ИЗДАТЕЛЬСТВО
САНКТ-ПЕТЕРБУРГСКОГО
УНИВЕРСИТЕТА
publishing.spbu.ru

Сумачёв Александр Эдуардович – кандидат технических наук, старший преподаватель кафедры гидрологии суши Санкт-Петербургского государственного университета. Области научных интересов: гидрология суши, ледовый режим, гидрологические расчеты и прогнозы.

Попов Сергей Викторович – доктор геолого-минералогических наук, ведущий геофизик Антарктической геофизической партии АО «Полярная морская геологоразведочная экспедиция», доцент кафедры гидрологии суши Санкт-Петербургского государственного университета. Профессиональная деятельность связана с радиолокационными исследованиями ледников Антарктиды, изучением подледниковых озер, в особенности озера Восток, а также инженерной геофизикой.

В учебно-методическом пособии приведены методы математической статистики и применение анализа при решении комплексных прикладных и научных задач в области гидрометеорологии; изложены общие положения теории вероятностей и математической статистики. В практических работах описано применение изложенных методов на практике. Первая работа посвящена всестороннему анализу случайных величин, показаны алгоритмы построения кривых обеспеченности, в том числе составных; вторая – подходам к прогнозированию экстремальных характеристик водного режима; третья – анализу и прогнозированию временных рядов различными методами.

9 785288 064258

СТАТИСТИЧЕСКАЯ ОБРАБОТКА ГИДРОМЕТЕОРОЛОГИЧЕСКИХ ДАННЫХ
А. Э. Сумачёв, С. В. Попов

Санкт-Петербургский
государственный
университет

А. Э. Сумачёв, С. В. Попов

# СТАТИСТИЧЕСКАЯ ОБРАБОТКА ГИДРОМЕТЕОРОЛОГИЧЕСКИХ ДАННЫХ

УЧЕБНО-МЕТОДИЧЕСКОЕ ПОСОБИЕ

САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

А. Э. Сумачёв, С. В. Попов

# СТАТИСТИЧЕСКАЯ ОБРАБОТКА ГИДРОМЕТЕОРОЛОГИЧЕСКИХ ДАННЫХ

*Учебно-методическое пособие*

ИЗДАТЕЛЬСТВО САНКТ-ПЕТЕРБУРГСКОГО УНИВЕРСИТЕТА

УДК 519.22
ББК 22.172
С89

Рецензенты:
канд. геогр. наук, доц. *П. Н. Священников* (С.-Петерб. гос. ун-т);
канд. геогр. наук, ст. науч. сотр. *Л. С. Банщиков* (Гос. гидролог. ин-т)

*Рекомендовано к публикации*
*Учебно-методической комиссией по УГСН 05.00.00 Науки о Земле*
*Санкт-Петербургского государственного университета*

**Сумачёв А. Э., Попов С. В.**

С89 Статистическая обработка гидрометеорологических данных: учеб.-метод. пособие. — СПб.: Изд-во С.-Петерб. ун-та, 2024. — 132 с.

ISBN 978-5-288-06425-8

В учебно-методическом пособии рассмотрены основные методы статистического анализа гидрометеорологических данных от их обработки до прогнозирования. Приведены исторические и теоретические аспекты теории вероятностей и математической статистики, даны наиболее важные определения. Рассмотрены методы математической статистики и анализа для решения реальных практических задач, связанных с гидрометеорологическими расчётами и прогнозами. Приведены три практические работы по анализу гидрологических наблюдений. В приложении дан справочный материал, необходимый для решения практических задач.

Предназначено для студентов программ бакалавриата и магистратуры гидрометеорологической специальности.

**УДК 519.22**
**ББК 22.172**

*Проект — победитель*
*ежегодного открытого конкурса учебных изданий СПбГУ*
*«Университетский заказ — 2023»*

ISBN 978-5-288-06425-8

# Содержание

# Предисловие

Учебно-методическое пособие состоит из трёх связанных частей. Первая часть охватывает теоретические аспекты математической статистики и анализа данных, приводятся важные определения, иллюстрируемые рисунками и графиками. Вторая часть посвящена практическому применению и закреплению теоретических навыков в ходе решения практических работ, имеющих комплексный характер. В третьей части рассматриваются три практические работы по анализу гидрологических наблюдений.

Практические работы направлены как на углубление уже имеющихся базовых знаний и умений обучающихся по статистике, так и на формирование новых компетенций, связанных с использованием методов кластерного, факторного и регрессионного анализов, методов машинного обучения при решении гидрометеорологических задач. Углубление имеющихся знаний связано в первую очередь с использованием специальных статистических программных продуктов и надстроек *Excel*, которые способны значительно облегчить вычислительные операции, для чего студентам необходимо выполнить комплексную работу по статистическому анализу гидрометеорологической информации. Освоение новых компетенций сопряжено с выполнением ряда практических работ, связанных с расчётом характеристик временны́х рядов и случайных процессов, кластеризации и классификации данных, прогнозирования случайных величин с использованием методов множественной регрессии и машинного обучения. В каждой работе приведены теоретические основы её выполнения, необходимые вычислительные функции, условия и порядок выполнения работы, отчётные материалы и контрольные вопросы.

В приложении приведен справочный материал, необходимый для решения практических задач.

Настоящее пособие предполагает хорошее знание методологии и источников получения исходных гидрометеорологических данных и в свою очередь является основой, на которой базируются специальные гидрометеорологические дисциплины.

# Введение

*Статистика* — отрасль знаний науки, в которой излагаются общие вопросы сбора, измерения, мониторинга, анализа массовых статистических (количественных или качественных) данных и их сравнение; изучение количественной стороны массовых общественных и природных явлений в числовой форме.

В гидрометеорологии великое множество характеристик определяются огромным числом факторов. Согласно центральной предельной теореме в случаях, когда эти факторы по отдельности слабы и ни один из них не является доминирующим, распределение подобных величин является нормальным или стремится к таковому. Наличие как минимум четырёх равнозначных факторов уже приводит к тому, что распределение случайной величины стремится к нормальному. Например, в соответствии с уравнением теплового баланса океана изменение температуры поверхностного слоя воды определяется восемью основными факторами. При этом ценность данных факторов в значительной степени зависит как от масштабов временно́го осреднения процессов формирования теплового баланса, так и от географического района океана. Например, при периоде осреднения, равном 1 мес., для большинства районов океана преобладающим фактором оказывается годовой ход коротковолнового притока солнечной радиации, который может значительно превышать вклад в изменения температуры воды других тепловых процессов. Именно вследствие преобладания этого фактора распределение среднемесячных значений температуры поверхности океана обычно не подчиняется нормальному закону. Если же в качестве масштаба временного осреднения взять один год, то радиационный фактор уже, как правило, не даёт преобладающего вклада в колебания температуры поверхности океана. Поэтому распределение средних годовых значений температуры, в отличие от среднемесячных величин, носит значительно более симметричный характер.

Таким образом, статистика является инструментом для количественного анализа и описания каких-либо процессов, происходящих в обществе, природе или в результате опыта. Задача статистического анализа заключается в численно-вероятностном описании сложных процессов, которые не могут быть описаны иными способами. Этим и определяется область применения статистических методов.

# 1. ОСНОВЫ ТЕОРИИ ВЕРОЯТНОСТЕЙ

*Теория вероятностей* — это математическая наука, изучающая закономерности, присущие массовым случайным явлениям. При этом изучаемые явления рассматриваются в абстрактной форме, независимо от их конкретной природы. Иначе говоря, теория вероятностей рассматривает не сами реальные явления, а их упрощённые схемы — *математические модели*. Предметом теории вероятностей являются математические модели случайных явлений. Под случайным явлением понимают такое явление, исход которого предсказать невозможно, поскольку при неоднократном воспроизведении одного и того же опыта он протекает каждый раз по-разному. В качестве примеров случайных явлений можно назвать выпадение герба при подбрасывании монеты, выигрыш по купленному лотерейному билету, результат измерения какой-либо величины, длительность работы телевизора и т. п. Цель теории вероятностей — осуществление прогноза в области случайных явлений, влияние на ход этих явлений, контроль их, ограничение сферы действия случайности [Письменный, 2004].

## 1.1. Историческая справка

У теории вероятностей, в отличие от других разделов математики, не было античных или средневековых предшественников, она целиком является созданием Нового времени, а её строгое обоснование было разработано лишь в 1929 г. В наше время теория вероятностей занимает одно из основных мест в прикладных науках, и нет почти ни одной естественной науки, в которой так или иначе не применялись бы вероятностные методы.

Первые задачи вероятностного характера возникли в различных азартных играх. Французский каноник XIII в. Ришар де Фурниваль (1201–1260) правильно подсчитал все возможные суммы очков после броска трёх костей и указал число способов, которыми может получиться каждая из этих сумм, что можно рассматривать как первую числовую меру ожидаемости события, аналогичную вероятности. В обширной математической энциклопедии «Сумма арифметики, геометрии, отношений и пропорций» итальянца Луки Пачоли (1445–1517) имеется множество оригинальных задач,

например о том, как разделить ставку между игроками, если серия игр прервана досрочно [Реньи, 1980]. Крупный учёный XVI в. Джероламо Кардано (1501–1576) посвятил анализу игры содержательную монографию «Книга об игре в кости» (1576), в которой он провёл полный и безошибочный комбинаторный анализ для значений суммы очков и указал для различных событий степень их вероятности [Майстров, 1967]. Аналогичными вопросами интересовался итальянский алгебраист Никколо Тарталья (1499–1557). Исследованием подобных математических задач занимался и Галилео Галилей (1564–1642), написавший трактат «О выходе очков при игре в кости» (1718). Изложение теории игры у него отличается исчерпывающей полнотой и ясностью. В своём основном трактате «Диалог о двух главнейших системах мира, птолемеевой и коперниковой» автор также указал на возможность оценки погрешности астрономических и иных измерений, допуская при этом, что небольшие ошибки вероятнее, чем значительные, отклонения в обе стороны равновероятны, а средний результат должен быть близок к истинному значению измеряемой величины. Эти качественные рассуждения стали первым в истории предсказанием закона нормального распределения ошибок [Майстров, 1967].

В XVII в. постепенно стало формироваться представление о проблематике теории вероятностей, появились первые математические методы решения вероятностных задач на основе комбинаторики. Основателями математической теории вероятностей стали Блез Паскаль (1623–1662) и Пьер Ферма (1601–1665) [Стройк, 1969]. Б. Паскаль в своих трудах далеко продвинул применение комбинаторных методов, которые систематизировал в книге «Трактат об арифметическом треугольнике» (опубл. 1665). Кроме того, опираясь на вероятностный подход, он даже доказывал, что быть верующим выгоднее, чем атеистом (так называемые пари Паскаля). Приведём фрагмент текста ввиду нетривиальности:

> …скажем так: “Бог либо есть, либо Его нет”. Какой же ответ мы изберём? Разум нам тут не помощник: между нами и Богом — бесконечность хаоса. На самом краю этой бесконечности идёт игра — что выпадет: орёл или решка? На что вы поставите? <…> Давайте подумаем. Поскольку выбор неизбежен, подумаем, что вас меньше затрагивает. Вам грозят два проигрыша: в одном случае проигрыш истины, в другом — блага… Взвесим наш возможный *выигрыш или проигрыш*, если вы поставите на орла, т. е. на Бога. Сопоставим тот и другой: выиграв — вы выиграете всё, проиграв — не потеряете ничего. Ставьте же, не колеблясь, на Бога! <…> Итак, чем вы рискуете, сделав такой выбор? Вы станете честным, неспособным к измене, смиренным, благодарным, творящим добро человеком, способным к нелицеприятной, искренней дружбе. Да, разумеется, для вас будут заказаны низменные наслаждения: слава, сладострастие, но разве вы ничего не получите взамен? Говорю вам, вы много выиграете даже в этой жизни, и с каждым шагом по избранному пути всё несомненнее будет для вас выигрыш и всё ничтожнее то, против чего вы поставили на несомненное и бесконечное, ничем при этом не пожертвовав [Паскаль, 2020. С. 451].

Над вопросами вероятности выигрыша также размышлял Христиан Гюйгенс (1629–1695), который опубликовал трактат «О расчётах при игре в кости» (1657), по сути первое глубокое исследование по теории вероятностей [Стройк, 1969]. В нём автор подробно изложил вопросы, рассмотренные Ферма и Паскалем, но также поставил и новые. Главным достижением голландского учёного стало введение понятия математического ожидания, а также ставший классическим способ его подсчёта [История математики, 1970]. К этому же периоду относятся публикации английских статистиков Джона Граунта (1620–1674) и Уильяма Петти (1623–1687). Обработав данные более чем за столетие, они показали, что многие демографические характеристики лондонского населения, несмотря на случайные колебания, имеют достаточно устойчивый характер. Дж. Граунт также впервые составил таблицы смертности, таблицы вероятности смерти как функции возраста [История математики, 1970]. Вопросами теории вероятностей и её применения к демографической статистике занялись также Иоганн Худде (1628–1704) и Ян де Витт (1625–1672), которые в 1671 г. также составили таблицы смертности и использовали их для вычисления размеров пожизненной ренты. Более подробно данный круг вопросов был изложен в 1693 г. Эдмундом Галлеем (1656–1742) [Реньи, 1980].

На трактат Гюйгенса опирались появившиеся в начале XVIII в. работы Пьера де Монмора (1678–1719) «Опыт исследования азартных игр» и Якоба Бернулли (1655–1705) «Искусство предположений», которые имели для теории вероятностей особо важное значение [Реньи, 1980]. Над книгой Я. Бернулли работал 20 лет, и она был первым систематическим изложением теории вероятностей, а одна из вероятностных схем и распределение случайных величин носят его имя. Достойный вклад в этот раздел математики также внесли Абрахам де Муавр (1667–1754) и Пьер-Симон Лаплас (1749–1827). Огромное значение как для теории вероятностей, так и для науки в целом имел доказанный Я. Бернулли первый вариант закона больших чисел (название закону дал позже С. Пуассон), который объясняет, почему статистическая частота при увеличении числа наблюдений сближается с теоретическим её значением. В дальнейшем закон больших чисел трудами многих математиков был значительно обобщён и уточнён [Майстров, 1967].

Трактат Я. Бернулли вызвал резкий подъём интереса к вероятностным проблемам и рост числа исследований новых задач. Абрахам де Муавр опубликовал ряд работ, среди которых статья «Об измерении случайности, или вероятностях результатов в азартных играх» (1711) и трактат «Учение о случаях» (1718). В трактате автор решил так называемую задачу о разорении игрока. Суть игры в следующем. Некто играет в орлянку. У него одна монета, а у его противника иное количество. Каковы шансы у игроков на выигрыш и на разорение? И какое количество партий до разорения одного из игроков? В другой работе («Аналитическая смесь») автор дал первый вариант теоремы Муавра — Лапласа, исследующей распределение возможных отклонений статистической частоты от вероятности. Ещё одним достижением А. де Муавра стало введение в науку нормального распределения (1733), которое появилось у него как ап-

проксимация биномиального распределения [Стройк, 1984]. Даниил Бернулли (1700–1782), племянник Я. Бернулли, также внёс вклад в эту науку. Он независимо от А. де Муавра исследовал нормальное распределение для ошибок наблюдений и первым применил к вероятностным задачам методы математического анализа.

Следующий важный шаг в теории вероятностей сделал английский математик Томас Симпсон (1710–1761), который в своей монографии «Природа и законы случая» (1740) использовал третье (наряду с классическим и статистическим) определение вероятности: геометрическое. Подход Симпсона развил Жорж-Луи де Бюффон (1707–1788), который в 1777 г. привёл классический пример подобного рода задач, так называемую задачу Бюффона о бросании иглы: плоскость разграфлена «в линейку», на неё наудачу бросается игла, требуется найти вероятность того, что игла пересечёт линию. В 1901 г. итальянский математик Марио Лаццарини использовал её для опытного определения числа π. Английским математиком Томасом Байесом (1702–1761) была решена важнейшая и фундаментальная задача о сложении вероятностей для нескольких несовместимых событий и получена основополагающая в теории вероятностей и статистике «формула Байеса» (опубл. 1763). К середине XVIII в. анализ игр всё ещё был привлекателен. В частности, Леонард Эйлер (1707–1783) дал подробный анализ разных типов лотерей, но центром внимания математиков всё в большей степени становятся демографическая статистика, страхование и оценка ошибок. Этим вопросам Эйлер посвятил множество работ, в частности решил задачу оценки вероятности того, что человек в возрасте *m* лет проживёт ещё *n* лет [Майстров, 1967].

В XIX в. число работ по теории вероятностей продолжало расти, а её математический аппарат продолжал совершенствоваться. Основной сферой её применения в тот период была математическая обработка данных, содержащих случайные погрешности, а также расчёты рисков в страховом деле и других статистических параметров. Уже к середине XIX в. формируется вероятностная теория артиллерийской стрельбы, а в большинстве крупных стран Европы создаются национальные статистические организации. В конце века область применения вероятностных методов начала успешно распространяться на физику, биологию, экономику, социологию [Стройк, 1969]. Карл Фридрих Гаусс (1777–1855) разработал вероятностную методику работы с измерениями, содержащими погрешности (1809). Он обосновал применение метода наименьших квадратов, а также глубоко изучил нормальное распределение и показал, что оно во многих практических ситуациях является предельным для случайных значений. Его вклад в теорию этого важнейшего распределения столь велик, что долгое время оно носило его имя. Основные достижения теории вероятностей подытожены в капитальном научном труде Лапласа «Аналитическая теория вероятностей», изданном в 1812 г. Симеон Дени Пуассон (1781–1840) в 1837 г. обобщил закон больших чисел. Он же опубликовал формулу Пуассона, удобную для описания схемы Бернулли в том случае, когда вероятность события близка к нулю или к еди-

нице. Распределение Пуассона (закон редких событий) является одним из основных в прикладных задачах.

До середины XIX в. практическое применение теории вероятностей было в основном ограничено статистикой и приближёнными вычислениями. Одним из первых случайных процессов в физике стало изученное Робертом Броуном (1773–1858) в 1827 г. под микроскопом хаотическое движение цветочной пыльцы, плававшей в воде (броуновское движение). Однако его математическая модель появилась лишь в начале XX в. Первые физические вероятностные модели возникли в статистической физике, которую разработали во второй половине XIX в. Людвиг Больцман (1844–1906), Джеймс Клерк Максвелл (1831–1879) и Джозайя Уиллард Гиббс (1839–1903). К концу XIX в. огромное практическое значение вероятностных методов стало общепризнанным фактом.

Математическая статистика, как основа для принятия надёжных решений о случайных величинах, возникла на рубеже XIX–XX вв. благодаря основополагающим работам Карла Пирсона (1857–1936), который разработал теорию корреляции, критерии согласия, регрессионный анализ, алгоритмы проверки гипотез, принятия решений и оценки параметров. Его алгоритмы нашли широкое применение в физике, медицине, биологии, социологии, сельском хозяйстве и смежных науках. Виднейшим продолжателем работ Пирсона по прикладной математической статистике стал Рональд Эйлмер Фишер (1890–1962). Он опубликовал работы по планированию эксперимента, разработал метод максимального правдоподобия, тест статистической значимости, дисперсионный анализ и решение ряда других практически важных статистических задач. Совместно с Ежи Нейманом (1894–1981) он разработал концепцию доверительного интервала (1937), а также является автором такого важного понятия, как «дисперсия случайной величины».

В России в первой половине XIX в. появились собственные серьёзные исследования по теории вероятностей. Первый учебный курс в Вильнюсском университете (1829) начал читать Зигмунд Ревковский. Там же в 1830 г. была создана первая в Российской империи кафедра теории вероятностей. В Петербургском университете лекции с 1837 г. читал сначала Викентий Александрович Анкудович (1790–1876), а с 1850 года Виктор Яковлевич Буняковский (1804–1889), который в 1846 г. опубликовал фундаментальный учебник «Основания математической теории вероятностей», а его русская терминология стала в нашей стране общепринятой. В Московском университете курс теории вероятностей появился в 1850 г. Лекции читал Август Юльевич Давидов (1823–1886), будущий президент Московского математического общества. Работы по этой тематике публиковали многие крупные математики России, в том числе Михаил Васильевич Остроградский (1801–1861), Николай Иванович Лобачевский (1792–1856), Николай Ефимович Зернов (1804–1862), Пафнутий Львович Чебышёв (1821–1894) и его ученики: Андрей Андреевич Марков (1856–1922), Александр Михайлович Ляпунов (1857–1918) и Андрей Николаевич Колмогоров (1903–1987).

## 1.2. Основы комбинаторики

*Комбинаторика* — это раздел математики, который изучает задачи выбора и расположения элементов из некоторого множества в соответствии с заданными правилами. Её формулы и принципы используются в теории вероятностей для подсчёта вероятности случайных событий и получения законов распределения случайных величин. Рассмотрим основные правила комбинаторики.

**Правило суммы.** Пусть имеется два действия $A$ и $B$, которые взаимно исключают друг друга, при этом действие $A$ можно выполнить $m$ способами, а $B$ — $n$ способами. Тогда выполнить любое из них, $A$ или $B$, можно $n + m$ способами.

*Пример*. В урне находятся 16 белых шаров и 10 чёрных. Сколькими способами можно вытащить один шар?

*Решение.* Шар можно вытащить или белый, или чёрный. По правилу суммы получаем, что это можно сделать $16 + 10 = 26$ способами.

**Правило произведения.** Пусть требуется выполнить последовательно $k$ действий. Если первое можно выполнить $n_1$ способами, второе — $n_2$ способами, третье — $n_3$ способами и так до $k$-го, которое можно выполнить $n_k$ способами, то все $k$ действий вместе могут быть выполнены $N$ способами:

$$N = \prod_{J=1}^{k} n_J.$$

*Пример.* В урне находятся 16 белых шаров и 10 чёрных. Сколькими способами можно вытащить два шара?

*Решение.* Первый шар можно вытащить либо белый, либо чёрный, и сделать это можно 26 способами. Второй шар может быть выбран из оставшихся 25 штук, т. е. 25 способами. При этом такое количество способов относится к каждому из выбранных шаров первый раз, т. е. общее количество способов составляет $26 \cdot 25 = 650$ вариантов.

**Сочетания без повторений**. Это классическая задача комбинаторики, которая отвечает на вопрос: сколькими способами можно выбрать $k$ из $n$ различных вариантов ( $C_n^k$ )?

$$C_n^k = \frac{n!}{k!(n-k)!}. \tag{1.1}$$

*Пример*. В урне находится 10 шаров. Сколькими способами можно вытащить четыре шара?

*Решение.* Поскольку порядок вытаскивания шаров значения не имеет, нужно найти число сочетаний 10 элементов по 4. Оно равно $10!/(6! \cdot 4!) = 210$ способов.

**Сочетания с повторениями.** Это несколько иная задача. Пусть имеется по $r$ одинаковых предметов каждого из $n$ различных типов. Сколькими способами можно выбрать $k$ из этих $n \cdot r$ предметов ( $\overline{C}_n^k$ )?

$$\overline{C}_n^k = C_{n+m-1}^m \frac{(n+m-1)!}{m!(n-1)!}.$$

*Пример. В* урне находятся шары четырёх цветов. Сколькими способами можно вытащить семь шаров?

*Решение.* Так как среди выбранных шаров могут быть шары одного цвета, то число способов определяется числом сочетаний с повторениями 7 по 4 и составляет $10!/(7! \cdot 3!) = 120$ вариантов.

**Размещения без повторений.** Это ещё одна классическая задача комбинаторики. Она заключается в определении количества способов размещения по различным местам $k$ из $n$ предметов $(A_n^k)$:

$$A_n^k = \frac{n!}{(n-k)!}. \qquad (1.2)$$

*Пример*. В газете 12 страниц. Сколькими способами можно разместить на них четыре фотографии и не более одной на странице?

*Решение.* Мы не просто выбираем фотографии, а размещаем их на определённых страницах газеты. Таким образом, задача сводится к определению числа размещений без повторений из 12 элементов по 4 и составляет $12!/8! = 11\,880$ вариантов.

**Размещения с повторением.** Также классической задачей комбинаторики является задача о числе размещений с повторениями, о том, сколькими способами можно выбрать и разместить по различным местам $m$ из $n$ предметов, среди которых есть одинаковые. Оно составляет $\overline{A}_n^k = n^k$.

*Пример. В* лифт восьмиэтажного дома вошли четыре человека. Сколькими способами они могут выйти (выход возможен на любом этаже, начиная со второго)?

*Решение. У* каждого пассажира лифта есть семь способов выхода на любом из этажей. Аналогичная возможность имеется и у остальных, поэтому общее число вариантов составляет $7^4 = 2401$.

**Перестановки без повторений.** Это ещё одна классическая задача о том, сколькими способами можно разместить $n$ различных предметов на $n$ различных местах ($P_n$):

$$P_n = n! \qquad (1.3)$$

*Пример*. Сколькими способами можно разместить пять монет?

*Решение.* Нужно определить общее число комбинаций размещения пяти предметов. Оно составляет $5! = 120$.

Соотношение (1.2), согласно (1.1) и (1.3), можно переписать: $A_n^k = C_n^k \cdot k!$.

**Перестановки с повторениями.** Это аналогичная задача о количестве способов перестановки $n$ предметов, расположенных на $n_j$ различных местах, если среди них имеется $k$ одинаковых. Оно составляет $P_{n_1, n_2, n_3 \ldots, n_k}$, $P_{n_1, n_2, n_{36} \ldots, n_k} = n!/(n_1!\, n_2! \ldots n_k!)$.

*Пример.* Сколько разных буквосочетаний можно сделать из букв слова «Миссисипи»?

*Решение.* Здесь одна буква «м», четыре буквы «и», три буквы «с» и одна буква «п». Всего девять букв. Следовательно, число перестановок с повторениями равно $9!/(1! \cdot 4! \cdot 3! \cdot 1!) = 2520$.

## 1.3. События, вероятность, действия над событиями

Определим основные понятия теории вероятностей. *Случайное событие* (или просто *событие*) — это исход некоторого *опыта* (*испытания*), которое может произойти либо не произойти. Например, в результате опыта (или испытания) подбрасывается кубик, и событием $A$ может быть выпадение пяти очков, событием $B$ — трёх очков. Непосредственные исходы опыта называются *элементарными событиями*. Множество всех элементарных событий называется *пространством элементарных событий*, или *пространством исходов*. Событие называется *достоверным*, если оно обязательно наступит в результате данного опыта. Событие называется *невозможным*, если оно заведомо не произойдёт в результате проведения опыта. Два события называются *несовместными*, если появление одного из них исключает появление другого события в одном и том же опыте, в противном случае события называются *совместными*. События называются *попарно несовместными*, если любые два из них несовместны. Несколько событий образуют полную группу, если они попарно несовместны и в результате каждого опыта происходит одно и только одно из них. Несколько событий в некотором опыте называются *равновозможными*, если ни одно из них не является объективно более возможным, чем другие.

*Вероятностью* события $A$ ($P(A)$) называется отношение количества элементарных благоприятных исходов к общему количеству всех равновозможных несовместных элементарных исходов. *Шансом* называется отношение вероятностей благоприятных и неблагоприятных событий. Например, в урне три шара: два белых и один чёрный. Шанс вытащить белый шар 2 : 1, а чёрный 1 : 2. Шансу 1:10 соответствует вероятность, равная $1/(1 + 10)$, или 9 %. Вероятности характеризуются рядом свойств. Вероятность достоверного события $A$: $P(A) = 1$. Вероятность невозможного события $A$: $P(A) = 0$. Вероятность случайного события $A$: $0 \le P(A) \le 1$.

*Суммой (или объединением) событий* $A$ и $B$ называется такое событие $C$, $C = A + B$ (или $C = A \cup B$), которое заключается в наступлении хотя бы одного из этих событий. *Произведением (или пересечением) событий* $A$ и $B$ называется такое событие $C$, $C = A \cdot B$ (или $C = A \cap B$), которое состоит в совместном наступлении этих событий. *Противоположным событием* $\overline{A}$ называется событие, которое происходит тогда и только тогда, когда не происходит событие $A$. Разностью событий $A$ и $B$ называется такое событие $C$, $C = A - B$, происходящее тогда и только тогда, когда происходит событие $A$, но не происходит событие $B$. Иначе его можно записать так: $C = A + \overline{B}$. Событие $A$ *влечёт событие* $B$, если из того, что произошло событие $A$, следует наступление события $B$: $A \subseteq B$. Если $A \subseteq B$ и $B \subseteq A$, то события $A$ и $B$ равны, т. е. $A = B$. События и действия над ними наглядно иллюстрируются так называемыми *диаграммами Эйлера — Венна* (рис. 1.1).

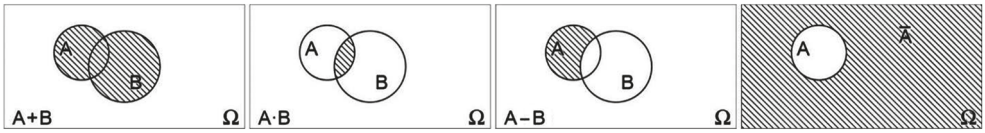



*Рис. 1.1.* Диаграммы Эйлера — Венна (Ω — достоверное событие) [Кузичев, 1968]

Из вышеизложенного имеется ряд следствий. В частности, сумма всех несовместных событий равна единице. Затем, если события независимы, то появление любого из них не меняет вероятности последующего.

Рассмотрим вероятность суммы двух *совместных событий*, т. е. тех, которые могут появляться одновременно. Напомним, что вероятность суммы двух несовместных $A$ и $B$ событий $P(A + B) = P(A) + P(B)$, при этом раз они несовместны, то $P(A \cdot B) = 0$. Представим вероятность суммы совместных событий $P(A + B)$ в качестве суммы вероятностей трёх несовместных: $P(A + \overline{B}) = P(A \cdot \overline{B}) + P(B \cdot \overline{A}) + P(A \cdot B)$. Каждое из них при этом можно также представить как сумму несовместных событий с вероятностями: $P(A) = P(A \cdot \overline{B}) + P(A \cdot B)$ и $P(B) = P(B \cdot \overline{A}) + P(A \cdot B)$. Подставив их в предыдущее соотношение, получим следующее выражение: $P(A + B) = P(A) - P(A \cdot B) + P(B) - P(A \cdot B) + P(A \cdot B)$, или

$$P(A + B) = P(A) + P(B) - P(A \cdot B), \tag{1.4}$$

которое верно для любых случаев. Вероятность суммы трёх событий (как совместных, так и несовместных) равна

$$P(A + B + C) = P(A) + P(B) + P(C) - P(A \cdot B) - P(A \cdot C) - P(B \cdot C) + P(A \cdot B \cdot C).$$

## 1.4. Условная вероятность, полная вероятность, теорема Байеса

До этого рассматривались безусловно-вероятностные события. Обратимся к несколько иному типу задач о вероятности появления некоторого события при условии появления другого. Условная вероятность $P(A \mid B)$ — это насту-

пление события $A$ при условии, что событие $B$ уже произошло. Например, вероятность того, что из колоды вытянута карта чёрной масти при условии, что до этого вытянута карта масти такого же цвета. Условная вероятность является одним из наиболее фундаментальных и наиболее важных понятий теории вероятностей.

*Условной вероятностью* события $A$ при условии, что произошло событие $B$, называется отношение вероятности произведения этих событий к вероятности события $B$, при этом $P(B) \neq 0$, т. е.

$$P(A \mid B) = \frac{P(A \cdot B)}{P(B)}. \tag{1.5}$$

Вероятность $P(B)$, в отличие от условной, называется *безусловной вероятностью*.

*Задача*. В урне два белых шара и семь чёрных. Из неё последовательно вынимают два шара. Какова вероятность того, что второй шар окажется белым, при условии, что первый шар был чёрным?

*Решение*. Пусть событие $A$ — появление первого чёрного шара, а событие $B$ — второго белого. Поскольку событие $A$ произошло, то в урне осталось на один шар меньше, т. е. восемь, причём из них два белых. Поэтому $P(B \mid A) = 2/8 = 0{,}25$.

Теперь решим эту же задачу с использованием формулы условной вероятности (1.5). Полная вероятность безусловного события $P(A)$, т. е. появление первого чёрного шара, составляет $P(A) = 7/9$. Найдём $P(A \cdot B)$. Событию $AB$ благоприятствуют, согласно (1.1), $C_2^1 \cdot C_7^1 = 14$ исходов; полное количество исходов $9 \cdot 8 = 72$. Таким образом, $P(A \cdot B) = 14/72 = 7/36$. Тогда $P(B \mid A) = 7/36 : 7/9 = 0{,}25$.

С темой условной вероятности непосредственно связано понятие *полной вероятности*. Пусть имеются события $A_1, A_2, A_3, \ldots, A_N$, при этом они образуют полную группу событий. Тогда вероятность любого события $B$

$$P(B) = \sum_{J=1}^{N} P(A_J) \cdot P(B \mid A_J). \tag{1.6}$$

*Задача.* В сборочный цех завода поступает 40 % деталей из первого цеха и 60 % — из второго. В первом цехе производится 90 % качественных деталей, а во втором — 95 %. Требуется найти вероятность того, что деталь, выбранная случайным образом, окажется качественной.

*Решение.* Выбор детали можно разбить на два этапа. Первый — это выбор цеха. Имеется две гипотезы: $A_1$ — деталь изготовлена первым цехом, $A_2$ — вторым. Следующий этап заключается в выборе детали. Событие $B$ — взятая наугад деталь является качественной. Понятно, что события $A_1$ и $A_2$ образуют полную группу, причём $P(A_1) = 0{,}4$ и $P(A_2) = 0{,}6$. Вероятности выбора деталей для цехов таковы: $P(B \mid A_1) = 0{,}90$ и $P(B \mid A_2) = 0{,}95$. По формуле (1.6) получаем, что $P(B) = 0{,}4 \cdot 0{,}9 + 0{,}6 \cdot 0{,}95 = 0{,}93$, или 93 %.

Важным следствием соотношений (1.5) и (1.6) является *теорема Байеса*, *формула Байеса* или *теорема гипотез* [Письменный, 2004]. Это одно из основных и наиболее значимых понятий теории вероятностей, названное в честь её автора Томаса Байеса (1702–1761), английского математика и священника. Сэр Гарольд Джеффрис (1891–1989), английский математик, статистик, геофизик и астроном, писал, что теорема Байеса для теории вероятностей то же, что теорема Пифагора для геометрии [Jeffreys, 1973]. Она позволяет переоценить вероятности гипотез, принятых до опыта (априорных), по результатам уже проведённого опыта, т. е. получить апостериорные оценки. Теорема Байеса звучит следующим образом. Пусть события $A_J$, $J=1\ldots N$ образуют полную группу событий. Тогда вероятность наступления события $AJ$ при условии, что произошло событие $B$, $P(A_J|B)$, такова, что

$$P(A_J \mid B) = \frac{P(A_J)\cdot P(B \mid A_J)}{P(B)}, \tag{1.7}$$

при этом $P(B)$ является полной вероятностью, которую можно вычислить по формуле (1.6).

*Задача*. Некто, почувствовав себя плохо, пришёл к врачу. Врач предложил больному сдать анализы, которые показали наличие некоторого заболевания. Какова вероятность того, что он действительно болен при условии, что анализ определяет факт заболевания в 80 % случаев, а в 40 % случаев даёт ложно положительный результат?

*Решение*. Для решения этой задачи воспользуемся теоремой Байеса об условной вероятности. Пусть событие $A$ заключается в том, что пациент действительно болен, а событие $B$ в том, что получен положительный тест. Тогда $P(A|B)$ равна произведению вероятности заразиться (т. е. априорной информации, до того, как пройден тест) $P(A)$ на вероятность $P(B|A)$ события в случае, если гипотеза правдива (т. е. на вероятность того, что пациент действительно болен и получил положительный тест), — произведению, поделённому на общую вероятность события $P(B)$ (т. е. на полную вероятность получения положительного теста вне зависимости от здоровья пациента). Но она складывается из вероятности двух событий: вероятности того, что положительный тест получен при условии болезни, т. е. $P(A)\cdot P(B|A)$, и в случае ошибки, т. е. $P(\overline{A})\cdot P(B|\overline{A})$. Окончательно:

$$P(A \mid B) = \frac{P(B \mid A)\cdot P(A)}{P(A)\cdot P(B \mid A) + P(\overline{A})\cdot P(B \mid \overline{A})}.$$

Обычно наиболее сложным вопросом является оценка изначальной вероятности появления события $P(A)$, но в нашем случае это может быть распространённость заболевания в популяции. Если подставить числа в полученную формулу: $(0{,}8\cdot 0{,}001)/(0{,}001\cdot 0{,}8+0{,}999\cdot 0{,}4)$, то получится, что вероятность реального заражения, при наличии положительного теста, составляет всего около 0,5 %.

# 2. ГИДРОМЕТЕОРОЛОГИЧЕСКИЕ ДАННЫЕ И ПОДХОДЫ К ИХ ИССЛЕДОВАНИЮ. ОСНОВЫ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ И АНАЛИЗА ДАННЫХ

## 2.1. Характеристика гидрометеорологических данных

С точки зрения охвата объекта исследования статистический анализ можно подразделить на два вида: *сплошной* и *выборочный*. Первый предполагает изучение генеральной совокупности данных, т. е. явления в целом, во всём его многообразии без распространения выводов на другие элементы, не входящие в анализируемую совокупность. Из названия следует, что этот вид анализа проводится по полным данным, которые охватывают всю возможную вариацию случайной величины. Этот тип статистического исследования является наиболее полным и точным. Примером сплошного статистического анализа можно назвать всеобщую перепись населения. Пограничными в этом плане являются различные опросы общественного мнения, референдумы и т. д. С одной стороны, это является полным анализом, но лишь при условии нераспространения мнения опрошенных на всё население или социальную группу. При экстраполяции результатов голосования на всё население мы уже имеем дело с выборочным анализом, а достоверность результатов подобных оценок зависит от репрезентативности выборки. Как бы там ни было, у сплошного наблюдения есть и очевидные минусы: на организацию и проведение исследования могут потребоваться значительные ресурсы.

Другим видом анализа является выборочный анализ. Название метода точно отражает его суть: из генеральной совокупности отбирается и анализируется только часть данных, а выводы распространяются на всю генеральную совокупность. Отбор данных происходит таким образом, чтобы выборка была репрезентативной, т. е. сохранила внутреннюю структуру и закономерности генеральной совокупности. При соблюдении этого условия есть основания рассчитывать на достаточно точное описание всей генеральной совокупности.

Анализ выборочных данных происходит так же, как и при сплошном наблюдении (рассчитываются различные показатели, делаются прогнозы и т. д.), только с поправкой на ошибку. Это значит, что, рассчитывая тот или иной показатель, мы понимаем, что при повторной выборке его значение будет другим. Допустим, опрос 1000 человек на улице дал некоторый результат. Вполне очевидно, что, опросив другую тысячу человек, будет получен другой результат,

отличный от первого. Однако, если обе выборки репрезентативны, то различия будут статистически незначимые. Положительный аспект заключается в том, что для проведения выборочного обследования требуется гораздо меньше ресурсов. Отрицательный же заключается в том, что выборочное наблюдение всегда ошибочно. Поэтому основная задача проведения подобного исследования в том, чтобы добиться максимальной точности при приемлемых затратах на его проведение. Рассчитывая выборочные параметры, необходимо оценить и параметры генеральной совокупности.

Подавляющее большинство гидрометеорологических характеристик являются выборочными, так как исследователь всегда имеет дело с данными, ограниченными периодом наблюдения, и распространяет свои выводы на все возможные значения случайной величины. К примеру, нормой стока какой-либо реки является среднее значение величины стока за многолетний период такой продолжительности, при увеличении которой полученное значение существенно не меняется. То есть, по сути, норма стока — это достаточно точная оценка математического ожидания случайной величины.

Гидрологическая информация представляется в серии ежегодных изданий водного кадастра. При этом статистическому анализу, как правило, подвергаются среднегодовые и экстремальные значения той или иной гидрометеорологической величины; при исследовании внутригодовых колебаний — среднемесячные или среднесуточные значения.

## 2.2. Математическая статистика: краткая историческая справка

Математическая статистика возникла в XVII в. и развивалась параллельно с теорией вероятностей. Дальнейшее развитие математической статистики (вторая половина XIX — начало XX в.) связано с именами К. Ф. Гаусса, П. Л. Чебышёва, А. А. Маркова, А. М. Ляпунова, а также с А. Кетле, Ф. Гальтоном, К. Пирсоном и др. Математическая статистика как наука начинается с работ знаменитого немецкого математика Карла Фридриха Гаусса (1777–1855), который на основе теории вероятностей исследовал и обосновал метод наименьших квадратов, созданный им в 1795 г. и применённый для обработки астрономических данных (с целью уточнения орбиты малой планеты Церера). Его именем часто называют одно из наиболее популярных распределений вероятностей — нормальное, а в теории случайных процессов основной объект изучения — гауссовские процессы.

В XX в. наиболее существенный вклад в математическую статистику был сделан советскими математиками (В. И. Романовский, Е. Е. Слуцкий, А. Н. Колмогоров, Н. В. Смирнов), а также английскими (Стьюдент (Уильям Сили Госсет), Р. Фишер, Э. Пирсон) и американскими (Ю. Нейман, А. Вальд) учёными.

В области гидрологических статистических расчётов особое место занимают работы С. Н. Крицкого и М. Ф. Менкеля, которым удалось разработать так называемое трёхпараметрическое гамма-распределение, называемое также распределением Крицкого — Менкеля. Дальнейшее развитие методов математической статистики в практике гидрологических расчётов можно отследить по работам Д. Л. Соколовского, А. В. Рождественского, В. А. Рожкова и других авторов.

## 2.3. Понятие закона распределения и методы оценки его параметров

С характеристикой гидрометеорологической информации и с одним из важнейших приёмов математической статистики и анализа — выборочности данных для будущего анализа — мы познакомились в подразд. 2.1. Классическая математическая статистика изучает так называемые случайные величины, т. е. величины, изменяемые случайным образом. Именно из таких величин извлекается выборка, которая также должна отвечать требованию случайности и быть репрезентативна генеральной совокупности. Исчерпывающей характеристикой случайной величины (генеральной совокупности) является её закон распределения, который может быть выражен в аналитическом, графическом или ином виде. Под законом распределения понимается закон, описывающий область значений случайной величины и соответствующие вероятности появления этих значений; в гидрометеорологической практике, как правило, применяется графическое представление законов распределения в виде аппроксимирующих эмпирических данных интегральных кривых превышения (обеспеченности) или непревышения случайной величиной некоторого заданного числа. В российской гидрометеорологической практике применяются именно кривые обеспеченности:

$$P_x = P\{X > x\}, \tag{2.1}$$

где $P_x$ — вероятность $P$ того, что случайная величина $X$ примет значение больше $x$.

Из формулы (2.1) совершенно очевидны свойства функции обеспеченности: при устремлении $x$ к бесконечности вероятность превышения этого значения стремится к нулю, при устремлении $x$ к нулю вероятность превышения этого значения стремится к единице; для всех возможных значений $x$ функция обеспеченности больше или равна нулю, если $x_1 > x_2$, то вероятность превышения $x_1$ меньше, чем $x_2$.

При этом надо разделять *аналитическую* и *эмпирическую* кривые обеспеченности. Эмпирическая кривая обеспеченности рассчитывается непосредственно по данным выборки, аналитическая кривая обеспеченности — по выборочным параметрам распределения и является характеристикой генеральной совокупности. Степень удачности аппроксимации аналитической

кривой обеспеченности эмпирических данных свидетельствует о соответствии выборки данному аналитическому закону распределения.

К наиболее распространённым законам распределения, используемым в практике гидрометеорологических расчётов, относятся: распределение Гаусса (нормальное распределение); логарифмически нормальное распределение, распределение Пирсона III типа, распределение Крицкого — Менкеля.

Нормальное распределение является наиболее типичным для многих естественных процессов. Распределение является двухпараметрическим, т. е. определяется двумя параметрами: математическим ожиданием и стандартным отклонением (симметричным), т. е. коэффициент асимметрии распределения равен нулю. Зачастую случайная величина может иметь несколько аномальных значений, придающих распределению асимметричный вид, приведение такого распределения к симметричному виду проводится путём логарифмирования. Таким образом, логарифмически нормальным распределением обладает такая случайная величина, логарифмы значений которой будут иметь нормальное распределение. Распределения Пирсона III типа и Крицкого — Менкеля в общем случае являются трёхпараметрическими и определяются тремя параметрами: математическим ожиданием, коэффициентом вариации и асимметрии. Соответственно, оба распределения являются асимметричными. Кривая Крицкого — Менкеля используется только для положительной асимметрии и имеет нижний предел, равный нулю, она наиболее подходит для оценки максимальных обеспеченных расходов воды. Кривая Пирсона может применяться как при отрицательной, так и при положительной асимметрии, что делает её пригодной для оценки высших обеспеченных уровней воды, которые зачастую могут иметь отрицательную асимметрию. Более подробно данные кривые обеспеченности рассмотрены в учебнике по методам статистической обработки гидрометеорологической информации [Сикан, 2007] и в соответствующих разделах данного учебно-методического пособия.

Как уже было сказано выше, для аналитического закона распределения следует оценить его параметры, которые не должны изменяться с течением времени. Параметры распределения играют важную самостоятельную роль и могут определяться вне рамок задачи аналитического закона распределения. К основным параметрам распределения относятся математическое ожидание, стандартное отклонение (коэффициент вариации), коэффициент асимметрии. Помимо прочего, перед началом построения кривых обеспеченности надо убедиться в однородности и стационарности ряда, отсутствии выбросов, одномодальности.

