

**В. С. Дружинин, А. В. Сикан**

## **МЕТОДЫ СТАТИСТИЧЕСКОЙ ОБРАБОТКИ ГИДРОМЕТЕОРОЛОГИЧЕСКОЙ ИНФОРМАЦИИ**

*Рекомендовано Учебно-методическим объединением по образованию  
в области гидрометеорологии в качестве учебного пособия для студентов*

*высших учебных заведений, обучающихся по направлению*

*"Гидрометеорология" и специальности "Гидрология"*

**Под редакцией проф. А.М. Владимирова**



Санкт-Петербург  
2001

УДК 519.24: 556.5

Дружинин В.С., Сикан А.В. Методы статистической обработки гидрометеорологической информации. Учебное пособие. Направление "Гидрометеорология". Специальность "Гидрология". – СПб.: изд. РГГМУ. 2001. – с.

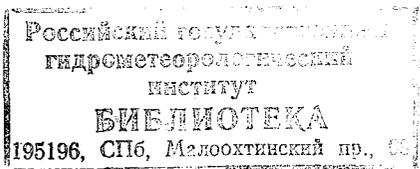
Рецензенты: д-р физ.-мат. наук С. А. Кондратьев (Институт озераедения РАН), д-р техн. наук, проф. А. В. Рождественский (Государственный гидрологический институт), канд. геогр. наук, доц. В. Л. Трушевский (Санкт-Петербургский государственный университет)

Излагаются методы статистической обработки гидрометеорологической информации. Рассматриваются кривые распределения и клетчатки вероятностей, применяемые в гидрологии, методы оценки однородности и случайности гидрологических рядов, техника расчета основных статистических характеристик, практические приемы использования аппарата парной и множественной линейной корреляции. В книге содержится большое количество примеров, позволяющих закрепить излагаемый теоретический материал.

Предназначена для студентов гидрометеорологических специальностей и специалистов-гидрологов.

The book contains the systematic account of the methods of statistical processing of the hydrometeorological information. The distribution curves and probability checks, applied in hydrology, methods of estimating of homogeneity and the random character of hydrological time series, methods of calculation of the basic statistical characteristics, practical methods of using of the device of pair and multiple linear correlation are examined. There are a lot of examples, allowing to fix theoretical material, stated in the book.

The book is intended for the students of hydrometeorological specializations and specialists in hydrology.



ISBN 5 – 86813 – 029 – 4

© В.С. Дружинин, 2001.

© А.В. Сикан, 2001.

© Российский государственный гидрометеорологический университет (РГГМУ), 2001.

Распределение Колмогорова

Вероятности превышения  $P(\lambda) = 2 \sum_{k=1}^{\infty} (-1)^{k-1} e^{-2k^2 \lambda^2}$  даны в долях единицы.

$\lambda$	0	1	2	3	4	5	6	7	8	9
0.3				0.9999	0.9988	0.9997	0.9995	0.9992	0.9987	0.9981
0.4	0.9972	0.9960	0.9945	0.9926	0.9903	0.9874	0.9840	0.9800	0.9753	0.9700
0.5	0.9639	0.9572	0.9497	0.9415	0.9325	0.9228	0.9124	0.9013	0.8896	0.8772
0.6	0.8643	0.8508	0.8368	0.8222	0.8073	0.7920	0.7764	0.7604	0.7442	0.7278
0.7	0.7112	0.6945	0.6777	0.6609	0.6440	0.6272	0.6104	0.5936	0.5770	0.5605
0.8	0.5441	0.5280	0.5120	0.4962	0.4806	0.4653	0.4503	0.4355	0.4209	0.4067
0.9	0.3927	0.3791	0.3657	0.3527	0.3399	0.3275	0.3154	0.3036	0.2921	0.2809
1.0	0.2700	0.2594	0.2492	0.2392	0.2296	0.2202	0.2111	0.2024	0.1939	0.1857
1.1	0.1777	0.1700	0.1626	0.1555	0.1486	0.1420	0.1356	0.1294	0.1235	0.1177
1.2	0.1122	0.1070	0.1019	0.0970	0.0924	0.0879	0.0836	0.0794	0.0755	0.0717
1.3	0.0681	0.0646	0.0613	0.0582	0.0551	0.0522	0.0495	0.0469	0.0444	0.0420
1.4	0.0397	0.0375	0.0354	0.0335	0.0316	0.0298	0.0282	0.0266	0.0250	0.0236
1.5	0.0222	0.0209	0.0197	0.0185	0.0174	0.0164	0.0154	0.0145	0.0136	0.0127
1.6	0.0120	0.0112	0.0105	0.0098	0.0092	0.0086	0.0081	0.0076	0.0071	0.0066
1.7	0.0062	0.0058	0.0054	0.0050	0.0047	0.0044	0.0041	0.0038	0.0035	0.0033
1.8	0.0031	0.0029	0.0027	0.0025	0.0023	0.0021	0.0020	0.0019	0.0017	0.0016
1.9	0.0015	0.0014	0.0013	0.0012	0.0011	0.0010	0.0009	0.0009	0.0008	0.0007
2.0	0.0007	0.0006	0.0006	0.0005	0.0005	0.0004	0.0004	0.0004	0.0003	0.0003
2.1	0.0003	0.0003	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0001	0.0001
2.2	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
2.3	0.0001	0.0000	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001

Эффект ложной корреляции для зависимости  $y = f(z)$  (по Бенсону)

Перемен - ная		Кoeffициент корреляции, $r_{y,z}$ (аппроксимация)	Особые случаи			
Y	Z		усло - вие	$r_{y,z}$ (аппроксимация)	коэффициент вариации, C	$r_{y,z}$ (значение)
$x_1/x_2$	$x_2$	$\frac{r_{1,2}C_1 + C_2}{(C_1^2 + C_2^2 + 2r_{1,2}C_1C_2)^{1/2}}$	$r_{1,2} = 0$	$\frac{C_2}{(C_1^2 + C_2^2)^{1/2}}$	$C_1 = C_2$ $C_1 > C_2$	0.71 > 0.71
$x_1/x_2$	$x_3/x_2$	$\frac{r_{1,3}C_1C_3 + r_{1,2}C_1C_2 + r_{2,3}C_2C_3 + C_2}{(C_1^2 + C_2^2 + 2r_{1,2}C_1C_2)^{1/2} (C_2^2 + C_3^2 + 2r_{2,3}C_2C_3)^{1/2}}$	$r_{1,2} = 0$ $r_{1,3} = 0$ $r_{2,3} = 0$	$\frac{C_2^2}{(C_1^2 + C_2^2)^{1/2} (C_2^2 + C_3^2)^{1/2}}$	$C_2 > C_1C_3$ $C_2 = 2C_1 = 2C_3$ $C_2 = 2C_1 = 3C_3$ $C_2 = 3C_1 = 3C_3$	> 0.50 0.80 0.85 0.90
$x_1/x_2$	$x_2$	$\frac{r_{1,2}C_1 + C_2}{(C_1^2 + C_2^2 - 2r_{1,2}C_1C_2)^{1/2}}$	$r_{1,2} = 0$	$\frac{-C_2}{(C_1^2 + C_2^2)^{1/2}}$	$C_2 > C_1$ $C_2 = 2C_1$ $C_2 = 3C_1$	< -0.70 - 0.89 - 0.95
$x_1/x_2$	$x_3/x_2$	$\frac{r_{1,3}C_1C_3 - r_{1,2}C_1C_2 - r_{2,3}C_2C_3 + C_2}{(C_1^2 + C_2^2 - 2r_{1,2}C_1C_2)^{1/2} (C_2^2 + C_3^2 - 2r_{2,3}C_2C_3)^{1/2}}$	$r_{1,2} = 0$ $r_{1,3} = 0$ $r_{2,3} = 0$	$\frac{C_2^2}{(C_1^2 + C_2^2)^{1/2} (C_2^2 + C_3^2)^{1/2}}$	$C_1 > C_2C_3$ $C_2 = 2C_1 = 2C_3$ $C_2 = 3C_1 = 3C_3$	> 0.50 0.80 0.90

<b>5.</b>	<b>СТАТИСТИЧЕСКИЙ АНАЛИЗ ЗАВИСИМОСТЕЙ МЕЖДУ ГИДРОЛОГИЧЕСКИМИ ПЕРЕМЕННЫМИ</b> .....	124
5.1.	Метод наименьших квадратов. Уравнение линейной регрессии для двух переменных .....	125
5.2.	Линеаризация нелинейных зависимостей .....	130
5.3.	Оценка точности уравнения линейной регрессии для двух переменных .....	132
5.4.	Интервальная оценка и оценка значимости параметров линейной регрессии для двух переменных .....	136
5.4.1.	Интервальная оценка коэффициента парной корреляции. Z-преобразование Фишера .....	136
5.4.2.	Интервальная оценка коэффициента регрессии .....	138
5.4.3.	Интервальная оценка свободного члена .....	139
5.4.4.	F - критерий значимости регрессии .....	140
5.4.5.	Построение доверительного интервала для уравнения линейной регрессии .....	141
5.5.	Множественная линейная корреляция .....	142
5.5.1.	Оценка точности уравнения множественной линейной регрессии .....	144
5.5.2.	Выбор "наилучшего" уравнения регрессии .....	145
5.5.3.	Ложная корреляция .....	147
	<b>СПИСОК ЛИТЕРАТУРЫ</b> .....	149
	<b>ПРИЛОЖЕНИЯ</b> .....	150
1.	Нормированные ординаты кривой обеспеченностей Пирсона III типа (непрерывное биномиальное распределение) .....	150
2.	Ординаты кривых обеспеченностей Крицкого – Менкеля (трехпараметрического гамма - распределения) .....	152
3.	Ординаты кривых обеспеченностей Н-распределения (А.В. Ежова) .....	158
4.	$\chi^2$ - распределение .....	160
5.	t-распределение (Стьюдента) .....	161
6.	F-распределение (Фишера) .....	162
7.	Распределение Колмогорова .....	165
8.	Эффект ложной корреляции для зависимости $y = f(z)$ (по Бенсону) .....	166
9.	Таблица для определения коэффициента вариации $C_v$ методом приближенного наибольшего правдоподобия в зависимости от статистики $\lambda_2$ при фиксированном соотношении $C_v/C_s$ (для кривой обеспеченностей Пирсона III типа) .....	167

## Table of Contents

	<b>The Foreword</b> .....	9
	<b>The List of Labels and Abridgements</b> .....	11
	<b>Introduction</b> .....	12
<b>1.</b>	<b>Some Items of Information From the Theory of Probabilities</b> ...	<b>13</b>
1.1.	Random variables and distribution functions .....	13
1.2.	Discrete and continuous random variables .....	15
1.3.	Random variable and time .....	17
1.4.	Numerical characteristics of random variables .....	18
1.4.1.	Modal value. Median. Expectation .....	18
1.4.2.	The moments of a random variable .....	20
1.4.3.	Variance. Standard deviation. Spread coefficient .....	21
1.4.4.	Skewness and kurtosis .....	22
1.4.5.	Properties of numerical characteristics of random variables .....	22
1.4.6.	Standard transformations of a random variable .....	24
1.4.7.	Quantiles of distribution .....	24
<b>2.</b>	<b>Analytical Distribution Functions Used in a Hydrology</b> .....	<b>26</b>
2.1.	Normal distribution .....	26
2.2.	The law of a uniform frequency .....	29
2.3.	Log-Normal distribution .....	30
2.4.	The law of distribution of extreme terms of a sample (distribution of Gumbel) .....	34
2.5.	Distribution of Pearson of the third type (continuous binomial distribution) .....	37
2.6.	Distribution of Kritsky – Menkel (three-parametric $I$ -distribution) .....	40
2.7.	Distribution of Johnson .....	43
2.8.	Graphic performance of distribution functions on standard probability paper .....	44
2.9.	Comments to the chapter .....	46
<b>3.</b>	<b>Construction of Cumulative Distribution Functions. Estimate of Distribution Parameters on the Empirical Data</b> .....	<b>50</b>
3.1.	Empirical cumulative distribution functions .....	52
3.2.	Methods of calculating of distribution parameters on empirical data .....	60
3.2.1.	Method of the moments .....	61

3.2.2.	Method of a maximum likelihood .....	64
3.2.3.	Method of quantiles .....	68
3.2.3.1.	Method of quantiles on the basis of a distribution of Pearson of the third type .....	68
3.2.3.2.	Method of quantiles on the basis of a three-parametric logarithmic normal distribution .....	70
3.2.3.3.	Method of quantiles on the basis of a distribution of Kritsky – Menkel .....	72
3.2.3.4.	Comments .....	75
3.2.4.	Graphic method .....	76
3.2.5.	Estimate of distribution parameters of Johnson .....	78
3.3.	Estimate of errors of empirical parameters of distribution .....	81
3.4.	Economical aspects of selection of calculated probabilities .....	83
3.4.1.	The purposes of selection of calculated probabilities .....	83
3.4.2.	Selection of optimum probabilities .....	84
<b>4.</b>	<b>Parameters Interval Estimation and Check of Statistical Hypotheses</b> .....	<b>89</b>
4.1.	Chi-squared distribution .....	89
4.2.	t – distribution (Student’s distribution) .....	90
4.3.	F – distribution (Fisher’s distribution) .....	92
4.4.	Interval estimations of distribution parameters .....	93
4.4.1.	Interval estimation of expectation .....	94
4.4.2.	Interval estimation of variance .....	95
4.5.	Statistical hypothesis. Test of a statistical hypothesis .....	96
4.6.	Tests used for check of a homogeneity of hydrological series .....	100
4.6.1.	Student’s test for check of a significance of distinction of mean values of two samples .....	100
4.6.2.	Test of equality of two variances (Fisher’s test) .....	102
4.6.3.	Rank-summarized test of Wilcoxon – Mann – Whitney .....	103
4.6.4.	Rank test of scattering of Siegel – Tukey .....	107
4.7.	Tests of consent .....	109
4.7.1.	$\chi^2$ – test (Pearson’s test) .....	110
4.7.2.	Kolmogorov’s test .....	114
4.7.3.	$n\omega^2$ – test (test of Cramer – Mises – Smirnov) .....	116
4.8.	Tests for conformity to random variable .....	118
4.8.1.	Criterion of sequences .....	119
4.8.1.1.	Test of common number of sequences .....	119
4.8.1.2.	Test of greatest length of sequence .....	120
4.8.2.	Test of number of raises and lowerings .....	121
4.8.3.	Test of number of extremums .....	122
<b>5.</b>	<b>Statistical Analysis of Associations Between Hydrological Variables</b> .....	<b>124</b>

