

Министерство науки и высшего образования Российской Федерации

---

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ  
ВЫСШЕГО ОБРАЗОВАНИЯ  
РОССИЙСКИЙ ГОСУДАРСТВЕННЫЙ  
ГИДРОМЕТЕОРОЛОГИЧЕСКИЙ УНИВЕРСИТЕТ

В.А. Шелутко, Е.С. Урусова

ПРАКТИКУМ ПО ДИСЦИПЛИНЕ  
«АНАЛИЗ И МЕТОДЫ ОБРАБОТКИ  
ГЕОЭКОЛОГИЧЕСКОЙ ИНФОРМАЦИИ»

Санкт-Петербург  
РГГМУ  
2020

**УДК 556.043**  
**ББК 26.22**

*Рецензент:* Шмакова М.В., канд. геогр. наук (Институт озераедения РАН)

**Шелутко В.А., Урсова Е.С.**

Ш44 Практикум по дисциплине «Анализ и методы обработки геоэкологической информации». – СПб.: РГГМУ, 2020. – 120 с.

Приведены содержание, план выполнения и примеры основных методов статистической обработки материалов наблюдений, используемых в научных исследованиях и практических расчетах.

Даются рекомендации по выполнению лабораторных работ для студентов экологического факультета. Направление подготовки 05.03.06 – Экология и природопользование.

УДК 556.043  
ББК 26.22

**ISBN 978-5-86813-518-7**

© В.А. Шелутко, Е.С. Урсова, 2020  
© Российский государственный гидрометеорологический университет (РГГМУ), 2020

# Предисловие

Результаты оценки состояния окружающей природной среды во многом зависят от методов обработки и анализа исходных рядов наблюдений. В результате научной и практической деятельности авторами накоплен богатый опыт по работе с регулярными и экспедиционными данными наблюдений за различными параметрами окружающей природной среды. Это позволило в рамках данного учебного пособия уделить внимание наиболее актуальным и востребованным методам обработки и анализа геоэкологической информации.

Основной задачей данного учебного пособия является закрепление и дополнение на практике теоретического материала, излагаемого в курсе лекций «Геоэкологическая информация: анализ и методы обработки». С этой целью в пособии приводится достаточно широкий круг заданий по применению методов статистического анализа, наиболее часто используемых в геоэкологических исследованиях.

В результате выполнения заданий, приведенных в пособии, студенты должны изучить на конкретных примерах основные методы обработки геоэкологических данных, которые широко применяются в научных исследованиях и на практике, а также лежат в основе многих руководящих документов.

Следует отметить, что часть расчетов, в приведенных заданиях, может быть выполнена с помощью распространенных статистических пакетов, например, «Статистика», «Статграф» и др. Однако, все задания носят учебный характер, а использование этих пакетов приведет к расчетам по схеме «черного ящика», так как алгоритм и физическая сущность используемых методов анализа информации и оценки различных характеристик рассматриваемых процессов останутся неизвестными. Между тем, основной задачей курса является именно освоение этих сведений. Поэтому задания в пособии сделаны таким образом, чтобы они могли выполняться в доступном и широко распространенном табличном процессоре, например Microsoft Excel.

В результате прохождения практического раздела дисциплины «Анализ и методы обработки геоэкологической информации» студенты могут считаться достаточно подготовленными для освоения других курсов учебной программы, широко использующих статистические методы.

Перед выполнением каждой лабораторной работы студенты должны изучить теоретический материал, изложенный в учебниках из списка литературы.

Всего в пособии приводятся 8 заданий, которые имеют самостоятельное назначение. Каждая лабораторная работа состоит из теоретической и расчетной части. В заключении работы следует сделать выводы из проведенного анализа. В качестве исходных данных могут использоваться ряды данных различных геоэкологических характеристик, объемом от 30 до 50 значений.

Выбор исходных данных осуществляется индивидуально для каждого студента преподавателем кафедры прикладной и системной экологии.

В учебном пособии при описании работ в качестве примера использованы ряды данных о концентрациях железа общего в водах р. Славянки в районе города Пушкин (створ 1 и створ 2) за период 2011–2016 гг. объемом в 42 значения.

Учитывая, что все расчеты в пособии выполнены с использованием выборочных рядов наблюдений, все полученные в результате выполнения работ параметры и характеристики носят выборочный характер. В связи с этим дополнительных обозначений для выборочных характеристик в формулах не вводится.

## **Лабораторная работа № 1.**

### **Предварительный анализ исходных данных**

*Цель работы:* изучить методы предварительной оценки исходных рядов наблюдения.

*Основные задачи:*

- 1) провести предварительную оценку законов распределения исследуемых рядов наблюдений;
- 2) провести оценку основных числовых характеристик ряда наблюдения.

*Исходные данные:* два временных ряда данных о состоянии объектов окружающей среды объемом от 30 до 50 значений.

*Содержание*

– Построить и описать графики изменений рассматриваемых процессов во времени.

– Рассчитать и построить гистограммы, эмпирические функции распределения и обеспеченности.

– Рассчитать оценки основных числовых характеристик исходных рядов методом моментов: моды, медианы, математического ожидания, дисперсии, среднего квадратического отклонения, коэффициента вариации, коэффициента асимметрии.

– Оценить погрешности рассчитанных оценок.

– Провести анализ полученных характеристик распределения и оценить особенности каждого из рассматриваемых рядов.

#### *Теоретические положения*

Для выполнения работы студенты должны изучить теоретический материал по учебникам и учебным пособиям из списка литературы. После изучения теоретического материала студенты должны знать следующие *характеристики*:

- 1) случайная величина (определения);
- 2) генеральная совокупность и выборка, требования, выдвигаемые к выборке из генеральной совокупности;
- 3) функция распределения и функция обеспеченности;
- 4) числовые характеристики случайных величин, виды числовых характеристик;
- 5) числовые характеристики положения, мода, медиана, математическое ожидание;
- 6) числовые характеристики рассеивания, дисперсия, среднее квадратическое отклонение, коэффициент вариации и коэффициент асимметрии.

При обработке рядов наблюдений используются различные методы статистического анализа. Выбор этих методов и способов их использования обычно производится на основе предварительного (разведочного) анализа исходных данных, который, в свою очередь, основывается на простейших формах описания и представления ряда значений случайной величины.

В результате этого анализа необходимо ответить на следующие вопросы.

1. Вид распределения – одномодальное или полимодальное, симметричное или асимметричное, положительная или отрицательная асимметрия.

2. Оценка вероятности возможных значений рассматриваемого процесса.

3. Пределы изменения и амплитуда колебаний наблюдаемых значений рассматриваемого процесса.

4. Рассеивание возможных значений ряда относительно центра распределения.

5. Возможная погрешность оценки основных числовых характеристик распределения.

Естественно, что на предварительном этапе применяются простейшие методы статистического анализа, учитывающие в той или иной степени выборочность исходных данных.

В работе представлен пример выполнения задания на основе ряда измеренных концентраций железа общего в р. Славянка в г. Пушкин (створ 2) за период с 2011 по 2016 г.

## План выполнения работы

**1. Построить график изменений рассматриваемого процесса во времени. Дать описание исходных данных.**

Исходные данные представить в таблице 1.1.

**Таблица 1.1. Значения концентраций железа общего, р. Славянка, г. Пушкин, 2011–2016 гг.**

№ п/п	Дата отбора	Значения концентраций $x_i$	№ п/п	Дата отбора	Значения концентраций $x_i$
1	03.05.2011	0,32	22	13.05.2014	0,22
2	23.06.2011	0,18	23	23.06.2014	0,115
3	07.07.2011	0,17	24	09.07.2014	0,13
4	02.08.2011	0,2	25	05.08.2014	0,21
5	22.09.2011	0,29	26	15.09.2014	0,21
6	17.10.2011	0,54	27	16.10.2014	0,11
7	15.11.2011	0,24	28	19.11.2014	0,19
8	15.05.2012	0,24	29	13.05.2015	0,4
9	18.06.2012	0,106	30	01.06.2015	0,11
10	09.07.2012	0,17	31	08.07.2015	0,21
11	06.08.2012	0,16	32	04.08.2015	0,117
12	06.09.2012	0,35	33	15.09.2015	0,16
13	10.10.2012	0,45	34	07.10.2015	0,21
14	11.11.2012	0,51	35	10.11.2015	0,22
15	20.05.2013	0,35	36	10.05.2016	0,24
16	12.06.2013	0,22	37	09.06.2016	0,11
17	02.07.2013	0,24	38	04.07.2016	0,22
18	07.08.2013	0,14	39	01.08.2016	0,02
19	03.09.2013	0,12	40	19.09.2016	0,23
20	04.10.2013	0,076	41	20.10.2016	0,24
21	14.11.2013	0,34	42	07.11.2016	0,30

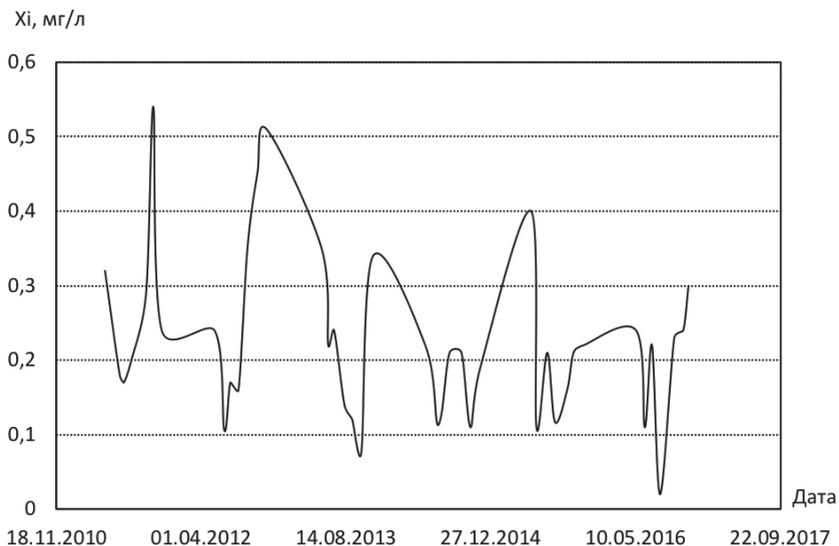


Рис. 1.1. Изменение концентраций железа общего, р. Славянка, г. Пушкин, 2011–2016 гг.

Для оценки изменения рассматриваемого процесса во времени необходимо на основе таблицы 1.1 построить график изменения значений концентраций во времени. На графике по оси абсцисс откладывается дата взятия пробы, а по оси ординат – значения концентраций.

На основе представленного рис. 1.1, можно сделать вывод о четко выраженном внутригодовом изменении значений концентраций железа общего. Просматривается также тенденция снижения значений концентраций во времени. Амплитуда колебаний ряда несколько более 0,5 мг/л. Максимальное значение концентрации составило 0,54 мг/л (17 октября 2011), а минимальное значение – 0,02 мг/л (1 августа 2016 года).

## **2. Расчет и построение гистограммы, эмпирических функций распределения и обеспеченности.**

*2.1. Рассчитать статистическую совокупность, функции распределения и обеспеченности значений концентраций.*

Простейшей табличной формой задания законов распределения случайной величины является ряд распределения. Для расчета значений ряда распределений необходимо разбить амплитуду колебаний ряда значений случайной величины на  $k$  интервалов, подсчитать количество значений  $m_j$  данной случайной величины

попадающих в каждый  $j$ -ый интервал, ( $j = 1, 2, \dots, k$ ) и рассчитать частоту каждого интервала:

$$p_j = \frac{m_j}{n}. \quad (1.1)$$

Следует отметить, что результаты анализа ряда наблюдений по гистограммам в некоторой степени зависят от числа интервалов. Чрезмерное увеличение числа интервалов при ограниченном материале наблюдений приведет к уменьшению материала о каждом интервале в отдельности и, следовательно, к большим погрешностям в определении частоты его значений. Уменьшение числа интервалов приведет к излишней схематизации анализа. Как показывает многолетний опыт работы с короткими рядами, для гидрохимических рядов наблюдения продолжительностью 30–50 значений оптимальным является 6–8 интервалов.

В данном случае для расчета статистической совокупности концентраций железа общего в реке Славянка необходимо определить амплитуду колебания значений ряда  $A$ :

$$A = C_{\max} - C_{\min}, \quad (1.2)$$

где  $C_{\max}$  и  $C_{\min}$  – соответственно максимальное и минимальное значение ряда.

По формуле 1.2 получаем:  $A = 0,54 - 0,02 = 0,52$  [мг/л].

Далее, необходимо рассчитать длину интервала  $L$ . Количество интервалов зависит от объема выборки. В представленном примере объем выборки  $n = 42$ . Таким образом, исходя из рекомендаций выше, определяем, что количество интервалов  $k = 6$ . Следовательно, расчетная длина интервала  $L$  определяется по формуле:

$$L = A/6. \quad (1.3)$$

В данном случае:  $L = 0,52/6 = 0,09$ .

Далее проводится расчет границ интервалов статистической совокупности:

$$l_j = C_{\min} + \sum_1^j L, \quad (1.4)$$

где  $l_j$  – верхняя граница  $j$ -го интервала ( $j = 1, 2, \dots, k$ ).

Следует отметить, что расчет статистической совокупности удобнее проводить по ранжированному в возрастающем порядке ряду.

Затем, по результатам расчётов строится таблица 1.2. Здесь в первой строке указываются границы интервалов, во второй строке – количество значений ряда  $m$ , попадающих в соответствующий

интервал, в третьей сточке приводится частота, рассчитанная по формуле 1.1.

**Таблица 1.2. Статистическая совокупность значений концентраций железа общего, р. Славянка, г. Пушкин, 2011–2016 гг.**

Границы интервала	0,02–0,11	0,11–0,19	0,19–0,28	0,28–0,37	0,37–0,45	0,45–0,54
$m$	6	11	15	6	2	2
$p = m/n$	0,143	0,262	0,357	0,143	0,048	0,048

Как видно из таблицы, максимальное количество значений (15) приходится на 3-й интервал (0,19–0,28), а наименьшее количество (2) – на 5-й и 6-й.

На основе статистических совокупностей могут быть рассчитаны оценки ординат функции обеспеченности и функции распределения.

**Функция распределения** показывает вероятность получить в результате испытания значения  $X$  меньше заданного значения –  $x$ .

**Функция обеспеченности** показывает вероятность получить в результате испытания значение  $X$  большее или равное заданному значению –  $x$ .

Расчет производится по форме таблицы 1.3. В первой строке таблицы указываются нижние границы интервалов статистической совокупности, в последнем столбце указывается верхняя граница последнего интервала. Во второй строке указывается  $k$  – количество значения ряда меньших границы интервала. В третьей строке приводятся оценки значений ординат функции распределения  $F(x)$ , рассчитанные по формуле (1.5):

$$F(x) = k/n. \quad (1.5)$$

В четвертой строке приводится оценка функции обеспеченности  $P(x)$ , расчет выполнен по формуле:

$$P(x) = 1 - F(x). \quad (1.6)$$

Функция обеспеченности, рассчитанная по конкретному ряду наблюдений, называется эмпирической функцией обеспеченности.

**Таблица 1.3. Расчет эмпирических функций распределения и обеспеченности значений концентраций железа общего, р. Славянка, г. Пушкин, 2011–2016 гг.**

Граница интервала	0,02	0,11	0,19	0,28	0,37	0,45	0,54
$k$	0	6	17	32	38	40	42
$F(x)$	0	0,143	0,405	0,762	0,905	0,952	1
$P(x)$	1	0,857	0,595	0,238	0,095	0,048	0

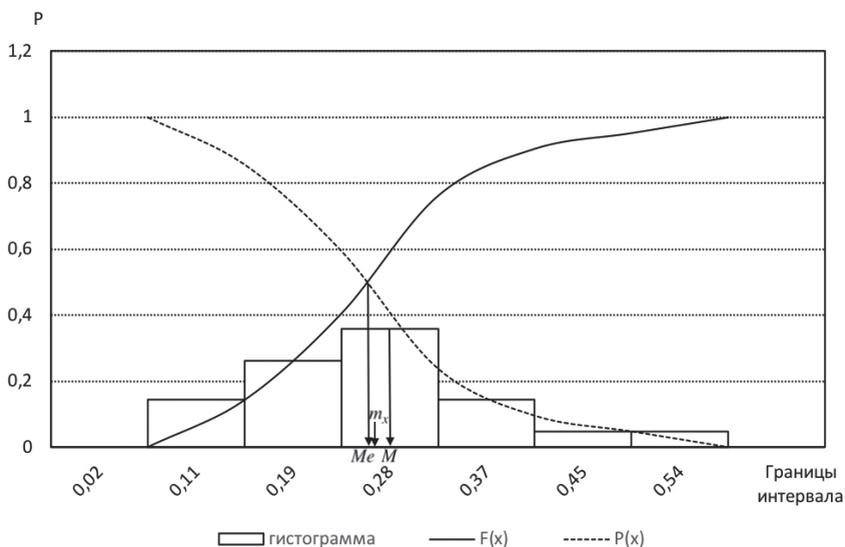


Рис. 1.2. Гистограмма, эмпирические функции распределения и обеспеченности значений концентраций железа общего, р. Славянка, г. Пушкин, 2011–2016 гг.

2.2. Построить в одних координатных осях гистограмму, эмпирические функции распределения и обеспеченности.

Построение рис. 1.2 рекомендуется выполнять с помощью диаграмм в программе Excel. Построения выполняются на основе таблиц 1.2 и 1.3. Для этого по оси абсцисс необходимо отметить границы интервалов значений ряда  $X$ . По оси ординат откладываются значения  $p$  из таблицы 1.2 и значения функций распределения и обеспеченности  $P(x)$  и  $F(x)$  из таблицы 1.3. В результате, для ряда концентраций железа общего на р. Славянка получился график, представленный на рис. 1.2.

Судя по рисунку, распределение ряда одномодальное с выраженной положительной асимметрией.

### 3. Оценка основных числовых характеристик.

Во многих случаях бывает вполне достаточно указать только отдельные параметры, характеризующие какие либо существенные стороны закона распределения исследуемой случайной величины.

*Характеристики, назначение которых – выразить в сжатой форме наиболее существенные особенности распределения, называются **числовыми характеристиками** или параметрами распределения случайной величины.*

Числовые характеристики разделяются на две группы: характеристики положения и рассеивания.

Оценка числовых характеристик по выборочным данным (имеющимся рядам наблюдений) может производиться разными методами (см. работу 3). В данном случае на этапе предварительного анализа эта оценка производится методом моментов.

### *3.1. Рассчитать характеристики положения.*

Числовые характеристики положения показывают расположение определенных характеристик распределения на оси абсцисс. Среди этих числовых характеристик важнейшую роль играют математическое ожидание случайной величины ( $m_x$ ), мода ( $M$ ), медиана ( $Me$ ).

*Для дискретной случайной величины мода – наиболее вероятное значение случайной величины.*

*Для непрерывной случайной величины мода – значение, имеющее наибольшую плотность вероятности.*

Для определения моды необходимо воспользоваться графиком, представленном на рис. 1.2. Значению моды соответствует середина самого высокого столбца гистограммы. В рассматриваемом примере мода равна 0,24 мг/л.

*Медиана непрерывной случайной величины – это такое её значение, для которого вероятность превышения и не превышения одинакова и равна 50 %.*

Медиана определяется по графику функции распределения или обеспеченности (рис. 1.2) как значение  $X$ , соответствующее точке их пересечения. В нашем примере медиана равна 0,21 мг/л.

Оценкой математического ожидания является среднеарифметическое значение:

$$m_x = \frac{\sum_{i=1}^n x_i}{n}. \quad (1.7)$$

Для нашего ряда  $m_x = 0,22$  мг/л.

Полученные характеристики положения необходимо нанести на рис. 1.2.

Взаимное положение характеристик положения на числовой оси может служить для оценки асимметричности распределения. Так, если  $m_x > M$ , то можно говорить о положительной асимметрии. Если  $m_x < M$ , то можно говорить об отрицательной асимметрии.

Однако, стоит отметить, что значение моды во многом зависит от того, какое количество интервалов (6, 7 или 8) будет выбрано для определения статистической совокупности. Следовательно, выводы об асимметрии, полученные по рис. 1.2, носят предварительный и во многом субъективный характер.

### 3.2. Рассчитать числовые характеристики рассеивания.

Числовые характеристика рассеивания характеризуют степень и форму рассеивания возможных значений  $X$  относительно математического ожидания. Среди числовых характеристик рассеивания важнейшую роль играют: дисперсия  $D_{x^2}$ , среднее квадратическое отклонение  $\sigma_{x^2}$ , коэффициент вариации  $Cv$ , коэффициент асимметрии  $Cs$ .

Расчет числовых характеристик рассеивания методом моментов производится в таблице 1.4. Здесь в 1-м столбце представлен порядковый номер значения, во 2-м столбце – значения ряда, в 3-м столбце отклонения значений ряда от среднего значения, в 4-м столбце – квадрат отклонений, в 5-м столбце – куб отклонений. Внизу под 2-м столбцом рассчитано среднее значение ряда, под 3-м, 4-м и 5-м столбцами рассчитана сумма всех значений.

**Таблица 1.4. Вспомогательные расчеты для оценки числовых характеристик ряда концентраций железа общего, р. Славянка, г. Пушкин, 2011–2016 гг.**

№ п/п	$x_i$	$(x_i - x_{cp})$	$(x_i - x_{cp})^2$	$(x_i - x_{cp})^3$
1	0,32	0,10	0,009	0,0009
2	0,18	-0,04	0,002	-0,0001
3	0,17	-0,05	0,003	-0,0002
4	0,2	-0,02	0,001	0,0000
5	0,29	0,07	0,004	0,0003
6	0,54	0,32	0,100	0,0317
7	0,24	0,02	0,000	0,0000
8	0,24	0,02	0,000	0,0000
9	0,106	-0,12	0,014	-0,0016
...	...	...	...	...
31	0,21	-0,01	0,000	0,0000
32	0,117	-0,11	0,011	-0,0012
33	0,16	-0,06	0,004	-0,0003
34	0,21	-0,01	0,000	0,0000
35	0,22	0,00	0,000	0,0000
36	0,24	0,02	0,000	0,0000
37	0,11	-0,11	0,013	-0,0015

№ п/п	$x_i$	$(x_i - x_{cp})$	$(x_i - x_{cp})^2$	$(x_i - x_{cp})^3$
38	0,22	0,00	0,000	0,0000
39	0,02	-0,20	0,041	-0,0084
40	0,23	0,01	0,000	0,0000
41	0,24	0,02	0,000	0,0000
42	0,30	0,08	0,006	0,0004
Среднее	0,22	–	–	–
Сумма	–	0,00	0,51	0,056

Далее приступаем к расчету непосредственно самих характеристик. По результатам, представленным в таблице 1.4, рассчитывается дисперсия.

*Дисперсия служит мерой разбросанности значений случайной величины около её математического ожидания.*

Для выборочных рядов применяется формула (1.8):

$$D_x = \frac{\sum_{i=1}^n (x_i - x_{cp})^2}{(n-1)}. \quad (1.8)$$

Среднее квадратическое отклонение  $\sigma_x$ , как и дисперсия, является мерой разбросанности значений случайной величины около её математического ожидания. Для выборочных рядов применяется формула (1.9):

$$\sigma_x = \sqrt{D_x} = \sqrt{\frac{\sum_{i=1}^n (x_i - x_{cp})^2}{(n-1)}}. \quad (1.9)$$

Для ряда концентраций железа общего в р. Славянка  $\sigma_x = 0,11$  мг/л.

Для заведомо положительных величин в геоэкологии часто применяется безразмерная характеристика рассеивания – коэффициент вариации  $Cv$ . Расчет коэффициента вариации проводится по формуле (1.10):

$$Cv = \frac{\sigma_x}{m_x} = \sqrt{\frac{\sum_{i=1}^n (x_i - x_{cp})^2}{(n-1) \cdot m_x^2}}. \quad (1.10)$$

Коэффициент вариации является безразмерной величиной и не зависит от среднего значения, поэтому может быть использован для сопоставления изменчивости различных рядов наблюдений. В нашем случае  $Cv = 0,50$ .

Коэффициент асимметрии  $Cs$  – это безразмерная величина, которая используется для характеристики асимметричности. Другими словами, коэффициент асимметрии показывает распределение значений ряда относительно математического ожидания. Расчет  $Cs$  для выборочных рядов производится по формуле (1.11):

$$Cs = \frac{\sum_{i=1}^n (x_i - x_{cp})^3}{(n-1) \cdot \sigma_x^3}. \quad (1.11)$$

Асимметрия может быть как положительной, так и отрицательной. Для положительной асимметрии  $Cs > 0$ , а для отрицательной соответственно  $Cs < 0$ . Так как коэффициент асимметрии безразмерная величина и не зависит от среднего значения, то она может быть использована для сравнения изменчивости различных рядов наблюдений. Для ряда концентраций железа общего  $Cs = 0,97$ , следовательно, мы можем сказать, что ряд обладает достаточно выраженной положительной асимметрией.

Результаты расчета основных числовых характеристик представлены в таблице 1.5.

**Таблица 1.5. Оценка основных числовых характеристик ряда концентраций железа общего, р. Славянка, г. Пушкин, 2011–2016 гг.**

$m_x$	$M$	$Me$	$D_x$	$\sigma_x$	$Cv$	$Cs$
0,22	0,24	0,21	0,013	0,11	0,50	0,97

*3.3. Рассчитать абсолютные и относительные погрешности оценок числовых характеристик.*

Для оценки погрешности расчетов числовых характеристик по выборочным данным используются средние квадратические погрешности. Следует отметить, что, вообще говоря, формулы средних квадратических погрешностей оценки большинства числовых характеристик зависят от закона распределения исследуемой величины. При предварительном анализе обычно используются наиболее простые формулы, не учитывающие внутрядные связи.

Абсолютные погрешности  $\sigma$  рассчитываются по формулам (1.12–1.14):

– погрешность оценки математического ожидания:

$$\sigma_{m_x} = \frac{\sigma_x}{\sqrt{n}}; \quad (1.12)$$

– погрешность оценки коэффициента вариации:

$$\sigma_{Cv} = Cv \cdot \sqrt{\frac{(1 + Cv^2)}{2 \cdot n}}; \quad (1.13)$$

– погрешность оценки коэффициента асимметрии:

$$\sigma_{Cs} = Cs \cdot \sqrt{\frac{6}{n} \cdot (1 + 6 \cdot Cv^2 + 5 \cdot Cv^4)}. \quad (1.14)$$

Для расчета относительных погрешностей  $\delta$ , %, можно воспользоваться универсальной формулой:

$$\delta_a = \frac{\sigma_a}{a} \cdot 100, \%, \quad (1.15)$$

где вместо  $a$  подставляете необходимую числовую характеристику.

Результаты расчета погрешностей оценок числовых характеристик представить в виде таблицы 1.6.

**Таблица 1.6. Расчет абсолютных и относительных погрешностей оценок числовых характеристик для ряда концентраций железа общего, р. Славянка, г. Пушкин, 2011–2016 гг.**

Характеристика	$m_x$	$Cv$	$Cs$
Оценка	0,22	0,50	0,97
$\sigma$	0,017	0,06	0,62
$\delta$ , %	7,7	12,21	64

Как видно из полученного результата, наименьшая погрешность получилась для математического ожидания (7,7 %), а наибольшая для коэффициента асимметрии (64 %). Полученный результат вполне типичен для гидрохимических рядов наблюдений.

## Лабораторная работа № 2. Исследование законов распределения и расчет значений ряда заданной обеспеченности

*Цель работы:* изучить основные законы распределения в геоэкологии и подобрать теоретический закон распределения для описания эмпирической кривой обеспеченности.

### *Основные задачи*

1. Рассчитать координаты и построить эмпирическую кривую обеспеченности в поле клетчатки вероятности с умеренной асимметрией.

2. Построить теоретические кривые обеспеченности основных законов распределения и выбрать теоретическую кривую наилучшим образом совпадающую с эмпирической кривой обеспеченности.

3. Определить по исходным рядам концентрации заданной обеспеченности.

4. Сопоставить результаты расчетов по различным кривым обеспеченности.

### *Исходные данные:*

– два временных ряда данных о состоянии объектов окружающей среды (ряды данных те же, что и в работе № 1);

– оценки числовых характеристик исходных рядов, полученные в первой работе.

### *Содержание работы*

– Рассчитать и построить по каждому ряду наблюдений эмпирическую кривую обеспеченности.

– Построить по исходным рядам наблюдений кривые обеспеченности нормального закона распределения, закона распределения Пирсона III типа, трехпараметрического гамма-распределения Крицкого–Менкеля, логарифмически нормального закона распределения, закона распределения Гамбела.

– Для закона распределения Пирсона III типа, опираясь на эмпирические кривые обеспеченности, подобрать оптимальное соотношение  $C_s$  к  $C_v$ .

– Подобрать оптимальный закон распределения, наилучшим образом совпадающий с эмпирическим законом распределения.

– Определить значения концентраций заданной обеспеченности.

### *Теоретические положения*

Для выполнения работы студенты должны изучить теоретический материал по учебникам и учебным пособиям из списка литературы. После изучения теоретического материала студенты должны знать следующие понятия:

- 1) закон распределения случайной величины;
- 2) функции распределения и обеспеченности;
- 3) основное назначение кривых обеспеченности в геоэкологии;
- 4) эмпирическая и теоретическая кривая обеспеченности;
- 5) клетчатка вероятности, её виды;
- 6) нормальный закон распределения, аналитические выражения, основная особенность, схема построения;
- 7) закон распределения Пирсона III типа, свойства, назначение, схема построения, ограничение применения;
- 8) закон распределения Крицкого–Менкеля (трехпараметрическое гамма-распределение), свойства, назначение, схема построения;
- 9) логарифмически-нормальное распределение, свойства, назначение, схема построения кривой обеспеченности;
- 10) закон распределения крайних членов выборки, кривые распределения Гамбела, свойства, назначение, схема построения.

*Эмпирической кривой обеспеченности называется кривая обеспеченности, координаты которой рассчитываются непосредственно по значениям исходного ряда наблюдений, причем каждому члену ряда ставится в соответствие эмпирическая обеспеченность, рассчитанная по одной из формул, учитывающих выборочность исходного ряда.*

Теоретические кривые обеспеченности используются для определения экстремальных расходов воды заданной обеспеченности. Кроме того, кривые обеспеченности используются при моделировании рядов стока по той или иной схеме и др.

В настоящее время наибольшее распространение получили кривые обеспеченности нормального закона распределения, закона распределения Пирсона третьего типа, кривые обеспеченности Крицкого–Менкеля (трехпараметрическое гамма-распределение), логарифмически-нормальные кривые обеспеченности, кривые обеспеченности Гамбела.

Оценка согласованности эмпирических и теоретических кривых обеспеченности может производиться визуально и с помощью различного рода количественных показателей. Визуальная оценка

имеет несколько субъективный характер, но позволяет оценить согласованность кривых отдельно в верхней и нижней части кривых обеспеченности.

В работе представлен пример выполнения задания на основе ряда измеренных концентраций железа общего в р. Славянка в г. Пушкин за период с 2011 по 2016 г.

## План выполнения задания

### 1. Построение эмпирической кривой обеспеченности значений концентраций.

Строить эмпирические и теоретические кривые обеспеченности можно как по самим значениям исходного ряда, так и по модульным коэффициентам. В данной работе построения следует производить по модульным коэффициентам.

Для расчета эмпирической кривой обеспеченности необходимо исходные значения ряда ранжировать в убывающем порядке (столбец 4 таблицы 2.1). Затем для каждого значения ранжированного ряда рассчитывается модульный коэффициент  $k_i$  по формуле (2.1) (столбец 5 таблицы 2.1):

$$k_i = \frac{x_i}{x_{cp}}, \quad (2.1)$$

где  $x_i$  – значение ранжированного ряда,  $x_{cp}$  – среднее значение ряда.

Расчет координат эмпирической кривой обеспеченности в процентах производится по формуле (2.2) (столбец 6 таблицы 2.1):

$$P_i = \frac{i}{(n+1)} \cdot 100, \quad (2.2)$$

где  $i$  – порядковый номер членов ранжированного ряда,  $n$  – число членов ряда.