Для расчёта параметров распределения используются два основных метода: *метод моментов* и *метод приближённого наибольшего правдоподобия*.

Суть метода моментов заключается в определении параметров аналитического закона распределения по выборке. Такие параметры называются выборочными. Как правило, для оценки аналитического закона распределения достаточно двух-трёх параметров, характеризующих центр распределения, вариацию распределения и его асимметричность. Соответственно, центром

аналитического распределения является математическое ожидание, которое определяется через первый начальный момент, соответствующий среднему арифметическому значению:

$$m_{1.0} = \frac{1}{N}\sum_{i=1}^{N} x_i = \overline{x}\,. \tag{2.2}$$

Данная оценка является эффективной и несмещённой.

Через второй центральный момент рассчитываются дисперсия, стандартное отклонение и коэффициент вариации:

$$m_2 = \frac{1}{N}\sum_{i=1}^{N}\left(x_i - \overline{x}\right)^2 = \frac{1}{N}\sum_{i=1}^{N}\left(k_i - 1\right)^2, \tag{2.3}$$

где $k_i$ — модульный коэффициент рассматриваемой характеристики,

$$k_i = \frac{x_i}{\overline{x}}. \tag{2.4}$$

Однако дисперсия по выборочным данным обладает смещённостью, что особенно заметно при $N < 50$. Смещённость данной оценки обусловлена тем фактом, что при её расчёте используется выборочное среднее значение, которое всегда несколько отличается от математического ожидания. Расстояние до выборочного среднего будет всегда меньше, чем расстояние до математического ожидания, что и обусловливает смещённость данной оценки. Однако при достаточно большой выборке выборочное среднее стремится к истинному значению математического ожидания, поэтому и дисперсия при достаточно большой выборке будет стремиться к своему истинному значению. Для достижения несмещённости оценки выборочной дисперсии и стандартного отклонения используют поправку $N - 1$:

$$D_{\text{несм}} = \frac{1}{N-1}\sum_{i=1}^{N}\left(x_i - \overline{x}\right)^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(k_i - 1\right)^2. \tag{2.5}$$

Стандартное отклонение в свою очередь равно квадратному корню из выборочной дисперсии:

$$\sigma_{\text{несм}} = \sqrt{D_{\text{несм}}}\,. \tag{2.6}$$

Коэффициентом вариации $C_v$ называется отношение стандартного отклонения к математическому ожиданию:

$$C_v = \frac{\sigma_{\text{несм}}}{\overline{x}}. \tag{2.7}$$

Все перечисленные величины, определяемые через второй центральный момент, являются характеристиками рассеивания случайной величины от-

носительно математического ожидания. При этом для разных целей анализа могут использоваться разные величины, например для сравнения изменчивости стока двух разных рек, обладающих разной водностью, целесообразно использовать коэффициент вариации. В то же время рассчитывать коэффициент вариации имеет смысл для величин, измеренных в абсолютной шкале, и для ненулевых средних, так как в противном случае значение коэффициента вариации будет сильно зависеть от выбора начала координат или уйдёт в область бесконечных значений. К примеру, для того чтобы корректно рассчитать коэффициент вариации для температуры воздуха, необходимо перевести градусы Цельсия в кельвины.

Зная математическое ожидание и стандартное отклонение, можно провести стандартизацию случайной величины таким образом, чтобы её среднее значение равнялось нулю, а стандартное отклонение — единице. Данный приём позволяет оценить разброс случайной величины в единицах стандартного отклонения и изучать взаимосвязи двух и более величин безотносительно их размерности. Также стоит отметить, что ординаты кривых обеспеченности, как правило, представляются либо в модульных коэффициентах, либо в стандартизованном виде. Процедура стандартизации проводится следующим образом: от каждого значения случайной величины отнимается величина математического ожидания, затем полученные разности делятся на стандартное отклонение.

На рис. 2.1 проиллюстрировано так называемое *правило трёх сигм*, которое говорит о том, что разброс нормального распределения с вероятностью 0,9973 лежит в диапазоне ± 3σ.

Коэффициент асимметрии определяется через третий центральный мо-

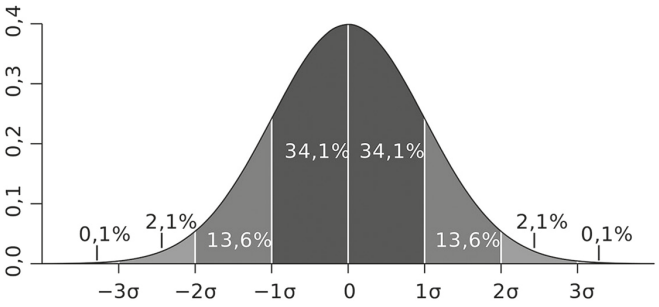



*Рис. 2.1.* Нормальное стандартное распределение.
Составлено по: [Сикан, 2007]

мент: он равен третьему центральному моменту, отнесённому к стандартному отклонению в третьей степени. С учётом поправки на смещённость коэффициент асимметрии $C_s$ примет вид

$$C_s = \frac{N\sum_{i=1}^{N}\left(x_i - \overline{x}\right)^3}{\sigma^3\left(N-1\right)\left(N-2\right)} = \frac{N\sum_{i=1}^{N}\left(k_i - 1\right)^3}{{C_v}^3\left(N-1\right)\left(N-2\right)}. \qquad (2.8)$$

Для трёхпараметрического распределения Крицкого — Менкеля, согласно Своду правил по проектированию и строительству (Определение основных расчётов гидрологических характеристик) СП 33-101-2003, коэффициент вариации и отношение $C_s$ к $C_v$ следует определять методом приближённого наибольшего правдоподобия. Суть метода заключается в выборе таких параметров распределения, при которых при данном законе распределения вероятность получить данную выборку наибольшая. Основным недостатком данного метода является необходимость точно знать аналитическое выражение закона распределения, что далеко не всегда возможно. Коэффициент вариации и коэффициент асимметрии для трёхпараметрического гамма-распределения Крицкого — Менкеля рассчитываются методом приближённого наибольшего правдоподобия в зависимости от статистик $\lambda_2$ и $\lambda_3$, вычисляемых по формулам

$$\lambda_2 = \frac{\sum_{i=1}^{N} \lg k_i}{N-1}, \tag{2.9}$$

$$\lambda_3 = \frac{\sum_{i=1}^{N} k_i \lg k_i}{N-1}. \tag{2.10}$$

Дальнейшие оценки могут проводиться различными способами, наиболее удобно рассчитывать коэффициенты вариации и асимметрии по специально разработанным номограммам. Принцип использования данных номограмм заключается в интерполяции значений $C_v$ и $C_s/C_v$ (рис. 2.2).

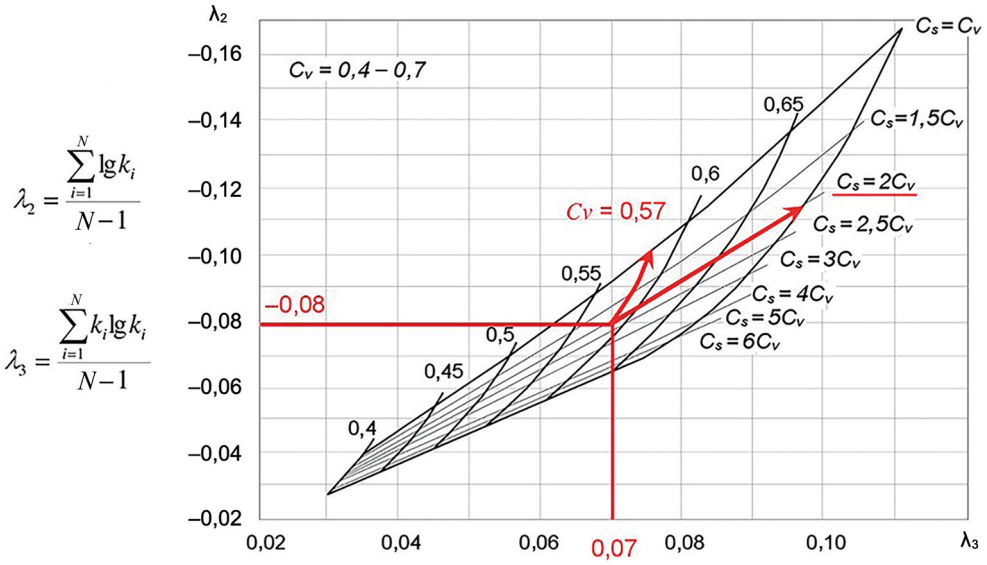



*Рис. 2.2.* Пример использования номограммы для определения параметров распределения Крицкого — Менкеля методом приближённого наибольшего правдоподобия [Сикан, 2007]

Полученные данным способом оценки параметров трёхпараметрического гамма-распределения являются состоятельными, эффективными и несмещёнными. В России номограммы для метода наибольшего правдоподобия разработаны применительно к распределению Крицкого — Менкеля. Однако на практике нередки ситуации, когда выборочные $\lambda_2$ и $\lambda_3$ приводят к выходу за пределы номограмм, т. е. решение отсутствует. В этом случае можно использовать сокращённый метод наибольшего правдоподобия. При реализации сокращённого метода статистика $\lambda_3$ не рассчитывается, а вместо неё используется районное соотношение $C_s/C_v$. Зная $\lambda_2$ и $C_s/C_v$, по номограмме можно рассчитать коэффициент вариации $C_v$.

Все без исключения перечисленные характеристики, в случае их определения по выборке из генеральной совокупности, имеют определённые погрешности, которые тем меньше, чем более репрезентативна выборка. Формулы для расчёта погрешностей зависят от метода оценки параметра и закона распределения исследуемой величины. При использовании методов моментов погрешности величины рассчитываются по следующим формулам.

*Абсолютная погрешность* математического ожидания определяется по формуле

$$\sigma_{m(x)} = \frac{\sigma}{\sqrt{N}}. \tag{2.11}$$

*Относительная погрешность* математического ожидания соответственно равна

$$\varepsilon_{m(x)} = \frac{\sigma}{\bar{x}\sqrt{N}} 100\,\% = \frac{C_v}{\sqrt{N}} 100\,\%. \tag{2.12}$$

Для расчёта абсолютной и относительной погрешностей коэффициента вариации используются формулы:

$$\sigma_{C_v} = \frac{C_v\sqrt{1 + aC_v^{\,2}}}{\sqrt{2N}}, \tag{2.13}$$

$$\varepsilon_{Cv} = \frac{\sqrt{1 + aC_v^{\,2}}}{\sqrt{2N}} 100\,\%, \tag{2.14}$$

где $a = 2$ — для нормального распределения; $a = 1$ — для двухпараметрического гамма-распределения. Напомним, что у двухпараметрического гамма-распределения $C_s/C_v = 2$, а у нормального распределения $C_s = 0$. Так как гидрологические ряды имеют, как правило, умеренную положительную асимметрию, эти формулы рекомендуется использовать при $a = 1$.

Погрешность коэффициента асимметрии определяется по формуле Крицкого — Менкеля:

$$\sigma_{C_s} = \frac{\sqrt{6\left(1 + 6C_v^{\,2} + 5C_v^{\,4}\right.}}{N} \tag{2.15}$$

или в относительных величинах:

$$\varepsilon_{Cs} = \frac{1}{\tilde{N}_s} \frac{\sqrt{6(1+6C_v^{\,2}+5C_v^{\,4}}}{N} 100\,\%. \qquad (2.16)$$

Формулы для расчёта погрешности коэффициента асимметрии (2.15) и (2.16) даны для $C_s = 2C_v$. Для распределений, у которых коэффициент асимметрии близок к нулю, эти формулы не применимы, так как множитель $1/C_s$ стремится к бесконечности. В этом случае относительная погрешность $C_s$ вообще не вычисляется.

При использовании метода наибольшего правдоподобия погрешность математического ожидания и коэффициента асимметрии рассчитываются по тем же формулам. Приближённая оценка *абсолютной* и *относительной погрешности коэффициента вариации* при использовании метода наибольшего правдоподобия определяется по формулам:

$$\sigma_{Cv} = \frac{C_v}{2N} \cdot \frac{\sqrt{3}}{3+C_v^{\,2}}, \qquad (2.17)$$

$$\varepsilon_{Cv} = \sqrt{\frac{3}{2N\left(3+C_v^{\,2}\right)}} 100\,\%. \qquad (2.18)$$

При использовании метода моментов погрешность тем больше, чем меньше ряд наблюдений и больше вариации. При определении коэффициента вариации методом наибольшего правдоподобия погрешность тем меньше, чем больше вариация.

Продолжительность периода наблюдений считают достаточной, если рассматриваемый период репрезентативен (представителен), а относительная средняя квадратическая погрешность расчётного значения исследуемой гидрологической характеристики не превышает 10 % для годовых и сезонных характеристик и 20 % — для экстремальных (максимальные и минимальные значения характеристик). Погрешность коэффициента асимметрии при имеющихся длинах ряда наблюдений, как правило, велика и может достигать 100 % и более, поэтому, как уже было сказано выше, на практике вместо выборочного значения коэффициента асимметрии рекомендуется использовать районное соотношение $C_s/C_v$. Имея районное соотношение $C_s/C_v$ и рассчитав выборочный коэффициент вариации, несложно получить и $C_s$ [Сикан, 2007].

## 2.4. Основные статистические критерии для оценки однородности и стационарности рядов наблюдений

Для оценки параметров распределения и вообще численных характеристик случайной величины необходимо убедиться в их неизменности во времени. Также иногда важно определить степень соответствия подобранного анали-

тического распределения эмпирическим данным или степень аномальности конкретного значения. Для этого используются специальные *статистические критерии*, которые часто называются статистическими тестами. В целом существует огромное множество статистических тестов, основанных на различных специальных распределениях, о которых будет сказано ниже. Суть, как правило, сводится к формулированию так называемой нулевой гипотезы, которая подвергается проверке и может быть отвергнута.

Нулевая гипотеза в подавляющем большинстве случаев утверждает о равенстве двух характеристик: либо соответствии данных выбранному закону распределения, либо независимости переменных друг от друга. Соответственно, в гидрологической практике параметры распределения не должны зависеть от времени, а их оценки за различные хронологические периоды должны быть примерно равны или, выражаясь языком математической статистики, разница между ними должна быть в пределах случайной погрешности и являться *статистически незначимой*. Как правило, существует формально определённая область принятия нулевой гипотезы, которая называется *доверительным интервалом*, а область опровержения нулевой гипотезы — критической областью, которая соответствует *уровню значимости*.

Совершенно очевидны и свойства данных характеристик: чем меньше доверительная область, тем статистический тест жёстче; в то же время возрастают уровень значимости и вероятность ложного отклонения нулевой гипотезы. Такая вероятность, а следовательно, и уровень значимости называются вероятностями совершения *ошибки первого рода.* Эта вероятность всегда заранее фиксируется определённым допустимым значением, которое зависит от целей анализа и количества доступной информации для анализа. *Ошибками второго рода* называют вероятность ложного принятия нулевой гипотезы; как правило, вероятность совершения ошибки второго рода убывает с увеличением количества анализируемых значений. Уровень значимости может быть односторонним или двухсторонним в зависимости от альтернативной гипотезы (рис. 2.3).

Все статистические тесты основываются на специальных распределениях, которым подчиняются производные той или иной характеристики.

Рассмотрим основные из них. Пусть имеется $N$ независимых случайных величин $X$, каждая из которых распределена по нормальному закону с нулевым средним значением и единичной диспер-

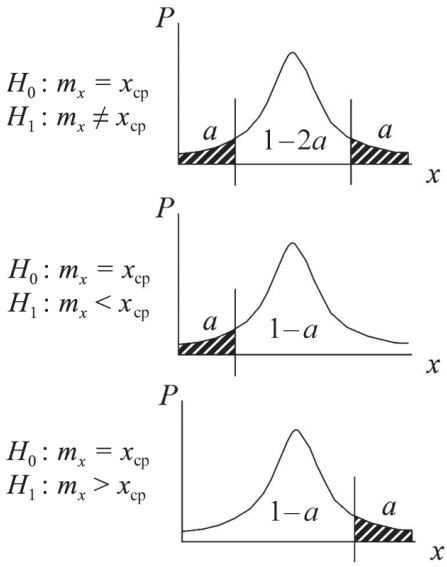



*Рис. 2.3.* Уровень значимости и доверительная область для различных альтернативных гипотез. Составлено по: [Сикан, 2007]

сией. Тогда распределением $\chi^2$ с $v$ степенями свободы называется распределение суммы квадратов независимых случайных величин: $\chi^2 = X_1^2 + X_2^2 + X_3^2 + X_n^2$. *Число степеней свободы* — это количество значений, функционально не связанных между собой, или, другими словами, число независимых параметров, в данном случае число степеней свободы численно равно $N$. При увеличении значения числа степеней свободы распределение медленно приближается к нормальному.

Пусть $Z$ и $E$ — независимые случайные величины, причём величина $Z$ является нормально распределённой с параметрами $M_Z = 0$; $D_Z = 1$, а $E$ — распределённой по закону $\chi^2$ с $v$ степенями свободы. Тогда случайная величина $t = Z / \sqrt{v / E}$ имеет распределение Стьюдента с $v$ степенями свободы. По мере увеличения числа степеней свободы распределение Стьюдента приближается к нормальному закону, причём скорость этого приближения выше, чем у распределения $\chi^2$. Из свойств распределения Стьюдента следует, что величина $\frac{\overline{X} - m_x}{\sigma / \sqrt{N}}$ имеет распределение Стьюдента.

Если $Z$ и $U$ — независимые случайные величины, обладающие распределением $\chi^2$ с $v_1$ и $v_2$ степенями свободы, то случайная величина $F = \frac{Z / v_1}{U / v_2}$ имеет *распределение Фишера* с $v_1$ и $v_2$ степенями свободы. Это распределение также называется $F$-распределением. Из свойств распределения Фишера следует, что отношение двух выборочных дисперсий будет подчиняться распределению Фишера.

Рассмотрим основные статистические тесты и критерии, применяемые в гидрометеорологии. Для проверки нулевой гипотезы о равенстве двух средних значений применяется критерий Стьюдента, основанный на одноимённом распределении. Пусть $X$ и $Y$ — выборки длиной $n$ и $m$ из нормальных распределений с неизвестными параметрами $m_x$, $\sigma_x$ и $m_y$, $\sigma_y$, но при этом известно, что $\sigma_x = \sigma_y$, т. е. они имеют одинаковые, хотя и неизвестные стандартные отклонения. Тогда эмпирическое значение статистики $t^*$ Стьюдента можно рассчитать по следующей формуле:

$$t^* = \frac{\overline{x} - \overline{y}}{\sigma_{\overline{x}\,\overline{y}}}, \tag{2.19}$$

где $\sigma_{xy}$ — стандартное отклонение разности средних $x$ и $y$.

В математической статистике доказано, что

$$\sigma_{\overline{x}-\overline{y}} = S\sqrt{\frac{m+n}{m \times n}}, \tag{2.20}$$

где $S$ — эмпирическая оценка $\sigma_{\overline{x}-\overline{y}}$.

Значение $S$ зависит от выборочных величин $D_x$ и $D_y$:

$$S = \sqrt{\frac{(n-1)D_X + (m-1)D_Y}{m+n-2}}. \tag{2.21}$$

В окончательном виде значение статистики Стьюдента рассчитывается по формуле

$$t^* = \frac{\overline{x} - \overline{y}}{S} \sqrt{\frac{n \times m}{n+m}}. \tag{2.22}$$

В практике гидрологических расчётов данный критерий применяется для оценки стационарности рядов по среднему значению. Весь ряд наблюдений разбивается на две части: длиной $n$ и $m$ соответственно. Эмпирическое значение статистики Стьюдента сравнивается с аналитическим значением, которое определяется по специально разработанным таблицам в зависимости от двухстороннего уровня значимости, равного 5 или 10 %, а также числа степеней свободы $\upsilon = n + m - 2$. В случаях, когда эмпирическое значение статистики Стьюдента по модулю больше теоретического, говорят, что нулевая гипотеза опровергается при заданном уровне значимости.

Для равенства двух дисперсий применяется критерий Фишера, основанный также на одноимённом распределении. Если $X$ и $Y$ — выборки из нормальных совокупностей с параметрами $m_x$, $D_x$ и $m_y$, $D_y$, то отношение их выборочных дисперсий $D_x/D_y$ подчиняется распределению Фишера с числом степеней свободы $\upsilon_1 = n - 1$ и $\upsilon_2 = m - 1$. Эмпирическое значение статистики Фишера рассчитывается по формуле

$$F^* = \frac{D_x}{D_y}, \tag{2.23}$$

где $D_x$ — большая из двух дисперсий.

Совершенно очевидно, что эмпирическое значение статистики Фишера принимает значения, равные или большие одного. В этом случае доверительная область при двухстороннем уровне значимости *2a* соответствует

$$1 \le \frac{D_x}{D_y} < F_{1-a}. \tag{2.24}$$

Теоретическое значение статистики Фишера определяется по специально разработанным таблицам в зависимости от уровня значимости и числа степеней свободы. В практике гидрологических расчётов данный критерий применяется для определения однородности гидрологического ряда по дисперсии, для чего аналогичным критерию Стьюдента образом весь ряд разбивается на две части. В случаях, когда эмпирическое значение статистики Фишера превышает теоретическое, говорят, что нулевая гипотеза о равенстве двух дисперсий опровергается при заданном уровне значимости. Так как критерий Стьюдента подразумевает равенство дисперсий, анализ следует начинать именно с определения равенства двух дисперсий по критерию Фишера. Критерии Стьюдента

и Фишера могут применяться также для так называемой интервальной оценки математического ожидания и дисперсии [Сикан, 2007].

Для оценки соответствия аналитической кривой обеспеченности эмпирическим данным необходимо использовать критерии согласия, основным из которых является критерий Пирсона, или критерий χ-квадрат. Критерий χ-квадрат был предложен в начале XX в. Карлом Пирсоном и в настоящее время является наиболее распространённым критерием согласия. Для его применения область допустимых значений (ОДЗ) исследуемой случайной величины $X$ разбивается на $k$ равнообеспеченных интервалов. При назначении границ интервалов наиболее удобно использовать следующую схему [Сикан, 2007]:

1) выбрать аналитическую кривую для аппроксимации закона распределения исследуемой случайной величины;
2) рассчитать по имеющейся выборке параметры распределения;
3) построить на клетчатке вероятности аналитическую кривую обеспеченностей;
4) разбить ось обеспеченностей на $k$ равных интервалов;
5) используя аналитическую кривую, определить границы интервалов для случайной величины $X$ в зависимости от границ интервалов на оси.

В качестве меры расхождения между эмпирическими данными и аналитической функцией распределения используется тестовая статистика:

$$\chi^2 = \frac{1}{m}\sum_{i=1}^{k}\left(m_i^*\right)^2 - N, \qquad (2.25)$$

где $m_i^*$ и $m$ — соответственно эмпирическое (фактическое) и теоретическое число случаев попадания значения случайной величины $X$ в $i$-й интервал, $m = n/k$.

Из выражения (2.25) видно, что чем выше значение статистики $\chi^2$, тем больше расхождение между эмпирической и аналитической кривыми. Поэтому при использовании критерия $\chi^2$ (Пирсона) назначают односторонний уровень значимости (обычно $a = 5\,\%$ или $a = 10\,\%$). Критерий $\chi^2$ может быть применён при выяснении вопроса о лучшем соответствии одной из нескольких аналитических кривых распределения одному и тому же эмпирическому ряду. При этом меньшее значение $\chi^2$ будет свидетельствовать о лучшем соответствии данной функции распределения эмпирическим данным.

Для оценки крайних членов выборки (максимальных и минимальных значений) соответствию нормальному распределению могут применяться критерии Смирнова — Граббса и Диксона. Данные критерии предназначены для проверки рядов наблюдений на выбросы — резко отличающиеся значения. По сути, определяется, относится ли конкретное значение к той же генеральной совокупности, что и остальные члены выборки, или нет. В первом случае отклонение можно объяснить тем, что данный гидрологический ряд содержит значение очень редкой повторяемости, во втором — ряд следует признать неоднородным. При использовании указанных критериев исходный ряд ранжируется в возрастающем порядке. Эмпирические значения статистики Смир-

нова — Граббса для максимального и минимального выборочного значения определяются по следующим формулам:

$$G^{*}_{\max} = \frac{x_{\max} - \overline{x}}{\sigma_x}, \tag{2.26}$$

$$G^{*}_{\min} = \frac{\overline{x} - x_{\min}}{\sigma_x}, \tag{2.27}$$

где $\sigma_x$ — выборочное стандартное отклонение, $x_{\max}$ и $x_{\min}$ — выборочные максимальные и минимальные значения соответственно.

Гипотеза об однородности рядов по критерию Смирнова — Граббса не опровергается, если:

$$G^{*} < G_a, \tag{2.28}$$

где $G_a$ — теоретическое значение статистики Смирнова — Граббса, определяемое в зависимости от уровня значимости коэффициентов автокорреляции и асимметрии, а также длины выборки (здесь, наверное, нужно говорить о присутствии в наблюдениях ошибки).

Более подробно ознакомиться со статистическими критериями можно в учебниках по статистике [Сикан, 2007; Малинин, 2008]. Применение данных и других критериев будет показано в разборе практического задания.

## 2.5. Основы и суть корреляционного и регрессионного анализов

Взаимосвязанности двух и более переменных определяются методом корреляционного и регрессионного анализов.

Парная корреляция позволяет оценить тесноту связи между двумя переменными. Для этого необходимо построить диаграмму рассеяния (рис. 2.4) и рассчитать коэффициент корреляции Пирсона. Коэффициент корреляции имеет диапазон значений от –1 до +1, и чем ближе по модулю коэффициент корреляции к единице, тем теснее линейная связь между ними. В случае отрицательного значения коэффициента корреляции говорят об обратной связи между переменными, в случае положительного — о прямой. Когда наблюдается обратная связь, одна переменная увеличивается при уменьшении другой, при прямой связи — обе переменные изменяются в одном направлении.

Связь между двумя переменными может быть линейной (рис. 2.4, *б* и *в*), тогда коэффициент линейной корреляции, или коэффициент корреляции Пирсона, рассчитывается по формуле

$$R_{XY} = \frac{\sum\left(X - \overline{X}\right)\left(Y - \overline{Y}\right)}{\sqrt{\sum\left(X - \overline{X}\right)^2 \sum\left(Y - \overline{Y}\right)^2}}. \tag{2.29}$$

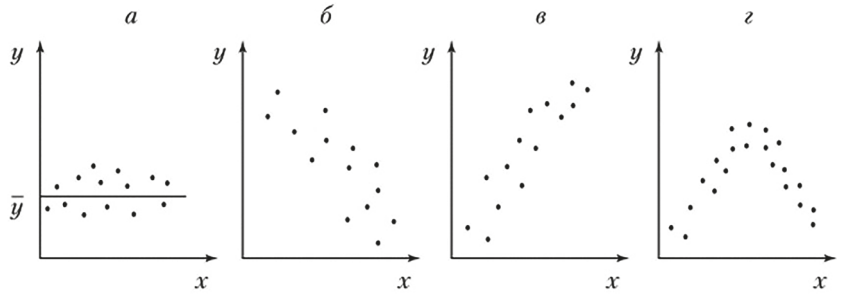



*Рис. 2.4.* Диаграммы рассеяния: *а* — корреляция отсутствует; *б* — корреляция линейная обратная; *в* — корреляция линейная прямая; *г* — корреляция нелинейная. Составлено по: [Сикан, 2007]

В некоторых случаях используется коэффициент детерминации, который в частном случае линейной зависимости представляет собой квадрат коэффициента корреляции. Коэффициент детерминации для модели с константой принимает значения от 0 до 1. Чем ближе значение коэффициента к единице, тем сильнее зависимость. При оценке регрессионных моделей это интерпретируется как соответствие модели данным. Для приемлемых моделей предполагается, что коэффициент детерминации должен быть не менее 0,50 (в этом случае коэффициент множественной корреляции превышает по модулю 0,70). Физический смысл коэффициента детерминации заключается в доле объяснённой дисперсии зависимой переменной. Оценка парной связи переменных называется корреляционным анализом и применяется для выявления наиболее информативных предикторов для уравнения регрессии. Для данных целей могут быть использованы пакет анализа программы *Excel* и соответствующая процедура «корреляционного анализа», где необходимо выбрать переменные для анализа, после чего будет построена корреляционная матрица, позволяющая выбрать наиболее информативные предикторы.

Оценив тесноту связи между двумя переменными, можно построить уравнение линейной регрессии. Уравнение линейной регрессии между двумя переменными имеет вид

$$Y = aX + C + \varepsilon, \tag{2.30}$$

где $Y$ — зависимая величина; $C$ — параметры регрессионной модели; $X$ — независимая величина; $\varepsilon$ — случайная ошибка регрессионной модели.

Коэффициенты $a$ и $C$ в уравнении линейной регрессии рассчитывают методом наименьших квадратов, суть которого сводится к минимизации суммы квадратов отклонений некоторой функции от искомой переменной. Коэффициент $a$ в уравнении линейной регрессии является угловым коэффициентом, или градиентом оценённой линии; он представляет собой величину, на которую $Y$ увеличивается в среднем, если мы увеличиваем $X$ на одну единицу, свободный член (пересечение) линии оценки — это значение $Y$, когда $X = 0$:

$$a = R\frac{\sigma_Y}{\sigma_X}, \tag{2.31}$$

$$C = \overline{Y} - a\overline{X}. \tag{2.32}$$

Выше рассмотрен случай парной линейной регрессии; в случае, когда больше одного предиктора, говорят о множественной регрессии. Уравнение множественной линейной регрессии имеет вид

$$Y = a_1X_1 + a_2X_2 + ... + a_nX_n + C. \tag{2.33}$$

Множественная регрессия предоставляет пользователю «соблазн» включить в качестве предикторов все возможные переменные для увеличения надёжности регрессионного уравнения. Однако увеличение числа членов регрессионного уравнения может привести к прямо противоположному результату. Таким образом, к независимым переменным, используемым в регрессионном уравнении, предъявляется ряд требований, а именно: их значимость — парные коэффициенты корреляции с зависимой переменной должны быть статистически значимы, отсутствие мультиколлинеарности — независимые переменные не должны коррелировать между собой. На практике максимальное число членов регрессионного уравнения зависит от длины рядов наблюдений, в практике гидрологических расчётов и прогнозов рекомендуется применять не более трёх членов регрессионного уравнения. При необходимости использования более трёх членов регрессионного уравнения и преодоления мультиколлинеарности уместна процедура факторного анализа (см. разд. 2.8).

Частным случаем корреляции является автокорреляция, представляющая собой статистическую взаимосвязь между последовательностями величин одного ряда, взятыми со сдвигом, например для случайного процесса — со сдвигом по времени. Коэффициенты автокорреляции имеют самостоятельное важное значение для моделей временны́х рядов авторегрессии проинтегрированного скользящего среднего (АРПСС).

Ограничения регрессионного анализа сводятся к аналитическому поиску лучших объясняющих переменных — необходимому условию стационарности рядов в настоящем и будущем, к сильному влиянию выбросов на коэффициенты регрессионного уравнения, а также к возможности так называемой ложной корреляции. Понятие ложной корреляции связано с ограничением всех методов регрессионного анализа, которое состоит в том, что они позволяют обнаружить только числовые зависимости, а не лежащие в их основе причинные связи, поэтому модели линейной регрессии называют моделями «чёрного ящика». Контроль регрессионных моделей заключается в использовании физически обоснованных предикторов, проверки моделей на независимом материале, а также проверки остатков модели на нормальность, стационарность и автокорреляцию. Наличие автокорреляции случайных ошибок регрессионной модели приводит к ухудшению качества оценок параметров регрессии, а также к завышению тестовых статистик, по которым проверяется

качество модели (т. е. создаётся искусственное улучшение качества модели относительно её действительного уровня точности).

Более подробно ознакомиться с процедурой корреляционного и регрессионного анализа можно в учебниках по статистике: [Сикан, 2007; Малинин, 2008].

## 2.6. Основы и суть машинного обучения, методы обучения искусственных нейронных сетей

Основным отличием методов обучения искусственных нейронных сетей (ИНС) от стандартных регрессионных моделей являются нелинейные преобразования внутри модели, а при прогнозировании временных рядов — использование и дополнительных предикторов, в отличие от моделей АРПСС. Также одним из преимуществ нейронных сетей является возможность сокращения числа предикторов непосредственно внутри модели. При использовании регрессионных моделей данная процедура выполняется посредством факторного анализа. Таким образом, ИНС являются достаточно гибким и универсальным инструментом при прогнозировании любых гидрологических величин.

Нейронная сеть представляет собой многослойную структуру из слоёв нейронов, причём каждый нейрон предыдущего слоя, как правило, связан с каждым нейроном последующего слоя. При отсутствии обратных связей нейронная сеть называется сетью прямого распространения. В самом простом виде это может быть один скрытый слой с определённым количеством скрытых нейронов, в котором происходит преобразование исходной информации посредством активационной функции и весовых коэффициентов, которые изменяются в процессе обучения с целью минимизации ошибки моделирования (в более сложном случае может быть несколько скрытых слоёв), выходного слоя предиктантов (рис. 2.5).

Нахождение оптимальных весовых коэффициентов и минимизация ошибки прогноза являются задачей, решаемой в процессе обучения искусственной нейронной сети. Активационные функции могут быть различны,

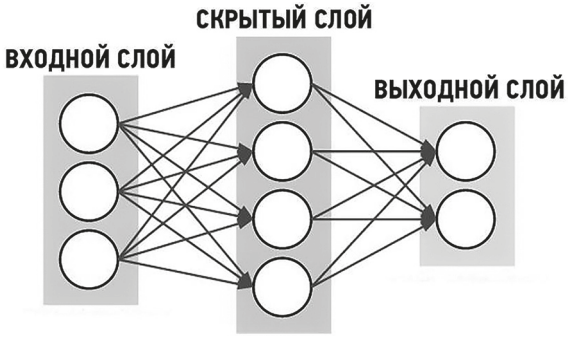



*Рис. 2.5.* Схематическое изображение искусственной нейронной сети

однако наиболее широко распространены линейная, гиперболический тангенс и *ReLu*. В скрытых слоях можно использовать любую из представленных функций, в выходном слое, для регрессионных задач, — линейное преобразование.

Определение начальных весов нейронов скрытого слоя называется процессом инициализации, а процесс их изменения — обучением. Инициализация нейронной сети должна выполняться таким образом, чтобы начальные веса для каждого нейрона были отличны, являлись случайным числом и были близки по своему значению к нулю. С точки зрения математики, обучение нейронных сетей — это многопараметрическая задача нелинейной оптимизации. Обучение нейронной сети происходит посредством оптимизации ошибки обучения, т. е. разницы между прогнозируемой величиной и фактически наблюдённым значением прогнозируемой величины.

В настоящее время существует большое количество методов оптимизации нейронных сетей. Суть методов оптимизации сводится к использованию относительно малых скоростей обучения (скорости изменения весов) и уменьшению этой скорости с приближением к локальному минимуму ошибки обучения. Уменьшение скорости обучения предназначено в первую очередь для невозможности преодоления минимума ошибки обучения, после чего ошибка обучения начинает возрастать.

Из наиболее распространённых методов оптимизации нейронных сетей можно назвать метод градиентного спуска и *Adam*, а также различные комбинации этих методов, например метод градиентного спуска с использованием момента Нестерова. Использование того или иного метода оптимизации ИНС в некоторых случаях способно значительно улучшить качество обучения, однако при решении регрессионных задач зачастую все методы дают примерно сопоставимые результаты. Критерием качества выпускаемых прогнозов при ИНС могут быть абсолютные ошибки прогноза, их процентное выражение или квадратические ошибки. В целях прогнозирования гидрологических величин уместно выбрать именно квадраты ошибок, потому что именно стандартная ошибка прогноза является критерием оценки качества выпускаемых прогнозов по методике Гидрометцентра и в критерии Нэша — Сатклиффа. К внешним параметрам модели, неизменяемым в ходе обучения, относятся количество скрытых слоёв и скрытых нейронов в них, выбранные метод инициализации и оптимизации, активационные функции для каждого слоя; к внутренним параметрам модели — весовые коэффициенты скрытых нейронов. Вышеописанные сети прямого распространения, имеющие один или несколько скрытых слоёв, называются многослойными персептронами, или *MLP*-сетями (*Multilayer perceptron*).

В настоящее время широко распространены методы глубокого обучения, или *deep learning*, что обусловило повышенный интерес к нейросетевому моделированию. *Deep learning* представляет собой развитие концепции обучения искусственных нейронных сетей с применением, в первую очередь, многослойной их структуры. Строго говоря, сеть, имеющую более одного скрытого слоя, можно

назвать глубокой. Зачастую глубокое обучение нейронных сетей связано с использованием и более замысловатой архитектуры нейронных сетей, например сетей с обратными связями, так называемых рекуррентных нейронных сетей. Рекуррентные нейронные сети применяются не только в задачах классификации и распознавании образов, но и в решении сложных регрессионных задач. Наиболее перспективными рекуррентными нейронными сетями для прогнозирования временных рядов являются искусственные нейронные сети с долгосрочной кратковременной памятью, так называемые *LSTM*-сети (англ. — *Long short-term memory*). Модель *Long short-term memory* является модификацией рекуррентной нейронной сети для глубокого обучения [Hochreiter, Schmidhuber, 1997] с так называемой долгой кратковременной памятью — дополнительными переменными состояния, сохраняющими веса отдельных нейронов и передающими их между расчётными шагами при обучении искусственной нейронной сети, если их значение приводит к снижению ошибки модели. *LSTM* сохраняет информацию о состояниях между расчётными шагами и на основании параметров регуляторов памяти определяет, когда и как долго сохранять эту информацию. Однако в отечественной гидрологии *LSTM*-сети не получили широкого распространения, поскольку требуются большой объём данных для их обучения и специализированные языки программирования.

Основным программным продуктом, позволяющим проводить обучение ИНС на уровне пользователя, является программный пакет *Statistica*. На уровне разработчика обучение нейронных сетей возможно на языке программирования *Python 3*, где доступно обучение в том числе и глубоких нейронных сетей. На языке *Python 3* написаны специализированные библиотеки для обучения искусственных нейронных сетей. Одной из таких открытых библиотек является *TensorFlow*, разработанная компанией *Google*. При работе с *TensorFlow* необходимо использовать библиотеку *Keras*, которая представляет собой надстройку над фреймворками *Deeplearning*, *TensorFlow* и *Theano*. Первым шагом при создании нейронных сетей на языке программирования *Python 3* является подключение к оперативной системе компьютера и импорт необходимых библиотек. После импорта данных необходимо провести стандартную процедуру разделения выборки на предикторы и предиктанты, а также на тестовую и обучающую подвыборки. Разделение на тестовую и обучающую подвыборки может осуществляться случайным образом, в данном случае требуется лишь указать объём тестовой выборки в процентах, альтернативным вариантом является использование каких-либо правил разделения. Также при необходимости может быть рекомендована выборка валидации. После определения размеров выборок нужно написать непосредственно код нейронной сети.

В настоящее время обучение нейронных сетей на языке программирования *Python 3* допустимо и в более автоматизированном варианте. К примеру, библиотека *Autokeras* позволяет использовать методику автоматической оптимизации всей архитектуры нейронной сети.

В программном пакете *Statistica 12* может быть реализован автоматический режим подбора архитектуры нейронной сети — необходимо лишь задать

диапазон количества скрытых нейронов для обучения. Нейронные сети, обучаемые в программном пакете *Statistica 12*, имеют всего один скрытый слой, однако автоматизированный режим позволяет подбирать архитектуру сети не уступающую, а в ряде случаев и превосходящую нейронные сети, обученные на языке программирования *Python 3*. В целом само по себе это неудивительно, так как алгоритмы обучения в *Statistica 12* являются достаточно надёжными, а задачи регрессии в большинстве случаев практики гидрологических прогнозов не требуют процессов глубокого обучения.

## 2.7. Кластерный анализ и задачи классификации

*Кластерный анализ* — многомерная статистическая процедура, упорядочивающая объекты в сравнительно однородные группы. Задача кластеризации относится к статистической обработке, а также к широкому классу задач обучения без учителя, т. е. количество, а главное — признаки кластеров, заранее неизвестны и определяются в процессе анализа. При решении задачи классификации, напротив, известно количество кластеров (классов объектов); целью же работы является нахождение алгоритма, по которому объект можно отнести к одному из известных классов. Задача классификации сводится к нахождению алгоритма отнесения объекта к одному из известных классов (кластеров). Зачастую задачи кластеризации и классификации могут решаться параллельно. Приведём такой пример: имеется некоторый набор гидрогеологических скважин, расположенных в различных широтах и обладающих различной глубиной. По всем этим скважинам есть полный набор гидрохимических показателей, по которому можно разделить скважину на несколько крупных групп — кластеров. Затем появляется или создаётся ещё какая-либо группа скважин, координаты и глубина которых известны, но нет гидрохимических показателей. Для отнесения этих скважин к одному из ранее определённых кластеров необходимо разработать алгоритм отнесения скважины к одному из кластеров лишь по данным координат и их глубины. Данный алгоритм может быть по-разному реализован, например методом обучения ИНС, но в сущности представляет собой логистическую регрессию. Таким образом, задача кластеризации — это так называемая задача обучения без учителя, когда правильные ответы заранее неизвестны, задача классификации — классическая задача «обучения с учителем».

Основные задачи, при которых необходимо использовать кластерный анализ:

- разработка типологии или классификации;
- исследование полезных концептуальных схем группирования объектов;
- порождение и проверка различных гипотез на основе исследования данных.

Первая задача решается, когда требуется научная классификация объектов на основании их признаков, например классификация рек в зависимости от множества признаков. Вторая задача — когда необходимо сокращение вы-

борки данных по каким-либо однородным группам, например гидрологическую информацию по большому количеству водных объектов можно сгруппировать и рассматривать характеристики кластеров, а не объектов. Для решения последней задачи у исследователя должна быть определённая гипотеза относительно исходных данных, например, используя методы кластерного анализа, можно проверить и уточнить существующие классификации.


*Рис. 2.6.* Наиболее распространённые метрики

Таким образом, методы кластерного анализа, с одной стороны, направлены на разработку, проверку и уточнение классификаций, с другой стороны — на сокращение объёмов исходных данных при работе с большими объёмами информации (*big data*). Для решения задачи кластеризации данных вводится понятие схожести отдельных наблюдений и кластеров. Под понятием схожести, в рамках кластерного анализа, имеется в виду близость отдельных наблюдений и кластеров, а в качестве меры близости принимается расстояние между объектами и кластерами. Множества, отнесённые к кластеру, не пересекаются, задача сводится к построению гиперплоскости, разделяющей кластеры. В настоящее время наиболее распространены следующие виды расстояний между объектами: евклидово расстояние, квадрат евклидова расстояния и расстояние городских кварталов (манхэттенское расстояние). На рис. 2.6 графически показано евклидово расстояние и расстояние Манхэттен от точки А до точки В, при этом можно заметить, что два этих расстояния для точек А — Г и А — Б совпадут.

Выбор метода определения расстояния зависит от исследователя; так, расстояние Манхэттен является линейным и менее подвержено влиянию выбросов, чем евклидово и квадрат евклидова расстояния. На первом шаге, когда каждый объект представляет собой отдельный кластер, расстояния между этими объектами определяются выбранной мерой. Однако, когда связываются вместе несколько объектов, необходимо правило объединения или связи для двух кластеров. Здесь имеются различные возможности: например, вы можете связать два кластера вместе, когда любые два объекта в двух кластерах ближе друг к другу, чем соответствующее расстояние связи. Другими словами, вы используете «метод ближайшего соседа» для определения расстояния между кластерами; этот метод называется методом одиночной связи. Это правило «волокнистых» кластеров, т. е. кластеров, «сцепленных вместе» только отдельными элементами, случайно оказавшимися ближе остальных друг к другу. Как альтернативу можно использовать «правило дальнего соседа», среднее невзвешенное или взвешенное расстояние, а также метод Варда, который отличается от всех других методов, поскольку в нём используются методы дисперсионного анализа для оценки расстояний между кластерами. В целом метод пред-

ставляется очень эффективным, если необходимо получить детальную кластеризацию с кластерами малого размера.

В настоящее время известно достаточно большое количество методов кластерного анализа, наибольшее распространение получили метод иерархического кластерного анализа и метод *k*-средних.

Алгоритм иерархического кластерного анализа выглядит следующим образом: на первом этапе каждый объект объявляется кластером, после чего два ближайших кластера объединяются в один, и т. д., пока вся выборка не объединится в один кластер; оптимальное число кластеров определяется по дендрограмме, где схематично изображаются кластеры и расстояния между ними (рис. 2.7).

