ГОСУДАРСТВЕННЫЙ СТАНДАРТ РОССИЙСКОЙ ФЕДЕРАЦИИ

Статистические методы

ПРОВЕРКА ОТКЛОНЕНИЯ РАСПРЕДЕЛЕНИЯ ВЕРОЯТНОСТЕЙ ОТ НОРМАЛЬНОГО РАСПРЕДЕЛЕНИЯ

Издание официальное

БЗ 11—2001/273

ГОССТАНДАРТ РОССИИ
Москва

gost.e-ecolog.ru

Предисловие

1 РАЗРАБОТАН И ВНЕСЕН Техническим комитетом по стандартизации ТК 125 «Статистические методы в управлении качеством продукции»;

Акционерным обществом «Научно-исследовательский центр контроля и диагностики технических систем» (АО НИЦ КД)

2 ПРИНЯТ И ВВЕДЕН В ДЕЙСТВИЕ Постановлением Госстандарта России от 22 января 2002 г. № 25-ст

3 Настоящий стандарт представляет собой аутентичный текст международного стандарта ИСО 5479 —97 «Статистическое представление данных. Проверка отклонения распределения вероятностей от нормального распределения»

4 ВВЕДЕН ВПЕРВЫЕ

© ИПК Издательство стандартов, 2002

Настоящий стандарт не может быть полностью или частично воспроизведен, тиражирован и распространен в качестве официального издания без разрешения Госстандарта России

Содержание

1 Область применения 1
2 Нормативные ссылки 1
3 Определения и обозначения 1
4 Общие положения 2
5 Графический метод 3
6 Направленные критерии 8
7 Совместный критерий, использующий статистики $\sqrt{b_1}$ и b_2 (многонаправленный критерий) 11
8 Многосторонние критерии 11
9 Совместный критерий, использующий несколько независимых выборок 16
10 Статистические таблицы 17
Приложение А Бланк бумаги для нормальных вероятностных графиков 25

Введение

Настоящий стандарт устанавливает критерии, с помощью которых можно проверить, подчиняется ли генеральная совокупность данных нормальному закону распределения. Это следующие виды критериев: графический метод, направленный критерий, многонаправленный критерий, многосторонний критерий, совместный критерий для нескольких независимых выборок.

Целью настоящего стандарта является — привести критерии, удобные для использования специалистами в промышленности для проверки на нормальность различных данных в ходе проведения измерений, контроля и испытаний.

В настоящем стандарте рассматриваются способы построения статистик и правила принятия решений для критериев проверки на нормальность.

ГОСУДАРСТВЕННЫЙ СТАНДАРТ РОССИЙСКОЙ ФЕДЕРАЦИИ

Статистические методы

ПРОВЕРКА ОТКЛОНЕНИЯ РАСПРЕДЕЛЕНИЯ ВЕРОЯТНОСТЕЙ ОТ НОРМАЛЬНОГО РАСПРЕДЕЛЕНИЯ

Statistical methods. Tests for departure of the probability distribution from the normal distribution

Дата введения 2002—07—01

1 Область применения

1.1 Настоящий стандарт устанавливает методы и критерии для проверки отклонения распределения вероятностей от нормального распределения при независимых наблюдениях.

1.2 Применение критерия на отклонение от нормального распределения необходимо во всех случаях, когда есть сомнение, нормально ли распределены наблюдения. В случае робастных методов (то есть таких, что результаты наблюдений незначительно изменяются, когда реальное распределение вероятностей наблюдений ненормально) критерий на отклонение от нормального распределения не применяют. Например случай, когда выборочное среднее арифметическое единственной случайной выборки, полученной в результате наблюдений, необходимо сопоставить с данным теоретическим значением с использованием *t*-критерия.

1.3 Необязательно использовать такой критерий при каждом обращении к статистическим методам, основанным на гипотезе нормальности. Существуют случаи, когда в нормальности распределения наблюдений нет сомнения: есть теоретические (например физические) обоснования, подтверждающие гипотезу, или гипотезу считают приемлемой согласно априорной информации.

1.4 Критерии на отклонение от нормального распределения, установленные в настоящем стандарте, в основном рассчитаны на полные, несгруппированные данные.

2 Нормативные ссылки

В настоящем стандарте использована ссылка на следующий стандарт:

ГОСТ Р 50779.10-2000 (ИСО 3534.1-93) Статистические методы. Вероятность и основы статистики. Термины и определения

3 Определения и обозначения

3.1 Определения

В настоящем стандарте применяются термины с соответствующими определениями по ГОСТ Р 50779.10.

3.2 Обозначения

В настоящем стандарте использованы следующие обозначения:

a_k — коэффициент критерия Шапиро-Уилка;

A; B — вспомогательные величины для критерия Эппса-Палли;

b_2 — эмпирическая кривизна;

$\sqrt{b_1}$ — эмпирическая асимметрия;

Издание официальное

E — математическое ожидание;
G_j — вспомогательная величина для совместного критерия, использующего несколько независимых выборок;
h — число последовательных выборок;
H_0 — нулевая гипотеза;
H_1 — альтернативная гипотеза;
k — порядковый номер значений x в выборке, упорядоченной в порядке неубывания;
m_j — выборочный центральный момент порядка j;
n — объем выборки;
p — вероятность, связанная с p-квантилью распределения вероятностей;
P — вероятность;
P_k — вероятность, связанная с X_k;
S — вспомогательная величина для критерия Шапиро-Уилка;
T — статистика критерия;
T_{EP} — статистика критерия Эппса-Палли;
u_p — p-квантиль стандартного нормального распределения;
v_j — вспомогательная величина для совместного критерия, использующего несколько независимых выборок;
W — статистика критерия Шапиро-Уилка;
W_j — вспомогательная величина для совместного критерия, использующего несколько независимых выборок;
X — случайная переменная;
x — значение случайной переменной X;
x_j — j-е значение в выборке, упорядоченной в порядке неубывания;
x_k — k-е значение в выборке, упорядоченной в порядке неубывания;
$\bar{x}$ — среднее арифметическое;
α — уровень значимости;
β — вероятность ошибки второго рода;
β_2 — кривизна совокупности;
$\beta_2 - 3$ — эксцесс совокупности;
$\sqrt{\beta_1}$ — асимметрия совокупности;
γ; δ; ε — вспомогательные величины для совместного критерия, использующего несколько независимых выборок;
γ_n; δ_n; ε_n — коэффициенты совместного критерия, использующего несколько независимых выборок;
μ — математическое ожидание (центральный момент первого порядка);
μ_2 — дисперсия совокупности (центральный момент второго порядка);
μ_3 — центральный момент совокупности третьего порядка;
μ_4 — центральный момент совокупности четвертого порядка;
σ — стандартное отклонение совокупности ($\sigma = \sqrt{\mu_2}$).

4 Общие положения

4.1 Существуют различные критерии на отклонение от нормальности. В настоящем стандарте установлены графические методы, моментные критерии, регрессионные критерии и критерии характеристических функций. Критерии хи-квадрат подходят только для сгруппированных данных, и так как группирование приводит к потере информации, в данном стандарте они не рассмотрены.

4.2 Если о выборке нет дополнительной информации, рекомендуется сначала построить нормальный вероятностный график, то есть построить кумулятивную функцию распределения значений, полученных в результате наблюдений, на бумаге для нормальных вероятностных графиков с осями координат, в которых кумулятивная функция нормального распределения представлена прямой линией.

Этот метод, установленный в разделе 5, позволяет сразу видеть, близко ли полученное распределение к нормальному. Используя данную дополнительную информацию, необходимо решить, какой критерий можно применить: направленный, регрессионный, критерий характеристической функции или никакой. Такое графическое представление нельзя рассматривать как строгий критерий, но даваемая им суммарная информация является существенным дополнением к любому критерию на отклонение от нормального распределения. В случае отклонения нулевой гипотезы эта информация дает возможность определить тип альтернативной гипотезы, которая могла бы быть применима.

4.3 Критерий на отклонение от нормального распределения имеет нулевую гипотезу, состоящую в том, что выборка содержит n значений независимых наблюдений, подчиняющихся одному и тому же нормальному распределению. Он заключается в вычислении функции T от этих значений, называемой статистикой критерия. Нулевую гипотезу о нормальности распределения принимают или отклоняют в зависимости от того, лежит ли статистика T в области ожидаемых значений, соответствующих нормальному распределению.

4.4 Критическая область критерия — это совокупность значений T, ведущих к отклонению нулевой гипотезы. Уровень значимости критерия — это вероятность P получения значения T в критической области, когда нулевая гипотеза верна. Этот уровень дает вероятность ошибочного отклонения нулевой гипотезы (вероятность ошибки первого рода).

Граница критической области (или в случае двустороннего критерия — границы критической области) — это критическое значение(ния) статистики критерия.

4.5 Мощность критерия — это вероятность отклонения нулевой гипотезы, когда она неверна. Высокая мощность соответствует низкой вероятности ошибочного применения нулевой гипотезы (вероятности ошибки второго рода).

Мощность критерия (то есть для данной ситуации вероятность, что нулевая гипотеза о нормальности распределения будет отклонена, если она неверна) возрастает с ростом числа наблюдений. Например, отклонение от нормального распределения, которое могло быть очевидным при использовании критерия с большой выборкой, можно не обнаружить при том же значении критерия с меньшей выборкой.

4.6 Существуют два вида критериев на отклонение от нормального распределения: направленный критерий — когда форму этого отклонения устанавливают в альтернативной гипотезе и многосторонний критерий — когда форму отклонения в ней не устанавливают.

В направленном критерии критическую область определяют так, чтобы мощность критерия достигала максимального значения. В многостороннем критерии необходимо отделить критическую область так, чтобы она состояла из значений статистики критерия, лежащих далеко от ожидаемого значения.

Если имеются предположения о форме отклонения от нормального распределения, то есть рассматривается распределение, у которого асимметрия или кривизна отличны от свойственных нормальному распределению, то следует применить направленный критерий, так как его мощность больше, чем у многостороннего.

4.7 Направленный критерий является односторонним. В случае асимметрии он сдвигается к положительной или отрицательной асимметрии. Если совместно рассматривают несколько альтернативных гипотез — это критерий многонаправленный. Такие критерии используют при совместном рассмотрении ненулевых асимметрии и кривизны, отличных от свойственных нормальному распределению.

4.8 Таблицы 8—14 и рисунок 9 позволяют применять критерии для наиболее распространенных уровней значимости α, то есть $\alpha = 0{,}05$ и $\alpha = 0{,}01$. До использования критерия следует установить уровень значимости. Критерий может привести к отклонению нулевой гипотезы при уровне значимости 0,05 и не отклонению той же гипотезы при уровне значимости 0,01.

4.9 При вычислении статистики критерия необходимо использовать не менее шести значащих цифр. Значения подсовокупностей, промежуточных результатов и вспомогательных величин следует округлять не менее чем до шести значащих цифр.

5 Графический метод

5.1 Кумулятивную функцию распределения наблюденных значений строят на бумаге для нормальных вероятностных графиков. Вертикальная ось имеет нелинейную шкалу, соответствующую площади под стандартной функцией нормального распределения и размечена значениями кумулятивной относительной частоты. Другая ось имеет линейную шкалу для упорядоченных значений x. Если кумулятивная функция распределения переменной X приближается к прямой линии, то распределение переменной X будет нормальным.

Иногда эти оси меняют местами. Если выполнено нормирование переменной X, линейную шкалу можно заменить логарифмической, квадратичной, обратной или другой шкалой.

На рисунке 1 приведен пример бумаги для нормальных вероятностных графиков. По вертикальной оси значения кумулятивной относительной частоты даны в процентах, а по горизонтальной — произвольная линейная шкала.

Чистый бланк бумаги для нормальных вероятностных графиков приведен в приложении А.