Модульные коэффициенты (столбец 5) и соответствующие им эмпирические обеспеченности (столбец 6) наносятся в поле клетчатки вероятности (рис. 2.1 и 2.2).

*Клетчаткой вероятности* называется поле графика, в котором по оси ординат откладываются в стандартном масштабе значения исходного ряда, а по оси абсцисс – значения обеспеченности.

В данной клетчатке масштаб по оси абсцисс подобран таким образом, чтобы кривая обеспеченности нормального закона

**Таблица 2.1. Расчет эмпирической кривой обеспеченности ряда концентраций железа общего, р. Славянка, г. Пушкин, 2011–2016 гг.**

№ п/п	$x_i$	Ранжированные значения $x_i$	$k_i$	$P_m$
1	0,32	0,54	2,42	2,3
2	0,18	0,51	2,28	4,7
3	0,17	0,45	2,01	7,0
4	0,2	0,4	1,79	9,3
5	0,29	0,35	1,57	11,6
6	0,54	0,35	1,57	14,0
7	0,24	0,34	1,52	16,3
8	0,24	0,32	1,43	18,6
9	0,106	0,30	1,34	20,9
10	0,17	0,29	1,30	23,3
...	...	...	...	...
30	0,11	0,16	0,72	69,8
31	0,21	0,16	0,72	72,1
32	0,117	0,14	0,63	74,4
33	0,16	0,13	0,58	76,7
34	0,21	0,12	0,54	79,1
35	0,22	0,117	0,52	81,4
36	0,24	0,115	0,51	83,7
37	0,11	0,11	0,49	86,0
38	0,22	0,11	0,49	88,4
39	0,02	0,11	0,49	90,7
40	0,23	0,106	0,47	93,0
41	0,24	0,076	0,34	95,3
42	0,30	0,02	0,09	97,7
Среднее значение	0,22	0,22	1,0	50,0

распределения принимала в поле клетчатки вероятностей вид прямой линии. Эта клетчатка называется *клетчаткой вероятностей с умеренной асимметричностью*.

При подборе масштаба по оси ординат необходимо учитывать, что в большинстве случаев теоретические кривые обеспеченности используются для экстраполяции значений за пределы наблюдаемых значений. В связи с этим необходимо предусмотреть на клетчатке вероятности свободное пространство выше и ниже наблюдаемых значений. Также следует учитывать, что минимум значений по оси ординат не может быть меньше 0 для заведомо положительных

величин, к которым относятся концентрации содержащихся в воде веществ. Практика показывает, что обычно полоса эмпирических точек должна занимать не более двух третей графика по оси ординат.

## 2. Расчет и построение теоретической кривой обеспеченности нормального закона распределения.

Для расчета координат кривой обеспечения нормального закона распределения необходимо воспользоваться приложением 1, в котором представлены нормированные ординаты кривых обеспеченности Пирсона III типа и нормального закона распределения. В первой строке таблицы приложения 1 представлены значения обеспеченности  $P$ , %. Во второй строке представлены соответствующие им значения нормированных ординат  $t_p$  при  $C_s = 0$ .

Результаты расчета координат для кривой обеспеченности нормального закона распределения представлены в таблице 2.2.

В 1-й строке таблицы 2.2 из приложения 1 представлены опорные значения обеспеченности  $P$ , %. Во 2-й строке – значения нормированных ординат  $t_p$ , для указанных обеспеченностей  $P$ , % при  $C_s = 0$ . В 3-й и 4-й строке – соответствующие значения модульных коэффициентов  $k_p$  и значения концентраций  $x_p$ , рассчитанные по формулам (2.3–2.4):

$$k_p = t_p \cdot C_v + 1, \quad (2.3)$$

$$x_p = m_x \cdot k_p, \quad (2.4)$$

где  $C_v$  и  $m_x$  – рассчитанные в первой работе оценки коэффициента вариации и математического ожидания.

**Таблица 2.2. Координаты кривой обеспеченности нормального закона распределения для ряда концентраций железа общего, р. Славянка, г. Пушкин, 2011–2016 гг.**

$P$	0,01	0,1	1	5	10	25	50	75	90	95	99	99,9
$t_p$	3,72	3,09	2,33	1,64	1,28	0,67	0	-0,67	-1,28	-1,64	-2,33	-3,09
$k_p$	2,87	2,55	2,17	1,82	1,64	1,34	1,00	0,66	0,36	0,18	-0,17	-0,55
$x_p$	0,64	0,57	0,48	0,41	0,37	0,30	0,22	0,15	0,08	0,04	-0,04	-0,12

Рассчитанные координаты нормальной кривой обеспеченности  $k_p$  наносятся в поле клетчатки вероятностей (рис. 2.1 и 2.2) и соединяются прямой линией. Линия должна проходить строго по рассчитанным координатам и не должна иметь переломов. В данном случае кривая обеспеченности нормального закона распределения проходит достаточно далеко от точек эмпирической кривой обеспеченности, особенно в верхней части, и уходит в область

отрицательных значений в нижней части, что противоречит физическому смыслу исследуемой величины. Следовательно, данный закон, вероятнее всего, не подходит для описания ряда концентраций железа общего на р. Славянка.

### **3. Расчет и построение кривых обеспеченности Пирсона III типа при различных отношениях $C_s/C_v$ .**

Закон Пирсона III типа предназначен для описания существенно положительных величин и определяется для разных обеспеченностей по величине  $C_s$ . Однако, учитывая большие погрешности расчетов  $C_s$  по имеющимся рядам наблюдений (см. работу 1), на практике определение  $C_s$  производится подбором, исходя из наилучшего совпадения теоретической и эмпирической кривой обеспеченности. В результате выполнения работы необходимо выбрать отношение  $C_s/C_v$ , при котором теоретическая кривая обеспеченности наилучшим образом соответствует значениям эмпирической кривой. Для решения этой задачи, как правило, необходимо построить кривые обеспеченности для 3–4 вариантов соотношений  $C_s/C_v$ . Следует учесть также, что наибольший интерес при исследовании значений концентраций чаще всего имеют их большие значения, следовательно наибольшее внимание при подборе теоретической кривой следует обращать на левую часть клетчатки вероятности. Следует отметить, что при  $C_s = 0$  кривая обеспеченности закона Пирсона совпадает с кривой нормального закона распределения. Как видно из рис. 2.2, кривая распределения Пирсона при  $C_s = 0$  проходит существенно ниже точек в левой части рисунка, значит, в нашем случае целесообразно рассмотреть закон Пирсона при соотношениях  $C_s/C_v = 2$ ,  $C_s/C_v = 3$  и  $C_s/C_v = 4$ .

Непосредственно расчет координат кривой обеспеченности Пирсона также проводится по опорным значениям обеспеченности  $P$  (таблица 2.3). На первом этапе определяются значения нормированных ординат  $t_p$  по приложению 1. При этом, учитывая большие погрешности расчетов  $C_s$ , в данном случае значения  $C_s$  определяются по коэффициентам вариации.

Так, в данном случае для соотношения  $C_s/C_v = 2$  и  $C_v = 0,5$  получаем  $C_s = 0,50 \cdot 2 = 1,0$ . По этому значению необходимо выписать из таблицы приложения 1 значения нормированных ординат  $t_p$ . Эти значения записываются в строку 2 в таблице 2.3.

Для соотношения  $C_s/C_v = 3$  необходимо выписать из таблицы приложения 1 значения нормированных ординат  $t_p$  из строчки  $C_s = 3C_v = 0,50 \cdot 3 = 1,5$ . В результате получим значения строки 5 в таблице 2.3.

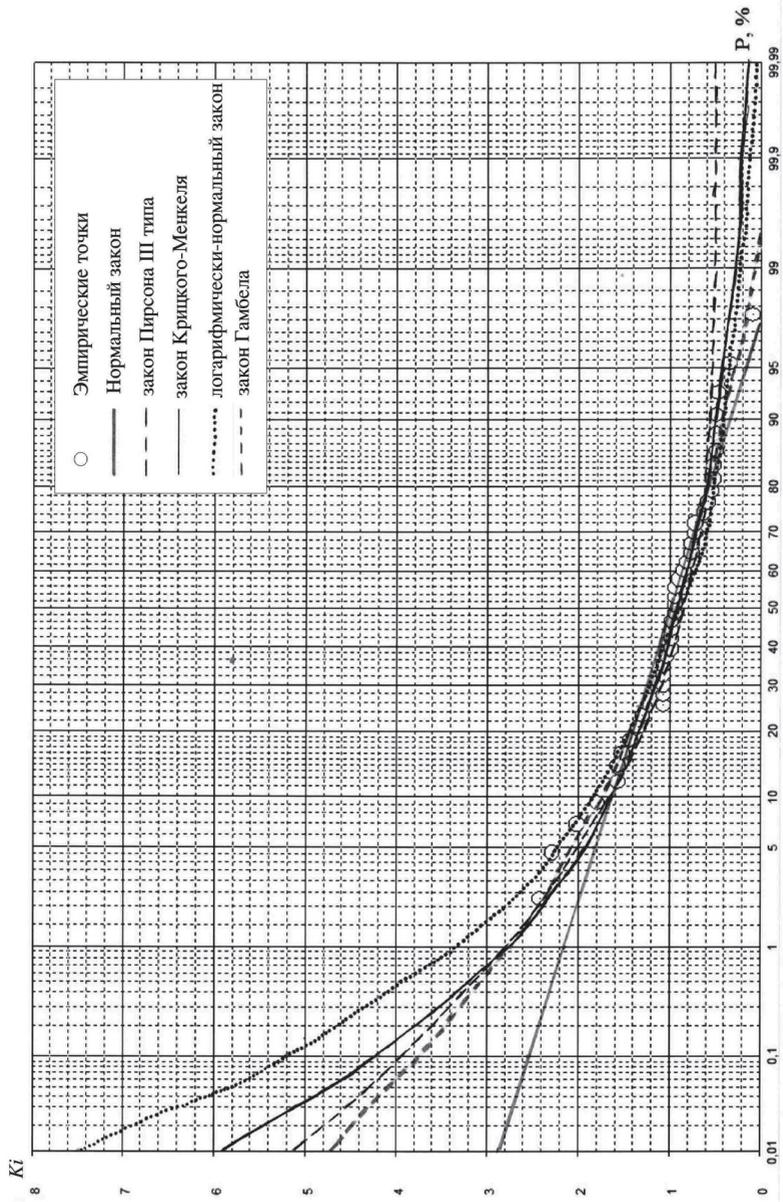


Рис. 2.1. Эмпирические и теоретические кривые обеспеченности для ряда концентраций железа общего, р. Славянка, г. Пушкин, 2011–2016 гг.

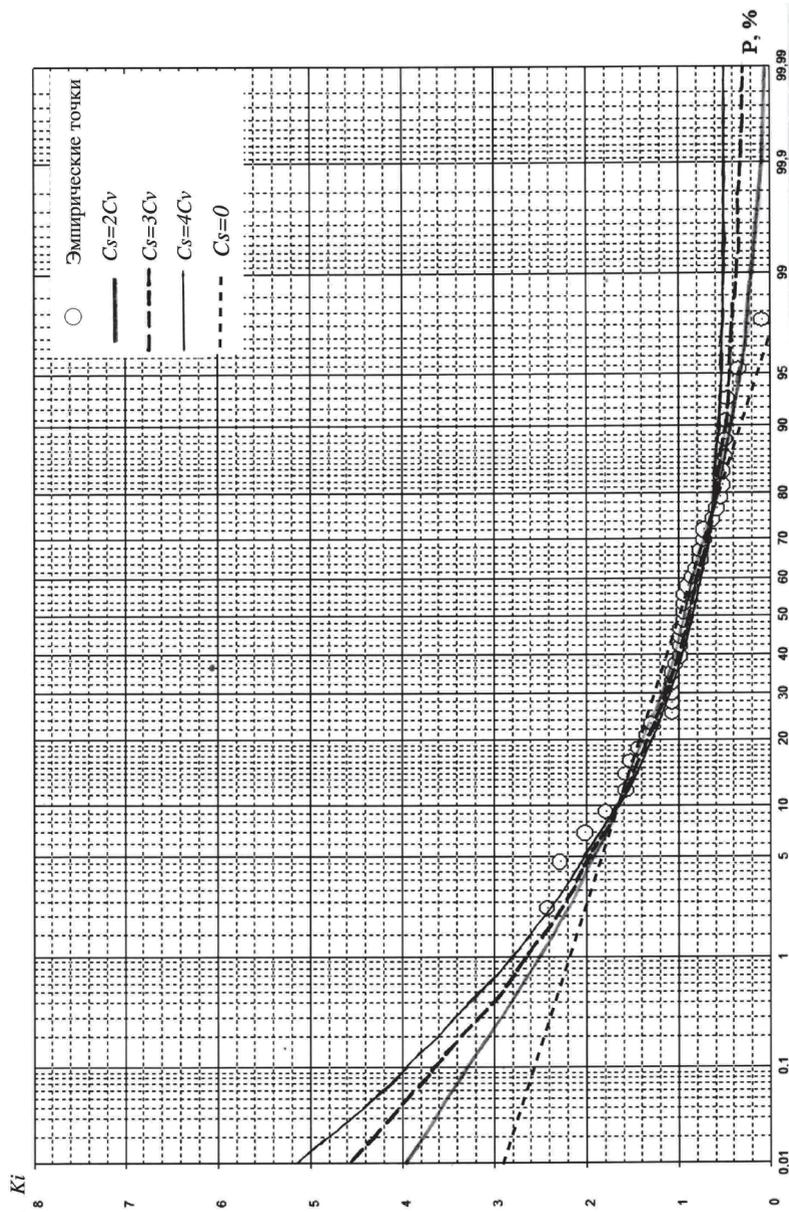


Рис. 2.2. Подбор соотношения  $C_s/C_v$  для теоретического закона Пирсона III типа для ряда концентраций железа общего, р. Славянка, г. Пушкин, 2011–2016 гг.

Для соотношения  $C_s/C_v = 4$  необходимо выписать из таблицы приложения 1 значения нормированных ординат  $t_p$  из строчки  $C_s = 4C_v = 0,50 \cdot 4 = 2,0$ . В результате получим значения строки 7 в таблице 2.3.

Следует отметить, что значение  $C_v$  обычно округляется до двух знаков после запятой, следовательно, если для исследуемого ряда значение попадает между представленными в приложении 1 значениями, то необходимо провести интерполяцию нормированных ординат при больших и меньших значениях  $C_s$ .

Далее по нормированным ординатам производится расчет модульных коэффициентов заданной обеспеченности  $k_p$  по формуле (2.3). Результаты расчета представлены в таблице 2.3.

Аналогично, значения концентраций заданной обеспеченности  $x_p$  рассчитываются по формуле (2.4).

**Таблица 2.3. Расчет координат кривой обеспеченности закона Пирсона III типа для ряда концентраций железа общего, р. Славянка, г. Пушкин, 2011–2016 гг.**

	$P$	0,01	0,1	1	5	10	25	50	75	90	95	99	99,9
$C_s/C_v = 2$	$t_p$	5,96	4,53	3,02	1,88	1,34	0,55	-0,16	-0,73	-1,13	-1,32	-1,59	-1,79
	$k_p$	3,99	3,27	2,51	1,94	1,67	1,28	0,92	0,63	0,43	0,34	0,20	0,10
	$x_p$	0,89	0,73	0,56	0,43	0,37	0,29	0,21	0,14	0,10	0,08	0,05	0,02
$C_s/C_v = 3$	$t_p$	7,09	5,28	3,33	1,96	1,33	0,47	-0,24	-0,73	-1,02	-1,13	-1,26	-1,31
	$k_p$	4,55	3,65	2,67	1,98	1,67	1,24	0,88	0,63	0,49	0,43	0,37	0,34
	$x_p$	1,02	0,81	0,60	0,44	0,37	0,28	0,20	0,14	0,11	0,10	0,08	0,08
$C_s/C_v = 4$	$t_p$	8,21	5,91	3,6	2	1,3	0,39	-0,31	-0,71	-0,9	-0,95	-0,99	-1
	$k_p$	5,12	3,96	2,81	2,00	1,65	1,20	0,84	0,64	0,55	0,52	0,50	0,50
	$x_p$	1,14	0,89	0,63	0,45	0,37	0,27	0,19	0,14	0,12	0,12	0,11	0,11

Найденные координаты кривой обеспеченности в модульных коэффициентах ( $k_p$ ) наносятся в поле клетчатки вероятностей на рис. 2.2 и соединяются плавной линией. Линия должна проходить строго по точкам и не иметь переломов. Таким образом, на рис. 2.2 мы получим четыре линии, соответствующие выбранным значениям  $C_s$ . Нам необходимо на основе визуального анализа определить то соотношение  $C_s$  к  $C_v$ , при котором теоретическая кривая наилучшим образом соответствует эмпирическим точкам. Как уже отмечалось ранее, при выборе необходимо особое внимание уделять левой части клетчатки вероятности.

В данном случае оптимальным для ряда концентраций является отношение  $Cs = 4Cv$ . Действительно, уменьшение или увеличение  $Cs$ , по сравнению с принятым, приводит к ухудшению совпадения эмпирической и теоретической кривой обеспеченности.

Один выбранный вариант кривой обеспеченности Пирсона III типа необходимо перенести на рис. 2.1, на котором будут представлены все исследуемые законы распределения.

#### 4. Расчет и построение теоретической кривой обеспеченности Крицкого–Менкеля.

Закон распределения Крицкого–Менкеля основан на модификации уравнения кривой Пирсона третьего типа при  $Cs = 2Cv$ .

Подбор отношения  $Cs$  к  $Cv$  для кривой обеспеченности Крицкого–Менкеля, в принципе, должен производиться по приведенной выше схеме подбора отношения для кривых Пирсона. В данном случае, для упрощения задачи построение теоретической кривой Крицкого–Менкеля проводится по отношению  $Cs$  к  $Cv$ , выбранному для кривой Пирсона III типа (см. пункт 3). В нашем случае оптимальным для описания ряда концентраций железа общего в реке Славянка было выбрано соотношение  $Cs = 4Cv$ .

Расчет координат кривой обеспеченности Крицкого–Менкеля производится в таблице 2.4. Здесь в первой строке, аналогично другим теоретическим законам распределения, даны опорные обеспеченности  $P$ , %.

Во вторую строку заносятся значения  $k_p$  для заданных обеспеченностей  $P$ , которые берутся из приложения 2 для  $Cs = 4Cv$  или приложения 3 для всех других соответствующую  $Cs/Cv$ . В приложении 3 следует выбирать таблицу, подходящую выбранному соотношению  $Cs$  к  $Cv$  (в нашем случае мы выбрали  $Cs/Cv = 4$ ). Значения выписываются из строки, соответствующей значению коэффициента вариации, полученного в первой работе (в нашем случае  $Cv = 0,50$ ).

По найденным величинам  $k_p$  и среднему значению рассчитываются значения концентраций соответствующей обеспеченности  $x_p$  и заносятся в третью строку таблицы 2.4.

**Таблица 2.4. Расчет координат кривой обеспеченности закона Крицкого–Менкеля для ряда концентраций железа общего, р. Славянка, г. Пушкин, 2011–2016 гг.**

$P$	0,01	0,1	1	5	10	25	50	75	90	95	99	99,9
$k_p$	5,9	4,2	2,77	1,93	1,61	1,21	0,89	0,66	0,51	0,44	0,34	0,25
$x_p$	1,32	0,94	0,62	0,43	0,36	0,27	0,20	0,15	0,11	0,10	0,08	0,06

Координаты кривой обеспеченности в модульных коэффициентах ( $k_p$ ) наносятся в поле клетчатки вероятностей (рис. 2.1). Как видно из рисунка, закон распределения Крицкого–Менкеля подходит достаточно близко к эмпирическим точкам, в том числе и в крайних областях графика.

Следует отметить, что при  $Cs = 2Cv$  кривые обеспеченности Крицкого–Менкеля и Пирсона III типа совпадают между собой.

### 5. Расчет и построение логарифмически-нормальной кривой обеспеченности.

Логарифмически-нормальный закон распределения получил свое распространение для анализа рядов с повышенной асимметрией. Для расчета координат теоретической кривой обеспеченности логарифмически-нормального закона распределения на первом этапе требуется провести логарифмирование исходного ряда наблюдений и рассчитать числовые характеристики ряда логарифмов (таблица 2.5).

Необходимо исходные значения перевести в натуральные логарифмы (столбец 3 таблицы 2.5). Далее необходимо рассчитать среднее значение ( $m_{\ln(x_i)}$ ) и среднее квадратическое отклонение ряда логарифмов.

Для этого в столбце 4 рассчитываются значения  $(\ln x_i - \ln x_{icp})^2$ . Под столбцом необходимо найти сумму.

**Таблица 2.5. Расчет статистических характеристик логарифмов ряда концентраций железа общего, р. Славянка, г. Пушкин, 2011–2016 гг.**

№ п/п	$x_i$	$\ln(x_i)$	$(\ln x_i - \ln x_{icp})^2$
1	0,32	-1,14	0,25
2	0,18	-1,71	0,01
3	0,17	-1,77	0,02
4	0,2	-1,61	0,00
5	0,29	-1,24	0,16
6	0,54	-0,62	1,04
7	0,24	-1,43	0,04
8	0,24	-1,43	0,04
9	0,106	-2,24	0,37
10	0,17	-1,77	0,02
...	...	...	...
31	0,21	-1,56	0,01
32	0,117	-2,15	0,26
33	0,16	-1,83	0,04
34	0,21	-1,56	0,01

№ п/п	$x_i$	$\ln(x_i)$	$(\ln x_i - \ln x_{icp})^2$
35	0,22	-1,51	0,01
36	0,24	-1,43	0,04
37	0,11	-2,21	0,33
38	0,22	-1,51	0,01
39	0,02	-3,91	5,19
40	0,23	-1,47	0,03
41	0,24	-1,43	0,04
42	0,30	-1,20	0,19
Среднее		-1,63	0,33
Сумма		-68,65	13,68

Среднее квадратическое отклонение ряда логарифмов  $\sigma_{\ln x_i}$  рассчитываем по формуле:

$$\sigma_{\ln x_i} = \sqrt{\frac{\sum_{i=1}^n (\ln x_i - \ln x_{icp})^2}{(n-1)}}. \quad (2.5)$$

Получаем для ряда логарифмов концентраций:  $m_{\ln x_i} = -1,63$ ;  $\sigma_{\ln x_i} = 0,58$ .

Далее в таблице 2.6 рассчитываем ординаты логарифмически-нормальной кривой обеспеченности. Здесь в строке 1 даны значения опорных обеспеченностей. В строке 2 представлены нормированные ординаты  $t_p$  аналогичные расчетам для нормального закона распределения (приложение 1 при  $C_s = 0$ ). Значения натуральных логарифмов  $x_p$  (строка 3) рассчитываются по формуле (2.6):

$$\ln(x_p) = t_p \cdot \sigma_{\ln x_i} + m_{\ln x_i}, \quad (2.6)$$

действительные значения  $x_p$  получаем путем потенцирования значений логарифмов:

$$x_p = \exp(\ln(x_p)). \quad (2.7)$$

Значения модульных коэффициентов  $k_p$  рассчитываем по формуле 2.1.

Рассчитанные координаты кривой обеспеченности логарифмически-нормального закона распределения наносятся на клетчатку вероятности (рис. 2.1). Очевидно, что кривая обеспеченности логарифмически-нормального закона распределения достаточно существенно отклоняется от эмпирической кривой обеспеченности.

**Таблица 2.6. Расчет координат кривой обеспеченности логарифмически-нормального закона для ряда концентраций железа общего, р. Славянка, г. Пушкин, 2011–2016 гг.**

$P$	0,01	0,1	1	5	10	25	50	75	90	95	99	99,9
$t_p$	3,72	3,09	2,33	1,64	1,28	0,67	0	-0,67	-1,28	-1,64	-2,33	-3,09
$\ln(x_p)$	0,51	0,15	-0,29	-0,69	-0,90	-1,25	-1,63	-2,02	-2,37	-2,58	-2,98	-3,42
$x_p$	1,67	1,16	0,75	0,50	0,41	0,29	0,20	0,13	0,09	0,08	0,05	0,03
$k_p$	7,49	5,20	3,35	2,25	1,83	1,29	0,87	0,59	0,42	0,34	0,23	0,15

### **6. Расчет и построение кривой обеспеченности Гамбела.**

Закон распределения Гамбела относится к группе законов распределения крайних членов выборки и применяется для расчетов экстремальных значений.

Для расчета координат кривой обеспеченности Гамбела необходимо определить и рассчитать ряд вспомогательных величин.

Во-первых, по таблице  $Y_p = f(p)$  (табл. 1 приложения 4) определить нормированные отклонения от моды  $Y_p$  для значений обеспеченностей  $P$ , представленных в первой строке таблицы 2.7. Полученные значения  $Y_p$  занести во вторую строку таблицы 2.7.

Во-вторых, по таблице 2 приложения 4  $(Y_n, \sigma_{y_n}) = f(n)$  определить вспомогательные величины  $Y_n$  и  $\sigma_{y_n}$  (среднее значение и среднее квадратическое отклонение величины  $Y$ ) в зависимости от объема выборки  $n$ . В данном случае:  $n = 42$ , следовательно  $Y_n = 0,5448$ ,  $\sigma_{y_n} = 1,1458$ .

Рассчитать коэффициенты  $\alpha$  и  $q$  по формулам:

$$\alpha = \frac{\sigma_{y_n}}{\sigma_x}, \quad (2.8)$$

$$q = m_x - \frac{Y_n}{\alpha}, \quad (2.9)$$

где  $\sigma_x$  – среднее квадратическое отклонение исходного ряда и  $m_x$  – среднее значение ряда, полученные в работе 1.

Определить координаты кривой обеспеченности Гамбела  $x_p$  для различных значений нормированных отклонений от моды  $Y_p$  (см. выше) по формуле:

$$x_p = q + \frac{Y_p}{\alpha}. \quad (2.10)$$

Значения  $x_p$  пересчитать в модульные коэффициенты  $k_p$  по формуле (2.1).

**Таблица 2.7. Расчет координат кривой обеспеченности Гамбела для ряда концентраций железа общего, р. Славянка, г. Пушкин, 2011–2016 гг.**

$P$	0,01	1	5	10	25	50	75	90	95	99	99,9
$Y_p$	9,1	4,6	2,97	2,25	1,25	0,37	-0,27	-0,83	-1	-1,53	-1,93
$k_p$	4,74	2,77	2,06	1,75	1,31	0,92	0,64	0,40	0,32	0,09	-0,08
$x_p$	1,06	0,62	0,46	0,39	0,29	0,21	0,14	0,09	0,07	0,02	-0,02

По координатам (строки 1 и 3) построить кривую обеспеченности Гамбела в поле клетчатки вероятности (рис. 2.1).

На рис. 2.1 видно, что закон уходит в отрицательную область значений, что противоречит физическому смыслу исследуемой величины. Следовательно, он не подходит для описания нашего ряда наблюдений.

### **7. Сопоставление значений концентраций заданной обеспеченности и выбор оптимального закона распределения.**

Заключительным этапом выполнения работы 2 является сопоставление значений концентраций заданной обеспеченности, полученных по различным законам распределения. Для этого необходимо в сводную таблицу 2.8 скопировать значения  $x_p$  из таблиц 2.2, 2.3, 2.4, 2.6 и 2.7. Последней строкой необходимо добавить значения эмпирической кривой обеспеченности. Для определения значений эмпирической кривой по опорным обеспеченностям необходимо воспользоваться интерполяцией.

**Таблица 2.8. Значения концентраций железа общего заданной обеспеченности для теоретических законов распределения.**

$P, \%$	0,01	0,1	1	5	10	25	50	75	90	95	99	99,9
Нормальный	0,64	0,57	0,48	0,41	0,37	0,30	0,22	0,15	0,08	0,04	-0,04	-0,12
Пирсона III типа	1,14	0,89	0,63	0,45	0,37	0,27	0,19	0,14	0,12	0,12	0,11	0,11
Крицкого–Менкеля	1,32	0,94	0,62	0,43	0,36	0,27	0,20	0,15	0,11	0,10	0,08	0,06
Логарифически-нормальный	1,67	1,16	0,75	0,50	0,41	0,29	0,20	0,13	0,09	0,08	0,05	0,03
Гамбела	1,06	–	0,62	0,46	0,39	0,29	0,21	0,14	0,09	0,07	0,02	-0,02
Эмпирическая кривая	–	–	–	0,5	0,38	0,25	0,21	0,14	0,11	0,08	–	–

В результате выполнения работы необходимо дать рекомендацию по выбору закона распределения для описания исходного ряда

наблюдения. Выбор делается на основании визуального сопоставления кривых обеспеченности теоретических законов распределения с эмпирической кривой на клетчатке вероятности, представленной на рис. 2.1. При этом наибольшее внимание уделяется левой части графика. Как видно из рис. 2.1, в центральной части графика практически все законы распределения хорошо описывают эмпирические точки. Однако они сильно разнятся в краевых частях графика. Наиболее подходящим законом распределения для описания рядов концентраций железа общего в р. Славянка на наш взгляд является закон Пирсона III типа, так как он более близко подходит к эмпирическим точкам.

Результат выбора оптимального закона распределения особенно сильно сказывается на анализе экстремальных значений концентраций малых или больших обеспеченностей. Так, из таблицы 2.8 видно, что при обеспеченности 0,01 % отличия значений концентраций по закону Пирсона от Нормального закона распределения составляют 44 %. Для логарифмически нормального закона это отличие составляет 46 % относительно закона Пирсона III типа. При обеспеченности 0,1 % различия в значениях концентраций также велики и доходят до 36 % для нормального закона распределения относительно закона Пирсона. Однако, наибольшие расхождения получены для обеспеченности 99,9 %. Здесь разница между концентрациями, полученными по закону распределения Пирсона третьего типа и по логарифмически-нормальному закону доходит до 70 %.

## **Лабораторная работа № 3. Статистические оценки числовых характеристик по имеющимся рядам наблюдений**

*Цель работы:* изучить и сопоставить методы оценки числовых характеристик по имеющимся рядам наблюдений.

*Основные задачи:*

- 1) рассчитать оценки параметров распределения методом моментов, методом наибольшего правдоподобия и методом квантилей;
- 2) рассчитать абсолютные и относительные погрешности оценок параметров распределения;
- 3) проанализировать полученные результаты и дать рекомендации по определению параметров распределения.

*Исходные данные:*

- два временных ряда данных о состоянии объектов окружающей среды (ряды данных те же, что и в работе № 1);
- оценки числовых характеристик методов моментов, полученные в первом задании.

*Содержание работы*

- Дать описание различных методов оценки параметров распределения.
- Рассчитать по исходным рядам наблюдений оценки числовых характеристик методом моментов, методом наибольшего правдоподобия и методом квантилей, их абсолютные и относительные погрешности.
- Сопоставить полученные по разным методам результаты оценок числовых характеристик, дать анализ полученных результатов, указав особенности каждого метода, дать рекомендации по использованию различных методов.

*Теоретические положения*

Для выполнения работы студенты должны изучить теоретический материал по учебникам и учебным пособиям из списка литературы. После изучения теоретического материала студенты должны знать:

- 1) основные требования к методам оценок числовых характеристик;
- 2) теоретические основы метода моментов и его соответствие требованиям, предъявляемым к методам оценки;
- 3) теоретические основы метода наибольшего правдоподобия и его соответствие требованиям, предъявляемым к методам оценки;
- 4) теоретические основы метода квантилей и его соответствие требованиям, предъявляемым к методам оценки.

При анализе различных методов оценки числовых характеристик обычно в качестве критерия используется показатель степени соответствия их следующим требованиям:

- 1) оценка должна быть несмещенной;
- 2) оценка должна быть состоятельной;
- 3) оценка должна быть эффективной.

В настоящее время в геоэкологических исследованиях используются в основном два метода оценки числовых характеристик: метод моментов и метод наибольшего правдоподобия. Кроме того, при анализе в полевых условиях и в научных исследованиях нередко используется графоаналитический метод Алексева (метод квантилей).

## План выполнения задания

### 1. Оценка числовых характеристик методом моментов.

Наибольшее распространение для вычисления оценок параметров распределения получил метод моментов.

Для оценки параметров распределения и погрешностей их определения по выборочным данным используются формулы (1.7–1.15), приведенные в первом задании. Таким образом, расчеты для метода моментов уже выполнены в первом задании, и их необходимо просто скопировать.