5.1.1.	Least square method. The equation of a linear regression for two variables .....	125
5.2.	Linearization of nonlinear associations .....	130
5.3.	Estimate of an exactitude of the equation of a linear regression for two variables .....	132
5.4.	Interval estimation and estimation of a significance of parameters of a linear regression for two variables .....	136
5.4.1.	Interval estimation of coefficient of correlation. Z-transformation of Fisher .....	136
5.4.2.	Interval estimation of a slope .....	138
5.4.3.	Interval estimation of an intercept .....	139
5.4.4.	F - test of significance of regression .....	140
5.4.5.	Construction of a confidence interval for the equation of a linear regression .....	141
5.5.	Multiple linear correlation .....	142
5.5.1.	Estimate of an exactitude of the equation of a multiple linear regression .....	144
5.5.2.	Selection of the "best" equation of a regression .....	145
5.5.3.	Nonsense correlation .....	147
	<b>Literature</b> .....	149
	<b>Applications</b> .....	150
1.	The normalized ordinates of distribution of Pearson of the third type (continuous binomial distribution) .....	150
2.	Ordinates of distribution of Kritsky – Menkel .....	152
3.	Ordinates of H-distribution ( of A. Ezhov) .....	158
4.	$\chi^2$ - distribution .....	160
5.	t - distribution ( of Student) .....	161
6.	F - distribution (of Fisher) .....	162
7.	Distribution of Kolmogorov. ....	165
8.	Effect of a nonsense correlation for association $y = f(z)$ ( according to Benson) .....	166
9.	The table for finding of coefficient of a spread $C$ , by a method of the maximum likelihood depending on a parameter $\lambda_2$ and relation $C_0/C_p$ (for distribution of Pearson of the third-type) .....	167

## Предисловие

В настоящее время статистические методы обработки гидрометеорологической информации являются важнейшей составной частью прикладной гидрологии. Методы математической статистики и теории вероятностей используются при проведении гидрологических, водохозяйственных и гидроэнергетических расчетов, при составлении гидрологических прогнозов, в процессе экологического мониторинга, при оценке экономической эффективности водохозяйственных проектов и т. д.

При подготовке учебного пособия авторы стремились в доступной и компактной форме изложить основные методы статистического анализа и обработки гидрометеорологических данных, используемые в гидрологической практике.

Книга состоит из пяти глав. Содержит девять приложений, список условных обозначений и сокращений.

В первой главе приводятся краткие сведения из теории вероятностей, необходимые для понимания излагаемых в дальнейшем расчетных методов.

Вторая глава целиком посвящена описанию аналитических функций распределения, используемых в практике гидрологических расчетов; анализируются их достоинства и недостатки. Наряду с традиционными распределениями здесь рассматриваются четырехпараметрическое распределение Джонсона и  $H$ -распределение, предложенное в 1986 г. А. В. Ежовым (ГТИ); впервые опубликованы ординаты этого распределения.

В третьей главе излагаются методы расчета оценок параметров распределения гидрологических характеристик и методы построения эмпирических кривых обеспеченностей. В книге впервые предложены практическая реализация графоаналитического метода оценки параметров распределения для кривой Крицкого-Менкеля и новая методика оценки параметров распределения Джонсона.

Четвертая глава посвящена интервальной оценке параметров распределения и проверке статистических гипотез. Здесь излагаются методы проверки рядов на случайность и однородность. Рассматриваются наиболее часто используемые критерии согласия.

В пятой главе излагаются основы парной и множественной линейной корреляции. Рассматриваются приемы, позволяющие оценить надежность получаемых по эмпирическим данным уравнений регрессии.

В тексте содержится большое количество примеров, позволяющих практически закрепить теоретический материал.

Книга ориентирована на студентов-гидрологов, изучающих дисциплину "Методы статистической обработки гидрометеорологической информации". Вместе с тем она может быть полезной и инженерам-гидрологам, а также другим специалистам, работающим с гидрометеорологическими данными.

Авторы выражают глубокую благодарность профессору А. В. Рождественскому, который внимательно прочитал рукопись и сделал ряд ценных замечаний, а также Г. Н. Минасвой за помощь при подготовке рукописи к печати.

ВВЕДЕНИЕ

Вопросы, связанные с изучением истории развития науки, являются одними из наиболее актуальных в настоящее время. Это связано с тем, что в настоящее время наблюдается стремительное развитие науки и техники, что приводит к возникновению новых проблем, требующих решения. В связи с этим становится все более важным изучение истории развития науки, чтобы понять, как она развивалась и какие проблемы она решала в прошлом. Это поможет нам лучше понять современную науку и ее проблемы, а также найти новые пути ее развития.

В данной работе мы рассмотрим историю развития науки с древних времен до наших дней. Мы рассмотрим основные этапы развития науки, а также проблемы, которые стояли перед учеными в разные периоды истории. Мы также рассмотрим роль философии в развитии науки и ее влияние на научное мышление. В заключение мы сделаем выводы о том, как наука развивалась и какие проблемы она решала в прошлом, а также о том, какие проблемы она решает в настоящее время.

История развития науки является сложным и многогранным процессом, который изучается с помощью различных методов. В данной работе мы рассмотрим историю развития науки с древних времен до наших дней. Мы рассмотрим основные этапы развития науки, а также проблемы, которые стояли перед учеными в разные периоды истории. Мы также рассмотрим роль философии в развитии науки и ее влияние на научное мышление. В заключение мы сделаем выводы о том, как наука развивалась и какие проблемы она решала в прошлом, а также о том, какие проблемы она решает в настоящее время.

История развития науки является сложным и многогранным процессом, который изучается с помощью различных методов. В данной работе мы рассмотрим историю развития науки с древних времен до наших дней. Мы рассмотрим основные этапы развития науки, а также проблемы, которые стояли перед учеными в разные периоды истории. Мы также рассмотрим роль философии в развитии науки и ее влияние на научное мышление. В заключение мы сделаем выводы о том, как наука развивалась и какие проблемы она решала в прошлом, а также о том, какие проблемы она решает в настоящее время.

## Список условных обозначений и сокращений

### Условные обозначения

$F(x)$  – интегральная функция распределения

$f(x)$  – дифференциальная функция распределения  
(функция плотности вероятности)

$p$  – значение вероятности

$P(x)$  – функция обеспеченностей

$Mo$  – мода

$Me$  – медиана

$\alpha_s$  – начальный момент  $s$ -го порядка

$\mu_s$  – центральный момент  $s$ -го порядка

$m_x$  – математическое ожидание

$\sigma$  – среднее квадратическое отклонение (стандарт)

$\sigma^*$  и  $S$  – выборочное среднее квадратическое отклонение

$D$  и  $\sigma^2$  – дисперсия

$D^*$  и  $S^2$  – выборочная дисперсия

$C_v$  – коэффициент вариации

$C_s$  – коэффициент асимметрии

$E_x$  – эксцесс

$k$  – модульный коэффициент

$t$  – стандартная нормированная величина

$\Gamma(\cdot)$  – гамма-функция

$\nu$  – число степеней свободы

$r$  – коэффициент парной корреляции

$R$  – коэффициент множественной корреляции

$P$  – оператор вероятности

$M$  – оператор математического ожидания

$D$  – оператор дисперсии

$\sigma$  – оператор среднее квадратического отклонения

### Сокращения

СВ – случайная величина

МО – математическое ожидание

СКО – среднее квадратическое отклонение

ОДЗ – область допустимых значений

МНК – метод наименьших квадратов

## Введение

Многие гидрологические характеристики (среднегодовой, максимальный и минимальный расходы воды; слой стока за половодье или паводок; продолжительность половодья или межени; дата начала и окончания ледостава на реке или озере и др.) определяются огромным числом факторов, степень влияния каждого из которых учесть практически невозможно. При этом конкретное значение характеристики есть результат случайного сочетания этих факторов. С учётом этого сама исследуемая характеристика должна рассматриваться как случайная величина (или случайный процесс), и для ее определения могут быть использованы методы теории вероятностей и математической статистики. Кроме того, в практике гидрологических расчетов довольно часто встречаются задачи, когда необходимо определить значения гидрологических величин, которые будут встречаться в будущем, например в период эксплуатации того или иного гидротехнического сооружения. Решение такого рода задач также невозможно без привлечения вероятностных (стохастических) методов.

Широкое применение методов теории вероятностей в гидрологической практике началось относительно недавно. Основу для этого заложили работы А. Хазена и А. Фостера, опубликованные в первой четверти XX в. В России толчок для развития этого направления дала работа Д. Л. Соколовского (1930 г.) "Применение кривых распределения к установлению вероятных колебаний годового стока рек европейской части СССР". Большой вклад в развитие стохастической гидрологии внесли также работы Г. А. Алексева, Е. Г. Блохинова, Г. П. Калинина, С. Н. Крицкого, М. Ф. Менкеля, А. В. Рождественского и др.

К настоящему времени накоплен достаточно большой опыт применения вероятностных методов в практике гидротехнического, дорожного и водохозяйственного проектирования.

В книге рассматриваются основные приемы обработки и анализа гидрологических данных, используемые в российской и мировой гидрологической практике. Излагаемый в книге материал соответствует программе дисциплины "Методы статистической обработки гидрометеорологической информации", которая читается на гидрологическом факультете РГТМУ.

# 1. НЕКОТОРЫЕ СВЕДЕНИЯ ИЗ ТЕОРИИ ВЕРоятНОСТЕЙ

## 1.1. Случайные величины и функции распределения

Величина, значение которой меняется от опыта к опыту случайным образом, называется *случайной величиной* (СВ).

В отличие от неслучайных (детерминированных) величин для СВ нельзя предсказать точно, какое она примет значение в определенных условиях, а можно только указать *закон распределения* СВ.

*Закон распределения* СВ считается заданным, если:

- указано множество возможных значений СВ;
- указан способ количественного определения вероятности попадания СВ в любую область из множества возможных значений.

Вероятность попадания СВ в заданную область, например в интервал  $[a; b]$ , можно определить следующим образом:

$$P_{(a,b)} = \lim_{N \rightarrow \infty} \left( \frac{m}{N} \right), \quad (1.1)$$

где  $m$  – число наблюдений случайной величины, оказавшихся в заданной области,  $N$  – общее число наблюдений.

Аналитическими выражениями законов распределения случайных величин являются *функции распределения* – интегральная и дифференциальная.

*Интегральная функция распределения*  $F(x)$  случайной величины  $X$  показывает вероятность того, что случайная величина не превысит некоторого заданного числа  $x$ , т. е.  $F(x) = P\{X \leq x\}$ .<sup>1</sup>

Следовательно, вероятность того, что значение случайной величины  $X$  заключено между  $x_1$  и  $x_2$ , равна разности значений функции распределения, вычисленных в этих двух точках:

$$P\{x_1 < X \leq x_2\} = F(x_2) - F(x_1). \quad (1.2)$$

Аналогично

$$P\{X > x\} = P\{+\infty > X > x\} = 1 - F(x). \quad (1.3)$$

В российской гидрологической практике вместо функции  $F(x)$  чаще используется так называемая функция обеспеченностей  $P(x)$ , которая определяется выражением (1.3), но с включением в интервал значения  $x$ :

$$P(x) = 1 - F(x) = P\{X \geq x\}. \quad (1.3a)$$

<sup>1</sup> Здесь и далее символ  $P$  будет обозначать оператор вероятности, а  $p$  – конкретное значение вероятности.

Таким образом, функция обеспеченностей СВ  $X$  показывает вероятность превышения некоторого заданного числа  $x$  (рис. 1.1.)

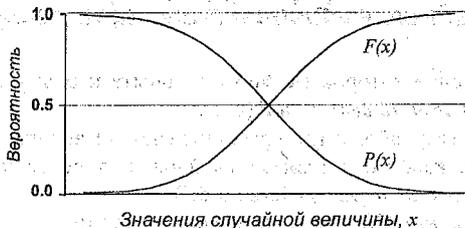


Рис. 1.1. Интегральная функция распределения  $F(x)$  и функция обеспеченностей  $P(x)$  случайной величины  $X$ .

Интегральная функция распределения  $F(x)$  и функция обеспеченностей  $P(x)$  обладают следующими свойствами:

1.  $\lim_{x \rightarrow -\infty} F(x) = 0$ ;  $\lim_{x \rightarrow -\infty} P(x) = 1$ .
2.  $\lim_{x \rightarrow +\infty} F(x) = 1$ ;  $\lim_{x \rightarrow +\infty} P(x) = 0$ .
3.  $F(x) \geq 0$  для всех  $x$ ;  $P(x) \geq 0$  для всех  $x$ .
4.  $F(x_2) \geq F(x_1)$ , если  $x_2 > x_1$ ;  $P(x_2) \leq P(x_1)$ , если  $x_2 > x_1$ .

Если функция  $F(x)$  дифференцируема для всех значений случайной величины  $X$ , то закон распределения вероятностей может быть выражен и в виде дифференциальной функции распределения вероятностей

$$f(x) = \frac{dF(x)}{dx} = \lim_{\Delta x \rightarrow 0} \frac{P\{x < X \leq x + \Delta x\}}{\Delta x}, \quad (1.4)$$

где  $\Delta x > 0$ . Таким образом, значение функции  $f(x)$  приближенно равно отношению вероятности попадания случайной величины  $X$  в интервал  $(x, x + \Delta x)$  к длине  $\Delta x$  этого интервала, когда  $\Delta x$  — бесконечно малая величина. Поэтому — функцию  $f(x)$  называют также функцией плотности распределения вероятностей (или короче — функцией **плотности вероятности**).

