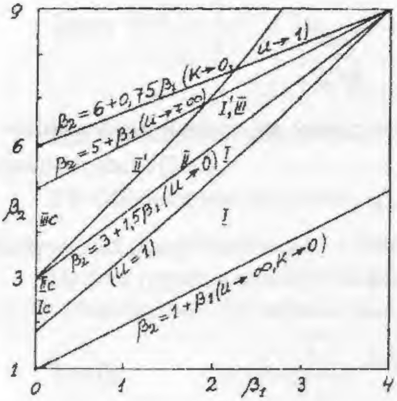

Рис. 2.3 – **Классификация распределений по критериям β_1, β_2.**

Распределения I' и III типа занимают одну область, ограниченную распределениями II и II' типов.

Симметричные распределения Ic типа при $ku = 1$ находятся на интервале оси ординат $1 \le \beta_2 < 3$. Для равномерного распределения ($u = 1$, $k = 1$) показатель $\beta_2 = 1{,}8$.

Выше кривых II' типа находится область распределений, для которых дискриминант $B^2 - 4AC < 0$. Эту область покрывают распределеиия III – V типов, заданные обобщенной плотностью (2.2).

2.10 Распределения III типа, заданные обобщенной плотностью (2.1) при $|\beta| = 1$, имеют центральные моменты до четвертого порядка включительно при условии

$$u > -\frac{1}{k+3}\ .$$

Это значит, что по методу моментов может быть найдена лишь незначительная часть распределений III типа при $|\beta| = 1$. Остальные распределения III, IV и V типов, заданные плотностью (2.1) при $|\beta| = 1$, остаются за пределами применимости метода моментов.

2.11. В связи с замечаниями 2.10 систему распределений, заданную обобщенной плотностью (2.1) при $|\beta| = 1$, следует дополнить распределениями III – V типов, заданными обобщенной плотностью (2.2), т. е.

$$p(x) = \frac{\beta(-\alpha\, u)^k\, \Gamma\left(1 - \frac{1}{u}\right)}{\Gamma(k)\Gamma\left(1 - \frac{1}{u} - k\right)}\ \frac{e^{k\beta x}}{\left(1 - \alpha\, u e^{\beta x}\right)^{1 - \frac{1}{u}}}\ , \qquad -\infty < x < \infty\,.$$

Эти распределения занимают на рис. 2.3 область выше кривой II' типа и ниже прямой $\beta_2 = 6 + 0{,}75\beta_1$.

2.12 Распределения III – V типов, заданные обобщенной плотностью (2.2), симметричны при условии

$$k = \frac{1}{2}\left(1 - \frac{1}{u}\right).$$

Они занимают отрезок оси ординат $3 < \beta_2 < 6$.

2.13 Диаграмма, представленная на рис. 2.3, справедлива для распределений с левосторонней асимметрией, у которых центральный момент 3-го порядка $\mu_3 < 0$.

Однако она успешно используется для установления типа выравнивающей кривой и при $\mu_3 > 0$, т. е. в случае распределений с правосторонней асимметрией.

2.14 Для вычисления различных показателей, характеризующих состояние технологического процесса, необходимо знать закон распределения контролируемого признака и оценки его параметров.

По Программе вначале вычисляются различные показатели статистического распределения, на основе которых вычисляется тип выравнивающей кривой, оценки параметров и нормирующий множитель.

2.15 Значения плотности распределения вероятностей вычисляются непосредственно по соответствующим формулам.

2.16 Для вычисления функции распределения $F(x) = \int_0^x p(x)dx$ плотность $p(x)$ разлагается в ряд и осуществляется почленное интегрирование.

2.17 Вычисление квантилей, доверительных интервалов и других величин осуществляется по методу Ньютона.

2.18 Система непрерывных распределений позволяет найти закон распределения среднего (по их центральным моментам) при известном законе распределения самой случайной величины X.

Центральные моменты среднего арифметического n одинаково распределенных взаимно независимых случайных величин (обозначим их $m_{r(n)}$) связаны с моментами отдельной случайной величины – μ_r следующими формулами:

$$m_{2(n)} = \frac{\mu_2}{n}$$

$$m_{3(n)} = \frac{\mu_3}{n^2}$$

$$m_{4(n)} = \frac{1}{n^3}(\mu_4 + 3(n-1)\mu_2^2).$$

При этом для средних арифметических справедливо равенство:

$$\nu_{1(n)} = \nu_1 .$$

2.19 Тип выравнивающей кривой распределения среднего $\bar{x}$ находится по соответствующим показателям асимметрии и островершинности:

$$\beta_{1(n)} = \frac{\beta_1}{n}; \qquad \beta_{2(n)} = 3 + \frac{\beta_2 - 3}{n}.$$

При $n \to \infty$ имеем: $\beta_{1(n)} = 0$; $\beta_{2(n)} = 3$, т. е. распределение среднего n независимых одинаково распределенных случайных величин с ростом n приближается к нормальному закону (**центральная предельная теорема**).