### 2. Оценка коэффициента вариации методом наибольшего правдоподобия.

По сравнению с другими методами, метод наибольшего правдоподобия обладает рядом важных достоинств, например, всегда дает эффективные и состоятельные оценки. Вместе с тем, этот метод имеет некоторые недостатки, в значительной степени сдерживающие его применение в повседневной практике и в научных исследованиях. Дело в том, что формулы оценки числовых характеристик этим методом зависят от закона распределения исследуемой величины. Но, так как закон распределения во многих случаях устанавливается по самим оценкам числовых характеристик, то это вносит некоторый элемент неопределенности в расчетную схему метода.

Для расчета значений коэффициента вариации по методу наибольшего правдоподобия необходимо вычислить десятичные логарифмы значений модульных коэффициентов (таблица 3.1).

**Таблица 3.1. Расчет значений коэффициента вариации по методу наибольшего правдоподобия для ряда концентраций железа общего, р. Славянка, г. Пушкин, 2011–2016 гг.**

№ п/п	$k_i$	$\lg(k_i)$	$\ln(k_i)$
1	2,42	0,383	0,882
2	2,28	0,358	0,825
3	2,01	0,304	0,700
4	1,79	0,253	0,582
5	1,57	0,195	0,449
6	1,57	0,195	0,449
7	1,52	0,182	0,420
8	1,43	0,156	0,359
9	1,34	0,128	0,295
...	...	...	...
31	0,72	-0,145	-0,334

№ п/п	$k_i$	$\lg(k_i)$	$\ln(k_i)$
32	0,63	-0,203	-0,467
33	0,58	-0,235	-0,542
34	0,54	-0,270	-0,622
35	0,52	-0,281	-0,647
36	0,51	-0,288	-0,664
37	0,49	-0,308	-0,709
38	0,49	-0,308	-0,709
39	0,49	-0,308	-0,709
40	0,47	-0,324	-0,746
41	0,34	-0,468	-1,078
42	0,09	-1,048	-2,413
Сумма	42	-2,48	-5,71

Модульные коэффициенты  $k_i$  для всех значений ряда концентраций были рассчитаны в работе 2 в таблице 1. Под таблицей необходимо найти сумму логарифмов. Затем по сумме логарифмов рассчитывается параметр  $\lambda$  для ряда десятичных логарифмов по формуле (3.1):

$$\lambda = \frac{\sum_{i=1}^n \lg k_i}{(n-1)}. \quad (3.1)$$

В данном случае  $\lambda = -0,0605$ . Значение  $Cv$  находится затем по таблице приложения 5 по заданному соотношению  $Cs/Cv$  и величине  $\lambda$ . Ранее для закона распределения Пирсона III типа было выбрано соотношение  $Cs = 4Cv$  (см. работу 2). В результате получаем  $Cv = 0,59$ .

В некоторых случаях, в частности при отсутствии специальных таблиц, для расчета  $Cv$  по  $\lambda$  может использоваться формула Семендяева:

$$Cv \approx \sqrt{\frac{-6 \cdot \lambda}{(3 - 2 \cdot \lambda)}}. \quad (3.2)$$

В этой формуле  $\lambda$  рассчитывается по натуральным логарифмам  $\ln(k_i)$ :

$$\lambda = \frac{\sum_{i=1}^n \ln k_i}{(n-1)}. \quad (3.3)$$

Ряд натуральных логарифмов  $\ln(k_i)$  заносится в столбец 4 таблицы 3.1. Получаем  $\lambda = -0,139$ , следовательно  $Cv = 0,50$ .

Оценка абсолютной погрешности расчетов  $Cv$  для обоих способов производится по формуле:

$$\sigma_{Cv} = Cv \cdot \sqrt{\frac{3}{2 \cdot n \cdot (3 + Cv^3)}}, \quad (3.4)$$

а относительной погрешности – по формуле:

$$\delta_{Cv} = \frac{\sigma_{Cv}}{Cv} \cdot 100, \%, \quad (3.5)$$

В данном случае мы получили по ряду натуральных логарифмов  $\delta_{Cv} = 10,68 \%$ , а по ряду десятичных логарифмов –  $\delta_{Cv} = 10,56 \%$ . Следовательно, в дальнейших расчетах используем значение коэффициента вариации с меньшей погрешностью, а именно полученное по ряду десятичных логарифмов  $Cv = 0,59$ .

### **3. Расчет параметров распределения методом квантилей.**

Метод квантилей является наименее трудоемким, что важно в полевых условиях при отсутствии вычислительной техники. Вместе с тем метод имеет ряд серьезных недостатков, в существенной степени сокращающих области его применения. Во-первых, он субъективен, так как результаты расчетов этим методом во многом зависят от проведения сглаживающей линии в поле точек эмпирической кривой обеспеченности. Во-вторых, большое влияние на результаты расчетов оказывают отдельные точки, расположенные в краевых частях графика. В третьих, обычно в гидрологических расчетах для рек однородных районов принимается фиксированное отношение  $Cs/Cv$ . При расчете методом квантилей это отношение индивидуально для каждого водного объекта, что в некоторой степени осложняет расчеты.

На первом этапе расчётов по методу квантилей необходимо на клетчатке вероятностей по эмпирическим точкам обеспеченности провести сглаживающую эмпирическую кривую обеспеченности (рис. 3.1). При этом необходимо использовать исходные значения ряда, а не модульные коэффициенты, как в работе 2.

Далее необходимо со сглаженной кривой обеспеченности снять квантили  $X_5$ ,  $X_{50}$  и  $X_{95}$  при  $P = 5, 50$  и  $95 \%$ .

Затем рассчитать коэффициент скошенности  $S$  по формуле (3.6):

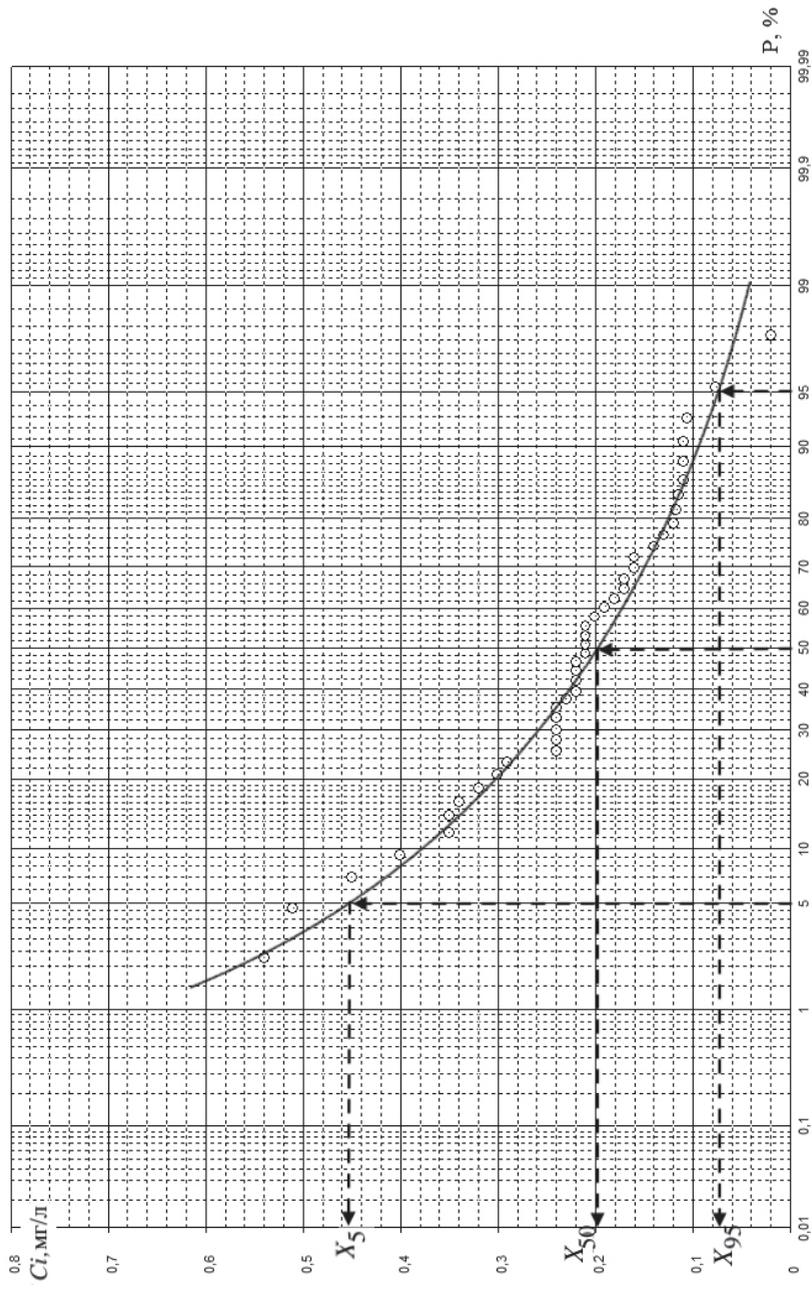


Рис. 3.1. График сглаживающей кривой для ряда концентраций железа общего, р. Славянка, г. Пушкин, 2011–2016 гг.

$$S = \frac{(X_5 + X_{95} - 2 \cdot X_{50})}{(X_5 - X_{95})}. \quad (3.6)$$

Результаты расчета представлены в таблице 3.2. Для ряда концентраций железа общего  $S = 0,32$ .

Пользуясь таблицей  $S = F(C_s)$  (приложение 6), определить оценку коэффициента асимметрии  $C_s$  и затем значения  $t_5$ ,  $t_{50}$ ,  $t_{95}$  и  $t_5 - t_{95}$  и занести их в таблицу. Если значение  $S$  попадает между представленными в таблице значениями, то необходимо выполнить интерполяцию.

Рассчитать оценку среднего квадратического отклонения по формуле (3.7):

$$\sigma_x = \frac{(X_5 - X_{95})}{(t_5 - t_{95})}. \quad (3.7)$$

Рассчитать оценку математического ожидания по формуле:

$$m_x = X_{50} - \sigma_x \cdot t_{50}. \quad (3.8)$$

Рассчитать оценку коэффициента вариации по формуле:

$$Cv = \frac{\sigma_x}{m_x}. \quad (3.9)$$

**Таблица 3.2. Расчет параметров распределения методом квантилей для ряда концентраций железа общего, р. Славянка, г. Пушкин, 2011–2016 гг.**

$X_5$	$X_{50}$	$X_{95}$	$S$	$C_s$	$t_5$	$t_{50}$	$t_{95}$	$t_5 - t_{95}$	$\sigma_x$	$m_x$	$Cv$
0,45	0,2	0,07	0,32	1,1	1,9	-0,18	-1,27	3,17	0,12	0,22	0,54

Расчет абсолютных и относительных погрешностей выполнить аналогично методу моментов, используя формулы 1.12–1.15 из работы 1.

#### **4. Сопоставление результатов расчета параметров распределения.**

Для сопоставления полученных оценок различными методами необходимо составить сводную таблицу результатов оценки параметров распределения (таблица 3.3).

Как видно из полученных результатов, наименьшие погрешности для математического ожидания и коэффициента асимметрии получились при расчетах методом моментов. Для коэффициента вариации наиболее точной оказалась оценка, полученная методом наибольшего правдоподобия.

**Таблица 3.3. Оценки параметров распределения, рассчитанные различными методами для ряда концентраций железа общего, р. Славянка, г. Пушкин, 2011–2016 гг.**

Характеристика	Метод моментов		Метод наибольшего правдоподобия		Метод квантилей	
	Оценка	$\delta$ , %	Оценка	$\delta$ , %	Оценка	$\delta$ , %
$m_x$	0,22	7,7	–	–	0,22	8,3
$C_v$	0,50	12,21	0,59	10,6	0,54	12,4
$C_s$	0,97	64	–	–	1,1	67

## **Лабораторная работа № 4. Применение критериев согласия для выбора закона распределения**

*Цель работы:* изучить методы количественной оценки обоснованности выбора закона распределения с помощью критериев согласия.

*Основные задачи:*

- 1) оценить правильность выбора закона распределения по критерию согласия Колмогорова;
- 2) оценить правильность выбора закона распределения по критерию согласия Пирсона;
- 3) оценить правильность выбора закона распределения по критерию согласия  $n\omega^2$ ;
- 4) сделать общий вывод об обоснованности теоретического закона распределения для описания рядов наблюдений.

*Исходные данные:*

- два временных ряда данных о состоянии объектов окружающей среды (ряды данных те же, что и в работе № 1);
- результаты выбора теоретического закона распределения, полученные в работе 2.

*Содержание работы*

- Рассчитать критерий согласия Колмогорова для выбранного ранее теоретического закона распределения.
- Рассчитать критерий согласия  $n\omega^2$  для выбранного теоретического закона распределения.
- Рассчитать критерий согласия Пирсона для выбранного теоретического закона распределения.

– Сопоставить полученные результаты и сделать вывод об обоснованности выбранного в работе 2 теоретического закона распределения.

#### *Теоретические положения*

Для выполнения работы студенты должны изучить теоретический материал по учебникам и учебным пособиям из списка литературы. После изучения теоретического материала студенты должны знать:

- 1) понятие гипотезы, виды гипотез;
- 2) уровень значимости;
- 3) теоретические основы критерия согласия Колмагорова, его основные достоинства и недостатки, схема расчёта;
- 4) теоретические основы критерия согласия Пирсона, его основные достоинства и недостатки, Схема расчёта;
- 5) теоретические основы критерия согласия  $\chi^2$ , его основные достоинства и недостатки, схема расчёта.

Для объективного суждения о соответствии того или иного закона распределения имеющимся данным наблюдений необходимо иметь какой-то количественный показатель, который измерял бы степень случайного расхождения (или согласия) между наблюдаемым рядом и предполагаемым законом распределения. Подобные показатели, называемые критериями согласия, представляют собой определенную меру расхождения между теоретическим (или гипотетическим) и эмпирическим распределением и основаны на непараметрических гипотезах вида  $H_0: P(x) = P^*(x)$ . Если расхождение в рассматриваемом случае превосходит установленный предел, то гипотеза опровергается. Имеется несколько видов критериев согласия, используемых в различных ситуациях.

## **План выполнения работы**

### **1. Оценка согласия теоретической и эмпирической кривой обеспеченности по критерию согласия Колмогорова.**

Критерий согласия Колмогорова основан на оценке максимального расхождения между теоретической и эмпирической обеспеченностями и предназначен для случая, когда  $n$  достаточно велико.

Для расчета критерия согласия Колмогорова необходимо для всех значений исходного ряда определить обеспеченности по теоретической ( $P_T$ ) и эмпирической ( $P$ ) кривым обеспеченности. Такое определение удобнее производить по клетчатке вероятности, содержащей только эмпирическую и выбранную в работе 2 теоретическую кривые обеспеченности (рис. 4.1). Для определения значений

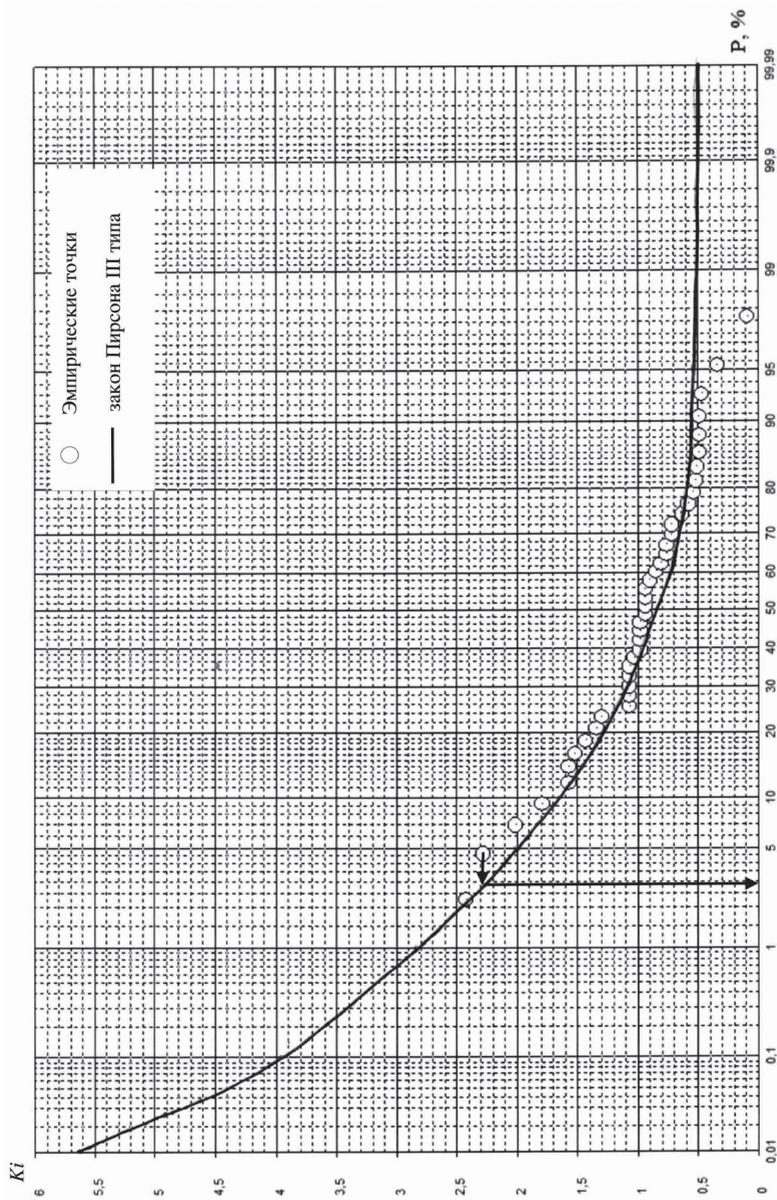


Рис. 4.1. Эмпирическая кривая обогащенности и кривая обогащенности Пирсона III типа по ряду концентраций железа общего, р. Славянка, г. Пушкин, 2011–2016 гг.

теоретических обеспеченностей необходимо каждую точку спроецировать на линию теоретического закона распределения параллельно оси  $OX$ , а затем определить значение полученной обеспеченности по оси  $X$ . Например, для точки  $k_i = 2,28$  значение  $P_T$  будет равно 2,7 % (на рис. 4.1 этот пример отмечен стрелками).

Определение максимального значения разностей обеспеченностей по теоретической и эмпирической кривым обеспеченности проводится в таблице 4.1. Здесь во 2-м столбце приведены значения модульных коэффициентов ряда концентраций ранжированные в убывающем порядке (таблица 2.1, работа 2). В 3-м столбце приведены значения обеспеченностей по эмпирической кривой обеспеченности (таблица 2.1). В 4-м столбце приведены значения обеспеченностей по теоретической кривой обеспеченности  $P_T$ . В 5-м столбце приведены разности этих обеспеченностей, выраженные в долях единицы:

$$D = \frac{|P - P_T|}{100}. \quad (4.1)$$

Далее необходимо определить максимальное значение  $D - D_{\max}$ . В нашем случае  $D_{\max} = 0,158$ .

**Таблица 4.1. Расчет критериев согласия Колмагорова и  $\chi^2$  для ряда концентраций железа общего, р. Славянка, г. Пушкин, 2011–2016 гг.**

№ п/п	$k_i$	$P$	$P_m$	$ P - P_T $	$P_i$	$(P_i - P_T)^2$
1	2,42	2,3	2	0,003	0,01	0,0001
2	2,28	4,7	2,7	0,020	0,04	0,0001
3	2,01	7,0	5	0,020	0,06	0,0001
4	1,79	9,3	7,2	0,021	0,08	0,0001
5	1,57	11,6	11,5	0,001	0,11	0,0001
...	...	...	...	...	...	...
22	0,94	51,2	40	0,112	0,51	0,0125
23	0,94	53,5	40	0,135	0,54	0,0184
24	0,94	55,8	40	0,158	0,56	0,0254
25	0,90	58,1	44	0,141	0,58	0,0205
...	...	...	...	...	...	...
36	0,51	83,7	93	0,093	0,85	0,0072
37	0,49	86,0	99,99	0,139	0,87	0,0171
38	0,49	88,4	99,99	0,116	0,89	0,0115
39	0,49	90,7	99,99	0,093	0,92	0,0069
40	0,47	93,0	99,99	0,070	0,94	0,0035

№ п/п	$k_i$	$P$	$P_m$	$ P - P_i $	$P_i$	$(P_i - P)^2$
41	0,34	95,3	99,99	0,046	0,96	0,0013
42	0,09	97,7	99,99	0,023	0,99	0,0001
$D_{\max}$				0,158	Сумма	0,2426

В качестве показателя степени согласия эмпирической и теоретической кривых обеспеченностей принята величина  $\lambda$ , определяемая по формуле:

$$\lambda = D_{\max} \cdot \sqrt{n}. \quad (4.2)$$

Для ряда концентраций железа общего получаем  $\lambda = 1,02$ .

Далее по таблице в приложении 7 определяем вероятность превышения  $\lambda - k(\lambda)$ . В нашем случае  $k(\lambda) = 0,7484$ .

Рассчитывается предельная вероятность превышения в процентах:

$$P_\lambda = (1 - k(\lambda)) \cdot 100. \quad (4.3)$$

Если полученное значение  $P_\lambda$  меньше или равно 10 %, то гипотеза о тождественности теоретического и эмпирического закона распределения опровергается.

В нашем случае  $1 - k(\lambda) = 25$  %, что больше 10 %, следовательно гипотеза о равенстве эмпирического и теоретического распределения не опровергается и выбранный нами закон распределения Пирсона III типа подходит для описания ряда концентраций железа общего.

## 2. Оценка согласия теоретической и эмпирической кривой обеспеченности по критерию согласия $n\omega^2$ .

Критерий согласия  $n\omega^2$  основывается на оценке расхождения обеспеченностей непосредственно наблюдаемых значений рассматриваемой величины  $X$ . Критерий  $n\omega^2$  обладает рядом преимуществ по сравнению с другими критериями. Во-первых, при его применении полнее используется информация, заключающаяся в данных выборки, так как его расчет основывается на анализе непосредственно наблюдаемых величин. Во-вторых, его распределение значительно быстрее сходится к предельному закону, особенно при больших значениях  $n\omega^2$ .

Для оценки гипотезы о законе распределения при использовании критерия  $n\omega^2$  необходимо для ранжированных в убывающем порядке значений модульных коэффициентов исходного ряда рассчитать эмпирические обеспеченности по формуле (4.4):

$$P_i = \frac{(i-0,5)}{n}, \quad (4.4)$$

где  $i$  – порядковый номер значения.

Затем определить обеспеченности значений  $k_i$  по предполагаемому закону распределения  $P_T$ . Так как значения теоретических обеспеченностей уже были определены ранее при выполнении критерия согласия Колмогорова, то расчеты по критерию согласия  $n\omega^2$  были добавлены в таблицу 4.1, где в столбце 6 рассчитывались эмпирические обеспеченности по формуле (4.4).

Расчет критерия  $n\omega^2$  производится по формуле:

$$n\omega^2 = \frac{1}{12 \cdot n} + \sum_{i=1}^n (P_i - P_T)^2, \quad (4.5)$$

где  $P_i$  и  $P_T$  – соответственно значения обеспеченности по эмпирической и предполагаемой теоретической кривым обеспеченностей. Значения  $(P_i - P_m)^2$  рассчитаны в столбце 7 таблицы 4.1. Внизу под таблицей найдена сумма этих значений. Таким образом, для ряда концентраций железа общего получаем:

$$n\omega^2 = 0,24.$$

При  $n \geq 40$  распределение произведения  $n\omega^2$  близко к некоторому предельному распределению, для которого определены верхние доверительные границы критерия  $n\omega^2$  (таблица 4.2).

**Таблица 4.2. Значения верхних доверительных границ критерия  $n\omega^2$  в зависимости от уровня значимости  $\alpha$ , %**

$\alpha$	20	10	5	3	2	1	0,1
$n\omega_\alpha^2$	0,2412	0,3473	0,4614	0,5489	0,6198	0,7435	1,168

Если  $n\omega^2 > n\omega_\alpha^2$ , то гипотеза о согласии кривых обеспеченности опровергается.

При уровне значимости  $\alpha = 10\%$  получаем, что значение  $n\omega^2$  меньше критического ( $0,24 < 0,3473$ ), следовательно, гипотеза о согласии теоретической и эмпирической кривых обеспеченности не опровергается и закон распределения Пирсона III типа подходит для описания ряда концентраций железа общего.

### **3. Оценка согласия теоретической и эмпирической кривой обеспеченности по критерию согласия Пирсона.**

Критерий согласия Пирсона основан на оценке согласия эмпирических данных и предполагаемого закона распределения с использованием расхождение между эмпирической и теоретической вероятностью интервалов значений случайной величины.

Для расчета критерия согласия Пирсона необходимо ряд разбить на интервалы. При этом необходимо соблюсти ряд требований:

- шаг разбиения должен быть одинаковым,
- количество значений в одном интервале не менее 3. Для соблюдения этого требования допускается объединять крайние интервалы с малым количеством значений в один.
- при объеме выборки 35–50 значений необходимо использовать 8–9 интервалов (до объединения крайних малочисленных интервалов).

Так как клетчатка вероятности и теоретический закон распределения построены в модульных коэффициентах, то разбивать на интервалы удобнее именно ряд  $k_i$ . Схема разбиения на интервалы подробно описана в работе 1. Для этого необходимо ранжировать ряд, определить амплитуду значений и определить шаг.

Ряд модульных коэффициентов концентраций железа общего был разбит на 8 интервалов: 0,09–0,38 (2 значения); 0,38–0,67 (9 значений); 0,67–0,96 (11 значений); 0,96–1,26 (10 значений); 1,26–1,55 (4 значения); 1,55–1,84 (3 значения); 1,84–2,13 (1 значение) и 2,13–2,42 (2 значения).

Так как в первом, седьмом и восьмом интервалах количество значений менее трех, то мы принимаем решение об объединении 1 и 2 интервалов и 7 и 8 интервалов. В итоге получаем в первом интервале 11 значений, а в последнем – 3 значения. Также общее количество интервалов сократится и составит 6 (см. таблицу 4.3).

Для расчета критерия согласия Пирсона полученные интервалы записываются в таблицу 4.3.

**Таблица 4.3. Расчет критерия согласия Пирсона для ряда концентраций железа общего, р. Славянка, г. Пушкин, 2011–2016 гг.**

Интервал	$m$	$P$	$P_i$	$P_{\tau}$	$\Delta P$	$\Delta P^2/P_m$
0,09–0,67	11	0,262	0,999	0,279	–0,017	0,001
0,67–0,96	11	0,262	0,72	0,340	–0,078	0,018
0,96–1,26	10	0,238	0,38	0,170	0,068	0,027
1,26–1,55	4	0,095	0,21	0,095	0,000	0,000
1,55–1,84	3	0,071	0,115	0,050	0,021	0,009
1,84–2,42	3	0,071	0,065	0,045	0,026	0,016
–	–	–	0,02	–	Сумма	0,071

Здесь в первом столбце указаны границы интервалов в модульных коэффициентах  $k_i$ . Во втором столбце записано количество значений эмпирической кривой в  $j$ -м интервале –  $m$ . В третьем столбце

эмпирическая вероятность или частота значений эмпирической кривой в  $j$ -том интервале  $P$ :

$$P = m_j / n. \quad (4.6)$$

В 4-м столбце во всех строках, за исключением последней, приводятся значения обеспеченности по теоретической кривой обеспеченности Пирсона III типа, соответствующие верхней границе каждого интервала. В последней строке 4-го столбца приводится значение обеспеченности, соответствующее нижней границе последнего интервала. В данном случае в последней строке 4-го столбца представлено значение теоретической обеспеченности для значения 2,42. Значения обеспеченностей по теоретической кривой обеспеченности определяются по рис. 4.1. В 5-м столбце приводятся теоретические вероятности интервалов интервала:

$$P_{\tau} = P_i - P_{i+1}. \quad (4.7)$$

В 6-м столбце приводится разность вероятностей интервалов по эмпирической и теоретической кривой обеспеченности:

$$\Delta P = P - P_{\tau}. \quad (4.8)$$

Наконец, в 7-м столбце приводится отношение  $\Delta P^2 / P_{\tau}$ . По значениям 7-го столбца рассчитывается сумма.

Критерий согласия Пирсона  $\chi^2$  рассчитывается по формуле:

$$\chi^2 = n \cdot \sum \frac{\Delta P^2}{P_{\tau}}. \quad (4.9)$$

Полученное значение  $\chi^2$  сравнивается с табличным значением (приложение 8) при заданном уровне значимости  $\alpha$  и числу степеней свободы  $\nu = k - 2$ , где  $k$  – количество интервалов. Если  $\chi^2$  окажется больше чем  $\chi^2_{\tau}$ , то гипотеза о тождественности кривых обеспеченности опровергается.

Для ряда концентраций железа общего получаем:

$$\chi^2 = 42 \cdot 0,071 = 2,99.$$

При уровне значимости  $\alpha = 10\%$  и  $\nu = 6 - 2 = 4$  значение  $\chi^2_{\tau} = 7,78$ .

Так как  $\chi^2$  меньше критического значения ( $2,99 < 7,78$ ), то гипотеза о тождественности теоретической и эмпирической кривых обеспеченности не опровергается и выбранный нами закон распределения Пирсона III типа подходит для описания ряда концентраций железа общего.

Считается, что использование распределения  $\chi^2$  в качестве критерия согласия возможно только в тех случаях, когда длина выборки

$n \geq 50$ . Помимо этого, недостатком метода является то, что группировка данных по интервалам в случае непрерывной случайной величины или большого числа возможных ее значений приводит к некоторой потере информации. Кроме того, элементы неопределенности и возможной неоднозначности решений вносятся при назначении числа интервалов и длины самих интервалов.

Для ряда концентраций железа общего получаем, что выбранный в работе 2 закон распределения Пирсона III типа может использоваться для его описания.

## **Лабораторная работа № 5. Оценка однородности исходных рядов наблюдений**

*Цель работы:* изучить методы оценки однородности рядов наблюдений и доверительных интервалов числовых характеристик.

*Основные задачи:*

- 1) оценить изменения доверительных интервалов действительного среднего значения по выборкам различной продолжительности;
- 2) оценить однородность исходных рядов по математическому ожиданию;
- 3) оценить однородность рядов по дисперсии;
- 4) дать общую оценку однородности исходных рядов наблюдений.

*Исходные данные:*

- два временных ряда данных о состоянии объектов окружающей среды (ряды данных те же, что и в работе № 1);
- оценки числовых характеристик исходных рядов, полученные в первой работе.

*Содержание работы*

- Определить доверительные интервалы математического ожидания и дисперсии.
- Оценить однородность исходных рядов по математическому ожиданию, используя критерий Стьюдента–Госсета.
- Оценить однородность исходных рядов по дисперсии используя критерии Фишера и Романовского.
- Оценить однородность исходных рядов стока, используя непараметрический критерий Манна–Уитни.
- Сделать общий вывод об однородности исходных рядов.

### *Теоретические положения*

Для выполнения работы студенты должны изучить теоретический материал по учебникам и учебным пособиям из списка литературы. После изучения теоретического материала студенты должны знать:

- 1) статистические гипотезы и их назначение;
- 2) уровни значимости и доверительные интервалы, назначение, определение;
- 3) параметрические и непараметрические гипотезы;
- 4) критерии согласия;
- 5) оценка однородности исходных рядов по математическому ожиданию;
- 6) оценка однородности исходных рядов по дисперсии;
- 7) доверительные интервалы числовых характеристик.

Одной из важных задач при геоэкологическом анализе является оценка однородности исходной информации. Так как расчет числовых характеристик ряда является корректным, только если ряд однороден.

*В данном случае под **однородностью** понимается принадлежность всех значений исходного ряда или двух и более отдельных рядов к одной генеральной совокупности.*

Другими словами, ряд однороден, если в течение всего периода наблюдений не менялись условия формирования изучаемого процесса. В математической статистике разработаны приемы такого анализа на основе гипотез того или иного вида, которые должны быть доказаны или опровергнуты в ходе проверки однородности.