Отметим основные свойства функции  $f(x)$ :

$$1. f(x) \geq 0; \quad 2. \int_{-\infty}^{+\infty} f(x) dx = 1; \quad 3. \int_{-\infty}^x f(z) dz = F(x); \quad 4. \lim_{x \rightarrow \infty} f(x) = 0.$$

С помощью дифференциальной функции распределения можно вычислить вероятность попадания случайной величины в любую заданную область из множества возможных значений. В частности (рис.1.2):

$$a) \quad P\{X \leq a_1\} = \int_{-\infty}^{a_1} f(x) dx;$$

$$б) \quad P\{X > a_2\} = \int_{a_2}^{+\infty} f(x) dx;$$

$$в) \quad P\{a_1 < X \leq a_2\} = \int_{a_1}^{a_2} f(x) dx.$$

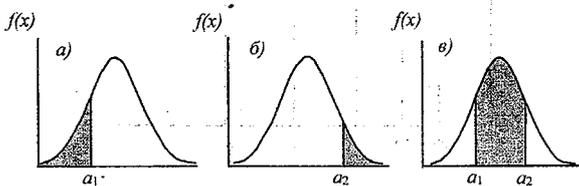


Рис.1.2. Графики функции плотности вероятности; заштрихованная площадь соответствует вероятности попадания СВ  $X$  в заданный интервал.

Как интегральная, так и дифференциальная функции распределения являются исчерпывающими характеристиками случайной величины.

## 1.2. Дискретные и непрерывные случайные величины

Следует различать дискретные и непрерывные случайные величины. Дискретная СВ  $X$  может принимать только конечное или счетное множество значений:  $x_1, x_2, x_3, \dots$ . Например, бросая игральную кость, мы генерируем дискретную случайную величину, которая может принимать только шесть значений: 1, 2, 3, ..., 6. В другом случае можно рассматривать дискретную СВ, которая принимает любые неотрицательные целочисленные значения. Здесь мы

уже имеем бесконечное множество возможных значений СВ  $X$ , но это множество является счетным.

Непрерывная СВ может принимать любые значения из некоторого замкнутого или открытого интервала, в том числе и бесконечного. Например, если область возможных значений непрерывной СВ ограничена интервалом  $[0, +1]$ , то СВ может принимать несчетное число значений. Перечислить их можно, только превратив непрерывную СВ в дискретную, вводя шаг дискретности. Если шаг дискретности 0,1, то СВ  $X$  округляется до одной десятой и будет принимать значения 0, 0,1, 0,2, ..., 0,9, 1,0. Если шаг дискретности 0,001, то СВ  $X$  будет принимать значения 0, 0,001, 0,002, ..., 0,999, 1,000.

Трактовка функции распределения, которая приведена в п. 1.1, в полной мере может быть отнесена только к непрерывным случайным величинам.

Интегральная функция распределения дискретной СВ  $X$  в практических ситуациях представляет собой ступенчатую функцию со скачками в точках  $x_1, x_2, x_3, \dots$  (рис. 1.3).

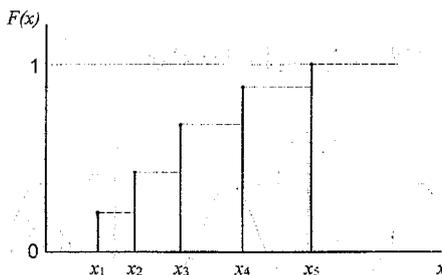


Рис. 1.3. Интегральная функция распределения дискретной случайной величины

Интегральная функция распределения  $F(x)$  дискретной СВ не дифференцируема, и вместо функции плотности вероятности используется ее дискретный аналог, который называют рядом распределения.

Ряд распределения дискретной случайной величины можно представить в виде таблицы, где для каждого значения СВ из области возможных значений указана его вероятность  $p_x$  (табл. 1.1).

Таблица 1.1

**Ряд распределения случайной величины  $X$**

$X$	...	$x_1$	$x_2$	$x_3$	...	$x_n$
$p_x$	...	$p_1$	$p_2$	$p_3$	...	$p_n$

На основании такой таблицы можно построить гистограмму распределения вероятностей дискретной СВ (рис. 1.4).

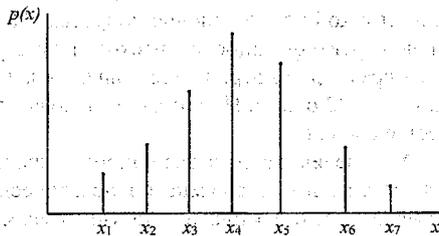


Рис.1.4. Гистограмма распределения вероятностей дискретной случайной величины

Для ряда распределения дискретной СВ должно выполняться равенство

$$\sum_{i=1}^n p(x_i) = 1, \text{ что эквивалентно свойству 2 плотности вероятности.}$$

В гидрометеорологической практике встречаются и дискретные, и непрерывные случайные величины. Например, число дней с отрицательной температурой в третьей декаде марта можно рассматривать как дискретную СВ с областью возможных значений: 0, 1, 2, 3, ..., 10. А максимальный в году расход воды можно рассматривать как непрерывную СВ, область возможных значений которой ограничена интервалом  $[0, +\infty)$ .

### 1.3. Случайная величина и время

В гидрологической практике исследователь имеет дело с рядами наблюдений за различными гидрометеорологическими характеристиками. Например, 32-летний ряд наблюдений за максимальными расходами дождевых паводков выглядит так, как показано в табл. 1.2, т. е. мы имеем отсчеты времени (с шагом в один год) и соответствующие паводочные расходы воды.

Таблица 1.2

Максимальные расходы воды дождевых паводков ( $\text{м}^3/\text{с}$ );  
р. Луга – ст. Толмачево,  $F = 6350 \text{ км}^2, n = 32$

Год	Расход	Год	Расход	Год	Расход	Год	Расход
1954	71.4	1962	121	1970	84.5	1978	130
1955	22.1	1963	25.0	1971	37.5	1979	22.0
1956	32.7	1964	31.4	1972	24.0	1980	54.0
1957	145	1965	17.3	1973	22.1	1981	100
1958	46.0	1966	54.8	1974	130	1982	78.1
1959	29.9	1967	29.9	1975	21.3	1983	39.2
1960	28.4	1968	104	1976	58.2	1984	66.0
1961	73.8	1969	27.5	1977	58.7	1985	84.4

Российский государственный  
гидрометеорологический  
институт  
БИБЛИОТЕКА  
195196, СПб, Малоохтинский пр., 21

Таким образом, максимальный паводочный расход можно рассматривать, как функцию времени, и такой подход вполне допустим. Но если мы пытаемся описать вероятностную структуру гидрологического ряда в рамках модели случайной величины, то время рассматривается только в качестве формального счетчика опытов: 32 года – 32 опыта. При этом последовательность расположения расходов не имеет значения.

Безусловно, случайная величина является наиболее простой вероятностной моделью, но она позволяет решить очень многие практические задачи.

Более сложные вероятностные модели, где учитывается фактор времени, рассматриваются в *теории случайных процессов*.

#### 1.4. Числовые характеристики случайных величин

Как отмечалось выше, исчерпывающей характеристикой любой случайной величины является ее функция распределения (интегральная или дифференциальная). Но основные свойства СВ (хотя и не все) могут быть описаны более компактно с помощью нескольких числовых характеристик (параметров распределения).

##### 1.4.1. Мода. Медиана. Математическое ожидание.

На практике одной из наиболее важных числовых характеристик является *центр распределения*. Однако определить его можно по-разному. Существует, по крайней мере, три параметра, характеризующих положение центра распределения (или короче – три характеристики положения): мода, медиана и математическое ожидание.

**Модой**  $M_o$  непрерывной СВ  $X$  называется такое ее значение, которому соответствует максимум плотности вероятности (рис. 1.5 а).

В случае дискретного распределения модой называется наиболее вероятное значение случайной величины (рис. 1.5 б). Если распределение имеет более чем один максимум, такое распределение называется *многомодальным* (полимодальным).

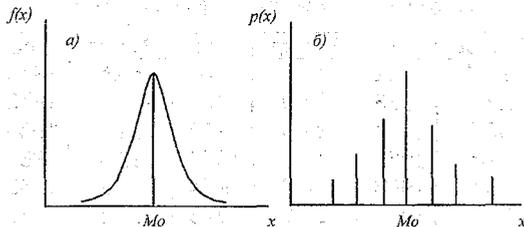


Рис. 1.5. Мода непрерывной (а) и дискретной (б) случайных величин.

**Медианой**  $Me$  непрерывной СВ  $X$  называется такое ее значение, при котором

$$P\{X < Me\} = P\{X > Me\} = 0.5. \quad (1.5)$$

Можно также сказать, что  $Me$  – это такое значение СВ, при котором значение функции обеспеченностей равно значению интегральной функции распределения (рис. 1.6)

$$P(Me) = F(Me) = 0.5. \quad (1.6)$$

Для дискретных СВ медиана определяется неоднозначно и практически не употребляется.

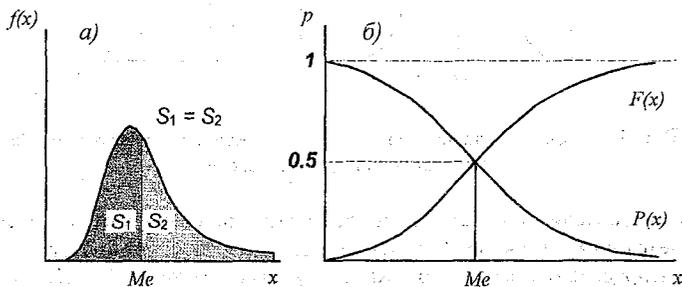


Рис. 1.6. Положение медианы на графиках дифференциальной (а) и интегральной (б) функций распределения.

**Математическое ожидание** (МО) случайной величины определяется следующими формулами:

для дискретной СВ 
$$m_x = \sum_i x_i p_i, \quad (1.7)$$

для непрерывной СВ 
$$m_x = \int_{-\infty}^{+\infty} x f(x) dx. \quad (1.8)$$

Как следует из (1.8), МО можно трактовать как центр тяжести плотности вероятности.

Математическое ожидание является наиболее употребляемой характеристикой положения, и в дальнейшем в качестве центра распределения мы будем подразумевать именно его. При этом следует отметить, что термин *математическое ожидание* применим не только к некоторой случайной величине  $X$ , но и к случайным величинам, которые являются функциями  $X$ , например, к  $X^2$  или  $\ln(X)$ . В этом случае в качестве символа МО мы будем использовать обозначение  $M[\cdot]$ . Таким образом, для СВ  $X$  можно записать также  $m_x \sim M[X]$ .

Кроме термина *математическое ожидание* используется и другой – *генеральное среднее*, в этом случае для обозначения МО используется символ  $\bar{x}_N$ , где  $N \rightarrow \infty$ .

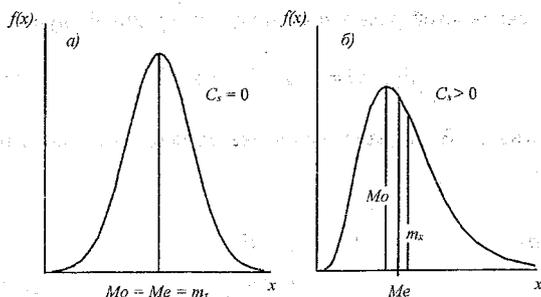


Рис. 1.7. Графики функций плотностей вероятности для симметричного (а) и несимметричного (б) распределений.

Если мода, медиана и математическое ожидание совпадают, то распределение является симметричным (рис.1.7 а). Если МО расположено правее медианы, то распределение имеет положительную асимметрию (рис.1.7 б), в противном случае – отрицательную.

#### 1.4.2. Моменты случайной величины

Большинство числовых характеристик, используемых на практике, связаны с понятием момента СВ. Различают начальные и центральные моменты СВ. **Начальный момент**  $S$ -го порядка СВ определяется формулой

$$\alpha_s = M[X^S], \quad (1.9)$$

**центральный момент**  $S$ -го порядка СВ  $X$  – формулой

$$\mu_s = M[(X - m_x)^S]. \quad (1.10)$$

Из формулы (1.9), в частности, следует, что МО есть первый начальный момент, т.е.  $m_x = M[X^1] = \alpha_1$ .

Начальные и центральные моменты связаны следующими соотношениями:

$$\left. \begin{aligned} \mu_1 &= 0 \\ \mu_2 &= \alpha_2 - \alpha_1^2 \\ \mu_3 &= \alpha_3 - 3\alpha_2\alpha_1 + 2\alpha_1^3 \\ \mu_4 &= \alpha_4 - 4\alpha_3\alpha_1 + 6\alpha_2\alpha_1^2 - 3\alpha_1^4 \end{aligned} \right\} \quad (1.11)$$

### 1.4.3. Дисперсия. Среднеквадратическое отклонение. Коэффициент вариации

Вторую группу наиболее часто используемых на практике параметров составляют параметры, характеризующие степень рассеяния СВ относительно центра распределения. К ним относятся: дисперсия, среднеквадратическое отклонение и коэффициент вариации.

*Дисперсия* СВ  $X$ , представляет собой второй центральный момент, т.е.

$$D_x = \mu_2 = \mathbf{M}[(X - m_x)^2], \quad (1.12)$$

в частности, для непрерывной СВ  $X$  дисперсия определяется формулой

$$D_x = \int_{-\infty}^{+\infty} (X - m_x)^2 f(x) dx. \quad (1.12a)$$

*Среднеквадратическое отклонение* (СКО) СВ  $X$  есть квадратный корень из дисперсии

$$\sigma_x = \sqrt{D_x} \quad \text{или} \quad \sigma_x^2 = D_x. \quad (1.13)$$

Эту величину также называют *стандартным отклонением*, или *стандартом*. Для описания рассеяния положительных СВ можно использовать безразмерную характеристику — коэффициент вариации.

*Коэффициент вариации*  $C_v$  СВ  $X$  есть отношение СКО к МО:

$$C_v = \frac{\sigma_x}{m_x} = \frac{\sqrt{D_x}}{m_x}. \quad (1.14)$$

Как следует из формул (1.12) – (1.14), дисперсия, СКО и коэффициент вариации взаимосвязаны, и на практике обычно используется только одна (любая) из перечисленных характеристик. Выбор того или иного параметра зависит от специфики задачи. Влияние коэффициента вариации (а следовательно, дисперсии и СКО) на функцию распределения иллюстрируется на рис. 1.8 а.

