

Министерство образования и науки Российской Федерации

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ
РОССИЙСКИЙ ГОСУДАРСТВЕННЫЙ ГИДРОМЕТЕОРОЛОГИЧЕСКИЙ
УНИВЕРСИТЕТ

А.В. Сикан, Н.Г. Малышева, И.О. Винокуров

СТАТИСТИЧЕСКИЕ МЕТОДЫ АНАЛИЗА ГИДРОМЕТЕОРОЛОГИЧЕСКОЙ ИНФОРМАЦИИ

Лабораторный практикум



Санкт-Петербург
2014

УДК 556.48

Рецензент: Лобанова А.Г., канд. техн. наук (ГГИ).

Сикан А.В., Малышева Н.Г., Винокуров И.О. Статистические методы анализа гидрометеорологической информации. Лабораторный практикум. — СПб.: РГГМУ, 2014. — 76 с.

ISBN 978-5-86813-410-4

Даются рекомендации по выполнению лабораторных работ, которые выполняются студентами III курса гидрологического факультета дневной формы обучения. Направление подготовки 280400 — прикладная гидрометеорология. Профиль подготовки — прикладная гидрология.

Sikan A.V., Malysheva N.G., Vinokurov I.O. Statistical methods of hydrometeorological information analysis. Laboratory practical work. — St.Petersburg, RSHU Publishers, 2014. — 76 pp.

The book considers methodological guidance for performance of tasks for students of III course of hydrological faculty.

*Одобрено научно-методической комиссией
гидрологического факультета РГГМУ*

ISBN 978-5-86813-410-4

© Сикан А.В., Малышева Н.Г., Винокуров И.О., 2014
© Российский государственный гидрометеорологический университет (РГГМУ), 2014

ПРЕДИСЛОВИЕ

Практикум составлен в соответствии с программой дисциплины «Статистические методы анализа гидрометеорологической информации».

Основной задачей дисциплины «Статистические методы анализа гидрометеорологической информации» является ознакомление студентов с методами обработки и анализа гидрометеорологической информации, используемых в российской и мировой гидрологической практике. В процессе изучения дисциплины студенты должны освоить теоретический материал и практические приемы статистической обработки данных гидрометеорологических наблюдений.

Изучив дисциплину, студенты должны уметь решать широкий круг инженерных задач, возникающих при проведении водохозяйственных и гидроэнергетических расчетов, при гидротехническом и дорожном проектировании, при составлении гидрологических прогнозов и т. д.

В Практикуме приводятся планы лабораторных работ, выполняемых студентами III курса гидрологического факультета РГГМУ, при изучении дисциплины «Статистические методы анализа гидрометеорологической информации». Все лабораторные работы содержат теоретическую и расчетную части. Перед выполнением расчетной части студенты должны проработать определенные разделы теории и ответить на контрольные вопросы. В конце Практикума приводится список литературы, необходимой для изучения настоящей дисциплины.

Каждый студент выполняет расчет по индивидуальному варианту, который согласовывается с преподавателем кафедры гидрологии суши.

К экзамену по дисциплине студенты допускаются только после получения зачета по всем контрольным работам.

В настоящем Практикуме при изложении планов работ в качестве примера использованы данные по среднегодовым расходам р. Воложба в створе д. Воложба.

ЛАБОРАТОРНАЯ РАБОТА № 1

Исследование числовых характеристик и эмпирических функций распределения ряда минимальных суточных зимних (летних) расходов воды реки _____ в створе _____

Требуется построить гистограмму эмпирических частот, интегральную функцию распределения и функцию обеспеченностей; рассчитать значения основных числовых характеристик исследуемого ряда и их погрешности.

ИСХОДНЫЕ ДАННЫЕ

В качестве исходных данных в работе используются минимальные суточные зимние или летние расходы воды за период 40 лет по одной реке.

Выбор исходных данных осуществляется каждым студентом в индивидуальном порядке по согласованию с преподавателем кафедры гидрологии суши.

Данные выписываются из справочников серии «Государственный водный кадастр. Многолетние данные о режиме и ресурсах поверхностных вод суши» (МДС). Не следует использовать данные по зарегулированным рекам и рекам, где имеет место межбассейновая переброска стока. Площадь водосбора реки в расчетном створе должна быть в пределах 500 - 50000 км². Ряд не должен содержать пропусков в наблюдениях более чем 3 - 4 года. Исходные данные оформляются в виде таблицы и прилагаются к работе (табл.1.1). Кроме того, в каждой работе должны быть представлены сведения о том, где находится река (региональная принадлежность) и ее основные гидрографические характеристики (табл.1.2).

Таблица 1.1

Среднегодовые расходы воды р. Воложба – д. Воложба

№ п/п	Год	Расход, м ³ /с	№ п/п	Год	Расход, м ³ /с
1	1936	9,61	28	1963	9,70
2	1937	<u>6,89</u>	29	1964	10,1
3	1938	8,66	30	1965	10,4
4	1939	7,37	31	1966	17,2
5	1940	8,48	32	1967	12,1
6	1941	8,16	33	1968	14,3
7	1942	12,4	34	1969	15,6
8	1943	11,8	35	1970	9,10
9	1944	6,94	36	1971	9,37
10	1945	10,7	37	1972	7,58
11	1946	11,8	38	1973	7,27
12	1947	8,87	39	1974	10,4
13	1948	10,4	40	1975	8,62
14	1949	10,8	41	1976	12,5
15	1950	11,2	42	1977	13,0
16	1951	9,74	43	1978	13,5
17	1952	16,1	44	1979	8,99
18	1953	<u>18,1</u>	45	1980	9,30
19	1954	11,0	46	1981	12,4
20	1955	15,9	47	1982	14,8
21	1956	13,1	48	1983	14,2
22	1957	17,4	49	1984	14,5
23	1958	14,4	50	1985	11,2
24	1959	11,0	51	1986	13,6
25	1960	7,86	52	1987	12,8
26	1961	11,5	53	1988	11,6
27	1962	16,3			

Таблица 1.2

Региональная принадлежность и основные гидрографические характеристики р. Воложба – д. Воложба

Региональная принадлежность	ОГХ или МДС	Длина ряда n	Площадь водосбора F , км ²	Длина реки L , км	Уклон реки I , ‰	Озерность $f_{оз}$, %
Северо-Запад РФ	ОГХ т.2	53	1330	108	1,68	1

ТЕОРЕТИЧЕСКАЯ ЧАСТЬ

Для выполнения работы студенты должны изучить теоретический материал по учебному пособию [1]. Необходимо проработать следующие разделы: Введение, 1.1-1.2; 1.4.1-1.4.4; 1.4.6; 3 (вводная часть к разделу); 3.1(первый метод); 3.2.1; 3.3.

После изучения теоретического материала студенты должны знать ответы на следующие вопросы:

1. Что называется случайной величиной и как эта модель используется в практике гидрологических расчетов?
2. Приведите определения интегральной функции распределения, функции обеспеченностей и функции плотности вероятности.
3. Как связаны между собой интегральная функция распределения и функция обеспеченностей? Вероятность чего характеризует каждая из этих функций?
4. Перечислите основные числовые характеристики случайной величины и дайте их определения.
5. Что такое модульный коэффициент?
6. От чего зависит погрешность расчета числовых характеристик?

РАСЧЕТНАЯ ЧАСТЬ

1.1. ПОСТРОЕНИЕ ГИСТОГРАММЫ ЭМПИРИЧЕСКИХ ЧАСТОТ И ЭМПИРИЧЕСКИХ ФУНКЦИЙ РАСПРЕДЕЛЕНИЯ

1.1.1. Определение размаха R

Размах – это разность между наибольшим и наименьшим значением статистического ряда:

$$R = Q_{\max} - Q_{\min} . \quad (1.1)$$

В рассматриваемом примере максимальное значение расхода воды $18,1 \text{ м}^3/\text{с}$, минимальное значение $6,89 \text{ м}^3/\text{с}$, следовательно, $R = 18,1 - 6,89 = 11,21$.

1.1.2. Определение длины расчетного интервала L

Для построения гистограммы эмпирических частот размах статистического ряда разбивается на m равных интервалов. Коли-

чество интервалов зависит от длины ряда n и приближенно определяется по эмпирической формуле Стерджесса:

$$m = 1 + 3,322 \lg(n). \quad (1.2)$$

Полученное значение m округляется до целых значений. В примере длина ряда $n = 53$, следовательно, $m = 1 + 3,322 \lg(53) = 6,55 \approx 7$.

Длина расчетного интервала определяется по формуле

$$L = R/m. \quad (1.3)$$

Полученное значение L рекомендуется округлить, но так чтобы длина расчетного интервала изменилась не более чем на 15 %. Для р. Воложба – д. Воложба получаем:

$$L = \frac{11,21}{7} = 1,601 \approx 1,6.$$

1.1.3. Построение гистограммы эмпирических частот

Для построения гистограммы эмпирических частот заполняется табл. 1.3. В первый столбец таблицы записываются границы расчетных интервалов. При этом минимальный расход должен попасть в первый расчетный интервал.

Таблица 1.3

**Расчет эмпирических частот для ряда среднегодовых расходов
р. Воложба – д. Воложба**

Интервал значений расходов воды, м³/с	Число значений в интервале	Частота, p	
		в долях единицы	в процентах
6,8 - 8,4	7	0,132	13,2
8,4 - 10,0	11	0,208	20,8
10,0 - 11,6	12	0,226	22,6
11,6 - 13,2	9	0,170	17,0
13,2 - 14,8	7	0,132	13,2
14,8 - 16,4	4	0,075	7,5
16,4 - 18,0	2	0,038	3,8
18,0 - 19,6	1	0,019	1,9
Сумма	53	1	100

В нашем примере $L = 1,6$, $Q_{\min} = 6,89$, следовательно, можно рассматривать интервалы (6,8-8,4), (8,4-10,0), (10,0-11,6) и т. д.

Следует отметить, что за счет округления длины интервала число интервалов m может измениться на единицу в большую или меньшую сторону, что не имеет принципиального значения.

Во второй столбец таблицы записывается количество расходов z_i , попадающих в конкретный интервал. Если значение расхода попадает точно на границу интервала, будем относить его к правой границе интервала. При этом общее число значений во всех интервалах должно равняться длине ряда n .

В третий столбец таблицы записывается относительное число значений в интервале

$$p_i = z_i / n \quad (1.4)$$

Величину p_i называют эмпирической частотой. Эмпирическую частоту можно выражать либо в долях единицы, либо в процентах (последний столбец табл. 1.3):

$$p_{i,\%} = \frac{z_i}{n} 100\% . \quad (1.5)$$

По данным столбцов 1 и 3 табл. 1.3 строится гистограмма эмпирических частот (рис.1.1).

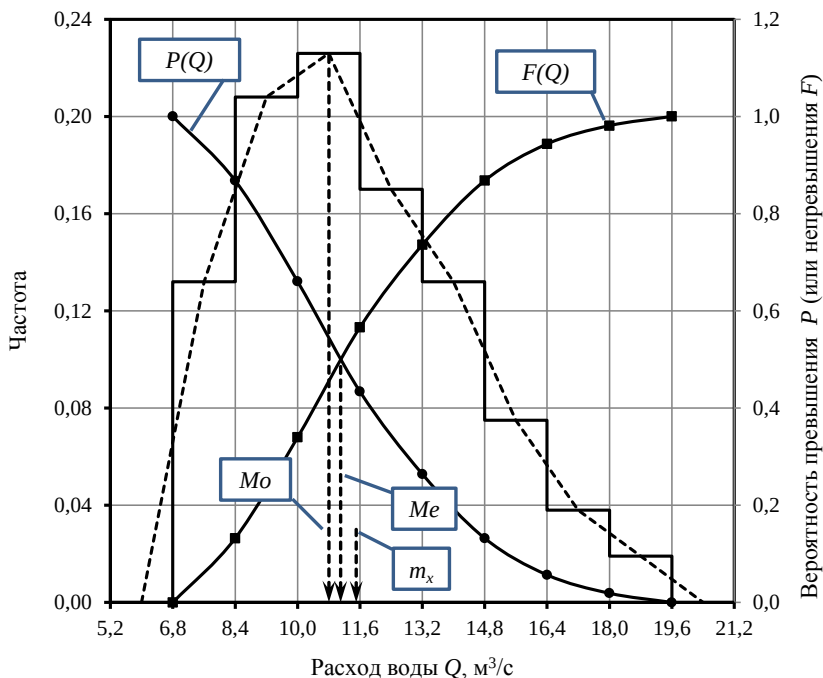


Рис.1.1. Гистограмма эмпирических частот и графики эмпирических функций распределения среднегодовых расходов воды, р. Воложба – д. Воложба. $F(Q)$ – интегральная функция распределения; $P(Q)$ – функция обеспеченностей;

1.1.4. Построение интегральной функции распределения и функции обеспеченностей

Для расчета интегральной функции распределения и функции обеспеченностей заполняется табл. 1.4. В первый столбец таблицы выписываются расходы воды, соответствующие границам всех использованных интервалов.

Во второй столбец записывается число случаев непревышения $z_{\text{нп}}$ для соответствующего расхода. Число случаев непревышения можно получить, последовательно суммируя значения из столбца 2 табл. 1.3.

В третий столбец записывается число случаев превышения $z_{\text{пр}}$ для соответствующего расхода. Получить эти значения можно, вычитая число случаев превышения из общей длины ряда. Так, в рассмотренном примере для расхода $6,8 \text{ м}^3/\text{с}$ число случаев непревышения равно 0, следовательно, число случаев превышения равно $53 - 0 = 53$; для расхода $8,4 \text{ м}^3/\text{с}$ число случаев непревышения равно 7, следовательно, число случаев превышения равно $53 - 7 = 46$ и т. д.

Таблица 1.4

Расчет координат интегральной функции распределения и функции обеспеченностей

Значение расхода воды, $\text{м}^3/\text{с}$	Число случаев непревышения	Число случаев превышения	Относительное число случаев непревышения, $F(Q)$	Относительное число случаев превышения, $P(Q)$
6,8	0	53	0	1
8,4	7	46	0,132	0,868
10	18	35	0,340	0,660
11,6	30	23	0,566	0,434
13,2	39	14	0,736	0,264
14,8	46	7	0,868	0,132
16,4	50	3	0,943	0,057
18	52	1	0,981	0,019
19,6	53	0	1	0

В четвертый столбец таблицы записывается относительное число случаев непревышения:

$$F(Q) = z_{\text{нп}}/n. \quad (1.6)$$

Полученная таким образом функция $F(Q)$ представляет собой эмпирическую интегральную функцию распределения вероятностей.

В пятый столбец записывается относительное число случаев превышения:

$$P(Q) = z_{\text{пр}}/n. \quad (1.7)$$

Функцию $P(Q)$ называют эмпирической функцией обеспеченностей. Поскольку функции $F(Q)$ и $P(Q)$ связаны соотношением

$$P(Q) = 1 - F(Q), \quad (1.8)$$

то вместо формулы (1.7) можно использовать формулу (1.8).

По данным столбцов 1 и 4 табл. 1.4 строится график интегральной функции распределения, по данным столбцов 1 и 5 – график функции обеспеченностей (см. рис.1.1).

Обратите внимание на то, что при построении графиков функций $F(Q)$ и $P(Q)$ на рис.1.1 использовалась правая шкала.

1.2. РАСЧЕТ ЧИСЛОВЫХ ХАРАКТЕРИСТИК И ИХ ПОГРЕШНОСТЕЙ

По выборке ограниченного объема нельзя точно рассчитать числовые характеристики, но можно получить приблизительные значения этих характеристик. Значения числовых характеристик, полученные по выборке, называются оценками числовых характеристик.

1.2.1. Расчет характеристик положения (моды, медианы, математического ожидания)

Модой непрерывной случайной величины называется такое ее значение, которому соответствует максимум плотности вероятностей. Для определения моды воспользуемся графиком плотности вероятностей представленном на рис. 1.1. В рассматриваемом примере $M_0 = 10,8 \text{ м}^3/\text{с}$.

Медианой непрерывной случайной величины называется такое ее значение, для которого вероятность превышения равна вероятности непревышения и равна 0,5.

Медиану также можно определить по графику представленному на рис.1.1. На графике медиана соответствует точке, где пересекаются интегральная функция распределения и функция обеспеченностей. В примере $M_e = 11,1$ (см. рис.1.1).