При этом точка с координатами ($\beta_{1(n)}$, $\beta_{2(n)}$) на рис. 2.3 с ростом n перемещается по прямой

$$\beta_{2(n)} = 3 + \frac{\beta_2 - 3}{\beta_1}\beta_{1(n)}$$

от точки (β_1, β_2) исходного распределения (случайной величины X_i) к точке (0, 3) нормального закона.

Степень приближения к нормальному закону при больших n зависит от исходных значений показателей β_1 и β_2 случайной величины X_i.

2.20 Минимально необходимое значение n (объем выборки), при котором распределение среднего арифметического можно считать нормальным, вычисляется по формулам автора:

$$n_1 = \frac{\beta_1}{\delta_1}; \qquad n_2 = \frac{\left|\beta_2 - 3 - 1{,}75\beta_1\right|}{\delta_2}.$$

Из двух найденных значений n принимается большее.

Значения δ_1 и δ_2 могут быть заранее заданы исходя из необходимой степени близости искомого распределения к нормальному.

Приложение 3
(справочное)

3 Программа SNR11M97

3.1 Настоящая программа является одной из серии программ, разработанных автором системы непрерывных распределений с 1988 по 1997 гг. Предыдущий ее вариант под названием SNR0MM97 зарегистрирован в Комитете по авторским и смежным правам при Министерстве юстиции РБ от 5 августа 1997 г. за № 287 и охраняется законом.

Программа предназначена для аппроксимации (выравнивания) широкого класса статистических распределений, заданных на всей числовой оси.

Программа устанавливает вид и тип выравнивающей кривой, выдает выравнивающее распределение с указанием интервала, на котором задана случайная величина, вычисляет по методу моментов точечные оценки параметров, выдает таблицу значений плотности вероятностей и функции распределения, вычисляет квантили, процентили, доверительные вероятности и доверительные интервалы, координаты моды и точек перегиба, строит кривую распределения (выравнивающую – пунктирной линией, а эмпирическую – отдельными точками) и, наконец, вычисляет показатели уровня качества продукции (см. Приложение 4).

Для этого достаточно ввести построчно через запятую значения случайной величины в серединах интервалов, частоту и при необходимости – ширину интервалов (при сгруппированных данных) или по одному значению случайной величины, если данные не сгруппированы.

После ввода всех строк программа выдает таблицу «Показатели статистического распределения», а также тип выравнивающей кривой и уравнение плотности распределения с указанием оценок параметров и границ интервала, на котором задано распределение.

После нажатия клавиши F5 на экране появляется меню:

ВВЕДИТЕ ВИД РАСЧЕТА
1 – плотности вероятностей и функции распределения
2 – квантилей
3 – процентилей
4 – доверительных интервалов
5 – доверительной вероятности
6 – координат моды C и точек перегиба A, B
7 – построение кривой распределения
8 – показателей уровня качества
9 – на начало программы
10 – окончание работы.

Для выполнения нужного расчета достаточно ввести соответствующее своему виду расчета число и при необходимости ответить на запрос (см. распечатку в Приложении 4).

Другие программы построены по такому же принципу.

При этом в одних программах для вычисления оценок параметров выравнивающих распределений используется метод моментов (например, SNR11M97), в других – общий устойчивый метод автора (например, SNR1V97).

Тиражирование любой программы автора, переработка, перевод на другой язык программирования и т. д. без заключения авторского договора запрещается. Автор готов к заключению таких договоров с начала 2001 года.

3.2 В Программе реализованы обобщенные распределения, заданные плотностями

$$p(x) = N(x-l)^{k\beta-1}\left[1-\alpha u(x-l)^{\beta}\right]^{\frac{1}{u}-1},$$

$$p(x) = Ne^{k\beta x}\left(1-\alpha u e^{\beta x}\right)^{\frac{1}{u}-1}.$$

Для первой плотности принято $|\beta| = 1$ и $|\beta| = 2$ (в случае симметричного распределения IIс типа, т. е. нормального закона).

3.3 Вычисление типа выравнивающей кривой и оценок параметров α, β, k, u, l осуществляется по методу центральных моментов (классическому или универсальному). В Программе принято обозначение $\gamma = k\beta$.

3.4 Метод центральных моментов позволяет обрабатывать как сгруппированные, так и несгруппированные статистические данные.