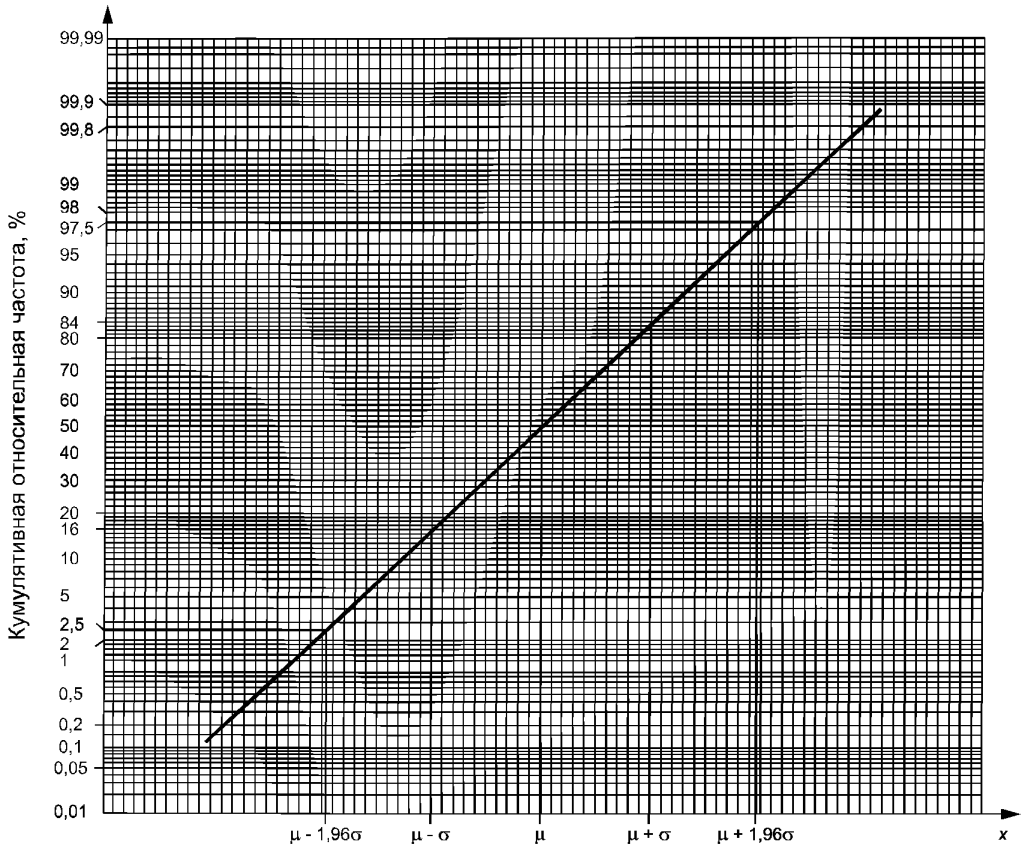


Рисунок 1 — Бумага для нормальных вероятностных графиков

Если график на этой бумаге представлен набором точек, которые рассеянны около прямой линии, то это дает первое подтверждение утверждению, что генеральная совокупность, из которой взята выборка, подчиняется нормальному закону распределения.

Этот подход важен тем, что дает наглядную информацию по типу отклонения от нормального распределения.

Если график показывает, что данные подчинены другому распределению, не имеющему отношения к нормальному (например график кумулятивной функции распределения такой, как на рисунке 5 или 6), то в некоторых случаях к нормальному распределению можно перейти с помощью специального преобразования.

Если график показывает, что данные не подчиняются простому однородному распределению, а скорее всего принадлежат смеси двух или нескольких однородных подсовокупностей (например, если график кумулятивной функции распределения такой, как на рисунке 7), то рекомендуется выявить подсовокупности и анализ каждой из них проводить отдельно.

Этот графический метод не является критерием на отклонение от нормального распределения в строгом смысле. Например, в случае малых выборок с его помощью можно получить выраженные кривые нормальных распределений, но для больших выборок кривые могут представлять ненормальные распределения.

5.2 Графическая процедура состоит в расположении наблюденных значений (x_1, x_2, ..., x_n) в неубывающем порядке и затем в нанесении значений вероятности P_k, рассчитанных по формуле

$$P_k = (k - 3/8)/(n + 1/4) , \quad (1)$$

на бумагу для нормальных вероятностных графиков (где k — порядковый номер x; $k = 1, ..., n$).

Примечание 1 — Часто вместо формулы (1) применяют следующие формулы: $P_k = (k - 1/2)/n$ и $P_k = k/(n + 1)$. Применение данных формул не рекомендуется, так как они дают плохую аппроксимацию нормальной функции распределения ожидаемой порядковой статистики $F[E(X_k)]$.

5.3 Пример использования бумаги для нормальных вероятностных графиков показан на рисунке 2.

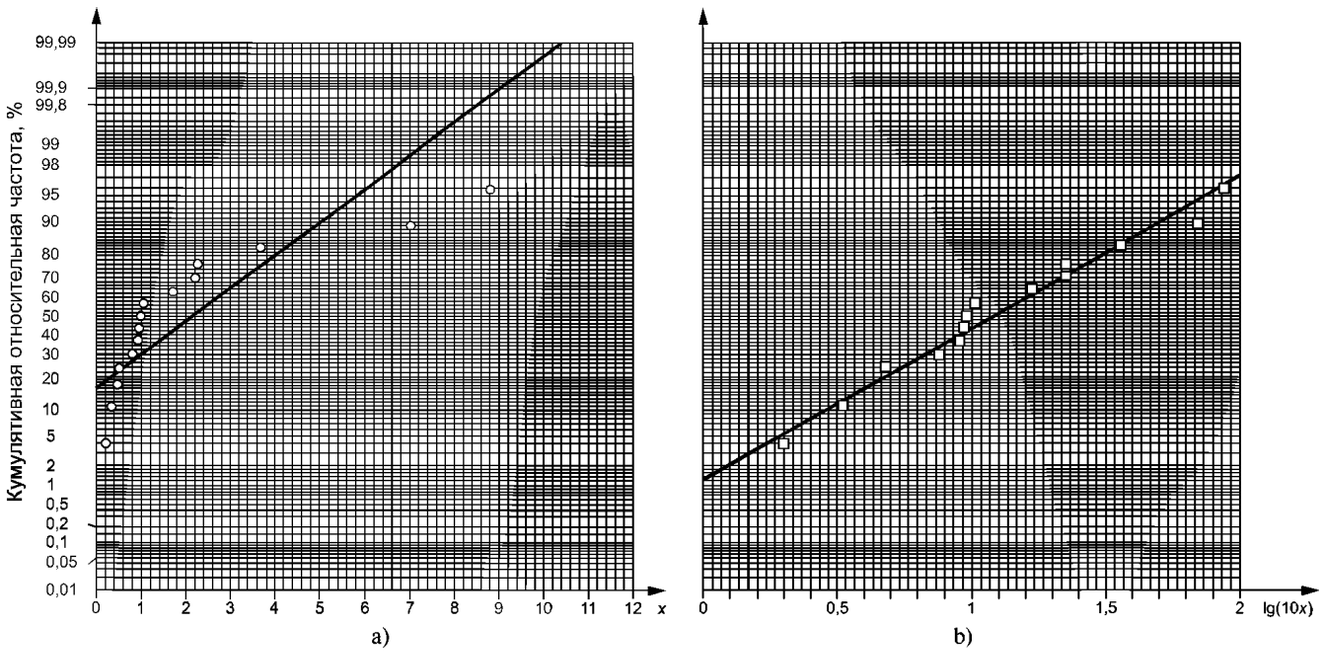


Рисунок 2 — График серии наблюдений на бумаге для нормальных вероятностных графиков

В таблице 1 приведены значения x_k в порядке неубывания как результат серии из 15 независимых наблюдений при испытаниях на усталость вращающегося соединения.

Серию точек, показанную на рисунке 2а) получают на основе значений вероятностей $P_k = (k - 3/8)/(n + 1/4)$ и x_k. Из графика видно, что эти точки не образуют прямой линии. Если x_k заменить на lg (10 x_k), то новый график на рисунке 2b) близок к прямой линии. Из этого следует, что гипотеза нормального распределения для логарифма наблюдений адекватна.

5.4 Рассеяние полученных экстремальных значений больше, чем у средних арифметических значений, поэтому шкала кумулятивной относительной частоты расширяется к краям. Наличие небольшого числа значений на любом

Таблица 1 — Значения x_k серии из 15 независимых наблюдений и соответствующие им значения lg (10 x_k)

k	$P = \frac{k - 3/8}{n + 1/4}$	x_k	lg (10 x_k)
1	0,041	0,200	0,301
2	0,107	0,330	0,519
3	0,172	0,445	0,648
4	0,238	0,490	0,690
5	0,303	0,780	0,892
6	0,369	0,920	0,964
7	0,343	0,950	0,978
8	0,500	0,970	0,987
9	0,566	1,040	1,017
10	0,631	1,710	1,233
11	0,697	2,220	1,346
12	0,762	2,275	1,357
13	0,828	3,650	1,562
14	0,893	7,000	1,845
15	0,959	8,800	1,944

Примечание 2 — В таблице 1 и последующих примерах единицы величин опущены, так как это несущественно для рассматриваемых критериев в данном стандарте.

конце графика функции кумулятивного распределения, которые заметно отклоняются от прямой линии, определяемой средними арифметическими значениями, нельзя рассматривать как показатель отклонения от нормального распределения.

Чем больше объем выборки, тем более надежны заключения, которые можно вывести из вида графика функции распределения.

Если график кумулятивной функции распределения наблюдаемых значений таков, что большие значения находятся ниже прямой линии, образуемой другими значениями, то такое преобразование как $y = \log x$ или $y = \sqrt{x}$ приведет к графику, более соответствующему прямой линии [см. рисунки 2b) и 5].

На рисунках 3—7 верхние графики представляют собой кумулятивные функции распределения; для сравнения нижние графики представляют соответствующие функции плотности распределения.

Если график кумулятивной функции распределения полученных значений подобен графикам на рисунках 3 или 4, то соответствующее частотное распределение имеет меньшую кривизну (график более плоский) или большую кривизну (график более выпуклый) соответственно.

Графики кумулятивных функций распределения на рисунках 5 и 6 относятся к функциям плотности распределения с положительной и отрицательной асимметрией.

На рисунке 7 представлены кумулятивная функция распределения и функция плотности распределения, полученные от наложения двух различных функций плотности.