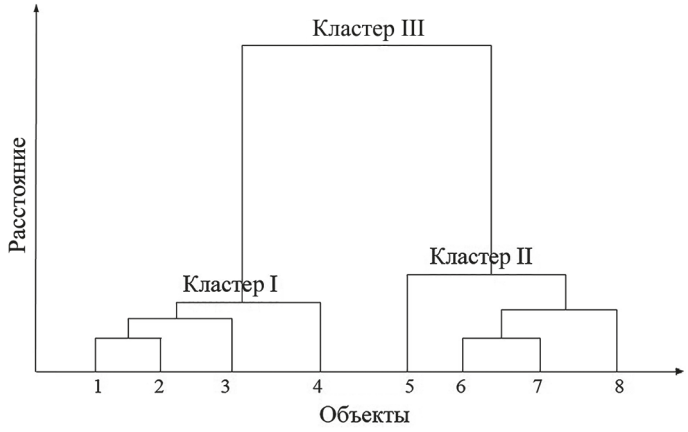



*Рис. 2.7.* Общий вид дендрограммы при использовании метода иерархического кластерного анализа. Составлено авторами

По виду дендрограммы можно судить о числе кластеров. Так, расстояние между объектами, входящими в первый и во второй кластеры, увеличивается постепенно. Это свидетельствует о том, что объекты действительно представляют собой два кластера данных, объединение этих двух кластеров в один третий кластер невозможно из-за значительного расстояния между ними. Таким образом, на данной дендрограмме фактически представлено два кластера. В целом иерархический кластерный анализ является универсальным, однако его использование затруднено при большом количестве кластеров, когда дендрограмма становится нечитаемой.

Альтернативным методом кластерного анализа является метод *k*-средних (*k-means*). Данный метод представляет собой алгоритм, минимизирующий суммарное квадратичное отклонение точек кластеров от центров этих кластеров. Основная идея заключается в том, что на каждой итерации перевычисляется центр масс для каждого кластера, полученного на предыдущем шаге, затем векторы разбиваются на кластеры вновь в соответствии с тем, какой из новых центров оказался ближе по выбранной метрике, при этом не

гарантируется достижение глобального минимума суммарного квадратичного отклонения, а только одного из локальных минимумов. Начальное положение центров кластеров может определяться либо случайным образом, либо из числа наблюдений, что является предпочтительнее, и т. д. При этом результат зависит от выбора исходных центров кластеров — их оптимальный выбор неизвестен. Число кластеров надо знать заранее. Несмотря на перечисленные недостатки, метод *k*-средних является на сегодняшний день наиболее популярным методом кластеризации данных.

Процедура кластерного анализа может быть одномерной — когда каждый столбец является отдельным объектом и целью кластерного анализа является группировка данных, также возможна многомерная (многофакторная) кластеризация — когда каждый столбец является характеристикой объекта. Например, для кластеризации качества воды может использоваться несколько химических показателей, которые при этом могут иметь разный порядок значений; в случаях, когда важно учесть вклад каждой характеристики вне зависимости от её абсолютного значения, обязательно проводится предварительная процедура стандартизации данных, чтобы среднее каждой из характеристик равнялось нулю, а стандартное отклонение — единице.

Таким образом, кластерный анализ является не столько обычным статистическим методом, сколько «набором» различных алгоритмов «распределения объектов по кластерам». Кластерный анализ определяет наиболее возможное значимое решение, поэтому проверка статистической значимости в действительности здесь неприменима. Процедура кластерного анализа сходится даже при отсутствии кластеров данных, поэтому основой проверки результатов является возможность интерпретации полученных кластеров.

В качестве примера рассмотрим многофакторную кластеризацию рек бассейнов Белого и Баренцева морей по ледовому режиму, проведённую в программном продукте *Statistica 12*. Исходными данными послужили 23 характеристики ледового режима (даты и уровни воды) по 34 гидрологическим пунктам; каждая строка представляет собой гидрологический пункт, каждый столбец — осреднённую характеристику ледового режима для конкретного пункта. Многомерная кластеризация данных требует обязательной стандартизации исходной информации. Стандартизация данных может проводиться непосредственно в программном пакете *Statistica*. На вкладке «Данные» необходимо выбрать процедуру стандартизации. После этого можно приступить к кластерному анализу: на вкладке «Статистик» (*statistics*) выбрать вкладку многомерного анализа (*mul/Exploratory*) — кластерный анализ (*cluster*) — иерархический кластерный анализ (*joining*). Затем на вкладке параметров иерархического кластерного анализа найти все характеристики выбранных объектов, а также способы определения расстояний между объектами и кластерами объектов.

Дендрограмма (рис. 2.8) читается сверху вниз. Из анализа дендрограммы следует, что реки и участки рек объединяются в некоторые группы, однако есть и группы, состоящие лишь из одного объекта. В целом определение

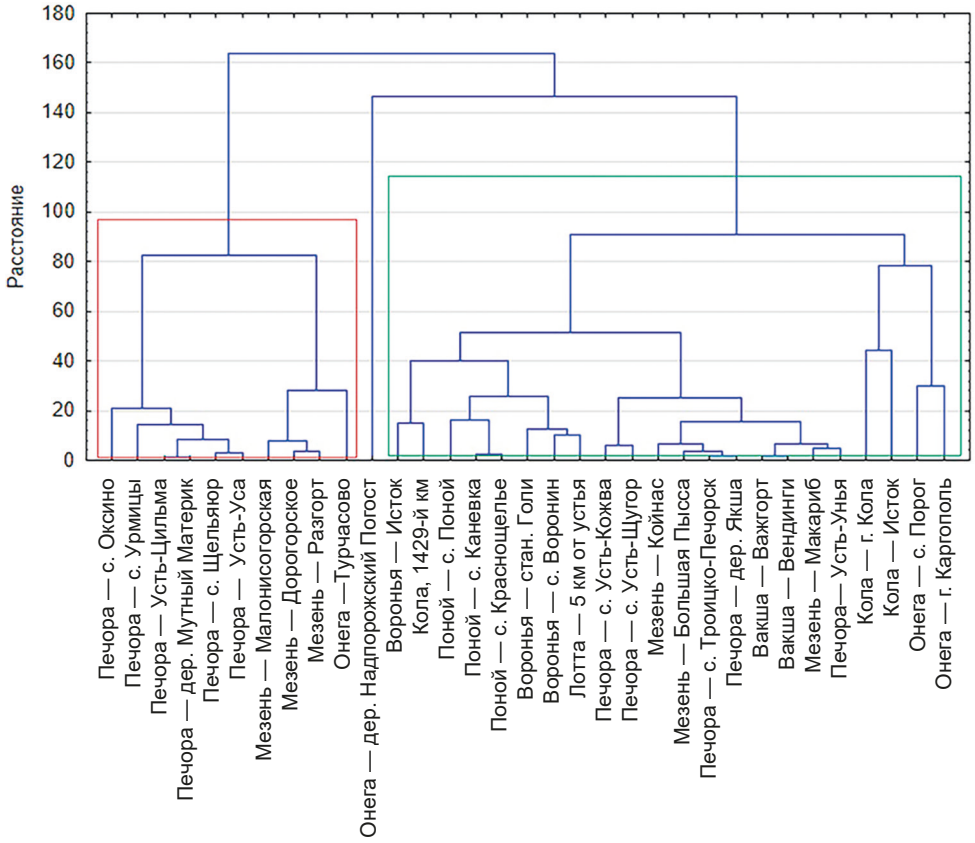



*Рис. 2.8.* Пример дендрограммы, построенной в программе *Statistica*, в качестве метода определения расстояния между объектами (использован квадрат евклидова расстояния); расстояние между кластерами определяется методом полной связи

кластеров по дендрограмме субъективно: на данной дендрограмме можно увидеть от трёх до шести кластеров данных. Наиболее правдоподобными представляются три кластера данных: первый кластер данных выделен на рисунке красным квадратом, второй — зелёным, последний кластер данных состоит из одного объекта — дер. Надпорожский Погост, при этом может быть и большее число кластеров, например кластер данных, выделенный красным цветом, может быть разделён на два. Таким образом, процедура анализа дендрограммы начинается сверху вниз, определение количества кластеров зависит от желаемой степени обобщённости. Далее можно воспользоваться процедурой кластерного анализа методом *k*-средних. Для этого необходимо задать два кластера данных, дер. Надпорожский Погост будет отнесена к одному из этих кластеров принудительно. Напомним, что в методе *k*-средних используется евклидово расстояние между объектами, поэтому данные кластеры могут несколько отличаться по составу. Тем не менее состав кластеров, полученный методом *k*-средних, не претерпел значительных изменений: было выделено два кластера данных, соответству-

ющих крупным незарегулированным и зарегулированным участкам рек. При проведении кластерного анализа особое внимание необходимо уделить оценке статистической значимости характеристик для отнесения объекта к определённому кластеру (незначимые характеристики можно удалить), а главное — возможности интерпретации полученных кластеров. В настоящее время программа *Statistica 12* автоматизированно определяет оптимальное число кластеров при использовании метода *k*-средних. Для этого необходимо перейти во вкладку «Добыча данных» (*data mining*), выбрать кластерный анализ, и на вкладке «Валидация» (*validation*) определить параметры кросс-валидации.

## 2.8. Факторный анализ и метод главных компонент

Рассмотренный в предыдущем разделе кластерный анализ предназначается прежде всего для разработки той или иной классификации данных. Однако при разработке моделей прогнозирования иногда необходимо не только объединение переменных в кластеры для сокращения числа предикторов, но и обеспечение независимости этих кластеров друг от друга для преодоления проблемы мультиколлинеарности. Такие независимые кластеры принято называть факторами, а процедуру их получения — факторным анализом. Таким образом, использование факторного анализа преследует несколько целей: сокращение числа переменных, построение обобщенных показателей (кластеризация данных), преодоление мультиколлинеарности предикторов и т. д.

Процедура факторного анализа сводится к объединению сильно коррелирующих между собой переменных в факторы, как следствие, происходит перераспределение дисперсии. Парная корреляция между фактором и исходными переменными называется факторной нагрузкой. Анализ факторной нагрузки помогает интерпретировать результаты факторного анализа.

Суть данного метода состоит в замене большого числа коррелированных между собой переменных меньшим числом некоррелированных факторов. Для выявления особенно значимых факторов наиболее оправдан метод главных компонент. Важной характеристикой метода является возможность ограничиться наиболее информативными главными компонентами и исключить остальные из анализа, что упрощает интерпретацию результатов. Достоинство метода главных компонент (МГК) также в том, что он единственный математически обоснованный метод факторного анализа.

Для применения факторного анализа в программе *Statistica 12* откроем исходные данные, затем на вкладке «Статистик» (*statistics*) выбираем вкладку «Многомерный анализ» (*mul/Exploratory*) — факторный анализ (*factor*). Выбрав все исходные данные, переходим на следующую вкладку и задаём максимальное число факторов, после чего вся процедура выполнится автоматически. Для завершения факторного анализа необходимо выбрать метод вра-

щения факторов, обеспечивающий интерпретацию полученных результатов на основе анализа факторных нагрузок, которые представляют собой коэффициент корреляции между фактором и исходной переменной. Анализируя, какими именно переменными нагружен тот или иной фактор, можно интерпретировать полученные результаты.

Распространённые методы вращений для факторного анализа:

1. *Метод варимакс*. Ортогональный метод вращения, минимизирующий число переменных с высокими нагрузками на каждый фактор. Этот метод упрощает интерпретацию факторов.

2. *Метод прямой облимин*. Метод косоугольного (неортогонального) вращения. Косоугольное решение соответствует разности, равной нулю (по умолчанию). По мере того как разность отклоняется в отрицательную сторону, факторы становятся более ортогональными.

3. *Метод квартимакс*. Метод вращения, который минимизирует число факторов, необходимых для объяснения каждой переменной. Этот метод упрощает интерпретацию наблюдённых переменных.

4. *Метод эквимакс*. Метод вращения, объединяющий методы варимакс, упрощающий интерпретацию факторов, и квартимакс, упрощающий интерпретацию наблюдённых переменных. Минимизируются число переменных с большими факторными нагрузками и число факторов, требуемых для объяснения переменной.

5. *Вращение типа промакс*. Косоугольное вращение в предположении, что факторы могут коррелировать между собой. Оно производится быстрее, чем вращение типа прямой облимин, поэтому оно полезно для больших наборов данных.

## 2.9. Введение в теорию случайных процессов и анализ временны́х рядов

Гидрологические расчёты или прогнозы предполагают знание о том, какая математическая модель используется для описания вероятностной структуры гидрологического ряда [Сикан, 2007]. Классическая математическая статистика и теория вероятностей, как правило, связаны со случайными величинами, основные свойства которых рассмотрены нами в предыдущих разделах. При этом отметим, что случайная величина принимается априори не зависящей от времени — время в данном случае не более чем формальный счётчик опытов. На практике, как правило, случайными величинами являются экстремальные ежегодно повторяющиеся значения расходов уровней воды и других гидрометеорологических характеристик, также к случайным величинам следует отнести ряды среднегодовых расходов. Теперь для понимания понятия случайного процесса перейдём от среднегодовых значений к среднемесячным.

При рассмотрении уровней или расходов воды месячной дискретности мы получим определённое количество гидрографов стока, каждый из ко-

торых является *реализацией* случайного процесса, а вся возможная их совокупность *случайным процессом*. Совершенно очевидно, что значения этого процесса за конкретный месяц могут быть различные, но и возможны вполне определённые значения: например, в мае наблюдается половодье и среднемесячные уровни воды за май являются высшими в году. Таким образом можно рассмотреть каждый из 12 мес. года и заключить, что значения случайного процесса за каждый отдельный месяц года являются случайной величиной. Соответственно, данный случайный процесс может быть разложен на 12 случайных величин, которые будут называться *сечением* случайного процесса. При этом данные случайные величины будут зависимы друг от друга (рис. 2.9).

Сформулируем ряд формальных определений:

- случайным процессом $X_{(t)}$ называется процесс, значение которого при любом фиксированном $t = t_i$ является случайной величиной $X(t_i)$;
- реализацией случайного процесса $X_{(t)}$ называется неслучайная функция $X_{(t)}$, в которую превращается случайный процесс $X_{(t)}$ в результате опыта;
- случайная величина $X(t_i)$, в которую обращается случайный процесс при $t = t_i$, называется сечением случайного процесса $X_{(t)}$, соответствующим данному значению аргумента.

Большинство гидрологических процессов являются процессами с непрерывными состояниями и непрерывным временем. Например, расход воды

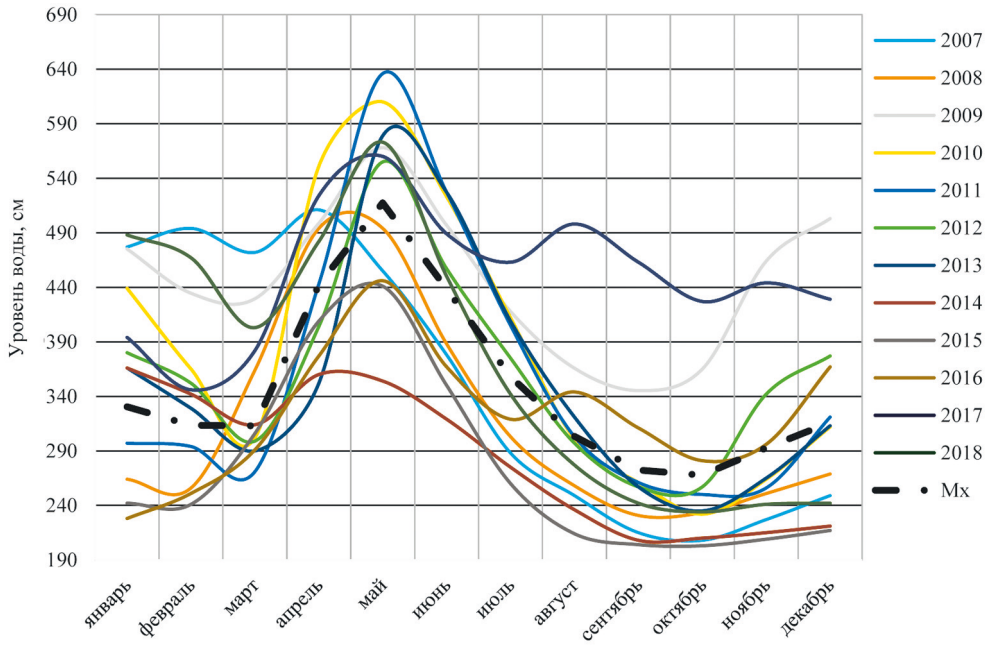



*Рис. 2.9.* Реализации случайного процесса и математическое ожидание ($m_x$) этого процесса

может изменяться в любой момент времени и принимать любые значения из некоторого интервала, границы которого зависят от размера реки и климатических особенностей региона. При этом надо учитывать, что на практике расходы воды осредняют за некоторый интервал времени (год, месяц, сутки и т. д.). Вводя шаг дискретности по времени, мы заменяем процесс с непрерывным временем на процесс с дискретным временем. Таким образом, проводя измерения гидрологических процессов, мы чаще всего используем модель случайного процесса с дискретным временем и непрерывными состояниями. Однако в гидрометеорологии используются и другие модели. Процесс количественного изменения облачности обычно представляют в виде процесса с дискретным временем и дискретными состояниями, так как наблюдения за облаками проводятся в фиксированные сроки, а их количество округляется до целых баллов (по 9-балльной шкале) [Сикан, 2007].

В общем случае исчерпывающей характеристикой случайного процесса является $n$-мерный закон распределения. Под $n$-мерной функцией распределения понимается совместное распределение $n$-го количества сечений случайного процесса. Однако определение даже двухмерных законов распределения является достаточно трудоёмкой задачей, вышеприведённый случайный процесс должен быть описан 12-мерным законом распределения. Поэтому на практике, как правило, вместо многомерных законов распределения используют основные характеристики случайных процессов, которые описывают случайный процесс частично [Сикан, 2007].

Так же, как и для случайной величины, для случайного процесса можно рассчитать основные моменты и статистические характеристики, только для случайного процесса искомые характеристики будут функциями.

Математическим ожиданием случайного процесса $X_{(t)}$ называется неслучайная функция $m_{x(t)}$, которая при любом значении аргумента $t$ равна математическому ожиданию соответствующего сечения случайного процесса. Таким образом, математическое ожидание случайного процесса представляет собой некоторую «среднюю» функцию (а в случае среднемесячных расходов воды средний за многолетний период — гидрограф), вокруг которой происходит разброс случайного процесса (рис. 2.9).

Дисперсией случайного процесса $X_{(t)}$ называется неслучайная функция $D_{x(t)}$, которая при любом значении аргумента $t$ равна дисперсии соответствующего сечения случайного процесса $X_{(t)}$. Соответственно, *стандартное отклонение* представляет собой квадратный корень из дисперсии.

Представленные выше характеристики случайного процесса не являются исчерпывающими, так как не учитывают взаимосвязь между сечениями. Такую связь характеризует корреляционная функция. *Корреляционной* (*или ковариационной*) *функцией* случайного процесса $X_{(t)}$ называется неслучайная функция $K_x(t, t')$, которая при каждой паре значений аргументов $t$ и $t'$ равна ковариации соответствующих сечений $x_{(t)}$ и $x_{(t')}$.

Помимо вышеперечисленного, случайный процесс может характеризоваться величиной периода (сезонности) и цикличности. Случайный процесс,

представленный в виде непрерывного ряда гидрографов, расположенных в хронологическом порядке, будет называться *временны́м рядом*. Особенности анализа и методы прогнозирования случайных процессов (временных рядов) разобраны в соответствующих работах, представленных в настоящем учебно-методическом пособии. Более подробную теоретическую информацию о случайных процессах и методах их моделирования можно найти в учебниках по статистике [Сикан, 2007; Рожков, 2001, 2002].

# 3. РАБОТА I. КОМПЛЕКСНЫЙ СТАТИСТИЧЕСКИЙ АНАЛИЗ ГИДРОМЕТЕОРОЛОГИЧЕСКОЙ ИНФОРМАЦИИ

В настоящей работе рассмотрен пример применения базовых методов математической статистики для комплексного описания случайной величины.

*Исходные данные* должны представлять собой случайную величину за период наблюдения не менее 50 лет. Они могут иметь различный генезис, как правило, это экстремальные или среднегодовые значения какой-либо гидрометеорологической характеристики. Для студентов-гидрологов это могут быть данные о максимальных и среднегодовых расходах воды либо высшие и среднегодовые уровни воды по конкретному гидрологическому посту. Для студентов других специальностей — температуры воздуха и океана, значения солёности и т. д. Исходные данные для анализа готовятся непосредственно студентами на основе ежегодных изданий Государственного водного кадастра и других возможных источников.

*Цель работы* — получение практических навыков в области статистического анализа гидрометеорологической информации. Для достижения поставленной цели студентам необходимо научиться применять встроенные функции и надстройки программного продукта *Excel* для расчёта основных статистических характеристик, а также правильно интерпретировать полученные результаты анализа.

*Задание.* По данным многолетних наблюдений построить хронологический график и гистограмму эмпирического распределения, выполнив их анализ. При анализе хронологического графика проанализировать характер ряда, наличие трендов и выбросов в ряде данных. При подозрении на наличие выбросов провести статистический тест на определение выбросов в ряде. Анализируя гистограмму, обращать внимание на её общий вид, асимметрию, количество мод и размах значений (для стандартизованной гистограммы).

Далее оценить стационарность и однородность рядов по критериям Фишера и Стьюдента, определить статистическую значимость линейного тренда. Расчёты выполнять при двухстороннем уровне значимости $2\alpha = 10\,\%$.

Затем рассчитать статистические характеристики (параметры распределений) рядов наблюдений и их статистические погрешности. К основным статистическим характеристикам относятся: математическое ожидание, оцениваемое через среднее арифметическое, медиана, мода, дисперсия, стандартное отклонение, коэффициенты вариации и асимметрии. Статистические погрешности рассчитываются для математического ожидания, коэффициентов вариации

и асимметрии. Расчёты выполняются методом моментов и методом приближённого максимального правдоподобия (для распределения Крицкого — Менкеля).

На заключительном этапе работы рассчитать эмпирические обеспеченности и по определённым ранее параметрам распределения построить аналитические кривые обеспеченности, провести статистическую оценку соответствия построенных аналитических кривых обеспеченности эмпирическим данным. Используя значение статистики $\chi^2$, выбрать наиболее подходящую кривую обеспеченности.

### *Порядок выполнения работы и отчётные материалы*

В качестве примера рассмотрим статистический анализ максимальных расходов воды за год по гидрологическому посту дер. Абрамково (р. Северная Двина). Исходные данные следует представить в виде таблицы с подписями характеристик и величин измерения (табл. 3.1).

Предварительный анализ данных таблицы выполняется визуально. В ходе анализа следует отметить анализируемую характеристику, период наблюдения, пропуски в данных, очевидно аномальные значения. В настоящем примере рассматриваются максимальные расходы воды Северной Двины по гидрологическому посту в дер. Абрамково за период наблюдений с 1950 по 2015 г. Пропуски и аномальные значения в данных отсутствуют, и можно приступить к построению хронологического графика. На хронологический график наносятся непосредственно значения исследуемой величины, а также в настройках диаграммы настраивается отображение линии тренда (рис. 3.1, пунктирная линия) и коэффициента детерминации линейного тренда.

Анализ графика показывает, что максимальные расходы воды в целом изменяются случайным образом, статистически значимый тренд, очевидно, отсутствует, в то же время с 1980 г. заметно уменьшается размах данных. Минимальное значение расхода воды составило 5860 м$^3$/с (1967 г.), максимальное —18 900 м$^3$/с (1974 г.). Отметим, что минимальное значение несколько больше, чем должно быть по отношению к общей вероятностной структуре ряда, что свидетельствует о том, что данное значение, возможно, является аномальным.

Построение гистограммы эмпирического распределения может выполняться по-разному. При ручном расчёте следует ранжировать исходный ряд, рассчитать количество интервалов и частоту попадания значений в заданный интервал. В программном продукте *Excel* возможно более автоматизированное построение гистограммы эмпирического распределения через пакет анализа данных, который входит в стандартные надстройки данного программного продукта. При построении гистограммы эмпирического распределения через пакет анализа данных следует выбрать соответствующий пункт меню и перейти к выбору данных. Для построения гистограммы достаточно показать входной (исходные данные для построения) и выходной (место вывода данных построения) интервалы, а также указать на необходимость вывода

*Таблица 3.1.* **Максимальные расходы воды р. Северная Двина по гидрологическому посту в дер. Абрамково за 1950–2015 гг.**

| Год | $Q_{max}$, м³/с | Год | $Q_{max}$, м³/с |
|---|---|---|---|
| 1950 | 8020 | 1983 | 9140 |
| 1951 | 7520 | 1984 | 9610 |
| 1952 | 15 000 | 1985 | 9710 |
| 1953 | 13 900 | 1986 | 15 500 |
| 1954 | 8290 | 1987 | 12 100 |
| 1955 | 15 500 | 1988 | 9530 |
| 1956 | 10 800 | 1989 | 8640 |
| 1957 | 18 400 | 1990 | 11 800 |
| 1958 | 13 100 | 1991 | 15 200 |
| 1959 | 10 600 | 1992 | 14 300 |
| 1960 | 8040 | 1993 | 16 700 |
| 1961 | 17 000 | 1994 | 14 500 |
| 1962 | 12 300 | 1995 | 17 500 |
| 1963 | 8190 | 1996 | 8540 |
| 1964 | 13 100 | 1997 | 13 100 |
| 1965 | 9710 | 1998 | 15 500 |
| 1966 | 11 400 | 1999 | 9930 |
| 1967 | 5860 | 2000 | 15 800 |
| 1968 | 15 300 | 2001 | 12 400 |
| 1969 | 11 400 | 2002 | 13 200 |
| 1970 | 8020 | 2003 | 11 300 |
| 1971 | 11 900 | 2004 | 12 000 |
| 1972 | 9610 | 2005 | 12 800 |
| 1973 | 7830 | 2006 | 11 700 |
| 1974 | 18 900 | 2007 | 8940 |
| 1975 | 8160 | 2008 | 11300 |
| 1976 | 14 800 | 2009 | 9300 |
| 1977 | 11 400 | 2010 | 11 000 |
| 1978 | 7710 | 2011 | 9660 |
| 1979 | 15 200 | 2012 | 18 000 |
| 1980 | 11 700 | 2013 | 10 400 |
| 1981 | 16 400 | 2014 | 10 100 |
| 1982 | 12 400 | 2015 | 9500 |

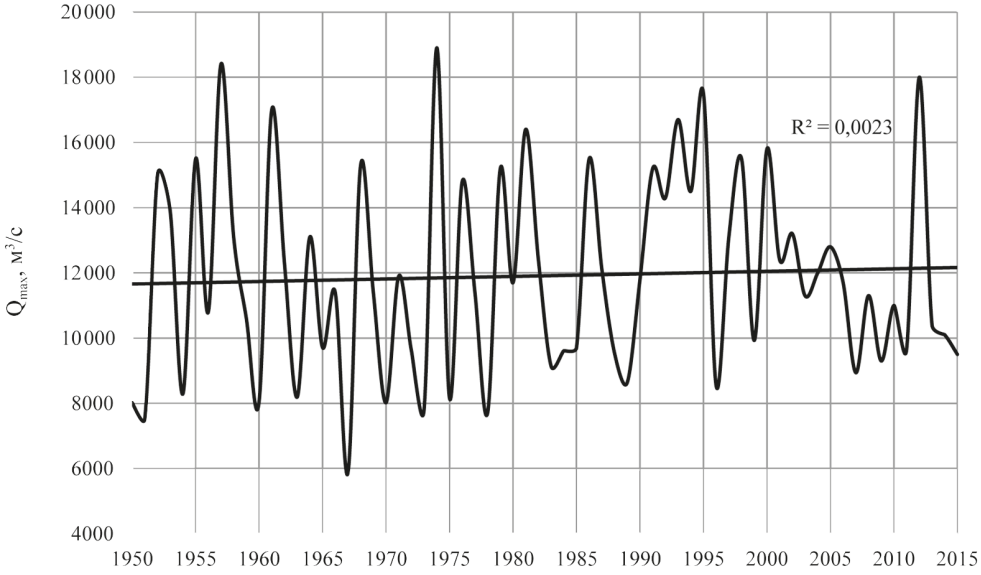



*Рис. 3.1.* Хронологический график максимальных расходов воды за 1950–2015 гг. по гидрологическому посту, расположенному в дер. Абрамково

графика, поставив галочку в соответствующем пункте. Помимо прочего, можно заблаговременно самостоятельно рассчитать желаемое число интервалов и интервал карманов и при построении гистограммы указать диапазон ячеек, в котором показаны граничные интервалы диапазонов (карманы). Можно поэкспериментировать с количеством интервалов для определения их оптимального числа. Построенная гистограмма, однако, не лишена недостатков и нуждается в некоторой «косметической» доработке. Во-первых, необходимо перевести частоту в доли единицы или проценты, во-вторых, по оси абсцисс желательно указать именно диапазон значений, а не верхнее значение диапазона, как это рассчитывается программой автоматически.

Количество интервалов рассчитывается по формуле

$$k = 4 \cdot \lg(n) = 4 \cdot \lg(66) \approx 7{,}3,$$

где $n$ — число членов ряда.

Размах данных, определяемый как разность максимального и минимального значений, составляет 13 040, что можно округлить до 13 000. Ширина одного диапазона равна отношению размаха данных к числу интервалов, в данном случае 1862, что можно округлить до 2000. Соответственно, имеем в общей сложности восемь диапазонов до 20 000 м$^3$/с. Заметим, что гистограмма эмпирического распределения аналогичным образом может быть рассчитана и для стандартизованной случайной величины (для которой среднее значение равно нулю, а стандартное отклонение — единице), в этом случае интервалы значений будут представлены величиной стандартного отклонения выборки, методика расчёта которого показана далее. Для приведения случайной величины к стан-

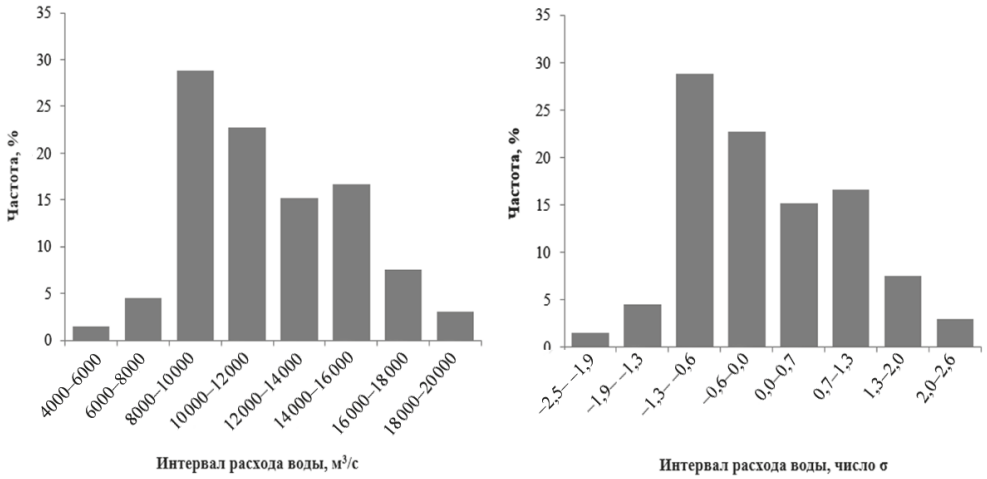



*Рис. 3.2.* Эмпирические гистограммы распределения максимальных расходов воды р. Северной Двины по гидрологическому посту, расположенному в дер. Абрамково

дартному виду необходимо от каждого её значения отнять математическое ожидание и эту разность поделить на стандартное отклонение. Предварительная стандартизация данных является важной, так как более точно позволяет определить соответствие эмпирической гистограммы нормальному закону распределения. В отчёте гистограммы представляются в следующем виде (рис. 3.2).

Анализ данных гистограмм позволяет сделать следующие выводы: распределение максимальных расходов воды близ дер. Абрамково является одномодальным (с модой в диапазоне от 8000 до 10 000 м$^3$/с), имеет умеренную положительную асимметрию, так как правый хвост распределения длиннее левого, размах распределения не превышает шести стандартных отклонений, что свидетельствует об отсутствии выбросов (ошибок измерений), значения менее 8000 м$^3$/м встречаются относительно редко (лишь 6 % значений меньше заданного числа). Таким образом, распределение случайной величины близко к распределению Пирсона III типа или Крицкого — Менкеля при соответствующих параметрах коэффициента асимметрии ($C_s$) и коэффициента вариации ($C_v$). В то же время существенных отклонений от нормального распределения, к которым относятся, например, двумодальность, наличие выбросов и т. д., не обнаружено. Это позволяет применять статистические критерии, обязательным условием для которых является близость распределения случайной величины к нормальному распределению.

Прежде чем определять параметры распределения и строить аналитические кривые обеспеченности, необходимо проверить ряд на стационарность и однородность. Под однородностью и стационарностью в данном случае подразумеваются постоянство математического ожидания и дисперсия, а также отсутствие статистически значимых трендов. В природе достаточно часто в связи с климатическими изменениями и антропогенной нагрузкой можно наблюдать нарушение однородности и стационарности рядов, в этом случае

расчёт параметров распределения по рядам с нарушенной однородностью и стационарностью заведомо является грубой ошибкой. Выявив неоднородность рядов, необходимо проанализировать её причины и при необходимости привести ряд к однородному виду.

Приведение рядов к однородному виду может выполняться различными способами в зависимости от целей анализа и природы нарушения однородности и стационарности. Одной из распространённых причин нарушения однородности рядов является постройка гидротехнических сооружений, которые значительным образом трансформируют гидрологический режим (например, могут быть существенно снижены максимальные значения расходов воды).

Когда совершенно очевидно, что река уже не может быть приведена к своему прошлому состоянию, целесообразно исключить из анализа предшествующую часть наблюдений либо вычислить разность средних значений и привести весь ряд к новому среднему значению путём добавления разности средних значений к каждому значению случайной величины за период, предшествующий постройке гидротехнического сооружения. Подобная процедура также может выполняться для уровней воды при изменении отметки нуля поста (прибавлением разности отметок нуля поста). При неоднородности и по дисперсии, и по среднему значению для предшествующей нарушению однородности части ряда необходимо провести процедуру стандартизации, после чего обратным пересчётом привести её к новому среднему значению и дисперсии. В случае статистически значимого тренда целесообразен анализ разностей с единичным лагом (анализ приращений). Все перечисленные процедуры нужно выполнять с большой осторожностью при полной уверенности в своих действиях.

Выполним проверку случайной величины на однородность и стационарность по критериям Фишера (прил. 1) и Стьюдента (прил. 2), оценим значимость линейного тренда при двухстороннем уровне значимости $2\alpha = 10\,\%$.

Для оценки однородности по дисперсии применим критерий Фишера. Так как предположительная дата нарушения однородности неизвестна, разделим ряд на две равные части: с 1950 по 1982 г. и с 1983 по 2015 г., количество значений в рядах $n = m = 33$. Строго говоря, разделение рядов может быть выполнено не пополам, а с учётом гипотетической даты нарушения однородности. Также рассчитаем значение *p-value* (*P-значение*), которое представляет собой минимальный уровень значимости, при котором опровергается нулевая гипотеза, которая в данном случае формулируется следующим образом: различие дисперсий двух выборок является статистически незначимым. Для проверки данной гипотезы рассчитаем дисперсии двух выборок, используя встроенную функцию *Excel* ДИСП.В.

Дисперсия выборки с 1950 по 1982 г.:

$$\text{ДИСП.В} = D_{\text{несм}} = \frac{\sum_{i=1}^{N}(x_i - \overline{x})^2}{n-1} = 12\,084\,605\ \text{м}^3/\text{с}^2.$$

Дисперсия выборки с 1983 по 2015 г.:

$$\text{ДИСП.В} = D_{\text{несм}} = \frac{\sum_{i=1}^{n}(y_i - \overline{y})^2}{m-1} = 7\ 554\ 409\ \text{м}^3/\text{с}^2.$$

Эмпирическое значение статистики Фишера рассчитывается как отношение двух дисперсий, причём в числителе всегда указывается большая из дисперсий:

$$F^* = \frac{12084605}{7554409} = 1{,}59.$$

Теоретическое значение статистики Фишера $F$ находится по специальным таблицам, которые непосредственно связаны с распределением Фишера (прил. 1). Значение $F$ зависит от принятого уровня значимости и числа степеней свободы двух выборок. В нашем случае уровень значимости 10 %, а число степеней свободы $v$ $(n(m)-1)$ равно 32 для обеих частей выборки. По таблице (прил. 1) $F \approx 1{,}8$. Так как $F^* < F$, нулевая гипотеза об однородности ряда по дисперсии (равенстве дисперсий) не опровергается при заданном уровне значимости.

Определим минимальный уровень значимости, при котором нулевая гипотеза отклоняется. Для этого в программном продукте *Excel* рассчитаем *p-value*, используя встроенную функцию F.ТЕСТ, для которой необходимо задать лишь два диапазона данных:

$$\text{F.ТЕСТ} = 0{,}189 \approx 19\,\%.$$

Таким образом, минимальный уровень значимости, при котором мы вынуждены отклонить нулевую гипотезу, в два раза больше принятого, что свидетельствует о возможности принятия нулевой гипотезы.

После проверки ряда на однородность по критерию Фишера определяем однородность по критерию Стьюдента о равенстве средних значений двух половин ряда. Ряд разбивается аналогичным образом. Выполним необходимые расчёты вручную с использованием пакета встроенного анализа. Эмпирическое значение статистики Стьюдента рассчитываем по формуле

$$t^* = \frac{\overline{x}_n - \overline{y}_m}{\sqrt{\dfrac{(n-1)D_n + (m-1)D_m}{m+n-2}}}\sqrt{\frac{n \times m}{n+m}} = -0{,}44,$$

где $\overline{x}_n$ и $\overline{y}_m$ — средние значения по первой и второй части выборки; $n$ — длина первой части выборки; $m$ — длина второй части выборки; $D_n$ и $D_m$ — дисперсии по первой и второй части выборки.

Теоретическое значение статистики Стьюдента определяется по соответствующим таблицам для двухстороннего уровня значимости 10 % и числа степеней свободы $v$, равного 66 $(m+n-2)$ (прил. 2) Значение $t$, рассчитанное по таблицам (прил. 2), $t \approx 1{,}7$. При этом гипотеза об однородности рядов по критерию Стьюдента

не опровергается, если $|t^*| < t$. Так как $|t^*| < t$, нулевая гипотеза об однородности ряда по математическому ожиданию (равенстве средних значений) не опровергается при заданном уровне значимости. Определим минимальный уровень значимости, при котором нулевая гипотеза отклоняется. Для этого в программном продукте *Excel* рассчитаем *p-value*, используя встроенную функцию СТЬЮДЕНТ.ТЕСТ, для которой необходимо задать два диапазона данных: количество хвостов распределения (двухсторонний или односторонний уровень значимости) и виды выполняемого *t*-теста (для одинаковых и разных дисперсий). Параметр «хвосты» является обязательным. Если значение «хвосты» равно 1, функция СТЬЮДЕНТ.ТЕСТ возвращает одностороннее распределение, если равно 2, — возвращает двухстороннее распределение. Вид выполняемого *t*-теста также является обязательным параметром (2 — тест выполняется для выборки с равными дисперсиями, 3 — для выборки с различными дисперсиями). Так как критерий Фишера показал, что дисперсии различаются статистически незначимо, можно применять двухсторонний *t*-тест для выборок с равными дисперсиями:

$$\text{СТЬЮДЕНТ.ТЕСТ} = 0{,}66 = 66\,\%.$$

Таким образом, минимальный уровень значимости, при котором нулевая гипотеза опровергается, почти в 7 раз больше, чем принятый, свидетельствует о хорошей стационарности ряда по среднему значению.

Аналогичные результаты можно получить и при использовании пакета анализа: в меню анализа выбирается *t*-тест с одинаковыми или различными значениями дисперсии (в данном случае с одинаковыми). Результаты представляются в виде таблицы, в которой приводятся все статистические оценки, включая уровень *p-value* (табл. 3.2).

Оценить статистическую значимость линейного тренда необходимо для того, чтобы убедиться в отсутствии детерминированного тренда к увеличению или уменьшению значений случайной величины. Уровень значимости применяется тот же, что и в предыдущих тестах. Определять статистическую значимость линейного тренда удобнее всего сравнением коэффициента корреляции с его критическим значением, при котором тренд принимается статистически значимым. Для заданного уровня значимости рассчитывается критическое значение коэффициента корреляции $R_{крит}$ по следующей формуле [Малинин, 2008]:

$$R_{крит} \approx \frac{2}{\sqrt{N+2}} \approx 0{,}24.$$

Коэффициент детерминации $R^2$ был определён ранее и выведен на хронологическом графике. Коэффициент корреляции равен квадратному корню из коэффициента детерминации и в данном случае составляет 0,05. Так как коэффициент корреляции меньше критического значения, статистически значимые тренды отсутствуют (как и было указано при визуальном анализе хронологического графика).

Таким образом, выполненный тест свидетельствует о том, что ряд максимальных расходов воды близ дер. Абрамково однороден и стационарен.

*Таблица 3.2.* **Результаты статистического теста Стьюдента**

| Характеристика | Переменная 1 | Переменная 2 |
|---|---|---|
| Среднее | 11 741 | 12 081 |
| Дисперсия | 12 084 604 | 7 554 409 |
| Наблюдения | 33 | 33 |
| Объединённая дисперсия | 9 819 507 | |
| Гипотетическая разность средних | 0 | |
| *Df* | 64 | |
| *t*-статистика | –0,44 | |
| $P(T<=t)$ одностороннее (*p-value*) | 0,33 | |
| *t*-критическое одностороннее | 1,29 | |
| $P(T<=t)$ двухстороннее (*p-value*) | 0,66 | |
| *t*-критическое двухстороннее | 1,66 | |

Рассчитаем статистические параметры распределения случайной величины. Параметры распределения могут рассчитываться методом приближённого наибольшего правдоподобия для распределения Крицкого — Менкеля (прил. 3) и иногда для распределения Пирсона III типа (прил. 4), а также методом моментов для любого из распределений, существенным образом не отклоняющихся от нормального. Удачность выбора параметров распределения и самого распределения может оцениваться визуально или формально по критерию χ-квадрата (прил. 5). Проще всего статистические параметры рассчитывать методом моментов, потому что он является общепринятым и даёт достаточно точные оценки статистических характеристик. Но применять этот метод без введения поправок допустимо при коэффициенте вариации менее 0,6 и коэффициенте асимметрии менее 1. При превышении данных значений рекомендуется введение соответствующих поправок согласно Своду правил по проектированию и строительству СП 33-101-2003. В программном продукте *Excel* математическое ожидание, дисперсию, стандартное отклонение и коэффициент асимметрии можно рассчитать с помощью соответствующих функций:

$$\text{СРЗНАЧ} = m_x = \frac{\sum_{i=1}^{N} x_i}{N} = 11\,900 \text{ м}^3/\text{с}^2,$$

$$\text{ДИСП.В} = D_{\text{несм}} = \frac{\sum_{i=1}^{N} \left(x_i - \overline{x}\right)^2}{N-1} = 970\,000 \text{ м}^3/\text{с}^2,$$

$$\text{СТАНДОТКЛОН.В} = \sigma = 3110 \text{ м}^3/\text{с},$$

$$C_v = \frac{\sigma}{\overline{x}} = 0,26,$$

$$\text{СКОС} = C_s = \frac{N\sum_{i=1}^{N}\left(x_i - \overline{x}\right)^3}{\sigma^3 (N-1)(N-2)} = 0,38,$$

$$C_s / C_v = 1,46.$$

Так как расход воды измеряется с точностью до трёх значащих цифр, результаты всех вычислений в отчёте целесообразно записывать с аналогичной точностью. Значение коэффициентов асимметрии и вариации не превышают допустимых для метода моментов, поэтому их уточнение можно не проводить. Далее покажем алгоритм вычисления параметров распределения методом приближённого наибольшего правдоподобия (табл. 3.3).

*Таблица 3.3.* **Фрагмент вспомогательной таблицы для оценки параметров распределения методом наибольшего правдоподобия**

| **Год** | $Q$, м$^3$/с | $k_i$ ($Q_i$ / $Q_{ср}$) | lg$k$ | $k \cdot$lg$k$ |
|---|---|---|---|---|
| 1950 | 8020 | 0,67 | –0,17 | –0,12 |
| 1951 | 7520 | 0,63 | –0,20 | –0,13 |
| 1952 | 15 000 | 1,26 | 0,10 | 0,13 |
| 1953 | 13 900 | 1,17 | 0,07 | 0,08 |
| 1954 | 8290 | 0,70 | –0,16 | –0,11 |
| 1955 | 15 500 | 1,30 | 0,11 | 0,15 |
| …… | …… | …… | …… | …… |
| 2010 | 11 000 | 0,92 | –0,03 | –0,03 |
| 2011 | 9660 | 0,81 | –0,09 | –0,07 |
| 2012 | 18 000 | 1,51 | 0,18 | 0,27 |
| 2013 | 10 400 | 0,87 | –0,06 | –0,05 |
| 2014 | 10 100 | 0,85 | –0,07 | –0,06 |
| 2015 | 9500 | 0,80 | –0,10 | –0,08 |
| Сумма | – | – | –0,97 | 0,96 |

Метод приближённого наибольшего правдоподобия рекомендован в качестве основного при расчёте параметров распределения Крицкого — Менкеля (прил. 6) с учётом Свода правил по проектированию и строительству СП 33-101-2003. Согласно действующим рекомендациям, математическое ожидание определяется методом моментов. Для оценки коэффициента вариации и отношения коэффициента асимметрии к коэффициенту вариации рассчитаем статистики $\lambda_2$ и $\lambda_3$.

Для определения данных статистики целесообразно выполнить промежуточные расчёты, при этом случайная величина переводится в модульные коэффициенты $k_i$.