Оценкой *математического ожидания* является среднее арифметическое значение:

$$m_Q \approx \bar{Q} = \frac{\sum_{i=1}^n Q_i}{n}. \quad (1.9)$$

Для р. Воложба – д. Воложба $\bar{Q} = 11,5$.

Полученные значения моды, медианы и математического ожидания следует нанести на график (см. рис.1.1).

1.2.2. Расчет характеристик рассеивания

Требуется рассчитать дисперсию D , среднеквадратическое отклонение σ , коэффициент вариации C_v и коэффициент асимметрии C_s . Расчет произвести методом моментов.

Для выполнения расчетов заполняется вспомогательная табл. 1.5. Переменная k в табл. 1.5 – модульный коэффициент, который определяется по формуле

$$k_i = \frac{Q_i}{Q}. \quad (1.10)$$

Таблица 1.5

Вспомогательная таблица для расчета основных статистических характеристик ряда среднегодовых расходов воды р. Воложба – д. Воложба.

№ п/п	$Q, \text{ м}^3/\text{с}$	k	$(k - 1)$	$(k - 1)^2$	$(k - 1)^3$
1	9,61	0,8341	-0,1659	0,0275	-0,0046
2	6,89	0,5980	-0,4020	0,1616	-0,0649
3	8,66	0,7517	-0,2483	0,0617	-0,0153
4	7,37	0,6397	-0,3603	0,1298	-0,0468
5	8,48	0,7361	-0,2639	0,0697	-0,0184
6	8,16	0,7083	-0,2917	0,0851	-0,0248
7	12,4	1,0763	0,0763	0,0058	0,0004
...	...	...	...	...	...
48	14,2	1,2325	0,2325	0,0541	0,0126
49	14,5	1,2586	0,2586	0,0669	0,0173
50	11,2	0,9721	-0,0279	0,0008	0,0000
51	13,6	1,1805	0,1805	0,0326	0,0059
52	12,8	1,1110	0,1110	0,0123	0,0014
53	11,6	1,0069	0,0069	0,0000	0,0000
Сумма	610,61	53,0	0,00	3,320	0,339
Среднее	11,52	1,00	0,00		

После заполнения и обработки табл. 1.5 рассчитываются требуемые статистические характеристики.

$$C_v = \sqrt{\frac{\sum_{i=1}^n (k_i - 1)^2}{n-1}}; \quad (1.11) \quad C_s = \frac{n \sum_{i=1}^n (k_i - 1)^3}{(n-1)(n-2)C_v^3}; \quad (1.12)$$

$$\sigma = C_v \bar{Q}; \quad (1.13) \quad D = \sigma^2. \quad (1.14)$$

Для р. Воложба – д. Воложба получаем:

$$C_v = \sqrt{\frac{3,320}{53-1}} = 0,25; \quad C_s = \frac{53 \cdot 0,339}{(53-1)(53-2)(0,253)^3} = 0,42;$$

$$\sigma = 0,253 \cdot 11,52 = 2,91; \quad D = (2,91)^2 = 8,47.$$

1.2.3. Сводная таблица основных статистических характеристик гидрологического ряда

При составлении сводной табл. (табл. 1.6) все значения числовых характеристик исследуемого ряда округляются до 3 значащих цифр, причем не более двух после десятичной запятой.

Таблица 1.6

Основные статистические характеристики ряда среднегодовых расходов воды р. Воложба – д. Воложба

Mo	Me	$\bar{Q}$	C_v	C_s	C_s / C_v	σ	D
10,8	11,1	11,5	0,25	0,42	1,68	2,91	8,47

1.2.4. Расчет абсолютных и относительных погрешностей среднего значения, коэффициента вариации и коэффициента асимметрии

Абсолютные погрешности рассчитываются по формулам:

- для среднего значения

$$\varepsilon_Q = \frac{\sigma_Q}{\sqrt{n}}, \quad (1.15)$$

- для коэффициента вариации

$$\varepsilon_{C_v} = \frac{C_v}{n + 4C_v^2} \sqrt{\frac{n(1 + C_v^2)}{2}} \quad (1.16)$$

- для коэффициента асимметрии

$$\varepsilon_{C_s} = \sqrt{\frac{6}{n} (1 + 6C_v^2 + 5C_v^4)} . \quad (1.17)$$

Относительные погрешности рассчитываются по формулам:

- для среднего значения

$$\varepsilon_{Q, \%} = \frac{C_v}{\sqrt{n}} 100 \% , \quad (1.18)$$

- для коэффициента вариации

$$\varepsilon_{C_v} = \frac{1}{n + 4C_v^2} \sqrt{\frac{n(1 + C_v^2)}{2}} 100\% \quad (1.19)$$

- для коэффициента асимметрии

$$\varepsilon_{C_s, \%} = \frac{1}{C_s} \sqrt{\frac{6}{n} (1 + 6C_v^2 + 5C_v^4)} 100\% . \quad (1.20)$$

Результаты расчета представить в виде табл. (табл.1.7).

Таблица 1.7

Оценка точности расчета числовых характеристик ряда среднегодовых расходов воды, р. Воложба – д. Воложба

Числовая характеристика	Значение	Абсолютная погрешность	Относительная погрешность, %
Среднее значение, $\bar{Q}$	11,5	0,39	3,4
Коэффициент вариации, C_v	0,25	0,025	10,0
Коэффициент асимметрии, C_s	0,42	0,40	95

ВЫВОДЫ

В выводах следует:

1. Охарактеризовать полученную эмпирическую функцию плотности вероятности (одномодальная или многомодальная; степень ее асимметричности; расположение моды, медианы и математического ожидания).
2. Отметить, какова надежность полученных статистических характеристик (среднего значения, коэффициента вариации, коэффициента асимметрии). От чего зависит погрешность расчета указанных характеристик?

ЛАБОРАТОРНАЯ РАБОТА № 2

Исследование законов распределения и расчет расходов воды различной обеспеченности

Требуется построить на клетчатке вероятности эмпирическую кривую обеспеченностей и аналитические кривые нормального закона распределения, логнормального, Гумбеля, Пирсона III типа, Крицкого – Менкеля.

ИСХОДНЫЕ ДАННЫЕ

1. В качестве исходных данных используются минимальные суточные зимние или летние расходы воды за период 40 лет по одной реке (тот же ряд, что и в работе № 1).
2. Числовые характеристики, рассчитанные в работе № 1 (среднее значение, среднеквадратическое отклонение, коэффициент вариации, коэффициент асимметрии, отношение коэффициента асимметрии к коэффициенту вариации).

ТЕОРЕТИЧЕСКАЯ ЧАСТЬ

Для выполнения работы студенты должны изучить теоретический материал по учебнику [1]. Необходимо проработать следующие разделы: 1.5; 2.1; 2.3-2.6; 2.8-2.9, 3.1(второй метод).

После изучения теоретического материала студенты должны знать ответы на следующие вопросы:

1. Что такое нормированная случайная величина? Как изменятся среднее значение и дисперсия, если провести нормировку ряда?
2. Что такое модульный коэффициент?
3. Что такое клетчатка вероятностей? Что представляет собой клетчатка вероятностей с умеренной асимметричностью?
4. Как рассчитать и построить эмпирическую кривую обеспеченностей?
5. Для чего в практике гидрологических расчетов используются аналитические кривые обеспеченностей?
6. Какие параметры распределения необходимо знать для построения каждой из рассмотренных аналитических кривых обеспеченностей?
7. Как различаются по структуре таблицы ординат аналитических кривых обеспеченностей Пирсона III типа и Крицкого – Менкеля?
8. Дайте краткую характеристику каждой из рассмотренных аналитических кривых обеспеченностей (количество параметров, ОДЗ, наличие асимметрии, достоинства и недостатки).

РАСЧЕТНАЯ ЧАСТЬ

2.1. ПОСТРОЕНИЕ ЭМПИРИЧЕСКОЙ КРИВОЙ ОБЕСПЕЧЕННОСТЕЙ

В практике гидрологических расчетов для построения эмпирической кривой обеспеченностей используется методика, отличная от той, которая была рассмотрена в работе № 1.

Исходный ряд расходов воды ранжируется – расходы располагаются в порядке убывания (табл.2.1). Затем для каждого значения ранжированного ряда рассчитывается эмпирическая обеспеченность по формуле:

$$P = \frac{m}{n+1} 100\%, \quad (2.1)$$

где m – порядковый номер расхода в ранжированном ряду; n – длина ряда.

Таблица 2.1

Расчет ординат эмпирической кривой обеспеченностей среднегодовых расходов р. Воложба – д. Воложба, $\bar{Q} = 11,5 \text{ м}^3/\text{с}$

№ п/п	Расходы воды, $Q, \text{ м}^3/\text{с}$	Ранжированные расходы воды, Q_R $\text{м}^3/\text{с}$	$k = \frac{Q_R}{\bar{Q}}$	P, %
1	9,61	18,1	1,57	1,85
2	6,89	17,4	1,51	3,70
3	8,66	17,2	1,49	5,56
4	7,37	16,3	1,41	7,41
5	8,48	16,1	1,40	9,26
6	8,16	15,9	1,38	11,1
...	...	...	...	...
47	14,8	8,16	0,71	87,0
48	14,2	7,86	0,68	88,9
49	14,5	7,58	0,66	90,7
50	11,2	7,37	0,64	92,6
51	13,6	7,27	0,63	94,4
52	12,8	6,94	0,60	96,3
53	11,6	6,89	0,60	98,1

В табл. 2.1 рассчитывается также столбец модульных коэффициентов (столбец 4). Строить эмпирические и аналитические кривые обеспеченностей можно как в расходах, так и в модульных коэффициентах. В данной работе требуется построить кривые обеспеченностей в модульных коэффициентах.

Кривые обеспеченностей строятся на специальной бумаге – клетчатке вероятностей. Существует несколько видов клетчатки вероятностей. В данной работе используется наиболее распространенная в России клетчатка – с умеренной асимметричностью. Эта клетчатка спрямляет закон нормального распределения. Поскольку для гидрологических рядов характерна небольшая положительная асимметрия, то на этой клетчатке они будут иметь выпуклость, направленную вниз (рис.2.1).

При построении эмпирической кривой обеспеченностей необходимо учитывать следующее:

1. Вертикальный масштаб следует выбирать таким образом, чтобы максимальное значение модульного коэффициента (в примере $k_{\max} = 1,57$, табл. 2.1) не попадало на верхний край клет-

- чатки вероятностей; это необходимо, чтобы оставалась возможность строить аналитические кривые в области малых обеспеченностей.
2. Ноль отсчета для модульных коэффициентов должен совпадать с нижним краем клетчатки вероятностей, так как расходы воды могут быть только положительными.
 3. Шкала обеспеченностей на клетчатке вероятностей имеет неравномерный масштаб и при нанесении точек следует учитывать цену делений в каждом интервале.

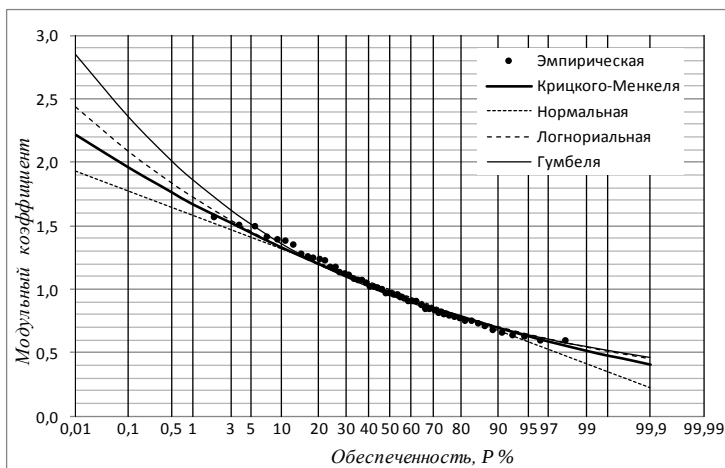


Рис.2.1. Эмпирическая и аналитические кривые обеспеченностей среднегодовых расходов воды в модульных коэффициентах р. Воложба – д. Воложба ($Q_{\text{ср}} = 11,5$; $C_v = 0,25$; $C_s/C_v = 2$).

Примечание: в рассматриваемом примере $C_s/C_v = 2$, поэтому кривые Крицкого – Менкеля и Пирсона III типа совпадают.

2.2. ПОСТРОЕНИЕ АНАЛИТИЧЕСКОЙ КРИВОЙ ОБЕСПЕЧЕННОСТЕЙ НОРМАЛЬНОГО ЗАКОНА РАСПРЕДЕЛЕНИЯ

Для расчета ординат аналитической кривой нормального закона распределения необходимо знать среднее значение исследуемого ряда и коэффициент вариации. Эти параметры были рассчитаны в работе № 1 п.1.2.3 – 1.2.4. В рассмотренном примере $\bar{Q} = 11,5$, $C_v = 0,25$ (рис. 2.1).

Расчет сводится в табл. 2.2. В первом столбце таблицы задаются опорные обеспеченности (одинаковые для всех). Для каждой опорной обеспеченности P из таблицы (с. 25 учебника [1])¹ выписываются нормированные ординаты кривой обеспеченностей нормального закона распределения t_p .

Таблица 2.2

Расчет координат аналитической кривой обеспеченностей нормального закона распределения для среднегодовых расходов воды

р. Воложба – д. Воложба; $\bar{Q} = 11,5$; $C_v = 0,25$

$P \%$	t_p	k_p	Q_p
0,01	3,72	1,93	22,2
0,1	3,09	1,77	20,4
1	2,33	1,58	18,2
5	1,64	1,41	16,2
10	1,28	1,32	15,2
20	0,84	1,21	13,9
30	0,52	1,13	13,0
50	0,00	1,00	11,5
70	– 0,52	0,87	10,0
80	– 0,84	0,79	9,09
90	– 1,28	0,68	7,82
95	– 1,64	0,59	6,79
99	– 2,33	0,42	4,80
99,9	– 3,09	0,23	2,62

В зависимости от t_p рассчитываются модульные коэффициенты по формуле:

$$k_p = t_p C_v + 1. \quad (2.2)$$

Переход от модульных коэффициентов расчетной обеспеченности к расходам расчетной обеспеченности производится по формуле:

$$Q_p = k_p \bar{Q}. \quad (2.3)$$

¹ Обратите внимание, что в таблице даны квантили распределения; для перехода к ординатам кривой обеспеченностей необходимо использовать формулу $P(x) = 100 - F(x)$.

По данным столбцов 1 и 3 (табл. 2.2) строится аналитическая кривая обеспеченностей нормального закона распределения. При правильном расчете и построении график на клетчатке вероятностей должен иметь вид прямой линии (см. рис.2.1).

2.3. ПОСТРОЕНИЕ АНАЛИТИЧЕСКОЙ КРИВОЙ ОБЕСПЕЧЕННОСТЕЙ ЛОГАРИФИЧЕСКИ НОРМАЛЬНОГО ЗАКОНА РАСПРЕДЕЛЕНИЯ

Преобразовать исходный ряд по формуле $z_i = \ln(Q_i)$. Для ряда z_i определить среднее значение и среднеквадратическое отклонение, с этой целью заполняется вспомогательная табл. 2.3.

Таблица 2.3

**Вспомогательная таблица для расчета статистических характеристик ряда z_i
р. Воложба – д. Воложба**

№ п/п	Q, м ³ /с	$z = \ln(Q)$	$z - \bar{z}$	$(z - \bar{z})^2$
1	9,61	2,263	-0,150	0,0225
2	6,89	1,930	-0,483	0,2331
3	8,66	2,159	-0,254	0,0646
4	7,37	1,997	-0,415	0,1726
5	8,48	2,138	-0,275	0,0757
6	8,16	2,099	-0,314	0,0983
7	12,4	2,518	0,105	0,0110
...	...	...	...	...
50	11,2	2,416	0,003	0,0000
51	13,6	2,610	0,197	0,0389
52	12,8	2,549	0,137	0,0187
53	11,6	2,451	0,038	0,0015
Сумма	610,6	127,9	0,000	3,344
Среднее	11,5	2,413	0,000	

Для ряда р. Воложба – д. Воложба получаем:

$$\bar{Q} = 11,5, \quad \bar{z} = 2,41; \quad \sigma_z = \sqrt{\frac{\sum_{i=1}^n (z_i - \bar{z})^2}{n-1}} = \sqrt{\frac{3,344}{53-1}} = 0,254.$$

Дальнейший расчет сводится в табл. 2.4, в первый столбец таблицы записываются опорные обеспеченности.