3.5 Минимальный объем выборки должен быть не менее 25 значений контролируемого параметра при статистическом регулировании технологического процесса и не менее 100 значений при статистическом анализе точности технологического процесса.

3.6 Максимальный объем выборки при несгруппированных данных составляет 256 замеров контролируемого параметра. Для сгруппированных данных объем выборки не ограничивается.

3.7 Программа обеспечивает вычисление различных показателей с точностью 5-6 значащих цифр, что вполне достаточно для практических расчетов.

3.8 Программа работает под управлением MS DOS. Язык программирования – GW-BASIC. Объем около 25 кбт исходного текста.

3.9 За справками обращаться к автору (тел. 8-0172) 51-06-35.

Приложение 4

Пример статистической обработки результатов замера контролируемого параметра по программе SNR11M97

Были измерены отклонения от номинального размера некоторой детали Ø 50 мм. Конструкторский допуск равен ± 0,012 мм. т. е. $T_н$ = – 0,012, $T_в$ = + 0,012. Ширина поля допуска $TX = T_в - T_н$ = 0,024 мм.

Результаты замеров (в микронах) представлены в контрольном листке (см. табл. 4.1). Точность измерительного средства равна 0,002 мм и составляет 1/12 ширины конструкторского допуска. Объем выборки n = 100.

Найдем выравнивающую кривую распределения по несгруппированным данным.

Запустив программу, введем по одному значению табличные данные, нажимая каждый раз клавишу “ENTER” (“Ввод”). При этом вместо значения Ø будем вводить 0,000001. Для завершения введем 0.

Получим множество показателей статистического распределения (см. распечатку, табл. 4.2). В этой таблице критерий L обозначен через $L1$.

Статистические данные контрольного листка можно представить в более компактном виде, если объединить одинаковые значения отклонений от номинального размера (табл. 4.3).

Такая форма представления данных замечательна тем, что из подобной таблицы видно, насколько тщательно были выполнены замеры контролируемого параметра. Например, в приведенной таблице резко выделяются частоты отклонений, кратных точности средства измерения (2 микрона), что свидетельствует о неправильном округлении результатов измерения.

В этом случае кривая распределения находится по сгруппированным данным, причем ширина интервала принимается равной единице. Вводить следует построчно пары чисел x_i, n_i через запятую, а для завершения вводится 0, 0.

Результаты расчетов остаются прежними за исключением двух строк: количество интервалов (20) и ширина интервала (1).

Поскольку критерий $L1 < 3$, то выравнивающее распределение относится к типу 1.1 и задается плотностью

$$p(x) = N(x-l)^{k-1}\left[1-\alpha u(x-l)\right]^{\frac{1}{u}-1} .$$

Оценки параметров и нормирующего множителя приведены в распечатке (см. табл. 4.2).

Случайная величина X задана на интервале

$$-20{,}7128 < X < 16{,}37711.$$

Теперь можно рассчитать любые интересующие нас показатели.

Таблица 4.1

Контрольный листок

Деталь N_______________________ Ø 50 мм ±0,012

(название)

Участок ________

Точность СИ 0,002 Дата __________ Время __________

№ п/п, i	Отклонение (мкм), x_i	x_i	x_i	x_i
1	3	8	0	8
2	8	6	4	0
3	2	- 4	10	2
4	4	6	6	6
5	6	- 3	4	8
6	6	8	4	2
7	- 2	10	6	6
8	8	10	8	7
9	-1	8	12	8
10	8	8	6	- 6
11	4	4	10	- 4
12	4	10	- 4	10
13	6	2	4	2
14	0	4	10	6
15	10	6	12	12
16	0	2	11	10
17	-2	1	0	2
18	0	8	2	8
19	2	1	- 2	6
20	2	- 10	- 4	- 3
21	14	0	2	- 2
22	4	4	- 8	0
23	6	- 2	1	1
24	0	5	7	2
25	10	0	0	6

Контролер ____________

Технолог цеха ____________

Таблица 4.2

ПОКАЗАТЕЛИ СТАТИСТИЧЕСКОГО РАСПРЕДЕЛЕНИЯ (SNR11M97)

К-во наблюдений	100
К-во интервалов	20
Ширина интервала	1
Среднее	3.92
Дисперсия	22.4736
Центр. момент 3-го порядка	- 37.34583
Центр. момент 4-го порядка	1408.746
Ср. квадратич. отклонение	4.740633
Коэф-т вариации %	120.9345
Показатель асимметрии	.1228758
Показатель островершинности	2.78925