После промежуточных расчётов статистики $\lambda_2$ и $\lambda_3$ равны:

$$\lambda_2 = \frac{\sum_{i=1}^{N} \lg k_i}{n-1} = \frac{-0,97}{65} = -0,15,$$

$$\lambda_3 = \frac{\sum_{i=1}^{N} k_i \lg k_i}{n-1} = \frac{0,97}{65} = 0,15.$$

По номограмме (прил. 6) в зависимости от $\lambda_2$ и $\lambda_3$ находим параметры распределения:

$$C_v = 0,26,$$

$$C_s/C_v = 2,31,$$

$$C_s = 0,60.$$

Кроме этого, в программном продукте *Excel* можно определить медиану при помощи функции МЕДИАНА, либо КВАРТИЛЬ.ВКЛ. Применяя последнюю функцию, можно рассчитать и значения остальных квартилей. Округлённая до трёх значащих цифр, медиана ряда будет равна 11 600. В случае положительной асимметрии медиана всегда меньше среднего значения.

Рассчитав основные статистические характеристики, необходимо выполнить оценку точности. При использовании метода моментов оценка точности определяется по приведённым ниже формулам.

Относительная погрешность оценки математического ожидания:

$$\varepsilon_{m(x)} = \frac{\sigma}{\bar{x}\sqrt{N}} 100\,\% = \frac{3110}{11900\sqrt{66}} 100\,\% = 3,2\,\%,$$

$$\varepsilon_{C_v} = \frac{\sqrt{1 + aC_v^{\,2}}}{\sqrt{2N}} \times 100\,\% = \frac{\sqrt{1 + 1 \times 0,26^2}}{\sqrt{2 \times 66}} \times 100\,\% = 8,9\,\%.$$

Погрешность коэффициента асимметрии может быть рассчитана по формуле Крицкого — Менкеля:

$$\varepsilon_{Cs} = \frac{1}{C_S}\sqrt{\frac{6\left(1 + 6C_v^{\,2} + 5C_v^{\,4}\right)}{N}} 100\,\% = \frac{1}{0,38}\sqrt{\frac{6\left(1 + 6 \cdot 0,26^2 + 5 \cdot 0,26^4\right)}{66}} 100\,\% = 94\,\%.$$

Погрешность коэффициента асимметрии при имеющихся длинах ряда, как правило, велика и может достигать 100 % и более, поэтому, как уже было сказано выше, на практике вместо выборочного значения коэффициента асимметрии рекомендуется использовать районное соотношение $C_s/C_v$. Имея последнее и рассчитав выборочную оценку коэффициента вариации, несложно получить и значение $C_s$.

При использовании метода наибольшего правдоподобия величина погрешности определения математического ожидания и коэффициент асимметрии определяются по тем же формулам, что и для метода моментов. Приближённая величина относительной погрешности коэффициента вариации при использовании метода наибольшего правдоподобия равна

$$\varepsilon_{Cv} = \frac{\sqrt{3}}{\sqrt{2N\left(3 + C_v^{\,2}\right)}} 100\,\% = 8{,}6\,\%.$$

В гидрологической практике принято считать, что расчёт параметров распределения выполнен надёжно, если погрешность среднего значения не превышает 10 %, а коэффициента вариации 15 %. Допустимая ошибка коэффициента асимметрии не регламентирована.

Рассчитав необходимые параметры распределений, можно приступить к оценке ординат эмпирической кривой обеспеченности и подбору аналитического закона распределения случайной величины. Для расчёта эмпирической обеспеченности исходный ряд ранжируется в убывающем порядке. Обеспеченность ординат рассчитывается по формуле Крицкого — Менкеля:

$$p = \frac{m}{N+1} 100\,\%,$$

где $m$ — порядковый номер расхода ранжированного ряда; $N$ — количество лет наблюдений. Значения обеспеченности записываются в процентах и показывают вероятность превышения случайной величиной заданного числа. Значения ординат и эмпирической обеспеченности приведены в табл. 3.4.

После определения эмпирических ординат и их обеспеченности рассчитывают ординаты аналитических кривых распределения. В настоящей работе были рассчитаны параметры распределения, необходимые для построения стандартных кривых обеспеченности Пирсона III типа и Крицкого — Менкеля. Для кривой Крицкого — Менкеля были определены параметры как методом моментов, так и методом приближённого наибольшего правдоподобия. Поэтому необходимо выбирать аналитическую кривую, которая бы наилучшим образом аппроксимировала ординаты эмпирической кривой обеспеченности. Для этого нужно поэтапно построить все три кривые и методами математической статистики выбрать лучшую из трёх.

Значения ординат аналитической кривой обеспеченности Пирсона III типа определяются по следующей формуле:

$$Q_{p\%} = x_{p\%}\sigma_x + m_x,$$

где $x_{p\%}$ — табличные значения нормированных ординат кривой обеспеченности Пирсона III типа; $\sigma_x$ — стандартное отклонение случайной величины; $m_x$ — математическое ожидание случайной величины.

Табличные значения нормированных ординат кривой обеспеченности Пирсона III типа определяются в зависимости от коэффициента асимметрии $C_s$ и значения обеспеченности $p$ (прил. 4).

Ординаты кривой обеспеченности Крицкого — Менкеля оцениваются по специализированным таблицам для модульных коэффициентов, которые

*Таблица 3.4.* **Фрагмент таблицы расчёта ординат эмпирической кривой обеспеченности**

| Год | Номер расхода воды в ранжированном ряде | $Q$, м$^3$/с | $p$, % |
|---|---|---|---|
| 1974 | 1 | 18 900 | 1,5 |
| 1957 | 2 | 18 400 | 3,0 |
| 2012 | 3 | 18 000 | 4,5 |
| 1995 | 4 | 17 500 | 6,0 |
| 1961 | 5 | 17 000 | 7,5 |
| 1993 | 6 | 16 700 | 9,0 |
| 1981 | 7 | 16 400 | 10,4 |
| 2000 | 8 | 15 800 | 11,9 |
| 1955 | 9 | 15 500 | 13,4 |
| …… | …… | …… | …… |
| 1954 | 57 | 8290 | 85,1 |
| 1963 | 58 | 8190 | 86,6 |
| 1975 | 59 | 8160 | 88,1 |
| 1960 | 60 | 8040 | 89,6 |
| 1950 | 61 | 8020 | 91,0 |
| 1970 | 62 | 8020 | 92,5 |
| 1973 | 63 | 7830 | 94,0 |
| 1978 | 64 | 7710 | 95,5 |
| 1951 | 65 | 7520 | 97,0 |
| 1967 | 66 | 5860 | 98,5 |

зависят от соотношения коэффициента асимметрии к коэффициенту вариации. В настоящей работе коэффициент вариации равен 0,26, соотношение $C_s/C_v$ может быть округлено при определении параметров распределения методом моментов до 1,5 и до 2,5 — методом приближённого наибольшего правдоподобия.

Расчёты без округления параметров достаточно трудоёмки, но могут быть выполнены в специализированных программных комплексах, например в *HydroStat*.

Значения ординат аналитической кривой обеспеченности Крицкого — Менкеля определяются по формуле

$$Q_{p\%} = x_{p\%} + m_x,$$

где $x_{p\%}$ — табличные значения ординат кривой обеспеченности Крицкого — Менкеля в модульных коэффициентах; $m_x$ — математическое ожидание случайной величины.

Значения ординат всех кривых приведены в табл. 3.5.

*Таблица 3.5.* **Ординаты кривых обеспеченности (КО) Пирсона III типа и Крицкого — Менкеля при различных значениях $C_s/C_v$**

| $p$, % | Нормированные ординаты КО Пирсона III типа | Ординаты КО Пирсона III типа, м$^3$/с | Нормированные ординаты КО Крицкого — Менкеля ($C_s/C_v = 1{,}5$) | Ординаты КО Крицкого — Менкеля ($C_s/C_v = 1{,}5$) | Нормированные ординаты КО Крицкого — Менкеля ($C_s/C_v = 2{,}5$) | Ординаты КО Крицкого — Менкеля ($C_s/C_v = 2{,}5$) |
|---|---|---|---|---|---|---|
| 0,01 | 4,6 | 26 200 | 2,18 | 26 000 | 2,40 | 28 600 |
| 0,1 | 3,6 | 23 200 | 1,96 | 23 300 | 2,07 | 24 700 |
| 1 | 2,6 | 20 000 | 1,68 | 20 000 | 1,74 | 20 700 |
| 5 | 1,7 | 17 300 | 1,46 | 17 400 | 1,47 | 17 500 |
| 10 | 1,3 | 16 000 | 1,34 | 16 000 | 1,34 | 16 000 |
| 20 | 0,8 | 14 500 | 1,22 | 14 500 | 1,20 | 14 300 |
| 25 | 0,6 | 13 900 | 1,17 | 13 900 | 1,16 | 13 800 |
| 30 | 0,5 | 13 400 | 1,12 | 13 400 | 1,12 | 13 300 |
| 40 | 0,2 | 12 500 | 1,05 | 12 500 | 1,04 | 12 400 |
| 50 | –0,1 | 11 700 | 0,98 | 11 700 | 0,97 | 11 600 |
| 60 | –0,3 | 11 000 | 0,92 | 10 900 | 0,91 | 10 800 |
| 70 | –0,6 | 10 100 | 0,85 | 10 100 | 0,85 | 10 100 |
| 75 | –0,7 | 9710 | 0,81 | 9700 | 0,81 | 9690 |
| 80 | –0,9 | 9260 | 0,77 | 9230 | 0,78 | 9260 |
| 90 | –1,2 | 8070 | 0,68 | 8070 | 0,69 | 8220 |
| 95 | –1,5 | 7160 | 0,60 | 7180 | 0,62 | 7440 |
| 98 | –1,8 | 6170 | 0,53 | 6260 | 0,56 | 6640 |
| 99 | –2,1 | 5500 | 0,48 | 5670 | 0,52 | 6140 |
| 99,9 | –2,6 | 3940 | 0,36 | 4260 | 0,41 | 4940 |

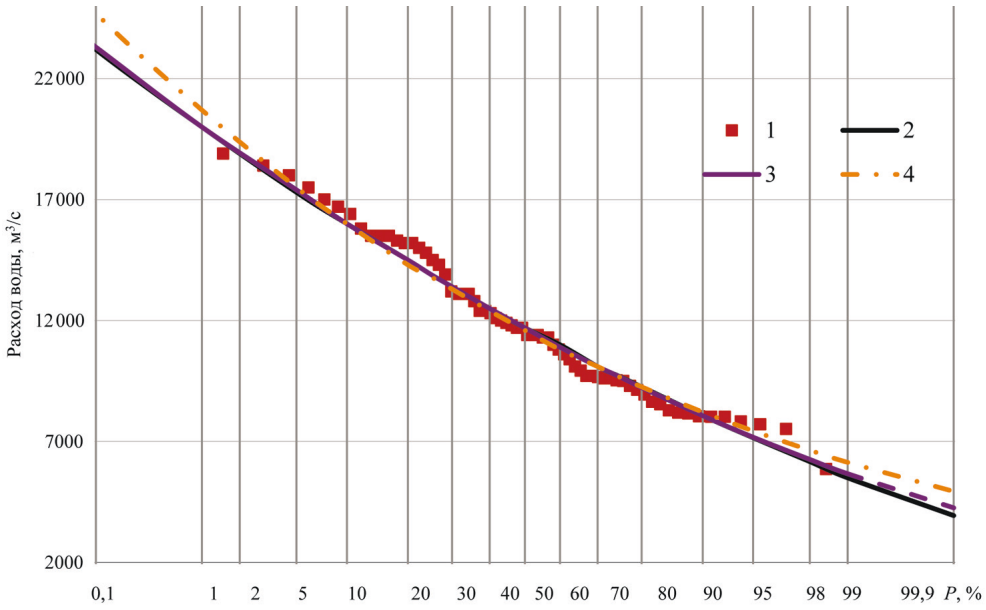



*Рис. 3.3.* Эмпирическая (1) и аналитические кривые распределения расходов воды Пирсона III типа (2), Крицкого — Менкеля при $C_s/C_v = 1{,}5$ (3) и $C_s/C_v = 2{,}5$ (4)

После расчёта ординат всех кривых обеспеченностей при различных параметрах все значения эмпирической и аналитической кривых обеспеченности наносятся на один график. Идеальным является нанесение всех кривых обеспеченности на специальную клетчатку вероятностей, спрямляющую нормальный закон распределения. В отсутствие клетчатки вероятности допустимо представить результаты на одном графике, построенном в *Excel*.

Соответствие той или иной кривой эмпирическим данным можно оценить визуально либо с помощью статистических критериев, в частности критерия $\chi^2$ Пирсона (табл. 3.6).

Для расчёта статистики $\chi^2$ нужно разделить ряд на 10 равнообеспеченных интервалов $k$, теоретическое число случаев попадания в каждый интервал будет равно 6,6 (66/10). Числовые границы интервалов могут быть установлены по таблице ординат обеспеченности. После этого следует подсчитать число попаданий ранжированной случайной величины в каждый интервал и, возведя полученную статистику в квадрат, просуммировать. В настоящей работе необходимо определить лучшее соответствие эмпирических данных одной из трёх кривых обеспеченностей (рис. 3.4).

Рассчитав все необходимые параметры для критерия $\chi^2$, можно приступить непосредственно к его определению (табл. 3.7, 3.8).

Для кривой обеспеченности Пирсона III типа значение статистики $\chi^2$ будет наименьшим:

$$\chi^2 = \frac{1}{m}\sum_{i=1}^{k}(m_i^*)^2 - n = \frac{1}{6{,}6}486 - 66 = 7{,}63.$$

*Таблица 3.6.* **Вспомогательная таблица расчёта критерия $\chi^2$ для кривой Пирсона III типа при $C_s = 0{,}38$; $C_v = 0{,}26$**

| $p$, % | Границы интервалов | Число случаев попадания расхода воды в интервал $m$ | $m^2$ |
|---|---|---|---|
| 0–10 | Беск. — 16 000 | 7 | 49 |
| 10–20 | 16 000–14 500 | 10 | 100 |
| 20–30 | 14 500–13 400 | 3 | 9 |
| 30–40 | 13 400–12 500 | 4 | 16 |
| 40–50 | 12 500–11 700 | 9 | 81 |
| 50–60 | 11 700–11 000 | 6 | 36 |
| 60–70 | 11 000–10 100 | 4 | 16 |
| 70–80 | 10 100–9260 | 9 | 81 |
| 80–90 | 9260–8070 | 7 | 49 |
| 90–100 | 8070 — отр. число | 7 | 49 |
| Сумма | — | 66 | 486 |

*Таблица 3.7.* **Вспомогательная таблица расчёта критерия $\chi^2$ для кривой Крицкого — Менкеля при $C_s/C_v = 1{,}5$; $C_v = 0{,}26$**

| $p$, % | Границы интервалов | Число случаев попадания расхода воды в интервал $m$ | $m^2$ |
|---|---|---|---|
| 0–10 | Беск. — 16 000 | 7 | 49 |
| 10–20 | 16 000–14 500 | 10 | 100 |
| 20–30 | 14 500–13 400 | 3 | 9 |
| 30–40 | 13 400–12 500 | 4 | 16 |
| 40–50 | 12 500–11 700 | 9 | 81 |
| 50–60 | 11 700–10 900 | 7 | 49 |
| 60–70 | 10 900–10 100 | 3 | 9 |
| 70–80 | 10 100–9230 | 9 | 81 |
| 80–90 | 9230–8070 | 8 | 64 |
| 90–100 | 8070–0 | 6 | 36 |
| Сумма | — | 66 | 494 |

Для кривой Крицкого — Менкеля эти значения составят 8,84 и 10,3 для $C_s/C_v = 1{,}5$ и $C_s/C_v = 2{,}5$ соответственно.

Критическое значение данной статистики зависит от количества параметров распределения и числа интервалов.

По прил. 5 находим теоретическое значение $\chi^2$ при числе степеней свободы $v = k - r - 1 = 10 - 3 - 1 = 6$ и уровне значимости $\alpha = 5\,\%$. В данном случае $\chi^2 = 12{,}6$.

Таким образом, нулевая гипотеза о соответствии эмпирических данных аналитическому закону распределения при заданном уровне значимости не опровергается для всех аналитических кривых. Тем не менее лучшее соотношение по критерию $\chi^2$ показала кривая Пирсона III типа (рис. 3.4).

*Таблица 3.8.* **Вспомогательная таблица расчёта критерия $\chi^2$ для кривой Крицкого — Менкеля при $C_s/C_v = 2{,}5$; $C_v = 0{,}26$**

| $p$, % | Границ интервалов | Число случаев попадания расхода воды в интервал $m$ | $m^2$ |
|---|---|---|---|
| 0–10 | Беск. — 16 000 | 7 | 49 |
| 10–20 | 16 000–14 300 | 11 | 121 |
| 20–30 | 14 300–13 300 | 2 | 4 |
| 30–40 | 13 300–12 400 | 6 | 36 |
| 40–50 | 12 400–11 600 | 7 | 49 |
| 50–60 | 11 600–10 800 | 7 | 49 |
| 60–70 | 10 800–10 100 | 3 | 9 |
| 70–80 | 10 100–9260 | 9 | 81 |
| 80–90 | 9260–8220 | 5 | 25 |
| 90–100 | 8220–0 | 9 | 81 |
| Сумма | — | 66 | 504 |

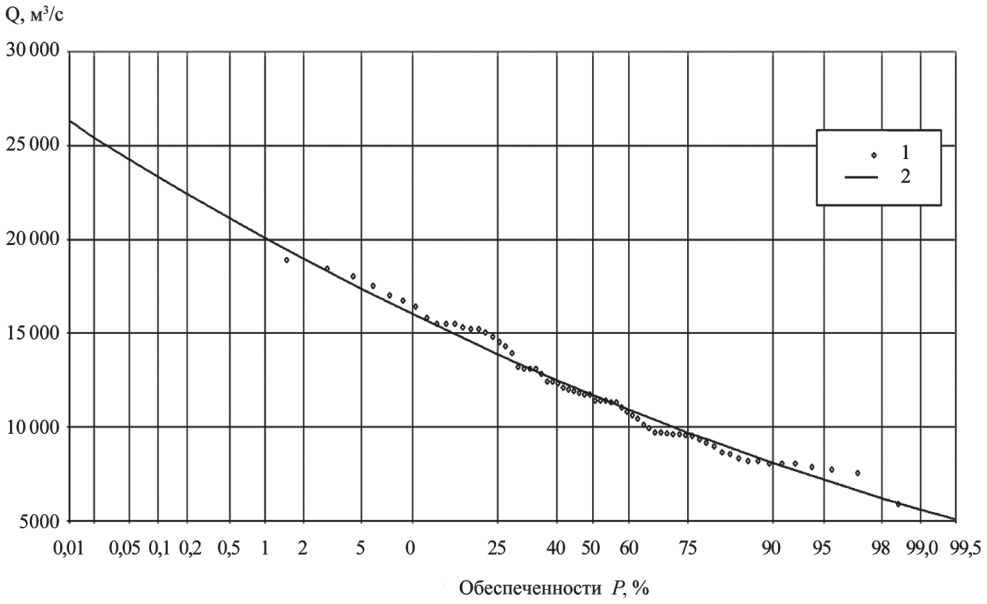



*Рис. 3.4.* Кривая обеспеченности Пирсона III типа (1) максимальных расходов воды р. Северная Двина (2) по гидрологическому посту, расположенному в дер. Абрамково, $C_v = 0{,}25$, $C_s = 0{,}38$, (кривая обеспеченности построена в программе *Hydrostat)*

Заключение по работе должно содержать основные выводы и полученные результаты. В настоящем примере проанализированы максимальные значения расходов воды по гидрологическому посту в дер. Абрамково за 1950–2015 гг. Установлено, что настоящий ряд является статистически однородным и стационарным при уровне значимости $2\alpha = 10\,\%$ как по дисперсии, так и по среднему значению, статистически значимые тренды и выбросы отсутствуют. Выборочное распределение отклоняется от нормального и является умеренно асимметричным. Параметры распределения оценивались двумя способами: методом моментов и методом наибольшего правдоподобия, статистические погрешности не превышают допустимых. Рассчитаны ординаты эмпирической и аналитической кривых обеспеченности, наибольшее соответствие эмпирическим данным показала кривая Пирсона III типа с параметрами распределения $C_v = 0{,}25$ и $C_s = 0{,}38$.

Частным случаем при выполнении данной работы может быть выявление нарушения стационарности рядов наблюдений по среднему значению или (и) дисперсии. Нарушение стационарности рядов наблюдений может быть вызвано различными причинами, в частности влиянием изменений климата или генетической неоднородностью, когда максимальные и средние значения расходов или уровней воды имеют различное происхождение. Например, высшие уровни воды могут формироваться в отдельные годы на чистой воде, в другие же годы имеют место подпорные явления, значительно увеличивающие величину уровней. В таких случаях принято пользоваться усечёнными или составными кривыми обеспеченности. В обоих случаях суть методов заключается в определении параметров распределения и построении кривых обеспеченностей лишь для отдельных частей рядов, с последующим пересчётом величины обеспеченности.

Рассмотрим пример построения таких кривых обеспеченностей на примере высших уровней р. Кундрючья по пос. Владимирская. Для начала построим и проанализируем хронологический график (рис. 3.5).

Анализ данного графика и результаты статистических тестов показали нарушение стационарности данного ряда как по среднему значению, так и по дисперсии (при разделении ряда на два периода: 1960–1990 гг. и 1991–2019 гг.). Отметим, что нарушение стационарности ряда, по сути, обусловлено отсутствием во второй его части высоких значений, которые периодически появлялись до 1990 г., минимальные же значения не претерпели значительных изменений. Это можно рассматривать по-разному: с одной стороны, можно говорить о необратимости этих изменений, например в связи со строительством какого-либо гидротехнического сооружения, с другой — о некоторой цикличности, что приведёт к увеличению уровней в будущем. В первом случае необходимо было бы привести ряд к однородному виду, во втором случае (или если невозможно установление однозначной необратимости этих изменений) применяются усечённые или составные кривые обеспеченности.

Первым этапом является построение эмпирической кривой обеспеченности (рис. 3.6).

На графике виден достаточно резкий перелом эмпирической кривой в районе 40 % обеспеченности, что можно принять за условную границу

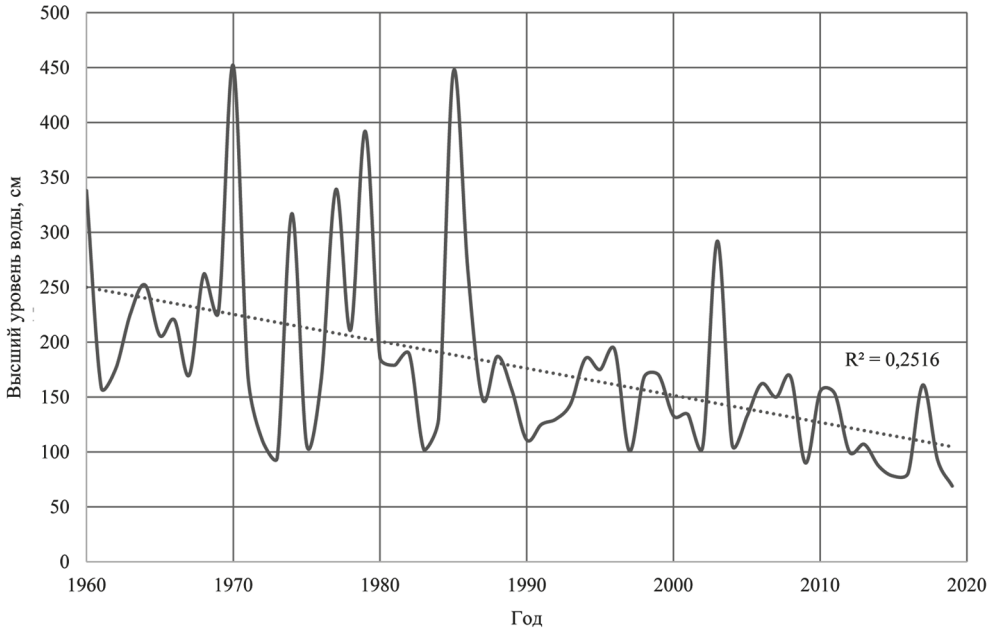



*Рис. 3.5.* Хронологический график высших уровней воды р. Кундрючья за 1960–2020 гг. по гидрологическому посту, расположенному в пос. Владимирская, с нанесённой линией тренда

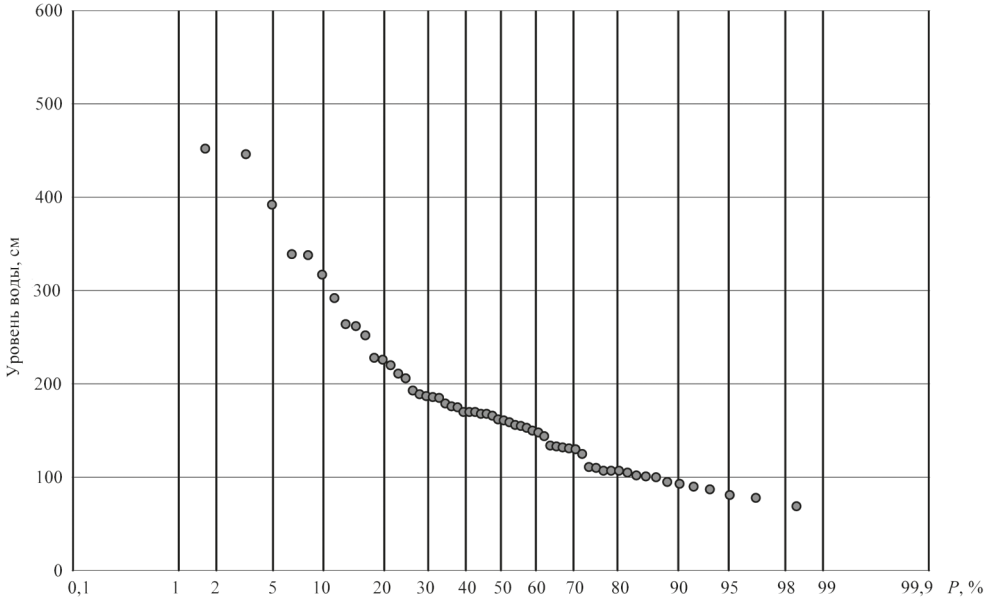



*Рис. 3.6.* Эмпирическая кривая обеспеченности высших уровней воды р. Кундрючья по пос. Владимирская

разделения данного ряда. Отметим, что в данном случае можно построить и обычную кривую обеспеченности при достаточно хорошем соответствии эмпирических данных аналитической кривой обеспеченности в верхней её части. Для построения кривой обеспеченности Пирсона III типа (рис. 3.7) использованы значения математического ожидания (177), СКО (86) и коэффициента асимметрии (1,5).

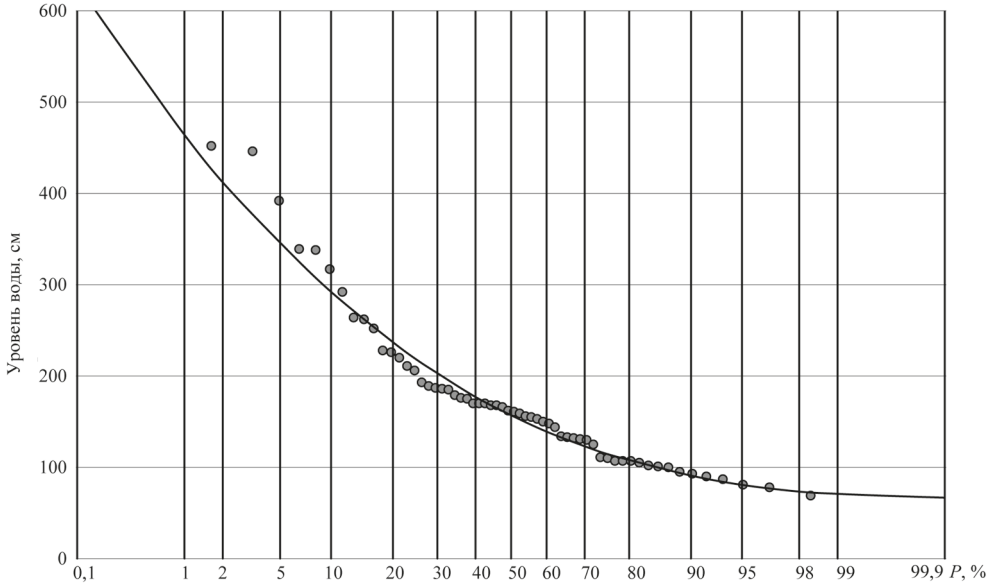



*Рис. 3.7.* Эмпирическая (кружочки) и аналитическая (сплошная линия) кривые обеспеченности Пирсона III типа для высших уровней воды р. Кундрючья по пос. Владимирская

Кривая обеспеченности Пирсона III типа плохо описывает эмпирические точки в верхней части, что свидетельствует о целесообразности построения усечённой или составной кривой обеспеченности.

Построим усечённую кривую обеспеченности для верхней части, приняв за условную границу отнесения к верхней части уровень воды, равный 170 см. Таким образом, в расчёте участвуют все данные более 170 см, исключаются 34 значения. Далее по данной части выборки следует рассчитать все параметры распределения (в примере использовано распределение Пирсона III типа) и, построив кривую обеспеченности, снять все значения уровней воды для опорных обеспеченностей. Пересчёт значения обеспеченности для исходной выборки выполняется по формуле

$$P_0 = \frac{P_1\left(N - m\right)}{N},$$

где $P_0$ — значение обеспеченности в исходном ряде длиной $N$; $P_1$ — значение обеспеченности в укороченном ряде на $m$ элементов.

Занесём значения в табл. 3.9.

*Таблица 3.9.* **Ординаты усечённой кривой обеспеченности Пирсона III типа**

| **Значение обеспеченности $P_1$, %** | **Высший уровень воды, см** | **Значение обеспеченности в исходной выборке $P_0$, %** |
|---|---|---|
| 0,01 | 957 | 0,004 |
| 0,02 | 889 | 0,009 |
| 0,05 | 813 | 0,022 |
| 0,1 | 756 | 0,043 |
| 0,2 | 698 | 0,087 |
| 0,5 | 618 | 0,22 |
| 1 | 555 | 0,43 |
| 2 | 495 | 0,87 |
| 5 | 418 | 2 |
| 10 | 357 | 4 |
| 20 | 298 | 9 |
| 25 | 280 | 11 |
| 30 | 263 | 13 |
| 40 | 239 | 17 |
| 50 | 220 | 22 |
| 60 | 205 | 26 |
| 70 | 193 | 30 |
| 75 | 187 | 33 |
| 80 | 181 | 35 |
| 90 | 172 | 39 |
| 95 | 168 | 41 |
| 98 | 165 | 42 |
| 99 | 165 | 43 |
| 99,5 | 165 | 43 |
| 99,8 | 164 | 43 |
| 99,9 | 164 | 43 |

По приведённым данным строится усечённая кривая обеспеченности (рис. 3.8).

Для высших уровней воды и максимальных экстремальных расходов воды обычно даются значения обеспеченности до 25–50 %. В данном случае кривая обеспеченности несколько лучше описывает верхнюю часть. Если расчётная задача не включает в себя определения значения обеспеченностей, больших, чем точка усечения, то задачу можно считать решённой, в противном случае необходимо рассчитать составную кривую обеспеченности, которая, по сути, собирается из двух усечённых кривых. Алгоритм сводится к следующим операциям:

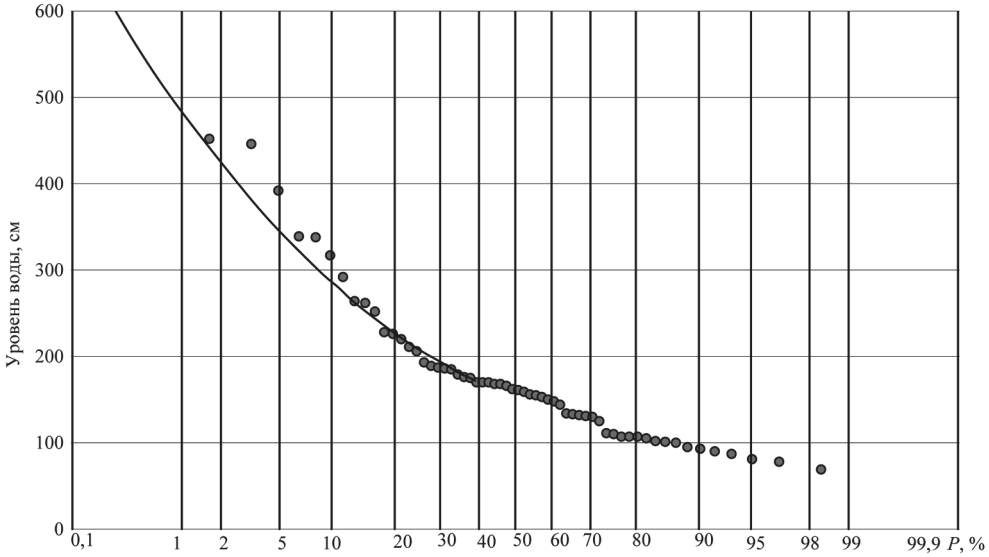



*Рис. 3.8.* Эмпирическая (кружочки) и усечённая аналитическая (сплошная линия) кривая обеспеченности Пирсона III типа высших уровней воды р. Кундрючья по пос. Владимирская

1) строим две (или более) кривые обеспеченности (любые) для верхней и нижней части ряда;
2) по этим кривым определяем максимальное и минимальное значения характеристики, размах; после этого разбиваем весь диапазон данных на *n*-е количество интервалов (необязательно равновеликих);
3) оцениваем обеспеченность каждого расхода воды по обеим кривым и вычисляем значение составной кривой обеспеченности.

Значение составной кривой обеспеченности рассчитывается по формуле

$$P_0 = \frac{P_1 n_1 + P_2 n_2 + P_N n_N}{N},$$

где $P_1$ — обеспеченность значения расхода воды по первой части ряда; $n_1$ — длина этой части ряда и т. д.

Выполним соответствующий расчёт, разделив ряд на две части, как в предыдущем случае. Вторая часть ряда обладает отрицательной асимметрией, поэтому целесообразно, так же, как и в первом случае, использовать распределение Пирсона III типа. После выполнения всех расчётов получаем минимальное значение уровня воды, равное 31 см (обеспеченность 99,9 %); максимальное значение уровня воды нам известно по первой кривой, построенной ранее, оно составляет 957 см (табл. 3.9). Задаём шаг по уровню воды, который может быть переменным; при решении данной задачи необходимо руководствоваться плавным изменением обеспеченности, после чего путём интерполяции для каждого значения уровня воды определяется обеспеченность по обеим кривым (табл. 3.10).

*Таблица 3.10*. **Ординаты составной кривой обеспеченности Пирсона III типа**

| Значение уровня воды, см | Значение обеспеченности, % | | |
|---|---|---|---|
| | для многоводных лет | для маловодных лет | составной кривой |
| 31 | 100 | 99,9 | 99,94 |
| 40 | 100 | 99,7 | 99,83 |
| 50 | 100 | 99,3 | 99,60 |
| 60 | 100 | 98,4 | 99,09 |
| 70 | 100 | 96,4 | 97,96 |
| 80 | 100 | 92,9 | 95,98 |
| 90 | 100 | 87 | 92,63 |
| 100 | 100 | 78 | 87,53 |
| 120 | 100 | 57 | 75,63 |
| 140 | 100 | 30 | 60,33 |
| 160 | 100 | 11,5 | 49,85 |
| …… | …… | …… | …… |
| 760 | 0,98 | 0 | 0,42 |
| 780 | 0,07 | 0 | 0,03 |
| 800 | 0,06 | 0 | 0,03 |
| 820 | 0,048 | 0 | 0,02 |
| 840 | 0,037 | 0 | 0,02 |
| 860 | 0,028 | 0 | 0,01 |
| 880 | 0,022 | 0 | 0,01 |
| 900 | 0,018 | 0 | 0,01 |
| 910 | 0,015 | 0 | 0,01 |
| 920 | 0,013 | 0 | 0,01 |
| 930 | 0,0125 | 0 | 0,01 |
| 940 | 0,012 | 0 | 0,01 |
| 957 | 0,01 | 0 | 0,004 |

После определения ординат построим составную кривую обеспеченности, для визуального сравнения добавим и простую кривую обеспеченности, рассчитанную ранее (рис. 3.9).

Визуальный анализ соответствия аналитических законов распределения эмпирическим данным наглядно показывает лучшее соотношение для составной кривой обеспеченности. Однако надо понимать, что подобный подход не часть классической статистики, а метод гидрологических расчётов, поэтому его применение требует большой осторожности. В частности, ряды данных для определения параметров распределения должны быть достаточной продолжительности, составные кривые должны иметь большее число параметров

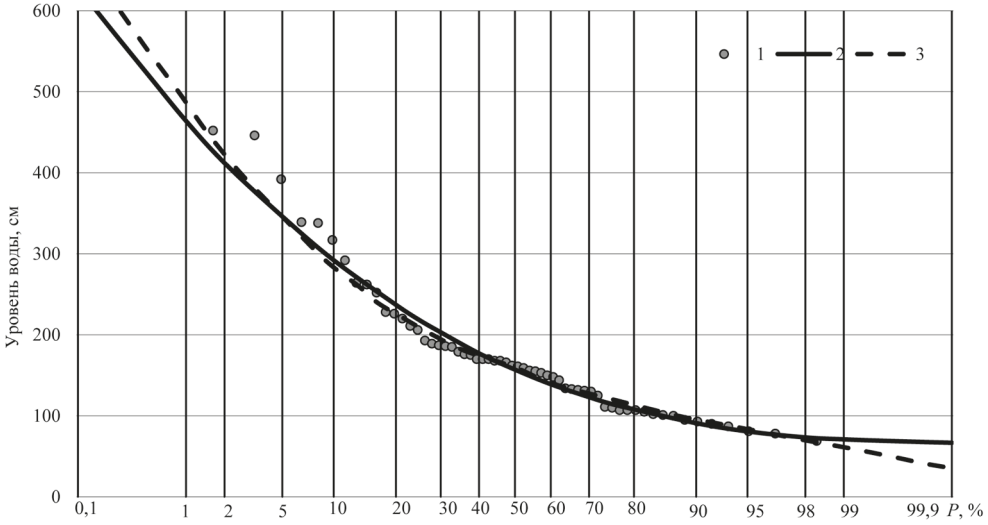



*Рис. 3.9.* Эмпирическая (1), аналитическая кривая обеспеченности Пирсона III типа (2) и составная (3) кривая обеспеченности Пирсона III типа высших уровней воды р. Кундрючья по пос. Владимирская

распределения (для каждой части ряда, по сути, определяются свои параметры распределения, при этом при достаточном обосновании могут использовать любые законы распределения). Поэтому применение составных кривых обеспеченности требует не только статистического, но и гидрологического обоснования.

## *Контрольные вопросы*

1. Что такое случайная величина?
2. Что такое хронологический график и что он показывает?
3. Что такое линия тренда, для чего она используется?
4. Что такое эмпирическая гистограмма распределения и каковы её свойства?
5. В чём сущность центральной предельной теоремы?
6. Как проводится стандартизация данных?
7. Что такое модульные коэффициенты?
8. Назовите способы оценки однородности и стационарности рядов.
9. Как рассчитывается и для чего используется критерий Стьюдента?
10. Как рассчитывается и для чего используется критерий Фишера?
11. Что такое уровень значимости и доверительная область?
12. Что такое и для чего используется *p-value*?
13. Для чего используются метод моментов и метод наибольшего правдоподобия?
14. Что называется несмещёнными оценками параметров распределения?
15. Назовите формулы для оценки параметров распределения методом моментов.

16. Какие существуют формы представления закона распределения случайной величины, что такое понятие обеспеченности?
17. Эмпирические и аналитические кривые распределения.
18. Назовите параметры, необходимые для расчёта ординат кривой обеспеченности Пирсона III типа.
19. Назовите параметры, необходимые для расчёта ординат кривой обеспеченности Крицкого — Менкеля.
20. Критерий $\chi^2$ для оценки соответствия аналитического закона распределения эмпирическим данным.

# 4. РАБОТА II. ПРОГНОЗИРОВАНИЕ ЭКСТРЕМАЛЬНЫХ ХАРАКТЕРИСТИК ВОДНОГО РЕЖИМА МЕТОДАМИ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ

Настоящая работа посвящена разработке методики прогнозирования высших уровней воды.

*Исходные данные* должны представлять собой случайные величины с датами регистрации (измерения). В качестве прогнозируемой величины предпочтительно использовать расходы воды на бесприточном участке реки за период наблюдений не менее 30–50 лет. В качестве предикторов можно взять данные о снегозапасах, расходах воды на вышележащих пунктах с лагом во времени и т. д. Оптимальный состав и количество предикторов определяется непосредственно студентами. Исходные данные для анализа готовятся непосредственно студентами на основе ежегодных изданий Государственного водного кадастра и других возможных источников.

*Цель работы* — получение практических навыков использования статистических методов для прогнозирования. Для достижения поставленной цели студентам необходимо освоить применение методов корреляционного и регрессионного анализов, а также метода обучения искусственных нейронных сетей.

*Задание.* По результатам совместных наблюдений за максимальными расходами (уровнями) воды и другими гидрометеорологическими характеристиками, анализа литературных источников определить оптимальный предиктивный состав прогностической модели. Оценить заблаговременность прогноза на основе метода множественной линейной регрессии и метода обучения искусственных нейронных сетей, построить и обучить прогностические модели. Оценить качество и эффективность полученных моделей.

### *Порядок выполнения работы и отчётные материалы*

При выполнении данной работы надо полагаться не только на методы математической статистики, но и на профессиональный опыт и знания о физических процессах формирования прогнозируемого события (прогнозируемая величина также может называться предиктантом). Безусловно, на формирование высших уровней воды оказывают влияние множество факторов, которые можно разделить на постоянно и переменно действующие.

К *постоянно действующим* факторам можно отнести характеристики и параметры водосбора и русла, учёт которых крайне важен в физико-математиче-

ских моделях. С точки зрения статистики, при разработке модели прогнозирования для конкретного створа данные факторы не играют роли, так как не оказывают влияния на вариацию прогнозируемой характеристики. С другой стороны, изменение этих параметров вследствие антропогенной нагрузки или иных воздействий приведёт к неустойчивости всей модели. Поэтому при использовании статистических методов прогнозирования принимается допущение о неизменности во времени состава грунтов и вклада этого состава в коэффициенты фильтрации, шероховатости, залесённости, заболоченности, озёрности.

К *переменно действующим* факторам (предикторам) относятся все условия текущего года, непосредственно оказывающие влияние на величину высшего уровня воды данного года в данном створе. К ним могут быть отнесены, с одной стороны, запасы воды в снеге, степень увлажнённости бассейна в осенний период, глубина промерзания почвогрунтов, максимальные расходы (уровни) воды на вышележащих створах, осадки, выпадающие в период половодья, — как факторы, характеризующие приток воды к створу (расходы воды). С другой стороны, это различные гидравлические факторы, определяющие величину переменного подпора и вероятность образования заторов льда. К данным факторам, как правило, относят высший уровень в начале ледостава как характеристику зашугованности русла, максимальные толщины льда как интегральную характеристику его прочности (что в целом весьма условно) и другие возможные предикторы, хорошо коррелирующие с прогнозируемой величиной. Общее правило об окончательном составе предиктивной модели должно исходить из максимальной их информативности для прогноза, лёгкости и прозрачности интерпретации, отсутствии несогласованного нарушения стационарности, мультиколлинеарности (предикторы не должны коррелировать между собой). Максимальное число предикторов при этом будет зависеть от длины выборки. В целом при длине выборки 50–100 лет допустимо использовать 3–4 предиктора. Если нужно большее количество предикторов, коррелируемых между собой, предикторы объединяются в некоррелируемые между собой факторы. Из методов факторного анализа наиболее распространённым является метод главных компонент.

Несомненно, что от выбора предикторов будет зависеть заблаговременность прогноза. Так, по информации о запасе воды в снеге, глубине промерзания и степени осеннего увлажнения можно прогнозировать высшие уровни воды с заблаговременностью более 1 мес., тогда как по значениям высших уровней вышележащих створов заблаговременность прогноза будет всего несколько суток. Под *заблаговременностью* подразумевается разница в днях от даты выпуска прогноза до наступления прогнозируемого события. Дата выпуска прогноза приурочивается к дате последнего события, используемого в качестве предиктора. В большинстве случаев с уменьшением заблаговременности прогноза растёт и его качество, которое оценивается стандартной ошибкой прогнозирования:

$$S = \sqrt{\frac{\sum_{i=1}^{N}\left(x_i - x^{*}_{\text{прогнозное}}\right)^2}{N}}. \qquad (4.1)$$

Несложно увидеть, что данная величина представляет собой среднее квадратическое отклонение фактических значений прогнозируемой величины от её прогнозных значений и по своему физическому смыслу близка к стандартному отклонению случайной величины. При этом из свойств среднего значения известно, что именно в зависимости от его фиксированного значения среднее квадратическое отклонение будет минимальным. Поэтому критерием качества и эффективности долгосрочных прогнозов является отношение стандартной ошибки прогнозирования $S$ к стандартному отклонению прогнозируемой величины σ. Прогностическая модель считается эффективной, если

$$\frac{S}{\sigma} < 0,8.$$

Для краткосрочных прогнозов используется отношение

$$\frac{S}{\sigma_\Delta} < 0,8,$$

где $\sigma_\Delta$ — среднее квадратическое отклонение фактических значений прогнозируемой величины от суммы фактических значений прогнозируемой величины на дату выпуска прогноза и среднего её изменения за период заблаговременности,

$$\sigma_\Delta = \sqrt{\frac{\sum_{i=1}^{N}\left(\Delta_i - \overline{\Delta}\right)^2}{N}}, \qquad (4.2)$$

$\Delta_i$ и $\overline{\Delta}$ — фактическое и среднее изменения прогнозируемой величины за период заблаговременности соответственно.