Для каждой опорной обеспеченности P из таблицы (стр. 25 учебника [1]) выписать нормированные ординаты кривой обеспеченностей нормального закона распределения t_p .

Таблица 2.4

Расчет координат аналитической кривой обеспеченностей логнормального закона распределения для среднегодовых расходов воды

р. Воложба – д. Воложба; $\bar{Q} = 11,5$, $\bar{z} = 2,41$, $\sigma_z = 0,25$

$P \%$	t_p	z_p	Q_p	k_p
0,01	3,72	3,35	28,5	2,48
0,1	3,09	3,19	24,4	2,12
1	2,33	3,00	20,1	1,75
5	1,64	2,83	16,9	1,47
10	1,28	2,74	15,4	1,34
20	0,84	2,62	13,8	1,20
30	0,52	2,54	12,7	1,10
50	0,00	2,41	11,1	0,97
70	– 0,52	2,28	9,8	0,85
80	– 0,84	2,20	9,0	0,78
90	– 1,28	2,08	8,0	0,70
95	– 1,64	1,99	7,3	0,64
99	– 2,33	1,82	6,2	0,53
99,9	– 3,09	1,63	5,1	0,44

В зависимости от t_p рассчитать значения z_p :

$$z_p = \sigma_z t_p + \bar{z}. \quad (2.4)$$

Для р. Воложба – д. Воложба при обеспеченности $P = 0,01 \%$ получаем $z_{0,01} = 0,254 \cdot 3,72 + 2,41 = 3,35$ и т. д.

В зависимости от z_p , рассчитать расходы расчетных обеспеченностей по формуле

$$Q_p = \exp(z_p). \quad (2.5)$$

Для р. Воложба – д. Воложба при обеспеченности $P = 0,01 \%$ получаем $Q_{0,01} = \exp(3,36) = 28,5$ и т. д.

Переход от расходов к модульным коэффициентам осуществляется по формуле

$$k_p = \frac{Q_p}{\bar{Q}}. \quad (2.6)$$

Для р. Воложба – д. Воложба при обеспеченности $P = 0,01\%$ получаем $k_{0,01} = \frac{28,5}{11,5} = 2,48$ и т. д.

По данным столбцов 1 и 5 табл. 2.4 на клетчатке вероятностей построить аналитическую кривую обеспеченностей логнормального закона распределения.

2.4. ПОСТРОЕНИЕ АНАЛИТИЧЕСКОЙ КРИВОЙ ОБЕСПЕЧЕННОСТЕЙ ГУМБЕЛЯ

Для расчета ординат аналитической кривой обеспеченностей закона распределения Гумбеля необходимо знать среднее значение исследуемого ряда и среднеквадратическое отклонение. Эти параметры были рассчитаны в работе № 1. В рассмотренном примере $\bar{Q} = 11,5$, $\sigma_Q = 2,91$. Расчет сводим в табл. 2.5.

В первый столбец табл. 2.5 записываются опорные обеспеченности.

Таблица 2.5

**Расчет координат аналитической кривой обеспеченностей Гумбеля для среднегодовых расходов воды р. Воложба – д. Воложба;
($Q = 11,5$, $\sigma_Q = 2,91$)**

$P \%$	y_p	Q_p	k_p
0,01	9,09	32,9	2,86
0,1	6,89	27,4	2,38
1	4,60	21,6	1,88
5	2,97	17,6	1,53
10	2,25	15,8	1,37
20	1,50	13,9	1,21
30	1,03	12,7	1,10
50	0,37	11,1	0,96
70	- 0,19	9,66	0,84
80	- 0,48	8,93	0,78
90	- 0,83	8,06	0,70
95	- 1,10	7,38	0,64
99	- 1,53	6,31	0,55
99,9	- 1,93	5,31	0,46

Во второй столбец записываются нормированные ординаты кривой обеспеченностей Гумбеля y_p . Ординаты y_p – это нормированные отклонения от моды, которые определяются по таблице (с. 34 учебника [1]).

Переход от y_p к расходам воды осуществляется по формуле

$$Q_p = q + \frac{1}{\alpha} y_p, \quad (2.7)$$

где q и α – параметры, для расчета которых необходимо знать статистические характеристики исходного ряда ($\bar{Q}$ и σ_Q), а также $\bar{y}$ и σ_y . Значения параметров $\bar{y}$ и σ_y определяются по таблице (с. 35 учебника [1]) в зависимости от длины исходного ряда. Для р. Воложба – д. Воложба $n = 53$, следовательно $\bar{y} = 0,55$; $\sigma_y = 1,17$.

Расчет параметров выражения (2.7) производится по формулам

$$\frac{1}{\alpha} = \frac{\sigma_Q}{\sigma_y}, \quad (2.8)$$

$$q = \bar{Q} - \bar{y} \left(\frac{1}{\alpha} \right). \quad (2.9)$$

Для р. Воложба – д. Воложба получаем:

$1/\alpha = 2,91/1,17 = 2,50$, $q = 11,5 - 0,55 \cdot 2,50 = 10,1$ и, следовательно, формула (2.7) в этом случае примет вид: $Q_p = 10,13 + 2,50 y_p$.

Переход от расходов воды к модульным коэффициентам осуществляется по формуле (2.6).

По данным столбцов 1 и 4 табл. 2.5 на клетчатке вероятностей построить аналитическую кривую обеспеченностей Гумбеля (см. рис.2.1).

2.5. ПОСТРОЕНИЕ АНАЛИТИЧЕСКОЙ КРИВОЙ ОБЕСПЕЧЕННОСТЕЙ ПИРСОНА III ТИПА

Для расчета ординат аналитической кривой Пирсона III типа необходимо знать среднее значение исследуемого ряда, коэффициент вариации и коэффициент асимметрии. Эти параметры были

рассчитаны в работе № 1, п. 1.2.4. В рассмотренном примере $\bar{Q} = 11,5$, $C_v = 0,25$, $C_s = 0,42$.

Следует отметить, что погрешность определения коэффициента асимметрии при современной длине рядов, как правило, существенно выше допустимой, поэтому на практике значение C_s , полученное по ряду наблюдений, не используется. Вместо этого значение коэффициента асимметрии определяется по среднему районному отношению коэффициента асимметрии к коэффициенту вариации (C_s/C_v).

В настоящей работе рекомендуется принять значение C_s/C_v исходя из соотношений представленных в табл.2.6.

Для р. Воложба – д. Воложба эмпирическое значение (C_s/C_v) = 1,68, следовательно, принимаем для расчетов (C_s/C_v) = 2. Тогда: $C_s = (C_s/C_v) C_v = 2 \cdot 0,25 = 0,50$.

Таблица 2.6

Рекомендуемые значения C_s/C_v

Эмпирическое значение	Рекомендуется принять
$(C_s/C_v) < 1,5$	$(C_s/C_v) = 1$
$1,5 \leq (C_s/C_v) \leq 2,5$	$(C_s/C_v) = 2$
$2,5 < (C_s/C_v) \leq 4$	$(C_s/C_v) = 3$
$(C_s/C_v) > 4$	$(C_s/C_v) = 4$

Полученное таким образом значение C_s используется в дальнейших расчетах. Расчет сводится в табл. 2.7.

В первом столбце таблицы задаются опорные обеспеченности.

Для каждой опорной обеспеченности P из таблицы (прил.1 учебника [1]) в зависимости от C_s выписываются нормированные ординаты кривой обеспеченностей Пирсона III типа t_p .

Затем по формуле (2.2) рассчитываются модульные коэффициенты.

Переход от модульных коэффициентов расчетной обеспеченности к расходам расчетной обеспеченности производится по формуле (2.3).

Таблица 2.7

Расчет координат аналитической кривой обеспеченностей Пирсона III типа для среднегодовых расходов воды р. Воложба – д. Воложба;

$$(\bar{Q} = 11,5, C_v = 0,25, C_s = 0,50)$$

$P \%$	t_p	k_p	Q_p
0,01	4,83	2,21	25,4
0,1	3,81	1,95	22,4
1	2,68	1,67	19,2
5	1,78	1,45	16,7
10	1,32	1,33	15,3
20	0,81	1,20	13,8
30	0,46	1,12	12,9
50	– 0,08	0,98	11,3
70	– 0,58	0,86	9,89
80	– 0,85	0,79	9,09
90	– 1,22	0,70	8,05
95	– 1,48	0,63	7,25
99	– 1,96	0,51	5,87
99,9	– 2,40	0,40	4,60

По данным столбцов 1 и 3 на клетчатке вероятностей строится аналитическая кривая обеспеченностей Пирсона III типа.

2.6. ПОСТРОЕНИЕ АНАЛИТИЧЕСКОЙ КРИВОЙ ОБЕСПЕЧЕННОСТЕЙ КРИЦКОГО-МЕНКЕЛЯ

Для расчета ординат аналитической кривой обеспеченностей Крицкого – Менкеля необходимо знать среднее значение исследуемого ряда, коэффициент вариации и отношение коэффициента асимметрии к коэффициенту вариации (C_s/C_v). Эти параметры были рассчитаны в работе № 1, п. 1.2.4.

В рассмотренном примере $\bar{Q} = 11,5$, $C_v = 0,25$, $(C_s/C_v) = 1,68$. Учитывая большую погрешность коэффициента асимметрии, значение C_s/C_v для р. Воложба – д. Воложба принималось в соответствии с табл. 2.6: $C_s/C_v = 2$. Это значение C_s/C_v использовалось в дальнейших расчетах.

Расчет сводится в табл. 2.8. В первую строку таблицы задаются опорные обеспеченности.

Таблица 2.8

Расчет координат аналитической кривой обеспеченностей Крицкого – Менкеля для среднегодовых расходов воды р. Воложба – д. Воложба;

$$\bar{Q} = 11,5, C_v = 0,25, C_s/C_v = 2$$

$P, \%$	0,01	0,1	1	5	10	20	30	50	70	80	90	95	99	99,9
k_p	2,22	1,96	1,67	1,45	1,33	1,20	1,11	0,98	0,86	0,79	0,70	0,63	0,52	0,40
Q_p	25,5	22,5	19,2	16,7	15,3	13,8	12,8	11,3	9,89	9,09	8,05	7,25	5,98	4,60

Для каждой опорной обеспеченности P из таблицы (прил.2 учебника [1]) в зависимости от соотношения (C_s/C_v) и коэффициента вариации выписываются модульные коэффициенты кривой обеспеченностей Крицкого – Менкеля k_p .

Переход от модульных коэффициентов к расходам расчетной обеспеченности производится по формуле (2.3).

По данным столбцов 1 и 3 (табл.2.8) на клетчатке вероятностей строится аналитическая кривая обеспеченностей Крицкого – Менкеля (см. рис.2.1).

Обратите внимание на то, что в случае рассматриваемого примера (р. Воложба – д. Воложба) кривые Пирсона III типа и Крицкого – Менкеля практически совпали. Это связано с тем, что при соотношении $C_s/C_v = 2$ обе эти кривые превращаются в двух-параметрическое гамма-распределение и поэтому должны совпадать.

ВЫВОДЫ

В выводах следует произвести сравнение всех рассмотренных кривых. Проанализировать, где между ними наблюдается наибольшее расхождение. Оценить, какая из аналитических кривых обеспеченностей лучше соответствует эмпирическим точкам.

ЛАБОРАТОРНАЯ РАБОТА № 3

Методы расчета оценок параметров распределения

Требуется рассчитать оценки параметров распределения методом моментов, методом наибольшего правдоподобия и графо-аналитическим методом.

ИСХОДНЫЕ ДАННЫЕ

В качестве исходных данных используются минимальные суточные зимние или летние расходы воды за период 40 лет по одной реке (тот же ряд, что и в работах № 1, № 2).

ТЕОРЕТИЧЕСКАЯ ЧАСТЬ

Для выполнения работы студенты должны изучить теоретический материал по учебнику [1]. Необходимо проработать следующие разделы: 3 (вводная часть); 3.2 (вводная часть); 3.2.1- 3.2.2; 3.2.3 (вводная часть); 3.2.3.1; 3.3.

После изучения теоретического материала студенты должны знать ответы на следующие вопросы:

1. В чем отличие истинных числовых характеристик случайной величины и числовых характеристик, рассчитанных по имеющемуся ряду?
2. Какие требования предъявляются к оценкам параметров распределения?
3. Перечислите методы расчета оценок параметров распределения? Каковы их достоинства и недостатки?
4. Что такое начальные и центральные статистические моменты? Чем отличаются теоретические и эмпирические статистические моменты?
5. Какой метод рекомендуется нормативными документами в качестве основного для расчета оценок параметров распределения гидрологических рядов?

РАСЧЕТНАЯ ЧАСТЬ

3.1. МЕТОД МОМЕНТОВ

Для ряда минимальных расходов рассчитать среднее значение, коэффициент вариации и коэффициент асимметрии методом моментов.

Рассчитать относительные погрешности среднего значения, коэффициента вариации и коэффициента асимметрии.

Примечание. Этот расчет уже выполнен в лабораторной работе № 1 (см. п. 1.2.4), поэтому ход решения и результаты расчетов можно просто переписать.

3.2. МЕТОД НАИБОЛЬШЕГО ПРАВДОПОДОБИЯ

При расчете оценок параметров распределения методом наибольшего правдоподобия необходимо заранее знать для какой аналитической кривой производится расчет. В данной работе в качестве расчетной используется кривая обеспеченностей Крицкого – Менкеля.

3.2.1. Расчет статистик λ_1 , λ_2 и λ_3

Для вычисления статистик λ_1 , λ_2 и λ_3 заполняется вспомогательная табл. 3.1. Расчет статистик λ_1 , λ_2 и λ_3 производится по формулам:

$$\lambda_1 = \bar{Q} = \frac{\sum_{i=1}^n Q_i}{n}, \quad (3.1) \quad \lambda_2 = \frac{\sum_{i=1}^n \lg(k_i)}{n-1}, \quad (3.2) \quad \lambda_3 = \frac{\sum_{i=1}^n k_i \lg(k_i)}{n-1}, \quad (3.3)$$

где $\bar{Q}$ – среднее значение; k_i – модульные коэффициенты; n – длина ряда.

Как видно из табл. 3.1 для р. Воложба – д. Воложба получаем:

$$\lambda_1 = \bar{Q} = \frac{610,61}{53} = 11,52; \quad \lambda_2 = \frac{-0,721}{52} = -0,0139; \quad \lambda_3 = \frac{0,713}{52} = 0,0137.$$

Таблица 3.1

Вспомогательная таблица для расчета статистик λ_1 , λ_2 и λ_3

№ п/п	Год	Q, м ³ /с	k	lg(k)	k lg(k)
1	1936	9,61	0,8341	-0,07876	-0,06570
2	1937	6,89	0,5980	-0,22327	-0,13352
3	1938	8,66	0,7517	-0,12397	-0,09319
4	1939	7,37	0,6397	-0,19402	-0,12412
5	1940	8,48	0,7361	-0,13309	-0,09796
6	1941	8,16	0,7083	-0,14980	-0,10610
7	1942	12,4	1,0763	0,03193	0,03437
8	1943	11,8	1,0242	0,01039	0,01065
...	...	...	...	...	...
46	1981	12,4	1,0763	0,03193	0,03437
47	1982	14,8	1,2846	0,10877	0,13973
48	1983	14,2	1,2325	0,09080	0,11191
49	1984	14,5	1,2586	0,09988	0,12571
50	1985	11,2	0,9721	-0,01227	-0,01193
51	1986	13,6	1,1805	0,07205	0,08505
52	1987	12,8	1,1110	0,04572	0,05080
53	1988	11,6	1,0069	0,00297	0,00299
Сумма		610,61		-0,721	0,713

3.2.2. Определение статистических характеристик ряда

Метод наибольшего правдоподобия для кривой обеспеченностей Крицкого – Менкеля позволяет определить в явном виде только среднее значение.

Коэффициент вариации (C_v) и отношение коэффициента асимметрии к коэффициенту вариации (C_s/C_v) определяются по специальным номограммам в зависимости от статистик λ_2 и λ_3 . Фрагмент такой номограммы показан на рис.3.1.