## **План выполнения задания**

### **1. Расчет числовых характеристик по всему ряду наблюдений и отдельным его частям.**

Для выполнения расчетов по оценке однородности исходных рядов необходимо ряд наблюдений разбить на две равные или неравные части, относительно которых имеются сомнения, что они относятся к и той же одной генеральной совокупности. В данном случае при общем объеме выборки в 42 значения исходный ряд разбивается на 2 части по 21 значению в каждой.

В таблице 5.1 представлены оценки числовых характеристик: среднее значение  $m_x$ , среднее квадратическое отклонение  $\sigma_x$ ,

дисперсия  $D_x$  и коэффициент вариации  $Cv$ , рассчитанные по названным частям ряда и по всему ряду. Для расчета использовать формулы (1.7–1.10).

**Таблица 5.1. Числовые характеристики по выборкам различной продолжительности ряда концентраций железа общего, р. Славянка, г. Пушкин, 2011–2016 гг.**

Ряд	$n$	$m_x$	$\sigma_x$	$Cv$	$D_x$
1 половина	21	0,26	0,13	0,50	0,017
2 половина	21	0,19	0,08	0,43	0,007
Весь ряд	42	0,22	0,11	0,50	0,013

Очевидно, что оценки числовых характеристик по представленным выборкам отличаются друг от друга. Однако возникает вопрос: объясняются ли эти отличия выборочностью исходных данных или являются следствием существенного изменения самого рассматриваемого процесса? Вместе с тем возникает и другой вопрос: в каких пределах могут изменяться числовые характеристики по выборочным данным, если описываемый процесс является стационарным?

## **2. Расчет доверительного интервала математического ожидания.**

Ответ на вопрос относительно математического ожидания даётся с помощью доверительных интервалов.

*Доверительный интервал математического ожидания* – это границы, внутри которых с определенной вероятностью находится действительное среднее значение генеральной совокупности.

Доверительный интервал в практических расчетах обычно определяется в предположении, что исходные ряды описываются нормальным законом распределения. Для этого используется формула:

$$m_x - t_\alpha \cdot \frac{\sigma_x}{\sqrt{n}} \leq m_x \leq m_x + t_\alpha \cdot \frac{\sigma_x}{\sqrt{n}}, \quad (5.1)$$

где значение  $t_\alpha$  – статистика Стьюдента–Госсета, которая определяется по таблице, представленной в приложении 9, в зависимости от числа степеней свободы  $\nu = n - 1$  и уровня значимости  $\alpha$ .

Например, для ряда наблюдений железа общего в результате оценки доверительных интервалов математических ожиданий при уровне значимости  $\alpha = 0,10$  по исходному ряду и его частям получаем:

- для всего ряда  $0,19 \leq m_x \leq 0,25$ ;
- для первой половины ряда  $0,21 \leq m_{x1} \leq 0,31$ ;
- для второй половины ряда  $0,16 \leq m_{x2} \leq 0,22$ .

В заключение, необходимо проанализировать изменение доверительных границ среднего значения в зависимости от  $n$  и  $C_v$ .

### **3. Расчет доверительного интервала дисперсии.**

Ответ на вопрос относительно дисперсии также дается с помощью доверительных интервалов.

*Доверительный интервал дисперсии, по аналогии с математическим ожиданием, это границы, внутри которых с определенной вероятностью находится действительное значение дисперсии генеральной совокупности.*

Доверительный интервал в практических расчетах обычно определяется в предположении, что исходные ряды описываются нормальным законом распределения, для этого используется формула:

$$\frac{(n-1) \cdot D_x}{\chi_{1-\alpha}^2} \leq D_x \leq \frac{(n-1) \cdot D_x}{\chi_{\alpha}^2}, \quad (5.2)$$

где  $\chi_{1-\alpha}^2$  и  $\chi_{\alpha}^2$  – значение  $\chi^2$  с вероятностью непревышения  $1 - \alpha$  и  $\alpha$ , определяется по таблице распределения  $\chi^2$ , представленной в приложении 8, в зависимости от числа степеней свободы  $\gamma = n - 1$  и уровня значимости  $\alpha$ . В данном случае  $\alpha = 0,10$ , следовательно, в приложении нам необходимо определить  $\chi_{0,9}^2$  и  $\chi_{0,1}^2$  для объемов выборки  $\gamma = 41$  и  $20$ .

Для ряда наблюдений железа общего в результате оценки доверительных интервалов дисперсий по исходному ряду и его частям получаем:

- для всего ряда  $0,010 \leq D_x \leq 0,018$ ;
- для первой половины ряда  $0,012 \leq D_{x1} \leq 0,027$ ;
- для второй половины ряда  $0,005 \leq D_{x2} \leq 0,011$ .

Также необходимо проанализировать изменение доверительных границ дисперсии в зависимости от объема выборки  $n$  и  $C_v$ .

### **4. Оценка однородности исходного ряда по математическому ожиданию с помощью критерия Стьюдента–Госсета.**

Оценка гипотез об однородности рядов наблюдений по математическому ожиданию осуществляется при помощи статистики  $t$

Стьюдента, являющейся нормированной разностью математических ожиданий двух рядов или двух частей одной выборки. Критерий Стьюдента относят к параметрическим критериям оценки однородности. В основе критерия лежит параметрическая нулевая гипотеза:

$$H_0: m_{x_1} = m_{x_2}.$$

Ряд считается неоднородным, если гипотеза при заданном уровне значимости опровергается. В противном случае гипотеза не опровергается, но это еще не значит, что она верна.

Для оценки однородности необходимо рассчитать значение статистики Стьюдента–Госсета  $t$  по формуле:

$$t = \frac{(m_{x_1} - m_{x_2})}{\sigma^*} \cdot \sqrt{\frac{n_1 \cdot n_2}{(n_1 + n_2)}}, \quad (5.3)$$

где

$$\sigma^* = \sqrt{\frac{((n_1 - 1) \cdot \sigma_{x_1}^2 + (n_2 - 1) \cdot \sigma_{x_2}^2)}{((n_1 - 1) + (n_2 - 1))}}; \quad (5.4)$$

$m_{x_1}$ ,  $m_{x_2}$  – среднее значение по 1-ой и 2-ой выборкам,  $\sigma_{x_1}$ ,  $\sigma_{x_2}$  – средние квадратические отклонения по 1-ой и 2-ой выборкам,  $n_1$ ,  $n_2$  – объем 1-ой и 2-ой выборки.

Дальше необходимо по приложению 9 определить критическое значение статистики Стьюдента  $t_\alpha$  при данном уровне значимости  $2\alpha$  (в нашем случае  $2\alpha = 10\%$ ) и числе степеней свободы  $\nu$ :

$$\nu = n_1 + n_2 - 2. \quad (5.5)$$

В нашем случае значение  $t_\alpha = 1,68$ .

Если  $t > t_\alpha$ , то гипотеза о равенстве математических ожиданий сопоставляемых выборок опровергается, ряд не является однородным. Если  $t \leq t_\alpha$ , то гипотеза не опровергается.

Результат оценки однородности для ряда концентраций железа общего в р. Славянка представлен в таблице 5.2.

**Таблица 5.2. Оценка однородности ряда концентраций железа общего, р. Славянка, г. Пушкин, 2011–2016 гг. по критерию Стьюдента**

$n_1$	$n_2$	$\sigma$	$\sigma^*$	$t$	$t_\alpha$	Результат
21	21	40	0,108	2,06	1,68	Гипотеза опровергается

Как видно из результатов, представленных в таблице 5.2 значение  $t > t_{\alpha}$ , следовательно, нулевая гипотеза  $H_0: m_{x1} = m_{x2}$  опровергается и ряд концентраций железа общего нельзя считать однородным.

### 5. Оценка однородности исходного ряда по дисперсии с помощью критериев Фишера и Романовского.

Оценка однородности рядов наблюдений по дисперсии основана на применении параметрических критериев Фишера и Романовского. Для обоих критериев нулевая гипотеза имеет вид:

$$H_0: D_{x1} = D_{x2}.$$

В основе критерия Фишера лежит оценка при помощи статистики  $F$ , которая рассчитывается по формуле:

$$F = \frac{D_{x1}}{D_{x2}}, \quad (5.6)$$

где  $D_{x1}$  и  $D_{x2}$  – оценки дисперсии по первой и второй выборке. Важно, что в числитель всегда ставится большее значение дисперсии, таким образом, критерий  $F$  всегда больше 1.

Далее, для вывода об однородности необходимо определить критическое значение статистики  $F_{\alpha}$  по приложению 10 при заданном уровне значимости  $\alpha$  ( $\alpha = 10\%$ ) и числа степеней свободы:

$$v_1 = n_1 - 1, v_2 = n_2 - 1. \quad (5.7)$$

Если  $F > F_{\alpha}$ , то гипотеза о равенстве дисперсий опровергается и ряд считать однородным нельзя. Если  $F \leq F_{\alpha}$ , то гипотеза  $H_0$  не опровергается.

Результаты расчета критерия Фишера для ряда концентраций железа общего представлены в таблице 5.3.

**Таблица 5.3. Оценка однородности ряда концентраций железа общего, р. Славянка, г. Пушкин, 2011–2016 гг. по критерию Фишера**

$D_{x1}$	$D_{x2}$	$F$	$F_{\alpha}$	Результат
0,017	0,007	2,56	1,81	Гипотеза опровергается

Как видно из таблицы 5.3,  $F > F_{\alpha}$ , следовательно, гипотеза о равенстве дисперсий опровергается, и ряд концентраций железа общего считать однородным нельзя.

В некоторых случаях для проверки гипотезы об однородности по дисперсии используют критерий Романовского, особенно если нет возможности воспользоваться таблицами критерия Фишера.

Критерий Романовского рассчитывается по формуле:

$$R = \frac{|\Theta - 1|}{\sigma_{\theta}} \geq 3, \quad (5.8)$$

где

$$\Theta = F \cdot \frac{(v_2 - 2)}{v_2}, \quad (5.9)$$

$$\sigma_{\theta} = \sqrt{\frac{2 \cdot (v_1 + v_2 - 2)}{v_1 \cdot (v_2 - 4)}}, \quad (5.10)$$

$v_1$  и  $v_2$  – число степеней свободы по первой и второй выборке (формула 5.7).

Если  $R \geq 3$ , то гипотеза о равенстве дисперсий опровергается и ряд считать однородным нельзя.

Результат оценки однородности по критерию Романовского для исследуемого ряда представлен в таблице 5.4

**Таблица 5.4. Оценка однородности ряда концентраций железа общего, р. Славянка, г. Пушкин, 2011–2016 гг. по критерию Романовского**

$\theta$	$\sigma_{\theta}$	$R$	Результат
2,30	0,49	2,7	Гипотеза не опровергается

Так как  $R$  меньше 3, то гипотеза о равенстве дисперсий по критерию Романовского не опровергается и ряд может считаться однородным.

### **6. Оценка однородности ряда наблюдений с помощью непараметрического критерия Манна–Уитни.**

Для оценки непараметрических гипотез об однородности наибольшее распространение получил критерий Манна–Уитни. В качестве критерия здесь применяется статистика  $U$ , равная общему числу инверсий в объединенной упорядоченной (ранжированной) выборке. Инверсией называется случай, когда какому либо значению  $X$  предшествует в объединенном ряду значение  $Y$ .

Для расчета инверсий необходимо разделить ряд наблюдений на две равные части (выборки) и обозначить все члены первой выборки через  $X$ , все члены второй выборки через  $Y$  (таблица 5.7). Данные обеих выборок объединить в общую последовательность и ранжировать в порядке убывания, сохраняя принятые обозначения ряда. В табличном процессоре Excel это проще всего сделать, окрасив значения

выборки  $X$  цветом. При этом при совместном ранжировании цвет останется. Например, в таблице 5.5 все значения выборки  $X$  окрашены серым цветом (первая половина исходного ряда), таким образом после объединения и ранжирования (столбец 4) значения  $X$  четко выделены в общей последовательности. Далее в столбце 5 записываются инверсии (сколько значений выборки  $Y$  стоит перед значением  $X$ ). Внизу под столбцом 5 находим значение  $U$  – сумму инверсий.

**Таблица 5.5. Оценка однородности ряда концентраций железа общего, р. Славянка, г. Пушкин, 2011–2016 гг. по критерию Манна–Уитни**

№ п/п	Исходный ряд $x_i$	Ранжированные значения $x_i$	Инверсии
1	0,32	0,54	0
2	0,18	0,51	0
3	0,17	0,45	0
4	0,2	0,4	
5	0,29	0,35	1
6	0,54	0,35	1
7	0,24	0,34	1
8	0,24	0,32	1
9	0,106	0,30	
10	0,17	0,29	2
11	0,16	0,24	2
12	0,35	0,24	2
13	0,45	0,24	2
14	0,51	0,24	
15	0,35	0,24	
16	0,22	0,23	
17	0,24	0,22	5
18	0,14	0,22	
19	0,12	0,22	
20	0,076	0,22	
21	0,34	0,21	
22	0,22	0,21	
23	0,115	0,21	
24	0,13	0,21	
25	0,21	0,2	12
26	0,21	0,19	
27	0,11	0,18	13
28	0,19	0,17	13
29	0,4	0,17	13
30	0,11	0,16	13

№ п/п	Исходный ряд $x_i$	Ранжированные значения $x_i$	Инверсии
31	0,21	0,16	
32	0,117	0,14	14
33	0,16	0,13	
34	0,21	0,12	15
35	0,22	0,117	
36	0,24	0,115	
37	0,11	0,11	
38	0,22	0,11	
39	0,02	0,11	
40	0,23	0,106	20
41	0,24	0,076	20
42	0,30	0,02	
$U$			150

В нашем случае для ряда концентраций железа общего получилось  $U = 150$ .

При объемах выборок  $n_1 > 10$  и  $n_2 > 10$  число инверсий  $U$  распределено приблизительно нормально с математическим ожиданием:

$$m_U = \frac{n_1 \cdot n_2}{2}, \quad (5.11)$$

и дисперсией:

$$D_U = \frac{n_1 \cdot n_2 \cdot (n_1 + n_2 + 1)}{12}. \quad (5.12)$$

Среднее квадратическое отклонение:

$$\sigma_U = \sqrt{D_U}. \quad (5.13)$$

Для ряда концентраций железа общего в реке Славянка получили:

$$m_U = 221, D_U = 1580, \sigma_U = 40.$$

Зная значения  $m_U$  и  $\sigma_U$ , можно легко установить критическую область для нулевой гипотезы об отсутствии существенных различий в сопоставляемых рядах:

$$U \leq m_U - t_p \cdot \sigma_U, \quad (5.14)$$

$$U \geq m_U + t_p \cdot \sigma_U, \quad (5.15)$$

где  $t_p$  нормированная ордината кривой обеспеченности нормального закона распределения при  $p = 1 - \alpha$ .

Если для рассчитанного для исходного ряда значения  $U$  будет выполняться одно из этих условий (5.14 или 5.15), то гипотеза об однородности анализируемого ряда опровергается. Если же полученное значение  $U$  будет лежать в пределах доверительного интервала, то гипотеза об однородности ряда не опровергается.

Для исследуемого ряда концентраций железа общего получили  $154 \leq U \leq 287$ . В таблице 5.7 было получено  $U = 150$ , следовательно условие 5.14 выполняется, рассчитанное значение  $U$  не попадает в доверительный интервал, и следовательно гипотеза об однородности ряда опровергается.

Общий вывод об однородности ряда делается по совокупности всех исследуемых критериев. Так если хотя бы по одному из критериев гипотеза опровергается, то ряд нельзя считать однородным.

Для ряда концентраций железа общего в р. Славянка гипотеза об однородности опровергается по трём критериям: Стьюдента, Фишера и Манна–Уитни, и не опровергается по критерию Романовского. Следовательно, общий вывод – ряд концентраций железа общего нельзя считать однородным.

## **Лабораторная работа № 6. Исследование взаимосвязи двух переменных величин**

*Цель работы:* изучить методы анализа взаимосвязи двух исследуемых процессов, представленных рядами  $Y$  и  $X$ .

*Задачи работы:*

- 1) исследовать взаимосвязь двух процессов ( $Y$  и  $X$ );
- 2) оценить значимость и стационарность связи;
- 3) определить возможность использования взаимосвязи для прогнозирования или восполнения пропусков в наблюдениях.

*Исходные данные:* два временных ряда данных о состоянии объектов окружающей среды (ряды данных те же, что и в работе № 1).

*Содержание работы*

- Построить в поле декартовых координат график зависимости  $Y = f(X)$ .
- Определить параметры уравнения регрессии  $Y$  по  $X$  и  $X$  по  $Y$ .
- Оценить погрешности расчетов и доверительные интервалы коэффициента корреляции и параметров уравнений регрессии.
- Оценить значимость коэффициента корреляции.

- Оценить стационарность связи.
- Оценить точность расчетов по уравнению регрессии на независимом материале.
- Оценить адекватность полученной модели парной линейной корреляции.

#### *Теоретические положения*

Для выполнения работы студенты должны изучить теоретический материал по учебникам и учебным пособиям из списка литературы. После изучения теоретического материала студенты должны знать:

- 1) взаимосвязь природных процессов, виды взаимосвязи;
- 2) функциональная и стохастическая связи;
- 3) парная корреляция, назначение, основные свойства, граничные условия;
- 4) коэффициент корреляции;
- 5) уравнение регрессии и его параметры;
- 6) оценка коэффициента корреляции, преобразование Фишера;
- 7) стационарность связи;
- 8) оценка адекватности модели парной линейной корреляции.

В геоэкологических исследованиях большое внимание уделяется анализу взаимосвязей между различными природными процессами. Дело в том, что геоэкологические процессы формируются цепью причинно-следственных связей, изменяющихся во времени и пространстве. Для изучения этих процессов необходимо установить их причины, движущие силы или источники и тенденцию развития.

Кроме того, исследование связей часто имеет и чисто практический выход. Эти исследования широко применяются в прогнозах и в расчетах состояния рассматриваемых процессов. Естественно, что именно при этом возникает необходимость в исследовании зависимостей, связывающих рассматриваемые процессы между собой и с другими процессами и явлениями. Взаимосвязь исследуемых процессов оценивается по соответствию изменений их значений во времени или пространстве. Эта взаимосвязь может быть линейной или нелинейной, более или менее тесной. По тесноте связи, или иначе, степени определенности одного из сопоставляемых процессов относительно другого, связи природных процессов, если они есть, могут быть разделены на функциональные и стохастические.

*Функциональными связями между процессами  $Y$  и  $X$  называются такие связи, когда каждому значению  $X$  соответствует только одно возможное значение  $Y$ .*

*Стохастическими (вероятностными) связями называются связи, когда каждому значению  $X$  соответствует определенное множество значений  $Y$ .*

Связи могут быть линейными и нелинейными. В данном случае будем рассматривать линейные связи.

## План выполнения работы

В качестве примера в работе используются два ряда наблюдений за концентрациями железа общего в р. Славянка в районе г. Пушкин на соседних створах (створ 1 и створ 2).

При построении модели парной линейной корреляции  $Y = f(X)$  необходимо определить какой из рядов выступит в роли  $X$ , а какой – в роли ряда  $Y$ . В данном случае логично предположить, что ряд, полученный в створе наблюдения выше по течению (створ 1), логичнее принять за  $X$ . А ряд, расположенный ниже по течению (створ 2), – за  $Y$ .

Для выполнения данной лабораторной работы необходимо до начала расчетов и построений из обоих рядов наблюдений исключить значения под номерами 1, 11, 21, 31 и 41. Все дальнейшие расчеты вести без них.

### 1. Визуальная оценка тесноты связи.

Для визуальной оценки тесноты связи необходимо нанести точки связи  $Y = f(X)$  в поле графика в декартовых координатах.

По оси  $X$  откладываем значения ряда  $X$ , по оси  $Y$  значения ряда  $Y$ . Необходимо построить точки связи ранжированных и не ранжированных значений (рис. 6.1). График связи ранжированных значений используется для более точного построения линии связи. После построения точек связи необходимо обвести облако точек и провести линию связи от руки. Подобрать масштаб таким образом, чтобы линия связи проходила примерно под углом 45 градусов.

Визуальная оценка тесноты связи проводится на основе вида облака точек. В целом, чем более широко разбросаны точки в поле графика, тем менее тесная связь. И наоборот, чем ближе расположены точки, тем связь теснее. В нашем случае для рядов концентраций железа общего можно сделать вывод что связь достаточно тесная.

### 2. Расчет коэффициента корреляции.

Коэффициент корреляции позволяет оценить характер (прямая или обратная) и тесноту связи.

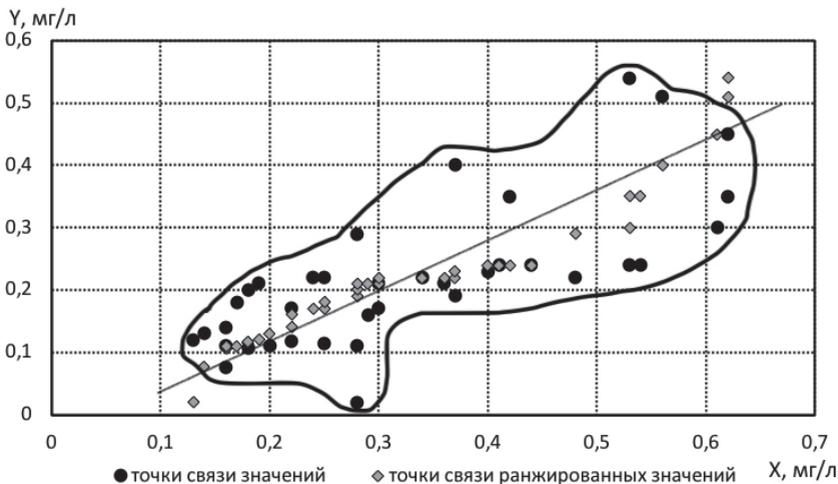


Рис. 6.1. График связи исходных и ранжированных значений концентраций железа общего, р. Славянка, г. Пушкин, 2011–2016 гг.

Расчет коэффициента корреляции  $r_{yx}$  между исходными рядами  $Y$  и  $X$  производится по формуле:

$$r_{xy} = \frac{\sum \delta y \cdot \delta x}{(n-1) \cdot \sigma_y \cdot \sigma_x} = \frac{\sum (y_i - y_{cp}) \cdot (x_i - x_{cp})}{(n-1) \cdot \sigma_y \cdot \sigma_x}, \quad (6.1)$$

где  $\sigma_y, \sigma_x$  – средние квадратические отклонения рядов  $Y$  и  $X$ ;  $\delta y$  и  $\delta x$  – центрированные значения случайных величин  $Y$  и  $X$ :

$$\delta y = y_i - y_{cp}, \quad \delta x = x_i - x_{cp}.$$

Расчет коэффициента корреляции производится в таблице 6.1.

**Таблица 6.1. Расчет коэффициента корреляции и параметров уравнения регрессии для рядов концентраций железа общего, р. Славянка, г. Пушкин (створ 1 и створ 2), 2011–2016 гг.**

№	$Y$	$X$	$\delta x$	$\delta y$	$\delta x^2$	$\delta y^2$	$\delta x \cdot \delta y$
1	0,18	0,17	-0,16	-0,04	0,0253	0,0015	0,0063
2	0,17	0,22	-0,11	-0,05	0,0119	0,0024	0,0054
3	0,2	0,18	-0,15	-0,02	0,0223	0,0004	0,0029
4	0,29	0,28	-0,05	0,07	0,0024	0,0050	-0,0035
5	0,54	0,53	0,20	0,32	0,0403	0,1029	0,0644
6	0,24	0,54	0,21	0,02	0,0444	0,0004	0,0044
7	0,24	0,44	0,11	0,02	0,0123	0,0004	0,0023
8	0,106	0,18	-0,15	-0,11	0,0223	0,0128	0,0169

№	Y	X	$\delta x$	$\delta y$	$\delta x^2$	$\delta y^2$	$\delta x \cdot \delta y$
9	0,17	0,3	-0,03	-0,05	0,0009	0,0024	0,0014
10	0,35	0,62	0,29	0,13	0,0846	0,0171	0,0380
...	...	...	...	...	...	...	...
28	0,117	0,22	-0,11	-0,10	0,0119	0,0105	0,0112
29	0,16	0,29	-0,04	-0,06	0,0015	0,0035	0,0023
30	0,21	0,3	-0,03	-0,01	0,0009	0,0001	0,0003
31	0,22	0,34	0,01	0,00	0,0001	0,0000	0,0000
32	0,24	0,53	0,20	0,02	0,0403	0,0004	0,0042
33	0,11	0,28	-0,05	-0,11	0,0024	0,0119	0,0054
34	0,22	0,48	0,15	0,00	0,0227	0,0000	0,0001
35	0,02	0,28	-0,05	-0,20	0,0024	0,0397	0,0098
36	0,23	0,40	0,07	0,01	0,0050	0,0001	0,0008
37	0,30	0,61	0,28	0,08	0,0789	0,0065	0,0227
Сумма	8,114	12,18	0,00	0,00	0,8077	0,4865	0,4651

Под таблицей необходимо найти суммы значений для столбцов 2–8. Затем для рядов X и Y необходимо рассчитать средние значения и средние квадратические отклонения.

Для ряда Y получаем:

$$m_y = 0,22; \sigma_y = 0,12.$$

Для ряда X:

$$m_x = 0,33; \sigma_x = 0,15.$$

Подставляя полученные значения в формулу (6.1), находим:

$$r = 0,4651 / (36 \cdot 0,15 \cdot 0,12) = 0,74.$$

Исходя из полученного значения коэффициента корреляции, можно утверждать, что, во-первых, связь прямая (так как  $r > 0$ ), то есть Y возрастает с возрастанием X. Во-вторых, значима, так как  $r$  намного превышает 0 (при функциональной связи  $r = 1$ , а при  $r = 0$  связь отсутствует).

### 3. Расчет параметров уравнения регрессии и построение линий регрессии $Y = f(X)$ и $X = f(Y)$ .

Следующим этапом исследования связи  $Y = f(X)$  является определение формы связи. Для описания линейной связи может быть использовано уравнение прямой линии:

$$Y = a \cdot X + b. \quad (6.2)$$

В уравнении (6.2) неизвестны параметры  $a$  и  $b$ . Расчет этих параметров выполняется по формулам (6.3–6.4):

$$a = \frac{\sum \delta y \cdot \delta x}{\sum \delta x^2} = \frac{\sum (y_i - y_{cp}) \cdot (x_i - x_{cp})}{\sum (x_i - x_{cp})^2}, \quad (6.3)$$

$$b = m_y - a \cdot m_x. \quad (6.4)$$

Для наших рядов, подставляя в формулы (6.3–6.4) значения, полученные в таблице 6.1, находим:

$$a = 0,4651/0,8077 = 0,58; \quad b = 0,22 - 0,58 \cdot 0,33 = 0,03.$$

Таким образом, для ряда концентраций железа общего в реке Славянка уравнение регрессии (6.2) имеет вид:

$$y = 0,58 \cdot x + 0,03. \quad (6.5)$$

Для описания линейной связи  $X = f(Y)$  применяется обратное уравнение регрессии вида:

$$X = a' \cdot Y + b'. \quad (6.6)$$

Параметры  $a'$  и  $b'$  определяются по формулам:

$$a' = \frac{\sum \delta y \cdot \delta x}{\sum \delta y^2} = \frac{\sum (y_i - y_{cp}) \cdot (x_i - x_{cp})}{\sum (y_i - y_{cp})^2}, \quad (6.7)$$

$$b' = m_x - a' \cdot m_y. \quad (6.8)$$

Для наших рядов, подставляя в формулы (6.7–6.8) значения из таблицы 6.1, получаем:

$$a' = 0,4651/0,4865 = 0,96; \quad b' = 0,33 - 0,96 \cdot 0,22 = 0,12.$$

Подставляя полученные значения в уравнение (6.6), находим:

$$x = 0,96 \cdot y + 0,12. \quad (6.9)$$

Для построения линий прямой и обратной регрессии необходимо вычислить их координаты. Расчет координат каждой прямой производится обычно для трех точек. Для прямой связи задаются  $X_{\min}$ ,  $X_{cp}$ ,  $X_{\max}$ , и по уравнению регрессии (6.5) рассчитываются значения  $Y$ . Для обратной связи задаются  $Y_{\min}$ ,  $Y_{cp}$ ,  $Y_{\max}$ , и по уравнению регрессии (6.9) рассчитываются значения  $X$  (таблица 6.2).

**Таблица 6.2. Координаты уравнений регрессии  $Y = f(X)$  и  $X = f(Y)$**

$y = f(x)$	$X_{\min}$	0,13	$Y(X_{\min})$	0,10
	$X_{cp}$	0,33	$Y(X_{cp})$	0,22
	$X_{\max}$	0,62	$Y(X_{\max})$	0,39
$x = f(y)$	$Y_{\min}$	0,02	$X(Y_{\min})$	0,14
	$Y_{cp}$	0,22	$X(Y_{cp})$	0,33
	$Y_{\max}$	0,54	$X(Y_{\max})$	0,64

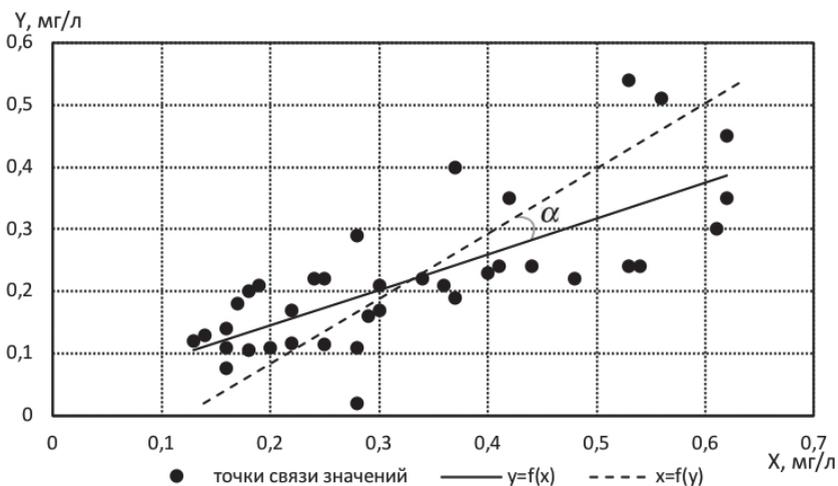


Рис. 6.2. График связи исходных значений и уравнения регрессии для рядов концентраций железа общего, р. Славянка, г. Пушкин, 2011–2016 гг.

Полученные координаты наносятся в поле графика, на котором уже представлены точки связи (рис. 6.2).

Из рис. 6.2 видно, что линии прямого и обратного уравнений регрессии пересекаются в точке  $(m_y, m_x)$  под углом  $\alpha$ . Чем более тесная связь между  $X$  и  $Y$ , тем меньше угол между линиями регрессии. Для функциональной связи угол  $\alpha = 0$ , а в случае отсутствия связи  $\alpha = 90^\circ$ . В нашем случае  $\alpha = 10^\circ$ , что говорит о том, что связь достаточно тесная.

#### 4. Оценка коэффициента корреляции и параметров уравнения регрессии по выборочным данным.

Коэффициент корреляции и параметры уравнения регрессии в практических расчетах определяются по выборкам и, естественно, что они носят выборочный характер и могут использоваться лишь как более или менее достоверные оценки действительных значений. Поэтому расчеты по уравнениям регрессии всегда сопровождаются оценкой погрешностей определения параметров и оценкой их устойчивости.

Для этого, прежде всего, необходимо рассчитать средние квадратические погрешности и доверительные интервалы коэффициента корреляции при уровне значимости  $2\alpha$  (в данном случае  $2\alpha = 10\%$ ).

Задача осложняется тем, что закон распределения выборочных значений коэффициента корреляции при больших его значениях является асимметричным. Поэтому оценка коэффициента корреляции производится разными методами в зависимости от величины коэффициента корреляции.

Так, при  $r_{yx} < 0,4$  или/и  $n > 30$  оценка производится на основе учета средней квадратической погрешности коэффициента корреляции:

$$\sigma_r = \frac{(1-r^2)}{\sqrt{(n-1)}}. \quad (6.10)$$

По значению  $\sigma_r$  определяется доверительный интервал коэффициента корреляции:

$$r_{yx} - t_\alpha \cdot \sigma_r < r_{yx} < r_{yx} + t_\alpha \cdot \sigma_r, \quad (6.11)$$

где  $t_\alpha$  – статистика Стьюдента (см. работу 5) при двухстороннем уровне значимости.