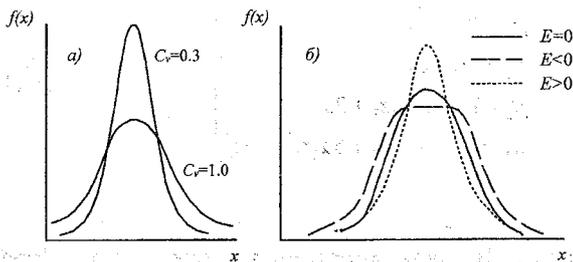


Рис. 1.8. Влияние коэффициента вариации (а) и эксцесса (б) на форму функции плотности вероятности

### 1.4.4. Асимметрия и эксцесс

Помимо уже перечисленных числовых характеристик рассмотрим еще две: коэффициент асимметрии и эксцесс.

**Коэффициент асимметрии**  $C_s$  является безразмерным параметром и характеризует степень симметричности рассеяния относительно математического ожидания (см. Рис.1.7). Коэффициент асимметрии определяется формулой

$$C_s = \frac{\mu_3}{\sigma_x^3} = M[(X - m_x)^3] / \sigma_x^3. \quad (1.15)$$

Для симметричных распределений коэффициент асимметрии равен нулю.

**Эксцесс**  $E_x$  также является безразмерным параметром и определяется формулой

$$E_x = \frac{\mu_4}{\sigma_x^4} - 3 = M[(X - m_x)^4] / \sigma_x^4 - 3. \quad (1.16)$$

Эксцесс позволяет оценить наличие островершинности, или наоборот туповершинности, функции плотности вероятности СВ  $X$  относительно нормального закона распределения (см. п.2.1), для которого  $E_x = 0$  (Рис. 1.8 б).

Как видно из формул (1.15) – (1.16), коэффициент асимметрии зависит от третьего, а эксцесс – от четвертого центральных моментов. Теоретически доказано, что погрешность оценки моментов по эмпирическим данным резко возрастает с увеличением порядка момента, поэтому в гидрологической практике эксцесс используется крайне редко.

### 1.4.5. Свойства числовых характеристик случайных величин

В данном параграфе приводятся без доказательства основные свойства МО и дисперсий. Доказательства приводимых соотношений можно найти в [1, 3, 11].

### Свойства математического ожидания

1. Математическое ожидание постоянной величины равно самой этой величине

$$\mathbf{M}[c] = c, \quad (1.17)$$

где  $c = \text{const}$ .

2. Постоянный множитель можно выносить за знак МО

$$\mathbf{M}[cX] = c\mathbf{M}[X]. \quad (1.18)$$

3. Математическое ожидание суммы независимых случайных величин равно сумме их МО

$$\mathbf{M}\left[\sum_{i=1}^N X_i\right] = \sum_{i=1}^N \mathbf{M}[X_i], \quad (1.19)$$

в частности, для двух случайных величин:  $\mathbf{M}[X + Y] = \mathbf{M}[X] + \mathbf{M}[Y]$ .

4. Математическое ожидание линейной функции от СВ выражается формулой

$$\mathbf{M}[aX + b] = a\mathbf{M}[X] + b, \quad (1.20)$$

где  $a$  и  $b = \text{const}$ .

Данное свойство является следствием свойств 1–3.

5. Математическое ожидание произведения независимых СВ равно произведению их МО

$$\mathbf{M}\left[\prod_{i=1}^N X_i\right] = \prod_{i=1}^N \mathbf{M}[X_i]. \quad (1.21)$$

### Свойства дисперсии

1. Дисперсия постоянной величины равна нулю<sup>1</sup>

$$\mathbf{D}[c] = 0, \quad (1.22)$$

где  $c = \text{const}$ .

2. Постоянную величину можно вынести за знак дисперсии, возводя ее в квадрат

$$\mathbf{D}[cX] = c^2 \mathbf{D}[X]. \quad (1.23)$$

Из выражения (1.23) также следует,

$$\sigma[cX] = c \cdot \sigma[X]. \quad (1.24)$$

3. Дисперсия суммы случайных величин равна сумме их дисперсий

<sup>1</sup> Здесь и далее  $\mathbf{D}$  и  $\sigma$  – соответственно операторы дисперсии и СКО.

$$D \left[ \sum_{i=1}^N X_i \right] = \sum_{i=1}^N D[X_i]. \quad (1.25)$$

4. Дисперсия линейной функции СВ

$$D[aX + b] = a^2 D[X]. \quad (1.26)$$

Данное свойство является следствием свойств 1-3.

#### 1.4.6. Стандартные преобразования случайной величины

Во многих случаях оказывается удобным рассматривать вместо СВ  $X$  другую случайную величину, полученную из СВ  $X$  на основе элементарного преобразования. В гидрологической практике наиболее часто используется замена СВ  $X$  модульными коэффициентами и замена СВ стандартной нормированной случайной величиной.

Модульным коэффициентом называется отношение значения СВ к ее математическому ожиданию

$$k_i = x_i / m_x. \quad (1.27)$$

Стандартная нормированная величина может быть получена из СВ по формуле

$$t_i = (x_i - m_x) / \sigma_x, \quad (1.28)$$

или с учетом формул (1.14), (1.27)

$$t_i = (k_i - 1) / C_v. \quad (1.28a)$$

Из свойств МО и дисперсии, изложенных в п. 1.4.5, следует, что формула (1.27) преобразует СВ  $X$  с МО равным  $m_x$ , в случайную величину с МО, равным единице, а преобразование (1.28) приводит к СВ с нулевым МО и  $\sigma_t = D_t = 1$ .

#### 1.4.7. Квантили распределения

В п. 1.1 мы определили интегральную функцию распределения  $F(x)$ , по которой для любого  $x$  можно установить вероятность его непревышения. Во многих практических случаях необходимо решить обратную задачу: по заданной вероятности непревышения  $F(x) = p$  определить величину  $x_p$ . Для обозначения  $x_p$ , в этом случае в математической статистике используется специальный термин — *квантиль*.

Таким образом,  $p$ -квантилем или  $p\%$ -квантилем (если  $p$  в %) называется значение случайной величины  $x_p$ , соответствующее заданному значению вероятности непревышения  $F(x) = p$ .

По аналогии с квантилями, в гидрологической практике используются  $\hat{p}$ -ординаты кривой обеспеченностей.<sup>1</sup>

Ординатой кривой обеспеченностей будем называть, такое значение СВ  $X$  (обозначим его  $\hat{x}_p$ ), которое соответствует заданной вероятности превышения  $P(x) = \hat{p}$ .

Как следует из (3.1а),  $P(x) = 1 - F(x)$ , и следовательно,  $p$  и  $\hat{p}$  связаны соотношением  $\hat{p} = 1 - p$  или (если  $p$  в %)  $\hat{p} = 100 - p$ .

<sup>1</sup> Для того чтобы подчеркнуть отличие квантилей от ординат кривой обеспеченностей, последние будем отмечать символом "Λ".

## 2. АНАЛИТИЧЕСКИЕ ФУНКЦИИ РАСПРЕДЕЛЕНИЯ, ИСПОЛЬЗУЕМЫЕ В ГИДРОЛОГИИ

Во многих случаях в качестве математической модели для описания гидрометеорологических явлений может использоваться случайная величина. В частности, в качестве СВ можно рассматривать среднегодовой расход воды, тогда среднегодовые расходы за отдельные годы, например за 30 лет, следует трактовать как последовательность значений СВ, полученную в результате 30 природных опытов.

Точно так в качестве СВ могут рассматриваться максимальный и минимальный расходы воды, годовая сумма осадков, снеготпасы на водосборе и т. д. Кроме того, в качестве случайных величин рассматриваются погрешности измерений гидрометеорологических характеристик (уровня воды, расхода, слоя осадков и т. д.), а также ошибки расчета различных числовых параметров, определяемых по эмпирическим данным.

Как уже отмечалось, каждая случайная величина исчерпывающим образом характеризуется своим законом распределения. Аналитическим выражением закона распределения является функция распределения (дифференциальная или интегральная), которую записывают в виде параметрического выражения, где в качестве параметров обычно используют числовые характеристики случайной величины (МО, СКО и т. д.). При такой записи для каждого значения СВ соответствующее значение функции распределения однозначно определяется параметрами аналитического выражения (параметрами распределения).

В гидрологической практике, как правило, рассматриваются случайные величины, функции распределения которых зависят от небольшого числа параметров — обычно двух или трех.

При этом в одних ситуациях, закон распределения можно вывести из теоретических соображений, в других — он не известен, и аналитическое выражение функции распределения является лишь аппроксимацией истинного распределения. В последнем случае можно предложить несколько аналитических выражений, и исследователь вынужден производить выбор подходящего, опираясь на практический опыт и некоторые априорные соображения.

Ниже рассматриваются аналитические функции распределения, наиболее часто используемые в практике гидрологических расчетов.

### 2.1. Нормальное распределение

В природе и различных областях человеческой деятельности весьма распространены СВ, которые представляют собой сумму большого числа независимых или слабо зависимых СВ, дисперсии которых малы по сравнению с дисперсией всей суммы. Как следует из центральной предельной теоремы [1, 3], распределение таких СВ при весьма общих дополнительных условиях

хорошо аппроксимируется нормальным распределением. Этим объясняется весьма широкое распространение последнего. Нормальное распределение применяется и в тех случаях, когда истинный закон распределения известен, но вычисления по этому закону затруднены, а аппроксимация его нормальным законом не приводит к большим ошибкам.

Функция плотности вероятности нормального распределения определяется выражением

$$f(x) = [1/(\sigma_x \sqrt{2\pi})] \cdot \exp[-(x - m_x)^2 / (2\sigma_x^2)]; \quad (2.1)$$

где  $\pi$  – число Пи. Таким образом, в общем случае нормальное распределение является двухпараметрическим, т. е. зависит от двух параметров: математического ожидания ( $m_x$ ) и среднеквадратического отклонения ( $\sigma_x$ ).

График плотности вероятности нормального распределения представлен на рис. 1.7 а. Как видно на рисунке, нормальное распределение является симметричным и, следовательно, для него коэффициент асимметрии равен нулю ( $C_s = 0$ ), а мода, медиана и МО совпадают. Область возможных значений СВ, подчиняющейся нормальному распределению, – интервал  $(-\infty, +\infty)$ .

Интегральная функция распределения имеет вид

$$F(x) = [1/(\sigma_x \sqrt{2\pi})] \int_{-\infty}^x \exp[-(z - m_x)^2 / (2\sigma_x^2)] dz. \quad (2.2)$$

где  $z$  – переменная интегрирования.

Интеграл, стоящий в правой части выражения (2.2), нельзя выразить через элементарные функции, поэтому ординаты нормального закона распределения обычно представляют в виде таблиц. Чтобы не публиковать множество таблиц, для различных сочетаний  $m_x$  и  $\sigma_x$  вместо СВ  $X$  рассматривают нормированную СВ  $t$  (табл.2.1), значения которой связаны со значениями  $X$  формулой (1.28).

Таблица 2.1

Квантили нормированной нормально распределенной случайной величины  $t$   
(значения  $F(t)$  даны в %)

$F(t)$	$t$	$F(t)$	$t$	$F(t)$	$t$
0.01	-3.72	20	-0.84	90	1.28
0.1	-3.05	25	-0.67	95	1.64
0.5	-2.58	30	-0.52	97	1.88
1	-2.33	40	-0.25	97.5	1.96
2	-2.02	50	0.00	98	2.02
2.5	-1.96	60	0.25	99	2.33
3	-1.88	70	0.52	99.5	2.58
5	-1.64	75	0.67	99.9	3.05
10	-1.28	80	0.84	99.99	3.72

Так как у нормированной СВ  $t: m_t = 0$  и  $\sigma_t = 1$ , функции распределения для нее с учетом формул (2.1) – (2.2), будут определяться выражениями

$$f(t) = (1/\sqrt{2\pi}) \cdot \exp(-t^2/2), \quad (2.3)$$

$$F(x) = (1/\sqrt{2\pi}) \int_{-\infty}^t \exp(-z^2/2) dz. \quad (2.4)$$

Переход от нормированных значений случайной величины к исходным в соответствии с выражением (1.28) производится по формуле

$$x_p = t_p \sigma_p + m_x, \quad (2.5)$$

а для модульных коэффициентов по формуле

$$k_p = t_p C_v + 1. \quad (2.5a)$$

**Пример 2.1.** Известно, что СВ  $X$  подчиняется нормальному распределению с параметрами  $m_x = 10$ ,  $\sigma_x = 5$ . Требуется определить значение  $x_p$ , соответствующее вероятности невыщшения 95 % (т.е. при  $F(x) = 95\%$ ).

**Решение:**

1. По табл. 2.1 определяем 95%-ный квантиль нормированной СВ  $t: t_{95} = 1.64$ .
2. По формуле (2.5) определяем  $x_{95} = 1.64 \cdot 5 + 10 = 18.2$ .

**Пример 2.2.** Для СВ  $X$  из предыдущего примера требуется определить значение  $x_p$ , соответствующее вероятности превышения 1 % (при  $P(x) = 1\%$ ).

**Решение.**

1. Так как в табл. 2.1 приводятся ординаты интегральной функции распределения (а не функции обеспеченности), сначала определяем соответствующую вероятность невыщшения для нормированной СВ  $t: F(t) = 100 - P(t) = 100 - 1 = 99\%$ .
2. По табл. 2.1 определяем 99%-ный квантиль нормированной СВ  $t: t_{99} = 2.33$ .
3. По формуле (2.5) определяем  $x_1: x_1 = x_{99} = 2.33 \cdot 5 + 10 = 21.65$ .

Несмотря на широкое распространение, нормальное распределение не универсально. В частности, если в качестве СВ рассматривать среднегодовые расходы воды, то налицо два несоответствия:

- расходы воды всегда положительны, а область возможных значений для нормальной СВ определяется интервалом  $(-\infty, +\infty)$ .
- для нормального закона распределения  $C_s = 0$ , а распределение среднегодовых расходов воды (и целого ряда других гидрологических характеристик), как правило, имеет умеренную положительную асимметрию.

## 2.2. Закон равномерной плотности

Пусть возможные значения СВ  $X$  лежат в интервале  $[a, b]$  и нет оснований для того, чтобы отдать предпочтение какому-либо из этих значений. При этих условиях СВ  $X$  подчиняется закону равномерной плотности (рис. 2.1). Это распределение называют также равномерным или прямоугольным.

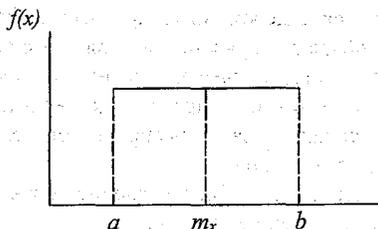


Рис. 2.1. Дифференциальная функция распределения закона равномерной плотности

Закон равномерной плотности определяется двумя параметрами: началом  $a$  и концом  $b$  интервала изменения СВ  $X$ .