Крупномасштабные номограммы входят в состав «Атласа расчетных гидрологических карт и номограмм», который является приложением к Пособию [4]. Мелкомасштабные номограммы для значений $C_v = 0,2-1,6$ представлены в Приложениях 1-5 настоящего практикума.

Как видно на рис. 3.1 для реки Воложба – д. Воложба $C_v = 0,25$; $C_s/C_v = 2,3$. Следовательно, $C_s = 2,3 \times 0,25 = 0,58$.

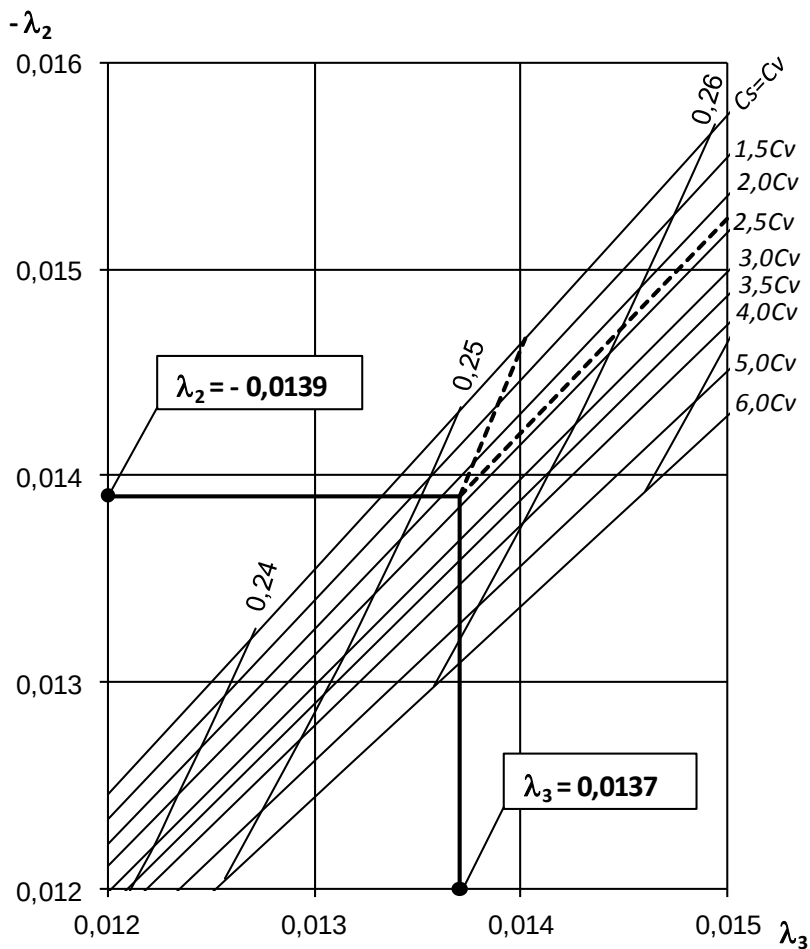


Рис. 3.1. Фрагмент номограммы для вычисления параметров трехпараметрического гамма-распределения (Крицкого – Менкеля) C_v и C_s методом приближенного наибольшего правдоподобия.

3.2.3. Оценка относительных погрешностей расчета статистических характеристик

Оценка относительных погрешностей среднего значения и коэффициента асимметрии производится с использованием тех же формул, что и в методе моментов. Используются соответственно формулы (1.18) и (1.20).

Относительная погрешность для коэффициента вариации рассчитывается по формуле

$$\varepsilon_{C_v, \%} = \sqrt{\frac{3}{2n(3 + C_v^2)}} 100\%. \quad (3.4)$$

Для р. Воложба – д. Воложба получаем:

$$\varepsilon_{Q, \%} = \frac{0,25}{\sqrt{53}} 100\% = 3,4\%;$$

$$\varepsilon_{C_v, \%} = \sqrt{\frac{3}{2 \cdot 53 \cdot (3 + 0,25^2)}} 100\% = 9,6\%;$$

$$\varepsilon_{\Sigma, \%} = \frac{1}{0,58} \sqrt{\frac{6}{53} (1 + 6 \cdot 0,25^2 + 5 \cdot 0,25^4)} 100\% = 68,5\%.$$

3.3. ГРАФО-АНАЛИТИЧЕСКИЙ МЕТОД (МЕТОД КВАНТИЛЕЙ)

В российской гидрологической практике при реализации графоаналитического метода в качестве опорной кривой принято использовать кривую обеспеченностей Пирсона III типа. Методика разработана Г. А. Алексеевым и опубликована в 1960 г.

3.3.1. Построение сглаженной эмпирической кривой обеспеченностей и определение трех опорных ординат

Построить эмпирическую кривую обеспеченностей. Техника построения эмпирической кривой обеспеченностей была рассмотрена в лабораторной работе № 2, п. 2.1.

В поле эмпирических точек от руки проводится сглаживающая линия. Эту линию называют сглаженной эмпирической кривой обеспеченностей. По сглаженной эмпирической кривой для опорных обеспеченностей $P = 5\%$, $P = 50\%$ и $P = 95\%$ определяются модульные коэффициенты $k_{5\%}$, $k_{50\%}$ и $k_{95\%}$ (рис. 3.2).

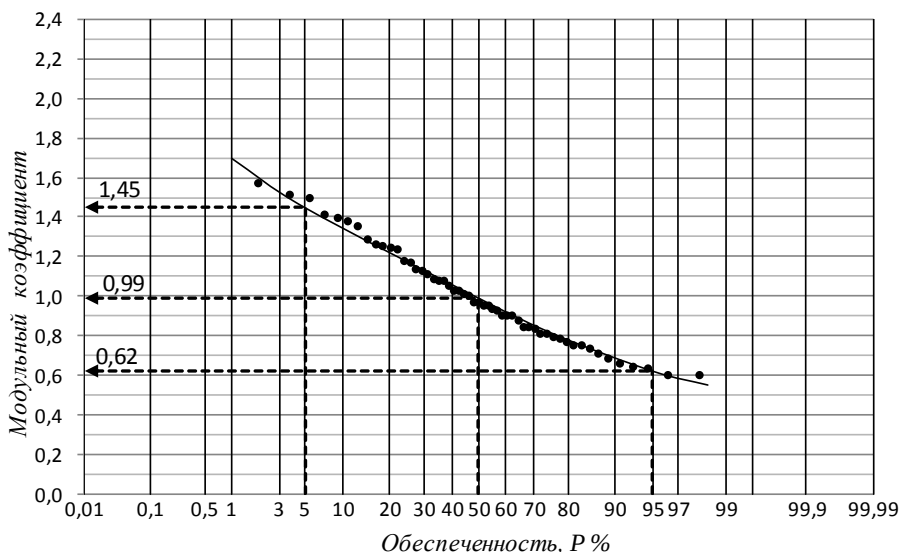


Рис. 3.2. Сглаженная эмпирическая кривая обеспеченностей среднегодовых расходов воды р. Воложба – д. Воложба

Умножая модульные коэффициенты на среднее значение, получаем расходы воды Q_5 , Q_{50} , Q_{95} .

Если кривая обеспеченностей строится не в модульных коэффициентах, а в расходах, то значения Q_5 , Q_{50} , Q_{95} определяются непосредственно по сглаженной эмпирической кривой.

Для р. Воложба – д. Воложба получаем:

$$k_{5\%} = 1,45; \quad k_{50\%} = 0,99; \quad k_{95\%} = 0,62; \quad Q_5 = 1,45 \cdot 11,5 = 16,7;$$

$$Q_{50} = 0,99 \cdot 11,5 = 11,4; \quad Q_{95} = 0,62 \cdot 11,5 = 7,13.$$

3.3.2. Расчет параметров распределения

Расчет осуществляется в следующей последовательности:

1. Рассчитывается коэффициент скошенности:

$$S = \frac{Q_5 + Q_{95} - 2 \cdot Q_{50}}{Q_5 - Q_{95}}. \quad (3.5)$$

Для р. Воложба – д. Воложба $S = 0,11$.

2. По таблице (прил. 1 пособия [1]) в зависимости от коэффициента скошенности определяются коэффициент асимметрии и нормированные ординаты кривой Пирсона III типа t_5 , t_{50} , и t_{95} .
Для р. Воложба – д. Воложба получаем:
 $C_s = 0,40$, $t_5 = 1,75$, $t_{50} = -0,07$, $t_{95} = -1,52$.

3. Рассчитывается среднеквадратическое отклонение по формуле

$$\sigma = \frac{Q_5 - Q_{95}}{t_5 - t_{95}}. \quad (3.6)$$

Для р. Воложба – д. Воложба $\sigma = \frac{16,7 - 7,13}{1,75 - (-1,52)} = 2,93$.

4. Рассчитывается новое среднее значение $\bar{Q}_r$ по формуле

$$\bar{Q}_{\text{граф.}} = Q_{50} - \sigma \cdot t_{50} \quad (3.7)$$

Для р. Воложба – д. Воложба получаем:

$$\bar{Q}_{\text{граф.}} = 11,4 - 2,93 \cdot (-0,07) = 11,6.$$

5. Определяется коэффициент вариации по формуле

$$C_v = \frac{\sigma}{\bar{Q}_{\text{граф.}}}. \quad (3.8)$$

Для р. Воложба – д. Воложба $C_v = 2,93/11,6 = 0,25$.

6. Определяется соотношение C_s/C_v .

Для р. Воложба – д. Воложба $C_s/C_v = 0,40/0,25 = 1,6$.

3.3.3. Анализ надежности расчета

При использовании графоаналитического метода абсолютные и относительные погрешности параметров распределения не определяются. Считается, что расчет проведен корректно, если выполняется условие

$$\left| \bar{Q} - \bar{Q}_{\text{граф.}} \right| < 0,02 \cdot \bar{Q}. \quad (3.9)$$

ВЫВОДЫ

Для выполнения анализа заполняется сводная табл. 3.2.

Таблица 3.2

Значения статистических характеристик среднегодовых расходов воды, рассчитанные разными методами р. Воложба – д. Воложба

Метод расчета	Стат. характеристика				Погрешность, %		
	$\bar{Q}$	C_v	C_s	C_s/C_v	$\bar{Q}$	C_v	C_s
Моментов	11,5	0,25	0,42	1,7	3,4	10,0	95
Наибольшего правдоподобия	11,5	0,25	0,58	2,3	3,4	9,6	69
Графоаналитический	11,6	0,25	0,40	1,6	–	–	–

В выводах следует сравнить значения параметров распределения, полученные разными методами. Описать достоинства и недостатки каждого из методов. Отметить, какой из методов нормативные документы рекомендуют в качестве основного.

ЛАБОРАТОРНАЯ РАБОТА № 4

Интервальные оценки параметров распределения и проверка гидрологических рядов на однородность

Требуется построить доверительные интервалы для математического ожидания, дисперсии и среднеквадратического отклонения; выполнить проверку однородности ряда с использованием

параметрических критериев Фишера и Стьюдента и непараметрических критериев Диксона и Смирнова – Граббса.

ИСХОДНЫЕ ДАННЫЕ

1. В качестве исходных данных используются минимальные 30-суточные зимние или летние расходы воды за период 40 лет по одной реке (тот же ряд, что и в работах № 1, № 2, № 3).
2. Числовые характеристики, рассчитанные в работе № 1 (среднее значение, среднеквадратическое отклонение, дисперсия).

ТЕОРЕТИЧЕСКАЯ ЧАСТЬ

Для выполнения работы студенты должны изучить теоретический материал по учебнику [1]. Необходимо проработать следующие разделы: 4.1 - 4.5; 4.6 (вводная часть); 4.6.1; 4.6.2.

После изучения теоретического материала студенты должны знать ответы на следующие вопросы:

1. Что такое интервальная оценка параметра распределения и чем она отличается от точечной оценки?
2. Что такое статистическая гипотеза? Для чего и как формулируются нулевая и альтернативная гипотезы?
3. Что такое доверительный интервал, критическая область, доверительная вероятность, уровень значимости?
4. Как проводится проверка однородности гидрологических рядов с использованием критерия Стьюдента?
5. Как проводится проверка однородности гидрологических рядов с использованием критерия Фишера?
6. С помощью каких критериев проводится анализ эмпирических точек, резко отклоняющихся от аналитической кривой обеспеченностей в области больших и малых значений?

РАСЧЕТНАЯ ЧАСТЬ

4.1. ИНТЕРВАЛЬНЫЕ ОЦЕНКИ ПАРАМЕТРОВ РАСПРЕДЕЛЕНИЯ

Интервальной оценкой некоторого параметра G называется интервал, границы которого являются функциями выборочных

значений $Q_1, Q_2, \dots, Q_n$ и который с заданной вероятностью покрывает оцениваемый параметр G .

Требуется построить интервальные оценки для математического ожидания, дисперсии и среднеквадратического отклонения исследуемого ряда. Оценки дать в виде 90%-ных доверительных интервалов.

4.1.1. Интервальная оценка математического ожидания

Теоретически доказано, что случайная величина $(\bar{Q} - m_Q)/(\sigma^*/\sqrt{n})$ имеет распределение Стьюдента с числом степеней свободы $\nu = n - 1$, где $\bar{Q}$ – среднее значение; m_Q – математическое ожидание; σ^* – выборочное среднеквадратическое отклонение; n – длина исследуемого ряда. Отсюда следует, что 90%-ный доверительный интервал для математического ожидания может быть представлен в виде:

$$\bar{Q} - \frac{t_{95\%}\sigma^*}{\sqrt{n}} \leq m_Q < \bar{Q} + \frac{t_{95\%}\sigma^*}{\sqrt{n}}, \quad (4.1)$$

где $t_{95\%}$ – 95%-ный квантиль распределения Стьюдента, который определяется по прил. 5 учебника [1].

Для р. Воложба – д. Воложба получаем:

$$11,5 - \frac{1,69 \cdot 2,91}{\sqrt{53}} \leq m_Q < 11,5 + \frac{1,69 \cdot 2,91}{\sqrt{53}},$$

$$10,82 \leq m_Q < 12,18.$$

4.1.2. Интервальная оценка дисперсии

Теоретически доказано, что случайная величина $\frac{(n-1)D^*}{D}$ имеет распределение хи-квадрат с $\nu = (n - 1)$ степенями свободы, где D^* и D – соответственно выборочное и теоретическое значение дисперсии. Отсюда следует, что 90%-ный доверительный интервал для дисперсии может быть представлен в виде:

$$\frac{(n-1)D^*}{\chi_{95\%}^2} < D \leq \frac{(n-1)D^*}{\chi_{5\%}^2}, \quad (4.2)$$

где $\chi_{5\%}^2$ и $\chi_{95\%}^2$ – соответственно 5%-ный и 95%-ный квантили распределения хи-квадрат.

Значения квантилей распределения хи-квадрат при числе степеней свободы $\nu \leq 30$ можно определить по прил. 4 учебника [1].

При $\nu > 30$ для определения расчетных квантилей используется формула

$$\chi_{p\%}^2 = \frac{1}{2}(\sqrt{2\nu-1} + t_{p\%})^2, \quad (4.3)$$

где $t_{p\%}$ – p %-ный квантиль нормального распределения, определяемый по табл. 2.1 учебника [1].

В рассматриваемом примере (р. Воложба – д. Воложба) число степеней свободы $\nu = 53 - 1 = 52$, следовательно, для определения расчетных квантилей используем формулу (4.3):

$$\chi_{5\%}^2 = \frac{1}{2}(\sqrt{2 \cdot 52 - 1} - 1,64)^2 = 36,2,$$

$$\chi_{95\%}^2 = \frac{1}{2}(\sqrt{2 \cdot 52 - 1} + 1,64)^2 = 69,5.$$

Таким образом, для р. Воложба – д. Воложба 90%-ный доверительный интервал для дисперсии будет иметь вид:

$$\frac{(53-1)8,47}{69,5} < D \leq \frac{(53-1)8,47}{36,2},$$

$$6,34 < D \leq 12,17.$$

4.1.3. Интервальная оценка среднеквадратического отклонения

Интервальная оценка для среднеквадратического отклонения получается из выражения (4.2) с учетом того, что $\sigma = \sqrt{D}$:

$$\sqrt{\frac{(n-1)D^*}{\chi_{95\%}^2}} < \sigma \leq \sqrt{\frac{(n-1)D^*}{\chi_{5\%}^2}}. \quad (4.4)$$

Для реки Воложба – д. Воложба получаем:

$$\sqrt{\frac{(53-1)8,47}{69,5}} < \sigma \leq \sqrt{\frac{(53-1)8,47}{36,2}},$$

$$2,52 < \sigma \leq 3,49.$$

По результатам выполненных расчетов составляется сводная табл. 4.1.