В зарубежной литературе можно встретить иные критерии качества, наиболее распространённым из которых является критерий Нэша — Сатклиффа — отношение суммы квадратов ошибок прогноза к сумме квадратов ошибок прогнозов, данных по среднему значению. В идеальном случае, когда прогнозные значения равны наблюдённым, дробь обращается в ноль, а коэффициент Нэша — Сатклиффа равен единице. Если прогностическая методика неэффективна по сравнению с ориентированием на среднее значение, коэффициент Нэша — Сатклиффа стремится к нулю:

$$NSE = 1 - \frac{\sum_{i=1}^{N}\left(x_i - X^{*}_{\text{прогнозное}}\right)^2}{\sum_{i=1}^{N}\left(x_i - \overline{\overline{X}}\right)^2}. \qquad (4.3)$$

Применение того или иного критерия качества должно отвечать наибольшей его строгости для критической оценки прогностической модели.

В качестве примера рассмотрим прогнозирование высших уровней воды за период весеннего ледохода, г. Котлас. Специфика формирования высших

уровней в Котласе обусловливается тем, что он расположен в месте слияния рек — Северной Двины и Вычегды. Однако предварительный анализ показал, что Вычегда вскрывается несколько позже Северной Двины и её основного притока — Сухоны. Таким образом, учёт ледовой ситуации на Вычегде невозможен ввиду отрицательной заблаговременности. Повторимся, что определение оптимального состава предиктивной модели является непосредственной задачей исследователя, и именно от этого и зависит конечная эффективность полученной модели. Для определения оптимального состава предикторов необходимо в первую очередь обратиться к литературным источникам, посвящённым решению той или иной проблемы прогнозирования. В случае же их отсутствия — полагаться на свой опыт и интуицию, статистические методы при этом являются вспомогательным инструментом, анализировать целесообразнее всего парные коэффициенты корреляции между предикторами и предиктантом, отбирая наиболее значимые.

Для Котласа ранее уже были разработаны методики прогнозирования высшего уровня воды за период весеннего ледохода. В частности, в работе [Бузин, 2015] предложена методика прогнозирования высшего уровня ледохода в Котласе в зависимости от максимального (высшего) уровня в начале ледостава ($H_{лс}$), максимальной толщины льда ($t_л$), расхода воды р. Сухоны в первый день ледохода у г. Тотьмы ($Q_{Тотьма}$), который при прогнозе вычисляется через уровень воды в Тотьме ($H_{Тотьма}$) и подъём уровня в этот же день ($\Delta H$), а также среднесуточной температуры воздуха на дату выпуска прогноза ($\theta$):

$$H_{лдх} = 0{,}0281 H_{лс} + 3{,}968 t_л + 0{,}273 Q_{Тотьма} + 4{,}93\theta - 27, \quad (4.4)$$

$$Q_{Тотьма} = 2{,}74 H_{Тотьма} - 0{,}783 \Delta H - 206. \quad (4.5)$$

Средняя заблаговременность такого прогноза составляет пять дней, стандартная ошибка 64 см, эффективность, оценённая по отношению к стандартному отклонению прогнозируемой величины, составила 0,59, к среднему изменению за период заблаговременности 0,47, что свидетельствует о достаточно высоком качестве данной методики [Бузин, 2015]. Среднесуточная температура воздуха как фактор, влияющий на формирование высшего уровня ледохода, как правило, остаётся незначимой, то же самое касается и максимальной толщины льда. Например, отмечается, что статистическая связь высшего уровня ледохода и максимальной толщины льда невысока, а коэффициент корреляции редко может достигать 0,45 [Бузин, 2015. С. 117]. Высший уровень в начале ледостава иногда может оказывать существенное влияние на формирование высшего уровня ледохода, однако для Котласа коэффициент корреляции между ними составляет всего 0,35.

Таким образом, основным информативным предиктором в уравнении (4.4) является расход воды р. Сухоны в первый день ледохода у г. Тотьмы, который может быть заменён уровнем воды на тот же день. При этом стоит отметить, что замена расхода воды соответствующим уровнем с гидрологической точки зрения является оправданной, так как расходы воды обычно вычисляются в конце года, а их предвычисление закладывает серьёзную ошибку и неод-

нозначность в текущий прогноз. Однако подобная модель содержит всего лишь один предиктор, она недостаточно эффективна, поэтому требуется найти как минимум ещё один предиктор. Таким предиктором может быть за октябрь — март низший уровень воды в районе Котласа. Минимальные уровни, с гидрологической точки зрения, с одной стороны, прямо связаны с дефицитом воды в русле, а с другой — обратно связаны с вероятностью образования заторов льда.

Проанализируем парные коэффициенты корреляции между предикторами и предиктантом, для чего в пакете анализа перейдём в раздел «Корреляция» и выберем все исходные данные, предназначенные для анализа, и укажем выходной интервал (табл. 4.1).

*Таблица 4.1.* **Результаты корреляционного анализа**

| **Переменная** | $H_{\text{min. Котлас}}(X_1)$ | $H_{\text{Тотьма}}(X_2)$ | $H_{\text{Котлас}}(Y)$ |
|---|---|---|---|
| $H_{\text{min. Котлас}}(X_1)$ | 1 | – | – |
| $H_{\text{Тотьма}}(X_2)$ | 0,33 | 1 | – |
| $H_{\text{Котлас}}(Y)$ | 0,52 | 0,74 | 1 |

Как известно, коэффициенты корреляции изменяются от –1 до +1, при этом чем ближе абсолютное значение коэффициента корреляции к единице, тем лучше связь между переменными. При прогнозировании рядов значимыми, как правило, можно считать коэффициенты корреляции более 0,50. Анализ представленных результатов показал корреляционную значимость данной модели, отсутствие выраженной мультиколлинеарности. Это позволяет использовать данную модель в дальнейшем при положительной заблаговременности прогноза.

В настоящей модели использовано два предиктора, дата выпуска прогноза приурочивается к первому дню ледохода в Тотьме. Следовательно, заблаговременность прогноза будет определяться разницей между этой датой и датой формирования высшего уровня воды в Котласе (табл. 4.2). Таким образом, средняя заблаговременность прогноза составила 6 дней. В отчёте представляют гистограмму распределения заблаговременности прогноза (рис. 4.1) (порядок построения гистограммы см. в разд. 3, работа I).

Анализ данной заблаговременности показывает модальное значение заблаговременности за 4–8 дней и сильную положительную асимметрию. Прогнозы, данные с заблаговременностью от 1 до 16 дней, можно считать среднесрочными, поэтому проверку качества надо проводить по обеим вышеописанным методикам.

После того как были рассчитаны заблаговременности прогноза и получены положительные результаты, свидетельствующие о корректности и адекватности прогностической модели, можно приступить непосредственно к разработке прогностической модели на основе метода множественной линейной регрессии. Для понимания множественной линейной регрессии следует обратиться к обычной парной линейной регрессии, которая представляет собой простую линейную аппроксимацию поля точек методом наименьших

*Таблица 4.2.* **Основные параметры для построения предиктивной модели и оценки заблаговременности и качества выпускаемых прогнозов**

| Год | Высший уровень воды (Котлас) | Уровень воды | Минимальный уровень воды (Котлас) за октябрь — март | Начало ледохода в г. Тотьма | Заблаговременность прогноза | Уровень воды в первый день ледохода (Тотьма) | Уровень воды на дату выпуска прогноза, г. Котлас | Изменение уровня воды за период заблаговременности (Δ) | $(\Delta - \Delta_{ср})^2$ |
|---|---|---|---|---|---|---|---|---|---|
| 1951 | 07.04.1951 | 562 | 41 | 05.04.1951 | 2 | 494 | 489 | 73 | 37 249 |
| 1954 | 23.04.1954 | 586 | 98 | 20.04.1954 | 3 | 492 | 236 | 350 | 7056 |
| 1987 | 03.05.1987 | 392 | 92 | 26.04.1987 | 7 | 269 | 172 | 220 | 2116 |
| 2016 | 22.04.2016 | 618 | 48 | 13.04.2016 | 9 | 475 | 350 | 268 | 4 |
| 2013 | 30.04.2013 | 502 | 95 | 23.04.2013 | 7 | 497 | 238 | 264 | 4 |
| 2010 | 21.04.2010 | 460 | 45 | 15.04.2010 | 6 | 408 | 195 | 265 | 1 |
| 1985 | 01.05.1985 | 475 | 65 | 27.04.1985 | 4 | 387 | 149 | 326 | 3600 |
| 1962 | 20.04.1962 | 534 | 97 | 14.04.1962 | 6 | 447 | 247 | 287 | 441 |
| 1990 | 14.04.1990 | 420 | 34 | 11.04.1990 | 3 | 478 | 274 | 146 | 14 400 |
| 1974 | 06.05.1974 | 439 | 47 | 27.04.1974 | 9 | 337 | 156 | 283 | 289 |
| …… | …… | …… | …… | …… | …… | …… | …… | …… | …… |
| 1958 | 29.04.1958 | 590 | 135 | 28.04.1958 | 1 | 502 | 340 | 250 | *256* |
| 1993 | 27.04.1993 | 468 | -15 | 25.04.1993 | 2 | 485 | 318 | 150 | *13 456* |
| 2003 | 23.04.2003 | 437 | 54 | 18.04.2003 | 5 | 180 | 166 | 271 | *25* |
| 1953 | 22.04.1953 | 605 | 108 | 17.04.1953 | 5 | 502 | 366 | 239 | *729* |
| 1971 | 05.05.1971 | 382 | 56 | 21.04.1971 | 14 | 316 | 235 | 147 | *14 161* |
| 1965 | 04.05.1965 | 384 | 86 | 25.04.1965 | 9 | 377 | 194 | 190 | 5776 |
| 1956 | 04.05.1956 | 635 | 51 | 01.05.1956 | 3 | 510 | 301 | 334 | 4624 |
| 1972 | 27.04.1972 | 629 | 45 | 24.04.1972 | 3 | 536 | 300 | 329 | 3969 |
| 1969 | 25.04.1969 | 585 | 59 | 21.04.1969 | 4 | 399 | 222 | 363 | 9409 |
| 2012 | 26.04.2012 | 605 | 6 | 23.04.2012 | 3 | 618 | 218 | 387 | 14 641 |

*Примечание.* Курсивом выделено тестовое подмножество данных.

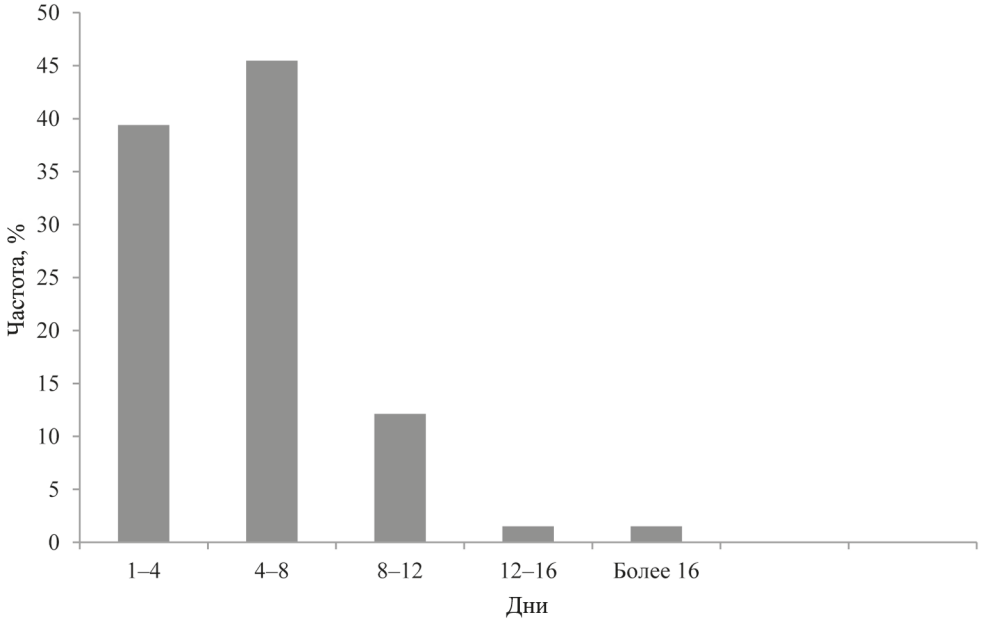



*Рис. 4.1.* Гистограмма распределения заблаговременности прогноза

квадратов. Разновидностью парной линейной регрессии является уравнение линии тренда, оценка значимости коэффициента корреляции которого была проведена в работе I. Построение линии регрессии проводится методом наименьших квадратов, суть которого заключается в минимизации квадратического отклонения наблюдённых значений от линии регрессии. Уравнение парной линейной регрессии имеет вид

$$Y = aX + C + \varepsilon,$$

где $Y$ — зависимая величина (предиктант или прогнозируемая величина в случае решения задачи прогнозирования); ε — случайная погрешность уравнения линейной регрессии, $a$ и $C$ — эмпирические коэффициенты, определяемые методом наименьших квадратов.

Коэффициент $a$ показывает, насколько в среднем изменится зависимая переменная $Y$ при изменении факторной переменной на единицу своего измерения, и в геометрическом смысле представляет собой тангенс угла наклона линии регрессии; коэффициент $C$ — расстояние от начала координат до точки пересечения оси ординат с линией регрессии. Данные коэффициенты определяются по следующим формулам:

$$a = R\frac{\sigma_Y}{\sigma_x},$$

$$C = \overline{Y} - a\overline{X},$$

где $R$ — коэффициент корреляции,

$$R_{XY} = \frac{\sum (X - \bar{X}) \times (Y - \bar{Y})}{\sqrt{\sum (X - \bar{X})^2 \times \sum (Y - \bar{Y})^2}}.$$

В случае множественной линейной регрессии зависимость результирующей переменной одновременно от нескольких объясняющих переменных описывает уравнение или модель. Таким образом, вместо линии регрессии в ней используется гиперплоскость. Как и в простой линейной регрессии, параметры модели $a_n$ и $C$ вычисляются методом наименьших квадратов. Разумеется, мы будем изучать построение модели множественной регрессии и её оценивание с использованием программных средств. Преимущество множественной линейной регрессии по сравнению с простой заключается в том, что использование в модели нескольких входных переменных позволяет увеличить долю объяснённой дисперсии выходной переменной и таким образом улучшить соответствие модели данным. То есть при добавлении в модель каждой новой переменной коэффициент детерминации растёт. Однако его рост с добавлением новых предикторов обозначает также и минус данного метода, поэтому в качестве меры надёжности уравнения множественной линейной регрессии рекомендуется использовать скорректированный коэффициент детерминации, который снижается при необоснованном добавлении предикторов. Как уже было сказано выше, при длительности рядов наблюдения 50–100 лет рекомендуется использовать не более трёх предикторов.

Для построения регрессионной модели в целях прогнозирования величин всю выборку следует разделить на две подвыборки, одна из которых используется для построения регрессионной модели, а вторая — для проверки на независимом материале. Разделение является обязательным условием проверки получившейся модели, так как именно проверка на независимом материале показывает устойчивость полученной модели и её параметров, свидетельствует о достаточности и неизбыточности количества предикторов. Считается, что данное разделение объективнее всего проводить случайным образом. Такое разделение можно выполнить, предварительно ранжировав ряд по сгенерированному случайному числу. В пакете *Excel* с помощью функции генерации случайных чисел, расположенной в пакете анализа, создаётся синтетическая случайная величина, имеющая нормальное распределение, и помещается в столбце рядом с прогнозируемой величиной. После этого необходимо выполнить сортировку (ранжирование) всех данных по столбцу со сгенерированной случайной величиной.

Таким образом, $n$ столбцов с годом, прогнозируемой величиной и предикторами расположатся в случайном порядке, а столбец со сгенерированной случайной величиной — в убывающем или возрастающем порядке. При описанном способе обучающая подвыборка составляет первые 70–85 % значений от всей выборки, тестовая — последние 15–30 %. Отметим, что не рекомендуется формировать обучающее и тестовое подмножество из первых и по-

следних значений выборки хронологического ряда без предварительного ранжирования рядов описанным выше способом, так как в этом случае в обучающее подмножество совсем не попадут современные наблюдения, что может сказаться на качестве модели (табл. 4.2).

После разделения выборки на два подмножества приступают к построению модели множественной линейной регрессии. В пакете *Excel* из окна «Анализ данных» нужно перейти в окно «Регрессионный анализ». В открывшемся окне указываются обучающий диапазон прогнозируемой величины (*Y*) и диапазон предикторов (*X*), целесообразно также показать необходимость вывода остатков модели. Результаты моделирования представляют в отчёте (табл. 4.3, 4.4).

Анализ полученных результатов показал следующее. Значение коэффициента множественной корреляции должно быть в пределах 0,7; нормированного коэффициента детерминации 0,5; *P*-значения — менее 0,05, что свидетельствует о статистической значимости полученных оценок. Также можно сразу оценить качество модели по критериям $S/\sigma$ и $S/\sigma_\Delta$ по обучающей выборке:

$$\frac{S}{\sigma} = \frac{65}{103} = 0,63,$$

$$\frac{S}{\sigma_\Delta} = \frac{65}{90} = 0,72.$$

*Таблица 4.3.* **Характеристики регрессионного уравнения**

| **Регрессионная статистика** | ***P*-значение** |
|---|---|
| Множественный *R* | 0,78 |
| *R*-квадрат | 0,61 |
| Нормированный *R*-квадрат | 0,60 |
| Стандартная ошибка | 65 |
| Наблюдения | 56 |

*Таблица 4.4.* **Параметры регрессионного уравнения**

| **Коэффициент** | **Коэффициенты уравнения** | ***P*-значение** |
|---|---|---|
| *C* (*Y*-пресечение) | 208 | 0 |
| Коэффициент при первом предикторе $a_1$ | 0,80 | 0,04 |
| Коэффициент при втором предикторе $a_2$ | 0,62 | 0 |

После оценки качества выпускаемых прогнозов на *зависимом* материале, необходимо рассчитать прогнозные значения для *независимого* материала и оценить качество выпускаемых прогнозов по аналогичным критериям. На *независимом* материале значения данных критериев равны 0,59 и 0,79 соответ-

ственно. Качество выпускаемых прогнозов может быть значительно улучшено при использовании метода искусственных нейронных сетей. Искусственная нейронная сеть (ИНС) — математическая модель, а также её программное или аппаратное воплощение, построенное по принципу организации и функционирования сетей нервных клеток живого организма. Это понятие возникло при изучении процессов, протекающих в мозге, и попытках их смоделировать. Первой такой попыткой были нейронные сети, представленные в работе [Мак-Каллок, Питтс, 1956].

После разработки алгоритмов обучения полученные модели стали использовать в практических целях: в задачах прогнозирования, для распознавания образов, в задачах управления и др. Методы обучения искусственных нейронных сетей для краткосрочного и долгосрочного прогнозирования элементов водного режима рек и озёр распространяются всё более широко и повсеместно. В настоящее время проведено много исследований применимости ИНС в целях краткосрочного прогнозирования расходов и уровней воды за период весеннего половодья и дождевых паводков.

Рассмотрим на нашем примере решение регрессионной задачи прогнозирования с применением метода обучения ИНС в программном продукте *Statistica 12.* Во вкладке «Статистик» перейдём в окно «Нейронные сети — регрессия», где на первом этапе обозначим целевую переменную и предикторы, а также зададим размер обучающей и валидационной выборки, которая является квазинезависимой. Тестовое подмножество не следует указывать, так как оно вообще не будет участвовать в процессе обучения. Это делается для того, чтобы разбиение на подмножества у двух методов было одинаковым и возможно было принципиальное сравнение результатов двух методов. После перехода в окно «Автоматическое обучение искусственных нейронных сетей» следует указать тип ИНС, минимальное и максимальное число скрытых нейронов и активационные функции, число обучающихся нейронных сетей и число сохраняемых (лучших сетей). Рекомендуется использовать сеть многослойного персептрона (*MLP*), максимальное число скрытых нейронов должно быть необходимым и достаточным — не следует создавать излишне сложную сеть; в качестве активационных можно взять комбинации гиперболического тангенса и линейной функции.

После настройки перечисленных параметров нужно запустить обучение, а по его окончании перейти в окно «Анализ результатов». В данном окне наибольший интерес вызывают производительность нейронной сети, представляющая собой коэффициент корреляции между прогнозными и фактическими значениями, прогнозные значения и абсолютные ошибки прогнозирования. После обучения ИНС следует выбрать сеть с меньшей квадратической ошибкой прогнозирования и сделать поверочные прогнозы на независимом материале.

В настоящем примере лучшей по данным параметрам оказалась искусственная нейронная сеть *MLP* 2-5-1, в качестве активационных функций использованы гиперболический тангенс на скрытом слое и линейная активаци-

онная функция. На зависимом материале нейронная сеть *MLP* 2-5-1 показала следующие результаты:

$$\frac{S}{\sigma} = \frac{50}{103} = 0,48,$$

$$\frac{S}{\sigma_{\Delta}} = \frac{50}{90} = 0,55.$$

На независимом материале данные значения составляют 0,36 и 0,51 соответственно. Полученные результаты свидетельствуют о высоком качестве прогностической модели. Хорошее соответствие прогностических и фактических значений на независимом материале можно показать на совмещённом графике (рис. 4.2).

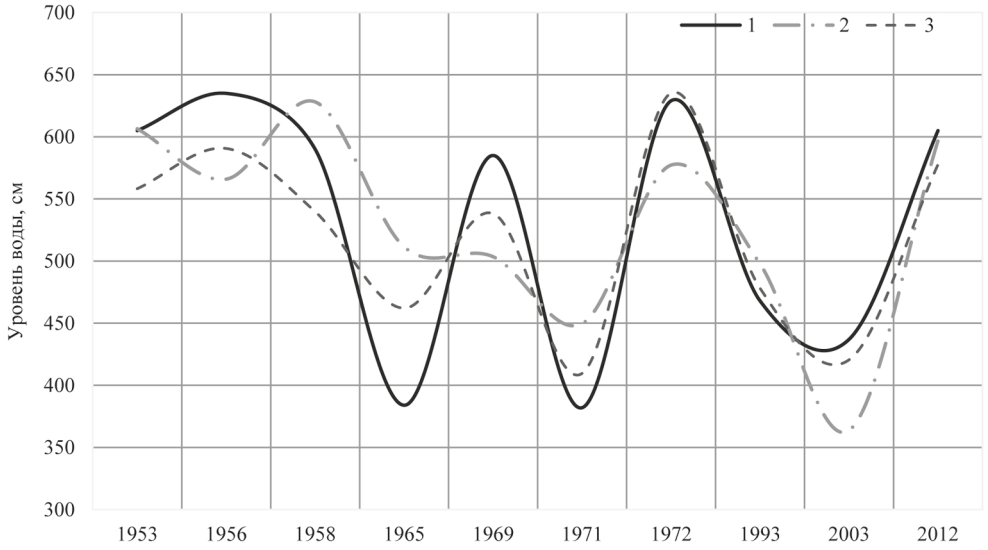



*Рис. 4.2.* Сопоставление фактических (1) и прогностических данных при использовании регрессионной (2) и нейросетевой (3) моделей высших уровней за период весеннего ледохода воды (Котлас)

Анализ представленных числовых значений показал, что искусственные нейронные сети обеспечивают существенное уменьшение стандартной ошибки прогнозирования (табл. 4.5).

В заключение к работе следует сделать выводы о соответствии фактических и прогностических данных, привести оценки качества моделей и среднюю заблаговременность прогноза. В данном случае искусственные нейронные сети показали лучший по сравнению с регрессионными методами результат, о чём можно судить по стандартной ошибке прогнозирования.

*Таблица 4.5.* **Оценка качества выпускаемых прогнозов для Северной Двины различными методами (г. Котлас)**

| Выборка | *S*, см | $S/\sigma_{ср}$ | $S/\sigma_{\Delta}$ | Допустимая ошибка, см $0{,}674 \cdot \sigma_{\Delta}$ | Число оправдавшихся прогнозов | Оправдываемость, % |
|---|---|---|---|---|---|---|
| | Искусственная нейронная сеть *MLP* 2-5-1 | | | | | |
| Полная | 48 | 0,47 | 0,54 | 60 | 52 | 79 |
| Обучающая | 50 | 0,48 | 0,55 | | 43 | 77 |
| Тестовая | 41 | 0,37 | 0,50 | | 9 | 90 |
| Уравнение множественной регрессии | | | | | | |
| Полная | 65 | 0,63 | 0,73 | 60 | 38 | 58 |
| Обучающая | 65 | 0,63 | 0,72 | | 33 | 59 |
| Тестовая | 65 | 0,59 | 0,79 | | 5 | 50 |

## *Контрольные вопросы*

1. В чём суть и задачи регрессионного и корреляционного анализа?
2. Суть метода наименьших квадратов.
3. Каковы меры для оценки связи между двумя величинами?
4. Суть парной линейной и нелинейной регрессии, запись уравнения линейной регрессии в общем виде.
5. Назовите формулы для определения коэффициентов уравнения линейной регрессии, их физический и геометрический смысл.
6. Стандартная ошибка регрессионного уравнения и его коэффициентов.
7. Суть стохастических методов прогнозирования, понятие заблаговременности.
8. Назовите критерии, применяемые для оценки качества выпускаемых прогнозов и их суть, допустимой оценки прогнозирования.
9. Суть метода обучения искусственных нейронных сетей при прогнозировании гидрометеорологических характеристик.
10. Понятия инициализации, активационных функций и обучения искусственной нейронной сети.

# 5. РАБОТА III. АНАЛИЗ И ПРОГНОЗИРОВАНИЕ ВРЕМЕННЫ́Х РЯДОВ С ИСПОЛЬЗОВАНИЕМ МОДЕЛИ АРПСС И ИСКУССТВЕННЫХ НЕЙРОННЫХ СЕТЕЙ

Настоящая работа посвящена оценке параметров временных рядов (случайных процессов) и их прогнозированию с использованием вероятностных подходов.

*Исходные данные* должны представлять собой случайный процесс за период наблюдения не менее 50 лет. В качестве исходных данных рекомендуется использовать средние месячные значения уровней воды крупных озёр либо расходы или уровни воды крупных рек с озёрным регулированием стока (Нева, Ангара и т. п.). Исходные данные для анализа готовятся непосредственно студентами на основе ежегодных изданий Государственного водного кадастра и других возможных источников.

*Цель работы* — получение практических навыков в области статистического анализа временных рядов и их прогнозирования. Студентам необходимо научиться применять встроенные функции и надстройки программного продукта *Excel* и *Statistica* для расчёта основных статистических характеристик и прогнозирования временных рядов; правильно интерпретировать полученные результаты анализа.

*Задание*. По результатам многолетних наблюдений определить и проанализировать характеристики случайного процесса. Оценить периодичность и цикличность исходных данных. Провести сезонную декомпозицию и на её основе построить авторегрессионную модель, учитывающую сезонные изменения. Определить параметры модели авторегрессии проинтегрированного скользящего среднего (АРПСС) и нейросетевых моделей многослойного персептрона для прогнозирования среднемесячных значений. Оценить качество выпускаемых прогнозов по всем трём моделям и обоснованно выбрать лучшую из них для решения конкретной задачи.

### *Порядок выполнения работы и отчётные материалы*

В качестве примера рассмотрим анализ и прогнозирование среднемесячных уровней воды оз. Ильмень с использованием модели АРПСС и искусственных нейронных сетей. Данный временной ряд является случайным процессом, который может быть представлен двенадцатью его реализациями (среднемесячными значениями уровня воды). Исходные данные представить в таблице (табл. 5.1).

Далее по исходным данным надо рассчитать основные статистические характеристики за каждый месяц (так же, как это делалось в разд. 3, работа I). Поскольку статистические характеристики нужны для описания ряда, а не для изучения его закона распределения, то расчёты следует выполнять методом моментов. Результаты расчёта приводятся в табл. 5.2.

Анализ таблицы показал нарушение стационарности временного ряда по дисперсии, для чего использовалась функция F.ТЕСТ, возвращающая значение *p-value*, и нарушение общей стационарности ряда по наличию статистически значимого тренда. Все расчёты следует выполнять для двухстороннего уровня значимости 10 %. В настоящем примере можно отметить наличие статистически значимого тренда к увеличению среднемесячных значений уровня воды с января по март, что приводит к увеличению уровней воды за год, однако в данном случае не наблюдается увеличения водности озера, так как подъём уровней воды может быть связан и с постепенным заилением. Наиболее заметное нарушение стационарности по дисперсии характерно для ноября и марта, что соответствует переходным периодам гидрологического режима, это можно увидеть по построенным спарклайнам. Смещение же сроков наступления данных фаз из-за глобального потепления вызывает неоднородность по дисперсии. Математическое ожидание случайного процесса представляет собой типичный гидрограф уровней воды, который показан в графе *спарклайн* (табл. 5.2).

Изучение распределений 12 случайных величин по гистограммам неудобно, так как гистограмма эмпирического распределения наглядно демонстрирует распределение одной случайной величины: при анализе нескольких случайных величин необходимо строить гистограммы для каждой из них. Для сравнительного анализа нескольких эмпирических распределений на практике используется диаграмма «ящики с усами» (*box-and-whiskers diagram/plot, box plot*), представляющая собой упрощённую модель гистограммы эмпирического распределения. Характеристики «ящика» и «усов» рассчитываются методами квантильного анализа. Алгоритм построения «ящиков с усами» сводится к определению медианы ряда, 1-го и 3-го квартилей (25-й и 75-й процентили соответственно); являясь верхней и нижней границами «ящика», 1-й и 3-й квартиль образуют тело «ящика». Затем наносятся выбросы, которые, как правило, могут быть определены как большие и меньшие на 1,5 межквартальных расстояния от границ «ящика» значения, выбросы обозначаются точками, максимальное и минимальное значения выборки, без учёта выбросов, соединяются усами; помимо медианы для оценки величины асимметрии в теле ящика наносится среднее арифметическое значение ряда. Однако не существует единого общего согласия, как конкретно строить «ящик с усами», при виде такого графика необходимо искать информацию в сопроводительном тексте, по каким параметрам «ящик с усами» строился.

Несмотря на свою простоту и удобство, первоначальная форма «ящика с усами» обладает и некоторыми недостатками. Один из таких существенных недостатков — отсутствие на графике информации о количестве наблюдений

*Таблица 5.1.* **Среднемесячные уровни воды оз. Ильмень по озёрному пункту дер. Войцы за 1960–2018 гг. (I–XII — месяцы)**

| Год | I | II | III | IV | V | VI | VII | VIII | IX | X | XI | XII | Средние годовые |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1960 | 201 | 177 | 156 | 297 | 503 | 406 | 337 | 284 | 277 | 264 | 296 | 391 | 299 |
| 1961 | 418 | 377 | 417 | 519 | 521 | 448 | 344 | 301 | 313 | 278 | 264 | 276 | 373 |
| 1962 | 263 | 252 | 233 | 457 | 628 | 533 | 502 | 462 | 432 | 414 | 402 | 464 | 420 |
| 1963 | 408 | 343 | 288 | 328 | 479 | 377 | 285 | 247 | 235 | 237 | 274 | 276 | 315 |
| 1964 | 242 | 210 | 181 | 255 | 433 | 383 | 296 | 239 | 201 | 187 | 204 | 224 | 255 |
| 1965 | 239 | 220 | 204 | 296 | 552 | 478 | 376 | 299 | 262 | 238 | 242 | 228 | 303 |
| 1966 | 220 | 201 | 203 | 492 | 574 | 606 | 454 | 333 | 274 | 267 | 267 | 252 | 345 |
| 1967 | 222 | 194 | 200 | 396 | 491 | 412 | 326 | 262 | 239 | 242 | 305 | 318 | 301 |
| 1968 | 274 | 231 | 206 | 562 | 609 | 508 | 400 | 325 | 266 | 252 | 298 | 282 | 351 |
| 1969 | 250 | 218 | 184 | 258 | 497 | 446 | 332 | 264 | 229 | 224 | 243 | 323 | 289 |
| 1970 | 303 | 258 | 213 | 350 | 535 | 421 | 309 | 249 | 224 | 227 | 241 | 254 | 299 |
| …… | …… | …… | …… | …… | …… | …… | …… | …… | …… | …… | …… | …… | …… | …… |
| 2008 | 264 | 256 | 364 | 494 | 494 | 388 | 303 | 258 | 231 | 233 | 251 | 269 | 317 |
| 2009 | 475 | 435 | 430 | 499 | 568 | 497 | 418 | 366 | 345 | 365 | 464 | 503 | 447 |
| 2010 | 439 | 365 | 302 | 551 | 610 | 523 | 412 | 301 | 261 | 232 | 264 | 312 | 381 |
| 2011 | 297 | 294 | 271 | 443 | 636 | 527 | 405 | 303 | 261 | 250 | 256 | 321 | 355 |
| 2012 | 380 | 352 | 299 | 404 | 555 | 457 | 375 | 297 | 257 | 257 | 342 | 377 | 363 |
| 2013 | 366 | 329 | 290 | 352 | 579 | 527 | 409 | 321 | 257 | 235 | 265 | 313 | 354 |
| 2014 | 366 | 342 | 314 | 360 | 354 | 319 | 275 | 236 | 208 | 210 | 215 | 221 | 285 |
| 2015 | 242 | 241 | 306 | 409 | 441 | 350 | 260 | 214 | 204 | 203 | 209 | 217 | 275 |
| 2016 | 228 | 251 | 291 | 377 | 446 | 367 | 319 | 344 | 311 | 281 | 296 | 367 | 323 |
| 2017 | 394 | 346 | 382 | 524 | 560 | 489 | 463 | 498 | 463 | 427 | 444 | 429 | 452 |
| 2018 | 488 | 467 | 403 | 482 | 573 | 450 | 342 | 277 | 242 | 234 | 241 | 242 | 370 |

*Таблица 5.2.* **Статистические характеристики среднемесячных значений уровней воды оз. Ильмень по озёрному гидрологическому пункту (дер. Войцы) за 1960–2018 гг. (I–XII — месяцы)**

| Характеристика | I | II | III | IV | V | VI | VII | VIII | IX | X | XI | XII | Средние годовые |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Mx* | 313 | 295 | 292 | 431 | 518 | 442 | 360 | 303 | 273 | 268 | 291 | 313 | 344 |
| Спарклайн | | | | | | | | | | | | | — |
| σ | 86 | 84 | 100 | 102 | 71 | 70 | 68 | 67 | 69 | 73 | 78 | 81 | 41 |
| Спарклайн | | | | | | | | | | | | | – |
| $C_v$ | 0,28 | 0,28 | 0,34 | 0,24 | 0,14 | 0,16 | 0,19 | 0,22 | 0,25 | 0,27 | 0,27 | 0,26 | 0,12 |
| $C_s$ | 0,5 | 0,5 | 0,8 | –0,3 | –0,4 | 0,2 | 0,8 | 1,5 | 1,7 | 2,1 | 1,3 | 0,9 | 0,3 |
| $F_{тест}$, % | 8 | 12 | 2 | 85 | 22 | 40 | 41 | 46 | 52 | 10 | 1 | 10 | 37 |
| $R_{тренда}$ | 0,30 | 0,39 | 0,42 | 0,24 | 0,02 | 0,03 | 0,11 | 0,14 | 0,09 | 0,05 | 0,06 | 0,06 | 0,32 |
| Значимость R | + | + | + | – | – | – | – | – | – | – | – | – | + |

по выборке, что в некоторых случаях может быть решено путём связи ширины ящика с величиной выборки. Безусловно, «ящики с усами» являются упрощённой версией гистограммы эмпирического распределения и могут быть использованы только в целях экспресс-анализа. «Ящики с усами», начиная с *Excel 2016*, являются стандартной диаграммой и имеют следующий путь: вставка — диаграммы — гистограмма — «ящик с усами». Для построения «ящиков с усами» необходимо выбрать все необходимые данные для анализа, после выполнения вышеуказанных процедур эти гистограммы построятся автоматически (рис. 5.1).

Случайные процессы, подобные данному, являются периодически скоррелированными, и представленные в виде 12 реализаций месяцы года хорошо коррелируют между собой. Для подобного корреляционного анализа необходимо рассчитать автокорреляционную матрицу с различными временными сдвигами, как правило, от 1 до 12 мес. Матрица корреляционных зависимостей внутригодовой изменчивости имеет две ветви связей: «вперёд» и «назад». Ветвь связи «вперёд» показывает коррелированность значений процесса в каждый *j*-й месяц года с величинами в последующие месяцы, ветвь связи «назад» — зависимость отсчётов процесса в каждый *j*-й месяц от значений в предшествующие ему месяцы. Подобный анализ надо провести для каждого месяца года [Бродская и др., 2015]. Таким образом, с одной стороны, получается корреляционная матрица января с январём, февралём, мартом и т. д., и января с декабрём, ноябрём и т. д. — с другой стороны. Результаты анализа представляются в виде графика корреляционной функции (рис. 5.2).

Анализ графика показывает быстрое снижение корреляционной функции уже после второго сдвига, когда значение коэффициентов корреляции падает

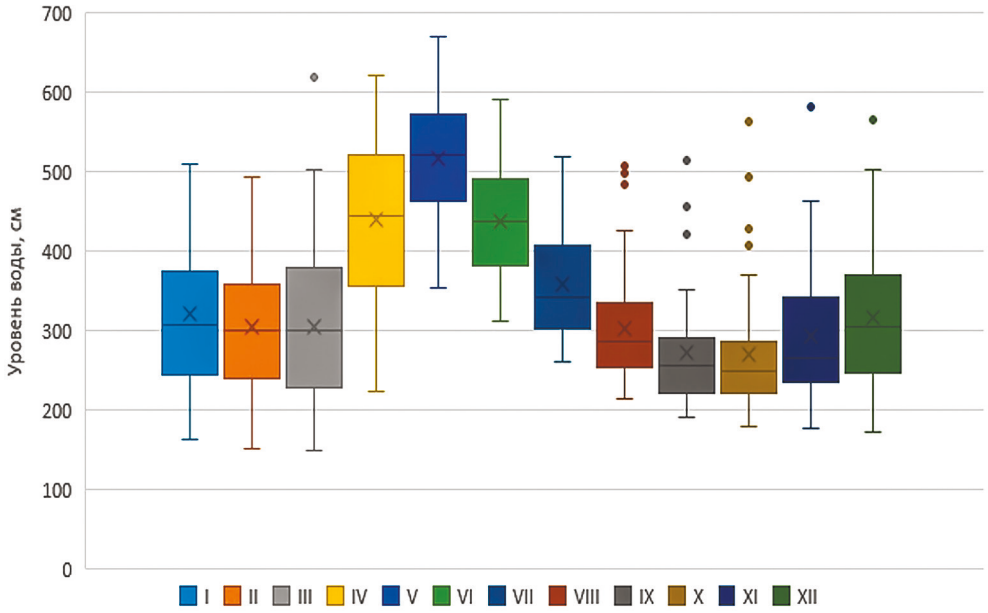



*Рис. 5.1.* Эмпирические гистограммы распределения среднемесячных (I–XII) уровней воды оз. Ильмень (дер. Войцы)

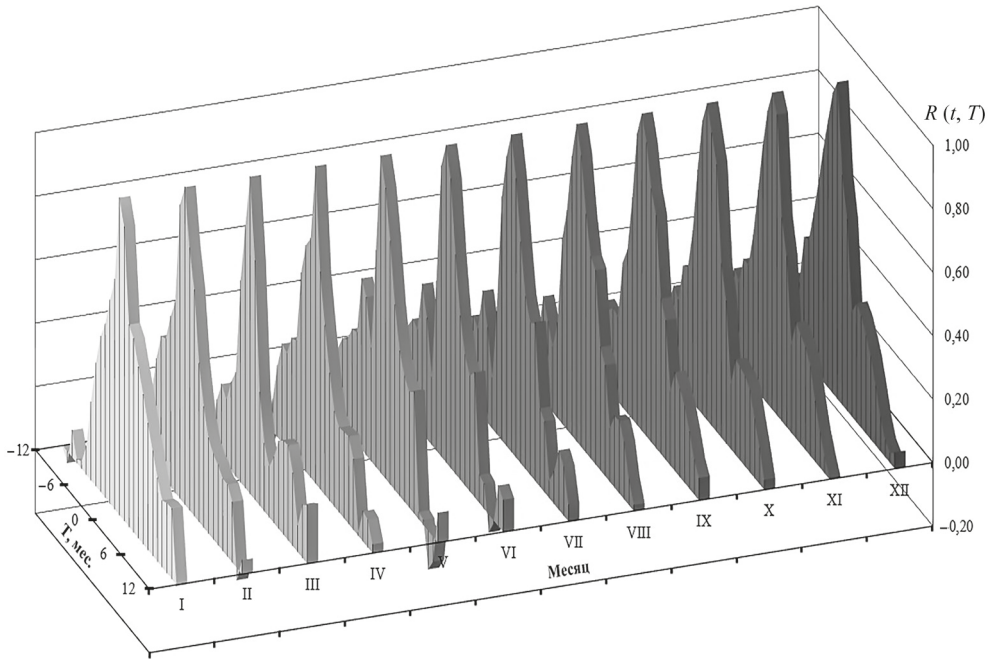



*Рис. 5.2.* График корреляционной функции периодически скоррелированного случайного процесса

до 0,6, что не всегда достаточно для прогнозирования. В целом также можно отметить однородную корреляционную структуру ряда, так как для каждого месяца года коррелограммы выглядят аналогичным образом.

После анализа характеристик случайного процесса, разместив последовательно все месяцы друг за другом в хронологическом порядке, строим график временного ряда (рис. 5.3).

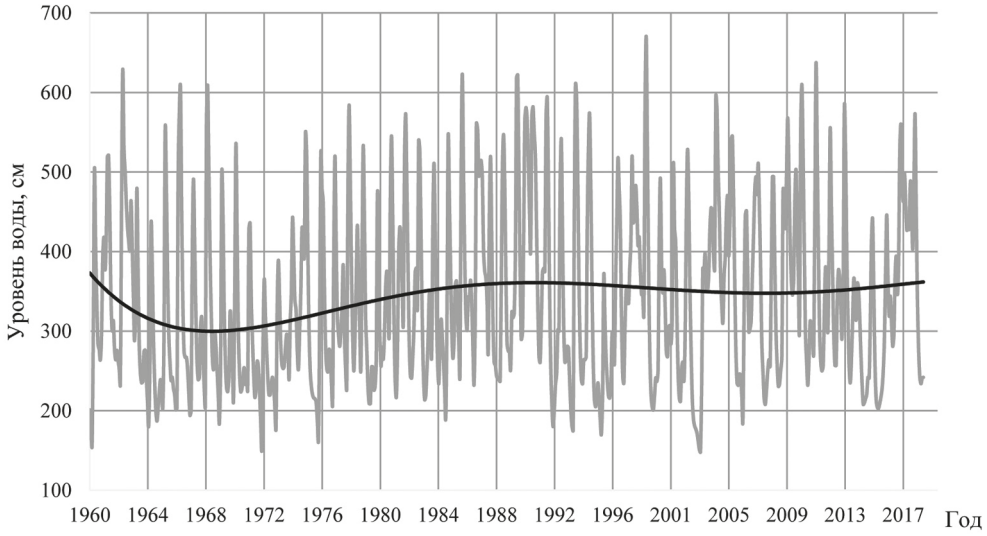



*Рис. 5.3.* Хронологический график среднемесячных значений уровней воды оз. Ильмень за 1960–2018 гг. (дер. Войцы)

При анализе временного ряда целесообразно построить не просто линейный тренд, а аппроксимировать ряд разночастотными фильтрами. Одним из таких фильтров является фильтр Баттерворта, приближённые значения также можно получить простым полиномиальным сглаживанием высокого порядка. Таким образом, на графике прослеживаются не только периодические колебания с периодом 12 мес., но и циклические колебания, учёт которых при прогнозировании может значительно улучшить качество выпускаемых прогнозов. Анализ циклических колебаний проводится с помощью программного продукта *Statistica*. Для реализации данного анализа на вкладке «Статистик» необходимо запустить «продвинутые» статистики и, выбрав анализ Фурье для одной переменной, построить периодограмму и вывести её значения. Анализ периодограммы можно выполнять непосредственно в программе *Statistica*. Отчётным материалом в данном случае является сама периодограмма или её ранжированные значения, отражающие циклические колебания (табл. 5.3).

Из таблицы видна наибольшая значимость годового, полугодового и сезонного периодов, а также примерно четырёхлетнего цикла. Наиболее важными с точки зрения дальнейшего анализа являются именно годовой период и четырёхлетние циклы.