Плотность вероятности равна

$$f(x) = \begin{cases} 0 & \text{при } x \leq a; \\ 1/(b-a) & \text{при } a < x \leq b; \\ 0 & \text{при } x > b. \end{cases} \quad (2.6)$$

Интегральная функция распределения определяется равенством

$$F(x) = \begin{cases} 0 & \text{при } x \leq a; \\ (x-a)/(b-a) & \text{при } a < x \leq b; \\ 1 & \text{при } x > b. \end{cases} \quad (2.7)$$

МО, дисперсия и СКО соответственно равны

$$m_x = (a+b)/2, \quad (2.8)$$

$$D_x = (a-b)^2/12, \quad (2.9)$$

$$\sigma_x = (b-a)/(2\sqrt{3}). \quad (2.10)$$

Медиана совпадает с МО. Закон равномерной плоскости моды не имеет, так как все значения плотности вероятности равны между собой.

Закон равномерной плотности играет важную роль при моделировании искусственных гидрологических рядов, поскольку значения любой случайной величины можно получить преобразованием значений СВ, равномерно распределенной на отрезке  $[0, 1]$ .

### 2.3. Логарифмически нормальное распределение

Многие гидрологические переменные, которые можно рассматривать как случайные величины, обладают правой, т.е. положительной асимметрией. Это объясняется тем, что гидрологические величины обычно превышают нуль либо некоторый другой нижний предел, но теоретически не ограничены верхним пределом. В таких случаях распределение случайной величины не соответствует нормальному закону.

Однако часто асимметричное распределение можно привести к нормальному путем замены переменной на ее логарифм.

Неотрицательная СВ  $X$  называется распределенной логарифмически нормально, если ее логарифм  $Z = \ln(X)$  распределен по нормальному закону (рис. 2.2).

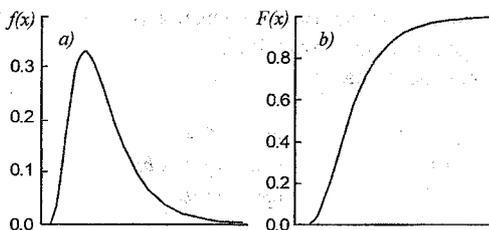


Рис. 2.2. Дифференциальная (а) и интегральная (б) функции распределения логнормального закона при  $m_z = 1$  и  $\sigma_z = 0.5$ .

Интегральная функция распределения для СВ  $X$  в этом случае имеет вид

$$F(x) = \begin{cases} 0 & \text{при } x \leq 0 \\ F(u) = (1/\sqrt{2\pi}) \int_{-\infty}^u \exp(-s^2/2) ds & \text{при } x > 0, \end{cases} \quad (2.11)$$

где  $u = (z - m_z)/\sigma_z$ ;  $z = \ln(x)$ ;  $s$  – переменная интегрирования.

Дифференцируя выражение (2.11), получим функцию плотности вероятности логнормального распределения (см. рис.2.2):

$$f(x) = \begin{cases} 0 & \text{при } x \leq 0, \\ [1/(\sigma_z x)] f(u) = [1/(\sigma_z x \sqrt{2\pi})] \exp(-u^2/2) & \text{при } x > 0. \end{cases} \quad (2.12)$$

Таким образом, данное распределение определяется двумя параметрами:  $m_z$  и  $\sigma_z$ . Величина  $m_z$  представляет собой МО случайной величины  $Z$ , а  $\sigma_z$  – ее среднеквадратическое отклонение.

Дисперсия СКО и МО случайных величин  $X$  и  $Z$  связаны соотношениями

$$D_z = \sigma_z^2 = \ln(\sigma_x^2/m_x^2 + 1), \quad (2.13)$$

$$m_z = \ln(m_x) - \sigma_z^2/2 = \ln(m_x) - [\ln(\sigma_x^2/m_x^2 + 1)]/2. \quad (2.14)$$

Если закон распределения СВ  $X$  не известен и логнормальное распределение используется для аппроксимации истинного распределения  $X$ , то значения  $m_z$  и  $\sigma_z$  для СВ  $Z$  могут отличаться от  $m_z$  и  $\sigma_z$ , полученных по формулам (2.13), (2.14) через  $m_x$  и  $\sigma_x$ . При этом расхождение будет тем больше, чем больше отклонение истинного распределения СВ  $X$  от логнормального.

Поскольку в гидрологической практике истинное распределение анализируемой СВ, как правило, не известно, то допускается два варианта расчета:

- 1) оценка  $m_z$  и  $\sigma_z$  производится по ряду значений СВ  $Z$ .
- 2) по ряду значений СВ  $X$  производится оценка  $m_x$  и  $\sigma_x$ , а затем по формулам (2.13), (2.14) определяются  $m_z$  и  $\sigma_z$ .

Коэффициент асимметрии логарифмически нормального распределения определяется по формуле

$$C_s = 3C_v + C_v^3, \quad (2.15)$$

где  $C_v = \sigma/m_x$ . Так как для большинства гидрологических переменных  $C_v < 1$ , то наиболее хорошего соответствия эмпирических данных с логнормальным распределением можно ожидать при  $C_s/C_v = 3 - 3.5$ .

Мода и медиана для СВ  $X$ , имеющей логнормальное распределение, определяются равенствами

$$Mo = \exp(m_z - \sigma_z^2), \quad (2.16)$$

$$Me = \exp(m_z). \quad (2.17)$$

Для определения ординат кривой обеспеченности логнормального закона используют таблицы стандартного нормального распределения, но в качестве случайной величины рассматривают  $Z = \ln(X)$ .

**Пример 2.3.** Определить максимальный паводочный расход однопроцентной обеспеченности ( $\hat{Q}_1$ ) на р. Луге в створе ст. Толмачево (см. табл. 1.2.), используя логнормальный закон распределения. Привести два варианта решения.

**Решение 1.**

1. Проводим преобразование исходного ряда по формуле  $z_i = \ln Q_i$ .
2. По ряду значений СВ  $Z$  приближенно определяем  $m_z$  и  $\sigma_z$  (методика приближенной оценки параметров распределения излагается в главе 3). Получаем:  $m_z \approx 3.87$ ;  $\sigma_z \approx 0.64$ .
3. Так как таблица стандартного нормального распределения (см. табл.2.1) приводится для интегральной функции распределения, заменяем обеспеченность на вероятность непревышения:  $F(t) = 100 - P(t) = 100 - 1 = 99\%$ .
4. По табл.2.1 определяем 99%-ный квантиль нормированной нормально распределенной СВ  $t$ :  $t_{99} = 2.33$ .
5. По формуле (2.5) определяем  $z_1$ :  

$$z_1 = z_{99} = \sigma_z t_{99} + m_z = 2.33 \cdot 0.64 + 3.87 = 5.36.$$
6. Так как  $z_1 = \ln \hat{Q}_1$ , то  $\hat{Q}_1 = \exp(z_1) = \exp(5.36) = 213 \text{ м}^3/\text{с}$ .

**Решение 2.**

1. Непосредственно по исходному ряду (см. табл. 1.2) приближенно определяем  $m_Q$  и  $\sigma_Q$ , получаем:  $m_Q \approx 58.44$ ;  $\sigma_Q \approx 37.19$ .
2. По формулам (2.13), (2.14) вычисляем  $m_z$  и  $\sigma_z$ :  

$$\sigma_z^2 = \ln[(37.19)^2 / (58.44)^2 + 1] = 0.34; \quad \sigma_z = 0.58;$$

$$m_z = \ln(58.44) - 0.34 / 2 = 3.90.$$
3. Дальнейший расчет ведется аналогично первому варианту:  
 $t_{99} = 2.33; \quad z_1 = z_{99} = 2.33 \cdot 0.58 + 3.90 = 5.25; \quad \hat{Q}_1 = \exp(5.25) = 191 \text{ м}^3/\text{с}.$

Заметим, что полученные по формулам (2.13), (2.14) значения  $m_z$  и  $\sigma_z$  несколько отличаются от значений этих параметров, полученных по первому варианту, а расчетные расходы (213 и 191) отличаются на 12%. Это косвенным образом подтверждает, что в данном случае истинное распределение может быть аппроксимировано логнормальным, однако такая аппроксимация не будет идеальной.

В практике гидрологических исследований логнормальное распределение наиболее часто используется при расчетах максимального паводочного стока. Такая практика имеет место в Болгарии, Великобритании, России, США, Японии и ряде других стран.

Наряду со стандартным логнормальным распределением используются различные его модификации. В частности, Г.А. Алексеев (1960 г.) предложил вместо преобразования  $Z = \ln(X)$ , использовать преобразование  $Z = \ln(X - a)$ , где  $a$  — дополнительный (третий) параметр.

Полученное на основе этого преобразования распределение, иногда называют трехпараметрическим логнормальным распределением. В отличие от

стандартного логнормального распределения, которое имеет нулевой нижний предел, это распределение имеет нижний предел, равный  $a$ .

Параметр  $a$  связан с коэффициентом асимметрии соотношением

$$C_{s,x} = \frac{3C_{v,x}}{1-k_0} + \left( \frac{C_{v,x}}{1-k_0} \right)^3, \quad (2.18)$$

где  $k_0$  – минимальный модульный коэффициент,  $k_0 = a/m_x$ . При  $k_0 = 0$  выражение (2.18) превращается в (2.15).

Распределение является более гибким по сравнению со стандартным, однако возможности его применения в гидрологии ограничены. Действительно, для того чтобы кривая распределения не уходила в область отрицательных значений, необходимо, чтобы значение  $k_0$  удовлетворяло условию  $k_0 \geq 0$ . В то же время минимальный модульный коэффициент всегда меньше единицы. Окончательно получаем  $0 \leq k_0 < 1$ . Таким образом, как видно из (2.18), логнормальное распределение с положительным нижним пределом ( $a > 0$ ) будет иметь более высокую асимметрию, чем обычное логнормальное, т.е.

$$C_s > 3C_v + C_v^3.$$

С учетом сказанного это распределение можно рекомендовать только для случайных величин с большой асимметрией  $C_s > 3.5C_v$ , но такие ситуации в гидрологической практике довольно редки.

**Пример 2.4.** Определить максимальный расход дождевого паводка однопроцентной обеспеченности ( $\hat{Q}_1$ ) на р. Луге в створе ст. Толмачево (см. табл.1.2), используя трехпараметрическое логнормальное распределение. Расчет вести по второму варианту.

**Решение.**

1. Непосредственно по исходному ряду приближенно определяем МО, СКО и коэффициент вариации :  $m_Q \approx 58.4$ ;  $\sigma_Q \approx 37.2$ ;  $C_v \approx 0.64$ .
2. Поскольку расчёт коэффициента асимметрии по эмпирическим данным производится с большой ошибкой, принимаем соотношение  $C_s/C_v$  осредненным по группе рек исследуемого района. В данном случае  $C_s/C_v = 2.5$ . Следовательно, принимаем  $C_{s,Q} \approx C_{v,Q} \cdot 2.5 \approx 0.64 \cdot 2.5 \approx 1.60$ .
3. В зависимости от  $C_{s,Q}$  по формуле (2.18) находим минимальный модульный коэффициент  $k_0$ . Решаем задачу методом последовательных приближений (табл. 2.2), т. е., задаваясь различными значениями  $k_0$  при  $C_v = 0.64$ , вычисляем теоретическое значение коэффициента асимметрии  $C_{s,T}$ . И в качестве расчётного принимаем такое значение  $k_0$ , при котором  $C_{s,T} = C_{s,Q}$ . Как видно из табл.2.2.  $k_0 = -0.3$ .

Зависимость  $k_0 = f(C_{s,T})$  при  $C_{v,Q} = 0.64$ 

$k_0$ .....	-0.5	-0.4	-0.3	-0.2	-0.1	0.0	0.1
$C_{s,T}$ .....	1.36	1.47	1.60	1.75	1.94	2.18	2.49

4. Зная  $k_0$ , находим параметр  $a$ :  $a = k_0 \bar{Q} = 58.4 \cdot (-0.3) = -17.5$ .

*Замечание.* Знак минус указывает на то, что при больших обеспеченностях кривая будет уходить в область отрицательных значений. Этого и следовало ожидать, так как асимметрия ряда не велика и не выполняется условие  $C_{s,Q} > 3C_{v,Q} + C_{v,Q}^3$ .

5. Для СВ  $U$ , которая связана с  $Q$  соотношением  $u_i = Q_i - a$ , рассчитываем числовые характеристики:

$$\bar{U} = \bar{Q} - a = 58.4 - (-17.5) = 58.4 + 17.5 = 75.9; \quad \sigma_u = \sigma_x = 37.2.$$

6. По формулам (2.13) и (2.14) вычисляем  $m_z$  и  $\sigma_z$ , где  $Z = \ln(U)$ :

$$\sigma_z^2 = \ln[(37.2)^2 / (75.9)^2 + 1] = 0.22; \quad \sigma_z = 0.47;$$

$$m_z = \ln(75.9) - (0.22/2) = 4.22.$$

7. По табл. 2.1 определяем 99 %-ный квантиль нормированной нормально распределённой СВ  $t$ :  $t_{99} = 2.33$ .

8. По формуле (2.5) определяем  $z_1 = z_{99} = \sigma_z t_{99} + m_z = 2.33 \cdot 0.47 + 4.22 = 5.32$ .

9. Так как  $z_1 = \ln(\hat{Q}_1 + a)$ , то  $\hat{Q}_1 = \exp(z_1) - a = \exp(5.32) - 17.5 = 187 \text{ м}^3/\text{с}$ .

Ещё одна модификация логнормального распределения будет рассмотрена в п. 2.7.

## 2.4. Закон распределения крайних членов выборки (распределение Гумбеля)

Закон распределения Гумбеля<sup>1</sup> разрабатывался для случая, когда в качестве случайной величины рассматриваются экстремальные характеристики гидрометеорологического режима, например максимальный в году срочный расход, минимальный в году среднесуточный расход, максимальный в году суточный слой осадков и т.д. Функция обеспеченности Гумбеля определяется выражением

$$P(x) = P\{X \geq x\} = 1 - \exp[-\exp(-y)], \quad (2.19)$$

где  $y$  - безразмерная величина, связанная с  $x$  выражением

$$y = \alpha(x - q), \quad (2.20)$$

<sup>1</sup> В литературе встречается также "распределение Гамбеля".