Для ряда концентраций железа общего получаем:

$$\sigma_r = 0,075.$$

Доверительный интервал коэффициента корреляции:

$$0,62 < r_{yx} < 0,87.$$

При малых значениях коэффициента корреляции возникает вопрос о реальности связи вообще, то есть вопрос о том, являются ли полученные коэффициенты корреляции значимыми и не объясняются ли они случайностями выборок. Обычно этот вопрос решается на основе нулевой гипотезы относительно рассматриваемой связи  $H_0: r = 0$ . Оценка гипотезы производится по выборочному коэффициенту корреляции. Для этого в соответствии с формулой (6.11) строится критическая область нулевой гипотезы:

$$|r| > t_\alpha \cdot \sigma_r. \quad (6.12)$$

Если неравенство 6.12 выполняется, то гипотеза о равенстве действительного значения коэффициента корреляции нулю опровергается, то есть связь является значимой. Если неравенство 6.12 не выполняется, то гипотеза не опровергается и коэффициент корреляции не значим.

Для ряда концентраций железа общего в реке Славянка  $0,74 > 0,13$ , следовательно, гипотеза  $H_0: r = 0$  опровергается, действительное значение коэффициента корреляции больше 0 и связь значима.

Если  $r_{yx} \geq 0,4$  или/и  $n \leq 30$ , то для оценки коэффициента корреляции и его доверительного интервала используется преобразование Фишера:

$$r = th(z), \quad (6.13)$$

где  $th(z)$  – гиперболический тангенс  $z$ .

Отсюда:

$$z = \frac{\ln \left[ \frac{(1+r)}{(1-r)} \right]}{2}. \quad (6.14)$$

Зная выборочное значение  $z$ , можно построить доверительный интервал:

$$z - t_{\alpha} \cdot \sigma_z \leq z < z + t_{\alpha} \cdot \sigma_z, \quad (6.15)$$

где

$$\sigma_z = \sqrt{\frac{1}{n-3}}. \quad (6.16)$$

Отсюда, переведя верхнюю и нижнюю доверительную границу  $z$  ( $z_n$  и  $z_b$ ) по формуле (6.13) в значения коэффициента корреляции, получаем доверительный интервал коэффициента корреляции.

При ручном счете переход от  $r$  к  $z$  и обратно является довольно трудоемким процессом. При расчете в табличном процессоре Excel определение  $z$  по  $r$  проводится по встроенной функции «ФИШЕР», а переход к  $r$  от  $z$  по функции обратного преобразования Фишера «ФИШЕРОБР». Также можно воспользоваться формулами (6.13–6.14).

Для ряда концентраций железа общего получаем:

$$z = 0,95; \sigma_z = 0,17.$$

Доверительный интервал  $z$ :

$$0,67 \leq z < 1,24.$$

Отсюда доверительный интервал  $r$ :

$$0,58 \leq r < 0,85.$$

Для оценки значимости коэффициента корреляции в данном случае (при  $r > 0,4$ ), как и выше, необходимо перейти к преобразованию Фишера. При этом проверяется нулевая гипотеза:  $H_0: z = 0$ .

Доверительный интервал этой гипотезы определяется по формуле (6.17):

$$z_{кр} = t_{\alpha} \cdot \sigma_z. \quad (6.17)$$

Если  $z > z_{кр}$ , то гипотеза об отсутствии связи опровергается. Если  $z < z_{кр}$ , то связь между двумя процессами не значима.

Для рядов концентраций железа общего получаем  $z_{кр} = 0,29$ . Так как  $z > z_{кр}$  ( $0,95 > 0,29$ ), следовательно гипотеза опровергается и связь между двумя процессами значима.

Заметим, что расчет для случая, когда  $r < 0,4$  приводится лишь как пример, и практического значения не имеет, так как в данном случае, учитывая величину коэффициента корреляции, необходимо использовать преобразование Фишера.

Оценка возможных погрешностей при определении параметров уравнения регрессии  $a$  и  $b$  производится по средней квадратической погрешности этих характеристик:

$$\sigma_a = \frac{\sigma_y}{\sigma_x} \cdot \sqrt{\frac{1-r^2}{n-1}}, \quad (6.18)$$

$$\sigma_b = \sigma_{\Delta} \cdot \sqrt{\frac{1}{n} + \frac{m_x^2}{\sigma_x^2 \cdot (n-1)}}. \quad (6.19)$$

Доверительные интервалы значений  $a$  и  $b$  определяются по формулам:

$$a - t_{\alpha} \cdot \sigma_a \leq a < a + t_{\alpha} \cdot \sigma_a, \quad (6.20)$$

$$b - t_{\alpha} \cdot \sigma_b \leq b < b + t_{\alpha} \cdot \sigma_b. \quad (6.21)$$

Для ряда концентраций железа общего получаем:

$$\sigma_a = 0,09, \quad \sigma_b = 0,03.$$

Доверительные интервалы:

$$0,43 \leq a < 0,72,$$

$$-0,02 \leq b < 0,08.$$

## 5. Расчет средней квадратической погрешности расчетов по уравнению регрессии.

Фактические значения  $Y$  могут существенно отклоняться от оценок, полученных по уравнению регрессии. При этом, чем меньше теснота связи, тем больше могут быть эти отклонения. Таким образом, оценка  $Y$  по  $X$  на основе уравнения регрессии всегда сопровождается погрешностями, во-первых, за счет аппроксимации стохастической связи прямой линией, во-вторых, за счет выборочности исходных данных. Для оценки погрешностей по уравнению регрессии обычно определяется средняя квадратическая погрешность

расчетов за счет разброса точек относительно аппроксимирующей линии связи:

$$\sigma_{\Delta} = \sigma_y \cdot \sqrt{(1-r^2)}. \quad (6.22)$$

При расчете по формуле (6.22) погрешность принимается постоянной и не зависит от  $X$ . Для ряда концентраций железа общего получаем:

$$\sigma_{\Delta} = 0,08.$$

В некоторых случаях, когда требуются более точные расчеты, в погрешностях определения  $Y$  по уравнению регрессии учитывается также погрешность аппроксимации стохастической связи прямой линией. В этом случае используется более общая и более точная формула:

$$\sigma_{y/x} = \sigma_{\Delta} \cdot \sqrt{1 + \frac{1}{n} + \frac{t^2}{n}}, \quad (6.23)$$

где  $t$  – нормированное значение  $x$ , определяемое по формуле:

$$t = \frac{(x_i - x_{cp})}{\sigma_x}. \quad (6.24)$$

Доверительный интервал уравнения регрессии в поле графика  $Y = f(X)$  рассчитывается по формуле:

$$y - t_{\alpha} \cdot \sigma_{y/x} < y \leq y + t_{\alpha} \cdot \sigma_{y/x}. \quad (6.25)$$

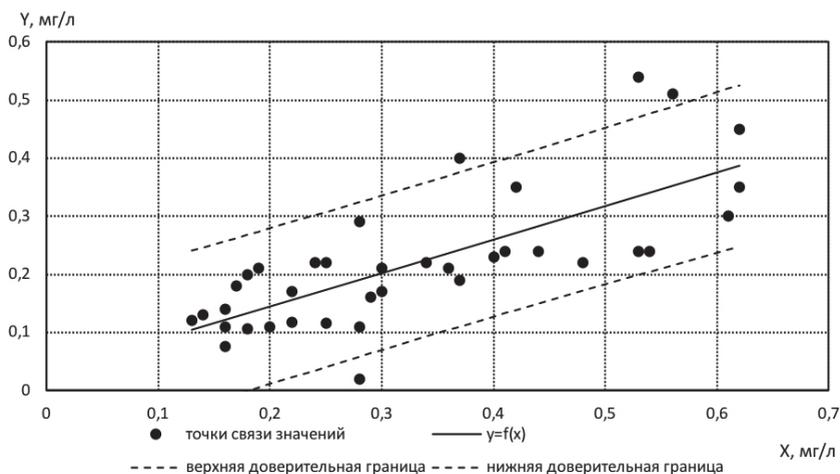
Расчет доверительного интервала уравнения регрессии для рядов концентраций железа общего представлен в таблице 6.3. Здесь во 2-м столбце приведены ранжированные значения ряда  $X$ . В 3-м столбце  $y(x)$  – значения рассчитываются по уравнению регрессии (6.2), используя значения ряда  $X$  и параметры  $a$  и  $b$ . В 4-м столбце  $t$  – нормированное значение, рассчитанное по формуле (6.24). Значение  $\sigma_{y/x}$  рассчитывается по формуле (6.23). В 6-м и 7-м столбцах представленные нижняя и верхняя доверительные границы рассчитаны по формулам (6.25).

Доверительные границы уравнения регрессии наносятся на рис. 6.3. Для нижней границы используются координаты из столбца 2 и 6, для верхней – из столбца 2 и 7.

Очевидно, что в данном случае три точки выходят за верхнюю доверительную границу и одна точка за нижнюю доверительную границу. Подобная ситуация требует дополнительного анализа,

**Таблица 6.3. Расчет доверительных границ уравнения регрессии для рядов концентраций железа общего, р. Славянка, г. Пушкин, 2011–2016 гг.**

№ п/п	Ранжированные значения ряда $X$	$y(x)$	$t$	$\sigma_{y/x}$	$y(x)_n$	$y(x)_g$
1	0,13	0,10	-1,33	0,081	-0,03	0,24
2	0,14	0,11	-1,26	0,081	-0,03	0,25
3	0,16	0,12	-1,13	0,080	-0,01	0,26
4	0,16	0,12	-1,13	0,080	-0,01	0,26
5	0,16	0,12	-1,13	0,080	-0,01	0,26
6	0,17	0,13	-1,06	0,080	-0,01	0,26
7	0,18	0,13	-1,00	0,080	0,00	0,27
8	0,18	0,13	-1,00	0,080	0,00	0,27
9	0,19	0,14	-0,93	0,080	0,00	0,27
10	0,2	0,14	-0,86	0,080	0,01	0,28
11	0,22	0,16	-0,73	0,080	0,02	0,29
12	0,22	0,16	-0,73	0,080	0,02	0,29
...	...	...	...	...	...	...
31	0,53	0,33	1,34	0,081	0,20	0,47
32	0,53	0,33	1,34	0,081	0,20	0,47
33	0,54	0,34	1,41	0,081	0,20	0,48
34	0,56	0,35	1,54	0,081	0,22	0,49
35	0,61	0,38	1,87	0,083	0,24	0,52
36	0,62	0,39	1,94	0,083	0,25	0,53
37	0,62	0,39	1,94	0,083	0,25	0,53



**Рис. 6.3. Точки связи, уравнение регрессии и его доверительные границы для рядов концентраций железа общего, р. Славянка, г. Пушкин, 2011–2016 гг.**

в частности необходимо оценить характер связи при исключении этих точек из исходных рядов наблюдений. В данном случае, при исключении этих 4 точек, значение коэффициента корреляции составило 0,79. Это говорит о том, что исключение отскакивающих точек влияет на характер связи, и это обстоятельство требует дальнейшего научного изучения.

### 6. Оценить стационарность связи.

Одним из основных условий возможности применения уравнения регрессии в практических расчетах является стационарность связи.

В настоящее время для проверки стационарности связи чаще всего используется метод, основанный на сопоставлении параметров уравнения регрессии по различным выборкам из имеющихся рядов наблюдений и проверки их равенства, исходя из нулевой гипотезы:

$$H_0: r_1 = r_2,$$

где  $r_1$  и  $r_2$  – действительные значения коэффициентов корреляции по первой и второй частям совместного периода наблюдений  $Y$  и  $X$ . Оценка гипотезы в этом случае производится с помощью критерия Стьюдента. Если величина статистики окажется меньше критической при данном уровне значимости, то гипотеза не опровергается. В противном случае гипотеза опровергается.

Таким образом, на первом этапе для оценки стационарности потребуется определить коэффициенты корреляции  $r_1$  и  $r_2$ . Для этого необходимо ряды  $X$  и  $Y$  (таблица 6.1) разбить на две примерно равные части. Так как общая продолжительность рядов в таблице 6.1  $n = 37$ , то первая часть – это значения с 1 по 18, а вторая часть – с 19 по 37. Таким образом:  $n_1 = 18$ ,  $n_2 = 19$ . Повторив для двух половин рядов действия из п. 2, получаем:

$$r_1 = 0,83; r_2 = 0,60.$$

При  $r < 0,4$  и/или  $n > 30$  оценка статистики производится непосредственно по коэффициенту корреляции:

$$t = \frac{(r_1 - r_2)}{\sigma_{r_1 - r_2}}, \quad (6.26)$$

где  $\sigma_{r_1 - r_2}$  – среднее квадратическое отклонение разности выборочных коэффициентов корреляции, рассчитанных по первой и второй части совместного периода наблюдений:

$$\sigma_{r_1 - r_2} = \sqrt{\sigma_{r_1}^2 + \sigma_{r_2}^2}. \quad (6.27)$$

Если  $t < t_\alpha$ , то гипотеза не опровергается и связь стационарна. Если  $t > t_\alpha$ , то гипотеза опровергается и связь не стационарна.

При  $r > 0,4$  и/или  $n < 30$  для расчета статистики используется преобразование Фишера. В этом случае:

$$t = \frac{(z_1 - z_2)}{\sigma_{z_1 - z_2}}, \quad (6.28)$$

где  $\sigma_{z_1 - z_2}$  – среднее квадратическое отклонение разности выборочных значений:

$$\sigma_{z_1 - z_2} = \sqrt{\sigma_{z_1}^2 + \sigma_{z_2}^2} = \sqrt{\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}}. \quad (6.29)$$

Если  $t < t_\alpha$ , то гипотеза не опровергается и связь стационарна. Если  $t > t_\alpha$ , то гипотеза опровергается, и связь не стационарна.

Исходя из значений  $r_1$  и  $r_2$ , полученных по двум половинам рядов концентраций железа общего, определяем, что оценку стационарности необходимо выполнить через преобразование Фишера используя формулы (6.28–6.29). Для расчета  $t$  необходимо определить значения  $z_1$  и  $z_2$  (см. формулу (6.14) п. 4). Для наших рядов получаем:

$$z_1 = 1,18; z_2 = 0,69; \sigma_{z_1 - z_2} = 0,36; t = 1,36; t_\alpha = 1,68.$$

Так как  $t < t_\alpha$ , то гипотеза не опровергается и связь стационарна.

### **7. Оценка точности расчетов по уравнению регрессии на независимом материале.**

Одним из условий адекватности математической модели парной линейной корреляции является соответствие между предсказаниями на основе математической модели и реальным поведением объекта. Для проверки этого условия необходимо оценить совпадение рассчитанных по уравнению регрессии значений  $Y_p$  с фактическими  $Y$  на независимом материале. То есть на материале, который не был использован при идентификации параметров уравнения регрессии. В данном случае при расчете параметров уравнения регрессии не использовались (см. выше) 1, 11, 21, 31, 40 значения рядов  $X$  и  $Y$ . Результаты расчетов на независимом материале приведены в таблице 6.4. Здесь во 2-м и 3-м столбце представлены фактические значения рядов. В 4-м столбце значения  $Y_p$  рассчитаны по уравнению регрессии (6.5). Далее в столбце 5 рассчитывают разницу фактических и рассчитанных значений:  $|Y - Y_p|$ .

**Таблица 6.4. Восстановление пропусков  
в исходных рядах наблюдений для ряда концентраций железа общего,  
р. Славянка, г. Пушкин, 2011–2016 гг.**

№	$X$	$Y$	$Y_p$	$ Y - Y_p $
1	0,49	0,32	0,31	0,01
11	0,25	0,16	0,17	0,01
21	0,74	0,34	0,46	0,12
31	0,31	0,21	0,21	0,00
41	0,49	0,24	0,31	0,07

По методике принятой в расчетах и прогнозах совпадение расчетной или прогнозной величины  $Y_p$  с фактической величиной  $Y$  считается удовлетворительным, если выполняются условия:

$$|Y - Y_p| < 0,674 \cdot \sigma_y;$$

В нашем случае:

$$0,674 \cdot \sigma_y = 0,076.$$

Таким образом, при проверке на независимом материале четыре результата из пяти оказались удовлетворительными.

### **8. Оценка адекватности математической модели парной линейной корреляции.**

В заключение исследования связи рассматриваемых процессов обычно проводится анализ адекватности подобранной математической модели линейной парной корреляции, т. е. определяется насколько хорошо и в достаточной ли мере определяет эта модель характеристики связи между рядами  $Y$  и  $X$  и может ли она использоваться на практике для определения  $Y$  по  $X$ .

Например, в гидрологических и геоэкологических прогнозах методика, основанная на той или иной математической модели, в том числе на математической модели линейной парной корреляции, во многих случаях считается эффективной (адекватной), если выполняются следующие условия:

$$\sigma_{\Delta}/\sigma_y \leq 0,70, \text{ при } n \leq 15;$$

$$\sigma_{\Delta}/\sigma_y \leq 0,75, \text{ при } 15 < n < 25;$$

$$\sigma_{\Delta}/\sigma_y \leq 0,80, \text{ при } n \geq 25,$$

Еще одним из критериев, при котором математическая модель парной корреляции может использоваться при решении ряда задач, является выполнение условия:

$$|r| \geq 0,7.$$

В результате расчетов нами получены следующие показатели адекватности используемой математической модели парной корреляции:

$$r_{yx} = 0,74; \sigma_y = 0,12; \sigma_{\Delta} = 0,08.$$

В данном случае отношение  $\sigma_{\Delta}/\sigma_y = 0,67$ . Отсюда, исходя из ранее изложенных требований, данное уравнение регрессии может использоваться на практике. Кроме того, учитывая что  $|r| \geq 0,7$ , мы также можем сделать вывод об адекватности математической модели. Об этом же говорят расчеты на независимом материале.

## Лабораторная работа № 7. Исследование взаимосвязи ряда переменных величин

*Цель работы:* изучить методы анализа взаимосвязи ряда переменных, представленных рядами  $Y$ ,  $X_1$  и  $X_2$ , на основе математической модели множественной линейной корреляции.

*Основные задачи:*

- 1) исследовать связь процессов  $Y$ ,  $X_1$  и  $X_2$  и определить ее характеристики на основе математической модели множественной линейной корреляции;
- 2) оценить стационарность связи;
- 3) определить адекватность математической модели множественной линейной корреляции.

*Исходные данные:* три временных ряда данных о концентрациях ( $Y$ ,  $X_1$ ,  $X_2$ ) за совместный период наблюдений, два из которых использовались в предыдущих работах.

*Содержание работы*

- Рассчитать коэффициенты корреляции и проанализировать связь между рядом  $Y$  и рядами  $X_j$  ( $j = 1, 2, \dots, m$ ). Исключить дублирующие и неэффективные аргументы, если они есть.
- Определить параметры уравнения множественной линейной корреляции.
- Оценить погрешности расчетов параметров уравнений регрессии и погрешности расчетов по уравнению регрессии.
- Определить с помощью сводного коэффициента корреляции общую определенность ряда  $Y$  по отношению к рядам  $X$ . Найти доверительные интервалы сводного коэффициента корреляции.

- Оценить стационарность связи  $Y = f(X_1, X_2)$ .
- Оценить точность расчетов по уравнению регрессии на независимом материале.
- Сделать выводы об адекватности полученной математической модели.

*Теоретические положения*

Для выполнения работы студенты должны изучить теоретический материал по учебникам и учебным пособиям из списка литературы. После изучения теоретического материала студенты должны знать:

- 1) этапы и задачи исследования многофакторных связей;
- 2) математическая модель множественной линейной корреляции, структура, параметры;
- 3) потенциальный набор переменных в уравнении множественной корреляции;
- 4) неэффективные и дублирующие переменные;
- 5) оценка структуры и параметров модели по выборочным данным;
- 6) граничные условия математической модели парной корреляции;
- 7) учет граничных условий в практических расчетах;
- 8) графические методы построения уравнений множественной корреляции;
- 9) оценка адекватности математической модели множественной линейной корреляции.

Наряду с рассмотренными в предыдущей лабораторной работе парными связями  $Y = f(X)$  в геоэкологических исследованиях часто встречаются ситуации, когда одна случайная величина  $Y$  каким-либо образом определена по отношению к множеству случайных величин  $X_j$  ( $j = 1, 2, \dots, m$ ,  $m$  – число случайных величин  $X$ ). Эта определенность используется для описания различных природных явлений и часто реализуется в виде математической модели линейной множественной корреляции:

$$\delta y_i = \sum_{j=1}^m a_j \cdot \delta x_{ji},$$

где  $i$  – порядковый номер члена ряда,  $j$  – номер ряда  $X$ ,  $m$  – число аргументов,  $a_j$  – коэффициент веса или влияния  $j$ -го аргумента.

Практическое использование названной математической модели обычно проводится в четыре этапа. На первом этапе обычно

производится набор потенциальных предикторов, на втором этапе исключаются неэффективные аргументы и один из пары дублирующих аргументов. На третьем этапе производится просеивание оставшихся аргументов и определяется расчетная формула уравнения множественной корреляции, включающая только эффективные аргументы. На четвертом этапе по известным функциям влияния каждого эффективного фактора определяются неизвестные значения исследуемого процесса.

## План выполнения работы

Теоретические и технические аспекты применения множественной корреляции в полном объеме изложены в учебнике [1, 2]. В данном случае необходимо детально рассмотреть наиболее простую схему использования множественной корреляции, когда учитываются только 2 аргумента, то есть решается зависимость  $Y = f(X_1, X_2)$ . В качестве примера используются ряды:  $Y$  – ряд концентраций железа общего в г. Пушкин (створ 2),  $X_1$  – ряд концентраций железа общего в створе выше по течению (створ 1) (см. работу 6). За ряд  $X_2$  принимаем ряд концентраций растворенного кислорода в р. Славянка в г. Пушкин (створ 1) за аналогичный период.

Для проверки полученного соотношения на независимом материале необходимо до начала расчетов и построений из трёх рядов наблюдений исключить значения под номерами 1, 11, 21, 31 и 41.

### **1. Расчет парных коэффициентов корреляции между исходными рядами, оценка эффективности предикторов и наличия дублирующих переменных.**

В общем случае набор потенциальных предикторов и их количество устанавливается перед началом расчетов. Далее производится проверка на наличие в наборе дублирующих аргументов. Затем из набора исключаются неэффективные предикторы.

Для решения этих задач на первом этапе необходимо определить парные коэффициенты корреляции и погрешности их расчета. Расчет парных коэффициентов корреляции можно выполнить по формуле (6.1, см. работу 6) или в табличном процессоре Microsoft Excel, воспользовавшись статистической функцией «КОРРЕЛ». Расчет погрешности  $\sigma$ , коэффициентов корреляции выполняется по формуле (6.10).

В результате для рядов железа общего и растворенного кислорода в р. Славянке получаем:

- коэффициент корреляции между рядом  $Y$  и  $X_1$ :  
 $r_{01} = 0,74; \sigma_{r_{01}} = 0,07;$
- коэффициент корреляции между рядом  $Y$  и  $X_2$ :  
 $r_{02} = 0,38; \sigma_{r_{02}} = 0,14;$
- коэффициент корреляции между рядами  $X_1$  и  $X_2$ :  
 $r_{12} = 0,27; \sigma_{r_{12}} = 0,15.$

Дальше производится проверка на наличие в наборе дублирующих и неэффективных аргументов.

Проверка на наличие дублирующих переменных проводится по формуле Алексеева (7.1). Переменные считаются дублирующими, если выполняется неравенство:

$$r_{12} > 1 - 2 \cdot \sigma_{r_{12}}. \quad (7.1)$$

Соотношение (7.1) не всегда достаточно оправдано, поэтому обычно рекомендуется считать переменные  $X$  дублирующими, если коэффициент корреляции между ними равен или больше 0,9.

Проверка показала, что критическое допустимое значение по формуле Алексеева равно 0,69. Полученное значение коэффициента корреляции  $r_{12}$  меньше этого значения, следовательно, переменные  $X_1$  и  $X_2$  не являются дублирующими, и обе могут войти в уравнение регрессии.

Далее исключаются неэффективные предикторы. В общем случае существует ряд критериев для выявления неэффективных предикторов. В данном случае мы принимаем, что предикторы являются неэффективными, если связь  $Y = f(X_1)$  и  $Y = f(X_2)$  незначима.

Оценка значимости коэффициентов корреляции выполняется по схеме, изложенной в работе 6. При этом необходимо учитывать величину коэффициентов корреляции. Таким образом, для коэффициента корреляции  $r_{01}$  необходимо использовать преобразование Фишера. Используя формулы (6.14), (6.16–6.17), получаем:

$$z_{01} = 0,95; \sigma_{z_{01}} = 1,17; t_{\alpha} \cdot \sigma_{z_{01}} = 0,28.$$

Так как  $|z_{01}| > t_{\alpha} \cdot \sigma_{z_{01}}$  ( $0,95 > 0,28$ ), то значение коэффициента корреляции входит в критическую область гипотезы о равенстве коэффициента корреляции нулю, следовательно, связь  $Y = f(X_1)$  значима.

Для  $r_{02}$  оценку проводят с использованием самого коэффициента корреляции. По формулам (6.10) и (6.12) получаем:

$$\sigma_{r_{02}} = 0,14; t_{\alpha} \cdot \sigma_{r_{02}} = 0,24.$$

Таким образом,  $|r_{02}| > t_{\alpha} \cdot \sigma_{r_{02}}$  ( $0,38 > 0,24$ ), следовательно, связь  $Y = f(X_2)$  значима.

В результате оказалось, что оба ряда  $X_1$  и  $X_2$  являются эффективными предикторами ряда  $Y$ .

## 2. Оценка сводного коэффициента корреляции.

Оценка общей связи исследуемого процесса с его предикторами проводится на основе сводного коэффициента корреляции  $-R_0$ :

$$R_0 = \sqrt{1 - \frac{D_0}{D_{00}}}, \quad (7.2)$$

где  $D_0$  – расширенный определитель матрицы парных коэффициентов корреляции;  $D_{00}$  – главный определитель матрицы парных коэффициентов корреляции.

При числе предикторов, равном 2, матрица парных коэффициентов корреляции имеет вид:

$$D = \begin{vmatrix} r_{00} & r_{01} & r_{02} \\ r_{01} & r_{11} & r_{12} \\ r_{02} & r_{12} & r_{22} \end{vmatrix}.$$

Или в нашем случае:

$$D = \begin{vmatrix} 1 & 0,74 & 0,38 \\ 0,74 & 1 & 0,27 \\ 0,38 & 0,27 & 1 \end{vmatrix}.$$

В таком случае  $D_0$  и  $D_{00}$  определяются по формулам:

$$D_0 = 1 - r_{12}^2 - r_{01}^2 - r_{02}^2 + 2 \cdot r_{12} \cdot r_{01} \cdot r_{02}, \quad (7.3)$$

$$D_{00} = 1 - r_{12}^2. \quad (7.4)$$

В результате расчетов по формулам (7.2–7.4) получаем:

$$D_0 = 0,38; D_{00} = 0,92; R_0 = 0,77.$$

Оценка средней квадратической погрешности сводного коэффициента корреляции проводится по формуле:

$$\sigma_{R_0} = \frac{(1 - R_0^2)}{\sqrt{n - m - 1}}, \quad (7.5)$$

где  $m$  – число предикторов.

Очевидно, что связь является значимой, так как коэффициент корреляции больше чем  $t_\alpha \cdot \sigma_{R_0}$  ( $0,77 > 0,12$ ).

Доверительный интервал сводного коэффициента корреляции по аналогии с парным коэффициентом, оценивается либо по

самому значению  $R_0$ , либо через преобразование Фишера. Схема такой оценки изложена в работе № 6.

В нашем случае построение доверительного интервала сводного коэффициента корреляции должно производиться на основе преобразования Фишера, так его величина больше, чем 0,4. В результате расчетов с использованием мастера функций Microsoft Excel имеем:

$$z = 1,01; \sigma_z = 0,17.$$

При двустороннем уровне значимости 10 % и  $t_\alpha = 1,68$  получаем:

$$0,72 < z < 1,30;$$

и затем, переводя нижнюю и верхнюю границу доверительного интервала  $z$  в значения нижней и верхней границы коэффициента корреляции (функция «ФИШЕРОБР»), получаем:

$$0,62 < R_0 < 0,86.$$

### 3. Расчет параметров уравнения регрессии.

Уравнение множественной линейной корреляции для  $Y=f(X_1; X_2)$  имеет вид:

$$y = a_1 \cdot x_1 + a_2 \cdot x_2 + \varepsilon, \quad (7.6)$$

где  $a_1$  и  $a_2$  – коэффициенты веса или влияния аргумента на исследуемый процесс,  $\varepsilon$  – случайная составляющая, независимая от  $X$ .

Коэффициенты веса в уравнении множественной линейной корреляции определяются по формулам:

$$a_1 = \frac{\sigma_y}{\sigma_{x_1}} \cdot \frac{D_{01}}{D_{00}}, \quad (7.7)$$

$$a_2 = \frac{\sigma_y}{\sigma_{x_2}} \cdot \frac{D_{02}}{D_{00}}, \quad (7.8)$$

где  $D_{01}$  и  $D_{02}$  – определители матрицы парных коэффициентов корреляции (см. выше), рассчитанные по формулам:

$$D_{01} = r_{01} - r_{02} \cdot r_{12}, \quad (7.9)$$

$$D_{02} = r_{02} - r_{01} \cdot r_{12}. \quad (7.10)$$

Случайная составляющая  $\varepsilon$  определяется по формуле (7.11):

$$\varepsilon = m_y - a_1 \cdot m_{x_1} - a_2 \cdot m_{x_2}. \quad (7.11)$$

где  $m_y$ ,  $m_{x_1}$  и  $m_{x_2}$  – математические ожидания (средние значения) рядов  $Y$ ,  $X_1$  и  $X_2$ , рассчитанные под таблицей 7.1.

Для рядов концентраций железа общего и растворенного кислорода в р. Славянка получаем:  $a_1 = 0,53$ ;  $a_2 = 0,02$ ;  $\varepsilon = -0,14$ . В итоге уравнение множественной линейной корреляции имеет вид:

$$y = 0,53 \cdot x_1 + 0,02 \cdot x_2 - 0,14.$$

Средняя квадратическая погрешность весовых коэффициентов регрессии рассчитывается по формулам:

$$\sigma_{a_1} = \frac{\sigma_y}{\sigma_{x_1}} \cdot \sqrt{\frac{(1 - R_0^2) \cdot \Delta_{01}}{(n - m - 1) \cdot D_{00}}}, \quad (7.12)$$

$$\sigma_{a_2} = \frac{\sigma_y}{\sigma_{x_2}} \cdot \sqrt{\frac{(1 - R_0^2) \cdot \Delta_{02}}{(n - m - 1) \cdot D_{00}}}, \quad (7.13)$$

где  $\Delta_{01}$  и  $\Delta_{02}$  – минор определителя  $D_{00}$ , соответствующий его элементу  $r_{11}$  и  $r_{22}$ . В нашем случае  $\Delta_{01}$  и  $\Delta_{02} = 1$ .

В практических приложениях доверительные интервалы коэффициентов веса обычно определяются по упрощенной формуле:

$$a - t_\alpha \cdot \sigma_a \leq a \leq a + t_\alpha \cdot \sigma_a. \quad (7.14)$$

Для рядов концентраций в р. Славянка получаем средние квадратические погрешности коэффициентов веса:

$$\sigma_{a_1} = 0,09; \sigma_{a_2} = 0,01;$$

и их доверительные интервалы:

$$0,38 \leq a_1 \leq 0,68;$$

$$0 \leq a_2 \leq 0,04.$$

#### 4. Расчет значений ряда $Y$ по уравнению регрессии.

По заданным значениям  $X_1$  и  $X_2$ , вошедшим в расчет параметров уравнения регрессии, найти значения  $Y - Y_p$ . Расчеты представлены в таблице 7.1.