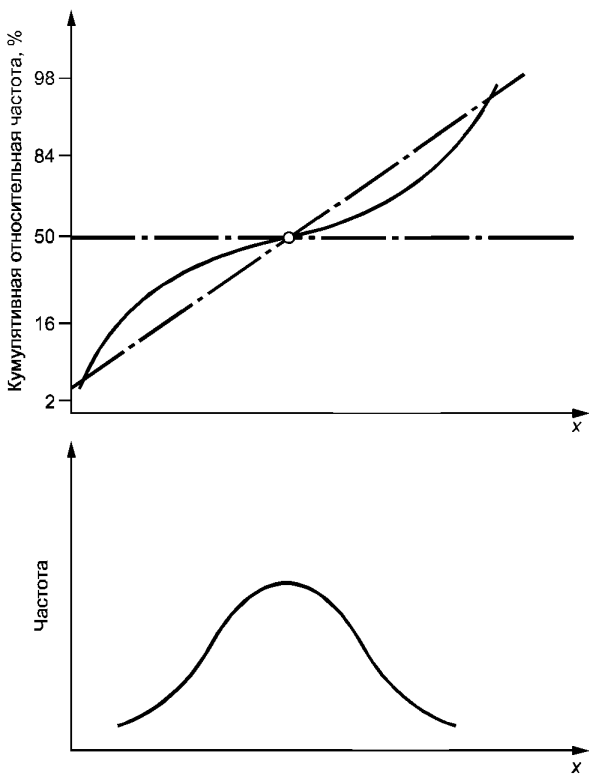


Рисунок 3 — Кумулятивная функция распределения (вверху) и функция плотности распределения с меньшей кривизной (внизу)

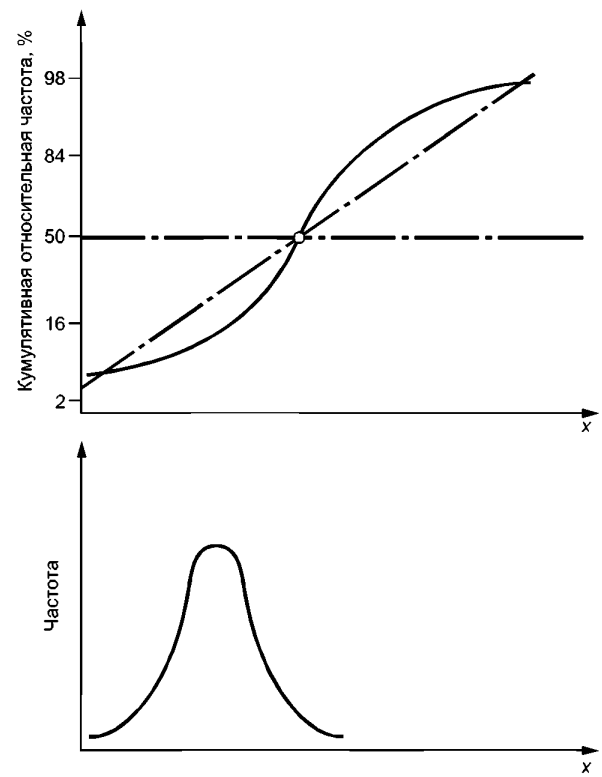


Рисунок 4 — Кумулятивная функция распределения (вверху) и функция плотности распределения с большой кривизной (внизу)

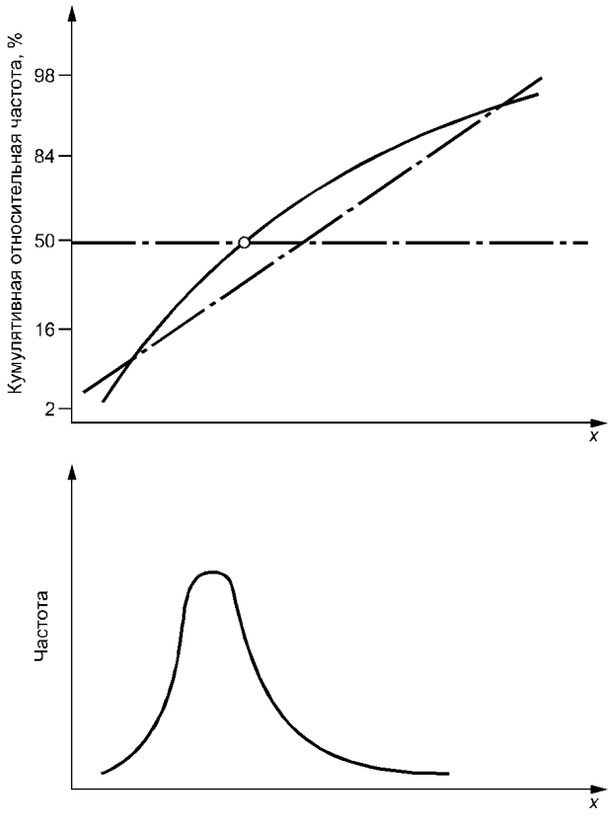


Рисунок 5 — Кумулятивная функция распределения (вверху) и функция плотности распределения с положительной асимметрией (внизу)

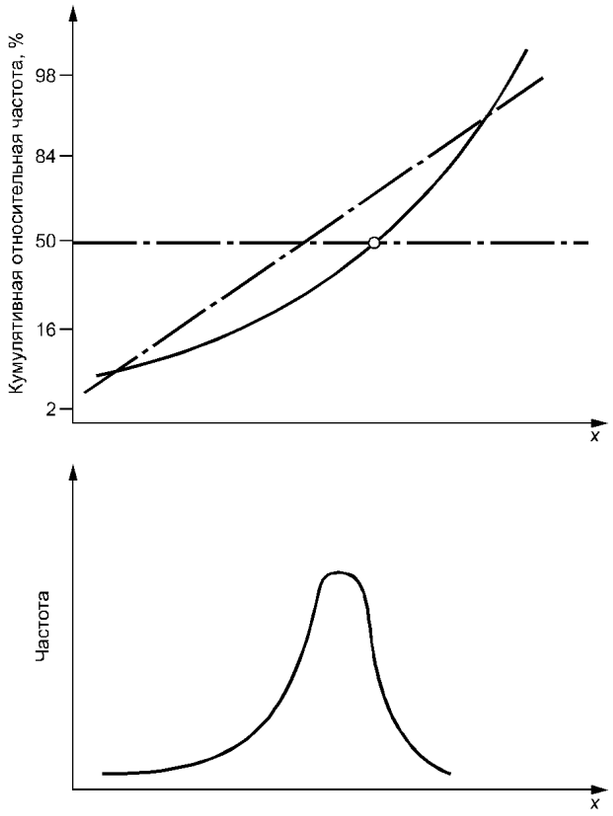


Рисунок 6 — Кумулятивная функция распределения (вверху) и функция плотности распределения с отрицательной асимметрией (внизу)

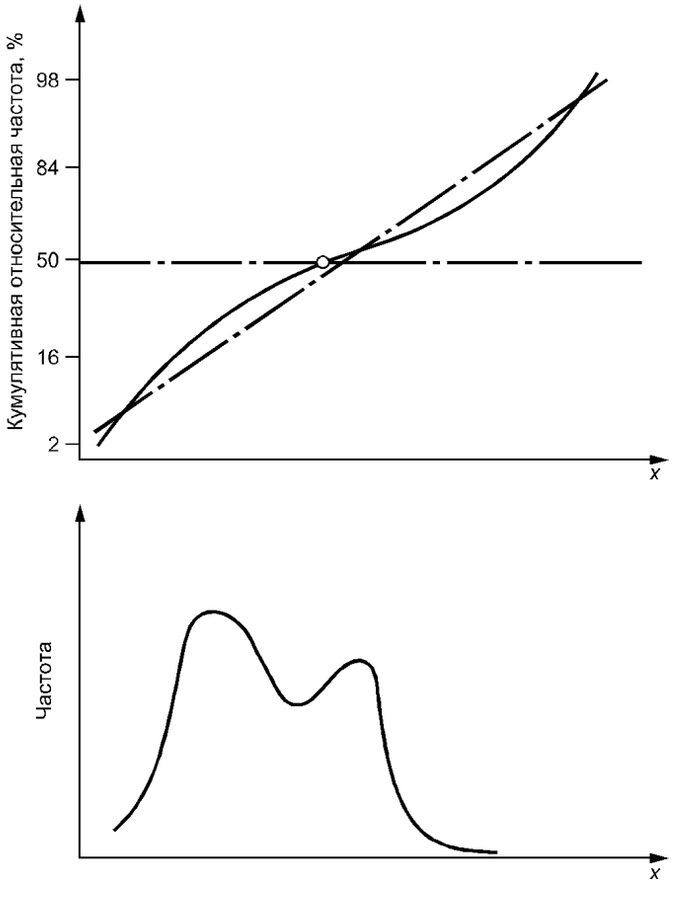


Рисунок 7 — Функции, полученные наложением двух различных кумулятивных функций распределения (вверху) и функций плотности распределения (внизу)

6 Направленные критерии

6.1 Общие положения

6.1.1 Направленные критерии относятся к характеристикам асимметрии или эксцесса распределения вероятностей наблюдений. Они основаны на фактах, что в случае нормальной случайной переменной X со средним $\mu = E(X)$:

центральный момент третьего порядка равен $\mu_3 = E[(X - \mu)^3] = 0$; (2)

нормированный центральный момент третьего порядка (асимметрия совокупности) равен

$$\sqrt{\beta_1} = E\left[\left(\frac{X-\mu}{\sigma}\right)^3\right] = \frac{\mu_3}{\mu_2^{3/2}} = \frac{\mu_3}{\sigma^3} = 0 \ ; \quad (3)$$

нормированный центральный момент четвертого порядка (кривизна совокупности) равен

$$\beta_2 = \mu_4 / \mu_2^2 = 3 \ , \quad (4)$$

где $\mu_2 = E[(X - m)^2]$ — момент второго порядка; (5)

$\mu_4 = E[(X - \mu)^4]$ — момент четвертого порядка; (6)

$\sqrt{\beta_1}$ — асимметрия совокупности, которая может быть большей, равной или меньшей чем нуль;

β_2 — кривизна совокупности (всегда положительная);

$(\beta_2 - 3)$ — эксцесс совокупности.

При этом всегда выполняется неравенство $\beta_2 \geq \left(\sqrt{\beta_1}\right)^2 + 1$.

6.1.2 В критерии на асимметричность альтернативную гипотезу можно задать в виде

H_1 ($\mu_3 > 0$), что эквивалентно $\sqrt{\beta_1} > 0$ [положительная асимметрия, (см. рисунок 5)] или

H_1 ($\mu_3 < 0$), что эквивалентно $\sqrt{\beta_1} < 0$ [отрицательная асимметрия, (см. рисунок 6)].

Распределение вероятностей с положительной асимметрией имеет повышенное рассеяние больших, а не малых значений переменных. Обратное верно для случая с отрицательной асимметрией.

6.1.3 В критерии на эксцесс совокупности альтернативную гипотезу можно задать в виде:

H_1 ($\beta_2 > 3$), что означает большую кривизну [функция плотности распределения более выпуклая (см. рисунок 4)] или

H_1 ($\beta_2 < 3$), что означает меньшую кривизну [функция плотности распределения более плоская (см. рисунок 3)].

По сравнению с нормальным распределением в распределении с большей кривизной преобладают значения переменной, близкие к среднему и к обоим краям. Обратное верно для меньшей кривизны.

6.1.4 Направленный критерий применяют только при наличии конкретной информации о том, как распределение может отличаться от нормального. Эта информация может исходить из физической природы данных или вида возмущения, которое может повлиять на процесс, генерирующий совокупность данных.

Например тот факт, что переменная неотрицательна со средним арифметическим, близким к нулю, в сравнении со значением стандартного отклонения, может быть основанием для положительной асимметрии реального распределения вероятностей . Аналогично, любое возмущение в генерирующем совокупность данных процессе, которое создает смесь нормальных совокупностей с одним средним, но разными дисперсиями, приводит к распределению вероятностей, отличному от нормального с $\beta_2 > 3$.

6.1.5 Выбор направленного критерия следует основывать на общих соображениях о природе наблюдений или процесса, генерирующего эти наблюдения, а не на конкретной форме распределения вероятностей наблюдаемых значений. В последнем случае могут считаться объективными только результаты многостороннего критерия.

6.1.6 Если $x_1, x_2, ..., x_n$ — значения случайной переменной X серии из n наблюдений, то среднее арифметическое или момент первого порядка рассчитывают по формуле

$$\bar{x} = \frac{1}{n} \sum_i x_i \ ; \tag{7}$$

моменты второго, третьего и четвертого порядков соответственно (j = 2, 3, 4) рассчитывают по формуле

$$m_j = \frac{1}{n} \sum_i \left(x_i - \bar{x}\right)^j \ . \tag{8}$$

Статистики критерия на асимметричность и кривизну представлены в следующем виде соответственно:

$$\left| \sqrt{b_1} \right| = \frac{m_3}{m_2^{3/2}} \tag{9}$$

и

$$b_2 = \frac{m_4}{m_2^2} \ . \tag{10}$$

6.2 Направленный критерий проверки на асимметрию, использующий статистику $\sqrt{b_1}$

Этот критерий применим при $n \geq 8$ (из практических соображений таблица 8 ограничена значением $n \leq 5000$).