Таблица 4.1

Сводная таблица интервальных и точечных оценок параметров распределения среднегодовых расходов воды р. Воложба – д. Воложба

Параметр	Точечная оценка	Интервальная оценка
m_Q	11.5	$10,8 \leq m_Q < 12,2$
D	8.47	$6,34 < D \leq 12,2$
σ	2.91	$2,52 < \sigma \leq 3,49$

4.2. ПРОВЕРКА ОДНОРОДНОСТИ ГИДРОЛОГИЧЕСКОГО РЯДА С ИСПОЛЬЗОВАНИЕМ ПАРАМЕТРИЧЕСКИХ КРИТЕРИЕВ ФИШЕРА И СТЬЮДЕНТА

Расчет числовых характеристик является корректным только в том случае, если ряд однороден. То есть в течение всего периода наблюдений условия формирования стока не изменялись.

Для проверки однородности гидрологических рядов используются специальные тесты – *критерии однородности*.

В данной работе для проверки однородности используются два параметрических критерия: критерий Стьюдента и критерий Фишера. Критерий Стьюдента позволяет провести проверку ряда на однородность по среднему значению, а критерий Фишера – по дисперсии.

Требуется провести проверку ряда на однородность при уровне значимости $2\alpha = 5\%$.

Для выполнения работы исходный ряд разбивается на две приблизительно равные части и для каждой части ряда рассчитываются среднее значение, среднееквадратическое отклонение и дисперсия (табл. 4.2).

Таблица 4.2

Основные статистические характеристики среднегодовых расходов воды по первой и второй частям ряда р. Воложба – д. Воложба

Выборка	Длина выборки	Среднее значение	СКО	Дисперсия
I часть ряда	26	11,2	3,13	9,80
II часть ряда	27	11,9	2,70	7,27
Весь ряд	53	11,5	2,91	8,47

4.2.1. Проверка однородности гидрологического ряда по дисперсии (критерий Фишера)

Рассчитать эмпирическое значение критерия Фишера:

$$F^* = \frac{D_1}{D_2}. \quad (4.5)$$

где D_1 и D_2 – дисперсии по одной и другой частям ряда, причем в числитель следует ставить большую из двух дисперсий.

Для р. Воложба – д. Воложба получаем: $F^* = 9,80 / 7,27 = 1,35$.

Эмпирическое значение статистики Фишера сравнивается с теоретическим F_T при уровне значимости $2\alpha = 5\%$. Теоретическое значение статистики Фишера определяется по таблице F -распределения (прил. 6 учебника [1]) в зависимости от принятого уровня значимости и числа степеней свободы ν_1 и ν_2 :

$$v_1 = n_1 - 1, \quad (4.6)$$

$$v_2 = n_2 - 1, \quad (4.7)$$

где n_1 – длина выборки с большей дисперсией; n_2 – длина выборки с меньшей дисперсией.

Для р. Воложба – д. Воложба получаем: $v_1 = 26 - 1 = 25$, $v_2 = 27 - 1 = 26$, $F_T = 2,17$. Поскольку в данном случае эмпирическое значение статистики Фишера меньше теоретического:

$$(F^* = 1,35) < (F_T = 2,17),$$

то можно считать различие в дисперсиях по отдельным частям ряда незначительным. В этом случае говорят, что гипотеза об однородности ряда по критерию Фишера при уровне значимости $2\alpha = 5\%$ не опровергается.

4.2.2. Проверка однородности гидрологического ряда по среднему значению (критерий Стьюдента)

Рассчитать эмпирическое значение критерия Стьюдента:

$$t^* = \left[(\bar{Q}_1 - \bar{Q}_2) / \sqrt{\frac{(n_1 - 1)\sigma_1^2 + (n_2 - 1)\sigma_2^2}{n_1 + n_2 - 2}} \right] \sqrt{\frac{n_1 n_2}{n_1 + n_2}}, \quad (4.8)$$

где n_1 и n_2 – длина первой и второй частей ряда; $\bar{Q}_1$ и $\bar{Q}_2$ – средние значения по первой и второй частям ряда; σ_1 и σ_2 – средне-квадратические отклонения по первой и второй частям ряда.

Для рассматриваемого примера получаем:

$$t^* = \frac{(11,2 - 11,9) \sqrt{(26 \cdot 27) / (26 + 27)}}{\sqrt{\frac{(26 - 1)(3,13)^2 + (27 - 1)(2,70)^2}{26 + 27 - 2}}} = -0,87.$$

Эмпирическое значение статистики Стьюдента сравнивается с теоретическим t_T при уровне значимости $2\alpha = 5\%$.

Теоретическое значение статистики Стьюдента определяется по таблице t -распределения (прил. 5 учебника [1]) в зависимости от принятого уровня значимости и числа степеней свободы ν :

$$\nu = n - 1, \quad (4.9)$$

где n – общая длина исследуемого ряда.

Для р. Воложба – д. Воложба получаем:

$$\nu = 53 - 1 = 52; \quad t_T = 2,007.$$

Поскольку в данном случае эмпирическое значение статистики Стьюдента по абсолютной величине меньше теоретического

$$(|t^*| = 0,87) < (t_T = 2,007),$$

то можно считать различие в средних значениях по отдельным частям ряда незначительным.

В этом случае говорят, что гипотеза об однородности ряда по критерию Стьюдента при уровне значимости $2\alpha = 5\%$ *не опровергается*.

По результатам проверки ряда на однородность с использованием критериев Фишера и Стьюдента составляется сводная табл. 4.3.

Таблица 4.3

Результаты проверки ряда на однородность по критериям Фишера и Стьюдента для среднегодовых расходов воды р. Воложба – д. Воложба при уровне значимости $2\alpha = 5\%$.

Критерий	Значение статистики		Нулевая гипотеза	
	эмпирическое	теоретическое	содержание	результат
Фишера	1,35	2,17	$D_1 = D_2$	не опровергается
Стьюдента	– 0,87	2,007	$\bar{Q}_1 = \bar{Q}_2$	не опровергается

4.3. ПРОВЕРКА ОДНОРОДНОСТИ ГИДРОЛОГИЧЕСКОГО РЯДА С ИСПОЛЬЗОВАНИЕМ НЕПАРАМЕТРИЧЕСКИХ КРИТЕРИЕВ ДИКСОНА И СМИРНОВА – ГРАББСА

Построение эмпирической и аналитической кривых обеспеченностей позволяет выявить точки, резко отклоняющиеся от аналитической кривой в области больших и малых значений. В этой ситуации необходимо выяснить, относится ли данная точка к той же генеральной совокупности, что и остальные члены выборки или нет.

В первом случае отклонение можно объяснить тем, что данный гидрологический ряд содержит значение очень редкой повторяемости, во втором – ряд следует признать неоднородным.

Для анализа таких ситуаций в СП 33-101-2003 включены два дополнительных критерия однородности – критерий Диксона и критерий Смирнова – Граббса.

Требуется провести проверку ряда на однородность при уровне значимости $\alpha = 5\%$ с использованием непараметрических критериев Диксона и Смирнова – Граббса

При использовании указанных критериев исходный ряд ранжируется в возрастающем порядке: $Q_1 < Q_2 < \dots < Q_n$ (табл.4.4).

Таблица 4.4

Исходный и ранжированный (в возрастающем порядке) ряды среднегодовых расходов воды р. Воложба – д. Воложба

№ п/п	Год	Расход воды Q_i , м ³ /с	
		Исходный ряд	Ранжированный ряд
1	1936	9,61	6,89
2	1937	6,89	6,94
3	1938	8,66	7,27
...	...	...	...
51	1986	13,60	17,20
52	1987	12,80	17,40
53	1988	11,60	18,10

Затем определяются значения Q_1 , Q_2 , Q_{n-1} , Q_n , а также среднее значение ряда и среднеквадратическое отклонение (табл. 4.5) . Эти

характеристики используются в дальнейшем для расчета статистик Диксона и Смирнова – Граббса.

Таблица 4.5

Вспомогательная таблица для расчета статистик Диксона и Смирнова – Граббса

Q_1	Q_2	Q_{n-1}	Q_n	$\bar{Q}$	σ^*
6,89	6,94	17,40	18,10	11,5	2,91

4.3.1. Проверка однородности гидрологического ряда по критерию Диксона

Статистики критерия Диксона для максимального и минимального членов выборки имеют вид:

$$D_{\max}^* = \frac{Q_n - Q_{n-1}}{Q_n - Q_1}, \quad (4.10)$$

$$D_{\min}^* = \frac{Q_1 - Q_2}{Q_1 - Q_n}. \quad (4.11)$$

Гипотеза об однородности ряда по критерию Диксона не опровергается, если

$$D^* < D_\alpha, \quad (4.12)$$

где D^* – эмпирическое значение статистики Диксона; D_α – теоретическое значение статистики Диксона, определяемое по приложению 10 учебника [1] в зависимости от уровня значимости α %, коэффициента асимметрии Cs , длины выборки n и коэффициента автокорреляции $r(1)$ (в данной работе принять: $r(1) = 0$) (табл. 4.6).

Таблица 4.6

Теоретические значения статистики Диксона для ряда среднегодовых расходов воды р. Воложба – д. Воложба

n	α	$r(1)$	Cs	Теоретическое значение статистики Диксона	
				$D_{\max}$	$D_{\min}$
53	5 %	0	0,42	0,28	0,15

Для рассматриваемого примера получаем:

$$D_{\max}^* = \frac{18,10 - 17,40}{18,10 - 6,89} = 0,062; \quad D_{\min}^* = \frac{6,89 - 6,94}{6,89 - 18,10} = 0,004.$$

Так как в данном случае

$$(D_{\max}^* = 0,062) < (D_{\max} = 0,28),$$

$$(D_{\min}^* = 0,004) < (D_{\min} = 0,15),$$

то гипотеза об однородности ряда по критерию Диксона не опровергается при уровне значимости 5 %.

4.3.2. Проверка однородности гидрологического ряда по критерию Смирнова – Граббса

Статистики критерия Смирнова – Граббса для максимального и минимального членов выборки имеют вид:

$$G_{\max}^* = \frac{Q_n - \bar{Q}}{\sigma^*}, \quad (4.13)$$

$$G_{\min}^* = \frac{\bar{Q} - Q_1}{\sigma^*}, \quad (4.14)$$

где $\bar{Q}$ – среднее значение; σ^* – выборочное среднеквадратическое отклонение.

Гипотеза об однородности ряда по критерию Смирнова – Граббса не опровергается, если

$$G^* < G_{\alpha}, \quad (4.15)$$

где G^* – эмпирическое значение статистики Смирнова – Граббса; G_{α} – теоретическое значение статистики Смирнова – Граббса, определяемое по приложению 11 учебника [1] в зависимости от уровня значимости α %, коэффициента асимметрии C_s , длины выборки n и коэффициента автокорреляции $r(1)$ (в данной работе принять: $r(1) = 0$) (табл. 4.7).

Для рассматриваемого примера получаем:

$$G_{\max}^* = \frac{18,10 - 11,5}{2,91} = 2,26; \quad G_{\min}^* = \frac{11,5 - 6,89}{2,91} = 1,59$$

Таблица 4.7

Теоретические значения статистики Смирнова – Граббса для ряда среднегодовых расходов воды р. Воложба – д. Воложба

n	α	r(1)	Cs	Теоретическое значение статистики Смирнова – Граббса	
				$D_{\max}$	$D_{\min}$
53	5 %	0	0,42	3,44	2,46

Так как в данном случае

$$(G_{\max}^* = 2,26) < (G_{\max} = 3,44),$$

$$(G_{\min}^* = 1,59) < (G_{\min} = 2,46),$$

то гипотеза об однородности ряда по критерию Смирнова – Граббса не опровергается при уровне значимости 5% .

По результатам проверки ряда на однородность с использованием критериев Диксона и Смирнова – Граббса составляется сводная табл. 4.8.

Таблица 4.8

Результаты проверки ряда на однородность по критериям Диксона и Смирнова – Граббса для среднегодовых расходов воды р. Воложба – д. Воложба при уровне значимости $\alpha = 5\%$.

Критерий	Значение статистики		Нулевая гипотеза	
	эмпирическое	теоретическое	содержание	результат
Диксона для $Q_{\max}$	0,062	0,28	$Q_{\max} \in Q$	Не опр.
Диксона для $Q_{\min}$	0,004	0,15	$Q_{\min} \in Q$	Не опр.
Смирнова – Граббса для $Q_{\max}$	2,26	3,44	$Q_{\max} \in Q$	Не опр.
Смирнова – Граббса для $Q_{\min}$	1,59	2,46	$Q_{\min} \in Q$	Не опр.

ВЫВОДЫ

В выводах следует отметить особенности точечных и интервальных оценок параметров.

При анализе однородности четко сформулировать для каждого критерия нулевую гипотезу и пояснить, какие результаты дала проверка этой гипотезы.

ЛАБОРАТОРНАЯ РАБОТА № 5

Анализ взаимосвязи двух переменных (парная корреляция)

Требуется построить график связи зависимой и независимой переменных, рассчитать параметры уравнения линейной регрессии и оценить их надежность; линию регрессии нанести на график и построить для нее 95%-ные доверительные границы.

ИСХОДНЫЕ ДАННЫЕ

В качестве исходных данных в работе используются два ряда продолжительностью по 20 лет: минимальные 30-суточные зимние (или летние) расходы воды и минимальные суточные зимние (или летние) расходы воды.

Ряд минимальных суточных расходов представляет собой фрагмент ряда, который использовался в работах № 1 – № 4. Ряды должны быть выписаны за совместный период наблюдений.

Минимальные 30-суточные расходы воды должны рассматриваться как независимая переменная x , а минимальные суточные расходы – как зависимая переменная y .

В примере, который приводится ниже, в качестве x и y фигурируют другие переменные. В качестве независимой переменной (x) рассматриваются среднегодовые расходы воды по р. Воложба – д. Воложба, а в качестве зависимой (y) – среднегодовые расходы воды по р. Воложба – д. Пареево (табл. 5.1).

Таблица 5.1

**Среднегодовые расходы воды (м³/с) р. Воложба в створе д. Воложба (x_i) и
р. Воложба в створе д. Пареево (y_i)**

№ п/п	Год	x_i	y_i	№ п/п	Год	x_i	y_i
1	1952	16,1	9,17	20	1971	9,37	5,85
2	1953	18,1	9,76	21	1972	7,58	4,84
3	1954	11,0	6,40	22	1973	7,27	4,95
4	1955	15,9	9,13	23	1974	10,4	6,68
5	1956	13,1	7,74	24	1975	8,62	5,56
6	1957	17,4	10,2	25	1976	12,5	7,35
7	1958	14,4	9,58	26	1977	13,0	7,32
8	1959	11,0	7,07	27	1978	13,5	8,26
9	1960	7,86	4,82	28	1979	8,99	5,74
10	1961	11,5	7,49	29	1980	9,30	6,11
11	1962	16,3	10,1	30	1981	12,4	7,73
12	1963	9,70	5,59	31	1982	14,8	9,47
13	1964	10,1	6,10	32	1983	14,2	8,89
14	1965	10,4	6,04	33	1984	14,5	8,89
15	1966	17,2	8,82	34	1985	11,2	7,08
16	1967	12,1	7,07	35	1986	13,6	8,05
17	1968	14,3	7,35	36	1987	12,8	7,60
18	1969	15,6	8,88	37	1988	11,6	7,47
19	1970	9,10	5,60				

ТЕОРЕТИЧЕСКАЯ ЧАСТЬ

Для выполнения работы студенты должны изучить теоретический материал по учебнику [1]. Необходимо проработать следующие разделы: 5 (вводная часть), 5.1, 5.3, 5.4.6.

После изучения теоретического материала студенты должны знать ответы на следующие вопросы:

1. В чем различие функциональных и стохастических связей и почему при описании гидрологических явлений мы, как правило, имеем дело со стохастическими зависимостями?
2. Какой метод используется для расчета параметров уравнения регрессии?
3. Что характеризует коэффициент парной корреляции? В каком диапазоне значений он может изменяться?
4. Как называются коэффициенты a и b в уравнении линейной регрессии $y = ax + b$?