где  $q$  – мода случайной величины  $X$ . Параметр  $q$  определяется в зависимости от среднего значения и СКО исходного ряда по формуле

$$q = x - 0.45\sigma_x, \quad (2.21)$$

а  $\alpha$  можно выразить как

$$\alpha = 1.28 / \sigma_x. \quad (2.22)$$

На практике вместо уравнения (2.20) используется уравнение, полученное при решении (2.20) относительно  $x$ :

$$\hat{x}_p = q + (1/\alpha)\hat{y}_p. \quad (2.23)$$

Значения  $\hat{y}_p$  можно получить из выражения (2.19) после двукратного логарифмирования

$$\hat{y}_p = -\ln\{-\ln[(100-p)/100]\}, \quad (2.24)$$

где  $p$  – расчетная обеспеченность в процентах. В табл. 2.3 приводятся значения  $\hat{y}_p$  для основных опорных обеспеченностей.

Таблица 2.3

Значения  $\hat{y}_p$  для различных обеспеченностей

$\hat{p} \%$	$\hat{y}_p$	$\hat{p} \%$	$\hat{y}_p$	$\hat{p} \%$	$\hat{y}_p$
0.01	9.09	10	2.25	80	-0.48
0.1	6.89	20	1.50	90	-0.83
0.5	5.29	30	1.03	95	-1.10
1	4.60	50	0.37	99.0	-1.53
5	2.97	70	-0.19	99.9	-1.93

Уравнения (2.21) и (2.22) связывают параметры  $q$  и  $\alpha$  с  $\bar{x}$  и  $\sigma_x$  при  $n \rightarrow \infty$ . Для конечных выборок из  $n$  членов Гумбель предложил следующие формулы:

$$(1/\alpha) = \sigma_x / \sigma_y, \quad (2.25)$$

$$q = \bar{x} - y(1/\alpha). \quad (2.26)$$

Параметры  $y$  и  $\sigma_y$  определяются в зависимости от длины анализируемого ряда (табл. 2.4).

С учетом формулы (2.20) выражение функции плотности вероятности распределения Гумбеля имеет вид

$$f(x) = \alpha \exp\{-\alpha(x-q) - \exp[-\alpha(x-q)]\}. \quad (2.27)$$

Как видно из (2.27), для распределения Гумбеля область возможных значений СВ  $X$  – интервал  $(-\infty, +\infty)$ . Распределение является двухпараметрическим, т.е. однозначно определяется параметрами  $\bar{x}$  и  $\sigma_x$ .

Таблица 2.4

Средние значения параметров  $\bar{y}$  и  $\sigma_y$  при различном числе членов ряда  $n$  (по Гумбелю)

$n$	$\bar{y}$	$\sigma_y$	$n$	$\bar{y}$	$\sigma_y$	$n$	$\bar{y}$	$\sigma_y$
20	0.524	1.063	40	0.544	1.141	60	0.552	1.175
22	0.527	1.076	42	0.545	1.146	65	0.554	1.180
24	0.530	1.086	44	0.546	1.150	70	0.555	1.185
26	0.532	1.096	46	0.547	1.154	75	0.556	1.190
28	0.534	1.105	48	0.548	1.157	80	0.557	1.194
30	0.536	1.112	50	0.548	1.161	85	0.558	1.197
32	0.538	1.119	52	0.549	1.164	90	0.559	1.201
34	0.540	1.126	54	0.550	1.167	95	0.559	1.204
36	0.541	1.131	56	0.551	1.170	100	0.560	1.206
38	0.542	1.136	58	0.552	1.172	$\infty$	0.577	1.282

Теоретические исследования, проведенные Г.А. Алексеевым, показали, что при  $n \rightarrow \infty$  третий центральный момент  $\mu_3 = 2.404$ . Следовательно,

$$C_{s,y} = C_{s,x} = \mu_3 / \sigma_y^3 = 2.404 / (1.282)^3 = 1.14. \quad (2.28)$$

Таким образом, коэффициент асимметрии для распределения Гумбеля является постоянным.

Это обстоятельство и то, что график функции  $P(x)$  при больших значениях обеспеченности уходит в область отрицательных значений, сильно ограничивает возможность применения распределения Гумбеля в практике гидрологических расчетов.

В настоящее время распределение Гумбеля используется главным образом при расчетах максимальных расходов дождевых паводков. Такая практика, в частности, существует в Нидерландах, США, Франции, Японии и ряде других стран.

**Пример 2.5.** Определить максимальный паводочный расход 1%-ной обеспеченности ( $\bar{Q}_{1\%}$ ) на р. Луге в створе ст.Толмачево (см. табл. 1.2), используя закон распределения Гумбеля.

**Решение.**

1. По ряду значений СВ  $Q$  приближенно определяем  $m_Q$  и  $\sigma_Q$   
 $m_Q \approx \bar{x} = 58.4$ ;  $\sigma_Q \approx 37.2$ .
2. По формуле (2.24) или по табл.2.3 определяем  $\hat{y}_p$ ;  $\hat{y}_{1\%} = 4.60$ .
3. В зависимости от длины ряда  $n$  по табл.2.4 определяем  $\bar{y}$  и  $\sigma_y$ .  
Так как в данном случае  $n = 32$ , то  $y = 0.54$ ;  $\sigma_y = 1.12$ .
4. По формулам (2.25) и (2.26) вычисляем параметры  $(1/\alpha)$  и  $q$ :  
 $1/\alpha = \sigma_Q/\sigma_y = 37.2/1.12 = 33.2$ ;  $q = 58.4 - 0.54 \cdot 33.2 = 40.5$ .
5. По формуле (2.23) вычисляем расчетный расход воды:  
 $\hat{Q}_{1\%} = 40.5 + 33.2 \cdot \hat{y}_{1\%} = 40.5 + 33.2 \cdot 4.60 = 193 \text{ м}^3/\text{с}$ .

### 2.5. Распределение Пирсона III типа (непрерывное биномиальное распределение)

Для аппроксимации закона распределения гидрометеорологических величин с одномодальной асимметричной функцией плотности вероятности А. Фостер предложил использовать общее дифференциальное уравнение Пирсона

$$\frac{dy}{dx} = \frac{y(z+a)}{b_0 + b_1z + b_2z^2}, \quad (2.29)$$

где  $Z$  – случайная величина, связанная с исходной СВ  $X$  соотношением  $z = x/m_x - 1 = k - 1$ ;  $k$  – модульный коэффициент;  $y$  – ордината функции плотности вероятности СВ  $Z$ ;  $a$  – расстояние от центра распределения ( $m_k$ ) до моды ( $M_0$ );  $b_0, b_1, b_2$  – параметры, изменяя которые можно получить различные типы кривых распределения (рис. 2.3).

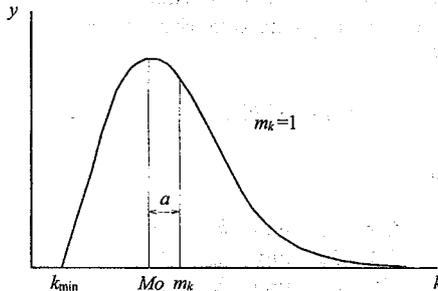


Рис. 2.3. Схема построения функции плотности вероятности Пирсона III типа

для случая  $C_s/C_v > 2$ ; ( $k_{\min} > 0$ ).

В практике гидрологических расчетов наибольшее распространение получила кривая *Пирсона III типа*, для которой  $b_2 = 0$ , и уравнение (2.29) принимает вид

$$\frac{dy}{dz} = \frac{y(z+a)}{b_0 + b_1 z} \quad (2.30)$$

Данному уравнению соответствует целое семейство функций. Для того чтобы  $y(z)$  действительно являлась искомой функцией плотности вероятности, необходимо ввести ряд дополнительных условий:

$$1) \int_{z_{\min}}^{\infty} y(z) dz = 1; \quad 2) y(z) = 0 \text{ при } z = z_{\min}; \quad 3) y(z) = 0 \text{ при } z \rightarrow \infty.$$

Интегрируя (2.30) и переходя от СВ  $Z$  к модульным коэффициентам, после ряда преобразований можно получить выражение для функции плотности вероятности

$$y(k) = \begin{cases} 0 & \text{при } k \leq k_{\min} \\ [\beta^\alpha / \Gamma(\alpha)] (k - k_{\min})^{\alpha-1} \exp[-\beta(k - k_{\min})] & \text{при } k > k_{\min} \end{cases} \quad (2.31)$$

где  $\Gamma(\cdot)$  – гамма-функция;  $\alpha$  и  $\beta$  – параметры распределения, связанные с  $C_v$  и  $C_s$  случайной величины  $X$  соотношениями

$$\alpha = (2/C_s)^2; \quad (3.32)$$

$$\beta = 2/(C_s \cdot C_v). \quad (2.33)$$

Заметим, что коэффициенты вариации и асимметрии не изменяются при замене значений СВ  $X$  на модульные коэффициенты.

Минимальное значение модульного коэффициента в выражении (2.31) определяется формулой

$$k_{\min} = 1 - 2C_v/C_s. \quad (2.34)$$

Из (2.34) следует, что

$$\begin{aligned} C_s &= 2C_v & \text{при } k_{\min} &= 0, \\ C_s &> 2C_v & \text{при } k_{\min} &> 0, \\ C_s &< 2C_v & \text{при } k_{\min} &< 0. \end{aligned}$$

Таким образом, дифференциальная кривая распределения Пирсона III типа при  $C_s = 2C_v$  начинается с нуля; при  $C_s > 2C_v$  — с некоторого положительного числа и, наконец, при  $C_s < 2C_v$  — уходит в область отрицательных чисел.

Мода функции (2.31) при  $C_s < 2$  выражается формулой

$$M_0 = 1 - (C_s C_v) / 2. \quad (2.35)$$

При больших значениях  $C_s$  кривая Пирсона III типа имеет весьма специфический вид. Из (2.32) следует, что при  $C_s = 2$   $\alpha = 1$  и функция (2.31) превращается в экспоненту, а при  $C_s > 2$   $\alpha < 1$  и функция (2.31) превращается в гиперболу.

Зная  $C_v$  и  $C_s$ , можно получить численные значения параметров  $k_{\min}$ ,  $\alpha$ ,  $\beta$  и записать выражение для вычисления обеспеченностей модульных коэффициентов

$$P(k) = \int_k^{\infty} y(s) ds = (\beta^\alpha / \Gamma(\alpha)) \int_k^{\infty} (s - k_{\min})^{\alpha-1} \exp[-\beta(s - k_{\min})] ds, \quad (2.36)$$

где  $s$  — переменная интегрирования.

В случае  $C_s = 2C_v$  имеем:  $k_{\min} = 0$ ,  $\alpha = 1/C_v^2$ ,  $\beta = 1/C_v^2$  и уравнения существенно упрощаются

$$y(k) = [\alpha^\alpha / \Gamma(\alpha)] k^{\alpha-1} \exp(-\alpha k), \quad (2.37)$$

$$P(k) = [\alpha^\alpha / \Gamma(\alpha)] \int_k^{\infty} s^{\alpha-1} \exp(-\alpha s) ds. \quad (2.38)$$

Выражение (2.37) называют двухпараметрическим гамма-распределением, или  $\Gamma$ -распределением.

Кривая обеспеченности Пирсона III типа (2.36) в общем случае является трехпараметрической и однозначно определяется параметрами  $C_v$  и  $C_s$ ; третий параметр —  $m_x$  — необходимо знать для перехода от модульных коэффициентов к значениям СВ  $X$ .

Кривая имеет нижний предел  $k_{\min}$  и не ограничена верхним пределом. При  $C_s \rightarrow 0$ , кривая Пирсона III типа стремится к нормальному распределению.

Численное решение уравнений (2.36), (2.38) — довольно трудоемкая задача, поэтому ординаты кривой обеспеченности Пирсона III типа обычно представляют в виде таблиц (Приложение 1).

В таблицах даны значения нормированной случайной величины в зависимости от обеспеченности и коэффициента асимметрии. Переход от нормиро-

ванных ординат к значениям искомой величины производится по формулам (2.5), (2.5а).

Кривая распределения Пирсона III типа имеет особое значение для гидрологов как первая кривая, широко внедренная в практику гидрологических расчетов. Впервые таблицы распределения Пирсона III типа были опубликованы А. Фостером в 1924 г.

В России приоритет внедрения этой кривой принадлежит Д. Л. Соколовскому (1930 г.). Таблицы А. Фостера были уточнены и расширены С. И. Рыбкиным (1938 г.), а затем и другими авторами.

В настоящее время кривая Пирсона остается одной из наиболее популярных в мировой практике гидрологических расчетов.

Наряду с обычной кривой Пирсона III типа в США и ряде других стран используется логарифмическая кривая Пирсона III типа. В этом случае считается, что распределению Пирсона подчиняется не исходная СВ  $X$ , а ее логарифм  $\ln(X)$ . Такой подход позволяет вывести кривую из отрицательной области при соотношении  $C_s/C_v < 2$ .

**Пример 2.6.** Определить максимальный паводочный расход 1%-ной обеспеченности ( $\dot{Q}_{1\%}$ ) на р. Луге в створе ст. Толмачево (см. табл. 1.2), используя кривую обеспеченностей Пирсона III типа.

**Решение:**

1. По ряду значений СВ  $Q$  приближенно определяем  $m_x \approx \bar{Q} = 58.4$ ;  
 $\sigma_Q \approx 37.2$ ;  $C_v \approx 0.64$ .
2. Поскольку по эмпирическим данным значение  $C_s$  определяется с большой погрешностью, примем соотношение  $C_s/C_v$  осредненным для всего рассматриваемого района. Исследования, проведенные на кафедре гидрологии суши, показали, что для паводочного стока рек Лужского бассейна  $C_s/C_v = 2.5$ . Теперь, зная  $C_v$  и  $C_s/C_v$ , определяем  $C_s$ :  
 $C_s = 2.5 \cdot 0.64 = 1.60$ .
3. По приложению 1 определяем нормированную ординату кривой обеспеченности Пирсона III типа:  $\hat{t}_{1\%} = 3.39$ .
4. По формуле (2.5) переходим от  $\hat{t}_{1\%}$  к  $\dot{Q}_{1\%}$ :  
 $\dot{Q}_{1\%} = \hat{t}_{1\%} \sigma_Q + \bar{Q} = 3.39 \cdot 37.2 + 58.4 = 185$ .