Если альтернативная гипотеза заключается в наличии положительной асимметрии, критерий следует выполнять только при условии $m_3 > 0$. Если альтернативная гипотеза заключается в наличии отрицательной асимметрии, критерий следует выполнять только при условии $m_3 < 0$.

В обоих случаях решение принимают в пользу отклонения нулевой гипотезы при уровне значимости α, когда статистика $\left| \sqrt{b_1} \right|$ превышает p-квантиль для $p = 1 - \alpha$.

В таблице 8 приведены значения статистики $\left| \sqrt{b_1} \right|$ для $p = 0{,}95$, $p = 0{,}99$ и объема выборки n.

Пример 1

Пример относится к применению направленного критерия с альтернативной гипотезой с использованием статистики $\sqrt{b_1}$. В таблице 2 приведена серия из 50 независимых измерений толщины заболони[1)] деревьев, предназначенных для использования в качестве телеграфных столбов. Поскольку эта толщина неотрицательна и близка к нулю, можно предположить положительную асимметрию. Поэтому необходимо применить соответствующий направленный критерий с альтернативной гипотезой H_1 ($\sqrt{b_1} > 0$).

Из полученных в ходе наблюдений значений, приведенных в таблице 2, вычислены следующие параметры:

$$\bar{x} = (1{,}25 + 1{,}35 + ... + 5{,}10)/50 = 2{,}873;$$
$$m_2 = [(1{,}25 - 2{,}873)^2 + ... +$$
$$+ (5{,}10 - 2{,}873)^2]/50 = 0{,}937\,921;$$
$$m_3 = [(1{,}25 - 2{,}873)^3 + ... +$$
$$+ (5{,}10 - 2{,}873)^3]/50 = 0{,}254\,559,$$

где $\bar{x}$ — среднее арифметическое;

m_2, m_3 — моменты второго и третьего порядков соответственно.

Таблица 2 — Значения толщины заболони для $n = 50$

Толщина заболони				
1,25	2,05	2,60	3,10	4,00
1,35	2,10	2,60	3,15	4,00
1,40	2,15	2,70	3,15	4,05
1,50	2,15	2,75	3,20	4,05
1,55	2,15	2,75	3,30	4,10
1,60	2,20	2,80	3,45	4,20
1,75	2,25	2,95	3,50	4,45
1,75	2,35	2,95	3,50	4,50
1,85	2,40	3,00	3,80	4,70
1,95	2,55	3,05	3,90	5,10

Примечание — Значения толщины заболони расположены в порядке неубывания.

[1)] Заболонь — наружный молодой, менее плотный слой древесины, лежащий непосредственно под корой.

Следовательно асимметрия совокупности равна $\sqrt{b_1} = \frac{m_3}{m_2^{3/2}} = 0{,}280$.

При уровне значимости $\alpha = 0{,}05$, то есть $p = 1 - \alpha = 0{,}95$, и объеме выборки $n = 50$ критическое значение статистики критерия (значение p-квантили) равно 0,53 (см. таблицу 8). Это значение более, чем вычисленное значение $\sqrt{b_1} = 0{,}280$. Значит нулевая гипотеза нормального распределения не отклоняется при выбранном уровне значимости.

6.3 Направленный критерий проверки на кривизну с использованием статистики b_2

Этот критерий применим при $n \geq 8$; из практических соображений таблица 9 ограничена значением $n \leq 5000$.

В критерии на большую кривизну альтернативная гипотеза представлена в следующем виде: H_1 ($\beta_2 > 3$). Если вычисленное значение b_2 превышает критическое значение статистики критерия (значение p-квантили) при $p = 1 - \alpha = 0{,}95$ или $p = 1 - \alpha = 0{,}99$ и объеме выборки n, то нулевая гипотеза должна быть отклонена при определенном уровне значимости, например $a = 0{,}05$ или $a = 0{,}01$.

В критерии на меньшую кривизну альтернативная гипотеза представлена в следующем виде: H_1 ($\beta_2 < 3$). Если вычисленное значение b_2 менее критического значения статистики критерия (значение p-квантили) при $p = \alpha = 0{,}05$ или $p = \alpha = 0{,}01$ и объеме выборки n, то нулевая гипотеза должна быть отклонена при определенном уровне значимости, например $a = 0{,}05$ или 0,01.

В таблице 9 приведены критические значения статистики критерия (значение p-квантили) b_2 для $p = 0{,}01$, 0,05, 0,95 и 0,99 и объема выборки n.

Пример 2

Пример относится к применению направленного критерия на кривизну с использованием статистики b_2. В таблице 3 приведена серия из 50 независимых измерений. В измерительном приборе существует дефект. Некоторые из измерений могли подвергнуться влиянию этого дефекта, приводящего к вариации рассеяния этих измерений.

Из-за упомянутого дефекта можно предположить, что для распределения наблюдений необходимо использовать статистику b_2, соответствующий направленный критерий и альтернативную гипотезу H_1 ($\beta_2 > 3$).

Таблица 3 — Серия из 50 измерений

Значения измерений				
9,5	5,1	5,7	16,6	12,9
14,4	5,8	10,8	20,9	13,3
10,2	9,2	22,5	21,5	8,5
4,2	12,9	5,5	9,1	3,3
17,1	6,3	8,6	11,9	1,4
4,4	3,1	7,4	12,9	12,9
4,5	12,9	6,9	26,6	16,3
8,5	11,9	7,9	7,5	15,6
9,9	11,4	3,6	5,4	11,4
7,7	5,9	7,3	32,0	6,0

На основе полученных значений, приведенных в таблице 3, вычислены следующие параметры:

$$\bar{x} = (9{,}5 + 14{,}4 + \ldots + 6{,}0)/50 = 10{,}542;$$

$$m_2 = [(9{,}5 - 10{,}542)^2 + \ldots + (6{,}0 - 10{,}542)^2]/50 = 37{,}996\ 4;$$

$$m_4 = [(9{,}5 - 10{,}542)^4 + \ldots + (6{,}0 - 10{,}542)^4]/50 = 7\ 098{,}04,$$

где $\bar{x}$ — среднее арифметическое;

m_2, m_4 — моменты второго и четвертого порядков соответственно.

Следовательно кривизна совокупности $b_2 = \frac{m_4}{m_2^2} = 4{,}916$.

При уровне значимости $\alpha = 0{,}05$, то есть $p = 1 - \alpha = 0{,}95$, и объеме выборки $n = 50$ критическое значение статистики (значение p-квантили) критерия равно 3,99 (см. таблицу 9). Вычисленное значение $b_2 = 4{,}916$ более этого критического значения, поэтому нулевая гипотеза отклоняется в пользу альтернативной при уровне значимости $\alpha = 0{,}05$. Это значит, что распределение значений, полученных в результате измерений, искажено и имеет большую кривизну.

Кроме того, критическое значение статистики при уровне значимости $\alpha = 0{,}01$ равно 4,88, поэтому отклонение нулевой гипотезы подтверждено и на этом уровне. Можно сделать вывод, что существование реального искажения данных более правдоподобно.

7 Совместный критерий, использующий статистики $\sqrt{b_1}$ и b_2 (многонаправленный критерий)

Совместный критерий применим при $20 \le n \le 1000$.

7.1 Альтернативная гипотеза состоит в следующем: распределение вероятностей имеет асимметрию, отличную от нуля, и (или) кривизна отлична от кривизны, свойственной нормальному распределению (без указания направления каждого отклонения). Альтернативная гипотеза имеет один из видов:

H_1 ($\sqrt{\beta_1} \neq 0$ и (или) $\beta_2 \neq 3$).

Каждое из сочетаний $\sqrt{\beta_1} \neq 0, \beta_2 \neq 3$ или $\sqrt{\beta_1} \neq 0, \beta_2 = 3$, или $\sqrt{\beta_1} = 0, \beta_2 \neq 3$ может быть выбрано на равных основаниях. Тест многонаправленный и предназначен выявить сочетание ненулевой асимметрии $\sqrt{\beta_1} \neq 0$ и (или) кривизны $\beta_2 \neq 3$.

Данный совместный критерий из-за выбора статистики нельзя считать многосторонним критерием в строгом смысле. Для направленных критериев его применение может быть оправдано только соображениями, основанными на природе наблюдений или процесса, генерирующего это наблюдение.

7.2 Статистика этого критерия образована парой значений статистик $\sqrt{b_1}$ и b_2, определенных в формулах (9) и (10) по 6.1.6. При нулевой гипотезе нормальности можно построить область вокруг точки (0; 3), в которую точка $\left(\left|\sqrt{b_1}\right|; b_2\right)$ попадает с вероятностью p (с осями координат $\sqrt{b_1}$, b_2). Кривые, ограничивающие эту область, показаны на рисунке 9a) при $p = 0{,}95$ и рисунке 9b) при $p = 0{,}99$ для различных объемов выборки n.

При уровне значимости $\alpha = 1 - p$ критическая область критерия образована точками, лежащими вне кривой, соответствующей объему выборки n.

Пример 3

Совместный критерий, использующий статистики $\sqrt{b_1}$ и b_2, можно применить к данным примера 2.

На основе значений, приведенных в таблице 3, вычисляют выборочный центральный момент третьего порядка:

$m_3 = [9{,}5 - 10{,}542)^3 + \ldots + (6{,}0 - 10{,}542)^3]/50 = 308{,}106$.

Отсюда $\sqrt{b_1} = m_3 / m_2^{3/2} = 1{,}315$.

На рисунке 9b) точка $\left(\left|\sqrt{b_1}\right| = 1{,}315; b_2 = 4{,}916\right)$ лежит далеко вне кривой, соответствующей объему выборки $n = 50$ для уровня значимости $\alpha = 0{,}01$.

Поэтому нулевую гипотезу нормального распределения отклоняют на этом уровне значимости в пользу альтернативной гипотезы. Это означает, что распределение вероятностей измеренной характеристики отлично от нормального.

8 Многосторонние критерии

8.1 Общие положения

8.1.1 Многосторонние критерии применяют в том случае, если нет априорной информации о типе отклонения от нормального распределения.

8.1.2 В настоящем стандарте представлены два многосторонних критерия: Шапиро-Уилка и Эппса-Палли. Критерий Шапиро-Уилка выбирают, когда на основе исходных данных можно выбрать альтернативную гипотезу следующего вида: примерно симметричное распределение с убывающей кри-

визной (то есть $\left|\sqrt{\beta_1}\right| < 1/2$ и $\beta_2 < 3$) или асимметричное распределение (например $\left|\sqrt{\beta_1}\right| > 1/2$), в противном случае выбирают критерий Эппса-Палли.

8.2 Критерий Шапиро-Уилка

Этот критерий применим при $8 \le n \le 50$. Малые выборки с $n < 8$ при обнаружении отклонений от нормального распределения не дают достоверных результатов.

Критерий основан на регрессионном анализе порядковых статистик по их ожидаемым значениям. Это критерий типа дисперсионного анализа для полной выборки. Статистика критерия — отношение квадрата суммы линейной разности выборочных порядковых статистик к обычной оценке дисперсии.

Критерий основан на упорядоченных наблюдениях. Если серию из n независимых наблюдений, расположенную в порядке неубывания, как указано в 5.3, обозначить символами $x_1, x_2, \ldots, x_n$, то вычисляют промежуточную сумму S по формуле:

$$S = \sum a_k \left[x_{(n+1-k)} - x_k\right], \qquad (11)$$

где k — индекс, имеющий значения от 1 до $n/2$ или от 1 до $(n-1)/2$ при четном и нечетном n соответственно;

a_k — коэффициент, имеющий специальные значения для объема выборки n (значения a_k приведены в таблице 10).

В этом случае статистика критерия W принимает вид:

$$W = S^2 / (nm_2), \qquad (12)$$

где $nm_2 = \sum (x_i - \bar{x})^2$;

n — объем выборки;

m_2 — выборочный центральный момент второго порядка.

Если значения некоторых наблюдений равны, упорядоченная серия нумеруется с повторением равных наблюдений столько раз, сколько они возникают в исходной серии.