*Таблица 5.3.* **Наиболее значимые характеристики циклических колебаний, полученные в ходе спектрального анализа Фурье**

| Частота | Период, мес. | Cos | Sin | Значение периодограммы | Спектральная плотность |
|---|---|---|---|---|---|
| 0,083333 | 12,0 | –45 | 73 | 2 597 967 | 1 172 917 |
| 0,166667 | 6,0 | –6 | –58 | 1 204 223 | 550 557 |
| 0,250000 | 4,0 | 27 | –7 | 273 571 | 125 250 |
| 0,021186 | 47,2 | –10 | –24 | 243 496 | 170 559 |
| 0,025424 | 39,3 | –23 | –8 | 211 114 | 104 021 |
| 0,019774 | 50,6 | –22 | 2 | 180 979 | 164 935 |

Как правило, анализ периодичности временного ряда не требуется, так как периодичность известна заранее. Для гидрологических и многих других величин периодичность составляет 12 мес. Зная периодичность временного ряда, можно проанализировать четыре составляющие временного ряда методом сезонной декомпозиции. Основная идея метода — отделить эти компоненты, т. е. разложить ряд на составляющую тренда, сезонную и тренд-циклическую компоненты, а также оставшуюся случайную составляющую. Приём, позволяющий выполнить такую декомпозицию, известен как метод *Census I.*

Основная идея сезонной декомпозиции проста. В общем случае временной ряд можно представить себе состоящим из четырёх различных компонент: 1) сезонной компоненты $S_t$ (где $t$ — момент времени); 2) тренда $T_t$; 3) циклической компоненты $C_t$; 4) случайной, нерегулярной компоненты, или флуктуации $I_t$. В методе *Census I* тренд и циклическую компоненту обычно объединяют в одну тренд-циклическую компоненту $TC_t$. Конкретные функциональные взаимосвязи между этими компонентами могут иметь самый разный вид.

Изменение временного ряда может быть выражено аддитивной и мультипликативной моделью, сезонной и тренд-циклической составляющей.

В аддитивном случае ряд будет иметь постоянные сезонные и тренд-циклические колебания, величина которых не зависит от общего уровня значений ряда; в мультипликативном — сезонные колебания будут меняться в зависимости от общего уровня значений ряда. То есть аддитивная модель подразумевает увеличение переменной на определённое число по отношению к предыдущему её значению, а в мультипликативном случае — увеличение переменной на определённый процент от предыдущего её значения. Таким образом, аддитивная модель имеет вид

$$X_t = TC_t + S_t + I_t,$$

а мультипликативная

$$X_t = TC_t\, S_t\, I_t.$$

Сезонная декомпозиция может быть использована для анализа по отдельности всех компонент ряда для выявления закономерностей и для прогнози-

рования временных рядов. При анализе трёх основных компонент ряда можно обратить внимание на нерегулярную компоненту. Анализ данной компоненты и факторов, на неё влияющих, может дать определённую информацию о ряде и облегчить прогнозирование.

Метод сезонной декомпозиции может применяться, как правило, для прогнозирования с заблаговременностью 1 мес., что относит прогнозы данной модели к долгосрочным. Суть при этом заключается в прогнозировании тренд-циклической компоненты с помощью авторегрессионной модели первого порядка, после чего к прогнозному значению прибавляется квазистационарная сезонная составляющая; нерегулярная компонента является ошибкой модели.

Для выполнения сезонной декомпозиции в программе *Statistica 12* необходимо предварительно определить тип модели и периодичность данных. С увеличением среднего значения повышение амплитуды колебаний уровней воды не наблюдается, поэтому можно предположить, что в данном случае ряд описывает аддитивная модель, период которой равен 12 мес. После определения типа модели оценивают сезонную декомпозицию. В программе *Statistica 12* выбирается пакет анализа временных рядов и процедура сезонной декомпозиции (метод *Census I*). После выполнения данной процедуры можно перейти к результатам анализа, которые будут представлены в таблице разложения исходного ряда на сезонную, тренд-циклическую и нерегулярную компоненту. Таким образом, мы получим две компоненты (сезонную и тренд-циклическую), значения которых могут быть вычислены с определённым упреждением. Заблаговременность такого прогноза будет зависеть от лага, на котором наблюдаются значимые коэффициенты корреляции для циклической компоненты.

Построим модель для прогнозирования среднемесячного уровня воды оз. Ильмень за период с 1960 по 2017 г. (2018 г. оставим для проверки на независимом материале) с заблаговременностью 1 мес. на основании уравнения линейной регрессии $TC_t$ от $TC_{t-1}$, которое имеет вид

$$TC_t = 0{,}94TC_{t-1} + 20.$$

Конечное уравнение для прогнозирования среднемесячного уровня воды имеет вид аддитивной модели без нерегулярной компоненты:

$$H_t = TC_t^* + S_t.$$

Сезонная компонента остаётся постоянной для всего ряда в настоящем и будущем, значения $TC_{t-1}$ пересчитываются для каждого нового данного прогноза на независимом материале (в будущем).

Стандартная ошибка прогнозирования представляет собой среднее квадратическое отклонение реальных значений уровней воды от прогнозируемых и рассчитывается по формуле

$$S = \sqrt{\frac{\sum_{i=1}^{n} (H_{\text{факт}} - H_{\text{прогнозное}})^2}{n}}.$$

Для ряда данных с 1960 по 2017 г. стандартная ошибка прогнозирования с заблаговременностью 1 мес. составила 39 см.

Качество выпускаемых долгосрочных прогнозов характеризуется отношением стандартной ошибки прогнозирования к стандартному отклонению случайной величины или случайного процесса. В данном случае прогнозируется случайный процесс, поэтому в качестве стандартного отклонения используется стандартное отклонение от математического ожидания случайного процесса, т. е. нормального гидрографа (12 математических ожиданий). Таким образом, стандартное отклонение составило 79 см, а $S/\sigma = 0{,}49$. Приемлемыми считаются прогнозы, для которых $S/\sigma < 0{,}8$. Однако качество выпускаемых прогнозов несколько понижается для независимой выборки, и для 2018 г. $S/\sigma = 0{,}54$. В качестве альтернативной модели также может быть применена простая модель авторегрессии первого порядка, хотя без сезонной составляющей данная модель является необоснованной. Для наглядности все прогнозы на независимом материале следует разместить на одном графике (рис. 5.4).

График показывает, что модель авторегрессии первого порядка без сезонной декомпозиции не описывает сезонную составляющую и уже поэтому непригодна для прогнозирования; в то же время аналогичная модель с сезонной декомпозицией, выполненной методом *Census 1*, достаточно точно описала сезонную составляющую, а ошибки прогнозирования данной модели значительно ниже, чем при ориентировании на математическое ожидание случайного про-

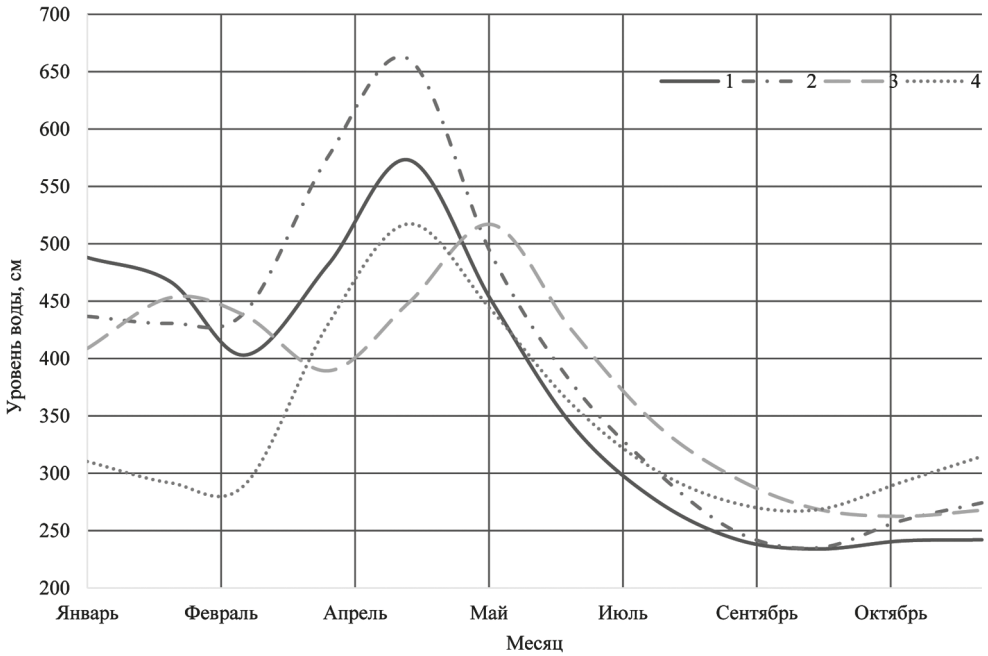



*Рис. 5.4.* Фактические (1) и прогнозные данные за 2018 г. при использовании метода сезонной декомпозиции (2), модели авторегрессии первого порядка (3) и средние многолетние (4) уровни воды оз. Ильмень (дер. Войцы)

цесса. Тем не менее модель *Census 1* не всегда удобна в использовании, так как требуется постоянный пересчёт тренд-циклической компоненты, прогнозирование же с использованием данной модели с заблаговременностью более 1 мес. нецелесообразно из-за существенного накопления ошибок.

Следующим шагом развития методов прогнозирования временных рядов являются методы АР (*AR*), СС (*MA*), АРСС (*ARMA*) и АРПСС. Модели АР, СС, АРСС, АРПСС в гидрологической практике используются, как правило, для прогнозирования среднемесячных уровней воды крупных озёр с упреждением от 1 до 12 мес. Методы теории АРСС разработаны и доведены до практического применения Дж. Боксом и Г. Дженкинсом [Бокс, Дженкинс, 1974]. Они позволяют не только описывать корреляционную и спектральную структуру временных рядов в терминах модельных процессов АРСС и отражать их статистическую взаимосвязь в терминах моделей передаточных функций, но и составлять прогноз, поскольку эти модели фактически являются прогнозирующей функцией. Обобщением модели АРСС на случай нестационарных временных рядов является модель авторегрессии — проинтегрированного скользящего среднего ($p$, $d$, $q$), где $p$ — параметр авторегрессии, $d$ — порядок операции взятия разностей, $q$ — параметр скользящего среднего. Обобщением модели АРПСС на случай сезонных нестационарных рядов является мультипликативная сезонная модель АРПСС ($p$, $d$, $q$) ($P_s$, $D_s$, $Q_s$), где к параметрам модели АРПСС ($p$, $d$, $q$) добавлены сезонные параметры: сезонный параметр авторегрессии — $P_s$, сезонная разность — $D_s$, сезонный параметр скользящего среднего — $Q_s$.

Для понимания модели АРПСС необходимо разобрать сущность процессов авторегрессии и скользящего среднего. Авторегрессионная модель — это модель временных рядов, в которой значения временного ряда в данный момент линейно зависят от предыдущих показателей этого же ряда. Как правило, значения частной автокорреляционной функции наиболее существенны на первом лаге и на лаге, равном сезонности, что визуально можно оценить по коррелограммам. Значимые частные коэффициенты корреляции на различных лагах позволяют оценить параметры $p$ и $P_s$. Согласно модели скользящего среднего, прогнозируемые члены ряда линейно зависят от текущего и прошлых значений, а также некоторого стохастического члена, который отражает вероятностный характер модели. Аналитически модель скользящего среднего для величины $X$ можно представить следующим образом:

$$X_t = \overline{X} + e_t + \beta_1 e_{t-1} + \beta_2 e_{t-2} + \beta_q e_{n-q},$$

где $X_t$ — прогнозируемая величина; $\overline{X}$ — среднее значение случайной величины; $e_t$ — белый шум в различные моменты времени; β — весовые коэффициенты.

Так как выборочные значения случайной величины $X$ нам известны, то данное уравнение может быть решено относительно величины белого шума. Весовые коэффициенты β находятся таким образом, чтобы обеспечивался минимум выражения $\min\limits_{\beta_1\beta_2} \sum e_j^2$. Соединение двух этих моделей даёт модель АРСС, основным недостатком которой является требование стационарности.

Взятие разности с различным лагом, как правило, приводят исходный ряд к стационарному виду. Полученная модель является моделью АРПСС (*p, d, q*) ($P_s$, $D_s$, $Q_s$).

Основная трудность в применении этих методов для решения практических задач гидрометеорологии заключается в идентификации модели для конкретного временного ряда, т. е. в подборе для него соответствующего модельного выражения, в терминах вероятностных характеристик которого осуществляется анализ либо прогнозирование. На практике каждый параметр модели АРПСС имеет порядок не больше второго, поэтому целесообразно проверить альтернативные модели и выбрать лучшую.

Параметры разности подбираются таким образом, чтобы выполнялось условие стационарности, как правило, это разности с лагом, кратным сезонности, и с лагом, равным единице. Затем можно определить начальные значения параметров *p*, *q* и $P_s$, $Q_s$. Большинство встречающихся на практике временных рядов можно с достаточной степенью точности аппроксимировать одной из пяти основных моделей, которые можно идентифицировать по виду автокорреляционной (АКФ) и частной автокорреляционной функции (ЧАКФ):

1. Один параметр авторегрессии (*p*): АКФ экспоненциально убывает; ЧАКФ имеет резко выделяющееся значение для лага 1, нет корреляций на других лагах.
2. Два параметра авторегрессии (*p*): АКФ имеет форму синусоиды или экспоненциально убывает; ЧАКФ имеет резко выделяющиеся значения на лагах 1, 2, нет корреляций на других лагах.
3. Один параметр скользящего среднего (*q*): АКФ имеет резко выделяющееся значение на лаге 1, нет корреляций на других лагах. ЧАКФ экспоненциально убывает.
4. Два параметра скользящего среднего (*q*): АКФ имеет резко выделяющиеся значения на лагах 1, 2, нет корреляций на других лагах. ЧАКФ имеет форму синусоиды или экспоненциально убывает.
5. Один параметр авторегрессии (*p*) и один параметр скользящего среднего (*q*): АКФ экспоненциально убывает с лага 1; ЧАКФ экспоненциально убывает с лага 1.

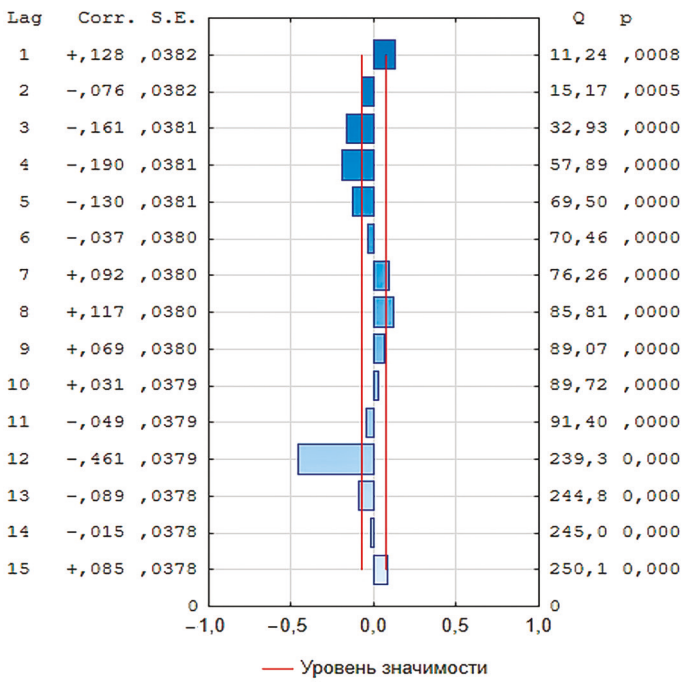
Функция частной автокорреляции при лаге *k* — это корреляция между рядами значений, отстоящих друг от друга на *k* интервалов со значениями интервалов в промежутке.

Прогнозирование с использованием модели АРПСС доступно в программе *Statistica* во вкладке «Временные ряды и прогнозирование» — АРПСС.

Для анализа количества параметров модели АРПСС необходимо провести анализ автокорреляционной и частной автокорреляционной функции, предварительно трансформировав ряд с разностями с лагами 1 и 12. Анализ автокорреляционной функции представить в отчёте (рис. 5.5).

Анализ графиков показывает существенные значения автокорреляционной и частной автокорреляционной функции на лагах 1 и 12, также наблю-

*а*




*б*

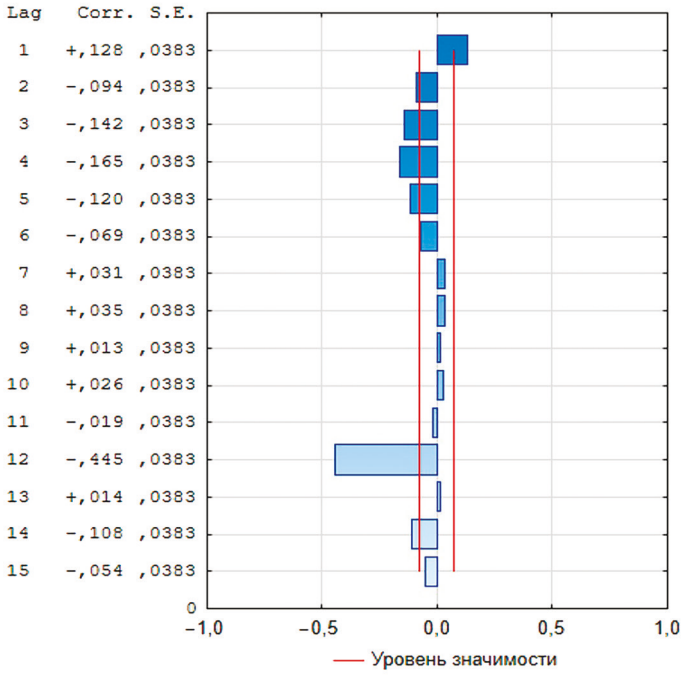



*Рис. 5.5.* График корреляционной (*а*) и частной корреляционной (*б*) функции ряда уровней воды после взятия разностей с лагами 1 и 12

даются значимые показатели на лагах 3 и 4, в целом обе функции имеют форму синусоиды. Таким образом, однозначно можно судить только о наличии сезонных параметров авторегрессии и скользящего среднего. Параметры $p$ и $q$ необходимо определить методом подбора так, чтобы порядок параметров оставался значимым, а ошибка модели была минимальна. Одна из лучших моделей для решения данной задачи может быть идентифицирована как АРПСС: (1,1,2); (2,1,2). Стандартная ошибка прогнозирования данной модели с упреждением в 1 мес. составила 49 см, отношение $S/\sigma = 0{,}62$, что близко к значению $S/\sigma$ при прогнозировании методом сезонной декомпозиции при проверке на независимом материале. Проверка обеих методик на независимом материале АРПСС даёт более устойчивые и лучшие результаты (рис. 5.6).

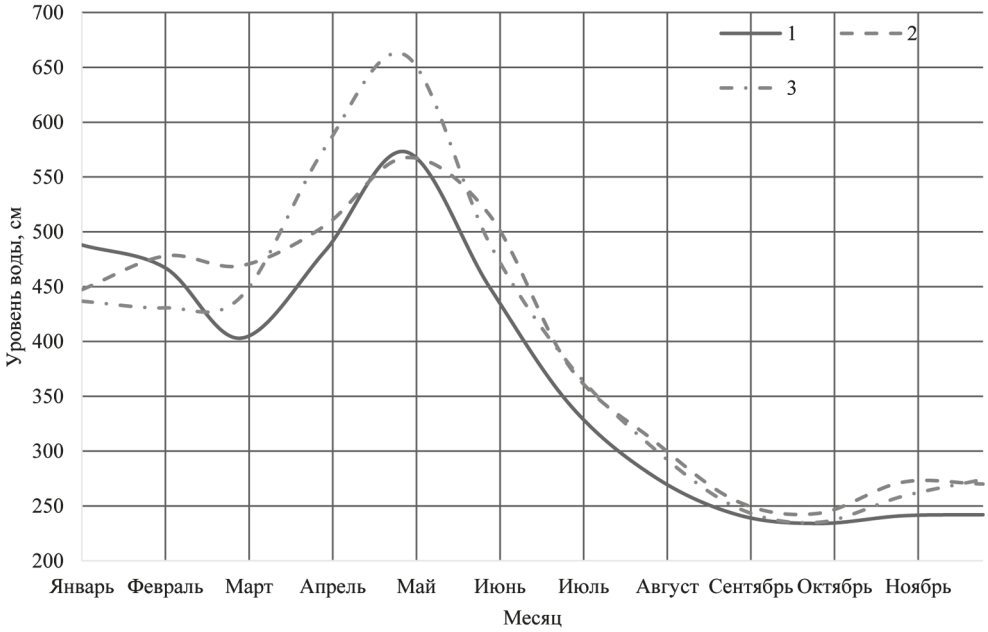



*Рис. 5.6.* Фактические (1) и прогностические данные (2018 г.) уровня воды оз. Ильмень (дер. Войцы) с использованием модели АРПСС (1,1,2; 2,1,2) (2) и метода сезонной декомпозиции (3)

Методом АРПСС можно прогнозировать с заблаговременностью более 1 мес. без существенного накопления ошибок прогнозирования, помимо прочего, АРПСС даёт намного более устойчивое решение и проще в расчётах. К недостаткам метода можно отнести: ручной подбор параметров модели, прогнозирование с минимальной дискретностью одна неделя, невозможность прогнозирования с суточной дискретностью и использования второстепенных предикторов. Тем не менее долгосрочное (стратегическое) прогнозирование методом АРПСС на сегодняшний день является общепризнанным.

Дальнейшее совершенствование и уменьшение ошибки прогнозирования возможно методом обучения искусственных нейронных сетей.

Рассмотрим прогнозирование временных рядов методом искусственных нейронных сетей в *Statistica 12*. Во вкладке искусственных нейронных сетей необходимо выбрать временные ряды — регрессию, а затем прогнозируемую переменную и задать количество наблюдений, используемых на входе в искусственную нейронную сеть. Как правило, используется количество наблюдений кратное сезонности, в данном случае — 12. Однако зачастую целесообразно брать большее число наблюдений, примерно равное цикличности; при этом количество скрытых нейронов стоит сократить для оптимальной аппроксимации. Цикличность в данном случае примерно 47 наблюдений, поэтому лучше использовать это количество наблюдений в качестве входных для прогнозирования с заблаговременностью 1 мес. и более. Обучив искусственные нейронные сети и выбрав лучшие из них, следует приступить к анализу ошибок. Так, стандартная ошибка лучшей искусственной нейронной сети *MLP* 50-4-1 составила 44 см, $S/\sigma = 0{,}55$, что несколько лучше, чем методом АРПСС. На независимом материале стандартная ошибка составила 48 см, что примерно равно стандартной ошибке при прогнозировании методом АРПСС. В 2018 г. ИНС также не показала превосходства в качестве выпускаемых прогнозов по сравнению с АРПСС: были получены схожие результаты (рис. 5.7).

Анализ графика показал, что модель достаточно хорошо описала сезонную составляющую и пригодна для прогнозирования. В целом прогнозирование с использованием ИНС является наиболее целесообразным и удобным способом, так как обеспечивает конкурентное качество выпускаемых прогнозов,

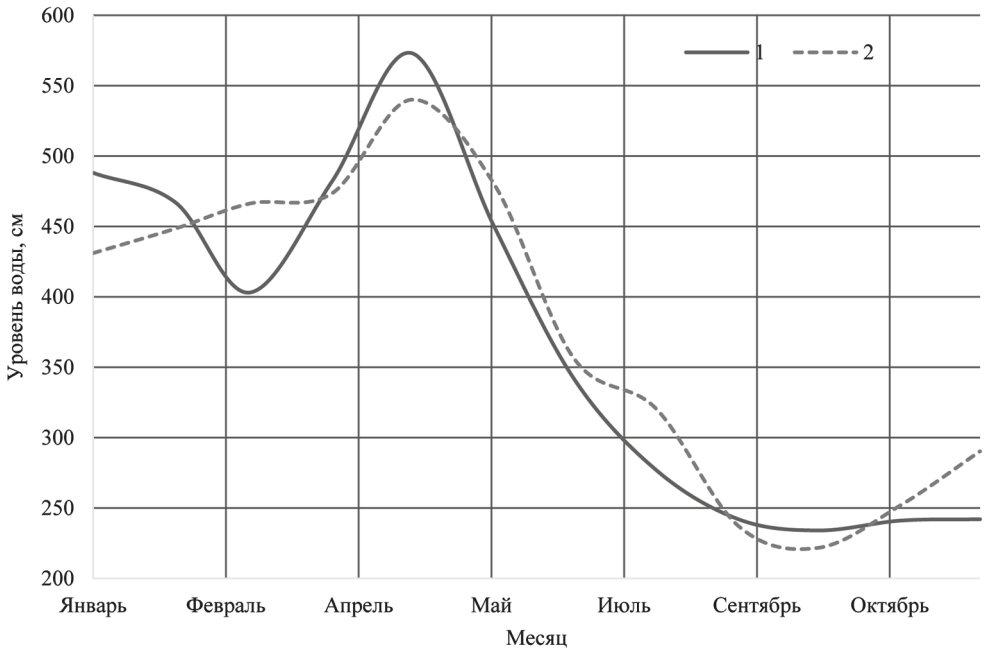



*Рис. 5.7.* Фактические (1) и прогностические данные (2018 г.) уровня воды оз. Ильмень (дер. Войцы) с использованием модели *MLP* 50-4-1 (2)

оптимизация же параметров обучения может проводиться полностью в автоматическом режиме, что минимизирует влияние человеческого фактора.

В заключение работы следует отразить основные результаты, полученные в ходе выполнения задания, использованные модели, их параметры и оценку точности. В частности, в настоящей работе проанализированы среднемесячные уровни воды оз. Ильмень за 1960–2018 гг. Установлены статистически значимые тренды, способствующие увеличению среднегодовых значений уровня воды, в первую очередь за счёт увеличения уровней за январь — март; на смещение фаз водного режима косвенно указывает сильная неоднородность рядов по дисперсии за март и ноябрь. Выявлены сезонность (12 мес.) и цикличность (примерно 47 мес.). Наиболее значимые для прогнозирования коэффициенты автокорреляции характерны для лага в 1 мес., что позволяет прогнозировать уровни воды озера с данной заблаговременностью. В работе рассматриваются три модели прогнозирования. Первая модель основана на принципе сезонной декомпозиции, вторая — на методе АРПСС, третья — на методе обучения искусственных нейронных сетей. Отмечается близость результатов, полученных двумя последними методами, при этом отношение стандартной ошибки прогнозирования к стандартному отклонению прогнозируемой величины для обеих моделей является приемлемым, что свидетельствует о возможности применения данных методов на практике для стратегического прогнозирования.

### *Контрольные вопросы*

1. Что такое случайный процесс?
2. Что такое временной ряд и каковы его отличия от случайной величины?
3. Что такое и как рассчитывается математическое ожидание и стандартное отклонение случайного процесса?
4. Что такое автокорреляционная и частная автокорреляционная функция?
5. Что такое и как определяется сезонность и цикличность?
6. Каковы принципы сезонной декомпозиции, как использовать данный метод для прогнозирования?
7. Что такое марковский случайный процесс?
8. Применение метода авторегрессии проинтегрированного скользящего среднего для прогнозирования периодически скоррелированных случайных процессов.
9. Каковы принципы определения параметров модели авторегрессии проинтегрированного скользящего среднего?
10. Искусственные нейронные сети для прогнозирования периодически скоррелированных случайных процессов и временных рядов.
11. Принципы определения параметров искусственной нейронной сети и их оптимизации.
12. Какие критерии применяются для оценки качества выпускаемых прогнозов?

# Список литературы


*Бокс Дж., Дженкинс Г.* Анализ временных рядов. Прогноз и управление. Кн. 1. / пер. с англ., под ред. В. Ф. Писаренко. М.: Мир, 1974.

*Бродская Н. А., Мякишева Н. В., Александрова К. В.* Оценка разномасштабного взаимодействия поверхностных и подземных вод // Учёные записки Российского государственного гидрометеорологического университета. 2015. № 38. С. 36.

*Бузин В. А.* Зажоры и заторы льда на реках России. СПб.: Гос. гидролог. ин-т, 2015.

*Бузин В. А.* Метод прогноза максимальных уровней воды при заторах льда на средних реках // Метеорология и гидрология. 2001. № 9. С. 84–89.

*Бузин В. А., Горошкова Н. И., Стриженок А. В., Палкина Д. А.* Зависимости для прогнозов максимальных заторных уровней воды Сухоны, Юга и Малой Северной Двины и влияние на них климатических и антропогенных факторов // Учёные записки Российского государственного гидрометеорологического университета. 2014. № 36. С. 12–21.

*Вентцель Е. С., Овчаров Л. А.* Теория вероятностей. М.: Наука, 1969.

История математики. Т. 1. С древнейших времён до начала Нового времени / под ред. А. П. Юшкевича. М.: Наука, 1970.

*Майстров Л. Е.* Теория вероятностей: исторический очерк. М.: Наука, 1967.

*Мак-Каллок У. С., Питтс В.* Логическое исчисление идей, относящихся к нервной активности. Архивная копия от 27 ноября 2007 на Wayback Machine // Автоматы / пер. с англ., под ред. К. Э. Шеннона, Дж. Маккарти. М.: Изд-во иностр. лит-ры, 1956. С. 363–384.

*Кузичев А. С.* Диаграммы Венна. История и применения. М.: Наука, 1968.

*Малинин В. Н.* Статистические методы анализа гидрометеорологической информации: учебник. СПб.: Рос. гос. гидметеороло. ун-т, 2008.

*Паскаль Б.* Мысли. М.: АСТ, 2020.

*Письменный Д. Т.* Конспект лекций по теории вероятностей и математической статистике. М.: Айрис-пресс, 2004.

*Попов Е. Г.* Основы гидрологических прогнозов. Л.: Гидрометеоиздат, 1968.

*Реньи А.* Трилогия о математике. М.: Мир, 1980.

*Рождественский А. В., Лобанов А. Г.* Методические рекомендации по определению расчётных гидрологических характеристик при наличии данных гидрометрических наблюдений. СПб.: Гос. гидролог. ин-т, 2009.

*Рожков В. А.* Теория и методы статистического оценивания вероятностных характеристик случайных величин и функций (с гидрометеорологическими примерами). Кн. 1. СПб.: Гидрометеоиздат, 2001.

*Рожков В. А.* Теория и методы статистического оценивания вероятностных характеристик случайных величин и функций (с гидрометеорологическими примерами). Кн. 2. СПб.: Гидрометеоиздат, 2002.

*Сикан А. В.* Вероятностные распределения в гидрологии. Специальные главы теории и практики гидрологических расчётов: учебник. СПб.: Рос. гос. гидметеороло г. ун-т, 2020.

*Сикан А. В.* Методы статистической обработки гидрометеорологической информации: учебник. СПб.: Рос. гос. гидметеороло г. ун-т, 2007.

*Сикан А. В.* Оптимизация параметров распределения при построении кривых обеспеченностей экстремальных расходов воды // Учёные записки Российского государственного гидрометеорологического университета. 2012. № 24. С. 26–32.

*Стройк Д. Я.* Краткий очерк истории математики. М.: Наука, 1969.

*Сумачёв А. Э., Банщикова Л. С.* Прогнозирование гидрологических характеристик с использованием нейронных сетей: труды III Всерос. конф. «Гидрометеорология и экология: достижения и перспективы развития». 2019. С. 812–815.

*Сумачёв А. Э., Банщикова Л. С.* Ледовый режим реки Печоры в современных климатических условиях и принципы прогнозирования высшего уровня воды за период весеннего ледохода // Успехи современного естествознания. 2021. № 10. С. 75–80.

*Сумачёв А. Э., Мякишева Н. В., Маргарян В. Г., Мисакян А. Э.* Долгосрочное прогнозирование уровней воды озера Ильмень с использованием вероятностных подходов // Естественные и технические науки. 2021. № 6 (157). С. 96–102.

*Hochreiter S., Schmidhuber J.* Long Short-Term Memory // Neural Computation. 1997. Vol. 9, no. 8. P. 1735–1780.

*Jeffreys H.* Scientific Inference, Third Edition. Cambridge: Cambridge University Press, 1973.

*Myakisheva N. V., Gaiducova E. V., Shanochkin S. V., Batmazova A. A.* Seasonal and Annual Forecasting of Water Levels in Large Lakes (Case Study of the Ladoga Lake) // International Letters of Natural Sciences. 2021. Vol. 82. P. 13–18.

*Sumachev A. E., Kuzmin V. A., Borodin E. S.* River flow forecasting using artificial neural networks // International Journal of Mechanical Engineering and Technology. 2018. Vol. 9, no. 10. P. 706–714.

# Приложения

*Приложение 1.* Значения статистики Фишера для различных уровней значимости и степеней свободы [Сикан, 2020]

| **2α = 0,10** | | **Число степеней свободы $\nu_1$ (для большей дисперсии)** | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 15 | 20 | 24 | 30 | 40 | 60 |
| Число степеней свободы $\nu_2$ (для меньшей дисперсии) | 1 | 161 | 200 | 216 | 225 | 230 | 234 | 237 | 239 | 271 | 242 | 243 | 244 | 246 | 248 | 249 | 250 | 251 | 252 |
| | 2 | 18,5 | 19,0 | 19,2 | 19,2 | 19,3 | 19,3 | 19,4 | 19,4 | 19,4 | 19,4 | 19,4 | 19,4 | 19,4 | 19,4 | 19,5 | 19,5 | 19,5 | 19,5 |
| | 3 | 10,10 | 9,55 | 9,28 | 9,12 | 9,01 | 8,94 | 8,89 | 8,85 | 8,81 | 8,79 | 8,76 | 8,74 | 8,70 | 8,66 | 8,64 | 8,62 | 8,59 | 8,57 |
| | 4 | 7,71 | 6,94 | 6,59 | 6,39 | 6,26 | 6,16 | 6,09 | 6,04 | 6,00 | 5,96 | 5,94 | 5,91 | 5,86 | 5,80 | 5,77 | 5,75 | 5,72 | 5,69 |
| | 5 | 6,61 | 5,79 | 5,41 | 5,19 | 5,05 | 4,95 | 4,88 | 4,82 | 4,77 | 4,74 | 4,71 | 4,68 | 4,62 | 4,56 | 4,53 | 4,50 | 4,46 | 4,43 |
| | 6 | 5,99 | 5,14 | 4,76 | 4,53 | 4,39 | 4,28 | 4,21 | 4,15 | 4,10 | 4,06 | 4,03 | 4,00 | 3,94 | 3,87 | 3,84 | 3,81 | 3,77 | 3,74 |
| | 7 | 5,59 | 4,74 | 4,35 | 4,12 | 3,97 | 3,87 | 3,79 | 3,73 | 3,68 | 3,64 | 3,60 | 3,57 | 3,51 | 3,44 | 3,41 | 3,38 | 3,34 | 3,30 |
| | 8 | 5,32 | 4,46 | 4,07 | 3,84 | 3,69 | 3,58 | 3,50 | 3,44 | 3,39 | 3,35 | 3,31 | 3,28 | 3,22 | 3,15 | 3,12 | 3,08 | 3,04 | 3,01 |
| | 9 | 5,12 | 4,26 | 3,86 | 3,63 | 3,48 | 3,37 | 3,29 | 3,23 | 3,18 | 3,14 | 3,10 | 3,07 | 3,01 | 2,94 | 2,90 | 2,86 | 2,83 | 2,79 |
| | 10 | 4,96 | 4,10 | 3,71 | 3,48 | 3,33 | 3,22 | 3,14 | 3,07 | 3,02 | 2,98 | 2,94 | 2,91 | 2,85 | 2,77 | 2,74 | 2,70 | 2,66 | 2,62 |
| | 11 | 4,84 | 3,98 | 3,59 | 3,36 | 3,20 | 3,09 | 3,01 | 2,95 | 2,90 | 2,85 | 2,82 | 2,79 | 2,72 | 2,65 | 2,61 | 2,57 | 2,53 | 2,49 |
| | 12 | 4,75 | 3,89 | 3,49 | 3,26 | 3,11 | 3,00 | 2,91 | 2,85 | 2,80 | 2,75 | 2,72 | 2,69 | 2,62 | 2,54 | 2,51 | 2,47 | 2,43 | 2,38 |
| | 13 | 4,67 | 3,81 | 3,41 | 3,18 | 3,03 | 2,92 | 2,83 | 2,77 | 2,71 | 2,67 | 2,63 | 2,60 | 2,53 | 2,46 | 2,42 | 2,38 | 2,34 | 2,30 |
| | 14 | 4,60 | 3,74 | 3,34 | 3,11 | 2,96 | 2,85 | 2,76 | 2,70 | 2,65 | 2,60 | 2,57 | 2,53 | 2,46 | 2,39 | 2,35 | 2,31 | 2,27 | 2,22 |
| | 15 | 4,54 | 3,68 | 3,29 | 3,06 | 2,90 | 2,79 | 2,71 | 2,64 | 2,59 | 2,54 | 2,51 | 2,48 | 3,52 | 3,37 | 3,29 | 3,21 | 3,13 | 3,05 |

*Продолжение приложения 1*

| **2α = 0,10** | | **Число степеней свободы $v_1$ (для большей дисперсии)** | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **11** | **12** | **15** | **20** | **24** | **30** | **40** | **60** |
| Число степеней свободы $v_2$ (для меньшей дисперсии) | 16 | 4,49 | 3,63 | 3,24 | 3,01 | 2,85 | 2,74 | 2,66 | 2,59 | 2,54 | 2,49 | 2,46 | 2,42 | 2,35 | 2,28 | 2,24 | 2,19 | 2,15 | 2,11 |
| | 17 | 4,45 | 3,59 | 3,20 | 2,96 | 2,81 | 2,70 | 2,61 | 2,55 | 2,49 | 2,45 | 2,41 | 2,38 | 2,31 | 2,23 | 2,19 | 2,15 | 2,10 | 2,06 |
| | 18 | 4,41 | 3,55 | 3,16 | 2,93 | 2,77 | 2,66 | 2,58 | 2,51 | 2,46 | 2,41 | 2,37 | 2,34 | 2,27 | 2,19 | 2,15 | 2,11 | 2,06 | 2,02 |
| | 19 | 4,38 | 3,52 | 3,13 | 2,90 | 2,74 | 2,63 | 2,54 | 2,48 | 2,42 | 2,38 | 2,34 | 2,31 | 2,23 | 2,16 | 2,11 | 2,07 | 2,03 | 1,98 |
| | 20 | 4,35 | 3,49 | 3,10 | 2,87 | 2,71 | 2,60 | 2,51 | 2,45 | 2,39 | 2,35 | 2,31 | 2,28 | 2,20 | 2,12 | 2,08 | 2,04 | 1,99 | 1,95 |
| | 22 | 4,30 | 3,44 | 3,05 | 2,82 | 2,66 | 2,55 | 2,46 | 2,40 | 2,34 | 2,30 | 2,26 | 2,23 | 2,15 | 2,07 | 2,03 | 1,98 | 1,94 | 1,89 |
| | 24 | 4,26 | 3,40 | 3,01 | 2,78 | 2,62 | 2,51 | 2,42 | 2,36 | 2,30 | 2,25 | 2,21 | 2,18 | 2,11 | 2,03 | 1,98 | 1,94 | 1,89 | 1,84 |
| | 26 | 4,23 | 3,37 | 2,98 | 2,74 | 2,59 | 2,47 | 2,39 | 2,32 | 2,27 | 2,22 | 2,18 | 2,15 | 2,07 | 1,99 | 1,95 | 1,90 | 1,85 | 1,80 |
| | 28 | 4,20 | 3,34 | 2,95 | 2,71 | 2,56 | 2,45 | 2,36 | 2,29 | 2,24 | 2,19 | 2,15 | 2,12 | 2,04 | 1,96 | 1,91 | 1,87 | 1,82 | 1,77 |
| | 30 | 4,17 | 3,32 | 2,92 | 2,69 | 2,53 | 2,42 | 2,33 | 2,27 | 2,21 | 2,16 | 2,13 | 2,09 | 2,01 | 1,93 | 1,89 | 1,84 | 1,79 | 1,74 |
| | 40 | 4,08 | 3,23 | 2,84 | 2,61 | 2,45 | 2,34 | 2,25 | 2,18 | 2,12 | 2,08 | 2,04 | 2,00 | 1,92 | 1,84 | 1,79 | 1,74 | 1,69 | 1,64 |
| | 60 | 4,00 | 3,15 | 2,76 | 2,53 | 2,37 | 2,25 | 2,17 | 2,10 | 2,04 | 1,99 | 1,95 | 1,92 | 1,84 | 1,75 | 1,70 | 1,65 | 1,59 | 1,53 |
| **2α = 0,20** | | **Число степеней свободы $v_1$ (для большей дисперсии)** | | | | | | | | | | | | | | | | | |
| | | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **11** | **12** | **15** | **20** | **24** | **30** | **40** | **60** |
| Число степеней свободы $v_2$ (для меньшей дисперсии) | 1 | 39,86 | 49,50 | 53,59 | 55,83 | 57,24 | 58,20 | 58,91 | 59,44 | 59,86 | 60,19 | 60,50 | 60,71 | 61,22 | 61,74 | 62,00 | 62,26 | 62,53 | 62,79 |
| | 2 | 8,53 | 9,00 | 9,16 | 9,24 | 9,29 | 9,33 | 9,35 | 9,37 | 9,38 | 9,39 | 9,40 | 9,41 | 9,42 | 9,44 | 9,45 | 9,46 | 9,47 | 9,47 |
| | 3 | 5,54 | 5,46 | 5,39 | 5,34 | 5,31 | 5,28 | 5,27 | 5,25 | 5,24 | 5,23 | 5,22 | 5,22 | 5,20 | 5,18 | 5,18 | 5,17 | 5,16 | 5,15 |
| | 4 | 4,54 | 4,32 | 4,19 | 4,11 | 4,05 | 4,01 | 3,98 | 3,95 | 3,94 | 3,92 | 3,91 | 3,90 | 3,87 | 3,84 | 3,83 | 3,82 | 3,80 | 3,79 |
| | 5 | 4,06 | 3,78 | 3,62 | 3,52 | 3,45 | 3,40 | 3,37 | 3,34 | 3,32 | 3,30 | 3,28 | 3,27 | 3,24 | 3,21 | 3,19 | 3,17 | 3,16 | 3,14 |
| | 6 | 3,78 | 3,46 | 3,29 | 3,18 | 3,11 | 3,05 | 3,01 | 2,98 | 2,96 | 2,94 | 2,92 | 2,90 | 2,87 | 2,84 | 2,82 | 2,80 | 2,78 | 2,76 |
| | 7 | 3,59 | 3,26 | 3,07 | 2,96 | 2,88 | 2,83 | 2,78 | 2,75 | 2,72 | 2,70 | 2,68 | 2,67 | 2,63 | 2,59 | 2,58 | 2,56 | 2,54 | 2,51 |

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Число степеней свободы $\nu_2$ (для меньшей дисперсии) | 8 | 3,46 | 3,11 | 2,92 | 2,81 | 2,73 | 2,67 | 2,62 | 2,59 | 2,56 | 2,54 | 2,52 | 2,50 | 2,46 | 2,42 | 2,40 | 2,38 | 2,36 | 2,34 |
| | 9 | 3,36 | 3,01 | 2,81 | 2,69 | 2,61 | 2,55 | 2,51 | 2,47 | 2,44 | 2,42 | 2,40 | 2,38 | 2,34 | 2,30 | 2,28 | 2,25 | 2,23 | 2,21 |
| | 10 | 3,29 | 2,92 | 2,73 | 2,61 | 2,52 | 2,46 | 2,41 | 2,38 | 2,35 | 2,32 | 2,30 | 2,28 | 2,24 | 2,20 | 2,18 | 2,16 | 2,13 | 2,11 |
| | 11 | 3,23 | 2,86 | 2,66 | 2,54 | 2,45 | 2,39 | 2,34 | 2,30 | 2,27 | 2,25 | 2,23 | 2,21 | 2,17 | 2,12 | 2,10 | 2,08 | 2,05 | 2,03 |
| | 12 | 3,18 | 2,81 | 2,61 | 2,48 | 2,39 | 2,33 | 2,28 | 2,24 | 2,21 | 2,19 | 2,17 | 2,15 | 2,10 | 2,06 | 2,04 | 2,01 | 1,99 | 1,96 |
| | 13 | 3,14 | 2,76 | 2,56 | 2,43 | 2,35 | 2,28 | 2,23 | 2,20 | 2,16 | 2,14 | 2,12 | 2,10 | 2,05 | 2,01 | 1,98 | 1,96 | 1,93 | 1,90 |
| | 14 | 3,10 | 2,73 | 2,52 | 2,39 | 2,31 | 2,24 | 2,19 | 2,15 | 2,12 | 2,10 | 2,08 | 2,05 | 2,01 | 1,96 | 1,94 | 1,91 | 1,89 | 1,86 |
| | 15 | 3,07 | 2,70 | 2,49 | 2,36 | 2,27 | 2,21 | 2,16 | 2,12 | 2,09 | 2,06 | 2,04 | 2,02 | 1,97 | 1,92 | 1,90 | 1,87 | 1,85 | 1,82 |
| | 16 | 3,05 | 2,67 | 2,46 | 2,33 | 2,24 | 2,18 | 2,13 | 2,09 | 2,06 | 2,03 | 2,01 | 1,99 | 1,94 | 1,89 | 1,87 | 1,84 | 1,81 | 1,78 |
| | 17 | 3,03 | 2,64 | 2,44 | 2,31 | 2,22 | 2,15 | 2,10 | 2,06 | 2,03 | 2,00 | 1,98 | 1,96 | 1,91 | 1,86 | 1,84 | 1,81 | 1,78 | 1,75 |
| | 18 | 3,01 | 2,62 | 2,42 | 2,29 | 2,20 | 2,13 | 2,08 | 2,04 | 2,00 | 1,98 | 1,96 | 1,93 | 1,89 | 1,84 | 1,81 | 1,78 | 1,75 | 1,72 |
| | 19 | 2,99 | 2,61 | 2,40 | 2,27 | 2,18 | 2,11 | 2,06 | 2,02 | 1,98 | 1,96 | 1,94 | 1,91 | 1,86 | 1,81 | 1,79 | 1,76 | 1,73 | 1,70 |
| | 20 | 2,97 | 2,59 | 2,38 | 2,25 | 2,16 | 2,09 | 2,04 | 2,00 | 1,96 | 1,94 | 1,92 | 1,89 | 1,84 | 1,79 | 1,77 | 1,74 | 1,71 | 1,68 |
| | 22 | 2,95 | 2,56 | 2,35 | 2,22 | 2,13 | 2,06 | 2,01 | 1,97 | 1,93 | 1,90 | 1,88 | 1,86 | 1,81 | 1,76 | 1,73 | 1,70 | 1,67 | 1,64 |
| | 24 | 2,93 | 2,54 | 2,33 | 2,19 | 2,10 | 2,04 | 1,98 | 1,94 | 1,91 | 1,88 | 1,85 | 1,83 | 1,78 | 1,73 | 1,70 | 1,67 | 1,64 | 1,61 |
| | 26 | 2,91 | 2,52 | 2,31 | 2,17 | 2,08 | 2,01 | 1,96 | 1,92 | 1,88 | 1,86 | 1,84 | 1,81 | 1,76 | 1,71 | 1,68 | 1,65 | 1,61 | 1,58 |
| | 28 | 2,89 | 2,50 | 2,29 | 2,16 | 2,06 | 2,00 | 1,94 | 1,90 | 1,87 | 1,84 | 1,81 | 1,79 | 1,74 | 1,69 | 1,66 | 1,63 | 1,59 | 1,56 |
| | 30 | 2,88 | 2,49 | 2,28 | 2,14 | 2,05 | 1,98 | 1,93 | 1,88 | 1,85 | 1,82 | 1,79 | 1,77 | 1,72 | 1,67 | 1,64 | 1,61 | 1,57 | 1,54 |
| | 40 | 2,84 | 2,44 | 2,23 | 2,09 | 2,00 | 1,93 | 1,87 | 1,83 | 1,79 | 1,76 | 1,73 | 1,71 | 1,66 | 1,61 | 1,57 | 1,54 | 1,51 | 1,47 |
| | 60 | 2,79 | 2,39 | 2,18 | 2,04 | 1,95 | 1,87 | 1,82 | 1,77 | 1,74 | 1,71 | 1,68 | 1,66 | 1,60 | 1,54 | 1,51 | 1,48 | 1,44 | 1,40 |