Можно действовать и иначе. Сначала по формуле (2.5а) получаем модульный коэффициент  $k_{1\%} = \hat{t}_{1\%} C_v + 1 = 3.39 \cdot 0.64 + 1 = 3.17$ . Затем, умножая его на  $\bar{x}$ , переходим к расходу:  $\dot{Q}_{1\%} = k_{1\%} \bar{x} = 3.17 \cdot 58.4 = 185 \text{ м}^3/\text{с}$ .

## 2.6. Распределение Крицкого – Менкеля (трехпараметрическое гамма-распределение)

Кривая распределения вероятностей Пирсона III типа, широко используемая в практике гидрологических расчетов, обладает, как было показано выше,

одним существенным недостатком: при  $C_s < 2C_v$  она уходит в область отрицательных значений.

В связи с этим неоднократно предпринимались попытки получения кривой для аппроксимации закона распределения гидрологических величин с диапазоном изменения  $(0 \leq x < \infty)$ , и пригодной для всех реально встречающихся на практике соотношений  $C_s$  и  $C_v$ .

Удачное решение рассматриваемой задачи в середине сороковых годов предложили С.Н. Крицкий и М.Ф. Менкель. В качестве исходной модели они приняли кривую Пирсона III типа при  $C_s = 2C_v$  (двухпараметрическое  $\Gamma$ -распределение):

$$G(z) = (\alpha^\alpha / \Gamma(\alpha)) \int_0^z s^{\alpha-1} e^{-\alpha s} ds, \quad (2.39)$$

где  $G(z)$  – интегральная функция  $\Gamma$ -распределения;  $s$  – переменная интегрирования;  $z = 1$ ;  $C_s = 2C_v$ ;  $\alpha = 1/C_{v,z}$ .

Авторы трансформировали аргумент  $z$  в новую переменную

$$k = az^b, \quad (2.40)$$

где  $a$  и  $b$  – параметры.

При этом предполагалось, что МО новой переменной равно единице, т.е.

$$M[k] = M[az^b] = 1. \quad (2.41)$$

Подставляя в выражение (2.39)  $z = (k/a)^{1/b}$  и имея ввиду, что  $f(k) = dG(k)/dk$ , Крицкий и Менкель получили новое распределение с плотностью вероятности вида

$$f(k) = [\alpha^\alpha / (\Gamma(\alpha) a^{\alpha/b} b)] \exp(-\alpha k^{1/b} a^{-1/b}) k^{(\alpha/b)-1}. \quad (2.42)$$

Начальный момент  $i$ -го порядка этого распределения связан с параметрами  $a$ ,  $a$  и  $b$  соотношением

$$\mu_i = [\Gamma(\alpha + ib) a^i] / [\Gamma(\alpha) \alpha^{ib}]. \quad (2.43)$$

Из выражения (2.43), в частности, следует, что

$$M[k] = \mu_1 = [\Gamma(\alpha + b) a] / [\Gamma(\alpha) \alpha^b]. \quad (2.44)$$

А так как по условию  $M[k] = 1$ , то, приравняв (2.44) единице, получаем выражение для параметра  $a$

$$a = [\Gamma(\alpha)\alpha^b] / \Gamma(\alpha + b). \quad (2.45)$$

Подставляя (2.45) в (2.42), можно получить выражение кривой плотности вероятности Крицкого–Менкеля, записанное через  $\Gamma$ -функцию. Это выражение определяется двумя параметрами  $a$  и  $b$ , которые с учетом (2.43) могут быть выражены через второй и третий начальные моменты. В свою очередь  $\mu_2$  и  $\mu_3$  можно выразить через  $C_v$  и  $C_s$  (см. п. 1.4.).

Таким образом, полученное распределение будет двухпараметрическим, однако для того, чтобы в дальнейшем перейти от модульных коэффициентов  $k$  к искомой СВ  $X$ , необходимо знать третий параметр –  $m_x$ . С учетом сказанного, распределение Крицкого–Менкеля называют также трехпараметрическим  $\Gamma$ -распределением.

Поскольку функцию обеспеченностей Крицкого–Менкеля нельзя выразить через элементарные функции, ее ординаты представляют в виде таблиц (приложение 2). Таблицы составлены в модульных коэффициентах и позволяют определить значение  $k_{p\%}$  в зависимости от  $(C_s/C_v)$ ,  $C_v$  и расчетной обеспеченности  $p$  (%). Впервые такие таблицы были опубликованы Д. В. Коренистым (1948 г.). Наиболее полные таблицы кривой обеспеченности Крицкого–Менкеля даны в [4]. Они охватывают диапазон изменения  $C_s/C_v$  от 0.5 до 6.0 и достаточно широкий диапазон  $C_v$ .

Действующие в настоящее время в России нормативные документы рекомендуют кривую Крицкого–Менкеля в качестве стандартной кривой при проведении гидрологических расчетов.

В заключение еще раз перечислим основные особенности данной кривой.

1. Кривая плотности вероятности является одномодальной с положительной асимметрией.
2. Нижним пределом кривой всегда является нуль.
3. Кривая Крицкого–Менкеля не ограничена верхним пределом.
4. При  $C_s = 2C_v$ , кривая превращается в двухпараметрическое  $\Gamma$ -распределение, т.е. совпадает с кривой Пирсона III типа.

**Пример 2.7.** Определить максимальный расход дождевого паводка 1%-ной обеспеченности  $\bar{Q}_{1\%}$  на р. Луге в створе ст. Толмачево (см. табл. 1.2), используя кривую обеспеченностей Крицкого–Менкеля.

**Решение.**

1. По эмпирическим данным определяем  $m_x \approx \bar{Q} = 58.4$ ;  $C_v \approx 0.64$ ;  $C_s / C_v = 2.5$  (см. пример 2.5).
2. По таблицам (приложение 2) определяем модульный коэффициент 1%-ной обеспеченности. При этом, если таблицы составлены только для целых соотношений  $C_s/C_v$ , необходимо провести интерполяцию.

Для  $C_s/C_v = 2$  при  $C_v = 0.64$  и  $p = 1\%$  имеем  $k_{1\%} = 3.05$ .

Для  $C_s/C_v = 3$  при  $C_v = 0.64$  и  $p = 1\%$  имеем  $k_{1\%} = 3.24$ .

Следовательно, для  $C_s/C_v = 2.5$   $k_{1\%} = 3.15$ .

3. Умножая модульный коэффициент на  $\bar{Q}$ , получаем расчетный расход

$$\hat{Q}_{1\%} = k_{1\%} \cdot \bar{Q} = 3.15 \cdot 58.4 = 184 \text{ м}^3/\text{с}.$$

## 2.7. Распределение Джонсона

В начале 70-х годов XX в. грузинские гидрологи Г.Г. Сванидзе и Г.Л. Григолия предложили использовать для аппроксимации закона распределения гидрологических величин кривую распределения, ограниченную с двух сторон. В качестве такой кривой они рекомендовали  $S_b$ -распределение Джонсона [2, 9].

В рамках этого подхода предполагается, что если исходную СВ  $X$  преобразовать по формуле

$$z = \ln [(x-a)/(b-x)], \quad (2.46)$$

то новая СВ  $Z$  будет иметь нормальное распределение, где  $a$  и  $b$  — соответственно нижний и верхний пределы СВ  $X$ .

Плотность распределения Джонсона имеет вид

$$f(x) = \frac{b-a}{\sigma_z \sqrt{2\pi(x-a)(b-x)}} \exp \left\{ -\frac{1}{2\sigma_z^2} \left[ \ln \left( \frac{x-a}{b-x} \right) - m_x \right]^2 \right\}. \quad (2.47)$$

Как видно из (2.47), это распределение является четырехпараметрическим и помимо параметров  $a$  и  $b$  содержит еще два параметра:  $m_x$  — математическое ожидание СВ  $X$  и  $\sigma_z$  — СКО СВ  $Z$ .

При расчете ординат кривой обеспеченности Джонсона используют таблицы нормированной нормально распределенной СВ  $t$  (см. табл. 2.1), по аналогии с тем, как это делалось в первом варианте для логарифмически нормального распределения (см. пример 2.3), с той лишь разницей, что для расчета параметров  $m_x$  и  $\sigma_z$  (при известных  $a$  и  $b$ ) исходный ряд преобразуется по формуле (2.46). Если все четыре параметра известны, то по табл. 2.1 определяется нормированная ордината  $f_p$ , затем вычисляется  $z_p$  и с учетом (2.46) определяется значение  $x_p$ :

$$x_p = [b \exp(z_p) + a] / [\exp(z_p) + 1]. \quad (2.48)$$

В гидрологической практике ситуации, когда верхний и нижний пределы известны заранее, довольно редки. В качестве примера здесь можно привести кривую обеспеченности коэффициентов паводочного стока, где  $a = 0$ ,  $b = 1$ . Однако гораздо чаще верхний и нижний пределы СВ не известны, поэтому параметры и приходится определять методом последовательных приближений на

основе наилучшего соответствия эмпирической и аналитической кривых обеспеченностей.<sup>1</sup>

**Пример 2.8.** Определить максимальный расход дождевого паводка 1%-ной обеспеченности  $Q_{1\%}$  на р. Луге в створе ст. Толмачево (см. табл. 1.2), используя кривую обеспеченности Джонсона, если параметры  $a$  и  $b$  известны (определены методом последовательных приближений:  $a = 11$ ,  $b = 490$ ).

**Решение.**

1. Производим преобразование исходного ряда по формуле (2.46)  
 $z_i = \ln[(Q_i - 11)/(490 - Q_i)]$ .
2. По ряду значений СВ  $Z$  приближенно определяем  $m_z$  и  $\sigma_z$  (методика оценки параметров распределения по эмпирическим данным излагается в главе 3):  
 $m_z = -2.54$ ;  $\sigma_z = 0.95$ .
3. Так как таблица стандартного нормального распределения (см. табл. 2.1) приводится для интегральной функции распределения, заменяем обеспеченность на вероятность непревышения:  $F(t) = 100 - P(t) = 100 - 1 = 99\%$ .
4. По табл. 2.1 определяем 99%-ный квантиль для СВ  $t$ :  $t_{99} = 2.33$ .
5. По формуле (2.5) определяем  $z_1 = z_{99} = \sigma_z t_{99} + m_z = 0.95 \cdot 2.33 - 2.54 = -0.33$ .
6. По формуле (2.47) определяем  $Q_{1\%}$

$$Q_1 = (490 e^{-0.33} + 11) / (e^{-0.33} + 1) = 363 / 1.72 = 211 \text{ м}^3/\text{с}.$$

В заключение отметим, что, несмотря на всю привлекательность данного распределения (интуитивно мы понимаем, что верхний и нижний пределы существуют), оно не нашло широкого распространения в гидрологической практике, так как очень редко удается надежно оценить параметры  $a$  и  $b$  по эмпирическим данным.

В то же время использование кривых, не ограниченных верхним пределом, не противоречит физической сути гидрологических явлений, так как недостоверно высокие значения СВ (например, расхода) при правильном выборе кривой обеспеченности будут соответствовать очень малой вероятности превышения (менее 0.01 %), т.е. будут практически не вероятными.

По мнению авторов, использование кривой Джонсона может быть эффективным лишь в том случае, когда хотя бы один из параметров ( $a$  или  $b$ ) известен априори или с достаточной точностью может быть назначен, исходя из физических соображений.

## 2.8. Графическое представление функций распределения на клетчатке вероятности

В тех случаях, когда закон распределения исследуемой гидрологической величины не известен, для его аппроксимации используется одно из перечисленных выше распределений.

<sup>1</sup> Об эмпирических кривых обеспеченности см. п. 3.1. Практические приемы определения параметров  $a$  и  $b$  описаны в п. 3.1.5.

Выбранный закон распределения задается функцией обеспеченности, которую в гидрологии принято называть аналитической кривой обеспеченности.

Аппроксимация считается достаточно надежной, если аналитическая кривая хорошо согласуется с эмпирическими данными.<sup>1</sup>

Для удобства проведения такого анализа используются специальные клетчатки вероятности. На клетчатке вероятности по оси абсцисс откладываются значения обеспеченности (в %). По оси ординат могут откладываться либо значения исследуемой СВ, либо ее модульные коэффициенты (1.27), либо ее нормированные значения (1.28).

Масштабы по осям на клетчатке вероятности трансформируются таким образом, чтобы график теоретической кривой обеспеченности представлял собой прямую линию. Очевидно, что клетчатки вероятности могут быть построены для распределений, имеющих только два изменяемых параметра: обычно используются среднее значение и СКО (или коэффициент вариации). Любые дополнительные параметры, такие, как коэффициент асимметрии, должны быть постоянными. Иными словами, для трехпараметрических распределений необходимо иметь спрямляющую клетчатку для каждого соотношения  $C_s/C_v$ .

Наиболее распространенной является клетчатка, спрямляющая нормальный закон распределения. Напомним, что нормальный закон является двухпараметрическим и для него  $C_s = 0$ .

Нормальная клетчатка может также использоваться для изображения кривых обеспеченности с умеренной асимметрией. В этом случае теоретическая кривая полностью не выпрямляется. Кривые с положительной асимметрией будут иметь вогнутую форму, а с отрицательной — выпуклую (Рис. 2.4).

Для построения клетчаток вероятности можно использовать аналитические и графические схемы. Более простым и наглядным является графический прием. Практические приемы построения спрямляющих клетчаток вероятности для различных функций распределения приводятся в работе [8].

При построении клетчатки, спрямляющей нормальный закон распределения, в качестве исходной принимается кривая обеспеченности в модульных коэффициентах с параметрами:  $k = 1$ ,  $C_v = 1$ ,  $C_s = 0$ . Данная клетчатка имеет равномерный масштаб по оси ординат и неравномерный масштаб по оси абсцисс. Если на такой клетчатке построить график нормальной функции обеспеченности с другим значением коэффициента вариации, то этот график все равно будет иметь вид прямой линии, но угол наклона изменится. При  $C_v < 1$  угол наклона уменьшится, при  $C_v > 1$  — увеличится.