При уровне значимости $\alpha = p$ критическая область критерия образована значениями, меньшими чем p-квантиль для $p = \alpha$. Таблица 11 содержит p-квантили статистики критерия W для $p = \alpha = 0{,}01$ и $p = \alpha = 0{,}05$.

Пример 4

Пример применения критерия Шапиро-Уилка. Таблица 4 содержит упорядоченную серию из $n = 44$ независимых значений годовых осадков, собранных на метеостанции. Используя табличные значения, можно вычислить:

$$\bar{x} = \sum x_k / 44 = 34545 / 44 = 785{,}114.$$

$$nm_2 = \sum (x_k - \bar{x})^2 = 630872.$$

Для облегчения вычислений в таблице приведены значения x_k, $x_{(n+1-k)}$ и $x_{(n+1-k)} - x_k$.

Используя значения коэффициента a_k, взятые из таблицы 10 и воспроизведенные в таблице 4, можно вычислить значения S:

$$S = \sum a_k \left[x_{(n+1-k)} - x_k\right] = 0{,}3872 \times 554 + 0{,}2667 \times 500 + \ldots + 0{,}0042 \times 9 = 787{,}263.$$

Следовательно $W = \dfrac{S^2}{nm_2} = (787{,}2627)^2 / 630872 = 0{,}982.$

Таблица 11 дает значение p-квантиля для $n = 44$ и $p = \alpha = 0{,}05$, равное 0,944. Поскольку это значение менее значения W, то нулевая гипотеза о нормальном распределении не отклоняется при уровне значимости $\alpha = 0{,}05$.

Таблица 4 — Ежегодные осадки, зафиксированные на метеостанции

k	x_k	$x_{(n+1-k)}$	$x_{(n+1-k)}-x_k$	a_k	k	x_k	$x_{(n+1-k)}$	$x_{(n+1-k)}-x_k$	a_k
1	520	1074	554	0,387 2	11	711	873	162	0,104 9
2	556	1056	500	0,266 7	12	713	862	149	0,094 3
3	561	963	402	0,232 3	13	714	851	137	0,084 2
4	616	952	336	0,207 2	14	719	837	118	0,074 5
5	635	926	291	0,186 8	15	727	834	107	0,065 1
6	669	922	253	0,169 5	16	735	826	91	0,056 0
7	686	904	218	0,154 2	17	740	822	82	0,047 1
8	692	900	208	0,140 5	18	744	821	77	0,038 3
9	704	889	185	0,127 8	19	745	794	49	0,029 6
10	707	879	172	0,116 0	20	750	791	41	0,021 1
					21	776	786	10	0,012 6
					22	777	786	9	0,004 2

Примечание — Упорядоченная серия из 44 наблюдений и соответствующие им значения коэффициентов критерия Шапиро-Уилка a_k, где k — номер индекса, $k = 1, 2, ..., 22$.

8.3 Критерий Эппса-Палли

Этот критерий применим при $n \geq 8$. Малые выборки с $n < 8$ при обнаружении отклонений от нормального распределения не дают достоверных результатов.

Многосторонний критерий с высокой мощностью при многих альтернативных гипотезах использует сумму квадратов модулей разности между характеристическими функциями на основе выборочных данных и номального распределения с весомыми коэффициентами.

По n наблюдениям x_j ($j = 1, 2, ..., n$) вычисляют следующие параметры:

$$\bar{x} = \frac{1}{n} \sum_{j=1}^{n} x_j \qquad (13)$$

и

$$m_2 = \frac{1}{n} \sum_{j=1}^{n} \left(x_j - \bar{x}\right)^2 , \qquad (14)$$

где $\bar{x}$ — среднее арифметическое;

m_2 — выборочный центральный момент второго порядка;

n — объем выборки.

Статистику критерия T_{EP} Эппса-Палли вычисляют по формуле

$$T_{EP} = 1 + \frac{n}{\sqrt{3}} + \frac{2}{n} \sum_{k=2}^{n} \sum_{j=1}^{k-1} \exp\left\{\frac{-\left(x_j - x_k\right)^2}{2m_2}\right\} - \sqrt{2} \sum_{j=1}^{n} \exp\left\{\frac{-\left(x_j - \bar{x}\right)^2}{4m_2}\right\}. \qquad (15)$$

Порядок значений произволен, но он должен оставаться неизменным в течение всех проводимых вычислений.

Алгоритм вычисления статистики критерия T_{EP} Эппса-Палли представлен на рисунке 8.

Нулевую гипотезу отклоняют, если вычисленное значение статистики T_{EP} превышает p-квантиль при данных уровне значимости α и объеме выборки n. p-Квантили статистики критерия T_{EP} при $p = 1 - \alpha = 0,90$; 0,95; 0,975 и 0,99 приведены в таблице 12.

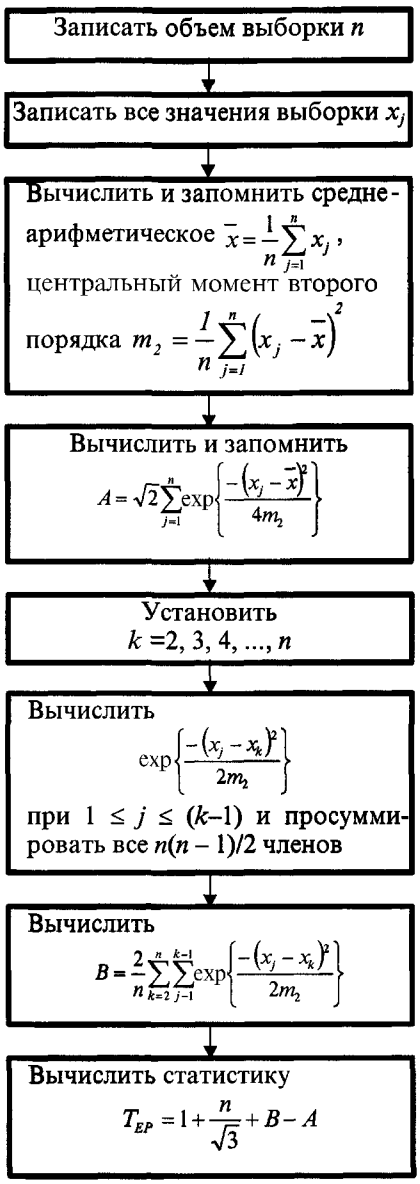


Рисунок 8 — Алгоритм вычисления статистики критерия Эппса-Палли T_{EP}

Пример 5

Пример применения критерия Эппса-Палли. Таблица 5 содержит серию из 25 значений x_j показателя прочности вискозной нити, измеренной при стандартных условиях в произвольных единицах. Дополнительно даны преобразованные значения $z_j = \lg (204 - x_j)$, которые рассеяны около прямой линии, нанесенной на бумаге для нормальных вероятностных графиков.

На основе данных таблицы 5 вычислено значение статистики критерия $T_{EP(x)} = 0{,}612$ с использованием калькулятора. Используем таблицу 12 для поиска значения квантиля уровня $p = 1 - \alpha = 0{,}99$ при $n = 25$. В таблице установлены следующие значения:

для $n = 20$ p-квантиль равен 0,564;

для $n = 30$ p-квантиль равен 0,569.

С помощью интерполяции значений, приведенных в таблице 12, можно оценить, что p-квантиль для $n = 25$ будет равен приблизительно 0,567. Вычисленное значение $T_{EP(x)}$ превышает это критическое значение, поэтому нулевая гипотеза отклоняется при уровне значимости $\alpha = 0{,}01$ для значений x_j.

На основе данных таблицы 5 найдено $T_{EP(z)} = 0{,}006$. Поскольку это значение менее критического значения для $n = 25$, то нулевая гипотеза для преобразованных значений z_j не отклоняется.

Этот пример подтверждает известный факт, что значения показателя прочности вискозной нити подчиняются логарифмически нормальному закону распределения.

Пример 6

Пример детально описывает процедуру вычисления статистики критерия T_{EP} в соответствии с формулой (15).

Второй столбец таблицы 6 содержит $n = 10$ значениям x_j, для которых должен быть проведен критерий Эппса-Палли. Согласно выражениям (13) и (14), получаем $\bar{x} = 10{,}4$ и $m_2 = 11{,}858$.

Таблица 5 — Значения показателя прочности вискозной нити

Измеренные значения x_j	Преобразованные значения z_j	Измеренные значения x_j	Преобразованные значения z_j
147	1,756	99	2,021
186	1,255	156	1,681
141	1,799	176	1,447
183	1,322	160	1,643
190	1,146	174	1,477
123	1,908	153	1,708
155	1,690	162	1,623
164	1,602	167	1,568
183	1,322	179	1,398
150	1,732	78	2,100
134	1,845	173	1,491
170	1,531	168	1,556
144	1,778		

Двойная сумма в третьем члене выражения (15) является конечной серией (n — 1) подсерий, первая из которых имеет один член, а последняя (n — 1) член.

Для первой подсерии установлен индекс $k = 2$ и единственный член суммы, равный

$$\exp\left\{\frac{-(x_1 - x_2)^2}{2m_2}\right\},$$ получен при $j = 1$. Во второй подсерии установлен индекс $k = 3$ и сумма имеет два члена, равные:

$$\exp\left\{\frac{-(x_1 - x_3)^2}{2m_2}\right\} \text{ и } \exp\left\{\frac{-(x_2 - x_3)^2}{2m_2}\right\},$$

которые получены при $j = 1$ и $j = 2$. Для последней подсерии фиксирован индекс $k = 10$, и сумма имеет 9 членов, равные:

$$\exp\left\{\frac{-(x_1 - x_{10})^2}{2m_2}\right\}, \ldots, \exp\left\{\frac{-(x_9 - x_{10})^2}{2m_2}\right\},$$

которые получены при $j = 1, 2, 3, \ldots, 9$.

Таблица 6 — Значения показателя прочности вискозной нити — вычисление статистики критерия T_{EP}

$\exp\left\{\frac{-(x_j - x_k)^2}{2m_2}\right\}$											$\exp\left\{\frac{-(x_j - \bar{x})^2}{4m_2}\right\}$
		$k = 2$	$k = 3$	$k = 4$	$k = 5$	$k = 6$	$k = 7$	$k = 8$	$k = 9$	$k = 10$	
j	x_j	$j =$	$j =$	$j =$	$j =$	$j =$	$j =$	$j =$	$j =$	$j =$	$j = 1, \ldots, 10$
1	4,9	0,9996	0,8977	0,2192	0,2083	0,1684	0,0769	0,0587	0,0304	0,0205	0,5285
2	5,0	—	0,9095	0,2304	0,2192	0,1778	0,0821	0,0629	0,0329	0,0222	0,5407
3	6,5	—	—	0,4421	0,4258	0,3633	0,1977	0,1593	0,0933	0,0673	0,7257
4	10,9	—	—	—	0,9996	0,9895	0,8723	0,8154	0,6668	0,5790	0,9947
5	11,0	—	—	—	—	0,9933	0,8853	0,8303	0,6842	0,5966	0,9924
6	11,4	—	—	—	—	—	0,9312	0,8853	0,7520	0,6668	0,9791
7	12,7	—	—	—	—	—	—	0,9933	0,9312	0,8723	0,8945
8	13,1	—	—	—	—	—	—	—	0,9664	0,9207	0,8575
9	14,0	—	—	—	—	—	—	—	—	0,9895	0,7609
10	14,5	—	—	—	—	—	—	—	—	—	0,7016
Сумма	104,0	0,9996	1,8072	0,8916	1,8528	2,6923	3,0455	3,8052	4,1573	4,7350	7,9757
Общая сумма						23,9865					

Значения членов для n—1=9 подсерий перечислены в столбцах 3—11 таблицы 6.

12-й столбец показывает $n = 10$ членам суммы в четвертом члене выражения (15).