Здесь в столбцах 2, 3 и 4 имеются исходные данные. В столбце 5 значения  $Y_p$  получены по уравнению (7.6) с использованием известных параметров  $a_1$ ,  $a_2$  и  $\varepsilon$ . Далее нам необходимо оценить адекватность полученного результата. Для этого в столбце 6 таблицы приведены отклонения полученных значений от фактических:  $|Y - Y_p|$ . Полученные отклонения сравниваем с критическим значением  $0,674 \cdot \sigma_y$ . Если  $|Y - Y_p| < 0,674 \cdot \sigma_y$ , то оценка считается удовлетворительной, если больше – то неудовлетворительной. Результат оценки каждого значения представлен в столбце 7 таблицы 7.1. Далее оценивается обеспеченность метода  $P = m/n \cdot 100$ , где  $m$  – число

удовлетворительных расчетов,  $n$  – общее число расчетов. Как и в методе контуров, метод считается удовлетворительным, если  $P$  больше или равно 80 %. Как видно из таблицы 7.1, для наших рядов метод не удовлетворительный, т. к.  $P = 25/37 \cdot 100 = 68 \%$ .

**Таблица 7.1. Расчет значений ряда  $Y$  по уравнению множественной линейной корреляции для рядов значений концентраций железа общего и растворенного кислорода в р. Славянка**

№	$Y$	$X_1$	$X_2$	$Y_p$	$ Y - Y_p $	Оцен-ка	Ранжирован-ные значения $Y_p$	$Y_p$ н.	$Y_p$ в.
1	0,18	0,17	8,09	0,12	0,058	уд	0,11	-0,018	0,234
2	0,17	0,22	7,40	0,13	0,036	уд	0,12	-0,005	0,247
3	0,2	0,18	8,29	0,13	0,069	уд	0,12	-0,004	0,247
4	0,29	0,28	8,11	0,18	0,109	неуд	0,13	0,001	0,253
5	0,54	0,53	9,50	0,34	0,196	неуд	0,13	0,003	0,254
6	0,24	0,54	9,60	0,35	0,111	неуд	0,13	0,005	0,257
7	0,24	0,44	8,34	0,27	0,031	уд	0,13	0,008	0,259
8	0,106	0,18	7,19	0,11	0,002	уд	0,14	0,015	0,266
9	0,17	0,3	7,84	0,19	0,016	уд	0,14	0,019	0,270
10	0,35	0,62	9,13	0,38	0,034	уд	0,15	0,021	0,272
11	0,45	0,62	8,80	0,38	0,073	уд	0,15	0,022	0,274
12	0,51	0,56	12,10	0,41	0,096	неуд	0,15	0,024	0,276
13	0,35	0,42	8,15	0,26	0,094	неуд	0,15	0,028	0,279
14	0,22	0,24	8,42	0,17	0,054	уд	0,16	0,039	0,290
15	0,24	0,41	7,95	0,25	0,007	уд	0,17	0,040	0,292
16	0,14	0,16	8,67	0,13	0,012	уд	0,18	0,054	0,306
17	0,12	0,13	9,08	0,12	0,001	уд	0,18	0,055	0,306
18	0,076	0,16	9,70	0,15	0,074	уд	0,19	0,060	0,311
19	0,22	0,25	6,96	0,14	0,080	неуд	0,19	0,061	0,312
20	0,115	0,25	8,09	0,16	0,049	уд	0,19	0,063	0,315
21	0,13	0,14	9,10	0,13	0,003	уд	0,20	0,076	0,327
22	0,21	0,36	6,48	0,19	0,021	уд	0,21	0,087	0,338
23	0,21	0,19	9,11	0,15	0,056	уд	0,25	0,121	0,373
24	0,11	0,16	9,60	0,15	0,038	уд	0,26	0,131	0,382
25	0,19	0,37	10,25	0,27	0,084	неуд	0,26	0,133	0,385
26	0,4	0,37	9,80	0,26	0,136	неуд	0,26	0,139	0,390
27	0,11	0,2	8,52	0,15	0,037	уд	0,27	0,140	0,392
28	0,117	0,22	7,90	0,14	0,027	уд	0,27	0,145	0,397
29	0,16	0,29	8,83	0,20	0,041	уд	0,27	0,148	0,400
30	0,21	0,3	9,10	0,21	0,002	уд	0,28	0,151	0,402
31	0,22	0,34	10,30	0,26	0,039	уд	0,33	0,209	0,460

№	$Y$	$X_1$	$X_2$	$Y_p$	$ Y - Y_p $	Оцен-ка	Ранжирован-ные значения $Y_p$	$Y_p$ н.	$Y_p$ в.
32	0,24	0,53	9,06	0,33	0,094	неуд	0,34	0,218	0,469
33	0,11	0,28	8,38	0,19	0,076	неуд	0,35	0,225	0,477
34	0,22	0,48	7,09	0,27	0,046	уд	0,38	0,251	0,503
35	0,02	0,28	8,08	0,18	0,160	неуд	0,38	0,251	0,503
36	0,23	0,40	9,60	0,28	0,046	уд	0,38	0,258	0,510
37	0,30	0,61	9,05	0,38	0,077	неуд	0,41	0,289	0,540

Также для оценки адекватности полученного результата необходимо построить график связи фактических и рассчитанных значений  $Y = f(Y_p)$ . Для этого значения  $Y_p$  откладываются по оси абсцисс, а значения  $Y$  – по оси ординат. Линию связи  $Y = f(Y_p)$  провести используя встроенные функции в мастере диаграмм Excel (добавить линейный тренд). Результат представлен на рис. 7.1. Из рисунка видно, что точки связи достаточно широко разбросаны относительно линии связи.

Расчеты по уравнениям регрессии всегда необходимо сопровождать оценкой погрешностей определения параметров и оценкой их устойчивости. Средняя квадратическая погрешность расчетов по

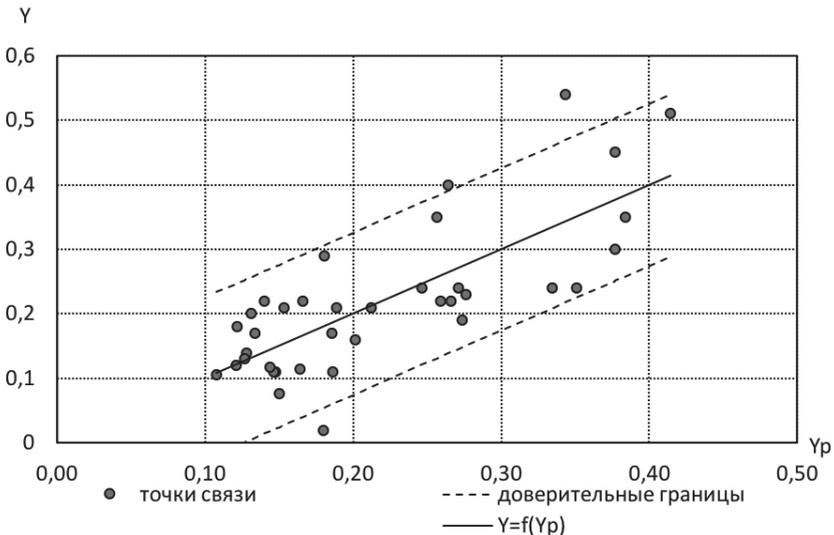


Рис.7.1. График связи фактических и рассчитанных значений ряда  $Y$

уравнению множественной линейной корреляции оценивается по формуле:

$$\sigma_{\Delta} \approx \sigma_y \cdot \sqrt{(1 - R_0^2)}. \quad (7.15)$$

Получаем  $\sigma_{\Delta} = 0,075$ .

Точно также по приближенной формуле рассчитывается доверительный интервал действительных значений  $Y$  при данных значениях  $X$ :

$$Y_p - t_{\alpha} \cdot \sigma_{\Delta} \leq Y_p \leq Y_p + t_{\alpha} \cdot \sigma_{\Delta}. \quad (7.16)$$

Расчет верхней и нижней границ уравнения  $Y = f(Y_p)$  при 10%-ном уровне значимости представлен в таблице 7.1 в столбцах 8–10. Для расчета сначала необходимо ранжировать значения  $Y_p$ , а затем по формуле (7.16) рассчитать отдельно верхнюю и нижнюю границы. Полученные результаты нанести на рис. 7.1. Как видно из рисунка, в критическую область связи выходит 3 точки, которые необходимо в дальнейшем проанализировать.

Интересно сопоставить оценки числовых характеристик (математического ожидания и среднеквадратического отклонения) исходного ряда и рассчитанного по уравнению регрессии ряда  $Y$ . Результаты оценки числовых характеристик представлены в таблице 7.2.

**Таблица 7.2. Оценка числовых характеристик исходного и рассчитанного ряда  $Y$**

	$Y$	$Y_{y.p.}$
$m_y$	0,22	0,22
$\sigma_y$	0,12	0,09

Из таблицы видно, что математические ожидания практически совпадают, а вот средние квадратические отклонения отличаются, причем весьма существенно.

### **6. Оценка стационарности связи.**

Оценка стационарности связи по сводному коэффициенту корреляции проводится по той же методике, что и для парного коэффициента. Методика оценки подробно изложена в пункте 6 работы № 6.

Для оценки стационарности разделим наши исходные ряды наблюдений  $Y$ ,  $X_1$  и  $X_2$  на две примерно равные части. Затем рассчитаем парные коэффициенты корреляции, определители  $D_0$  и  $D_{00}$  и сводный коэффициент корреляции. Метод оценки стационарности будет зависеть от величины  $R_0$ . В нашем случае необходимо

использовать преобразование Фишера (так как оба коэффициента корреляции  $> 0,4$ ). Результат оценки стационарности представлен в таблице 7.3.

**Таблица 7.3. Оценка стационарности связи для рядов значений концентраций железа общего и растворенного кислорода в р. Славянка**

	1 половина	2 половина
$n$	18	19
$r_{01}$	0,83	0,60
$r_{02}$	0,52	0,24
$r_{12}$	0,46	0,04
$D_0$	0,23	0,59
$D_{00}$	0,79	1,00
$R_0$	0,84	0,64
$z$	1,24	0,76
$\sigma_z$	0,26	0,25
$\sigma_{z-1-z2}$	0,36	
$t$	1,3	

Так как значение  $t = 1,3$  меньше, чем значение  $t_\alpha = 1,68$ , то гипотеза о стационарности связи не опровергается. Но значение  $t$  настолько велико, что желательно продолжить исследования в этом направлении.

### **7. Оценка точности расчетов по уравнению регрессии на независимом материале.**

Как отмечалось в предыдущем задании, одним из условий адекватности математической модели является соответствие между предсказаниями на ее основе и реальным поведением объекта. В данном случае наиболее удобно это соответствие проверить на так называемом независимом материале, то есть материале, который не был использован при идентификации модели.

Для этого по уравнению регрессии необходимо рассчитать значения  $Y$ , не вошедшие в период совместных наблюдений при оценке параметров уравнения множественной корреляции. В нашем случае это значения исходных рядов под номерами 1, 11, 21, 31 и 41, исключенные на первом этапе работы до начала расчетов.

В таблице 7.4 представлены результаты расчетов значений ряда  $Y$ . В этой таблице:  $Y$ ,  $X_1$  и  $X_2$  – значения исходных рядов;  $Y_p$  – значения  $Y$ , рассчитанные по формуле уравнению регрессии;  $|\hat{Y} - Y_p|$  – значения погрешностей расчетов на независимом

материале, определяемые как разность фактических и рассчитанных значений.

Оценка качества расчетов значений  $Y$  производилась по допустимой погрешности  $\Delta$ . В геоэкологических прогнозах в качестве допустимой погрешности принимается величина:

$$\Delta = 0,674 \cdot \sigma_y. \quad (7.17)$$

В данном случае по ряду  $Y$ :

$$\Delta = 0,076.$$

Если ошибка  $|Y - Y_p|$  не превышает  $\Delta$ , то расчет считается удовлетворительным (уд), если превышает – неудовлетворительным (неуд).

**Таблица 7.4. Оценка погрешностей расчетов по уравнению регрессии на независимом материале**

№	$X_1$	$Y$	$X_2$	$Y_p$	$ Y - Y_p $	$\Delta$
1	0,49	0,32	10,40	0,34	0,02	уд
11	0,25	0,16	6,12	0,12	0,04	уд
21	0,74	0,34	10,50	0,48	0,14	неуд
31	0,31	0,21	8,81	0,21	0,00	уд
41	0,49	0,24	10,60	0,35	0,11	неуд

Методика считается удовлетворительной, если число удовлетворительных расчетов (обеспеченность  $P$ ) составляет 80 или более процентов.

В данном случае  $P = 40\%$ , то есть для прогнозов, судя по проверке на независимом материале, эта методика (уравнение регрессии) неэффективна.

Необходимо обратить внимание на то, что испытание на независимом материале проведено на весьма ограниченных выборках, поэтому в данном случае результаты оценки носят приближенный характер. В действительности объем независимых исходных данных должен составлять не менее 10–20 членов.

### **8. Оценка адекватности модели рассматриваемой связи.**

Для суждения об адекватности математической модели связи будем ориентироваться в данном случае на следующие требования.

В геоэкологических прогнозах модель принимается адекватной если (см. работу 5):

$$\begin{aligned} \sigma_{\Delta}/\sigma_y &\leq 0,70 \text{ при } n \leq 15; \\ \sigma_{\Delta}/\sigma_y &\leq 0,75 \text{ при } 15 < n < 25; \\ \sigma_{\Delta}/\sigma_y &\leq 0,80 \text{ при } n \geq 25. \end{aligned}$$

В геоэкологических расчетах математическая модель множественной корреляции считается адекватной если выполняются следующие условия:

$$R_0 \geq 0,7; R_0/\sigma_{R_0} \geq 2; a_j/\sigma_{a_j} \geq 2; n \geq 10.$$

В данном случае отношение  $\sigma_{\Delta}/\sigma_y = 0,64$ , сводный коэффициент корреляции  $R_0 = 0,77$ , отношение  $R_0/\sigma_{R_0} = 10,8$ ;  $a_1/\sigma_{a_1} = 6,0$ ;  $a_2/\sigma_{a_2} = 1,7$ ;  $n = 37$ .

Таким образом, судя по результатам оценки на зависимом материале и с учетом оценок адекватности модели по методу контуров и расчетов по уравнению множественной линейной корреляции, а также учитывая полученные параметры, модель множественной линейной корреляции для рядов железа общего и растворенного кислорода не может использоваться в практических расчетах.

## **Лабораторная работа № 8.**

### **Проверка исходных рядов на однородность и случайность**

*Цель работы:* изучить методы проверки рядов наблюдений на соответствие требованиям случайности и однородности.

*Основные задачи:*

- 1) провести оценку стационарности рассматриваемого процесса по рядам наблюдений различными методами;
- 2) провести оценку даты нарушения однородности в исследуемых рядах наблюдений;
- 3) произвести проверку случайности рядов наблюдений по различным критериям;
- 4) выбрать математическую модель исходных рядов наблюдений по характеру внутрирядной связи.

*Исходные данные:*

- два временных ряда данных о состоянии объектов окружающей среды (ряды данных те же, что и в работе № 1);
- результаты расчетов оценок числовых характеристик методом моментов, представленные в первом задании.

*Содержание работы*

- Проверка стационарности рассматриваемых процессов по коэффициенту корреляции и по критерию Спирмена.

– Произвести оценку даты нарушения однородности, если она есть, с использованием графика интегральной кривой.

– Произвести проверку случайности рядов стока по критерию длин и числа серий по числу повышений и понижений и по числу экстремумов.

– Определить характер внутрирядных связей с использованием автокорреляционной функции.

#### *Теоретические положения*

Для выполнения работы студенты должны изучить теоретический материал по учебникам и учебным пособиям из списка литературы. После изучения теоретического материала студенты должны знать:

1) математические модели для описания изменения процесса во времени;

2) понятие стационарности ряда наблюдения;

3) способы оценки стационарности и выявления трендов в рядах наблюдений;

4) случайность, методы проверки рядов на случайность;

5) корреляционные функции стационарных процессов;

6) оценка корреляционных функций по выборочным данным.

Для описания изменений геоэкологических процессов во времени используются математические модели. Наиболее распространенными из них являются следующие.

1. Математическая модель исходного ряда в виде последовательности значений случайной величины. В геоэкологии, гидрометеорологии и в других областях науки эта последовательность значений часто называется случайным рядом, а сама случайная величина – чисто случайной величиной.

2. Математическая модель в виде простой цепи Маркова. В этом случае принимается, что каждое последующее значение ряда зависит только от предыдущего и не зависит от других предшествующих ему значений.

3. Математическая модель в виде последовательности значений сложной цепи Маркова. В этом случае принимается, что каждое последующее значение ряда зависит от  $m$  предшествующих, где  $m$  – радиус корреляции или период обратной связи.

Возможность использования той или иной математической модели, а следовательно и методы обработки временных рядов, определяются рядом признаков, среди которых главными являются однородность, стационарность и характер внутрирядных связей.

## План выполнения задания

### 1. Оценка стационарности ряда наблюдения по коэффициенту корреляции и критерию Спирмена.

В некоторых случаях нестационарность процесса может определяться наличием тренда.

Обычно **трендом** называют медленное, постепенное изменение случайной переменной  $X$  в течение анализируемого периода.

Для оценки наличия тренда обычно используется коэффициент корреляции рассматриваемого ряда  $X$  со временем и критерий Спирмена.

Предварительную оценку наличия или отсутствия тренда в рядах наблюдений можно выполнить на основе анализа графика значений концентраций. На рис. 8.1 представлен график изменения концентраций железа общего во времени. На рисунок добавлена линия тренда (линейного) с использованием встроенного мастера в Excel.

Как видно из рисунка, линия тренда идет на понижение. Однако о значимости тренда можно судить только после использования численных методов оценки.

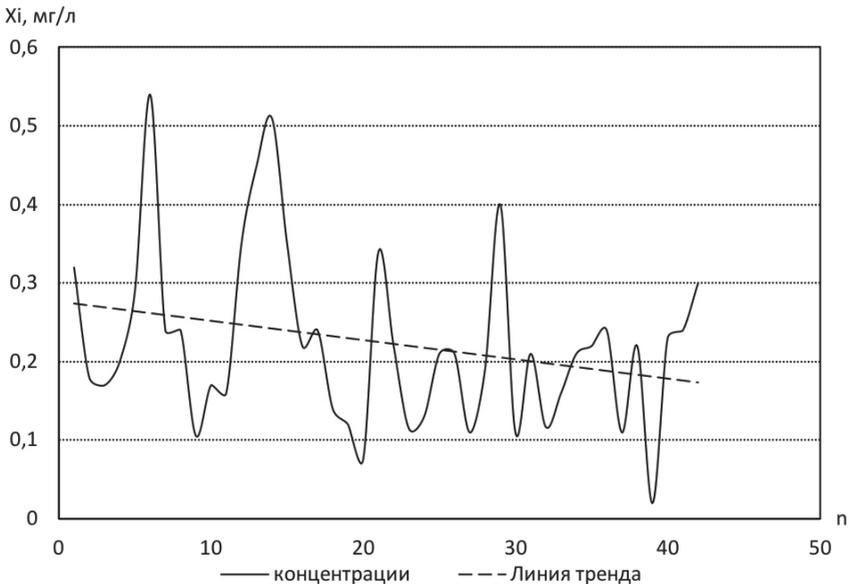


Рис. 8.1. Изменение концентраций железа общего во времени, р. Славянка, г. Пушкин, 2011–2016 гг.

### *1.1. Оценка тренда по коэффициенту корреляции.*

Оценка тренда по коэффициенту корреляции сводится к оценке коэффициента корреляции между значениями ряда концентраций и его порядковыми номерами. Для расчета коэффициента корреляции в программе Excel можно воспользоваться функцией «КОРРЕЛ» где в качестве первого аргумента принимаются значения ряда, а в качестве второго аргумента принимаются порядковые номера членов ряда (в нашем случае от 1 до 42). Далее необходимо провести оценку значимости полученного коэффициента корреляции по методике изложенной в работе 5. Если значение  $r$  окажется значимым, то очевидно наличие тренда в ряду наблюдений. Кроме того, знак коэффициента корреляции указывает направление тренда. Так если  $r < 0$ , то тренд идет на понижение, если  $r > 0$ , то тренд на увеличение.

Для исходного ряда получаем:

- коэффициент корреляции исходного ряда концентраций и порядковых номеров  $r = -0,27$ ;
- средняя квадратическая погрешность коэффициента корреляции  $\sigma_r = 0,14$ ;
- произведение  $t_\alpha \cdot \sigma_r = 0,24$ .

Так как  $|r| > t_\alpha \cdot \sigma_r$ , то тренд является значимым. Следовательно, ряд концентраций железа общего не стационарен. Кроме того, так как  $r < 0$ , то тренд идет на понижение, что соответствует выводам по рис. 8.1.

### *1.2. Оценка тренда по критерию Спирмена.*

Одним из наиболее распространенных методов анализа зависимости процесса от времени является ранговый критерий Спирмена.

Для расчета критерия  $\rho$  необходимо для каждого значения ряда определить его ранговый номер. Ранговый номер – это порядковый номер значения в ранжированном ряду. Для этого в таблице 8.1 в столбце 3 представлены ранжированные по возрастанию значения ряда концентраций железа общего. Затем каждому значению исходного (не ранжированного) ряда присваивается его ранг и заносится в столбец 4 таблицы 8.1. Здесь можно воспользоваться встроенной функцией Excel «РАНГ», которую необходимо применять с осторожностью, если в ряду есть повторяющиеся значения. Далее в столбце 5 рассчитывается значение  $d^2$  – это разница между порядковым номером и рангом каждого отдельного значения.

Ранговый критерий Спирмена рассчитывается по формуле (8.1):

$$\rho = 1 - \frac{6 \cdot \sum_{i=1}^n d_i^2}{(n^3 - n)}, \quad (8.1)$$

где  $d_i$  – разность между порядковым номером значения и его рангом.

Критическое значение критерия Спирмена определяется по формуле (8.2):

$$t_\alpha \cdot \sigma_\rho = t_\alpha \cdot \sqrt{\frac{1}{n-1}}. \quad (8.2)$$

Если  $|\rho| > t_\alpha \cdot \sigma_\rho$ , то тренд значим и ряд не стационарный. По аналогии с коэффициентом корреляции, знак  $\rho$  указывает направление тренда.

**Таблица 8.1. Расчет критерия Спирмена для ряда концентраций железа общего, р. Славянка, г. Пушкин, 2011–2016 гг.**

№ п/п	Исходный ряд $x_i$	Ранжированные значения $x_i$	Ранг	$d^2$
1	0,32	0,02	35	1156
2	0,18	0,076	16	196
3	0,17	0,106	14	121
4	0,2	0,11	18	196
5	0,29	0,11	33	784
6	0,54	0,11	42	1296
7	0,24	0,115	28	441
8	0,24	0,117	29	441
9	0,106	0,12	3	36
10	0,17	0,13	15	25
11	0,16	0,14	12	1
12	0,35	0,16	37	625
13	0,45	0,16	40	729
14	0,51	0,17	41	729
15	0,35	0,17	38	529
16	0,22	0,18	23	49
17	0,24	0,19	30	169
18	0,14	0,2	11	49
19	0,12	0,21	9	100
20	0,076	0,21	2	324
21	0,34	0,21	36	225

№ п/п	Исходный ряд $x_i$	Ранжированные значения $x_i$	Ранг	$d^2$
22	0,22	0,21	24	4
23	0,115	0,22	7	256
24	0,13	0,22	10	196
25	0,21	0,22	19	36
26	0,21	0,22	20	36
27	0,11	0,23	4	529
28	0,19	0,24	17	121
29	0,4	0,24	39	100
30	0,11	0,24	5	625
31	0,21	0,24	21	100
32	0,117	0,24	8	576
33	0,16	0,29	13	400
34	0,21	0,30	22	144
35	0,22	0,32	25	100
36	0,24	0,34	31	25
37	0,11	0,35	6	961
38	0,22	0,35	26	144
39	0,02	0,4	1	1444
40	0,23	0,45	27	169
41	0,24	0,51	32	81
42	0,30	0,54	34	64
Сумма				14332

Далее получаем по формуле (8.1)  $\rho = -0,16$ .

Критическое значение для критерия Спирмена  $t_\alpha \cdot \sigma_\rho = 1,68 \cdot 0,16 = 0,26$ .

Так как  $|\rho| < t_\alpha \cdot \sigma_\rho$ , то тренд не значим и ряд стационарный.

## 2. Определение наличия даты нарушения однородности.

Однородность – это принадлежность всех значений ряда к одной генеральной совокупности (см. работу 5). Ранее в работе 5 мы использовали параметрические и непараметрические критерии оценки однородности ряда. Однако часто перед исследователями стоит задача не просто оценить сам факт однородности или неоднородности, но и определить конкретную дату нарушения однородности, если она есть. Для определения даты нарушения однородности может использоваться график интегральной кривой.

Интегральная кривая представляет собой нарастающую сумму значений ряда:

$$U_i = U_{i-1} + x_{ip} \quad (i = 1, 2, 3, \dots, n), \quad (8.3)$$

где  $U_{i-1}$  – предыдущее значение интегральной кривой,  $x_i$  – текущее значение исходного ряда.

Расчет интегральной кривой для ряда концентраций железа общего представлен в таблице 8.2.

**Таблица 8.2. Расчет интегральной кривой для ряда концентраций железа общего, р. Славянка, г. Пушкин, 2011–2016 гг.**

№ п/п	Исходный ряд $x_i$	$U_i$
1	0,32	0,32
2	0,18	0,50
3	0,17	0,67
4	0,2	0,87
5	0,29	1,16
6	0,54	1,70
7	0,24	1,94
8	0,24	2,18
9	0,106	2,29
10	0,17	2,46
11	0,16	2,62
12	0,35	2,97
13	0,45	3,42
14	0,51	3,93
15	0,35	4,28
16	0,22	4,50
...	...	...
36	0,24	8,26
37	0,11	8,37
38	0,22	8,59
39	0,02	8,61
40	0,23	8,84
41	0,24	9,08
42	0,30	9,38

Графическое отображение интегральной кривой представлено на рис. 8.2.

При отсутствии резких скачков и изменений на графике (интегральная кривая практически укладывается в прямую линию) говорят, что дата нарушения однородности отсутствует. Если же на графике четко видны один или несколько скачков или явное изменение

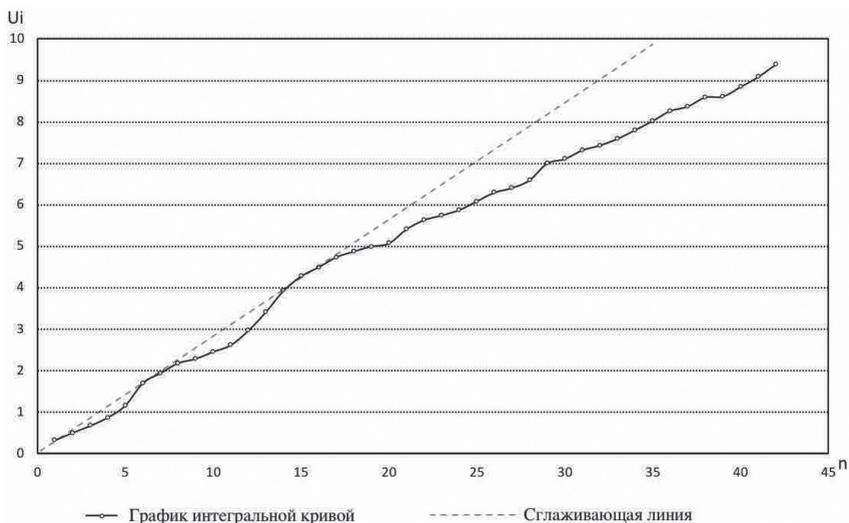


Рис. 8.2. График интегральной кривой для ряда концентраций железа общего, р. Славянка, г. Пушкин, 2011–2016 гг.

угла направления при  $n > 10$ , то эти точки и будут датами нарушения однородности.

В нашем случае скачки при значениях  $U_i$ , равных 1,16 и 1,7 (рис. 8.2), не считаются датами нарушения однородности, так как расположены в самом начале графика. Это связано с особенностью интегральной кривой при  $n < 10$ , когда сумма еще мала и каждое отдельное значение вносит существенный вклад в изменение направления кривой. Скачки при значениях  $U_i$ , равных 2,6 и 3,9, уже можно считать датами нарушения однородности. И действительно, если мы посмотрим на исходные ряды, то увидим что значения под номерами 13–15 достаточно велики, что и приводит к представленному на рисунке скачку между точками 2,6 и 3,9.

### 3. Проверка ряда наблюдений на случайность.

Существует большое количество тестов для проверки на «случайность». Ниже приводится тест, широко распространенный в практике как у нас, так и за рубежом, основанный на использовании следующих трех критериев: по длине и числу серий, по числу повышений и понижений ряда, по числу экстремумов. Проверка по всем трем критериям основана на нулевой гипотезе об отсутствии внутрядных связей:

$$H_0: r(\tau) = 0, \tau > 0.$$

Если рассматриваемый ряд является чисто случайным, то он обладает некоторыми свойствами, связанными главным образом с характером группировок (серий) членов ряда по тем или иным признакам. Если ряд с большой степенью достоверности обладает указанными свойствами, то он является случайным, если нет, то нулевая гипотеза опровергается, и с заданной степенью уверенности можно принять этот ряд неслучайным, т. е. внутрирядно связанным.

### 3.1. Проверка случайности по критерию длин и числа серий.

Проверка проводится путем сопоставления длин и числа серий анализируемой выборки с длиной и числом серий случайной величины.

**Серия** – последовательность элементов (членов ряда), одинаковых по какому-либо признаку, непосредственно перед которой и после которой стоят элементы другого вида.

В нашем случае в качестве такого критерия выступают значения медианы ( $Me$ ), полученные в первой работе. Для упрощения расчетов продолжительности и количества серий в каждом исходном ряду значения большие или равные  $Me$  отмечаются нулем, меньшие  $Me$  – единицей. В таблице 8.3 значения серий для ряда концентраций железа общего представлены 3 столбце. По полученной последовательности нулей и единиц подсчитывается количество серий значений 0 ( $R_0$ ) и 1 ( $R_1$ ) разной продолжительности  $k$  ( $k = 1, 2, 3, \dots$ ). При этом крайние, незамкнутые серии (серии в начале и конце ряда) необходимо отбросить, соответственно уменьшив число членов ряда. В таблице 8.3 крайние незамкнутые серии для наглядности выделены серым. Затем рассчитывается общее число серий  $R = R_0 + R_1$ , где  $R_0$  количество серий, обозначенных 0,  $R_1$  количество серий, обозначенных 1. Помимо количества серий необходимо определить максимальную длину серии  $k_{\max}$ .

Для ряда концентраций железа общего по таблице 8.3 получаем:

$$R_0 = 8; R_1 = 9; R = 17; k_{\max} = 6.$$

**Таблица 8.3. Оценка случайности ряда концентраций железа общего в р. Славянка, г. Пушкин, по различным критериям**

№ п/п	$C_i$	Серии	Повышения	Экстремумы
1	0,32	0		
2	0,18	1	–	
3	0,17	1	–	э
4	0,2	1	+	

Окончание табл. 8.3

№ п/п	$C_i$	Серии	Повышения	Экстремумы
5	0,29	0	+	
6	0,54	0	+	э
7	0,24	0	-	
8	0,24	0		
9	0,106	1	-	э
10	0,17	1	+	э
11	0,16	1	-	э
12	0,35	0	+	
13	0,45	0	-	
14	0,51	0	+	э
15	0,35	0	-	
16	0,22	0	-	э
17	0,24	0	+	э
18	0,14	1	-	
19	0,12	1	-	
20	0,076	1	-	э
21	0,34	0	+	э
22	0,22	0	-	
23	0,115	1	-	э
24	0,13	1	+	
25	0,21	0	+	э
26	0,21	0		
27	0,11	1	-	э
28	0,19	1	+	
29	0,4	0	+	э
30	0,11	1	-	э
31	0,21	0	+	э
32	0,117	1	-	э
33	0,16	1	+	
34	0,21	0	+	
35	0,22	0	+	
36	0,24	0	+	э
37	0,11	1	-	э
38	0,22	0	+	э
39	0,02	1	-	э
40	0,23	0	+	
41	0,24	0	+	
42	0,30	0	+	

Зная математическое ожидание и дисперсию числа серий различной продолжительности в рядах значений случайной величины, можно рассчитать доверительные границы гипотезы об отсутствии внутрирядных связей. Так в таблице 8.4 представлены граничные значения общего числа серий при двухсторонних уровнях значимости 5 и 10 %.