*Приложение 2. З*начения статистики Стьюдента для разных уровней значимости и чисел степеней свободы [Сикан, 2020]

| $\nu$ | Уровни значимости для одностороннего критерия (α) | | | | | |
|---|---|---|---|---|---|---|
| | 0,2 | 0,1 | 0,05 | 0,025 | 0,01 | 0,005 |
| | Уровни значимости для двухстороннего критерия (2α) | | | | | |
| | 0,4 | 0,2 | 0,1 | 0,05 | 0,02 | 0,01 |
| 1 | 1,38 | 3,08 | 6,31 | 12,71 | 31,82 | 63,66 |
| 2 | 1,06 | 1,89 | 2,92 | 4,30 | 6,96 | 9,92 |
| 3 | 0,98 | 1,64 | 2,35 | 3,18 | 4,54 | 5,84 |
| 4 | 0,94 | 1,53 | 2,13 | 2,78 | 3,75 | 4,60 |
| 5 | 0,92 | 1,48 | 2,02 | 2,57 | 3,36 | 4,03 |
| 6 | 0,91 | 1,44 | 1,94, | 2,45 | 3,14 | 3,71 |
| 7 | 0,90 | 1,41 | 1,89 | 2,36 | 3,00 | 3,50 |
| 8 | 0,89 | 1,40 | 1,86 | 2,31 | 2,90 | 3,36 |
| 9 | 0,88 | 1,38 | 1,83 | 2,26 | 2,82 | 3,25 |
| 10 | 0,88 | 1,37 | 1,81 | 2,23 | 2,76 | 3,17 |
| 11 | 0,88 | 1,36 | 1,80 | 2,20 | 2,72 | 3,11 |
| 12 | 0,87 | 1,36 | 1,78 | 2,18 | 2,68 | 3,05 |
| 13 | 0,87 | 1,35 | 1,77 | 2,16 | 2,65 | 3,01 |
| 14 | 0,87 | 1,34 | 1,76 | 2,14 | 2,62 | 2,98 |
| 15 | 0,87 | 1,34 | 1,75 | 2,13 | 2,60 | 2,95 |
| 16 | 0,87 | 1,34 | 1,75 | 2,12 | 2,58 | 2,92 |
| 17 | 0,86 | 1,33 | 1,74 | 2,11 | 2,57 | 2,90 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 18 | 0,86 | 1,33 | 1,73 | 2,10 | 2,55 | 2.88 |
| 19 | 0,86 | 1,33 | 1,73 | 2,09 | 2,54 | 2,86 |
| 20 | 0,86 | 1,33 | 1,72 | 2,09 | 2,53 | 2,85 |
| 21 | 0,86 | 1,32 | 1,72 | 2,08 | 2,52 | 2,83 |
| 22 | 0,86 | 1,32 | 1,72 | 2,07 | 2,51 | 2,82 |
| 23 | 0,86 | 1,32 | 1,71 | 2,07 | 2,50 | 2,81 |
| 24 | 0,86 | 1,32 | 1,71 | 2,06 | 2,49 | 2,80 |
| 25 | 0,86 | 1,32 | 1,71 | 2,06 | 2,49 | 2,79 |
| 26 | 0,86 | 1,32 | 1,71 | 2,06 | 2,48 | 2,78 |
| 27 | 0,86 | 1,31 | 1,70 | 2,05 | 2,47 | 2,77 |
| 28 | 0,86 | 1,31 | 1,70 | 2,05 | 2,47 | 2,76 |
| 29 | 0,85 | 1,31 | 1,70 | 2,05 | 2,46 | 2,76 |
| 30 | 0,85 | 1,31 | 1,70 | 2,04 | 2,46 | 2,75 |
| 40 | 0,85 | 1,30 | 1,68 | 2,02 | 2,42 | 2,70 |
| 60 | 0,85 | 1,30 | 1,67 | 2,00 | 2,39 | 2,66 |
| 120 | 0,84 | 1,29 | 1,67 | 1,98 | 2,36 | 2,62 |
| ∞ | 0,84 | 1,29 | 1,67 | 1,98 | 2,36 | 2,58 |

*Приложение 3.* **Ординаты кривых обеспеченности Крицкого — Менкеля (трёхпараметрического гамма-распределения) в модульных коэффициентах $K_p = f(C_v, C_s/C_v, P)$ [Сикан, 2020]**

| *P*, % | *Cv* | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | **0,1** | **0,2** | **0,3** | **0,4** | **0,5** | **0,6** | **0,7** | **0,8** | **0,9** | **1,0** |
| $C_s/C_v = 1{,}0$ | | | | | | | | | | |
| 0,001 | 1,44 | 1,94 | 2,46 | 2,97 | 3,47 | 3,95 | 4,35 | 4,72 | 5,02 | 5,30 |
| 0,01 | 1,40 | 1,81 | 2,25 | 2,70 | 3,15 | 3,57 | 3,91 | 4,31 | 4,63 | 4,91 |
| 0,03 | 1,36 | 1,74 | 2,15 | 2,56 | 2,97 | 3,37 | 3,74 | 4,11 | 4,44 | 4,72 |
| 0,05 | 1,34 | 1,71 | 2,11 | 2,49 | 2,89 | 3,27 | 3,62 | 3,98 | 4,30 | 4,60 |
| 0,1 | 1,32 | 1,67 | 2,03 | 2,39 | 2,77 | 3,14 | 3,48 | 3,82 | 4,13 | 4,44 |
| 0,3 | 1,29 | 1,59 | 1,90 | 2,23 | 2,55 | 2,89 | 3,21 | 3,53 | 3,85 | 4,17 |
| 0,5 | 1,27 | 1,55 | 1,84 | 2,15 | 2,45 | 2,76 | 3,06 | 3,37 | 3,68 | 4,00 |
| 1 | 1,24 | 1,49 | 1,75 | 2,03 | 2,31 | 2,59 | 2,87 | 3,15 | 3,45 | 3,78 |
| 3 | 1,19 | 1,39 | 1,59 | 1,81 | 2,03 | 2,27 | 2,51 | 2,75 | 3,02 | 3,32 |
| 5 | 1,17 | 1,34 | 1,52 | 1,70 | 1,90 | 2,10 | 2,31 | 2,52 | 2,76 | 3,04 |
| 10 | 1,13 | 1,26 | 1,39 | 1,53 | 1,68 | 1,83 | 1,99 | 2,16 | 2,35 | 2,57 |
| 20 | 1,08 | 1,17 | 1,25 | 1,34 | 1,42 | 1,51 | 1,59 | 1,69 | 1,78 | 1,88 |
| 25 | 1,06 | 1,13 | 1,19 | 1,26 | 1,33 | 1,41 | 1,47 | 1,52 | 1,58 | 1,62 |
| 30 | 1,05 | 1,10 | 1,15 | 1,20 | 1,24 | 1,29 | 1,34 | 1,38 | 1,40 | 1,39 |
| 40 | 1,02 | 1,04 | 1,06 | 1,08 | 1,09 | 1,10 | 1,10 | 1,10 | 1,05 | 0,99 |
| 50 | 1,00 | 0,99 | 0,99 | 0,97 | 0,96 | 0,93 | 0,89 | 0,83 | 0,76 | 0,67 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 60 | 0,97 | 0,94 | 0,90 | 0,87 | 0,83 | 0,79 | 0,71 | 0,61 | 0,51 | 0,40 |
| 70 | 0,95 | 0,89 | 0,83 | 0,77 | 0,70 | 0,62 | 0,51 | 0,41 | 0,30 | 0,21 |
| 75 | 0,93 | 0,86 | 0,78 | 0,71 | 0,62 | 0,53 | 0,42 | 0,31 | 0,21 | 0,14 |
| 80 | 0,91 | 0,83 | 0,74 | 0,65 | 0,55 | 0,45 | 0,35 | 0,24 | 0,15 | 0,09 |
| 90 | 0,88 | 0,75 | 0,63 | 0,50 | 0,38 | 0,26 | 0,17 | 0,09 | 0,04 | 0,02 |
| 95 | 0,84 | 0,68 | 0,53 | 0,38 | 0,26 | 0,15 | 0,08 | 0,01 | 0,01 | 0,00 |
| 97 | 0,82 | 0,64 | 0,48 | 0,33 | 0,21 | 0,11 | 0,05 | 0,02 | 0,00 | 0,00 |
| 99 | 0,78 | 0,57 | 0,38 | 0,23 | 0,12 | 0,05 | 0,01 | 0,00 | 0,00 | 0,00 |
| 99,5 | 0,76 | 0,53 | 0,31 | 0,18 | 0,09 | 0,03 | 0,00 | 0,00 | 0,00 | 0,00 |
| 99,7 | 0,74 | 0,50 | 0,31 | 0,15 | 0,07 | 0,02 | 0,00 | 0,00 | 0,00 | 0,00 |
| 99,9 | 0,70 | 0,45 | 0,25 | 0,11 | 0,04 | 0,01 | 0,00 | 0,00 | 0,00 | 0,00 |
| *P*, % | *Cv* | | | | | | | | | |
| | 1,1 | 1,2 | 1,3 | 1,4 | 1,5 | 1,6 | 1,7 | 1,8 | 1,9 | 2,0 |
| $C_s/C_v = 1{,}0$ | | | | | | | | | | |
| 0,01 | 5,16 | 5,34 | 5,46 | 5,58 | 5,68 | 5,76 | 5,82 | 5,88 | 5,92 | 5,96 |
| 0,1 | 4,69 | 4,92 | 5,06 | 5,18 | 5,29 | 5,37 | 5,44 | 5,49 | 5,54 | 5,58 |
| 0,3 | 4,44 | 4,74 | 4,92 | 5,06 | 5,16 | 5,24 | 5,31 | 5,36 | 5,42 | 5,46 |
| 0,5 | 4,29 | 4,58 | 4,75 | 4,91 | 5,02 | 5,11 | 5,18 | 5,24 | 5,28 | 5,32 |
| 1 | 4,06 | 4,36 | 4,55 | 4,72 | 4,84 | 4,94 | 5,00 | 5,07 | 5,12 | 5,16 |
| 3 | 3,59 | 3,92 | 4,14 | 4,33 | 4,46 | 4,58 | 4,68 | 4,76 | 4,84 | 4,92 |
| 5 | 3,31 | 3,63 | 3,84 | 4,02 | 4,16 | 4,28 | 4,40 | 4,50 | 4,60 | 4,69 |

*Продолжение приложения 3*

| $P$, % | $C_v$ | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | **1,1** | **1,2** | **1,3** | **1,4** | **1,5** | **1,6** | **1,7** | **1,8** | **1,9** | **2,0** |
| $C_s/C_v = 1,0$ | | | | | | | | | | |
| 10 | 2,78 | 3,03 | 3,26 | 3,46 | 3,64 | 3,81 | 3,94 | 4,05 | 4,15 | 4,25 |
| 20 | 2,00 | 2,10 | 2,20 | 2,32 | 2,44 | 2,56 | 2,67 | 2,80 | 2,92 | 3,03 |
| 25 | 1,68 | 1,69 | 1,70 | 1,70 | 1,68 | 1,66 | 1,61 | 1,56 | 1,51 | 1,46 |
| 30 | 1,39 | 1,34 | 1,26 | 1,17 | 1,07 | 0,96 | 0,84 | 0,72 | 0,60 | 0,45 |
| 40 | 0,916 | 0,808 | 0,72 | 0,60 | 0,50 | 0,38 | 0,28 | 0,20 | 0,11 | 0,04 |
| 50 | 0,559 | 0,446 | 0,34 | 0,26 | 0,20 | 0,15 | 0,105 | 0,07 | 0,04 | 0,01 |
| 60 | 0,306 | 0,216 | 0,19 | 0,13 | 0,10 | 0,075 | 0,055 | 0,035 | 0,015 | 0,000 |
| 70 | 0,141 | 0,085 | 0,06 | 0,045 | 0,035 | 0,025 | 0,015 | 0,010 | 0,000 | 0,000 |
| 75 | 0,086 | 0,046 | 0,025 | 0,020 | 0,010 | 0,005 | 0,000 | 0,000 | 0,000 | 0,000 |
| 80 | 0,047 | 0,023 | 0,015 | 0,005 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 |
| 90 | 0,007 | 0,002 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 |
| 95 | 0,001 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 |
| 97 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 |
| 99 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 |
| 99,5 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 |
| 99,7 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 |
| 99,9 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 |

| $P$, % | $C_v$ | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **0,1** | **0,2** | **0,3** | **0,4** | **0,5** | **0,6** | **0,7** | **0,8** | **0,9** | **1,0** | **1,1** | **1,2** |
| $C_s/C_v = 3,0$ | | | | | | | | | | | | |
| 0,001 | 1,50 | 2,28 | 3,35 | 4,69 | 6,30 | 8,21 | 10,4 | 12,9 | 15,5 | 18,3 | 21,3 | 24,6 |
| 0,01 | 1,42 | 2,06 | 2,86 | 3,78 | 5,00 | 6,28 | 7,70 | 9,21 | 11,0 | 12,9 | 14,8 | 16,9 |
| 0,03 | 1,39 | 1,99 | 2,62 | 3,41 | 4,34 | 5,48 | 6,59 | 7,74 | 9,14 | 10,6 | 12,2 | 13,8 |
| 0,05 | 1,36 | 1,88 | 2,50 | 3,23 | 4,10 | 5,06 | 6,07 | 7,11 | 8,32 | 9,66 | 11,0 | 12,4 |
| 0,1 | 1,35 | 1,80 | 2,36 | 3,00 | 3,75 | 4,58 | 5,43 | 6,31 | 7,33 | 8,43 | 9,54 | 10,7 |
| 0,3 | 1,31 | 1,69 | 2,12 | 2,64 | 3,22 | 3,82 | 4,44 | 5,11 | 5,84 | 6,62 | 7,40 | 8,21 |
| 0,5 | 1,29 | 1,63 | 2,02 | 2,48 | 3,00 | 3,50 | 4,00 | 4,58 | 5,21 | 5,85 | 6,50 | 7,16 |
| 1 | 1,25 | 1,55 | 1,88 | 2,25 | 2,66 | 3,07 | 3,49 | 3,92 | 4,40 | 4.88 | 5,37 | 5,85 |
| 3 | 1,21 | 1,42 | 1,67 | 1,91 | 2,17 | 2,42 | 2,70 | 2,44 | 3,22 | 3,47 | 3,74 | 3,99 |
| 5 | 1,17 | 1,36 | 1,54 | 1,75 | 1,94 | 2,14 | 2,35 | 2,51 | 2,70 | 2,89 | 3,05 | 3,23 |
| 10 | 1,14 | 1,26 | 1,39 | 1,52 | 1,63 | 1,76 | 1,87 | 1,97 | 2,09 | 2,15 | 2,24 | 2,31 |
| 20 | 1,09 | 1,16 | 1,23 | 1,29 | 1,33 | 1,38 | 1,42 | 1,45 | 1,47 | 1,49 | 1,49 | 1,59 |
| 25 | 1,07 | 1,12 | 1,17 | 1,21 | 1,23 | 1,26 | 1,27 | 1,29 | 1,28 | 1,28 | 1,27 | 1,27 |
| 30 | 1,05 | 1,09 | 1,12 | 1,14 | 1,15 | 1,15 | 1,16 | 1,15 | 1,14 | 1,13 | 1,11 | 1,08 |
| 40 | 1,02 | 1,03 | 1,03 | 1,03 | 1,01 | 1,00 | 0,97 | 0,95 | 0,91 | 0,88 | 0,85 | 0,81 |
| 50 | 0,99 | 0,98 | 0,96 | 0,93 | 0,90 | 0,86 | 0,82 | 0,78 | 0,74 | 0,70 | 0,66 | 0,61 |
| 60 | 0,97 | 0,93 | 0,89 | 0,84 | 0,79 | 0,74 | 0,69 | 0,65 | 0,60 | 0,55 | 0,50 | 0,46 |
| 70 | 0,94 | 0,88 | 0,82 | 0,77 | 0,71 | 0,66 | 0,61 | 0,56 | 0,52 | 0,48 | 0,41 | 0,41 |

*Продолжение приложения 3*

| $P$, % | $C_v$ | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **0,1** | **0,2** | **0,3** | **0,4** | **0,5** | **0,6** | **0,7** | **0,8** | **0,9** | **1,0** | **1,1** | **1,2** |
| $C_s/C_v = 3,0$ | | | | | | | | | | | | |
| 75 | 0,93 | 0,86 | 0,79 | 0.72 | 0,65 | 0,58 | 0,52 | 0,47 | 0,41 | 0,36 | 0,31 | 0,27 |
| 80 | 0,91 | 0,83 | 0,75 | 0,67 | 0,60 | 0,53 | 0,47 | 0,41 | 0,36 | 0,31 | 0,26 | 0,22 |
| 90 | 0,87 | 0,76 | 0,66 | 0,57 | 0,48 | 0,41 | 0,34 | 0,29 | 0,24 | 0,19 | 0,16 | 0,12 |
| 95 | 0,84 | 0,71 | 0,59 | 0,49 | 0,41 | 0,33 | 0,26 | 0,21 | 0,17 | 0,13 | 0,10 | 0,07 |
| 97 | 0,83 | 0,68 | 0,55 | 0,45 | 0,36 | 0,28 | 0,22 | 0,17 | 0,13 | 0,10 | 0,07 | 0,05 |
| 99 | 0,79 | 0,62 | 0,48 | 0,37 | 0,29 | 0,21 | 0,16 | 0,12 | 0,08 | 0,06 | 0,04 | 0,03 |
| 99,5 | 0,77 | 0,59 | 0,45 | 0,34 | 0,25 | 0,18 | 0,12 | 0,09 | 0,06 | 0,04 | 0,03 | 0,01 |
| 99,7 | 0,76 | 0,57 | 0,43 | 0,31 | 0,23 | 0,16 | 0,12 | 0,08 | 0,05 | 0,03 | 0,02 | 0,01 |
| 99,9 | 0,73 | 0,53 | 0,38 | 0,27 | 0,19 | 0,13 | 0,09 | 0.06 | 0,03 | 0,02 | 0,01 | 0,01 |

| $P$, % | $C_v$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | **1,3** | **1,4** | **1,5** | **1,6** | **1,7** | **1,8** | **1,9** | **2,0** |
| $C_s/C_v = 3,0$ | | | | | | | | |
| 0,01 | 19,0 | 21,2 | 23,5 | 25,9 | 28,4 | 31,0 | 33,7 | 36,5 |
| 0,1 | 11,8 | 13,0 | 14,2 | 15,4 | 16,7 | 18,0 | 19,4 | 20,8 |
| 0,3 | 8,94 | 9,75 | 10,6 | 11,4 | 12,3 | 13,1 | 14,0 | 14,8 |
| 0,5 | 7,75 | 8,41 | 9,07 | 9,74 | 10,4 | 11,1 | 11,8 | 12,4 |
| 1 | 6,26 | 6,74 | 7,21 | 7,68 | 8,14 | 8,61 | 9,07 | 9,53 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 3 | 4,20 | 4,44 | 4,67 | 4,89 | 5,10 | 5,31 | 5,51 | 5,70 |
| 5 | 3,37 | 3,52 | 3,66 | 3,80 | 3,92 | 4,04 | 4,15 | 4,26 |
| 10 | 2,36 | 2,42 | 2,47 | 2,51 | 2,55 | 2,58 | 2,60 | 2,62 |
| 20 | 1,50 | 1,49 | 1,48 | 1,46 | 1,45 | 1,42 | 1,40 | 1,37 |
| 25 | 1,25 | 1,23 | 1,20 | 1,18 | 1,15 | 1,12 | 1,08 | 1,05 |
| 30 | 1,06 | 1,03 | 0,997 | 0,964 | 0,929 | 0,892 | 0,855 | 0,818 |
| 40 | 0,775 | 0,736 | 0,697 | 0,659 | 0,620 | 0,581 | 0,544 | 0,507 |
| 50 | 0,572 | 0,531 | 0,491 | 0,452 | 0,415 | 0,379 | 0,345 | 0,313 |
| 60 | 0,417 | 0,337 | 0,339 | 0,304 | 0,271 | 0,240 | 0,212 | 0,186 |
| 70 | 0,293 | 0,257 | 0,224 | 0,194 | 0,166 | 0,142 | 0,121 | 0,102 |
| 75 | 0,239 | 0,206 | 0,176 | 0,149 | 0,125 | 0,105 | 0,087 | 0,071 |
| 80 | 0,190 | 0,160 | 0,133 | 0,110 | 0,090 | 0,073 | 0,059 | 0,047 |
| 90 | 0,100 | 0,078 | 0,061 | 0,047 | 0,035 | 0,026 | 0,019 | 0,014 |
| 95 | 0,057 | 0,042 | 0,030 | 0,022 | 0,015 | 0,010 | 0,007 | 0,004 |
| 97 | 0,038 | 0,027 | 0,018 | 0,012 | 0,008 | 0,005 | 0,003 | 0,002 |
| 99 | 0,017 | 0,011 | 0,007 | 0,004 | 0,002 | 0,001 | 0,001 | 0,000 |
| 99,5 | 0,011 | 0,006 | 0,004 | 0,002 | 0,001 | 0,000 | 0,000 | 0,000 |
| 99,7 | 0,008 | 0,004 | 0,002 | 0,001 | 0,000 | 0,000 | 0,000 | 0,000 |
| 99,9 | 0,004 | 0,002 | 0,001 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 |

*Продолжение приложения 3*

| $P$, % | $C_v$ | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **0,1** | **0,2** | **0,3** | **0,4** | **0,5** | **0,6** | **0,7** | **0,8** | **0,9** | **1,0** | **1,1** | **1,2** |
| $C_s/C_v$ = **4,0** | | | | | | | | | | | | |
| 0,001 | 1,58 | 2,50 | 3,82 | 5,60 | 8,10 | 11,0 | 14,2 | 17,5 | 20,6 | 24,0 | 27,5 | 32,9 |
| 0,01 | 1,51 | 2,20 | 3,15 | 4,35 | 5,90 | 7,70 | 9,57 | 11,4 | 13,6 | 15,6 | 17,6 | 20,7 |
| 0,03 | 1,45 | 2,05 | 2,87 | 3,85 | 5,05 | 6,35 | 7,81 | 9,15 | 10,7 | 12,2 | 13,7 | 16,0 |
| 0,05 | 1,40 | 1,97 | 2,72 | 3,60 | 4,70 | 5.75 | 7,00 | 8,20 | 9,46 | 10,9 | 12,10 | 14,0 |
| 0,1 | 1,38 | 1,87 | 2,53 | 3,29 | 4,20 | 5,07 | 6,05 | 7,02 | 8,12 | 9,25 | 10,4 | 11,6 |
| 0,3 | 1,34 | 1,73 | 2,23 | 2,81 | 3,45 | 4,09 | 4,76 | 5,46 | 6,18 | 6,94 | 7,71 | 8,53 |
| 0,5 | 1,30 | 1,67 | 2,10 | 2,60 | 3,13 | 3,69 | 4,25 | 4,81 | 5,38 | 6,02 | 6,65 | 7,31 |
| 1 | 1,25 | 1,58 | 1,94 | 2,34 | 2,77 | 3,17 | 3,59 | 4,01 | 4,43 | 4,90 | 5,35 | 5,82 |
| 3 | 1,19 | 1,43 | 1,67 | 1,92 | 2,18 | 2,44 | 2,67 | 2,90 | 3,12 | 3,35 | 3,60 | 3,84 |
| 5 | 1,17 | 1,36 | 1,55 | 1,75 | 1,93 | 2,11 | 2,28 | 2,45 | 2,60 | 2,77 | 2,92 | 3,07 |
| 10 | 1,11 | 1,26 | 1,38 | 1,51 | 1,61 | 1,72 | 1,82 | 1,90 | 2,00 | 2,05 | 2,12 | 2,18 |
| 20 | 1,08 | 1,15 | 1,21 | 1,26 | 1.31 | 1,34 | 1,37 | 1,40 | 1,41 | 1,42 | 1,43 | 1,43 |
| 25 | 1,06 | 1,11 | 1,15 | 1,19 | 1,21 | 1,23 | 1,23 | 1,24 | 1,25 | 1,21 | 1,24 | 1,22 |
| 30 | 1,05 | 1,08 | 1,10 | 1,12 | 1,13 | 1,13 | 1,12 | 1,12 | 1,10 | 1,09 | 1,07 | 1,06 |
| 40 | 1,02 | 0,03 | 1,02 | 1,01 | 0,99 | 0,97 | 0,95 | 0,93 | 0,90 | 0,87 | 0,85 | 0,81 |
| 50 | 0,99 | 0,98 | 0,95 | 0,92 | 0,89 | 0,85 | 0,82 | 0,78 | 0,75 | 0,71 | 0,67 | 0,63 |
| 60 | 0,97 | 0,93 | 0,89 | 0,84 | 0,79 | 0,75 | 0,70 | 0,66 | 0,62 | 0,57 | 0,53 | 0,49 |
| 70 | 0,94 | 0,88 | 0,82 | 0,76 | 0,71 | 0,65 | 0,60 | 0,55 | 0,60 | 0,46 | 0,42 | 0,38 |
| 75 | 0,93 | 0,86 | 0,79 | 0,72 | 0,66 | 0,60 | 0,55 | 0,50 | 0,45 | 0,40 | 0,37 | 0,32 |
| 80 | 0,91 | 0,83 | 0,75 | 0,68 | 0,61 | 0,55 | 0,50 | 0,45 | 0,40 | 0,36 | 0,31 | 0,27 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 90 | 0,88 | 0,77 | 0,67 | 0,59 | 0,51 | 0,44 | 0,38 | 0,33 | 0,29 | 0,25 | 0,21 | 0,18 |
| 95 | 0,85 | 0,72 | 0,61 | 0,52 | 0,44 | 0,37 | 0,32 | 0,26 | 0,22 | 0,18 | 0,15 | 0,12 |
| 97 | 0,83 | 0,69 | 0,58 | 0,48 | 0,40 | 0,33 | 0,27 | 0,23 | 0,18 | 0,15 | 0,12 | 0,10 |
| 99 | 0,80 | 0,64 | 0,52 | 0,42 | 0,34 | 0,27 | 0,22 | 0,17 | 0,14 | 0,11 | 0,08 | 0,06 |
| 99,5 | 0,78 | 0,61 | 0,49 | 0,39 | 0,30 | 0,24 | 0,19 | 0,15 | 0,11 | 0,08 | 0,06 | 0,05 |
| 99,7 | 0,77 | 0,60 | 0,47 | 0,37 | 0,29 | 0,22 | 0,17 | 0,13 | 0,10 | 0,07 | 0,05 | 0,04 |
| 99.9 | 0,75 | 0,56 | 0,43 | 0,33 | 0,25 | 0,19 | 0,11 | 0,10 | 0,08 | 0,05 | 0,04 | 0,03 |

| ***P*, %** | ***C*$_v$** | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | **1,3** | **1,4** | **1,5** | **1,6** | **1,7** | **1,8** | **1,9** | **2,0** |
| | ***C*$_s$/*C*$_v$ = 4,0** | | | | | | | |
| 0,01 | 22,8 | 25,4 | 28,0 | 30,8 | 33,6 | 36,5 | 39,4 | 42,4 |
| 0,1 | 12,8 | 14,0 | 15,3 | 16,6 | 17,9 | 19,2 | 20,6 | 21,9 |
| 0,3 | 9,31 | 10,1 | 10,9 | 11,7 | 12,5 | 13,3 | 14,2 | 15,0 |
| 0,5 | 7,91 | 8,53 | 9,16 | 9,79 | 10,4 | 11,0 | 11,7 | 12,3 |
| 1 | 6,22 | 6,66 | 7,09 | 7,52 | 7,95 | 8,73 | 8,78 | 9,19 |
| 3 | 4,04 | 4,25 | 4,45 | 4,64 | 4,83 | 5,01 | 5,18 | 5,34 |
| 5 | 3,21 | 3,34 | 3,46 | 3,57 | 3,68 | 3,78 | 3,87 | 3,96 |
| 10 | 2,24 | 2,28 | 2,32 | 2,36 | 2,39 | 2,42 | 2,4 | 2,45 |
| 20 | 1,43 | 1,43 | 1,42 | 1,41 | 1,39 | 1,38 | 1,36 | 1,33 |
| 25 | 1,21 | 1,19 | 1,17 | 1,15 | 1,13 | 1,10 | 1,08 | 1,05 |
| 30 | 1,04 | 1,01 | 0,985 | 0,958 | 0,929 | 0,90 | 0,871 | 0,841 |
| 40 | 0,781 | 0,748 | 0,716 | 0,684 | 0,652 | 0,62 | 0,588 | 0,558 |
| 50 | 0,598 | 0,562 | 0,529 | 0,495 | 0,464 | 0,433 | 0,403 | 0,375 |
| 60 | 0,457 | 0,421 | 0,388 | 0,356 | 0,327 | 0,299 | 0,273 | 0,249 |
| 70 | 0,341 | 0,308 | 0,227 | 0,248 | 0,223 | 0,199 | 0,177 | 0,157 |

*Продолжение приложения 3*

| $P$, % | $C_v$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | **1,3** | **1,4** | **1,5** | **1,6** | **1,7** | **1,8** | **1,9** | **2,0** |
| **$C_s/C_v = 4{,}0$** | | | | | | | | |
| 75 | 0,290 | 0,258 | 0,230 | 0,203 | 0,179 | 0,158 | 0,139 | 0,121 |
| 80 | 0,242 | 0,212 | 0,185 | 0,162 | 0,140 | 0,122 | 0,105 | 0,090 |
| 90 | 0,148 | 0,125 | 0,104 | 0,087 | 0,072 | 0,060 | 0,049 | 0,040 |
| 95 | 0,098 | 0,080 | 0,064 | 0,051 | 0,041 | 0,032 | 0,025 | 0,019 |
| 97 | 0,075 | 0,059 | 0,046 | 0,036 | 0,028 | 0,021 | 0,016 | 0,012 |
| 99 | 0,045 | 0,034 | 0,025 | 0,018 | 0,013 | 0,009 | 0,006 | 0,004 |
| 99,5 | 0,034 | 0,024 | 0,017 | 0,012 | 0,008 | 0,006 | 0,004 | 0,003 |
| 99,7 | 0,027 | 0,019 | 0,013 | 0,009 | 0,006 | 0,004 | 0,003 | 0,002 |
| 99.9 | 0,018 | 0,012 | 0,008 | 0,005 | 0,003 | 0,002 | 0,001 | 0,001 |

| $P$, % | $C_v$ | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **0,1** | **0,2** | **0,3** | **0,4** | **0,5** | **0,6** | **0,7** | **0,8** | **0,9** | **1,0** | **1,1** | **1,2** |
| **$C_s/C_v = 5{,}0$** | | | | | | | | | | | | |
| 0,001 | 1,67 | 2,75 | 4,38 | 6,87 | 9,90 | 13,4 | 17,0 | 21,2 | 25,3 | 28,5 | 33,8 | 38,2 |
| 0,01 | 1,54 | 2,34 | 3,43 | 4,91 | 6,65 | 8,70 | 10,7 | 12,7 | 15,0 | 17,4 | 20,0 | 22,7 |
| 0,03 | 1,47 | 2,15 | 3,07 | 4,23 | 5,50 | 6,95 | 8,43 | 9,96 | 11,6 | 13,3 | 15,2 | 17,2 |
| 0,05 | 1,43 | 2,06 | 2,87 | 3,90 | 5,05 | 6,24 | 7,51 | 8,82 | 10,2 | 11,7 | 13,2 | 14,8 |
| 0,1 | 1,40 | 1,95 | 2,66 | 3,51 | 4,44 | 5,40 | 6,43 | 7,54 | 8,64 | 9,83 | 11,0 | 12,1 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0,3 | 1,34 | 1,78 | 2,31 | 2,92 | 3,52 | 4,22 | 4,91 | 5,69 | 6,41 | 7,15 | 7,90 | 8,63 |
| 0,5 | 1,31 | 1,70 | 2,16 | 2,69 | 3,21 | 3,77 | 4,34 | 4,93 | 5,52 | 6,17 | 6,85 | 7,35 |
| 1 | 1,27 | 1,61 | 1,98 | 2,38 | 2,79 | 3,21 | 3,65 | 4,06 | 4,50 | 4,94 | 5,33 | 5,75 |
| 3 | 1,20 | 1,44 | 1,67 | 1,93 | 2,17 | 2,42 | 2,62 | 2,88 | 3,10 | 3,33 | 3,52 | 3,75 |
| 5 | 1,17 | 1,36 | 1,55 | 1,74 | 1,90 | 2,08 | 2,22 | 2,42 | 2,54 | 2,71 | 2,85 | 2,98 |
| 10 | 1,13 | 1,26 | 1,37 | 1,49 | 1,60 | 1,70 | 1,79 | 1,86 | 1,94 | 2,00 | 2,05 | 2,11 |
| 20 | 1,08 | 1,15 | 1,21 | 1,25 | 1,30 | 1,32 | 1,34 | 1,36 | 1,36 | 1,39 | 1,40 | 1,41 |
| 25 | 1,06 | 1,11 | 1,15 | 1,17 | 1,20 | 1,20 | 1,20 | 1,22 | 1,22 | 1,22 | 1,20 | 1,20 |
| 30 | 1,05 | 1,08 | 1,09 | 1,10 | 1,10 | 1,11 | 1,10 | 1,10 | 1,09 | 1,08 | 1,06 | 1,04 |
| 40 | 1,02 | 1,02 | 1,01 | 1,00 | 0,98 | 0,97 | 0,94 | 0,92 | 0,90 | 0,87 | 0,84 | 0,81 |
| 50 | 0,99 | 0,97 | 0,94 | 0,92 | 0,88 | 0,85 | 0,82 | 0,78 | 0,75 | 0,71 | 0,68 | 0,65 |
| 60 | 0,97 | 0,93 | 0,88 | 0,84 | 0,79 | 0,75 | 0,71 | 0,67 | 0,63 | 0,58 | 0,55 | 0,51 |
| 70 | 0,94 | 0,88 | 0,82 | 0,77 | 0,71 | 0,66 | 0,61 | 0,56 | 0,52 | 0,48 | 0,41 | 0,41 |
| 75 | 0,93 | 0,86 | 0,79 | 0,73 | 0,67 | 0,62 | 0,56 | 0,51 | 0,47 | 0,42 | 0,39 | 0,36 |
| 80 | 0,91 | 0,73 | 0,85 | 0,69 | 0,63 | 0,57 | 0,52 | 0,47 | 0,42 | 0,37 | 0,34 | 0,31 |
| 90 | 0,88 | 0,77 | 0,68 | 0,61 | 0,53 | 0,47 | 0,41 | 0,36 | 0,32 | 0,27 | 0,24 | 0,21 |
| 95 | 0,84 | 0,73 | 0,63 | 0,55 | 0,47 | 0,40 | 0,34 | 0,29 | 0,25 | 0,21 | 0,18 | 0,15 |
| 97 | 0,82 | 0,70 | 0,60 | 0,51 | 0,43 | 0,36 | 0,31 | 0,26 | 0,22 | 0,18 | 0,15 | 0,12 |
| 99 | 0,78 | 0,66 | 0,55 | 0,45 | 0,37 | 0,31 | 0,25 | 0,20 | 0,16 | 0,13 | 0,10 | 0,08 |
| 99,5 | 0,76 | 0,63 | 0,52 | 0,42 | 0,34 | 0,28 | 0,23 | 0,18 | 0,14 | 0,11 | 0,09 | 0,07 |
| 99,7 | 0,75 | 0,62 | 0,51 | 0,41 | 0,32 | 0,26 | 0,21 | 0,16 | 0,12 | 0,10 | 0,08 | 0,06 |
| 99,9 | 0,73 | 0,59 | 0,47 | 0,37 | 0,29 | 0,23 | 0,18 | 0,14 | 0,10 | 0,08 | 0,06 | 0,04 |

| *P*, % | $C_v$ | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **0,1** | **0,2** | **0,3** | **0,4** | **0,5** | **0,6** | **0,7** | **0,8** | **0,9** | **1,0** | **1,1** | **1,2** |
| $C_s/C_v = 6,0$ | | | | | | | | | | | | |
| 0,001 | 1,80 | 3,02 | 5,20 | 8,10 | 11,50 | 15,30 | 19,30 | 23,80 | 28,00 | 32,20 | 36,70 | 41,50 |
| 0,01 | 1,60 | 2,48 | 3,75 | 5,48 | 7,30 | 9,39 | 11,50 | 13,80 | 16,40 | 18,90 | 21,50 | 24,00 |
| 0,03 | 1,52 | 2,25 | 3,25 | 4,54 | 5,90 | 7,37 | 8,90 | 10,53 | 12,30 | 14,10 | 16,00 | 17,90 |
| 0,05 | 1,47 | 2,15 | 3,05 | 4,15 | 5,25 | 6,57 | 7,85 | 9,24 | 10,70 | 12,10 | 13,70 | 15,40 |
| 0,1 | 1,41 | 2,02 | 2,80 | 3,68 | 4,58 | 5,54 | 6,57 | 7,63 | 8,79 | 10,00 | 11,18 | 12,39 |
| 0,3 | 1,35 | 1,83 | 2,38 | 2,98 | 3,64 | 4,31 | 5,00 | 5,66 | 6,38 | 7,16 | 7,90 | 8,67 |
| 0,5 | 1,32 | 1,74 | 2,22 | 2,73 | 3,26 | 3,82 | 4,38 | 4,93 | 5,51 | 6,11 | 6,71 | 7,31 |
| 1 | 1,29 | 1,63 | 2,01 | 2,40 | 2,81 | 3,22 | 3,63 | 4,03 | 4,44 | 4,86 | 5,27 | 5,69 |
| 3 | 1,21 | 1,45 | 1,68 | 1,92 | 2,14 | 2,38 | 2,60 | 2,82 | 3,04 | 3,26 | 3,46 | 3,67 |
| 5 | 1,18 | 1,37 | 1,55 | 1,73 | 1,89 | 2,05 | 2,20 | 2,36 | 2,81 | 2,66 | 2,80 | 2,90 |
| 10 | 1,14 | 1,26 | 1,37 | 1,47 | 1,56 | 1,66 | 1,73 | 1,82 | 1,90 | 1,96 | 2,03 | 2,08 |
| 20 | 1,08 | 1,14 | 1,19 | 1,23 | 1,27 | 1,30 | 1,32 | 1,34 | 1,36 | 1,37 | 1,37 | 1,38 |
| 25 | 1,07 | 1,10 | 1,13 | 1,16 | 1,18 | 1,19 | 1,20 | 1,21 | 1,20 | 1,20 | 1,20 | 1,19 |
| 30 | 1,04 | 1,07 | 1,08 | 1,10 | 1,10 | 1,10 | 1,10 | 1,09 | 1,08 | 1,07 | 1,05 | 1,04 |
| 40 | 1,02 | 1,02 | 1,01 | 0,99 | 0,98 | 0,96 | 0,94 | 0,92 | 0,89 | 0,87 | 0,85 | 0,82 |
| 50 | 0,99 | 0,97 | 0,94 | 0,91 | 0,88 | 0,85 | 0,82 | 0,79 | 0,75 | 0,72 | 0,68 | 0,66 |
| 60 | 0,96 | 0,92 | 0,88 | 0,84 | 0,80 | 0,76 | 0,72 | 0,68 | 0,64 | 0,60 | 0,56 | 0,53 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 70 | 0,94 | 0,88 | 0,83 | 0,77 | 0,72 | 0,67 | 0,63 | 0,58 | 0,54 | 0,49 | 0,45 | 0,42 |
| 75 | 0,93 | 0,86 | 0,80 | 0,74 | 0,68 | 0,63 | 0,58 | 0,53 | 0,49 | 0,44 | 0,40 | 0,37 |
| 80 | 0,91 | 0,84 | 0,77 | 0,70 | 0,64 | 0,58 | 0,53 | 0,48 | 0,44 | 0,39 | 0,35 | 0,32 |
| 90 | 0,88 | 0,78 | 0,70 | 0,62 | 0,55 | 0,49 | 0,43 | 0,38 | 0,33 | 0,29 | 0,26 | 0,22 |
| 95 | 0,85 | 0,74 | 0,65 | 0,56 | 0,49 | 0,43 | 0,37 | 0,32 | 0,27 | 0,23 | 0,20 | 0,17 |
| 97 | 0,83 | 0,72 | 0,62 | 0,53 | 0,46 | 0,39 | 0,33 | 0,28 | 0,24 | 0,20 | 0,17 | 0,14 |
| 99 | 0,80 | 0,67 | 0,57 | 0,48 | 0,40 | 0,33 | 0,28 | 0,23 | 0,19 | 0,15 | 0,12 | 0,10 |
| 99,5 | 0,78 | 0,65 | 0,55 | 0,45 | 0,37 | 0,31 | 0,25 | 0,20 | 0,17 | 0,13 | 0,10 | 0,08 |
| 99,7 | 0,76 | 0,64 | 0,53 | 0,43 | 0,36 | 0,29 | 0,24 | 0,19 | 0,15 | 0,12 | 0,09 | 0,07 |
| 99,9 | 0,75 | 0,61 | 0,50 | 0,40 | 0,33 | 0,26 | 0,21 | 0,16 | 0,12 | 0,09 | 0,08 | 0,06 |