Для каждого из последних 10 столбцов таблицы 6 вычислены их суммы и указаны внизу столбца.

Все 45 членов, принадлежащих сумме в третьем члене выражения (15), после суммирования дали общую сумму $\sum_{k=2}^{10} \sum_{j=1}^{k-1} \exp\left\{\frac{-(x_j - x_k)^2}{2m_2}\right\} = 23{,}9865.$

Окончательно выражение (15) подсчитано и равно

$$T_{EP} = 1 + \frac{10}{\sqrt{3}} + \left(\frac{2}{10} \times 23{,}9865\right) - \left(\sqrt{2} \times 7{,}9757\right) = 0{,}2914.$$

При $n = 10$ таблица 12 содержит значение p-квантиля для $p = 1-\alpha = 0{,}95$, равное 0,357. Вычисленное значение $T_{EP} = 0{,}2914$ не превышает это критическое значение, поэтому в этом примере нулевая гипотеза не отклоняется при уровне значимости $\alpha = 0{,}05$.

9 Совместный критерий, использующий несколько независимых выборок

Критерий применим при нескольких выборках одинакового объема с $n \geq 8$, однако из практических соображений таблица 13 ограничена значениями коэффициентов для $n \leq 50$. Она основана на предположении, что независимые выборки взяты из одной совокупности.

Во многих случаях необходимо проверить отклонение от нормального распределения, используя несколько независимых выборок, поскольку каждая отдельная выборка слишком мала, чтобы обнаружить даже значительное отклонение от нормального распределения. В такой ситуации применяют модифицированный критерий Шапиро-Уилка.

Для h последовательных выборок объемом n каждая, отобранных из одной совокупности, подсчитывают значения W_j ($j = 1, 2, ..., h$) согласно выражению (12). Для совместного критерия вычисляют промежуточные значения статистики G_j совместного критерия по формуле

$$G_j = \gamma_n + \delta_n \nu_j , \tag{16}$$

где ν_j — вспомогательная величина совместного критерия, рассчитываемая по формуле

$$\nu_j = \ln\left\{\frac{W_j - \varepsilon_n}{1 - W_j}\right\} ; \tag{17}$$

γ_n, δ_n *и* ε_n — коэффициенты для преобразования W_j в G_j, приведенные в таблице 13.

В случае, если основное распределение вероятностей нормальное, переменные G_j приблизительно следуют стандартному нормальному распределению. Среднее арифметическое значение переменной G_j равно

$$\overline{G} = \frac{1}{h} \sum_{j=1}^{h} G_j , \tag{18}$$

и статистикой критерия является выражение $\sqrt{h} \times \overline{G}$, где h — число последовательных выборок.

Нулевая гипотеза отклоняется при уровне значимости α, если

$$\sqrt{h} \times \overline{G} < u_{1-\alpha} , \tag{19}$$

где $u_{1-a} = u_p$ — p-квантиль стандартного нормального закона распределения.

Пример 7

Пример применения совместного критерия, использующего несколько независимых выборок. Имеются 22 случайные выборки (h =22), каждая объемом $n = 20$, отобранные из одной совокупности. Измеряют значение случайной переменной X этих 20 объектов, которая в предположении имеет распределение, отличное от нормального. Для каждой выборки вычисляют соответствующие значения W_j ($j = 1, 2, ..., 22$) согласно выражению (12). 22 значения W_j приведены в таблице 7. Из таблицы 13 взяты следующие коэффициенты:

$\gamma_{20} = -5{,}153$; $\delta_{20} = 1{,}802$; $\varepsilon_{20} = 0{,}2359$.

Используя эти значения коэффициентов, вычисляют соответствующие 22 значения G_j согласно выражениям (16) и (17); значения W_j и G_j приведены в таблице 7.

Таблица 7 — Значения W_j и G_j для 22 выборок объемом $n = 20$, взятых из одной совокупности

Номер выборки	W_j	G_j	Номер выборки	W_j	G_j
1	0,9543	—0,189	12	0,9218	—1,240
2	0,9645	+0,292	13	0,9551	—0,155
3	0,9148	—1,413	14	0,9338	—0,909
4	0,8864	—2,008	15	0,9584	—0,009
5	0,9573	—0,059	16	0,9088	—1,552
6	0,9158	—1,389	17	0,9028	—1,683
7	0,9462	—0,503	18	0,8947	—1,849
8	0,9277	—1,083	19	0,9488	—0,407
9	0,9639	+0,260	20	0,9445	—0,563
10	0,9363	—0,833	21	0,9471	—0,470
11	0,9067	—1,598	22	0,9451	—0,542
			Сумма		—17,902

По таблице 11, находят значение статистики (значение p-квантили) W, равное 0,868 для $n = 20$ при уровне значимости $\alpha = 0,01$.

Из таблицы 14 значение p-квантили для $\sqrt{h} \times \overline{G}$ равно

$$-u_{1-\alpha} = -u_{0,99} = -2,326$$

при уровне значимости $\alpha = 0,01$.

По результатам обработки каждой из этих 22 выборок нельзя выявить отклонение от нормального распределения при данном уровне значимости $\alpha = 0,01$, поскольку ни одно из значений W_j не менее критического значения 0,868 и ни одно из значений G_j не менее критического значения минус 2,326.

Совместное оценивание всех 22 выборок дает $\overline{G} = -17,902 / 22 = -0,814$ и $\sqrt{h} \times \overline{G} = -3,82$.

Значение $\sqrt{h} \times \overline{G}$ сравнивается со значением $-u_p = -2,326$ при уровне значимости $\alpha = 0,01$. Поскольку вычисленное значение минус 3,82 лежит значительно ниже найденного в таблице 14 значения, то нулевую гипотезу отклоняют при уровне значимости $\alpha = 0,01$.

10 Статистические таблицы

Таблица 8 — Критерий проверки на асимметрию; значения p-квантили для статистики $\left|\sqrt{b_1}\right|$ при $p = 1 - \alpha = 0,95$ и 0,99

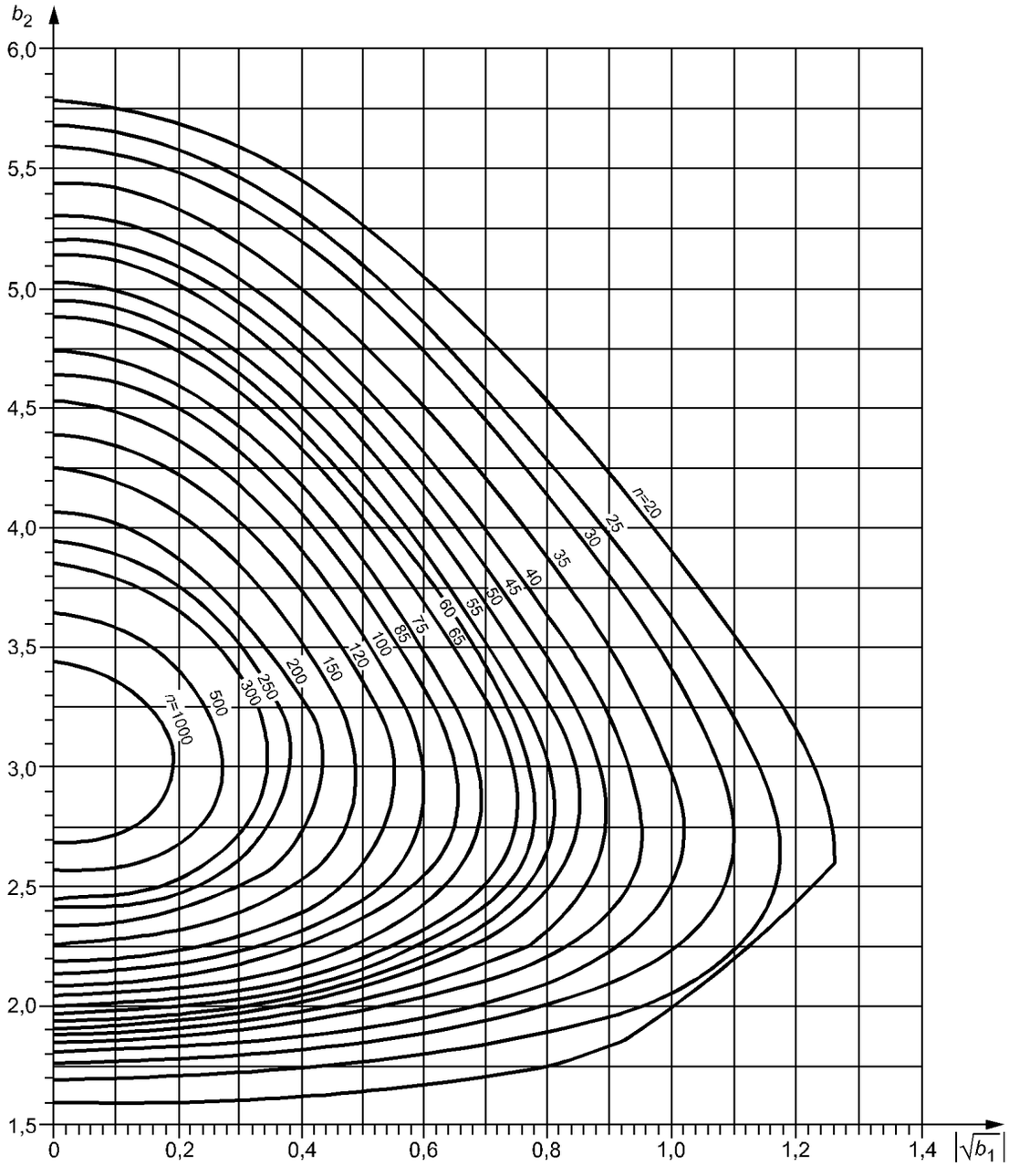
n	p		n	p	
	0,95	0,99		0,95	0,99
8	0,99	1,42	60	0,49	0,72
9	0,97	1,41	70	0,46	0,67
10	0,95	1,39	80	0,43	0,63
12	0,91	1,34	90	0,41	0,60
15	0,85	1,26	100	0,39	0,57
20	0,77	1,15	125	0,35	0,51
25	0,71	1,06	150	0,32	0,46
30	0,66	0,98	175	0,30	0,43
35	0,62	0,92	200	0,28	0,40
40	0,59	0,87	250	0,25	0,36
45	0,56	0,82	300	0,23	0,33
50	0,53	0,79	350	0,21	0,30

Окончание таблицы 8

n	p		n	p	
	0,95	0,99		0,95	0,99
400	0,20	0,28	1000	0,13	0,18
450	0,19	0,27	1200	0,12	0,16
500	0,18	0,26	1400	0,11	0,15
550	0,17	0,24	1600	0,10	0,14
600	0,16	0,23	1800	0,10	0,13
650	0,16	0,22	2000	0,09	0,13
700	0,15	0,22	2500	0,08	0,11
750	0,15	0,21	3000	0,07	0,10
800	0,14	0,20	3500	0,07	0,10
850	0,14	0,20	4000	0,06	0,09
900	0,13	0,19	4500	0,06	0,08
950	0,13	0,18	5000	0,06	0,08