**Таблица 8.4. Критические границы общего числа серий  $R$  в случайных рядах различной длительности  $n$  при двухстороннем уровне значимости  $2\alpha$ , равном 5 и 10 %**

Границы	Число наблюдений $n$												
	10	20	30	40	50	60	80	100	120	140	160	180	200
Уровень значимости $2\alpha = 5\%$													
Верхняя	9	15	21	27	33	39	50	61	72	83	93	104	115
Нижняя	2	6	10	14	18	22	31	40	49	58	68	77	86
Уровень значимости $2\alpha = 10\%$													
Верхняя	8	15	20	26	32	37	48	59	70	81	91	102	113
Нижняя	3	6	11	15	19	24	33	42	51	60	70	79	88

Если по исследуемому ряду наблюдений общее число серий окажется выше верхней критической границы или ниже нижней критической границы при данном уровне значимости  $2\alpha$ , то нулевая гипотеза о случайности ряда опровергается и ряд неслучаен. То есть в данном ряду имеются внутрирядные связи при том или ином радиусе корреляции.

Для ряда концентраций железа общего находим доверительные границы при уровне значимости  $2\alpha = 10\%$  и объеме выборки  $n = 38$ . Объем выборки берется без крайних не замкнутых серий. По таблице 8.4 получаем критические границы общего числа серий:

$$R_{\text{н}} = 14; R_{\text{в}} = 25.$$

Таким образом по числу серий ряд концентраций железа общего можно считать случайным при уровне значимости 10 %, т. к. общее число серий  $R$  входит в доверительные границы ( $14 < 17 < 25$ ).

Граничное значение критерия наибольшей длины серии  $k_{\alpha}$  определяется по формуле:

$$k_{\alpha} = \frac{\lg \left[ -\frac{n}{\ln(1-\alpha)} \right]}{\lg(2)} - 1, \quad (8.4)$$

где  $\alpha$  – уровень значимости в долях единицы.

Если  $k_{\max} < k_{\alpha}$ , то гипотеза о случайности ряда не опровергается и ряд считается случайным по критерию длины серии.

Для ряда концентраций железа общего при уровне значимости  $2\alpha = 10\%$  и объеме выборки  $n = 38$  получаем:

$$k_{\alpha} = 7.$$

Следовательно, по критерию длины серии ряд концентраций железа общего можно считать случайным ( $6 < 7$ ) при уровне значимости  $10\%$ .

### 3.2. Проверка случайности по числу повышений и понижений.

Сущность проверки состоит в следующем. Пусть имеется выборка  $x_1, x_2, \dots, x_n$ . Переход  $x_i > x_{i-1}$  называется повышением и отмечается знаком (+), а переход  $x_i < x_{i-1}$  называется понижением и отмечается знаком (-).

По исходным рядам концентраций в таблице 8.3 подсчитывается число повышений и понижений (столбец 4). Если имеются равные соседние значения, то они принимаются за одно повышение или понижение (в нашем случае таких значений оказалось 2, в таблице 8.3 они выделены серым). Далее рассчитывается общее количество повышений  $n_+$  и общее количество понижений  $n_-$ . В исходных рядах концентраций железа общего оказалось 21 повышение ( $n_+ = 21$ ) и 18 понижений ( $n_- = 18$ ).

Общее число повышений (или понижений) ряда значений случайной величины распределено асимптотически нормально с математическим ожиданием:

$$m_+ = m_- = n/2, \quad (8.5)$$

и дисперсией:

$$D_+ = D_- = (n + 1) / 12. \quad (8.6)$$

Зная  $m$  и  $D$ , можно рассчитать по данному ряду нормированную величину числа повышений или понижений:

$$t_+ = \frac{(n_+ - m_+)}{\sqrt{D_+}}, \quad (8.7)$$

$$t_- = \frac{(n_- - m_-)}{\sqrt{D_-}}. \quad (8.8)$$

Для ряда концентраций железа общего в р. Славянка, учитывая  $n = 42$ , получаем:

$$m_+ = m_- = 42/2 = 21;$$

$$D_+ = D_- = (42 + 1) / 12 = 3,6;$$

$$t_+ = 0;$$

$$t_- = -1,58.$$

Дальше, учитывая асимптотически нормальный закон распределения числа повышений и понижений, необходимо сравнить рассчитанные значения  $t_+$  и  $t_-$  со значениями номированных ординат нормального закона распределения при данном уровне значимости  $2\alpha$ . Если рассчитанные значения будут выходить за пределы доверительных границ, то гипотеза о случайности ряда наблюдений опровергается, и считается, что исходный ряд не случаен и имеет устойчивую тенденцию к повышению или понижению.

Доверительные границы числа понижений и повышений для данной гипотезы при уровне значимости  $2\alpha = 10\%$ , учитывая нормальный закон распределения этих характеристик, определяются по приложению 1. В данном случае при двухстороннем уровне значимости  $2\alpha$ :

$$t_{\text{в}} = t_{5\%} = 1,64, t_{\text{н}} = t_{95\%} = -1,64.$$

Таким образом доверительный интервал гипотезы о случайности исходных рядов находится в пределах от 1,64 до -1,64. Полученные значения  $t$  находятся в пределах этого интервала, т. е. гипотеза о случайности ряда концентраций железа общего не опровергается.

### 3.3. Проверка случайности по числу экстремумов.

Экстремумом называется любой элемент, принадлежащий последовательности  $x_1, x_2, \dots, x_n$ , для которой выполняется одно из неравенств:

$$x_{i-1} < x_i > x_{i+1} \text{ или } x_{i-1} > x_i < x_{i+1},$$

т. е. значение ряда  $x_i$  либо больше двух соседних значений и тогда такое значение считается за максимум, либо меньше двух соседних и тогда его принимают за минимум. Если имеются равные соседние значения, то они принимаются за один экстремум.

Определение экстремумов для ряда концентраций железа общего представлено в таблице 8.3 в столбце 5. Далее рассчитывается общее количество выявленных экстремумов  $n_3$ . В нашем случае  $n_3 = 21$ .

Общее число экстремумов случайного ряда распределено асимптотически нормально с математическим ожиданием:

$$m_3 = \frac{2 \cdot n}{3} \tag{8.9}$$

и дисперсией:

$$D_3 = \frac{(17 \cdot n - 29)}{90}. \quad (8.10)$$

Для проверки гипотезы о случайности исходного ряда достаточно рассчитать по этому ряду нормированное число экстремумов  $t_3$  по формуле:

$$t_3 = \frac{(n_3 - m_3)}{\sqrt{D_3}}. \quad (8.11)$$

Для ряда концентраций железа общего по формулам (8.9–8.11) получаем:

$$\begin{aligned} m_3 &= 2 \cdot n/3 = 2 \cdot 42/3 = 27; \\ D_3 &= (17 \cdot n - 29)/90 = (17 \cdot 42 - 29) / 90 = 7,1; \\ t_3 &= -2,4. \end{aligned}$$

Необходимо сравнить полученное значение с нормированными ординатами нормального закона распределения при заданном уровне значимости. Если полученное значение  $t_3$  не попадает в доверительный интервал, то гипотеза о случайности исходного ряда опровергается.

Оценка нулевой гипотезы о случайности исходного ряда концентраций проводится при уровне значимости  $2\alpha = 10\%$ . В данном случае, как и выше, доверительные границы гипотезы по числу экстремумов определяются по приложению 1:

$$t_{\alpha} = t_{5\%} = 1,64, \quad t_{1-\alpha} = t_{95\%} = -1,64.$$

Для ряда концентраций железа общего в р. Славянка полученное значение  $t_3$  выходит за доверительные границы. Следовательно, гипотеза о случайности этого ряда по характеру цикличности опровергается.

Таким образом, гипотеза о случайности ряда концентраций железа общего не опровергается по двум из трех критериев: по числу и количеству серий и по наличию тренда развития описываемого процесса, опровергается по третьему критерию – характеру цикличности. По-видимому, характер цикличности ряда существенно отличается от случайного ряда, но какого либо тренда в развитии процесса не прослеживается.

#### **4. Расчет автокорреляционной функции ряда наблюдений.**

Если в результате проверки случайности ряда наблюдений выяснилось наличие внутрирядных связей (ряд оказался не случаен),

то для описания изменений данного процесса во времени используется корреляционная функция.

*Корреляционная функция представляет собой последовательность коэффициентов корреляции значений ряда при временном сдвиге  $\tau$ .*

В общем случае корреляционная функция показывает степень зависимости значения ряда от одного или нескольких предшествующих значений. Следовательно, корреляционная функция характеризует динамику развития процесса во времени.

На практике временной сдвиг  $\tau$  обычно равен дискретности исходного ряда наблюдений (год, месяц, сутки и т.д.). В нашем случае дискретность ряда около одного месяца (сдвиг  $\tau = 1$  – это сдвиг на 1 значение ряда,  $\tau = 2$  – это сдвиг на 2 значение ряда, и т.д.). Для рядов продолжительностью от 30 до 50 значений обычно достаточно рассчитать корреляционную функцию до  $\tau = 10$ .

Так как корреляционная функция – это последовательность парных коэффициентов корреляции, то её расчет сводится к определению значений  $r(\tau)$  при временном сдвиге  $\tau$  от 1 до 10. Коэффициент корреляции  $r(\tau)$  при  $\tau = 0$  – это корреляция значений ряда самих с собой, следовательно, всегда при  $\tau = 0$   $r(\tau) = 1$ .

Расчет последовательности коэффициентов корреляции  $r(\tau)$  производится по следующей схеме. Пусть имеется ряд  $x_1, x_2, \dots, x_n$ . При  $\tau = 1$  нам необходимо найти коэффициент корреляции между значениями ряда от 1 до  $n-1$  и от 2 до  $n$ . При  $\tau = 2$  нам необходимо найти коэффициент корреляции между значениями ряда от 1 до  $n-2$  и от 3 до  $n$ . Расчет коэффициента корреляции в данном случае целесообразно выполнять с помощью встроенной функции Excel «КОРРЕЛ».

Для ряда концентраций железа общего ( $n = 42$ ) расчет корреляционной функции представлен в таблице 8.5. Здесь в 1 столбце представлены значения  $\tau$  от 0 до 10. Во втором столбце рассчитывались коэффициенты корреляции  $r(\tau)$ . Расчет производился через функцию «КОРРЕЛ». При  $\tau = 1$  в качестве первого аргумента функции выбираются значения ряда от 1 до 41, а в качестве второго аргумента значения от 2 до 42. При  $\tau = 2$  в качестве первого аргумента функции выбираются значения ряда от 1 до 40, а в качестве второго аргумента значения от 3 до 42. И так далее. В итоге, при  $\tau = 10$  в качестве первого аргумента функции выбираются значения ряда от 1 до 32, а в качестве второго аргумента значения от 11 до 42.

Оценка достоверности корреляционной функции проводится с помощью доверительного интервала, построенного на основании нулевой гипотезы о равенстве действительных значений корреляционной функции нулю:

$$H_0: r(\tau) = 0, \text{ при } \tau \neq 0.$$

Расчет доверительных границ в предположении, что  $r(\tau) = 0$ , производится по формуле (8.12):

$$-t_\alpha \cdot \sigma_{r(\tau)} < r(\tau) < +t_\alpha \cdot \sigma_{r(\tau)}, \quad (8.12)$$

где  $t_\alpha$  – статистика Стьюдента при уровне значимости  $\alpha$  и числе степеней свободы  $n-\tau-1$ ;  $\sigma_{r(\tau)}$  – среднее квадратическое отклонение ординат корреляционной функции, определяемое по формуле:

$$\sigma_{r(\tau)} = \frac{1}{\sqrt{(n-\tau-1)}}. \quad (8.13)$$

Если при данном  $\tau$  значение  $r(\tau)$  выходит за доверительные границы и попадает в критическую область, то гипотеза  $H_0$  для данного  $r(\tau)$  опровергается и он считается значимым. Результаты расчета доверительных границ корреляционной функции для ряда концентраций железа общего представлены в столбцах 3–6 таблицы 8.5.

**Таблица 8.5. Расчет корреляционной функции для ряда концентраций железа общего в р. Славянка, г. Пушкин**

$\tau$	$r(\tau)$	$\sigma_{r(\tau)}$	$t_\alpha$	$r(\tau)_н$	$r(\tau)_в$
0	1	0,16	1,68	-0,26	0,26
1	0,31	0,16	1,68	-0,27	0,27
2	0,10	0,16	1,68	-0,27	0,27
3	-0,22	0,16	1,68	-0,27	0,27
4	-0,16	0,16	1,69	-0,28	0,28
5	-0,13	0,17	1,69	-0,28	0,28
6	-0,03	0,17	1,69	-0,29	0,29
7	0,38	0,17	1,69	-0,29	0,29
8	0,32	0,17	1,69	-0,29	0,29
9	0,12	0,18	1,70	-0,30	0,30
10	-0,22	0,18	1,70	-0,31	0,31

Для наглядности и удобства анализа построим график корреляционной функции и её доверительных границ. По оси абсцисс откладываем значения  $\tau$ , а по оси ординат – значения  $r(\tau)$ . График корреляционной функции для ряда концентраций железа общего представлен на рис. 8.3.

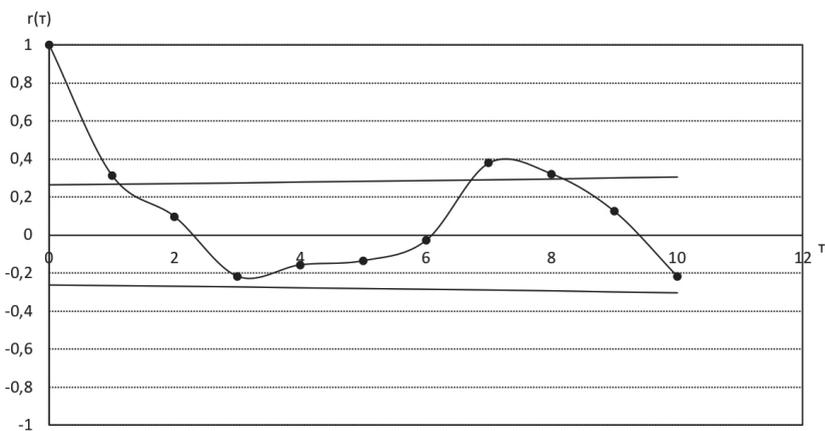


Рис. 8.3. Корреляционная функция ряда концентраций железа общего в р. Славянка, г. Пушкин

На основе анализа графика корреляционной функции можно предположить характер внутрирядных связей. Так, если при  $\tau > 0$  все значения  $r(\tau)$  лежат внутри доверительных границ, то можно предположить, что внутрирядные связи отсутствуют и для описания ряда подходит математическая модель случайной величины. Если при  $\tau > 0$  некоторые значения  $r(\tau)$  выходят за пределы доверительных границ, то можно предположить, что ряд не случаен. Так, если за доверительную границу выходит только значение  $r(\tau)$  при  $\tau = 1$ , а другие значения лежат внутри доверительных границ, то можно предположить, что каждое значение ряда зависит от одного предыдущего значения и для описания такого ряда можно использовать модель по схеме простой цепи Маркова. Если же за доверительные границы выходит одно или несколько значений  $r(\tau)$  при  $\tau > 1$ , то можно предположить, что каждое значение ряда зависит от нескольких предыдущих значений и для описания такого ряда можно использовать модель по схеме сложной цепи Маркова.

Важно отметить, что при  $\tau = 0$  значение  $r(\tau) = 1$  и всегда выходит за пределы доверительных границ.

На рис. 8.3 за пределы доверительных границ выходят значения  $r(\tau)$  при  $\tau = 1, 7$  и  $8$ . По-видимому, ряд не случаен и в какой-то степени отражает сезонные изменения концентраций.

Таким образом, ряд концентраций железа общего в р. Славянка имеет признаки не стационарности и неоднородности. По-видимому, для ряда также характерна цикличность, которая, учитывая дискретность в один месяц, может быть связана с сезонами года.

## Литература

1. *Алексеев Д.К., Гальцова В.В., Дмитриев В.В.* Экологический мониторинг: современное состояние, подходы и методы. – СПб.: РГГМУ, 2011. – 302 с.
2. *Алексеев Д.К., Зуева Н.В., Розенкова И.В., Урусова Е.С., Шелутко В.А.* Экологические и гидрометеорологические проблемы больших городов и промышленных зон: новые горизонты // *Метеорологический вестник.* – 2017. – Т. 9. № 2. – С. 1–8.
3. *Бендат Дж., Пирсол А.* Применение корреляционного и спектрального анализа. – М.: Мир, 1983. – 312 с.
4. *Гнеденко Б.В.* Курс теории вероятностей. – М.: Наука, 1969. – 400 с.
5. *Дмитриев В.В.* Оценка состояния и устойчивости наземных и водных геосистем: учебно-методическое пособие. – СПб.: ООО «Медиапапир», 2020 – 200 с.
6. *Евстигнеев В.М., Магрицкий Д.В.* Речной сток. Методические основы современной практики гидрологических расчетов. – М.: Триумф, 2016. – 224 с.
7. *Кайсл Ч.* Анализ временных рядов гидрологических данных. – Л.: Гидрометеоздат, 1972. – 138 с.
8. *Кучмент Л.С.* Речной сток (генезис, моделирование, предвычисление). – М.: Наука, 2008. – 394 с.
9. Определение основных расчетных гидрологических характеристик. СП 33-101-2003. – М.: Стройиздат, 2004. – 72 с.
10. *Примак Е.А., Зуева Н.В., Алексеев Д.К., Воякина Е.Ю.* Нормирование и снижение негативного воздействия на водные экосистемы: учебное пособие для высших учебных заведений. – СПб.: РГГМУ, 2020. – 116 с.
11. *Рожков В.А.* Теория и методы статистического оценивания вероятностных характеристик случайных величин и функций с гидрометеорологическими приложениями. – СПб.: Гидрометеоздат, 2002. – Кн. 1. – 340 с.; Кн. 2. – 440 с.
12. *Сикан А.В.* Методы статистической обработки гидрометеорологической информации. – СПб.: РГГМУ, 2007. – 279 с.
13. *Урусова Е.С.* Оценка загрязненности реки Охта в пределах Санкт-Петербурга на основе применения интегральных кривых // *Общество. Среда. Развитие.* – 2015. – № 4 (37). – С. 171–175.
14. *Урусова Е.С.* Вопросы применения методов статистической обработки данных гидрохимических наблюдений // *Метеорологический вестник.* – 2017. – Т. 9. № 2. – С. 216–220.
15. *Чичасов Г.Н.* Численные методы обработки и анализа гидрометеорологической информации. – М.: Боргес, 2013. – 235 с.
16. *Шелутко В.А.* Методы обработки и анализа геоэкологической информации. – СПб.: РГГМУ, 2020. – 296 с.
17. *Шелутко В.А.* Численные методы в гидрологии. – Л.: Гидрометеоздат, 1991. – 238 с.
18. *Шелутко В.А., Урусова Е.С.* Вопросы применения метода Монте-Карло для оценки погрешностей выборочных оценок числовых характеристик по временным рядам наблюдений // *Гидрометеорология и экология: научные достижения и перспективы развития.* Труды II Всероссийской конференции. – 2018. – С. 713–714.

# Приложения

## Приложение 1. Нормированные ординаты биномиальной кривой обеспеченности Пирсона III типа

$$t_p = (x_p - x) / \sigma_x = f(P, C_s)$$

C <sub>s</sub>	Обеспеченность, P %								
	0,01	0,1	1,0	3	5	10	20	25	30
-4,0	0,500	0,500	0,500	0,500	0,500	0,500	0,500	0,49	0,49
-3,8	0,527	0,527	0,526	0,526	0,526	0,526	0,520	0,52	0,51
-3,6	0,556	0,556	0,556	0,556	0,556	0,555	0,550	0,54	0,54
-3,4	0,588	0,588	0,588	0,588	0,587	0,586	0,580	0,57	0,55
-3,2	0,625	0,625	0,625	0,625	0,625	0,621	0,610	0,59	0,57
-3,0	0,667	0,667	0,666	0,666	0,665	0,661	0,640	0,62	0,59
-2,8	0,715	0,715	0,715	0,714	0,711	0,703	0,670	0,64	0,60
-2,6	0,770	0,770	0,770	0,766	0,764	0,746	0,700	0,66	0,61
-2,4	0,835	0,833	0,830	0,826	0,820	0,792	0,720	0,67	0,62
-2,2	0,914	0,910	0,905	0,895	0,882	0,842	0,750	0,69	0,64
-2,0	1,01	1,00	0,990	0,970	0,950	0,900	0,780	0,71	0,64
-1,8	1,11	1,11	1,09	1,06	1,02	0,940	0,800	0,72	0,64
-1,6	1,26	1,24	1,20	1,14	1,10	0,990	0,810	0,73	0,64
-1,4	1,41	1,39	1,32	1,23	1,17	1,04	0,830	0,73	0,64
-1,2	1,68	1,58	1,45	1,33	1,24	1,08	0,840	0,74	0,63
-1,0	1,92	1,79	1,59	1,42	1,32	1,13	0,850	0,73	0,62
-0,8	2,23	2,02	1,74	1,52	1,38	1,17	0,860	0,73	0,60
-0,6	2,57	2,27	1,88	1,61	1,45	1,20	0,850	0,72	0,59
-0,4	2,98	2,54	2,03	1,70	1,52	1,23	0,850	0,71	0,57
-0,2	3,37	2,81	2,18	1,79	1,58	1,26	0,850	0,69	0,55
0,0	3,72	3,09	2,33	1,88	1,64	1,28	0,840	0,67	0,52
0,2	4,16	3,38	2,47	1,96	1,70	1,30	0,83	0,65	0,50
0,4	4,61	3,66	2,61	2,04	1,75	1,32	0,82	0,63	0,47
0,6	5,05	3,96	2,75	2,12	1,80	1,33	0,80	0,61	0,44
0,8	5,50	4,24	2,89	2,18	1,84	1,34	0,78	0,58	0,41
1,0	5,96	4,53	3,02	2,25	1,88	1,34	0,76	0,55	0,38
1,2	6,41	4,81	3,15	2,31	1,92	1,34	0,73	0,52	0,35
1,4	6,87	5,09	3,27	2,37	1,95	1,34	0,71	0,49	0,31
1,6	7,31	5,37	3,39	2,42	1,97	1,33	0,68	0,46	0,28
1,8	7,76	5,64	3,50	2,46	1,99	1,32	0,64	0,42	0,24
2,0	8,21	5,91	3,60	2,51	2,00	1,30	0,61	0,39	0,20
2,2	8,63	6,14	3,68	2,54	2,02	1,27	0,57	0,35	0,16

Cs	Обеспеченность, P %								
	0,01	0,1	1,0	3	5	10	20	25	30
2,4	9,00	6,37	3,78	2,60	2,00	1,25	0,52	0,29	0,12
2,6	9,39	6,54	3,86	2,63	2,00	1,21	0,48	0,25	0,085
2,8	9,77	6,86	3,96	2,65	2,00	1,18	0,44	0,22	0,057
3,0	10,16	7,10	4,05	2,66	1,97	1,13	0,39	0,19	0,027
3,2	10,55	7,35	4,11	2,66	1,96	1,09	0,35	0,15	-0,006
3,4	10,9	7,54	4,18	2,66	1,94	1,06	0,31	0,11	-0,036
3,6	11,3	7,72	4,24	2,66	1,93	1,03	0,28	0,064	-0,072
3,8	11,67	7,97	4,29	2,65	1,90	1,00	0,24	0,032	-0,095
4,0	12,02	8,17	4,34	2,65	1,90	0,96	0,21	0,010	-0,120
4,2	12,40	8,38	4,39	2,64	1,88	0,93	0,19	-0,010	-0,13
4,4	12,76	8,60	4,42	2,63	1,86	0,91	0,15	-0,032	-0,15
4,6	13,12	8,79	4,46	2,62	1,84	0,87	0,13	-0,052	-0,17
4,8	13,51	8,96	4,50	2,60	1,81	0,82	0,10	-0,075	-0,19
5,0	13,87	9,12	4,54	2,60	1,78	0,78	0,068	-0,099	-0,20
5,2	14,25	9,27	4,59	2,60	1,74	0,73	0,035	-0,120	-0,21
5,4	14,60	9,42	4,62	2,60	1,70	0,67	0,02	-0,100	-0,21
5,6	14,95	9,59	4,65	2,60	1,67	0,62	0,0	-0,120	-0,21
5,8	15,32	9,70	4,70	2,60	1,64	0,57	-0,02	-0,140	-0,21
6,0	15,67	9,84	4,70	2,60	1,60	0,51	-0,05	-0,150	-0,21
6,2	16,04	9,95	4,71	2,60	1,56	0,47	-0,05	-0,150	-0,21
6,4	16,40	10,05	4,71	2,60	1,52	0,42	-0,05	-0,150	-0,21

Приложение 1 (продолжение)

Cs	Обеспеченность, P %										
	40	50	60	70	75	80	90	95	97	99	99,9
-4,0	0,46	0,41	0,31	-0,120	-0,010	-0,21	-0,96	-1,90	-2,65	-4,34	-8,17
-3,8	0,48	0,42	0,30	-0,095	-0,032	-0,24	-1,00	-1,90	-2,65	-4,29	-7,97
-3,6	0,49	0,42	0,28	-0,072	-0,064	-0,28	-1,03	-1,93	-2,66	-4,24	-7,72
-3,4	0,50	0,41	0,27	-0,036	-0,11	-0,31	-1,06	-1,94	-2,66	-4,18	-7,54
-3,2	0,51	0,41	0,25	-0,006	-0,15	-0,35	-1,09	-1,96	-2,66	-4,11	-7,35
-3,0	0,51	0,40	0,22	-0,027	-0,19	-0,39	-1,13	-1,97	-2,66	-4,05	-7,10
-2,8	0,51	0,39	0,20	-0,057	-0,22	-0,44	-1,18	-2,00	-2,65	-3,86	-6,86
-2,6	0,51	0,37	0,17	-0,085	-0,25	-0,48	-1,21	-2,00	-2,63	-3,86	-6,54
-2,4	0,51	0,35	0,17	-0,12	-0,29	-0,52	-1,25	-2,00	-2,60	-3,78	-6,37
-2,2	0,50	0,33	0,12	-0,16	-0,35	-0,57	-1,27	-2,02	-2,54	-3,68	-6,14
-2,0	0,49	0,31	0,09	-0,20	-0,39	-0,61	-1,30	-2,00	-2,51	-3,60	-5,91
-1,8	0,48	0,28	0,05	-0,24	-0,42	-0,64	-1,32	-1,99	-2,46	-3,50	-5,64
-1,6	0,46	0,25	0,02	-0,28	-0,46	-0,68	-1,33	-1,97	-2,42	-3,39	-5,37
-1,4	0,44	0,22	-0,20	-0,31	-0,49	-0,71	-1,34	-1,95	-2,37	-3,27	-5,09

Cs	Обеспеченность, P %										
	40	50	60	70	75	80	90	95	97	99	99,9
-1,2	0,42	0,19	-0,05	-0,35	-0,52	-0,73	-1,34	-1,92	-2,31	-3,15	-4,81
-1,0	0,39	0,16	-0,09	-0,38	-0,55	-0,76	-1,34	-1,88	-2,25	-3,02	-4,53
-0,8	0,37	0,13	-0,12	-0,41	-0,58	-0,79	-1,34	-1,84	-2,18	-2,89	-4,24
-0,6	0,34	0,10	-0,16	-0,44	-0,61	-0,80	-1,33	-1,80	-2,12	-2,75	-3,96
-0,4	0,31	0,07	-0,19	-0,47	-0,63	-0,82	-1,32	-1,75	-2,04	-2,61	-3,66
-0,2	0,28	0,03	-0,22	-0,50	-0,65	-0,83	-1,30	-1,70	-1,96	-2,47	-3,38
0,0	0,25	0,00	-0,25	-0,52	-0,67	-0,84	-1,28	-1,64	-1,88	-2,33	-3,09
0,2	0,22	-0,03	-0,28	-0,55	-0,69	-0,85	-1,26	-1,58	-1,79	-2,81	-2,81
0,4	0,19	-0,07	-0,31	-0,57	-0,71	-0,85	-1,23	-1,52	-1,70	-2,03	-2,54
0,6	0,16	-0,10	-0,34	-0,59	-0,72	-0,85	-1,20	-1,45	-1,61	-1,88	-2,27
0,8	0,12	-0,13	-0,37	-0,60	-0,73	-0,86	-1,17	-1,38	-1,52	-1,74	-2,02
1,0	0,09	-0,16	-0,39	-0,62	-0,73	-0,85	-1,13	-1,32	-1,42	-1,59	-1,79
1,2	0,05	-0,19	-0,42	-0,63	-0,74	-0,84	-1,08	-1,24	-1,33	-1,45	-1,58
1,4	0,02	-0,22	-0,44	-0,64	-0,73	-0,83	-1,04	-1,17	-1,23	-1,32	-1,39
1,6	-0,02	-0,25	-0,46	-0,64	-0,73	-0,81	-0,99	-1,10	-1,14	-1,20	-1,24
1,8	-0,05	-0,28	-0,48	-0,64	-0,72	-0,80	-0,94	-1,02	-1,06	-1,09	-1,11
2,0	-0,08	-0,31	-0,49	-0,64	-0,71	-0,78	-0,90	-0,95	-0,97	-0,99	-1,00
2,2	-0,12	-0,33	-0,50	-0,64	-0,69	-0,75	-0,842	-0,882	-0,895	-0,905	-0,910
2,4	-0,14	-0,35	-0,51	-0,62	-0,67	-0,72	-0,792	-0,820	-0,826	-0,830	-0,833
2,6	-0,17	-0,37	-0,51	-0,61	-0,66	-0,70	-0,746	-0,764	-0,766	-0,770	-0,770
2,8	-0,20	-0,39	-0,51	-0,60	-0,6	-0,67	-0,703	-0,711	-0,714	-0,715	-0,715
3,0	-0,22	-0,40	-0,51	-0,59	-0,62	-0,64	-0,661	-0,665	-0,666	0,666	-0,667
3,2	-0,25	-0,41	-0,51	-0,57	-0,59	-0,61	-0,621	0,625	0,625	0,625	-0,625
3,4	-0,27	-0,41	-0,50	-0,55	-0,57	-0,58	-0,586	-0,587	-0,588	-0,588	-0,588
3,6	-0,28	-0,42	-0,49	-0,54	-0,54	-0,55	-0,555	-0,556	-0,556	-0,556	-0,556
3,8	-0,30	-0,42	-0,48	-0,51	-0,52	-0,52	-0,526	-0,526	-0,526	0,526	-0,527
4,0	-0,31	-0,41	-0,46	-0,49	-0,49	-0,50	-0,500	-0,500	0,500	-0,500	-0,500
4,2	-0,31	-0,41	-0,45	-0,47	-0,473	-0,475	-0,476	-0,476	-0,476	-0,477	-0,477
4,4	-0,32	-0,40	-0,44	-0,451	-0,454	-0,455	-0,455	-0,455	-0,455	-0,455	-0,455
4,6	-0,32	-0,40	-0,42	-0,432	-0,454	-0,435	-0,435	-0,435	-0,435	-0,435	-0,435
4,8	-0,32	-0,39	-0,41	-0,416	-0,416	-0,416	-0,416	-0,416	-0,416	-0,417	-0,417
5,0	-0,33	-0,38	-0,40	-0,399	-0,400	-0,400	-0,400	-0,400	-0,400	-0,400	-0,400
5,2	-0,30	-0,37	-0,38	-0,384	-0,385	-0,385	-0,385	-0,385	-0,385	-0,385	-0,385
5,4	-0,30	-0,37	-0,37	-0,37	-0,37	-0,37	-0,37	-0,37	-0,37	-0,37	-0,37
5,6	-0,30	-0,36	-0,36	-0,36	-0,36	-0,36	-0,36	-0,36	-0,36	-0,36	-0,36
5,8	-0,30	-0,35	-0,35	-0,35	-0,35	-0,35	-0,35	-0,35	-0,35	-0,35	-0,35
6,0	-0,30	-0,34	-0,34	-0,34	-0,34	-0,34	-0,34	-0,34	-0,34	-0,34	-0,34
6,2	-0,30	-0,34	-0,34	-0,34	-0,34	-0,34	-0,34	-0,34	-0,34	-0,34	-0,34
6,4	-0,30	-0,33	-0,39	-0,33	-0,33	-0,33	-0,33	-0,33	-0,33	-0,33	-0,33