L 1 = 2.616711

РАСПРЕДЕЛЕНИЕ ТИПА 1.1 С ПАРАМЕТРАМИ

AU = 2.696148E - 02 K = 8.403986 U = .2352945 L = - 20.71284
N = 1.367198E - 10

Случайная величина X задана на интервале - 20.71284 < X < 16.37711

P(X)=N* (X-L)^(K-1)* (1-AU* (X-L))^(1/U-1)

ПОКАЗАТЕЛИ УРОВНЯ КАЧЕСТВА (SNR11M97)

Кол-во наблюдений	100
Среднее из опыта W1	3.92
Среднее квадратич. отклонение (с. к. о.) Sx	4.740633
Центр поля допуска T0	0
Нижняя граница поля допуска TH	- 12
Верхняя граница поля допуска TB	12
Ширина поля допуска TX = TB - TH	24
Отношение допуска к с. к. о. (TX/SX)	5.062615
Смещение центров рассеян. и допуска E = W1 – T0	3.92
Ширина поля рассеяния RX при P =. 9973	25.42721
Нижняя граница поля рассеяния	- 11.01229
Верхняя граница поля рассеяния	14.41492
Отношение поля рассеяния к с. к. о. (RX/SX)	5.363674
Показатель уровня настройки KN = E/TX	.163333
Показатель точности KT = RX/TX	1.059467
Прогнозируемый процент брака Q %	2.76925
в т. ч. на нижней границе поля допуска	.060416
на верхней границе поля допуска	2.708835

Таблица 4.3.

Отклонения от номинального размера детали “NN” Ø 50 ± 0,012.

Значения отклонений	Частота n_i	Значения отклонений	Частота n_i
- 10	1	3	1
- 8	1	4	11
- 6	1	5	1
- 4	4	6	15
- 3	2	7	2
- 2	5	8	13
- 1	1	10	10
0	11	11	1
1	4	12	3
2	12	14	1

$$\Sigma n_i = 100$$

Например, выбрав в меню “ВИД РАСЧЕТА” пункт 4 (вычисление доверительных интервалов) и введя вероятность $P = 0,9973$, найдем границы технологического допуска

$$-11,01229 < X < 14,41492.$$

Для вычисления ожидаемого процента брака необходимо выбрать в меню “ВИД РАСЧЕТА” пункт 8 (вычисление показателей уровня качества). Введя затем середину поля допуска $T_0 = 0$ и отклонение $\Delta T = 12$, получим ряд показателей качества (см. распечатку), в том числе:

Смещение центров рассеяния и допуска $E = 3,92$
Показатель уровня настройки $K_н = 0,163333$
Показатель точности $K_T = 1,059467$
Прогнозируемый процент брака $Q\,\% = 2,76925$
в том числе:
на нижней границе поля допуска $Q\,\% = 0,060416$
на верхней границе поля допуска $Q\,\% = 2,708835.$

Поскольку в данном случае показатель точности (коэффициент рассеяния) $K_T > 1$, то для полного исключения брака требуется уменьшить рассеяние.

Однако только за счет смещения центров рассеяния и допуска можно добиться значительного уменьшения брака.

Например, при вводе $T_0 = 3,92$, $\Delta T = 12$ получим процент брака 0,911558.

При этом весь брак будет на нижней границе поля допуска.

Следовательно, для уменьшения брака следует осуществить настройку оборудования, т. е. **совместить центры полей рассеяния и допуска.**

Ввиду того, что в данном случае выравнивающая кривая распределения имеет небольшую левостороннюю асимметрию ($\mu_3 < 0$), то **необязательно точно совмещать центры полей рассеяния и допуска.** Например, при вводе $T_0 = 2$, $\Delta T = 12$ получим еще меньший процент брака, равный 0,5648. При этом он примерно одинаков на обеих границах конструкторского допуска.

Для уменьшения рассеяния необходимо принимать другие меры. Эти вопросы решает отдел главного технолога.

Сделаем некоторые **выводы по рассмотренному примеру**.

Во-первых, **выравнивающее распределение оказалось далеко не нормальным**, которое характеризуется показателями: $\beta_1 = 0$; $\beta_2 = 3$; $L1 = 3$.

Как показывает практика, **нормальный закон распределения реализуется довольно редко.**

Если бы в данном примере в качестве выравнивающего распределения **был принят нормальный закон, то ожидаемый процент брака был бы равен 4,456 %, т. е. в 1,6 раза больше фактического**. И это при небольшой асимметрии выравнивающей кривой! В других случаях нормальный закон может давать значительно большую ошибку (до 10 раз по сравнению с фактическим браком) (см. рис. 4.1).