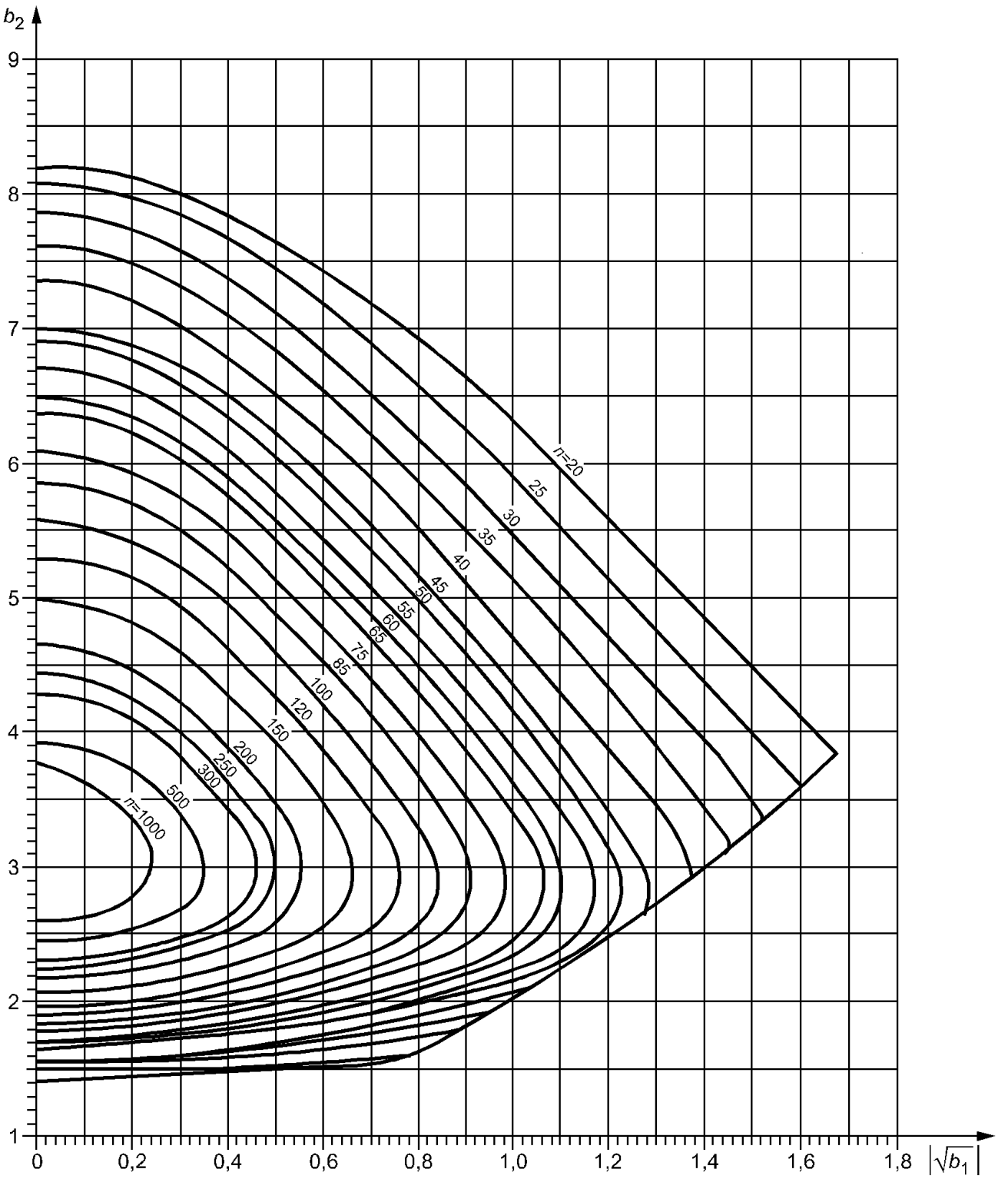
Таблица 9 — Критерий проверки на кривизну; значения p-квантили для статистики b_2 при $p = \alpha = 0{,}01$ и $0{,}05$ и $p = 1 - \alpha = 0{,}95$ и $0{,}99$

n	p				n	p			
	0,01	0,05	0,95	0,99		0,01	0,05	0,95	0,99
8	1,31	1,46	3,70	4,53	600	2,60	2,70	3,34	3,54
9	1,35	1,53	3,86	4,82	650	2,61	2,71	3,33	3,52
10	1,39	1,56	3,95	5,00	700	2,62	2,72	3,31	3,50
12	1,46	1,64	4,05	5,20	750	2,64	2,73	3,30	3,48
					800	2,65	2,74	3,29	3,46
15	1,55	1,72	4,13	5,30	850	2,66	2,74	3,28	3,45
20	1,65	1,82	4,17	5,36	900	2,66	2,75	3,28	3,43
25	1,72	1,91	4,16	5,30	950	2,67	2,76	3,27	3,42
30	1,79	1,98	4,11	5,21	1000	2,68	2,76	3,26	3,41
35	1,84	2,03	4,10	5,13	1200	2,71	2,78	3,24	3,37
40	1,89	2,07	4,05	5,04	1400	2,72	2,80	3,22	3,34
45	1,93	2,11	4,00	4,94	1600	2,74	2,81	3,21	3,32
50	1,95	2,15	3,99	4,88	1800	2,76	2,82	3,20	3,30
75	2,08	2,27	3,87	4,59	2000	2,77	2,83	3,18	3,28
100	2,18	2,35	3,77	4,39	2500	2,79	2,85	3,16	3,25
125	2,24	2,40	3,71	4,24	3000	2,81	2,86	3,15	3,22
150	2,29	2,45	3,65	4,13	3500	2,82	2,87	3,14	3,21
200	2,37	2,51	3,57	3,98	4000	2,83	2,88	3,13	3,19
250	2,42	2,55	3,52	3,87	4500	2,84	2,88	3,12	3,18
300	2,46	2,59	3,47	3,79	5000	2,85	2,89	3,12	3,17
350	2,50	2,62	3,44	3,72					
400	2,52	2,64	3,41	3,67					
450	2,55	2,66	3,39	3,63					
500	2,57	2,67	3,37	3,60					
550	2,58	2,69	3,35	3,57					



а) Кривые, выделяющие критическую область на уровне значимости $\alpha = 0{,}05$

Рисунок 9 — Совместный критерий, использующий статистики $\left|\sqrt{b_1}\right|$ и b_2 (многонаправленный критерий), лист 1



b) Кривые, выделяющие критическую область на уровне значимости $\alpha = 0{,}01$

Рисунок 9. Лист 2

Таблица 10 — Значения коэффициента критерия Шапиро-Уилка для вычисления статистики критерия W

k	n									
								8	9	10
1	—	—	—	—	—	—	—	0,605 2	0,588 8	0,573 9
2	—	—	—	—	—	—	—	0,316 4	0,324 4	0,329 1
3	—	—	—	—	—	—	—	0,174 3	0,197 6	0,214 1
4	—	—	—	—	—	—	—	0,056 1	0,094 7	0,122 4
5	—	—	—	—	—	—	—	—	—	0,039 9
	11	12	13	14	15	16	17	18	19	20
1	0,560 1	0,547 5	0,535 9	0,525 1	0,515 0	0,505 6	0,496 8	0,488 6	0,480 8	0,473 4
2	0,331 5	0,332 5	0,332 5	0,331 8	0,330 6	0,329 0	0,327 3	0,325 3	0,323 2	0,321 1
3	0,226 0	0,234 7	0,241 2	0,246 0	0,249 5	0,252 1	0,254 0	0,255 3	0,256 1	0,256 5
4	0,142 9	0,158 6	0,170 7	0,180 2	0,187 8	0,193 9	0,198 8	0,202 7	0,205 9	0,208 5
5	0,069 5	0,092 2	0,109 9	0,124 0	0,135 3	0,144 7	0,152 4	0,158 7	0,164 1	0,168 6
6	—	0,030 3	0,053 9	0,072 7	0,098 0	0,100 5	0,110 9	0,119 7	0,127 1	0,133 4
7	—	—	—	0,024 0	0,043 3	0,059 3	0,072 5	0,073 7	0,093 2	0,101 3
8	—	—	—	—	—	0,019 6	0,035 9	0,049 6	0,061 2	0,071 1
9	—	—	—	—	—	—	—	0,016 3	0,030 3	0,042 2
10	—	—	—	—	—	—	—	—	—	0,014 0
	21	22	23	24	25	26	27	28	29	30
1	0,464 3	0,459 0	0,454 2	0,449 3	0,445 0	0,440 7	0,436 6	0,432 8	0,429 1	0,425 4
2	0,318 5	0,315 6	0,312 6	0,309 8	0,306 9	0,304 3	0,301 8	0,299 2	0,296 8	0,294 4
3	0,257 8	0,257 1	0,256 3	0,255 4	0,254 3	0,253 3	0,252 2	0,251 0	0,249 9	0,248 7
4	0,211 9	0,213 1	0,213 9	0,214 5	0,214 8	0,215 1	0,215 2	0,215 1	0,215 0	0,214 8
5	0,173 6	0,176 4	0,178 7	0,180 7	0,182 2	0,183 6	0,184 8	0,185 7	0,186 4	0,187 0
6	0,139 9	0,144 3	0,148 0	0,151 2	0,153 9	0,156 3	0,158 4	0,160 1	0,161 6	0,163 0
7	0,109 2	0,115 0	0,120 1	0,124 5	0,128 3	0,131 6	0,134 6	0,137 2	0,139 5	0,141 5
8	0,080 4	0,087 8	0,094 1	0,099 7	0,104 6	0,108 9	0,112 8	0,116 2	0,119 2	0,121 9
9	0,053 0	0,061 8	0,069 6	0,076 4	0,082 3	0,087 6	0,092 3	0,096 5	0,100 2	0,103 6
10	0,026 3	0,036 8	0,045 9	0,053 9	0,061 0	0,067 2	0,072 8	0,077 8	0,082 2	0,086 2
11	—	0,012 2	0,022 8	0,032 1	0,040 3	0,047 6	0,054 0	0,059 8	0,065 0	0,069 7
12	—	—	—	0,010 7	0,020 0	0,028 4	0,035 8	0,042 4	0,048 3	0,053 7
13	—	—	—	—	—	0,009 4	0,017 8	0,025 3	0,032 0	0,038 1
14	—	—	—	—	—	—	—	0,008 4	0,015 9	0,022 7
15	—	—	—	—	—	—	—	—	—	0,007 6
	31	32	33	34	35	36	37	38	39	40
1	0,422 0	0,418 8	0,415 6	0,412 7	0,409 8	0,406 8	0,404 0	0,401 5	0,398 9	0,396 4
2	0,292 1	0,289 8	0,278 6	0,285 4	0,283 4	0,281 3	0,279 4	0,277 4	0,275 5	0,273 7
3	0,247 5	0,246 3	0,245 1	0,243 9	0,242 7	0,241 5	0,240 3	0,239 1	0,238 0	0,236 8
4	0,215 5	0,214 1	0,213 7	0,213 2	0,212 7	0,212 1	0,211 6	0,211 0	0,210 4	0,209 8
5	0,187 4	0,187 8	0,188 0	0,188 2	0,188 3	0,188 3	0,188 3	0,188 1	0,188 0	0,187 8
6	0,164 1	0,165 1	0,166 0	0,166 7	0,167 3	0,167 8	0,168 3	0,168 6	0,168 9	0,169 1
7	0,143 3	0,144 9	0,146 3	0,147 5	0,148 7	0,149 6	0,150 5	0,151 3	0,152 0	0,152 6
8	0,124 3	0,126 5	0,128 4	0,130 1	0,131 7	0,133 1	0,134 4	0,135 6	0,136 6	0,137 6
9	0,106 6	0,109 3	0,111 8	0,114 0	0,116 0	0,117 9	0,119 6	0,121 1	0,122 5	0,123 7
10	0,089 9	0,093 1	0,096 1	0,098 8	0,101 3	0,103 6	0,105 6	0,107 5	0,109 2	0,110 8