**Приложение 2. Ординаты кривых обеспеченности Пирсона  
III типа при  $C_s = 2C_v$  (в модульных коэффициентах)**

$C_v$	Обеспеченность, $P\%$													
	0,01	0,1	1	3	5	10	25	50	75	90	95	97	99	99,9
0,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
0,05	1,20	1,16	1,12	1,10	1,08	1,07	1,03	1,00	0,97	0,94	0,92	0,91	0,89	0,85
0,10	1,42	1,34	1,25	1,20	1,17	1,13	1,07	1,00	0,93	0,87	0,84	0,82	0,78	0,72
0,15	1,66	1,53	1,38	1,30	1,26	1,20	1,10	0,99	0,90	0,81	0,77	0,74	0,67	0,60
0,20	1,92	1,73	1,52	1,41	1,35	1,26	1,13	0,99	0,86	0,75	0,70	0,66	0,59	0,49
0,25	2,21	1,95	1,67	1,52	1,44	1,33	1,16	0,98	0,82	0,70	0,63	0,59	0,51	0,40
0,30	2,51	2,17	1,83	1,64	1,54	1,40	1,18	0,97	0,79	0,64	0,56	0,52	0,44	0,32
0,35	2,85	2,44	1,99	1,75	1,64	1,47	1,21	0,96	0,75	0,59	0,50	0,45	0,37	0,25
0,40	3,20	2,70	2,16	1,88	1,74	1,53	1,23	0,95	0,71	0,53	0,45	0,39	0,31	0,19
0,45	3,58	2,97	2,33	2,09	1,84	1,60	1,26	0,93	0,67	0,48	0,39	0,34	0,25	0,15
0,50	3,98	3,26	2,51	2,13	1,94	1,67	1,28	0,92	0,63	0,44	0,34	0,29	0,21	0,11
0,55	4,40	3,57	2,70	2,26	2,01	1,74	1,30	0,90	0,60	0,39	0,30	0,24	0,17	0,08
0,60	4,85	3,89	2,89	2,39	2,15	1,81	1,31	0,88	0,56	0,35	0,25	0,20	0,13	0,05
0,65	5,32	4,22	3,09	2,52	2,25	1,87	1,33	0,86	0,52	0,31	0,22	0,17	0,10	0,04
0,70	5,81	4,57	3,29	2,66	2,36	1,94	1,34	0,84	0,49	0,27	0,18	0,14	0,08	0,02
0,75	6,32	4,93	3,50	2,80	2,46	2,00	1,36	0,82	0,45	0,24	0,15	0,11	0,06	0,05
0,80	6,85	5,30	3,71	2,94	2,57	2,06	1,37	0,80	0,42	0,21	0,13	0,09	0,04	0,10
0,85	7,41	5,68	3,93	3,08	2,68	2,13	1,38	0,77	0,33	0,18	0,10	0,07	0,03	0,06
0,90	7,89	6,08	4,15	3,22	2,78	2,19	1,38	0,75	0,35	0,15	0,08	0,05	0,02	0,03
0,95	8,59	6,49	4,37	3,36	2,89	2,29	1,39	0,72	0,32	0,13	0,07	0,04	0,015	0,02
1,00	9,21	6,91	4,61	3,51	3,00	2,30	1,39	0,69	0,29	0,11	0,09	0,03	0,01	0,01

**Приложение 3. Ординаты интегральных кривых  
обеспеченности Крицкого–Менкеля (трехпараметриче-  
ского гамма-распределения) в модульных коэффициентах**

$$k_p = f(C_s/C_v, C_s/C_v, P)$$

$$C_s/C_v = 1,0$$

$P\%$	$C_v$										
	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	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	

P %	Cv									
	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
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

Приложение 3 (продолжение)

$Cs/Cv = 1,0$

P %	Cv									
	1,1	1,2	1,3	1,4	1,5	1,6	1,7	1,8	1,9	2,0
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
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

P %	Cv									
	1,1	1,2	1,3	1,4	1,5	1,6	1,7	1,8	1,9	2,0
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

Приложение 3 (продолжение)

$Cs/Cv = 3,0$

P %	Cv											
	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0	1,1	1,2
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	0,76	0,70	0,64	0,58	0,53	0,47	0,42	0,37	0,33

P %	Cv											
	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0	1,1	1,2
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

Приложение 3 (продолжение)

$$Cs/Cv = 3,0$$

P %	Cv								
	1,3	1,4	1,5	1,6	1,7	1,8	1,9	2,0	
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	

Приложение 3 (продолжение)

$$C_s/C_v = 4,0$$

P %	C <sub>v</sub>											
	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0	1,1	1,2
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	0,93	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

Приложение 3 (продолжение)

$$C_s/C_v = 4,0$$

P %	C <sub>v</sub>								
	1,3	1,4	1,5	1,6	1,7	1,8	1,9	2,0	
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	

P %	Cv							
	1,3	1,4	1,5	1,6	1,7	1,8	1,9	2,0
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
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

Приложение 3 (продолжение)

$Cs/Cv = 5,0$

P %	Cv											
	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0	1,1	1,2
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

P %	Cv											
	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0	1,1	1,2
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

Приложение 3 (продолжение)

$Cs/Cv = 6,0$

P %	Cv											
	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0	1,1	1,2
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

P %	Cv											
	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0	1,1	1,2
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. Таблицы для построения закона распределения Гамбела

**Таблица 1. Значение функции  $P_n = \exp(-e^y)$  для различных значений  $y$**

$y_p$	$P_n$								
9,10	0,01	4,00	1,80	1,50	20,0	0,09	60,0	-1,00	95
6,90	0,12	3,49	3,00	1,25	25,0	0,00	63,0	-1,10	97
6,00	0,24	3,00	4,90	1,03	30,0	-0,19	70,0	-1,53	99
5,30	0,50	2,97	5,00	1,00	35,0	-0,27	75,0	-1,93	99,9
5,00	0,67	2,25	10,0	0,67	40,0	-0,48	80,0	-2,00	99,99
4,60	1,00	2,00	13,0	0,37	50,0	-0,83	90,0		

**Таблица 2. Средние значения  $y_n$   
и средние квадратические отклонения  $\sigma_{yn}$  величин  $y_k$**

$n$	$y_n$	$\sigma_{yn}$	$n$	$y_n$	$\sigma_{yn}$	$n$	$y_n$	$\sigma_{yn}$
20	0,5236	1,0628	41	0,5442	1,1436	62	0,5527	1,1770
21	0,5252	1,0695	42	0,5448	1,1458	64	0,5533	1,1793
22	0,5268	1,0755	43	0,5453	1,1480	66	0,5538	1,1814
23	0,5282	1,0812	44	0,5458	1,1499	68	0,5543	1,1834
24	0,5296	1,0865	45	0,5463	1,1519	70	0,5548	1,1854
25	0,5306	1,0915	46	0,5468	1,1538	72	0,5552	1,1873
26	0,5320	1,0961	47	0,5473	1,1557	74	0,5557	1,1890
27	0,5332	1,1004	48	0,5477	1,1574	76	0,5561	1,1906
28	0,5343	1,1047	49	0,5481	1,1590	78	0,5565	1,1923
29	0,5353	1,1086	50	0,5485	1,1607	80	0,5569	1,1938
30	0,5362	1,1124	51	0,5489	1,1623	82	0,5572	1,1953
31	0,5371	1,1159	52	0,5493	1,1638	84	0,5576	1,1967

$n$	$y_n$	$\sigma_{yn}$	$n$	$y_n$	$\sigma_{yn}$	$n$	$y_n$	$\sigma_{yn}$
32	0,5380	1,1193	53	0,5497	1,1653	86	0,5580	1,1980
33	0,5388	1,1226	54	0,5501	1,1667	88	0,5583	1,1994
34	0,5396	1,1255	55	0,5504	1,1681	90	0,5586	1,2007
35	0,5403	1,1285	56	0,5508	1,1696	92	0,5589	1,2020
36	0,5410	1,1313	57	0,5511	1,1708	94	0,5592	1,2032
37	0,5418	1,1339	58	0,5515	1,1721	96	0,5595	1,2044
38	0,5424	1,1363	59	0,5518	1,1734	98	0,5598	1,2055
39	0,5430	1,1388	60	0,5521	1,1747	100	0,5600	1,2065
40	0,5436	1,1413						

**Приложение 5. Зависимость  $Cv = f(\lambda)$  для вычисления  $Cv$  методом приближенно наибольшего правдоподобия при фиксированных значениях отношения  $Cs$  к  $Cv$**

$Cv$	0	1	2	3	4	5	6	7	8	9
$Cs = Cv$										
0,10	0,0022	0,0026	0,0032	0,0038	0,0044	0,0050	0,0058	0,0066	0,0074	0,0082
0,20	0,0090	0,0100	0,0111	0,0122	0,0132	0,0143	0,0157	0,0171	0,0184	0,0199
0,30	0,0212	0,0229	0,0246	0,0263	0,0280	0,0297	0,0318	0,0339	0,036	0,0381
0,40	0,0402	0,0428	0,0454	0,0479	0,0505	0,0530	0,0562	0,0593	0,0624	0,0655
0,50	0,0686	0,0723	0,0760	0,0797	0,0835	0,0872	0,0916	0,0961	0,101	0,105
0,60	0,109	0,115	0,120	0,125	0,130	0,136	0,142	0,148	0,154	0,160
0,70	0,166	0,173	0,181	0,188	0,195	0,202	0,210	0,218	0,227	0,235
0,80	0,243	0,252	0,262	0,271	0,281	0,290	0,300	0,311	0,321	0,332
0,90	0,342	0,354	0,366	0,377	0,389	0,401	0,414	0,427	0,440	0,453
1,00	0,466	0,480	0,495	0,509	0,523	0,538	0,553	0,569	0,585	0,600
1,10	0,616	0,633	0,650	0,668	0,684	0,701	0,719	0,737	0,755	0,774
1,20	0,792	–	–	–	–	–	–	–	–	–
$Cs = 2Cv$										
0,10	0,0022	0,0026	0,0031	0,0037	0,0043	0,0049	0,0056	0,0063	0,0071	0,0079
0,20	0,0087	0,0096	0,0106	0,0116	0,0126	0,0137	0,0148	0,0160	0,0172	0,0185
0,30	0,0198	0,0212	0,0296	0,0241	0,0256	0,0271	0,0287	0,304	0,0321	0,0339
0,40	0,0357	0,0375	0,0394	0,0414	0,0434	0,0454	0,0476	0,0497	0,0519	0,0542
0,50	0,0565	0,0589	0,0613	0,0638	0,0664	0,0690	0,0716	0,0743	0,0771	0,0799
0,60	0,0828	0,0858	0,0887	0,0918	0,0949	0,0981	0,101,	0,105	0,108	0,111
0,70	0,115	0,118	0,122	0,126	0,129	0,133	0,137	0,141	0,145	0,149
0,80	0,153	0,157	0,162	0,166	0,170	0,175	0,180	0,184	0,180	0,194

Cv	0	1	2	3	4	5	6	7	8	9
0,90	0,198	0,203	0,208	0,213	0,218	0,224	0,229	0,234	0,240	0,245
1,00	0,251	0,256	0,262	0,268	0,274	0,280	0,286	0,292	0,298	0,304
1,10	0,311	0,317	0,323	0,330	0,337	0,343	0,350	0,357	0,364	0,371
1,20	0,378	0,386	0,393	0,400	0,408	0,415	0,423	0,431	0,438	0,446
1,30	0,454	0,462	0,171	0,479	0,487	0,496	0,504	0,513	0,521	0,530
1,40	0,539	0,548	0,557	0,566	0,575	0,584	0,594	0,603	0,613	0,623
1,50	0,632	0,642	0,652	0,662	0,672	0,682	0,692	0,703	0,713	0,724
1,60	0,734	0,745	0,756	0,767	0,778	0,789	0,800	0,811	0,823	0,834
1,70	0,864	0,857	0,860	0,881	0,893	0,905	0,917	0,929	0,911	0,953
1,80	0,966	0,978	0,991	0,004	1,016	1,029	1,042	1,055	1,069	1,082
1,90	1,095	1,109	1,122	1,136	1,150	1,163	1,177	1,191	1,205	1,220
$C_s = 3C_v$										
0,10	0,0022	0,0026	0,0031	0,0037	0,0043	0,0048	0,0055	0,0062	0,0070	0,0078
0,20	0,0085	0,0094	0,0104	0,0113	0,0122	0,0132	0,0143	0,0154	0,0165	0,0176
0,30	0,0188	0,0201	0,0214	0,0227	0,0240	0,0253	0,0267	0,0282	0,0297	0,0311
0,40	0,0326	0,0342	0,0359	0,0375	0,0391	0,0408	0,0425	0,0443	0,0461	0,0479
0,50	0,0497	0,0526	0,0535	0,0555	0,0574	0,0594	0,0614	0,0635	0,0656	0,0677
0,60	0,0698	0,0720	0,0742	0,0764	0,0786	0,0808	0,0832	0,0856	0,0879	0,0903
0,70	0,0926	0,0951	0,0976	0,100	0,103	0,105	0,108	0,110	0,113	0,116
0,80	0,118	0,121	0,124	0,127	0,129	0,132	0,135	0,138	0,141	0,144
0,90	0,147	0,150	0,153	0,156	0,159	0,162	0,165	0,168	0,171	0,174
1,00	0,177	0,181	0,184	0,187	0,191	0,194	0,197	0,201	0,204	0,208
1,10	0,211	0,215	0,218	0,222	0,225	0,229	0,232	0,236	0,240	0,243
1,20	0,247	0,251	0,255	0,258	0,262	0,266	0,270	0,274	0,277	0,281
1,30	0,286	0,289	0,294	0,298	0,302	0,306	0,311	0,315	0,319	0,323
1,40	0,327	0,332	0,336	0,340	0,345	0,349	0,354	0,358	0,363	0,367
1,50	0,372	0,376	0,381	0,386	0,390	0,395	0,400	0,404	0,409	0,414
1,60	0,419	0,424	0,428	0,433	0,438	0,443	0,448	0,453	0,458	0,463
1,70	0,468	0,474	0,479	0,484	0,489	0,495	0,500	0,505	0,511	0,516
1,80	0,521	0,527	0,532	0,538	0,534	0,549	0,555	0,560	0,566	0,572
1,90	0,578	0,584	0,589	0,595	0,600	0,606	0,612	0,618	0,624	0,630
$C_s = 4C_v$										
0,10	0,0022	0,0026	0,0031	0,0037	0,0043	0,0048	0,0055	0,0062	0,0070	0,0078
0,20	0,0083	0,0092	0,0101	0,0110	0,0119	0,0128	0,0138	0,0148	0,0159	0,0169
0,30	0,0180	0,0192	0,0203	0,0215	0,0227	0,0239	0,0252	0,0265	0,0279	0,0292
0,40	0,0305	0,0319	0,0334	0,0348	0,0363	0,0377	0,0392	0,0408	0,0423	0,0439
0,50	0,0454	0,0471	0,0488	0,0505	0,0522	0,0538	0,0556	0,0574	0,0592	0,0609
0,60	0,0627	0,0646	0,0665	0,0683	0,0702	0,0721	0,0741	0,0760	0,0780	0,0800
0,70	0,0820	0,0840	0,0861	0,0882	0,0903	0,0923	0,0946	0,0966	0,0989	0,101

Cv	0	1	2	3	4	5	6	7	8	9
0,80	0,103	0,105	0,108	0,110	0,112	0,114	0,117	0,119	0,121	0,124
0,90	0,126	0,128	0,131	0,133	0,136	0,138	0,141	0,143	0,146	0,148
1,00	0,151	0,153	0,156	0,158	0,161	0,163	0,166	0,169	0,171	0,174
1,10	0,177	0,179	0,182	0,185	0,188	0,190	0,193	0,196	0,199	0,201
1,20	0,204	0,207	0,210	0,213	0,216	0,219	0,222	0,224	0,227	0,230
1,30	0,233	0,236	0,239	0,242	0,245	0,248	0,251	0,254	0,258	0,261
1,40	0,264	0,267	0,270	0,273	0,276	0,279	0,283	0,286	0,289	0,292
1,50	0,296	0,299	0,302	0,305	0,309	0,312	0,315	0,319	0,322	0,325
1,60	0,329	0,332	0,336	0,339	0,342	0,346	0,349	0,353	0,356	0,360
1,70	0,363	0,367	0,370	0,374	0,377	0,381	0,384	0,388	0,392	0,395
1,80	0,399	0,403	0,406	0,410	0,414	0,417	0,421	0,425	0,429	0,432
1,90	0,436	0,440	0,444	0,447	0,451	0,455	0,459	0,463	0,467	0,471

**Приложение 6. Значения коэффициента асимметрии  $C_s$   
и коэффициента скошенности  $S$   
кривой обеспеченности Пирсона III типа**

$C_s$	$t_p = (x_p - x)/\sigma_x$				$S$
	$t_5$	$t_{50}$	$t_{95}$	$t_5 - t_{95}$	
1	2	3	4	5	6
0,0	1,64	0,00	-1,64	3,28	0,00
0,1	1,67	-0,02	-1,61	3,28	0,03
0,2	1,70	-0,03	-1,58	3,28	0,06
0,3	-1,72	-0,05	-1,55	3,27	0,08
0,4	1,75	-0,07	-1,52	3,27	0,11
0,5	1,77	-0,08	-1,49	3,26	0,11
0,6	1,80	-0,10	-1,45	3,25	0,17
0,7	1,82	-0,12	-1,42	3,24	0,20
0,8	1,84	-0,13	-1,38	3,22	0,22
0,9	1,86	-0,15	-1,35	3,21	0,25
1,0	1,88	-0,16	-1,32	3,20	0,28
1,1	1,89	-0,18	-1,28	3,17	0,31
1,2	1,92	-0,19	-1,24	3,16	0,34
1,3	1,94	-0,21	-1,20	3,14	0,37
1,4	1,95	-0,22	-1,17	3,12	0,39
1,5	1,96	-0,24	-1,13	3,09	0,42
1,6	1,47	-0,25	-1,10	3,07	0,45

Cs	$t_p = (x_p - \bar{x})/\sigma_x$				S
	$t_5$	$t_{50}$	$t_{95}$	$t_5 - t_{95}$	
1	2	3	4	5	6
1,7	1,98	-0,27	-1,06	3,04	0,48
1,8	1,99	-0,28	-1,02	3,01	0,51
1,9	2,00	-0,29	-0,98	2,98	0,54
2,0	2,00	-0,31	-0,95	2,95	0,57
2,1	2,01	-0,32	-0,91	2,92	0,59
2,2	2,02	-0,33	-0,88	2,89	0,63
2,3-	2,01	-0,34	-0,85	2,86	0,64
2,4	2,00	-0,35	-0,82	2,82	0,67
2,5	2,00	-0,36	-0,79	2,79	0,69
2,6	2,00	-0,37	-0,76	2,76	0,72
2,7	2,00	-0,38	-0,74	2,74	0,74
2,8	2,00	-0,39	-0,71	2,71	0,76
2,9	1,99	-0,39	-0,69	2,68	0,78
3,0	1,97	-0,40	-0,67	2,64	0,80
3,1	1,97	-0,40	-0,64	2,62	0,81
3,2	1,96	-0,41	-0,62	2,59	0,83
3,3	1,95	-0,41	-0,60	2,56	0,85
3,4	1,94	-0,41	-0,59	2,53	0,86
3,5	1,93	-0,41	-0,57	2,50	0,87
3,6	1,93	-0,42	-0,56	2,48	0,89
3,7	1,91	-0,42	-0,54	2,45	0,90
3,8	1,90	-0,42	-0,53	2,43	0,91
3,9	1,90	-0,41	-6,51	2,41	0,92
4,0	1,90	-0,41	-0,50	2,40	0,92
4,1	1,89	-0,11	-0,49	2,38	0,93
4,2	1,88	-0,41	-0,48	2,36	0,94
4,3	1,87	-0,40	-0,47	2,34	0,94
4,4	1,86	-0,40	-0,46	2,32	0,95
4,5	1,85	-0,40	-0,45	2,30	0,96
4,6	1,84	-0,40	-0,44	2,28	0,97
4,7	1,83	-0,40	-0,43	2,26	0,97
4,8	1,81	-0,39	-0,42	2,23	0,98
4,9	1,80	-0,39	-0,41	2,21	0,98
5,0	1,78	-0,38	-0,40	2,18	0,98
5,1	1,76	-0,38	-0,39	2,15	0,98
5,2	1,74	-0,37	-0,38	2,15	0,98

### Приложение 7. Значения функции распределения Колмогорова

$\lambda$	$k(\lambda)$	$\lambda$	$k(\lambda)$	$\lambda$	$k(\lambda)$	$\lambda$	$k(\lambda)$
0,30	0,00	1,22	0,898	1,42	0,964	1,62	0,989
0,40	0,003	1,24	0,910	1,44	0,968	1,64	0,991
0,50	0,036	1,26	0,916	1,46	0,972	1,66	0,992
0,60	0,136	1,28	0,925	1,48	0,975	1,68	0,993
0,70	0,289	1,30	0,932	1,50	0,978	1,70	0,994
0,80	0,456	1,32	0,939	1,52	0,980	1,72	0,995
0,90	0,607	1,34	0,945	1,54	0,983	1,74	0,995
1,00	0,730	1,36	0,951	1,56	0,985	1,76	0,996
1,10	0,822	1,38	0,956	1,58	0,986	1,78	0,996
1,20	0,888	1,40	0,960	1,60	0,988	1,80	0,997

### Приложение 8. Значения $\chi^2$ , соответствующие заданным вероятностям превышения $P(\chi^2 > \chi^2_q)$ и числам степеней свободы $\nu$

$\nu$	$P(\chi^2 > \chi^2_q)$										
	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

v	$P(\chi^2 > \chi^2_q)$										
	0,99	0,95	0,90	0,50	0,25	0,10	0,05	0,025	0,01	0,005	0,001
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 9. Значение статистики Стьюдента  $t$   
при разных уровнях значимости  $\alpha$   
и числах степеней свободы  $\nu$**

v	Уровни значимости для одностороннего критерия ( $\alpha$ )					
	0,2	0,1	0,05	0,025	0,01	0,005
	Уровни значимости для двустороннего критерия ( $2\alpha$ )					
	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

v	Уровни значимости для одностороннего критерия ( $\alpha$ )					
	0,2	0,1	0,05	0,025	0,01	0,005
	Уровни значимости для двустороннего критерия ( $2\alpha$ )					
	0,4	0,2	0,1	0,05	0,02	0,01
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
$\infty$	0,84	1,28	1,64	1,96	2,33	2,58

## Приложение 10. Критические значения распределения $F$

$$F(\alpha = 0,05)$$

v <sub>2</sub>	v <sub>1</sub>								
	3	4	5	6	8	12	24	120	$\infty$
3	9,277	9,117	9,015	8,941	8,845	8,745	8,639	8,549	8,527
4	6,591	6,388	6,256	6,163	6,041	5,912	5,774	2,65S	5,628
5	5,410	5,192	5,050	4,950	4,818	4,678	4,527	4,398	4,365
6	4,757	4,534	4,387	4,284	4,147	4,000	3,842	3,705	3,669
7	4,347	4,120	3,972	3,866	3,726	3,575	3,411	3,267	3,230
8	4,066	3,838	3,688	3,581	3,438	3,284	3,115	2,967	2,928
9	3,863	3,633	3,482	3,374	3,230	3,073	2,901	2,748	2,707
10	3,708	3,478	3,326	3,217	3,072	2,913	2,737	2,580	2,538
11	3,587	3,357	3,204	3,095	2,948	2,788	2,609	2,448	2,405
12	3,490	3,259	3,106	2,996	2,849	2,687	2,506	2,341	2,296
13	3,411	3,179	3,025	2,915	2,767	2,604	2,420	2,252	2,206
14	3,334	3,112	2,958	2,848	2,699	2,534	2,349	2,178	2,131
15	3,287	3,056	2,901	2,791	2,641	2,475	2,288	2,114	2,066
16	3,239	3,007	2,852	2,741	2,591	2,425	2,235	2,059	2,010
17	3,197	2,965	2,810	2,699	2,548	2,381	2,190	2,011	1,960
18	3,160	2,928	2,773	2,661	2,510	2,342	2,150	1,968	1,917
19	3,127	2,895	2,740	2,528	2,477	2,308	2,114	1,930	1,878
20	3,098	2,866	2,711	2,599	2,447	2,278	2,083	1,896	1,843
21	3,073	2,840	2,685	2,573	2,421	2,250	2,054	1,866	1,812

$v_2$	$v_1$								
	3	4	5	6	8	12	24	120	$\infty$
22	3,049	2,817	2,661	2,549	2,397	2,226	2,028	1,838	1,783
23	3,028	2,796	2,640	2,528	2,375	2,204	2,005	1,813	1,757
24	3,009	2,776	2,621	2,508	2,355	2,183	1,984	1,790	1,733
25	2,991	2,759	2,603	2,490	2,337	2,165	1,964	1,768	1,711
26	2,975	2,743	2,587	2,474	2,321	2,148	1,946	1,749	1,691
27	2,960	2,728	2,572	2,459	2,305	2,132	1,830	1,731	1,672
28	2,947	2,714	2,558	2,445	2,291	2,118	1,915	1,714	1,654
29	2,934	2,701	2,545	2,432	2,278	2,105	1,901	1,698	1,638
30	2,922	2,690	2,534	2,421	2,266	2,092	1,887	1,684	1,622
40	2,839	2,606	2,450	2,336	2,180	2,004	1,793	1,577	1,509
60	2,758	2,525	2,368	2,254	2,097	1,917	1,700	1,467	1,389
120	2,680	2,447	2,290	2,175	2,016	1,834	1,608	1,354	1,254
$\infty$	2,605	2,372	2,214	2,099	1,938	1,752	1,517	1,221	1,000

Приложение 10 (продолжение)

$F(\alpha = 0,10)$

$v_2$	$v_1$								
	3	4	5	6	8	12	24	120	$\infty$
3	5,391	5,343	5,309	5,285	5,252	5,216	5,176	5,143	5,134
4	4,191	4,107	4,051	4,010	3,955	3,896	3,831	3,775	3,761
5	3,620	3,520	3,453	3,405	3,339	3,268	3,191	3,123	3,105
6	3,289	3,181	3,108	3,055	2,983	2,905	2,818	2,742	2,722
7	3,074	2,961	2,883	2,827	2,752	2,668	2,575	2,493	2,471
8	2,924	2,806	2,727	2,668	2,589	2,502	2,404	2,316	2,293
9	2,813	2,693	2,611	2,551	2,469	2,379	2,277	2,184	2,159
10	2,728	2,605	2,522	2,461	2,377	2,284	2,178	2,082	2,055
11	2,660	2,536	2,451	2,389	2,304	2,209	2,100	2,000	1,972
12	2,606	2,480	2,391	2,331	2,245	2,147	2,036	1,932	1,904
13	2,560	2,434	2,347	2,283	2,195	2,097	1,983	1,876	1,846
14	2,522	2,395	2,307	2,243	2,154	2,054	1,938	1,828	1,797
15	2,490	2,361	2,273	2,208	2,119	2,017	1,899	1,787	1,755
16	2,462	2,333	2,244	2,178	2,088	1,985	1,866	1,751	1,718
17	2,437	2,308	2,218	2,152	2,061	1,958	1,836	1,719	1,686
18	2,416	2,286	2,196	2,130	2,038	1,933	1,810	1,691	1,657
19	2,397	2,266	2,176	2,109	2,017	1,912	1,787	1,666	1,631
20	2,380	2,249	2,158	2,091	1,999	1,892	1,767	1,643	1,607
21	2,365	2,233	2,142	2,075	1,982	1,875	1,748	1,623	1,586

$v_2$	$v_1$								
	3	4	5	6	8	12	24	120	$\infty$
22	2,351	2,219	2,128	2,061	1,967	1,859	1,731	1,604	1,567
23	2,331	2,207	2,115	2,047	1,953	1,845	1,716	1,587	1,549
24	2,327	2,195	2,103	2,035	1,941	1,832	1,702	1,572	1,533
25	2,317	2,184	2,092	2,024	1,929	1,820	1,689	1,557	1,518
26	2,308	2,175	2,082	2,014	1,919	1,809	1,677	1,544	1,504
27	2,299	2,166	2,073	2,005	1,909	1,799	1,666	1,531	1,491
28	2,291	2,157	2,065	1,996	1,900	1,790	1,656	1,520	1,478
29	2,283	2,149	2,057	1,988	1,892	1,781	1,647	1,509	1,467
30	2,276	2,142	2,019	1,980	1,884	1,773	1,638	1,499	1,456
40	2,226	2,091	1,997	1,927	1,829	1,715	1,774	1,425	1,377
60	2,177	2,04.1	1,946	1,875	1,775	1,657	1,511	1,348	1,292
120	2,130	2,992	1,896	1,824	1,722	1,601	1,447	1,265	1,193
$\infty$	2,084	2,945	1,847	1,774	1,670	1,546	1,383	1,169	1,000

# Содержание

Предисловие . . . . .	3
Лабораторная работа № 1. Предварительный анализ исходных данных . . . . .	4
Лабораторная работа № 2. Исследование законов распределения и расчет значений ряда заданной обеспеченности . . . . .	16
Лабораторная работа № 3. Статистические оценки числовых характеристик по имеющимся рядам наблюдений . . . . .	30
Лабораторная работа № 4. Применение критериев согласия для выбора закона распределения . . . . .	37
Лабораторная работа № 5. Оценка однородности исходных рядов наблюдений . . . . .	45
Лабораторная работа № 6. Исследование взаимосвязи двух переменных величин . . . . .	54
Лабораторная работа № 7. Исследование взаимосвязи ряда переменных величин . . . . .	69
Лабораторная работа № 8. Проверка исходных рядов на однородность и случайность . . . . .	81
Литература . . . . .	98
Приложения . . . . .	99
Приложение 1. Нормированные ординаты биномиальной кривой обеспеченности Пирсона III типа . . . . .	99
Приложение 2. Ординаты кривых обеспеченности Пирсона III типа при $C_s = 2C_v$ (в модульных коэффициентах) . . . . .	102
Приложение 3. Ординаты интегральных кривых обеспеченности Крицкого–Менкеля (трехпараметрического гамма-распределения) в модульных коэффициентах . . . . .	102
Приложение 4. Таблицы для построения закона распределения Гамбела . . . . .	109
Приложение 5. Зависимость $C_v = f(\lambda)$ для вычисления $C_v$ методом приближенно наибольшего правдоподобия при фиксированных значениях отношения $C_s$ к $C_v$ . . . . .	110
Приложение 6. Значения коэффициента асимметрии $C_s$ и коэффициента скошенности $S$ кривой обеспеченности Пирсона III типа . . . . .	112
Приложение 7. Значения функции распределения Колмогорова. . . . .	114
Приложение 8. Значения $\chi^2$ , соответствующие заданным вероятностям превышения $P(\chi^2 > \chi^2_q)$ и числам степеней свободы $\nu$ . . . . .	114
Приложение 9. Значение статистики Стьюдента $t$ при разных уровнях значимости $\alpha$ и числах степеней свободы $\nu$ . . . . .	115
Приложение 10. Критические значения распределения $F$ . . . . .	116

*Учебное издание*

**Шелутко Владислав Аркадьевич**, проф., д-р геогр. наук  
**Урусова Елена Сергеевна**, канд. геогр. наук

**ПРАКТИКУМ ПО ДИСЦИПЛИНЕ  
« АНАЛИЗ И МЕТОДЫ ОБРАБОТКИ  
ГЕОЭКОЛОГИЧЕСКОЙ ИНФОРМАЦИИ»**

*Начальник РИО А.В. Ляхтейнен  
Редактор Л.Ю. Кладова  
Верстка М.В. Ивановой*

---

Подписано в печать 30.12.2020. Формат 60×90 1/16. Гарнитура Times New Roman.  
Печать цифровая. Усл. печ. л. 7,5. Тираж 30 экз. Заказ № 1029.  
РГГМУ, 192007, Санкт-Петербург, Воронежская ул., 79.

---