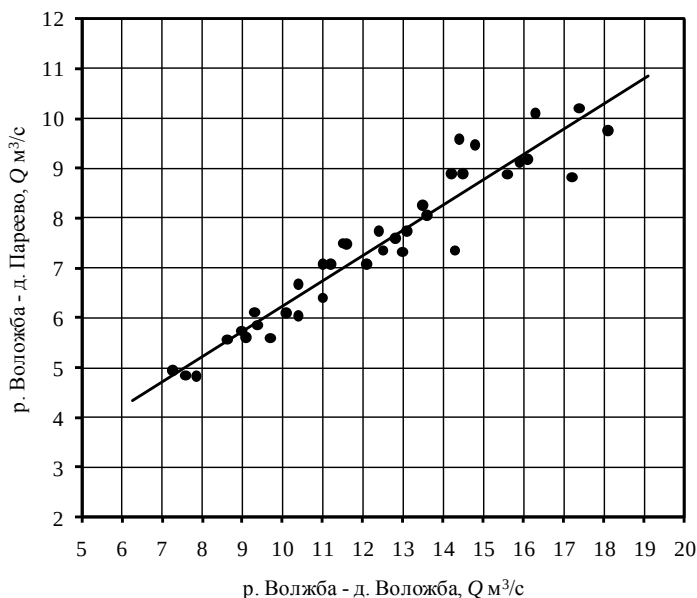
5. Когда уравнение регрессии считается надежным и его можно использовать для практических расчетов?

РАСЧЕТНАЯ ЧАСТЬ

В случае линейной зависимости между двумя переменными можно построить график связи вида $y = f(x)$ и провести в поле точек прямую линию $y = ax + b$. Линия, относительно которой наблюдается наименьший разброс точек, называется *линией регрессии*, а соответствующее ей аналитическое выражение – *уравнением регрессии*.

5.1. ПОСТРОЕНИЕ ГРАФИКА СВЯЗИ $y = f(x)$

График связи $y = f(x)$ строится на миллиметровой бумаге формата А4 по данным табл. 5.1. (рис.5.1).



**Рис. 5.1. График связи среднегодовых расходов воды
р. Волжба – д. Волжба и р. Волжба – д. Пареево.**

5.2. РАСЧЕТ ПАРАМЕТРОВ УРАВНЕНИЯ ЛИНЕЙНОЙ РЕГРЕССИИ

Для расчета параметров уравнения линейной регрессии заполняется вспомогательная табл. 5.2.

Таблица 5.2

Вспомогательная таблица для расчета параметров уравнения линейной регрессии для двух переменных

№ п/п	x_i	y_i	$\Delta x = x_i - \bar{x}$	$\Delta y = y_i - \bar{y}$	$(\Delta x)^2$	$(\Delta y)^2$	$\Delta x \Delta y$	$(\Delta x + \Delta y)^2$
1	16,1	9,17	3,754	1,744	14,095	3,043	6,549	30,235
2	18,1	9,76	5,754	2,334	33,112	5,449	13,432	65,426
3	11,0	6,40	-1,346	-1,026	1,811	1,052	1,380	5,623
...	...	...	...	...	...	...	...	...
36	12,8	7,60	0,454	0,174	0,206	0,030	0,079	0,395
37	11,6	7,47	-0,746	0,044	0,556	0,002	-0,033	0,492
Сумма	456,8	274,75	0,00	0,00	309,74	87,25	157,34	711,67
Среднее	12,35	7,43	0,00	0,00	8,37	2,36	4,25	19,23

$$\text{Контроль: } \sum (\Delta x)^2 + \sum (\Delta y)^2 + 2 \cdot \sum \Delta x \Delta y = \sum (\Delta x + \Delta y)^2 \rightarrow 309,74 + 87,25 + 2 \cdot 157,34 = 711,67 \rightarrow 711,67 = 711,67.$$

По данным табл. 5.2 рассчитываются средние значения и среднеквадратические отклонения исследуемых рядов, а также коэффициент корреляции – R .

$$\bar{x} = \frac{\sum x_i}{n} = \frac{456,8}{37} = 12,35; \quad \bar{y} = \frac{\sum y_i}{n} = \frac{274,75}{37} = 7,43;$$

$$\sigma_x = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{\sum (\Delta x)^2}{n-1}} = \sqrt{\frac{309,7}{36}} = 2,93;$$

$$\sigma_y = \sqrt{\frac{\sum (y_i - \bar{y})^2}{n-1}} = \sqrt{\frac{\sum (\Delta y)^2}{n-1}} = \sqrt{\frac{87,25}{36}} = 1,56.$$

Коэффициент корреляции R представляет собой эмпирическую меру линейной зависимости между переменными Y и X и рассчитывается по формуле

$$R = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}} = \frac{\sum (\Delta x \Delta y)}{\sqrt{\sum (\Delta x)^2 \sum (\Delta y)^2}}. \quad (5.1)$$

Выражение (5.1) можно также представить в виде:

$$R = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{(n-1)\sigma_x \sigma_y}. \quad (5.2)$$

В этом случае для рассматриваемого примера получаем

$$R = \frac{157,34}{36 \cdot 2,93 \cdot 1,56} = 0,96.$$

С использованием полученных числовых характеристик рассчитываются параметры уравнения линейной регрессии

$$y = ax + b. \quad (5.3)$$

Параметрами этого уравнения являются коэффициент регрессии – a и свободный член – b :

$$a = R \frac{\sigma_y}{\sigma_x}, \quad (5.4)$$

$$b = \bar{y} - a\bar{x}. \quad (5.5)$$

Для рассматриваемого примера получаем:

$$a = 0,96 (1,56/2,93) = 0,51; \quad b = 7,43 - 0,51 \cdot 12,3 = 1,16.$$

Таким образом, расчетное уравнение принимает вид

$$y = 0,51x + 1,16.$$

Прямая линия, соответствующая этому уравнению, наносится на график связи $y = f(x)$ (см. рис.5.1).

5.3. ОЦЕНКА ТОЧНОСТИ УРАВНЕНИЯ ЛИНЕЙНОЙ РЕГРЕССИИ

Полученное на основе эмпирических данных уравнение линейной регрессии может использоваться в качестве прогностической зависимости или расчетной формулы, однако это является корректным только в случае, если данное уравнение обеспечивает необходимую точность расчета или, проще говоря, является надежным.

В гидрологической практике уравнение линейной регрессии считается надежным, если выполняются следующие условия:

$$1) n \geq 10; \quad 2) |R| \geq 0,7; \quad 3) \frac{R}{\sigma_R} \geq 2; \quad 4) \frac{a}{\sigma_a} \geq 2, \quad (5.6)$$

где n – длина используемых рядов; R – коэффициент парной корреляции; a – коэффициент регрессии; σ_R и σ_a – абсолютные средне-квадратические ошибки соответственно коэффициента корреляции и коэффициента регрессии, которые рассчитываются по формулам:

$$\sigma_r = \frac{1 - R^2}{\sqrt{n - 1}}, \quad (5.7)$$

$$\sigma_a = \frac{\sigma_y}{\sigma_x} \sqrt{\frac{1 - R^2}{n - 2}}. \quad (5.8)$$

Для рассматриваемого примера условия (5.6) выполняются:

$$1) n = 37 \geq 10; \quad 2) R = 0,96 \geq 0,7;$$

$$3) \sigma_R = \frac{1 - 0,96^2}{\sqrt{37 - 1}} = 0,013, \quad \frac{R}{\sigma_R} = \frac{0,96}{0,013} = 74 \geq 2;$$

$$4) \sigma_a = \frac{1,56}{2,93} \sqrt{\frac{1 - 0,96^2}{37 - 2}} = 0,025, \quad \frac{a}{\sigma_a} = \frac{0,51}{0,025} = 20 \geq 2.$$

5.4. РАСЧЕТ И ПОСТРОЕНИЕ ДОВЕРИТЕЛЬНЫХ ГРАНИЦ ДЛЯ УРАВНЕНИЯ ЛИНЕЙНОЙ РЕГРЕССИИ

При использовании уравнения линейной регрессии произвольно взятому значению x_k соответствует значение $y_k = ax_k + b$. При этом в силу выборочности исходных данных положение линии регрессии определяется с некоторой ошибкой и истинное значение $y_{k,0}$ представляет собой случайную величину с МО, равным y_k , и стандартным отклонением

$$\sigma_{\text{перп}} = \sigma_{y(x)} \sqrt{\frac{1}{n} + \frac{(x_k - \bar{x})^2}{\sum (x_i - \bar{x})^2}} = \sigma_{y(x)} \sqrt{\frac{1}{n} + \frac{(x_k - \bar{x})^2}{\sigma_x^2(n-1)}}, \quad (5.9)$$

где σ_x – среднеквадратическое отклонение ряда X ; $\bar{x}$ – среднее значение ряда X ; n – длина рядов; $\sigma_{y(x)}$ – стандартная ошибка уравнения линейной регрессии, определяемая по формуле

$$\sigma_{y(x)} = \sigma_y \sqrt{1 - R^2}, \quad (5.10)$$

где σ_y – среднеквадратическое отклонение ряда Y .

С учетом выражений (5.9)-(5.10) 95%-ные доверительные границы для уравнения линейной регрессии будут иметь вид:

$$y_{k,0} = y_k \pm t_{97,5} \sigma_{\text{перп}}, \quad (5.11)$$

где $t_{97,5}$ – квантиль стандартного нормального распределения, соответствующий вероятности непревышения 97,5 %, $t_{97,5} = 1,96$.

В рассматриваемом примере получаем:

$$\begin{aligned} \sigma_{y(x)} &= \sigma_y \sqrt{1 - R^2} = 1,56 \sqrt{1 - 0,96^2} = 0,44, \\ \sigma_{\text{перп}} &= \sigma_{y(x)} \sqrt{\frac{1}{n} + \frac{(x_k - \bar{x})^2}{\sigma_x^2(n-1)}} = 0,44 \sqrt{\frac{1}{37} + \frac{(x_k - 12,35)^2}{2,93^2(37-1)}} \end{aligned} \quad (5.12)$$

Для построения доверительных границ заполняется вспомогательная табл. 5.3.

Таблица 5.3

Расчет 95%-ных доверительных границ для линии регрессии

x_k	$(x_k - \bar{x})^2$	$\sigma_{\text{перп}}$	y_k	$y_k + t_{97,5} \sigma_{\text{перп}}$	$y_k - t_{97,5} \sigma_{\text{перп}}$
0	152,5	0,317	1,16	1,78	0,54

5	54,0	0,198	3,71	4,10	3,32
10	5,5	0,093	6,26	6,44	6,08
15	7,0	0,10	8,81	9,00	8,62
20	58,5	0,205	11,36	11,76	10,96
25	160,0	0,325	13,91	14,55	13,27

По данным табл. 5.3 на график связи наносятся линия уравнения регрессии $y_k = f(x_k)$ и 95%-ные доверительные границы (рис. 5.2).

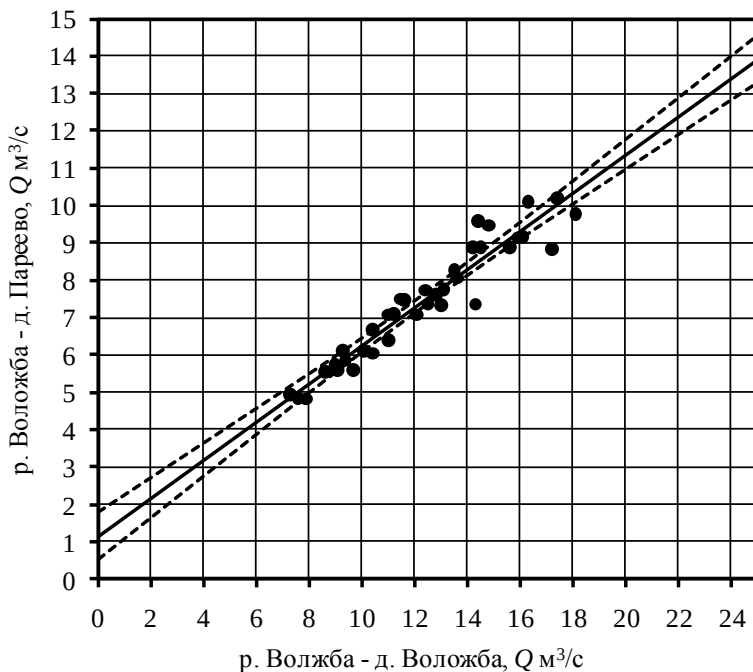


Рис. 5.2. График связи среднегодовых расходов воды р. Воложба – д. Воложба и р. Воложба – д. Пареево.

Линия регрессии и 95%-ные доверительные границы (пунктирные линии).

ВЫВОДЫ

В выводах следует охарактеризовать разброс точек на графике, оценить надежность полученного уравнения регрессии с учетом неравенств (5.6).

ЛАБОРАТОРНАЯ РАБОТА № 6

Проверка гидрологического ряда на случайность

Требуется выполнить проверку ряда на случайность с использованием критериев «общего числа серий», «наибольшей длины серии», «повышений-понижений», «экстремумов» и «значимости коэффициента автокорреляции». Принять уровень значимости равным 10%.

ИСХОДНЫЕ ДАННЫЕ

В качестве исходных данных используются минимальные суточные зимние или летние расходы воды за период 40 лет по одной реке (тот же ряд, что и в работах № 1, № 2, № 3, № 4).

ТЕОРЕТИЧЕСКАЯ ЧАСТЬ

Для выполнения работы студенты должны изучить теоретический материал по учебному пособию [1]. Необходимо проработать разделы 4.8 и 7.2 указанного пособия.

После изучения теоретического материала студенты должны знать ответы на следующие вопросы:

1. С какой целью проводится проверка гидрологических рядов на случайность?
2. Какие критерии используются для проверки гидрологических рядов на случайность?

РАСЧЕТНАЯ ЧАСТЬ

Многие приемы обработки эмпирических данных, строго говоря, пригодны лишь для статистических совокупностей, вероятностная структура которых может быть описана в рамках математической модели *случайной величины*. Для проверки соответствует ли этому требованию конкретная выборка служат критерии случайности.

Проверка гидрологических рядов на случайность проводится в рамках общей схемы статистической проверки гипотез. В качестве нулевой принимается гипотеза о том, что имеющаяся выборка представляет собой последовательность независимых значений некоторой случайной величины.

Применение критериев случайности в общем случае основано на сопоставлении конкретных статистик эмпирического ряда с соответствующими теоретическими статистиками случайных совокупностей.

6.1. КРИТЕРИЙ «ОБЩЕГО ЧИСЛА СЕРИЙ»

Под *серией* в данном случае будем понимать всякий участок последовательности, состоящий из элементов одного и того же рода. К серии из элементов «А» будем относить члены последовательности, значения которых превышают выборочное среднее значение, а к серии из элементов «В» – значения которых меньше среднего. Длиной серии называется число элементов, входящих в серию. Обработанная таким образом выборка будет представлять собой последовательность серий различной продолжительности.

Для обработки ряда на предмет выделения серий заполняется табл. 6.1.

По данным табл. 6.1. рассчитывается статистика R^* , представляющая собой сумму серий из элементов «А» и элементов «В» (длина серий i при этом значения не имеет):

$$R^* = R_A + R_B. \quad (6.1)$$

Как видно из табл. 6.1, для р. Воложба – д. Воложба $R_A = 9$ и $R_B = 9$, следовательно: $R^* = 9 + 9 = 18$.