Окончание таблицы 10

k	n									
	31	32	33	34	35	36	37	38	39	40
11	0,073 9	0,077 7	0,081 2	0,084 4	0,087 3	0,090 0	0,092 4	0,094 7	0,096 7	0,098 6
12	0,058 5	0,062 9	0,066 9	0,070 6	0,073 9	0,077 0	0,079 8	0,082 4	0,084 8	0,087 0
13	0,043 5	0,048 5	0,053 0	0,057 2	0,061 0	0,064 5	0,067 7	0,070 6	0,073 3	0,075 9
14	0,028 9	0,034 4	0,039 5	0,044 1	0,048 4	0,052 3	0,055 9	0,059 2	0,062 2	0,065 1
15	0,014 4	0,020 6	0,026 2	0,031 4	0,036 1	0,040 4	0,044 4	0,048 1	0,051 5	0,054 6
16	—	0,006 8	0,013 1	0,018 7	0,023 9	0,028 7	0,033 1	0,037 2	0,040 9	0,044 4
17	—	—	—	0,006 2	0,011 9	0,017 2	0,022 0	0,026 4	0,030 5	0,034 3
18	—	—	—	—	—	0,005 7	0,011 0	0,015 8	0,020 3	0,024 4
19	—	—	—	—	—	—	—	0,005 3	0,010 1	0,014 6
20	—	—	—	—	—	—	—	—	—	0,004 9
	41	42	43	44	45	46	47	48	49	50
1	0,394 0	0,391 7	0,389 4	0,387 2	0,385 0	0,383 0	0,380 8	0,378 9	0,377 0	0,365 1
2	0,271 9	0,270 1	0,268 4	0,266 7	0,265 1	0,263 5	0,262 0	0,260 4	0,258 9	0,257 4
3	0,235 7	0,234 5	0,233 4	0,232 3	0,231 3	0,230 2	0,229 1	0,228 1	0,227 1	0,226 0
4	0,209 1	0,208 5	0,207 8	0,207 2	0,206 5	0,205 8	0,205 2	0,204 5	0,203 8	0,203 2
5	0,187 6	0,187 4	0,187 1	0,186 8	0,186 8	0,186 2	0,186 9	0,185 5	0,185 1	0,184 7
6	0,169 3	0,169 4	0,169 5	0,169 5	0,169 5	0,169 5	0,169 5	0,169 3	0,169 2	0,169 1
7	0,153 1	0,153 5	0,153 9	0,154 2	0,154 5	0,154 8	0,155 0	0,155 1	0,155 3	0,155 4
8	0,138 4	0,139 2	0,139 8	0,140 5	0,141 0	0,141 5	0,142 0	0,142 3	0,142 7	0,143 0
9	0,124 9	0,125 9	0,126 9	0,127 8	0,128 6	0,129 3	0,130 0	0,130 6	0,131 2	0,131 7
10	0,112 3	0,113 6	0,114 9	0,116 0	0,117 0	0,118 0	0,118 9	0,119 7	0,120 5	0,121 2
11	0,100 4	0,102 0	0,103 5	0,104 9	0,106 2	0,107 3	0,108 5	0,109 5	0,110 5	0,111 3
12	0,089 1	0,090 9	0,092 7	0,094 3	0,095 9	0,097 2	0,098 6	0,099 8	0,101 0	0,102 0
13	0,078 2	0,080 4	0,082 4	0,084 2	0,086 0	0,087 6	0,089 2	0,090 6	0,091 9	0,093 2
14	0,067 7	0,070 1	0,072 4	0,074 5	0,076 5	0,078 3	0,080 1	0,081 7	0,083 2	0,084 6
15	0,057 5	0,060 2	0,062 8	0,065 1	0,067 3	0,069 4	0,071 3	0,073 1	0,074 8	0,076 4
16	0,047 6	0,050 6	0,053 4	0,056 0	0,058 4	0,060 7	0,062 8	0,064 8	0,066 7	0,068 5
17	0,037 9	0,041 1	0,044 2	0,047 1	0,049 7	0,052 2	0,054 6	0,056 8	0,058 8	0,060 8
18	0,028 3	0,031 8	0,035 2	0,038 3	0,041 2	0,043 9	0,046 5	0,048 9	0,051 1	0,053 2
19	0,018 8	0,022 7	0,026 3	0,029 6	0,032 8	0,035 7	0,038 5	0,041 1	0,043 6	0,045 9
20	0,009 4	0,013 6	0,017 5	0,021 1	0,024 5	0,027 7	0,030 7	0,033 5	0,036 1	0,038 6
21	—	0,004 5	0,008 7	0,012 6	0,016 3	0,019 7	0,022 9	0,022 9	0,028 8	0,031 4
22	—	—	—	0,004 2	0,008 1	0,011 8	0,015 3	0,018 5	0,021 5	0,024 4
23	—	—	—	—	—	0,003 9	0,007 6	0,011 1	0,014 3	0,017 4
24	—	—	—	—	—	—	—	0,003 7	0,007 1	0,010 4
25	—	—	—	—	—	—	—	—	—	0,003 5

Таблица 11 — Критерий Шапиро-Уилка; p-квантили статистики критерия W для $p = \alpha = 0{,}01$ и $0{,}05$

n	p		n	p	
	0,01	0,05		0,01	0,05
			26	0,891	0,920
			27	0,894	0,923
			28	0,896	0,924
			29	0,898	0,926
			30	0,900	0,927
			31	0,902	0,929
			32	0,904	0,930
8	0,749	0,818	33	0,906	0,931
9	0,764	0,829	34	0,908	0,933
10	0,781	0,842	35	0,910	0,934
11	0,792	0,850	36	0,912	0,935
12	0,805	0,859	37	0,914	0,936
13	0,814	0,866	38	0,916	0,938
14	0,825	0,874	39	0,917	0,939
15	0,835	0,881	40	0,919	0,940
16	0,844	0,887	41	0,920	0,941
17	0,851	0,892	42	0,922	0,942
18	0,858	0,897	43	0,923	0,943
19	0,863	0,901	44	0,924	0,944
20	0,868	0,905	45	0,926	0,945
21	0,873	0,908	46	0,927	0,945
22	0,878	0,911	47	0,928	0,946
23	0,881	0,914	48	0,929	0,947
24	0,884	0,916	49	0,929	0,947
25	0,888	0,918	50	0,930	0,947

Таблица 12 — Критерий Эппса-Палли: p-квантили статистики критерия T_{EP} для $p = 1 - \alpha = 0{,}90$; 0,95; 0,975 и 0,99

n	p				n	p			
	0,90	0,95	0,975	0,99		0,90	0,95	0,975	0,99
8	0,271	0,347	0,426	0,526					
9	0,275	0,350	0,428	0,537	50	0,290	0,374	0,460	0,574
10	0,279	0,357	0,437	0,545	100	0,291	0,376	0,464	0,583
15	0,284	0,366	0,447	0,560	200	0,290	0,379	0,467	0,590
20	0,287	0,368	0,450	0,564					
30	0,289	0,371	0,459	0,569					

Таблица 13 — Совместный критерий, использующий несколько независимых выборок: коэффициенты совместного критерия для преобразования W в стандартизованную нормальную переменную для объема выборки n

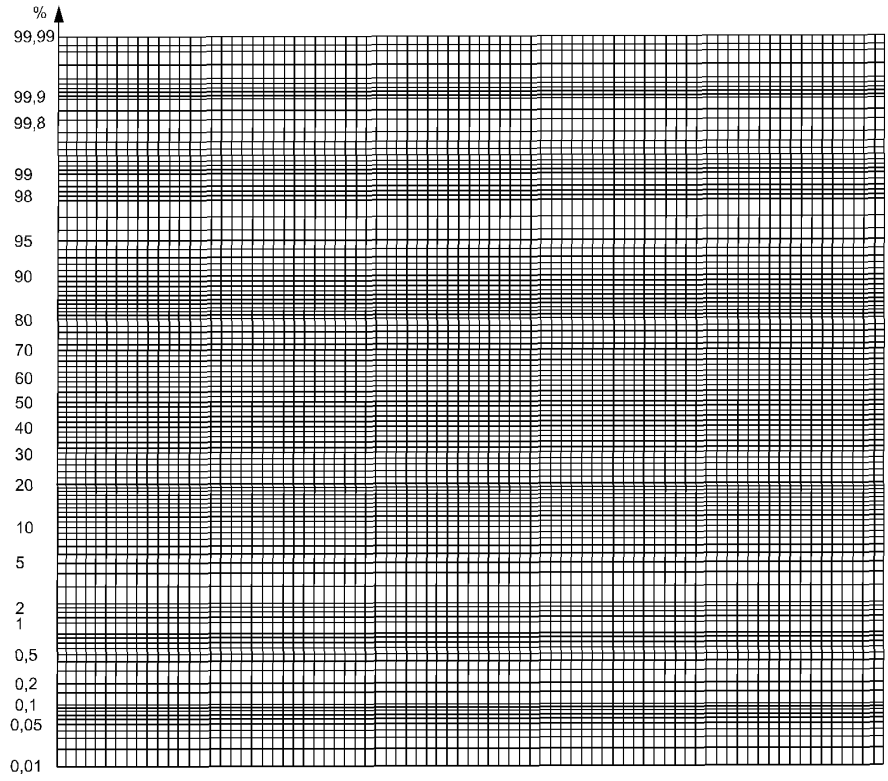
n	γ_n	δ_n	ε_n	n	γ_n	δ_n	ε_n
8	—2,696	1,333	0,4186				
9	—2,968	1,400	0,3900				
10	—3,262	1,471	0,3660				
11	—3,485	1,515	0,3451	31	—6,248	1,965	0,1840
12	—3,731	1,571	0,3270	32	—6,324	1,976	0,1811
13	—3,936	1,613	0,3111	33	—6,402	1,988	0,1781
14	—4,155	1,655	0,2969	34	—6,480	2,000	0,1755
15	—4,373	1,695	0,2842	35	—6,559	2,012	0,1727
16	—4,567	1,724	0,2727	36	—6,640	2,024	0,1702
17	—4,713	1,739	0,2622	37	—6,721	2,037	0,1677
18	—4,885	1,770	0,2528	38	—6,803	2,049	0,1656
19	—5,018	1,786	0,2440	39	—6,887	2,062	0,1633
20	—5,153	1,802	0,2359	40	—6,961	2,075	0,1612
21	—5,291	1,818	0,2264	41	—7,035	2,088	0,1591
22	—5,413	1,835	0,2207	42	—7,111	2,101	0,1572
23	—5,508	1,848	0,2157	43	—7,188	2,114	0,1552
24	—5,605	1,862	0,2106	44	—7,266	2,128	0,1534
25	—5,704	1,876	0,2063	45	—7,345	2,141	0,1516
26	—5,803	1,890	0,2020	46	—7,414	2,155	0,1499
27	—5,905	1,905	0,1980	47	—7,484	2,169	0,1482
28	—5,988	1,919	0,1943	48	—7,555	2,183	0,1466
29	—6,074	1,934	0,1907	49	—7,615	2,198	0,1451
30	—6,150	1,949	0,1872	50	—7,677	2,212	0,1436

Таблица 14 — Значения p-квантилей стандартного нормального закона распределения вероятностей u_p

p, %	α	u_p	p, %	α	u_p
90,0	0,10	1,282	99,0	0,01	2,326
95,0	0,05	1,645	99,5	0,005	2,576
97,5	0,025	1,960			

ПРИЛОЖЕНИЕ А
(справочное)

Бланк бумаги для нормальных вероятностных графиков



УДК 658.562:519.2:006.354 ОКС 03.120.30 Т59 ОКСТУ 0011

Ключевые слова: статистические методы, критерии на отклонение от нормального распределения, мощность критерия, направленный критерий, совместный критерий, многонаправленный критерий

Редактор *Т. С. Шеко*
Технический редактор *В. Н. Прусакова*
Корректор *С. И. Фирсова*
Компьютерная верстка *В. Н. Романовой*

Изд. лиц. № 02354 от 14.07.2000. Подписано в печать 27.03.2002. Усл. печ. л. 3,72. Уч.-изд. л. 3,00. Тираж 446 экз.
С 4803. Зак. 270

ИПК Издательство стандартов,107076 Москва, Колодезный пер., 14.
http://gost.e-ecolog.ru e-mail:gost.e-ecolog.ru
Набрано в Калужской типографии стандартов на ПЭВМ.
Филиал ИПК Издательство стандартов — тип. «Московский печатник», 103062 Москва, Лялин пер., 6.
Плр № 080102

gost.e-ecolog.ru