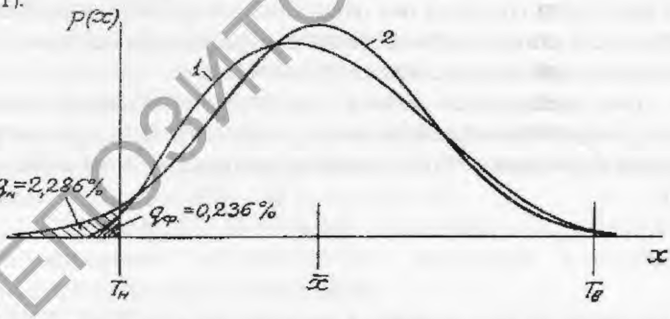

Рис. 4.1 **Ожидаемый процент брака для двух законов распределения: 1 – обобщенного; 2 – нормального. Нормальный закон показал брак в 9,7 раза больше фактического.**

Отсюда следует вывод, что **самым важным вопросом, который необходимо всякий раз решать при статистическом анализе точности технологического процесса – это вопрос установления наилучшей выравнивающей кривой. Он может быть решен только при использовании обобщенных распределений.**

Во-вторых, из рассмотренного примера видно, что **разладка технологического процесса определяется двумя факторами: смещением среднего значения контролируемого параметра и рассеиванием его значений относительно среднего.**

В-третьих, **регулирование настройки технологического процесса дает возможность уменьшить брак в 3 – 6 раз, хотя не исключает его полностью, если коэффициент точности $K_T > 1$.**

В-четвертых, **при наладке технологического процесса одним из основных критериев следует принимать** не только величину смещения центров рассеяния и допуска, но и **минимальный ожидаемый процент брака**, особенно в случае асимметричного выравнивающего распределения. С помощью Программы нетрудно подобрать оптимальное смещение центров рассеяния и допуска, при котором ожидаемый процент брака будет минимальным при заданном коэффициенте точности.

В-пятых, **Программа может быть использована при статистическом регулировании технологических процессов**. В этом случае через определенные интервалы времени достаточно отбирать выборку из 25 – 30 изделий и по результатам замеров контролируемого параметра вычислять необходимые характеристики: среднее, среднее квадратическое отклонение, размах и т. д. и наносить их на контрольную карту.

Таким образом, обобщенные распределения, доведенные до программной реализации (в виде серии программ под общим названием SNR), являются мощным инструментом для статистического анализа и регулирования технологических процессов, а также для решения многих других задач.

При этом достигается весьма высокая вероятность вычисления наилучшей выравнивающей кривой, что подтверждено апробацией программ на большом статистическом материале в течение ряда лет (с 1990 г.).

ЛИТЕРАТУРА

1. Богатырев А.А., Филиппов Ю.Д. Стандартизация статистических методов управления качеством. – М.: Издательство стандартов, 1989. – 120 с.
2. Венецкий И.Г., Венецкая В.И. Основные математико-статистические формулы в экономическом анализе: Справочник; 2-е изд., перераб. и доп. – М.: Статистика, 1979. – 447 с.
3. Герасимович А.И., Матвеева Я.И. Математическая статистика. Минск: Вышэйшая школа, 1978. – 200 с.
4. Гончаров Э.Н., Козлов В.В., Круглова Е.Д. Контроль качества продукции. – М.: Изд-во стандартов,1987. – 120 с.
5. Крамер Г. Математические методы статистики. – М.: Мир, 1975.
6. Методика организации внедрения статистических методов контроля качества продукции на промышленном предприятии. – М.: Изд-во стандартов, 1977. – 40 с.
7. Нешитой В.В. Построение системы непрерывных распределений/ БелНИИНТИ. – Минск, 1979. – 193 с. – деп. в БелНИИНТИ 30.07.80, № 174.
8. Нешитой В.В. Оценивание параметров обобщенных распределений/БелНИИНТИ. – Минск, 1983. – 96 с. – деп. в БелНИИНТИ 11.03.84, № 857.
9. Нешитой В.В. Об аппроксимации статистических распределений обобщенными плотностями// Надежность и контроль качества. – М.: ВНИИНМАШ, 1988. – № 2. – с. 7 – 13.
10. Нешитой В.В. Статистический анализ технологических процессов на базе обобщенных распределений// Техника, экономика, организация. – Минск, 1998. – № 2. – с. 36 – 39.
11. СТ СЭВ 3946-82. Внедрение статистических методов анализа, регулирования технологических процессов и статистических методов приемочного контроля.
12. ГОСТ 27.202-83. Надежность в технике. Технологические системы. Методы оценки надежности по параметрам качества изготовляемой продукции.