Теоретически доказано, что в рамках модели случайной величины выборочные значения статистики R подчиняются нормальному распределению с параметрами:

$$m_R = \frac{n+1}{2}, \quad (6.2)$$

$$\sigma_R = \frac{\sqrt{n-1}}{2}. \quad (6.3)$$

Таблица 6.1

Расчет наибольшей длины и общего числа серий для ряда среднегодовых расходов воды р. Воложба – д. Воложба

Год	Q, м³/с	Индекс	Длина серии	Год	Q, м³/с	Индекс	Длина серии
1936	9,61	B	6	1963	9,70	B	3
1937	6,89	B		1964	10,1	B	
1938	8,66	B		1965	10,4	B	
1939	7,37	B		1966	17,2	A	4
1940	8,48	B		1967	12,1	A	
1941	8,16	B		1968	14,3	A	
1942	12,4	A	2	1969	15,6	A	6
1943	11,8	A		1970	9,10	B	
1944	6,94	B	2	1971	9,37	B	
1945	10,7	B		1972	7,58	B	
1946	11,8	A	1	1973	7,27	B	
1947	8,87	B	5	1974	10,4	B	
1948	10,4	B		1975	8,62	B	
1949	10,8	B		1976	12,5	A	3
1950	11,2	B		1977	13,0	A	
1951	9,74	B		1978	13,5	A	
1952	16,1	A	2	1979	8,99	B	2
1953	18,1	A		1980	9,30	B	
1954	11,0	B	1	1981	12,4	A	4
1955	15,9	A	4	1982	14,8	A	
1956	13,1	A		1983	14,2	A	
	17,4	A		1984	14,5	A	
1958	14,4	A		1985	11,2	B	1
1959	11,0	B	3	1986	13,6	A	3
1960	7,86	B		1987	12,8	A	
1961	11,5	B		1988	11,6	A	
1962	16,3	A	1				

В этом случае 90%-ный доверительный интервал для статистики R имеет вид:

$$t_{5\%} \leq \frac{R - m_R}{\sigma_R} < t_{95\%}, \quad (6.4)$$

где $t_{5\%}$ и $t_{95\%}$ – квантили нормированного нормального распределения, определяемые по табл.2.1 учебника [1].

Учитывая, что для нормированного нормального распределения $t_{5\%} = -t_{95\%}$ в окончательном виде имеем:

$$m_R - t_{95\%}\sigma_R \leq R < m_R + t_{95\%}\sigma_R. \quad (6.5)$$

Для р. Воложба – д. Воложба получаем:

$$m_R = \frac{53+1}{2} = 27; \quad \sigma_R = \frac{\sqrt{53-1}}{2} = 3,61;$$

$$(27 - 1,64 \cdot 3,61) \leq R < (27 + 1,64 \cdot 3,61);$$

$$21 \leq R < 33.$$

В данном случае эмпирическое значение статистики R^* выходит за пределы доверительного интервала ($R^* = 18$). Следовательно, гипотеза о случайности ряда по критерию «общего числа серий» опровергается при уровне значимости $2\alpha = 10\%$.

Малое число серий указывает на то, что колебания водности на рассматриваемой реке носят более плавный характер, по сравнению с флуктуациями значений случайной величины.

6.2. КРИТЕРИЙ «НАИБОЛЬШЕЙ ДЛИНЫ СЕРИИ»

Критерий наибольшей длины серий использует в качестве тестовой статистики наибольшую длину серии из элементов "А" или "В": $K = i_{max}$. Теоретически доказано, что в рамках модели случайной величины критическое значение статистики K выражается формулой

$$K_{\alpha} = \frac{\lg \left[-\frac{n}{\ln(1-\alpha)} \right]}{\lg 2} - 1, \quad (6.6)$$

где α – односторонний уровень значимости в долях единицы.

Для р. Воложба – д. Воложба при $\alpha = 0.1$ получаем:

$$K_{0.05} = \frac{\lg \left[-\frac{53}{\ln(1-0.1)} \right]}{\lg 2} - 1 = 8. \quad (6.7)$$

По табл. 6.1 находим максимальную длину серии: $K^* = 6$.

Так как в данном случае $(K^* = 6) < (K_{0.1} = 8)$, то гипотеза о случайности ряда по критерию «наибольшей длины серий» не опровергается при уровне значимости $\alpha = 10\%$.

6.3. КРИТЕРИЙ «ПОВЫШЕНИЙ – ПОНИЖЕНИЙ»

Если для ряда расходов воды $Q_{i-1} < Q_i$ будем называть переход от Q_{i-1} к Q_i повышением и обозначать его знаком «+», если $Q_{i-1} > Q_i$ будем называть переход от Q_{i-1} к Q_i понижением и обозначать знаком «-».

Для подсчета числа повышений и понижений заполняется табл. 6.2.

Теоретически доказано, что в рамках модели случайной величины число повышений (или понижений) N распределяется асимптотически нормально с параметрами:

$$m_+ = m_- = \frac{n}{2}, \quad (6.8)$$

$$\sigma_+ = \sigma_- = \sqrt{\frac{n+1}{12}}. \quad (6.9)$$

В этом случае 90%-ный доверительный интервал для статистики N имеет вид:

$$m - t_{95}\sigma \leq N < m + t_{95}\sigma, \quad (6.10)$$

где $t_{95\%}$ – квантиль нормированного нормального распределения, определяемый по табл. 2.1 учебника [1].

Рассмотрим пример расчета по р. Воложба – д. Воложба:

1. По табл. 6.2 определяем число повышений $N_+ = 29$.
2. По табл. 6.2 определяем число понижений $N_- = 23$.
3. По формулам (6.8) и (6.9) рассчитываем математическое ожидание и среднеквадратическое отклонение статистики N :

$$m_+ = m_- = 53/2 = 26,5, \quad \sigma_+ = \sigma_- = \sqrt{(53+1)/12} = 2,12.$$

Таблица 6.2

Расчет числа повышений «+», понижений «-» и экстремумов «э» для ряда среднегодовых расходов воды р. Воложба – д. Воложба, $\bar{Q} = 11,5 \text{ м}^3/\text{с}$

Год	$Q, \text{ м}^3/\text{с}$	+/-	Э	Год	$Q, \text{ м}^3/\text{с}$	+/-	Э
1936	9,61			1963	9,70	–	э
1937	6,89	–	э	1964	10,1	+	
1938	8,66	+	э	1965	10,4	+	
1939	7,37	–	э	1966	17,2	+	э
1940	8,48	+	э	1967	12,1	–	э
1941	8,16	–	э	1968	14,3	+	
1942	12,4	+	э	1969	15,6	+	э
1943	11,8	–		1970	9,10	–	э
1944	6,94	–	э	1971	9,37	+	э
1945	10,7	+		1972	7,58	–	
1946	11,8	+	э	1973	7,27	–	э
1947	8,87	–	э	1974	10,4	+	э
1948	10,4	+		1975	8,62	–	э
1949	10,8	+		1976	12,5	+	
1950	11,2	+	э	1977	13,0	+	
1951	9,74	–	э	1978	13,5	+	э
1952	16,1	+		1979	8,99	–	э
1953	18,1	+	э	1980	9,30	+	
1954	11,0	–	э	1981	12,4	+	
1955	15,9	+	э	1982	14,8	+	э
1956	13,1	–	э	1983	14,2	–	э
1957	17,4	+	э	1984	14,5	+	э
1958	14,4	–		1985	11,2	–	э
1959	11,0	–		1986	13,6	+	э
1960	7,86	–	э	1987	12,8	–	
1961	11,5	+		1988	11,6	–	
1962	16,3	+	э				

4. Используя формулу (6.10), строим 90%-ный доверительный интервал для статистики N :

$$26,5 - 1,64 \cdot 2,12 \leq N < 26,5 + 1,64 \cdot 2,12,$$

$$23 \leq N < 30.$$

Так как в данном случае N_+ и N_- не выходят за пределы доверительного интервала, то гипотеза о случайности ряда по критерию "повышений-понижений" не опровергается при уровне значимости $2\alpha = 10\%$.

6.4. КРИТЕРИЙ «ЭКСТРЕМУМОВ»

Экстремум – элемент последовательности $Q_1, Q_2, Q_3 \dots, Q_n$, для которого выполняется одно из неравенств:

$Q_{i-1} < Q_i > Q_{i+1}$ – максимум, $Q_{i-1} > Q_i < Q_{i+1}$ – минимум.

Подсчет числа экстремумов производим в табл. 6.2.

Теоретически доказано, что в рамках модели случайной величины число экстремумов $N_{\text{э}}$ в выборке распределено асимптотически нормально с параметрами:

$$m_{\text{э}} = \frac{2n}{3}, \quad (6.11)$$

$$\sigma = \sqrt{\frac{16n - 29}{90}}. \quad (6.12)$$

В этом случае 90%-ный доверительный интервал для статистики $N_{\text{э}}$ имеет вид:

$$m_{\text{э}} - t_{95}\sigma \leq N_{\text{э}} < m_{\text{э}} + t_{95}\sigma, \quad (6.13)$$

где $t_{95\%}$ – квантиль нормированного нормального распределения, определяемый по табл. 2.1 учебника [1].

Рассмотрим пример расчета по р. Воложба – д. Воложба:

1. По табл. 6.2 определяем число экстремумов $N_{\text{э}}^* = 34$.

2. По формулам (6.11) и (6.12) рассчитываем математическое ожидание и среднеквадратическое отклонение статистики $N_{\mathcal{I}}$:

$$m_{\mathcal{I}} = \frac{2 \cdot 53}{3} = 35,3,$$

$$\sigma = \sqrt{\frac{16 \cdot 53 - 29}{90}} = 3,02.$$

3. Используя формулу (6.13), строим 90%-ный доверительный интервал для статистики $N_{\mathcal{I}}$:

$$35,3 - 1,64 \cdot 3,02 \leq N_{\mathcal{I}} < 35,3 + 1,64 \cdot 3,02,$$

$$30 \leq N_{\mathcal{I}} < 40.$$

Так как в данном случае $N_{\mathcal{I}}^*$ не выходит за пределы доверительного интервала ($N_{\mathcal{I}}^* = 34$), то гипотеза о случайности ряда по критерию «экстремумов» не опровергается при уровне значимости $2\alpha = 10\%$.

6.5. КРИТЕРИЙ «ЗНАЧИМОСТИ КОЭФФИЦИЕНТА АВТОКОРРЕЛЯЦИИ»

При использовании этого критерия ряд сдвигается (вверх) относительно самого себя на величину $\tau = 1$. Исходный ряд обозначим x_i , ряд, полученный при сдвиге, обозначим x_{i+1} (табл. 6.3). Совместный период для полученных рядов равен $(n - 1)$, так как у ряда x_i теряется последнее значение, а у ряда x_{i+1} – первое.

Для полученных таким образом рядов рассчитываем коэффициент парной корреляции. Этот коэффициент называется коэффициентом автокорреляции и обозначается $r(1)$. Он характеризует тесноту связи между смежными членами ряда:

$$r(1) = \frac{\sum_{i=1}^{i=n-1} (x_i - \bar{x})(x_{i+1} - \bar{x})}{(n-2)D_x}, \quad (6.14)$$

где $\bar{x}$ и D_x – соответственно среднее значение и дисперсия исходного ряда.

Таблица 6.3

Таблица для расчета коэффициента автокорреляции ряда среднегодовых расходов воды р. Воложба – д. Воложба; $\bar{x} = 11,5 \text{ м}^3/\text{с}$, $D_x = 8,47$

№ п/п	x_i	x_{i+1}	$x_i - \bar{x}$	$x_{i+1} - \bar{x}$	$(x_i - \bar{x})(x_{i+1} - \bar{x})$
1	9,61	9,61 6,89	-1,89	-4,61	8,713
2	6,89	8,66	-4,61	-2,84	13,092
3	8,66	7,37	-2,84	-4,13	11,729
...					
50	11,2	13,6	-0,3	2,1	-0,630
51	13,6	12,8	2,1	1,3	2,730
52	12,8	11,6	1,3	0,1	0,130
	11,6				
Сумма					164,94

Для полученных таким образом рядов рассчитываем коэффициент парной корреляции. Этот коэффициент называется коэффициентом автокорреляции и обозначается $r(1)$. Он характеризует тесноту связи между смежными членами ряда:

$$r(1) = \frac{\sum_{i=1}^{i=n-1} (x_i - \bar{x})(x_{i+1} - \bar{x})}{(n-2)D_x}, \quad (6.14)$$

где $\bar{x}$ и D_x – соответственно среднее значение и дисперсия исходного ряда.

Для реки Воложба – д. Воложба $\bar{Q} = 11,5 \text{ м}^3/\text{с}$, $D_x = 8,47$, следовательно, с учетом данных табл. 6.3 получаем

$$r(1) = \frac{164,94}{(53-2)8,47} = 0,38.$$

Если ряд соответствует модели случайной величины, то коэффициент автокорреляции должен быть близок к нулю, т.е. он должен быть статистически незначимым; 90%-ный доверительный интервал при нулевой гипотезе $H_0: r(1) = 0$ имеет вид:

$$|r(1)| \leq \sigma_r t_{95\%}, \quad (6.15)$$

где $t_{95\%}$ – квантиль нормированного нормального распределения, определяемый по табл. 2.1 учебника [1]; σ_r – среднеквадратическая погрешность коэффициента автокорреляции, которая определяется по формуле

$$\sigma_r = \frac{1-r(1)^2}{\sqrt{n-2}}. \quad (6.16)$$

Если коэффициент автокорреляции выходит за пределы доверительного интервала, то он является статистически значимым и гипотеза о случайности ряда опровергается.

Проведем оценку значимости коэффициента автокорреляции для р. Воложба – д. Воложба:

$$\sigma_r = \frac{1-0,38^2}{\sqrt{53-2}} = 0,12; \quad t_{95\%} = 1,64; \quad \sigma_r t_{95\%} = 0,20.$$

Так как в данном случае $0,38 > 0,20$, то есть $|r(1)| > \sigma_r t_{95\%}$, то гипотеза о случайности ряда опровергается при уровне значимости $2\alpha = 10\%$.

Данный критерий наиболее часто используется в практике гидрологических расчетов.

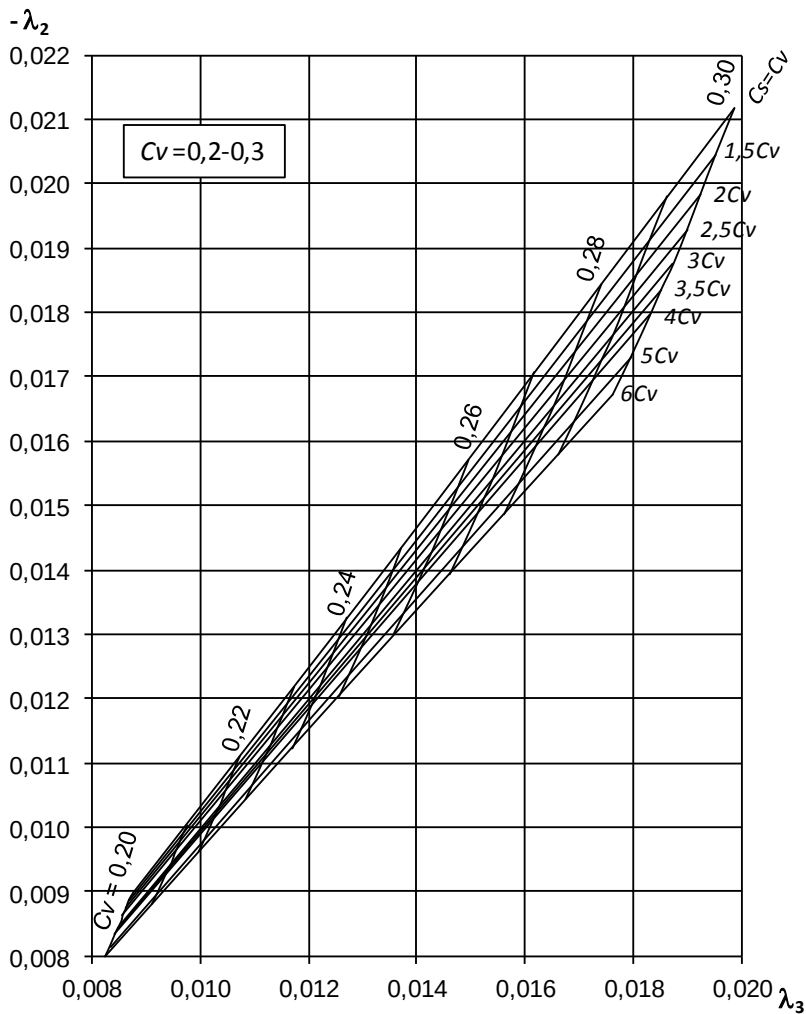
ВЫВОДЫ

В выводах следует перечислить, по каким критериям гипотеза о случайности ряда опровергается, а по каким – нет.