*Приложение 4.* **Нормированные ординаты распределения Пирсона III типа $T_{P\%}=(X_{P\%}-X_{CP})/\sigma$ (биномиальная кривая распределения) [Сикан, 2020]**

| $C_s$ | *P*, % | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **0,01** | **0,1** | **1** | **3** | **5** | **10** | **20** | **25** | **30** | **40** | **50** | **60** | **70** | **75** | **80** | **90** | **95** | **97** | **99** | **99,9** |
| –4 | 0,5 | 0,5 | 0,5 | 0,5 | 0,5 | 0,5 | 0,5 | 0,49 | 0,49 | 0,46 | 0,41 | 0,31 | –0,12 | –0,01 | –0,21 | –0,96 | –1,9 | –2,65 | –4,34 | –8,17 |
| –3,8 | 0,527 | 0,527 | 0,526 | 0,526 | 0,526 | 0,526 | 0,52 | 0,52 | 0,51 | 0,48 | 0,42 | 0,3 | –0,095 | –0,032 | –0,24 | –1 | –1,9 | –2,65 | –4,29 | –7,97 |
| –3,6 | 0,556 | 0,556 | 0,556 | 0,556 | 0,556 | 0,555 | 0,55 | 0,54 | 0,54 | 0,49 | 0,42 | 0,28 | –0,072 | –0,064 | –0,28 | –1,03 | –1,93 | –2,66 | –4,24 | –7,72 |
| –3,4 | 0,588 | 0,588 | 0,588 | 0,588 | 0,587 | 0,586 | 0,58 | 0,57 | 0,55 | 0,5 | 0,41 | 0,27 | –0,036 | –0,11 | –0,31 | –1,06 | –1,94 | –2,66 | –4,18 | –7,54 |
| –3,2 | 0,625 | 0,625 | 0,625 | 0,625 | 0,625 | 0,621 | 0,61 | 0,59 | 0,57 | 0,51 | 0,41 | 0,25 | –0,006 | –0,15 | –0,35 | –1,09 | –1,96 | –2,66 | –4,11 | –7,35 |
| –3 | 0,667 | 0,667 | 0,666 | 0,666 | 0,665 | 0,661 | 0,64 | 0,62 | 0,59 | 0,51 | 0,4 | 0,22 | –0,027 | –0,19 | –0,39 | –1,13 | –1,97 | –2,66 | –4,05 | –7,1 |
| –2,8 | 0,715 | 0,715 | 0,715 | 0,714 | 0,711 | 0,703 | 0,67 | 0,64 | 0,6 | 0,51 | 0,39 | 0,2 | –0,057 | –0,22 | –0,44 | –1,18 | –2 | –2,65 | –3,86 | –6,86 |
| –2,6 | 0,77 | 0,77 | 0,77 | 0,766 | 0,764 | 0,746 | 0,7 | 0,66 | 0,61 | 0,51 | 0,37 | 0,17 | –0,085 | –0,25 | –0,48 | –1,21 | –2 | –2,63 | –3,86 | –6,54 |
| –2,4 | 0,835 | 0,833 | 0,83 | 0,826 | 0,82 | 0,792 | 0,72 | 0,67 | 0,62 | 0,51 | 0,35 | 0,17 | –0,12 | –0,29 | –0,52 | –1,25 | –2 | –2,6 | –3,78 | –6,37 |
| –2,2 | 0,914 | 0,91 | 0,905 | 0,895 | 0,882 | 0,842 | 0,75 | 0,69 | 0,64 | 0,5 | 0,33 | 0,12 | –0,16 | –0,35 | –0,57 | –1,27 | –2,02 | –2,54 | –3,68 | –6,14 |
| –2 | 1,01 | 1 | 0,99 | 0,97 | 0,95 | 0,9 | 0,78 | 0,71 | 0,64 | 0,49 | 0,31 | 0,09 | –0,2 | –0,39 | –0,61 | –1,3 | –2 | –2,51 | –3,6 | –5,91 |
| –1,8 | 1,11 | 1,11 | 1,09 | 1,06 | 1,02 | 0,94 | 0,8 | 0,72 | 0,64 | 0,48 | 0,28 | 0,05 | –0,24 | –0,42 | –0,64 | –1,32 | –1,99 | –2,46 | –3,5 | –5,64 |
| –1,6 | 1,26 | 1,24 | 1,2 | 1,14 | 1,1 | 0,99 | 0,81 | 0,73 | 0,64 | 0,46 | 0,25 | 0,02 | –0,28 | –0,46 | –0,68 | –1,33 | –1,97 | –2,42 | –3,39 | –5,37 |
| –1,4 | 1,41 | 1,39 | 1,32 | 1,23 | 1,17 | 1,04 | 0,83 | 0,73 | 0,64 | 0,44 | 0,22 | 0,2 | –0,31 | –0,49 | –0,71 | –1,34 | –1,95 | –2,37 | –3,27 | –5,09 |
| –1,2 | 1,68 | 1,58 | 1,45 | 1,33 | 1,24 | 1,08 | 0,84 | 0,74 | 0,63 | 0,42 | 0,19 | –0,05 | –0,35 | –0,52 | –0,73 | –1,34 | 1,92 | –2,31 | –3,15 | –4,81 |
| –1 | 1,92 | 1,79 | 1,59 | 1,42 | 1,32 | 1,13 | 0,85 | 0,73 | 0,62 | 0,39 | 0,16 | –0,09 | –0,38 | –0,55 | –0,76 | –1,34 | –1,88 | –2,25 | –3,02 | –4,53 |
| –0,8 | 2,23 | 2,02 | 1,74 | 1,52 | 1,38 | 1,17 | 0,86 | 0,73 | 0,6 | 0,37 | 0,13 | –0,12 | –0,41 | –0,58 | –0,79 | –1,34 | –1,84 | –2,18 | –2,89 | –4,24 |

| | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| –0,6 | 2,57 | 2,27 | 1,88 | 1,61 | 1,45 | 1,2 | 0,85 | 0,72 | 0,59 | 0,34 | 0,1 | –0,16 | –0,44 | –0,61 | –0,8 | –1,33 | –1,8 | –2,12 | –2,75 | –3,96 |
| –0,4 | 2,98 | 2,54 | 2,03 | 1,7 | 1,52 | 1,23 | 0,85 | 0,71 | 0,57 | 0,31 | 0,07 | –0,19 | –0,47 | –0,63 | –0,82 | –1,32 | –1,75 | –2,04 | –2,61 | –0,366 |
| –0,2 | 3,37 | 2,81 | 2,18 | 1,79 | 1,58 | 1,26 | 0,85 | 0,69 | 0,55 | 0,28 | 0,03 | –0,22 | –0,5 | –0,65 | –0,83 | –1,3 | –1,7 | –1,96 | –2,47 | –3,38 |
| 0 | 3,72 | 3,09 | 2,33 | 1,88 | 1,64 | 1,28 | 0,84 | 0,67 | 0,52 | 0,25 | 0 | –0,25 | –0,52 | –0,67 | –0,84 | –1,28 | –1,64 | –1,88 | –2,33 | –3,09 |
| 0,2 | 4,16 | 3,38 | 2,47 | 1,96 | 1,7 | 1,3 | 0,83 | 0,65 | 0,5 | 0,22 | –0,03 | –0,28 | –0,55 | –0,69 | –0,85 | –1,26 | –1,58 | –1,79 | –2,18 | –2,81 |
| 0,4 | 4,61 | 3,66 | 2,61 | 2,04 | 1,75 | 1,32 | 0,82 | 0,63 | 0,47 | 0,19 | –0,07 | –0,31 | –0,57 | –0,71 | –0,85 | –1,23 | –1,52 | –1,7 | –2,03 | –2,54 |
| 0,6 | 5,05 | 3,96 | 2,75 | 2,12 | 1,8 | 1,33 | 0,8 | 0,61 | 0,44 | 0,16 | –0,1 | –0,34 | –0,59 | –0,72 | –0,85 | –1,2 | –1,45 | –1,61 | –1,88 | –2,27 |
| 0,8 | 5,5 | 4,24 | 2,89 | 2,18 | 1,84 | 1,34 | 0,78 | 0,58 | 0,41 | 0,12 | –0,13 | –0,37 | –0,6 | –0,73 | –0,86 | –1,17 | –1,38 | –1,52 | –1,74 | –2,02 |
| 1 | 5,96 | 4,53 | 3,02 | 2,25 | 1,88 | 1,34 | 0,76 | 0,55 | 0,38 | 0,09 | –0,16 | –0,39 | –0,62 | –0,73 | –0,85 | –1,13 | –1,32 | –1,42 | –1,59 | –1,79 |
| 1,2 | 6,41 | 4,81 | 3,15 | 2,31 | 1,92 | 1,34 | 0,73 | 0,52 | 0,35 | 0,05 | –0,19 | –0,42 | –0,63 | –0,74 | –0,84 | –1,08 | –1,24 | –1,33 | –1,45 | –1,58 |
| 1,4 | 6,87 | 5,09 | 3,27 | 2,37 | 1,95 | 1,34 | 0,71 | 0,49 | 0,31 | 0,02 | –0,22 | –0,44 | –0,64 | –0,73 | –0,83 | –1,04 | –1,17 | –1,23 | –1,32 | –1,39 |
| 1,6 | 7,31 | 5,37 | 3,39 | 2,42 | 1,97 | 1,33 | 0,68 | 0,46 | 0,28 | -0,02 | –0,25 | –0,46 | –0,64 | –0,73 | –0,81 | –0,99 | –1,1 | –1,14 | –1,2 | –1,24 |
| 1,8 | 7,76 | 5,64 | 3,5 | 2,46 | 1,99 | 1,32 | 0,64 | 0,42 | 0,24 | -0,05 | –0,28 | –0,48 | –0,64 | –0,72 | –0,8 | –0,94 | –1,02 | –1,06 | –1,09 | –1,11 |
| 2 | 8,21 | 5,91 | 3,6 | 2,51 | 2 | 1,3 | 0,61 | 0,39 | 0,2 | –0,08 | –0,31 | –0,49 | –0,64 | –0,71 | –0,78 | –0,9 | –0,95 | –0,97 | –0,99 | –1 |
| 2,2 | 8,63 | 6,14 | 3,68 | 2,54 | 2,02 | 1,27 | 0,57 | 0,35 | 0,16 | –0,12 | –0,33 | –0,5 | –0,64 | –0,69 | –0,75 | –0,842 | –0,882 | –0,895 | –0,905 | –0,91 |
| 2,4 | 9 | 6,37 | 3,78 | 2,6 | 2 | 1,25 | 0,52 | 0,29 | 0,12 | –0,14 | –0,35 | –0,51 | –0,62 | –0,67 | –0,72 | –0,792 | –0,82 | –0,826 | –0,83 | –0,833 |
| 2,6 | 9,39 | 6,54 | 3,86 | 2,63 | 2 | 1,21 | 0,48 | 0,25 | 0,085 | –0,17 | –0,37 | –0,51 | –0,61 | –0,66 | –0,7 | –0,746 | –0,764 | –0,766 | –0,77 | –0,77 |
| 2,8 | 9,77 | 6,86 | 3,96 | 2,65 | 2 | 1,18 | 0,44 | 0,22 | 0,057 | –0,2 | –0,39 | –0,51 | –0,6 | –0,64 | –0,67 | –0,703 | –0,711 | –0,714 | –0,715 | –0,715 |
| 3 | 10,16 | 7,1 | 4,05 | 2,66 | 1,97 | 1,13 | 0,39 | 0,19 | 0,027 | –0,22 | –0,4 | –0,51 | –0,59 | –0,62 | –0,64 | –0,661 | –0,665 | –0,666 | –0,666 | –0,667 |
| 3,2 | 10,55 | 7,35 | 4,11 | 2,66 | 1,96 | 1,09 | 0,35 | 0,15 | –0,006 | –0,25 | –0,41 | –0,51 | –0,57 | –0,59 | –0,61 | –0,621 | –0,625 | –0,625 | –0,625 | –0,625 |

*Окончание приложения 4*

| $C_s$ | *P*, % | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **0,01** | **0,1** | **1** | **3** | **5** | **10** | **20** | **25** | **30** | **40** | **50** | **60** | **70** | **75** | **80** | **90** | **95** | **97** | **99** | **99,9** |
| 3,4 | 10,9 | 7,54 | 4,18 | 2,66 | 1,94 | 1,06 | 0,31 | 0,11 | –0,036 | –0,27 | –0,41 | –0,5 | –0,55 | –0,57 | –0,58 | –0,586 | –0,587 | –0,588 | –0,588 | –0,588 |
| 3,6 | 11,3 | 7,72 | 4,24 | 2,66 | 1,93 | 1,03 | 0,28 | 0,064 | –0,072 | –0,28 | –0,42 | –0,49 | –0,54 | –0,54 | –0,55 | –0,555 | –0,556 | –0,556 | –0,556 | –0,556 |
| 3,8 | 11,67 | 7,97 | 4,29 | 2,65 | 1,9 | 1 | 0,24 | 0,032 | –0,095 | –0,3 | –0,42 | –0,48 | –0,51 | –0,52 | –0,52 | –0,526 | –0,526 | –0,526 | –0,526 | –0,527 |
| 4 | 12,02 | 8,17 | 4,34 | 2,65 | 1,9 | 0,96 | 0,21 | 0,01 | –0,12 | –0,31 | –0,41 | –0,46 | –0,49 | –0,49 | –0,5 | –0,5 | –0,5 | –0,5 | –0,5 | –0,5 |
| 4,2 | 12,4 | 8,38 | 4,39 | 2,64 | 1,88 | 0,93 | 0,19 | –0,01 | –0,13 | –0,31 | –0,41 | –0,45 | –0,47 | –0,473 | –0,475 | –0,476 | –0,476 | –0,476 | –0,477 | –0,477 |
| 4,4 | 12,76 | 8,6 | 4,42 | 2,63 | 1,86 | 0,91 | 0,15 | –0,032 | –0,15 | –0,32 | –0,4 | –0,44 | –0,451 | –0,454 | –0,455 | –0,455 | –0,455 | –0,455 | –0,455 | –0,455 |
| 4,6 | 13,12 | 8,79 | 4,46 | 2,62 | 1,84 | 0,87 | 0,13 | –0,052 | –0,17 | –0,32 | –0,4 | –0,42 | –0,432 | –0,454 | –0,435 | –0,435 | –0,435 | –0,435 | –0,435 | –0,435 |
| 4,8 | 13,51 | 8,96 | 4,5 | 2,6 | 1,81 | 0,82 | 0,1 | –0,075 | –0,19 | –0,32 | –0,39 | –0,41 | –0,416 | –0,416 | –0,416 | –0,416 | –0,416 | –0,416 | –0,417 | –0,417 |
| 5 | 13,87 | 9,12 | 4,54 | 2,6 | 1,78 | 0,78 | 0,068 | –0,099 | –0,2 | –0,33 | –0,38 | –0,4 | –0,399 | –0,4 | –0,4 | –0,4 | –0,4 | –0,4 | –0,4 | –0,4 |
| 5,2 | 14,25 | 9,27 | 4,59 | 2,6 | 1,74 | 0,73 | 0,035 | –0,12 | –0,21 | –0,3 | –0,37 | –0,38 | –0,384 | –0,385 | –0,385 | –0,385 | –0,385 | –0,385 | –0,385 | –0,385 |
| 5,4 | 14,6 | 9,42 | 4,62 | 2,6 | 1,7 | 0,67 | 0,02 | –0,1 | –0,21 | –0,3 | –0,37 | –0,37 | –0,37 | –0,37 | –0,37 | –0,37 | –0,37 | –0,37 | –0,37 | –0,37 |
| 5,6 | 14,95 | 9,59 | 4,65 | 2,6 | 1,67 | 0,62 | 0 | –0,12 | –0,21 | –0,3 | –0,36 | –0,36 | –0,36 | –0,36 | –0,36 | –0,36 | –0,36 | –0,36 | –0,36 | –0,36 |
| 5,8 | 15,32 | 9,7 | 4,7 | 2,6 | 1,64 | 0,57 | –0,02 | –0,14 | –0,21 | –0,3 | –0,35 | –0,35 | –0,35 | –0,35 | –0,35 | –0,35 | –0,35 | –0,35 | –0,35 | –0,35 |
| 6 | 15,67 | 9,84 | 4,7 | 2,6 | 1,6 | 0,51 | –0,05 | –0,15 | –0,21 | –0,3 | –0,34 | –0,34 | –0,34 | –0,34 | –0,34 | –0,34 | –0,34 | –0,34 | –0,34 | –0,34 |
| 6,2 | 16,04 | 9,95 | 4,71 | 2,6 | 1,56 | 0,47 | –0,05 | –0,15 | –0,21 | –0,3 | –0,34 | –0,34 | –0,34 | –0,34 | –0,34 | –0,34 | –0,34 | –0,34 | –0,34 | –0,34 |
| 6,4 | 16,4 | 10,05 | 4,71 | 2,6 | 1,52 | 0,42 | –0,05 | –0,15 | –0,21 | –0,3 | –0,33 | –0,39 | –0,33 | –0,33 | –0,33 | –0,33 | –0,33 | –0,33 | –0,33 | –0,33 |

*Приложение 5.* **$\chi^2$-распределение (ординаты даны в зависимости от числа степеней свободы и уровня значимости) [Сикан, 2020]**

| *v* | *a* | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | **0,99** | **0,95** | **0,90** | **0,50** | **0,25** | **0,10** | **0,05** | **0,025** | **0,01** | **0,005** | **0,001** |
| 1 | 0,0002 | 0,04 | 0,02 | 0,46 | 1,32 | 2,71 | 3,84 | 5,02 | 6,63 | 7,88 | 10,8 |
| 2 | 0,02 | 0,10 | 0,21 | 1,39 | 2,77 | 4,61 | 5,99 | 7,38 | 9,21 | 10,6 | 13,8 |
| 3 | 0,12 | 0,35 | 0,58 | 2,37 | 4,11 | 6,25 | 7,81 | 9,35 | 11,3 | 12,8 | 16,3 |
| 4 | 0,30 | 0,71 | 1,06 | 3,36 | 5,39 | 7,78 | 9,49 | 11,1 | 13,3 | 14,9 | 18,5 |
| 5 | 0,55 | 1,15 | 1,61 | 4,35 | 6,63 | 9,24 | 11,1 | 12,8 | 15,1 | 16,7 | 20,5 |
| 6 | 0,87 | 1,64 | 2,20 | 5,35 | 7,84 | 10,6 | 12,6 | 14,4 | 16,8 | 18,5 | 22,5 |
| 7 | 1,24 | 2,17 | 2,83 | 6,35 | 9,04 | 12,0 | 14,1 | 16,0 | 18,5 | 20,3 | 24,3 |
| 8 | 1,65 | 2,73 | 3,49 | 7,34 | 10,2 | 13,4 | 15,5 | 17,5 | 20,1 | 22,0 | 26,1 |
| 9 | 2,09 | 3,33 | 4,17 | 8,34 | 11 4 | 14,7 | 16,9 | 19,0 | 21,7 | 23,6 | 27,9 |
| 10 | 2,56 | 3,94 | 4,87 | 9,34 | 12,5 | 16,0 | 18,4 | 20,5 | 23,2 | 25,2 | 29,6 |
| 11 | 3,05 | 4,57 | 5,58 | 10,3 | 13,7 | 17,3 | 19,7 | 21,9 | 24,7 | 26,8 | 31,3 |
| 12 | 3,57 | 5,23 | 6,30 | 11,3 | 14,8 | 18,5 | 21,0 | 23,3 | 26,2 | 28,3 | 32,9 |
| 13 | 4,11 | 5,89 | 7,04 | 12,3 | 16,0 | 19,8 | 22,4 | 24,7 | 27,7 | 29,8 | 34,5 |
| 14 | 4,66 | 6,57 | 7,79 | 13,3 | 17,1 | 21,1 | 23,7 | 26,1 | 29,1 | 31,3 | 36,1 |
| 15 | 5,23 | 7,26 | 8,55 | 14,3 | 18,2 | 22,3 | 25,0 | 27,5 | 30,6 | 32,8 | 37,7 |
| 16 | 5,81 | 7,96 | 9,31 | 15,3 | 19,4 | 23,5 | 26,3 | 28,8 | 32,0 | 34,3 | 39,3 |
| 17 | 6,41 | 8,67 | 10,1 | 16,3 | 20,5 | 24,8 | 27,6 | 30,2 | 33,4 | 35,7 | 40,8 |
| 18 | 7,01 | 9,39 | 10,9 | 17,3 | 21,6 | 26,0 | 28,9 | 31,5 | 34,8 | 37,2 | 42,3 |
| 19 | 7,63 | 10,1 | 11,7 | 18,3 | 22,7 | 27,2 | 30,1 | 32,9 | 36,2 | 38,6 | 43,8 |
| 20 | 8,26 | 10,9 | 12,4 | 19,3 | 23,8 | 28,4 | 31,4 | 34,2 | 37,6 | 40,0 | 45,3 |
| 30 | 15,0 | 18,5 | 20,6 | 29,3 | 34,8 | 40,3 | 43,8 | 47,0 | 50,9 | 53,7 | 59,7 |
| 40 | 22,2 | 26,5 | 29,1 | 39,3, | 45,8 | 51,8 | 55,8 | 59,3 | 63,7 | 66,8 | 73,4 |

## Приложение 6. Номограммы для определения параметров распределения Крицкого — Менкеля методом приближённого наибольшего правдоподобия [Сикан, 2020]

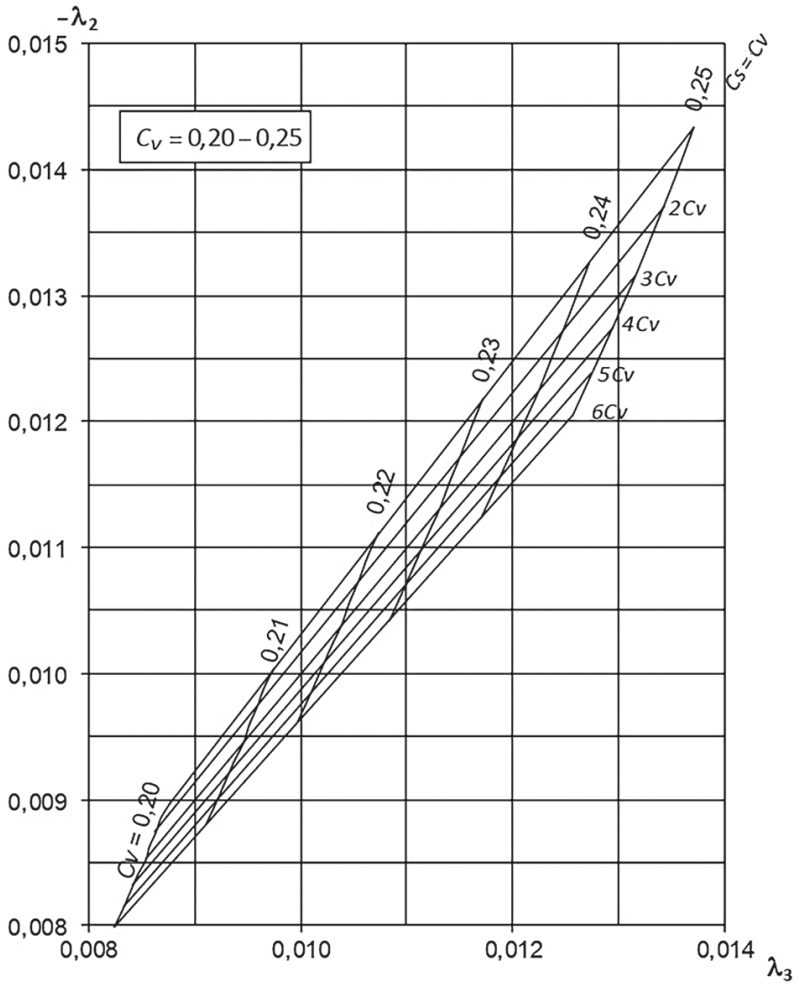



*Рис. 1.* Определение параметров распределения Крицкого — Менкеля при $C_v = 0,20–0,25$

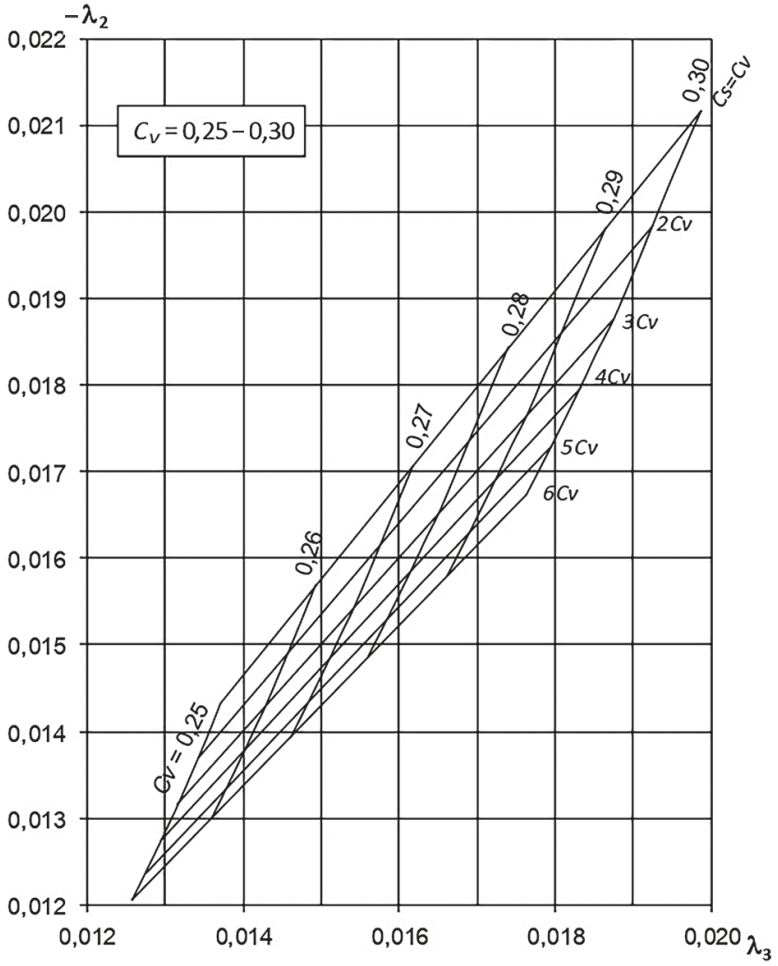



*Рис. 2.* Определение параметров распределения Крицкого — Менкеля при $C_v = 0{,}25–0{,}30$

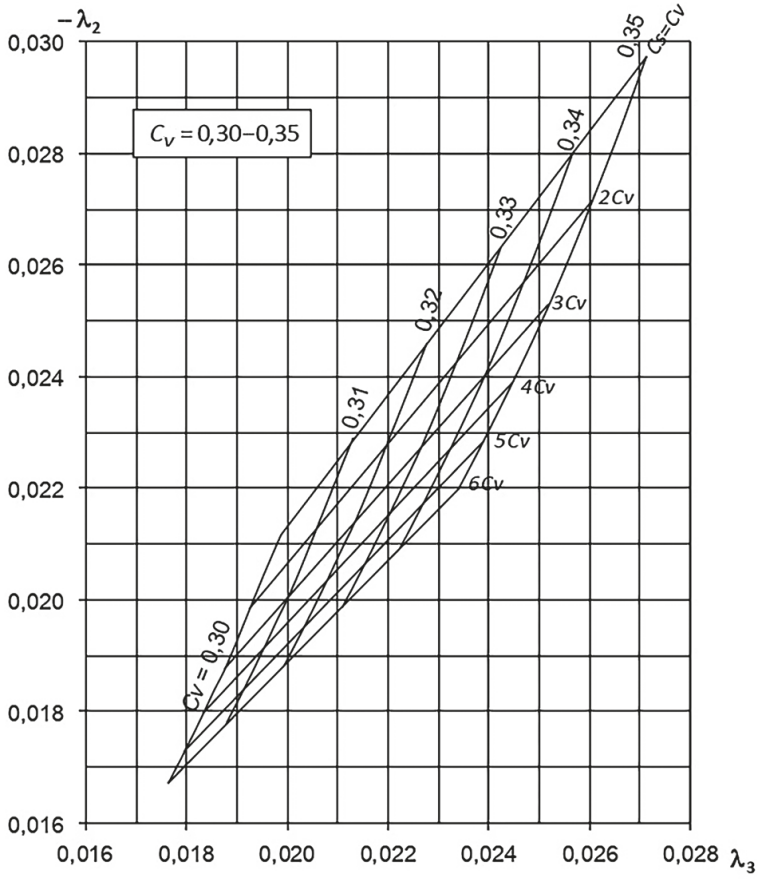



*Рис. 3.* Определение параметров распределения Крицкого — Менкеля при $C_v = 0{,}30–0{,}35$

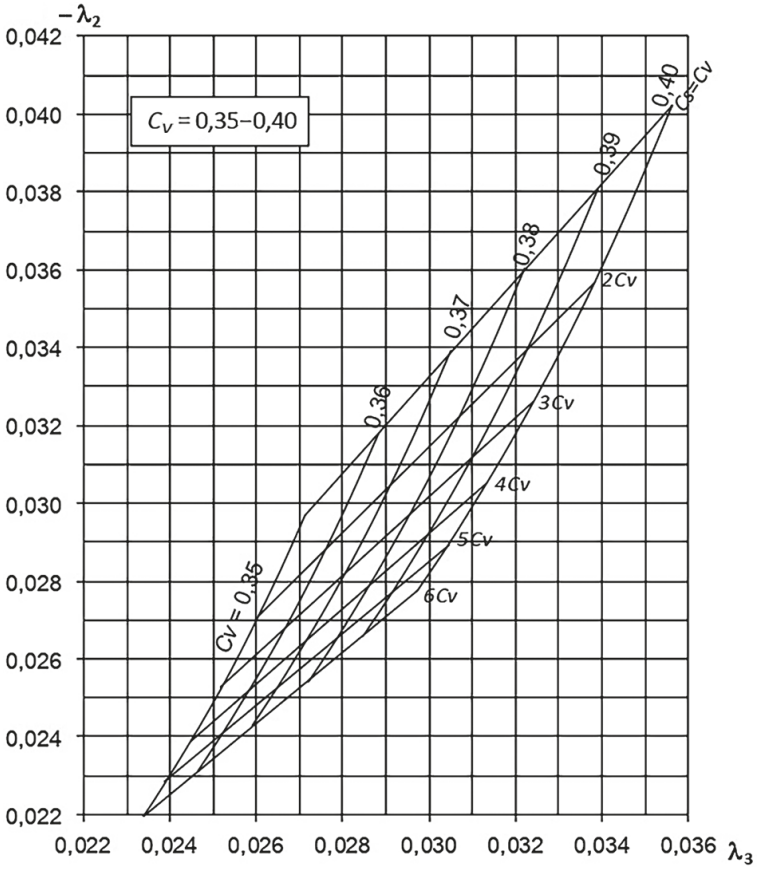



*Рис. 4.* Определение параметров распределения Крицкого — Менкеля при $C_v = 0{,}35–0{,}40$

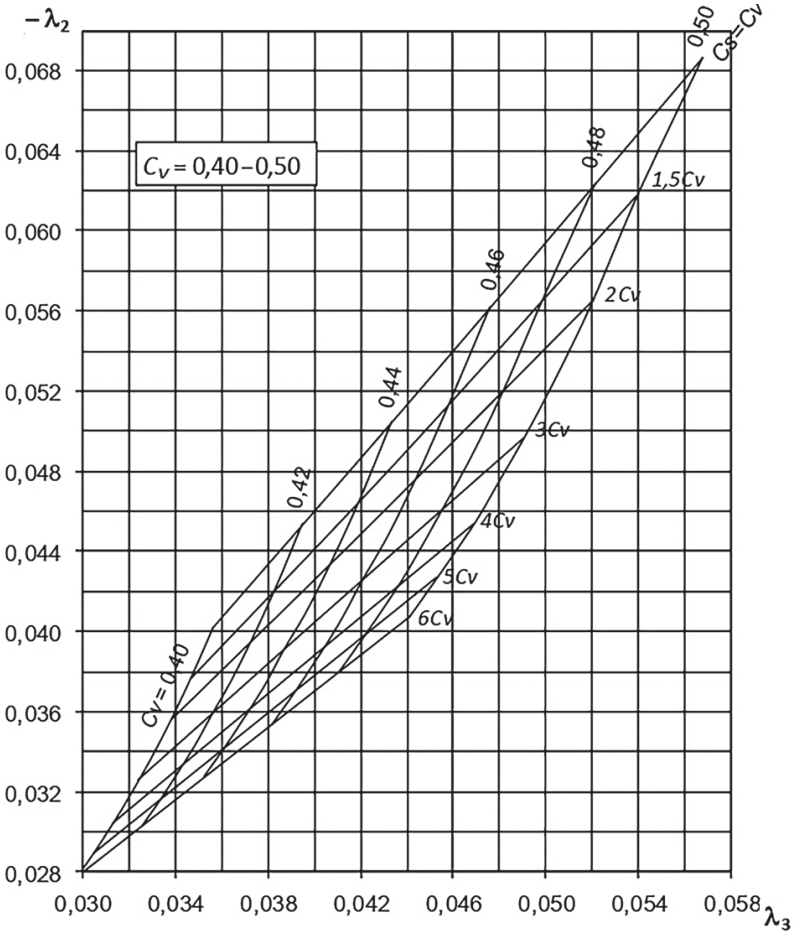



*Рис. 5.* Определение параметров распределения Крицкого — Менкеля при $C_v = 0,40–0,50$

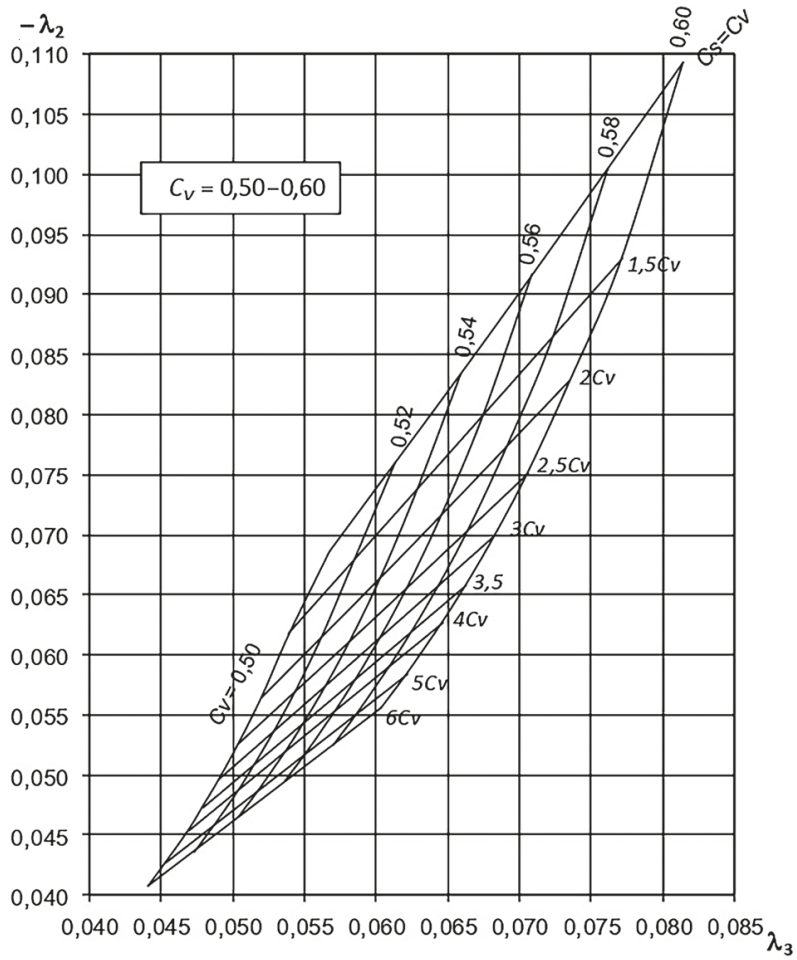



*Рис. 6.* Определение параметров распределения Крицкого — Менкеля при $C_v = 0{,}50–0{,}60$

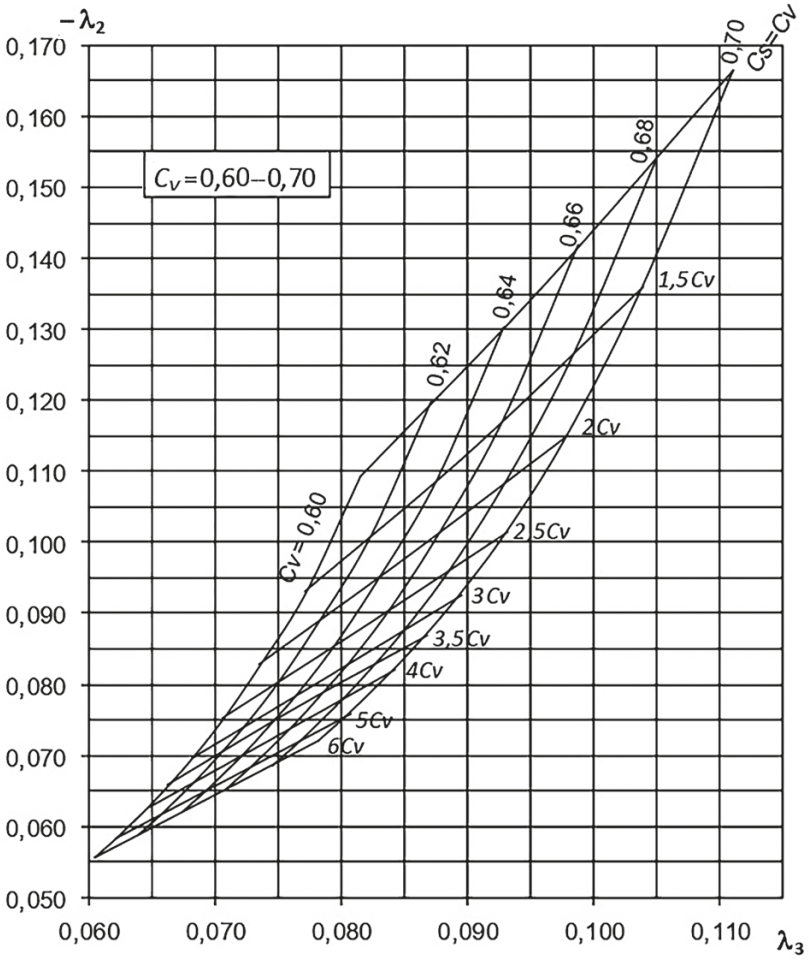



*Рис. 7.* Определение параметров распределения Крицкого — Менкеля при $C_v$ = 0,60–0,70

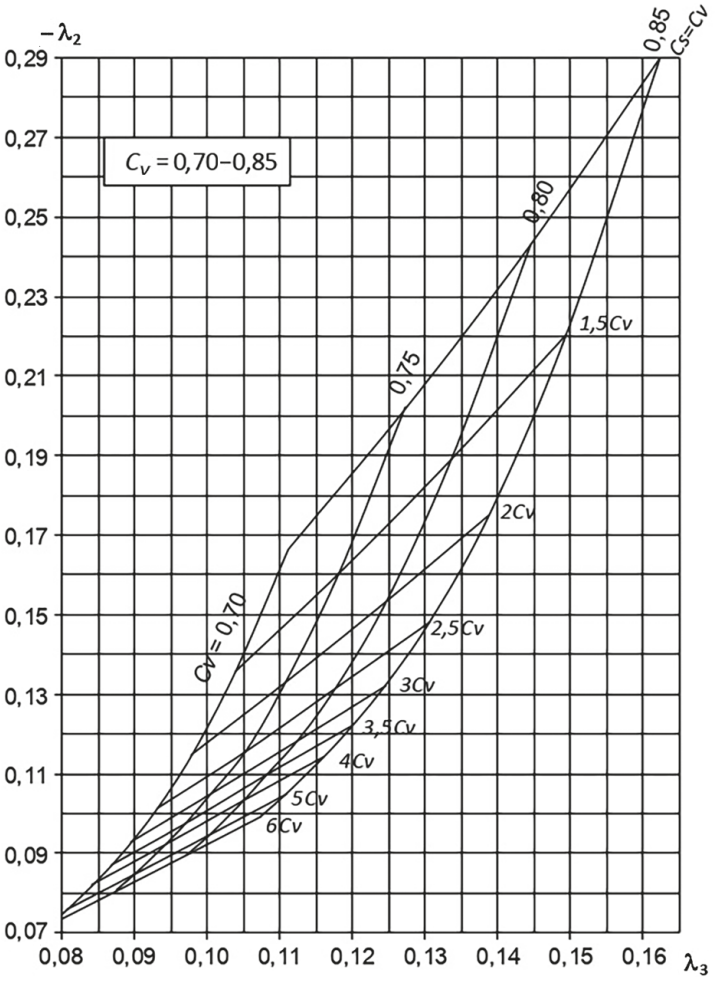



*Рис. 8.* Определение параметров распределения Крицкого — Менкеля при $C_v = 0,70–0,85$

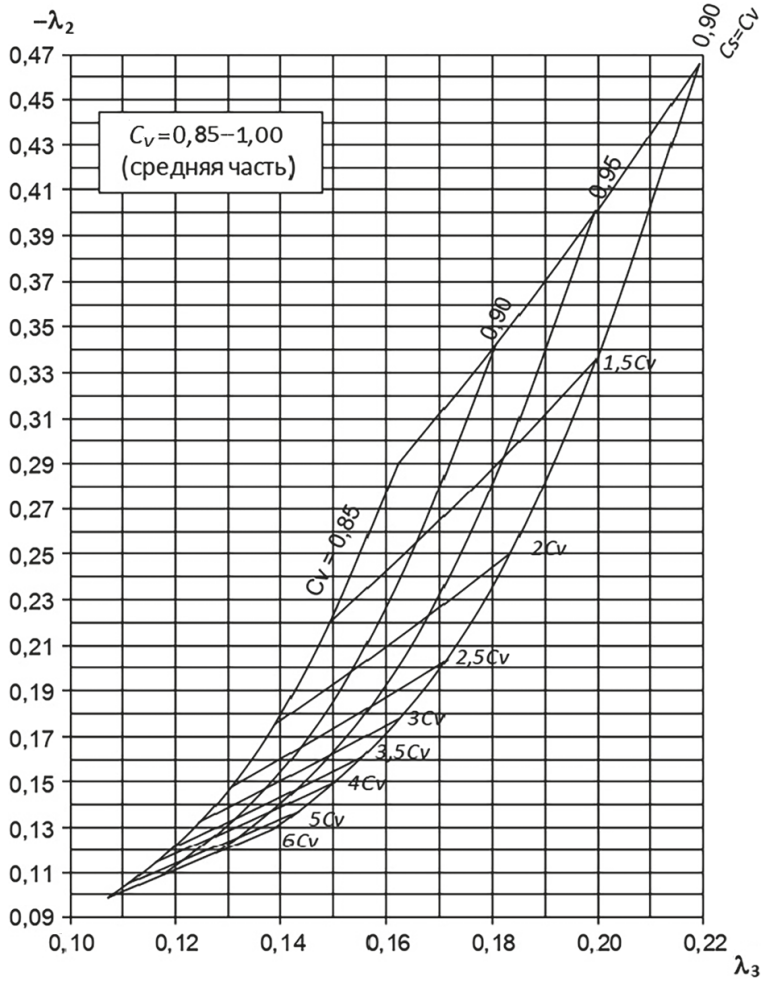



*Рис. 9.* Определение параметров распределения Крицкого — Менкеля при $C_v$ = 0,85–1,00

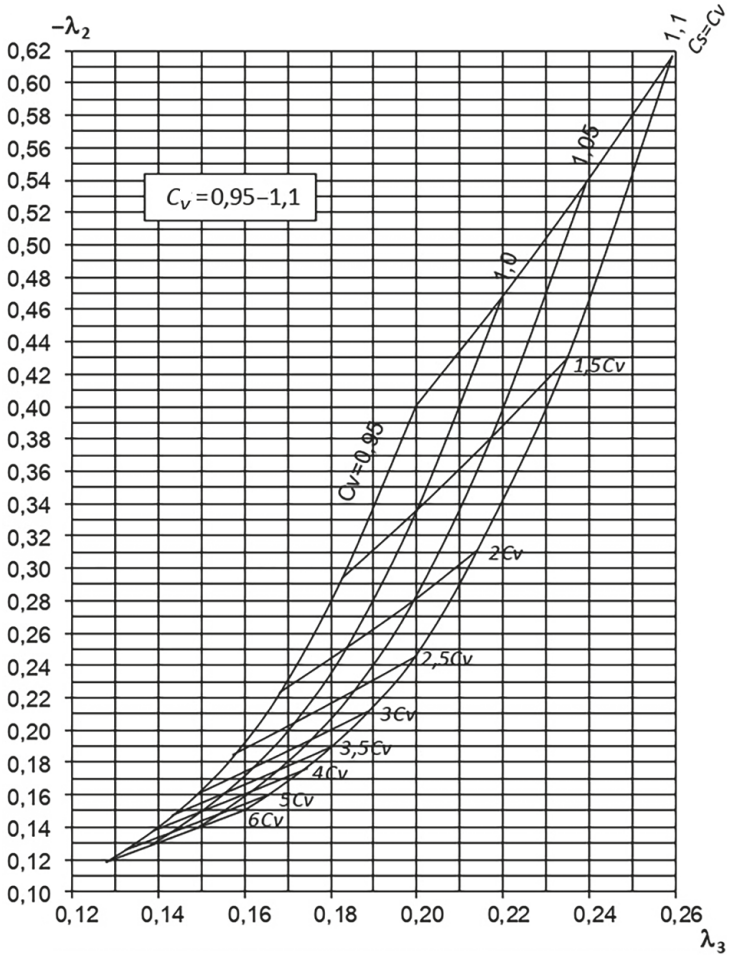



*Рис. 10.* Определение параметров распределения Крицкого — Менкеля при $C_v = 0{,}95–1{,}1$

**Книги и журналы СПбГУ** можно приобрести:

по издательской цене

в интернет-магазине: **publishing.spbu.ru**

и

в сети магазинов «Дом университетской книги», Санкт-Петербург:

Менделеевская линия, д. 5

6-я линия, д. 15

Университетская наб., д. 11

Справки: +7(812)328-44-22, publishing.spbu.ru

Книги СПбГУ продаются в центральных книжных магазинах РФ, интернет-магазинах **amazon.com**, **ozon.ru**, **bookvoed.ru**, **biblio-globus.ru**, **books.ru**, **URSS.ru**

В электронном формате: **litres.ru**

---

Учебное издание

*СУМАЧЁВ Александр Эдуардович, ПОПОВ Сергей Викторович*

СТАТИСТИЧЕСКАЯ ОБРАБОТКА
ГИДРОМЕТЕОРОЛОГИЧЕСКИХ ДАННЫХ

*Учебно-методическое пособие*

Редактор *Н. И. Сочивко*
Корректоры *Т. В. Иванкова*, *И. П. Журова*
Компьютерная верстка *Е. М. Воронковой*
Обложка *И. А. Колтушиной*

Подписано в печать 21.08.2024. Формат 70×100 $^{1}/_{16}$.
Усл. печ. л. 10,7. Тираж 1000 экз. Print-on-Demand. Заказ № .

Издательство Санкт-Петербургского университета.
199004, С.-Петербург, В. О., 6-я линия, 11.
Тел./факс +7(812)328-44-22
publishing@spbu.ru

**publishing.spbu.ru**

Типография Издательства СПбГУ. 199034, С.-Петербург, Менделеевская линия, д. 5.