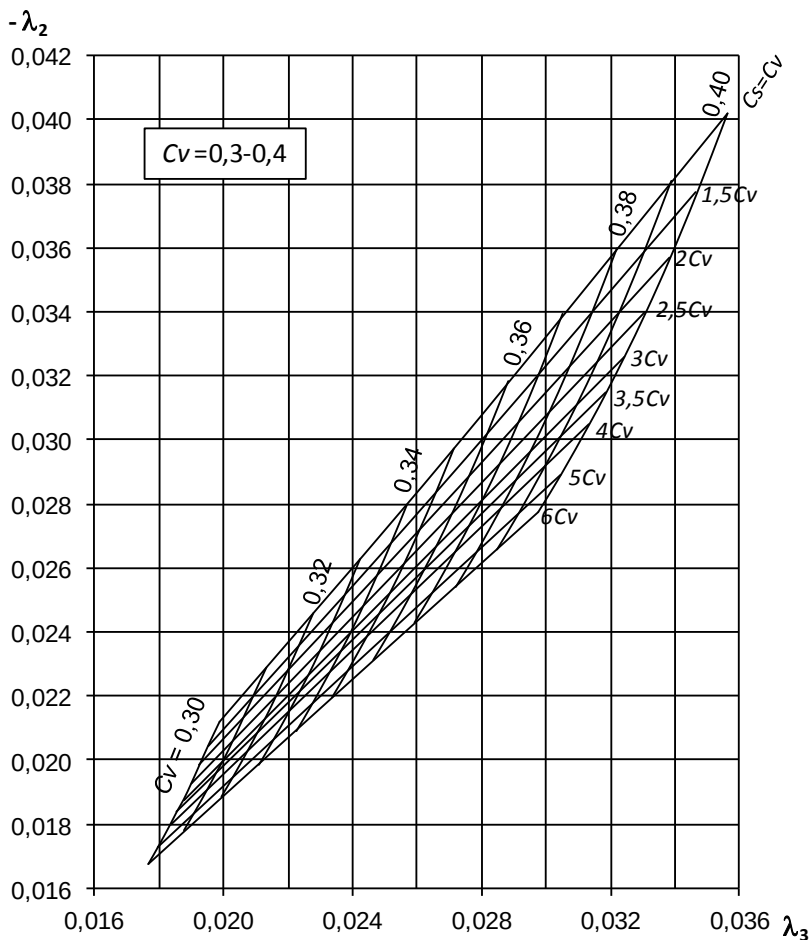
В случае опровержения гипотезы о случайности по какому-то из критериев нужно попытаться проанализировать, какие свойства

исследуемого ряда не характерны для выборок из случайных совокупностей.

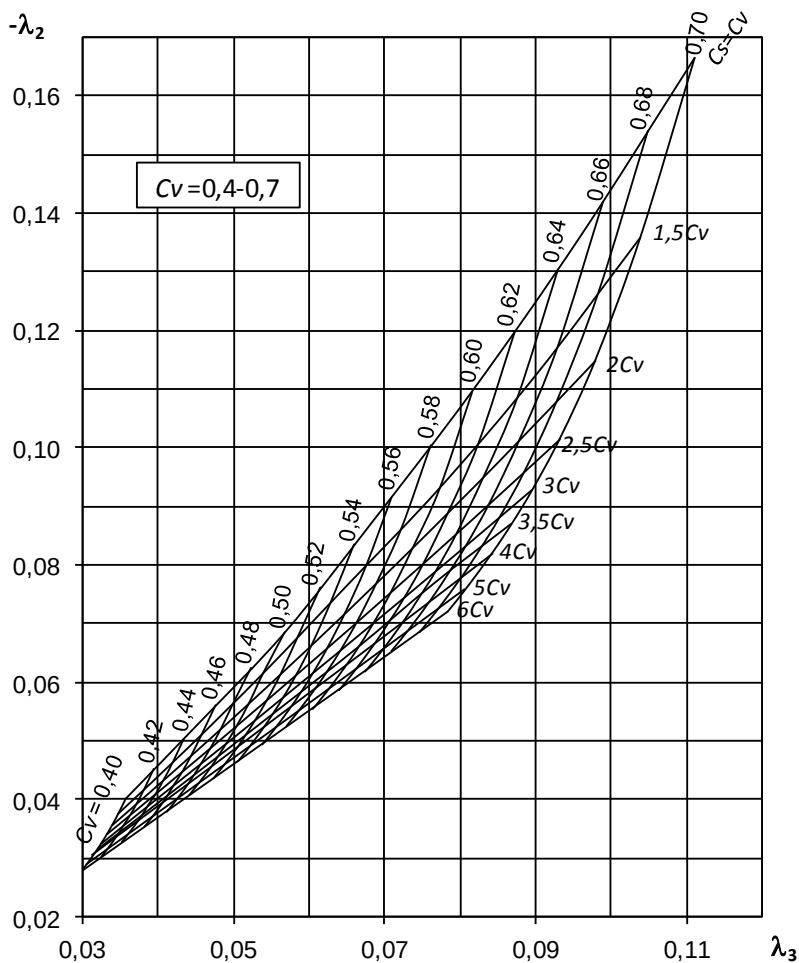
Номограмма для определения параметров распределения Крицкого – Менкеля C_v и C_s методом наибольшего правдоподобия при $C_v = 0,20-0,30$



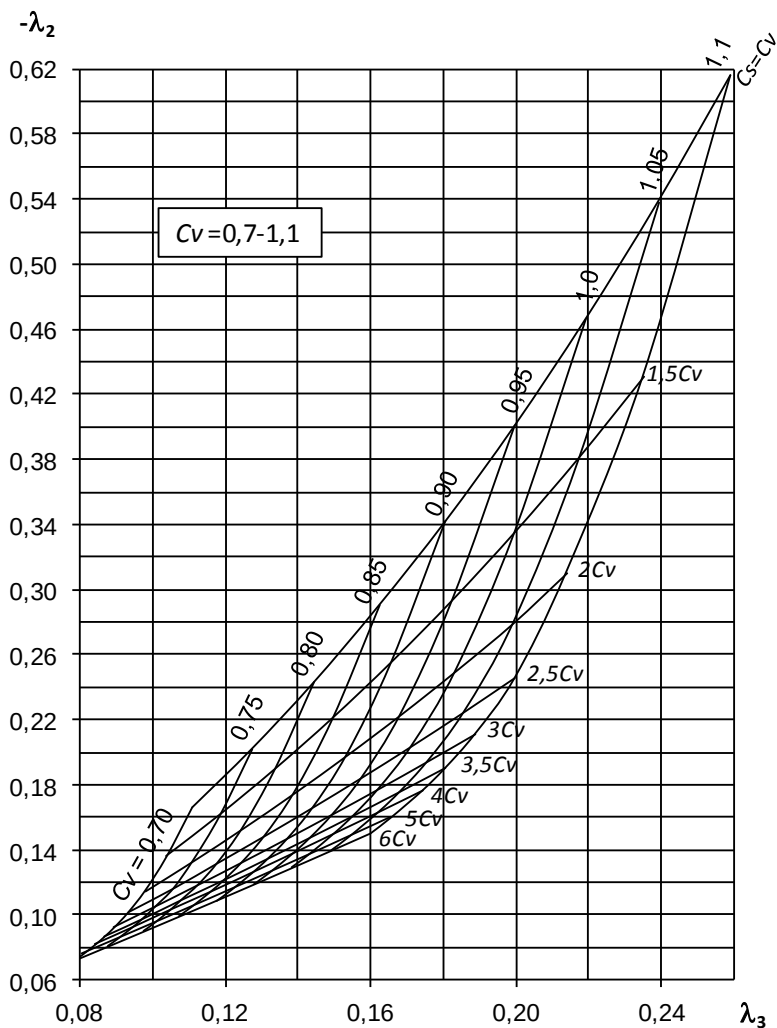
Номограмма для определения параметров распределения Крицкого – Менкеля C_v и C_s методом наибольшего правдоподобия при $C_v = 0,30-0,40$



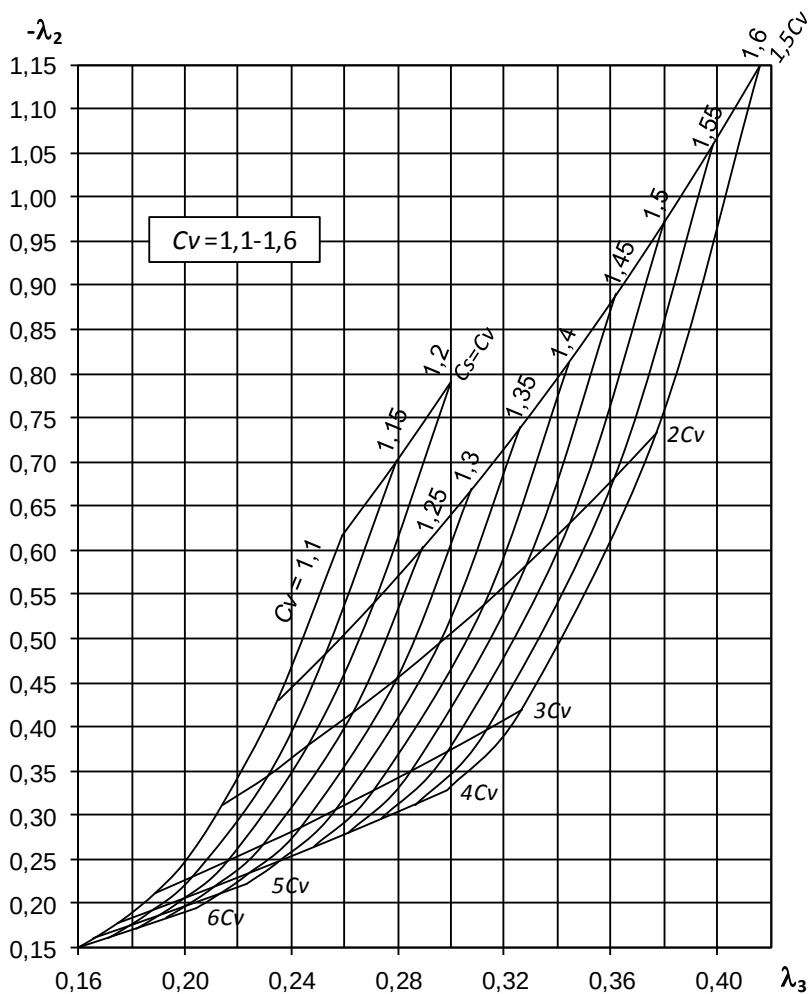
Номограмма для определения параметров распределения Крицкого – Менкеля C_v и C_s методом наибольшего правдоподобия при $C_v = 0,40-0,70$



Номограмма для определения параметров распределения Крицкого – Мен-
келя C_v и C_s методом наибольшего правдоподобия при $C_v = 0,70-1,1$



Номограмма для определения параметров распределения Крицкого – Менкеля C_v и C_s методом наибольшего правдоподобия при $C_v = 1,1-1,6$



ЛИТЕРАТУРА

Основная

1. Сикан А.В. Методы статистической обработки гидрометеорологической информации. – СПб., изд. РГГМУ, 2007.–279 с.

Дополнительная

2. Методические рекомендации по оценке однородности гидрологических характеристик и определению их расчетных значений по неоднородным данным. – СПб: Нестор-История, 2010.–162 с.
3. Методические рекомендации по определению расчетных гидрологических характеристик при наличии данных гидрометрических наблюдений. – Нижний Новгород: Вектор-ТиС, 2007. – 134 с.
4. Пособие по определению расчетных гидрологических характеристик. – Л.: Гидрометеоиздат, 1984. – 444 с.
5. Рождественский А.В., Чеботарев А.И. Статистические методы в гидрологии. – Л.: Гидрометеоиздат, 1974. – 424 с.
6. Свод правил СП 33-101-2003. Определение основных расчетных гидрологических характеристик. – М.: Стройиздат, 2004. – 72 с.

СОДЕРЖАНИЕ

Стр

ПРЕДИСЛОВИЕ.....

ЛАБОРАТОРНАЯ РАБОТА № 1

***Исследование числовых характеристик и эмпирических функций
распределения ряда минимальных суточных зимних (летних)
расходов воды реки _____ в створе _____***.....

- 1.1 Построение гистограммы эмпирических частот и эмпирических функций распределения
- 1.1.1 Определение размаха R
- 1.1.2 Определение длины расчетного интервала L
- 1.1.3 Построение гистограммы эмпирических частот
- 1.1.4 Построение интегральной функции распределения и функции обеспеченностей
- 1.2 Расчет числовых характеристик и их погрешностей
- 1.2.1 Расчет характеристик положения (моды, медианы, математического ожидания).....
- 1.2.2 Расчет характеристик рассеивания
- 1.2.3 Сводная таблица основных статистических характеристик гидрологического ряда
- 1.2.4 Расчет абсолютных и относительных погрешностей среднего значения, коэффициента вариации и коэффициента асимметрии
- Выводы

ЛАБОРАТОРНАЯ РАБОТА № 2

***Исследование законов распределения и расчет расходов воды
различной обеспеченности***.....

- 2.1 Построение эмпирической кривой обеспеченностей
- 2.2 Построение аналитической кривой обеспеченностей нормального закона распределения
- 2.3 Построение аналитической кривой обеспеченностей логарифмически нормального закона распределения.....
- 2.4 Построение аналитической кривой обеспеченностей Гумбеля.....
- 2.5 Построение аналитической кривой обеспеченностей Пирсона III типа.....
- 2.6 Построение аналитической кривой обеспеченностей Крицкого – Менкеля
- Выводы

ЛАБОРАТОРНАЯ РАБОТА № 3

<i>Методы расчета оценок параметров распределения</i>	
3.1 Метод моментов	
3.2 Метод наибольшего правдоподобия	
3.2.1 Расчет статистик λ_1 , λ_2 и λ_3	
3.2.2 Определение статистических характеристик ряда.....	
3.2.3 Оценка относительных погрешностей расчета статистических характеристик.....	
3.3 Графо-аналитический метод (метод квантилей).....	
3.3.1 Построение сглаженной эмпирической кривой обеспеченностей и определение трех опорных ординат.....	
3.3.2 Расчет параметров распределения.....	
3.3.3 Анализ надежности расчета.....	
Выводы	

ЛАБОРАТОРНАЯ РАБОТА № 4

<i>Интервальные оценки параметров распределения и проверка гидрологических рядов на однородность</i>	
4.1 Интервальные оценки параметров распределения	
4.1.1 Интервальная оценка математического ожидания	
4.1.2 Интервальная оценка дисперсии	
4.1.3 Интервальная оценка среднеквадратического отклонения	
4.2 Проверка однородности гидрологического ряда с использованием параметрических критериев Фишера и Стьюдента	
4.2.1 Проверка однородности гидрологического ряда по дисперсии (критерий Фишера)	
4.2.2 Проверка однородности гидрологического ряда по среднему значению (критерий Стьюдента)	
4.3 Проверка однородности гидрологического ряда с использованием непараметрических критериев Диксона и Смирнова – Граббса.....	
4.3.1 Проверка однородности гидрологического ряда по критерию Диксона.....	
4.3.2 Проверка однородности гидрологического ряда по критерию Смирнова – Граббса.....	
Выводы	

ЛАБОРАТОРНАЯ РАБОТА № 5

<i>Анализ взаимосвязи двух переменных (парная корреляция)</i>	
5.1 Построение графика связи $y = f(x)$	
5.2 Расчет параметров уравнения линейной регрессии.....	

5.3	Оценка точности уравнения линейной регрессии.....
5.4	Расчет и построение доверительных границ для уравнения линейной регрессии.....
	Выводы

ЛАБОРАТОРНАЯ РАБОТА № 6

	<i>Проверка гидрологического ряда на случайность</i>
6.1	Критерий «общего числа серий»
6.2	Критерий «наибольшей длины серии»
6.3	Критерий «повышений – понижений»
6.4	Критерий «экстремумов»
6.5	Критерий «значимости коэффициента автокорреляции».....
	Выводы
	ПРИЛОЖЕНИЯ
	ЛИТЕРАТУРА

Учебное издание

ПРАКТИКУМ
по дисциплине

«Статистические методы анализа гидрометеорологической информации»

Составители: Сикан Александр Владимирович
Малышева Наталья Геннадьевна
Винокуров Игорь Олегович

Редактор: Крайнова О.С.

Компьютерная верстка:

ЛР № 020309 от 30.19.96.

Подписано в печать	Формат $60 \times 90^{1/16}$	Бумага кн.-жур.	Печать офсетная.
Печ. л.	Уч.-изд. л.	Тираж	Зак.

195196, СПб, Малоохтинский пр. 98. РГТМУ.

Отпечатано