В российской гидрологической практике в основном используются нормальные клетчатки (их также называют клетчатками с умеренной асимметричностью) и клетчатки, спрямляющие кривую обеспеченности Крицкого–Менкеля. Образцы этих клетчаток для различных соотношений  $C_s$  и  $C_v$  даны в приложении к Пособию [4].

<sup>1</sup> О технике построения эмпирических кривых обеспеченностей см. п. 3.1.

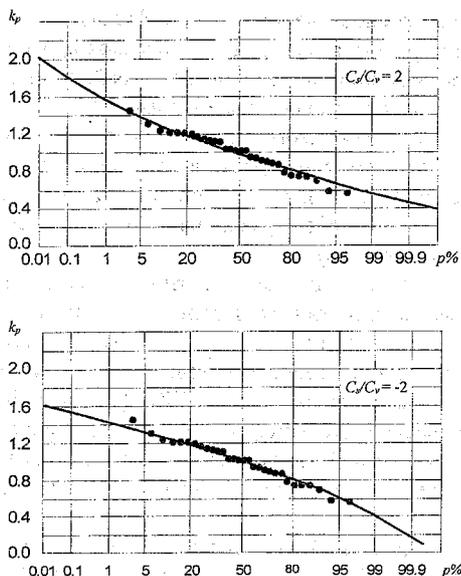


Рис. 2.4. Кривые обеспеченности Пирсона III типа на клетчатке с умеренной асимметричностью, построенные для одного и того же ряда ( $n = 30$ ;  $\bar{Q} = 98.5$ ;  $C_v = 0.22$ ), при различных соотношениях  $C_s/C_v$ .

## 2.9. Комментарии к разделу

В этом разделе были рассмотрены распределения, наиболее часто используемые в практике гидрологических расчетов. Основные характеристики этих функций приводятся в табл. 2.5.

Для аппроксимации законов распределения конкретных гидрологических величин выбор типа функции распределения следует производить с учетом области изменения ее аргумента. Так, для заведомо положительных величин (расход воды, слой осадков и т.д.) наиболее подходящими будут кривые лог-нормального распределения, Крицкого–Менкеля и Пирсона III типа, имеющие нижний предел, но не ограниченные сверху. В то же время, для температуры воздуха или воды в большей степени подходят кривые с диапазоном изменения аргумента  $(-\infty, +\infty)$  и т.д.

Помимо перечисленных распределений в Российской гидрологической литературе [5, 8 и др.] есть ссылки на распределения Р.Д. Гудрича (1927 г.) и Г.Н. Бровковича (1941 г.), однако, в настоящее время эти распределения практически не используются.

Основные характеристики функций распределения,  
используемых в гидрологических расчетах

Тип распределения	Число параметров	Область изменения аргумента	Примечание
Равномерное	2	$[a, b]$	Прямоугольное
Нормальное	2	$(-\infty, +\infty)$	Симметричное
Логнормальное	2	$[0, +\infty)$	Асимметричное
Трехпараметрическое логнормальное	3	$[a, +\infty)$	Асимметричное, $a \geq 0$ при $C_s \geq 3C_v + C_v^3$
Гумбеля	2	$(-\infty, +\infty)$	Асимметричное
Пирсона III типа	3	$[a, +\infty)$	Асимметричное, $a \geq 0$ при $(C_s / C_v) \geq 2$
Крицкого-Менкеля	3	$[0, +\infty)$	Асимметричное
$S_b$ -Джонсона	4	$[a, b]$	Асимметричное, $a \leq x_{\min}; b > x_{\max}$

Из современных разработок можно отметить  $H$ -распределение А. В. Ежова [6]. В основу вывода этого распределения были заложены следующие принципы:

- 1) максимально возможная гибкость функции распределения при соблюдении некоторых естественных ограничений.
- 2) максимальная согласованность состава ограничений с выбираемой функцией распределения.

С позиций теории информации, сказанное можно сформулировать в виде критерия максимума условной энтропии искомого распределения при данных ограничениях.

Полученная А. В. Ежовым на основе этого подхода функция распределения является трехпараметрической с областью изменения аргумента  $[0, +\infty)$ . Распределение обладает рядом положительных свойств. В частности, для него совпадают оценки параметров для метода моментов и метода наибольшего правдоподобия.<sup>1</sup> Однако, как показали исследования, проведенные на кафедре гидрологии суши РГТМУ,  $H$ -распределение накладывает весьма жесткие ограничения на параметры распределения. В частности, для того чтобы функция

<sup>1</sup> Об оценках параметров распределения см. п. 3.2.

плотности вероятности этого распределения была одномодальной и при этом удовлетворяла условию  $f(0) = 0$ , требуется, чтобы  $C_v$  и  $C_s/C_v$  удовлетворяли системе ограничений

$$\begin{cases} 0 < C_v \leq 1.0 \\ 0 < (C_s / C_v) \leq 2 \\ (C_s / C_v) \geq (2.7C_v - 0.7) \end{cases}$$

Кроме того, его ординаты (приложение 3), в интервале реально используемых обеспеченностей весьма незначительно отличаются от ординат кривой Крицкого–Менкеля при тех же значениях  $C_s$  и  $C_v$ .

Все перечисленные выше распределения являются непрерывными. Дискретные распределения в гидрологии используются весьма редко. Здесь можно назвать распределение Пуассона и дискретное биномиальное распределение.

Следует отметить, что распределение Пирсона III типа можно получить из дискретного биномиального распределения на основе предельного перехода, устремив шаг дискретности к нулю и при  $n \rightarrow \infty$  ( $n$  – число опытов). Поэтому распределение Пирсона III типа называют также непрерывным биномиальным распределением.

В заключение приведем сравнительный анализ рассмотренных в этом разделе функций распределения. В табл. 2.6 представлены максимальные паводочные расходы воды различной обеспеченности, полученные с использованием семи типов функций распределения для створа р. Луга – ст. Толмачево.

Как видно из табл. 2.6, выбор типа распределения оказывает ощутимое влияние на значение расчетного расхода. При одной и той же обеспеченности расходы могут существенно различаться. Причем надо иметь в виду, что наиболее заметным это различие становится при больших значениях коэффициента вариации ( $C_v > 0.5$ ).

Сопоставляя кривые, можно отметить, что в интервале обеспеченностей от 10 до 75 % все кривые дают весьма близкие результаты, а кривые Пирсона III типа и Крицкого–Менкеля практически совпадают.

В зоне малых обеспеченностей ( $p < 1$  %) расхождение становится более ощутимым и может достигать нескольких десятков процентов; но наибольшее расхождение кривых наблюдается в зоне больших обеспеченностей. Так, кривая Пирсона III типа (при  $C_s/C_v = 2.5$ ) имеет нижний предел (см. формулу 3.34):  $Q_{\min} = \bar{Q} k_{\min} = \bar{Q}(1 - 2C_v/C_s) = 58.4 \cdot 0.2 = 11.7 \text{ м}^3/\text{с}$  и при любых значениях обеспеченности не может дать расчетный расход ниже этого значения. То же самое в данном случае можно сказать и о кривой Джонсона, для которой нижний предел  $b = 11$ . В то же время кривая Крицкого–Менкеля имеет нулевой нижний предел и дает расчетный расход значительно ниже, чем две предыдущие кривые.

Основные характеристики функций распределения,  
используемых в гидрологических расчетах

Тип распределения	Число параметров	Область изменения аргумента	Примечание
Равномерное	2	$[a, b]$	Прямоугольное
Нормальное	2	$(-\infty, +\infty)$	Симметричное
Логнормальное	2	$[0, +\infty)$	Асимметричное
Трехпараметрическое логнормальное	3	$[a, +\infty)$	Асимметричное, $a \geq 0$ при $C_s \geq 3C_v + C_v^3$
Гумбеля	2	$(-\infty, +\infty)$	Асимметричное
Пирсона III типа	3	$[a, +\infty)$	Асимметричное, $a \geq 0$ при $(C_s / C_v) \geq 2$
Крицкого-Менкеля	3	$[0, +\infty)$	Асимметричное
$S_b$ -Джонсона	4	$[a, b]$	Асимметричное, $a \leq x_{\min}; b > x_{\max}$

Из современных разработок можно отметить  $H$ -распределение А. В. Ежова [6]. В основу вывода этого распределения были заложены следующие принципы:

- 1) максимально возможная гибкость функции распределения при соблюдении некоторых естественных ограничений.
- 2) максимальная согласованность состава ограничений с выбираемой функцией распределения.

С позиций теории информации, сказанное можно сформулировать в виде критерия максимума условной энтропии искомого распределения при данных ограничениях.

Полученная А. В. Ежовым на основе этого подхода функция распределения является трехпараметрической с областью изменения аргумента  $[0, +\infty)$ . Распределение обладает рядом положительных свойств. В частности, для него совпадают оценки параметров для метода моментов и метода наибольшего правдоподобия.<sup>1</sup> Однако, как показали исследования, проведенные на кафедре гидрологии суши РГГМУ,  $H$ -распределение накладывает весьма жесткие ограничения на параметры распределения. В частности, для того чтобы функция

<sup>1</sup> Об оценках параметров распределения см. п. 3.2.

плотности вероятности этого распределения была одномодальной и при этом удовлетворяла условию  $f(0) = 0$ , требуется, чтобы  $C_v$  и  $C_s/C_v$  удовлетворяли системе ограничений

$$\begin{cases} 0 < C_v \leq 1.0 \\ 0 < (C_s/C_v) \leq 2 \\ (C_s/C_v) \geq (2.7C_v - 0.7). \end{cases}$$

Кроме того, его ординаты (приложение 3), в интервале реально используемых обеспеченностей весьма незначительно отличаются от ординат кривой Крицкого–Менкеля при тех же значениях  $C_s$  и  $C_v$ .

Все перечисленные выше распределения являются непрерывными. Дискретные распределения в гидрологии используются весьма редко. Здесь можно назвать распределение Пуассона и дискретное биномиальное распределение.

Следует отметить, что распределение Пирсона III типа можно получить из дискретного биномиального распределения на основе предельного перехода, устремив шаг дискретности к нулю и при  $n \rightarrow \infty$  ( $n$  – число опытов). Поэтому распределение Пирсона III типа называют также непрерывным биномиальным распределением.

В заключение приведем сравнительный анализ рассмотренных в этом разделе функций распределения. В табл. 2.6 представлены максимальные паводочные расходы воды различной обеспеченности, полученные с использованием семи типов функций распределения для створа р. Луга – ст. Толмачево.

Как видно из табл. 2.6, выбор типа распределения оказывает ощутимое влияние на значение расчетного расхода. При одной и той же обеспеченности расходы могут существенно различаться. Причем надо иметь в виду, что наиболее заметным это различие становится при больших значениях коэффициента вариации ( $C_v > 0.5$ ).

Сопоставляя кривые, можно отметить, что в интервале обеспеченностей от 10 до 75 % все кривые дают весьма близкие результаты, а кривые Пирсона III типа и Крицкого–Менкеля практически совпадают.

В зоне малых обеспеченностей ( $p < 1\%$ ) расхождение становится более ощутимым и может достигать нескольких десятков процентов; но наибольшее расхождение кривых наблюдается в зоне больших обеспеченностей. Так, кривая Пирсона III типа (при  $C_s/C_v = 2.5$ ) имеет нижний предел (см. формулу 3.34):  $Q_{\min} = \bar{Q} k_{\min} = \bar{Q}(1 - 2C_v/C_s) = 58.4 \cdot 0.2 = 11.7 \text{ м}^3/\text{с}$  и при любых значениях обеспеченности не может дать расчетный расход ниже этого значения. То же самое в данном случае можно сказать и о кривой Джонсона, для которой нижний предел  $b = 11$ . В то же время кривая Крицкого–Менкеля имеет нулевой нижний предел и дает расчетный расход значительно ниже, чем две предыдущие кривые.

Максимальные расходы дождевых паводков различной обеспеченности  
на р. Луге – ст. Толмачево

Распределение	Обеспеченность, $p$ %								
	0.1	1	10	25	50	75	90	99	99.9
Нормальное	172	145	106	83.3	58.4	33.5	10.8	-28.3	-55.1
Логнормальное									
1-й вариант	338	213	109	73.7	47.9	31.2	21.1	10.8	6.81
2-й вариант	290	191	104	73.0	49.4	33.4	23.5	12.8	8.42
Трехпараметрическое логнормальное	267	187	106	75.7	50.5	32.2	19.8	5.26	-1.28
Гумбеля	269	193	115	82.5	52.3	29.4	12.9	-10.3	-23.6
Пирсона III типа ( $C_s/C_v = 2.5$ )	258	185	108	75.5	49.1	31.2	21.6	13.8	12.3
Крицкого-Менкеля ( $C_s/C_v = 2.5$ )	266	184	108	77.7	49.9	31.0	19.6	8.94	4.55
$S_b$ -Джонсона ( $a = 11, b = 490$ )	293	211	112	73.1	46.0	30.2	21.9	15.1	13.1

Нормальная кривая и кривая Гумбеля не ограничены нижним пределом и при больших значениях обеспеченностей вообще уходят в область отрицательных значений. Чтобы избежать разночтений при аппроксимации функций распределения гидрологических величин, в России введен стандарт, закрепленный в нормативных документах. В качестве стандартной кривой обеспеченности рекомендуется использовать кривую Крицкого-Менкеля. Однако при достаточном обосновании допускается применять кривую распределения Пирсона III типа или другие распределения.

### 3. Построение кривых обеспеченностей и оценка параметров распределения по эмпирическим данным

При решении многих практических задач функция распределения СВ не может быть определена теоретическим путем. В таких случаях используются результаты наблюдений за СВ, позволяющие (при достаточном их числе и надлежащей обработке) определить с известной степенью достоверности вид функции распределения и оценить ее числовые характеристики. Для решения такого рода задач используются методы математической статистики.

Известный в этой области математик А. Уайльд определяет математическую статистику как совокупность методов, которые дают нам возможность принимать оптимальные решения в условиях неопределенности.

В качестве более строгого определения можно привести следующее: математическая статистика – раздел математики, посвященный математическим методам систематизации, обработки и использования статистических данных для научных и практических выводов.

Фундаментальными понятиями статистической теории являются понятия *генеральной совокупности* и *выборки*.

*Генеральная совокупность* обычно интерпретируется как совокупность всех возможных значений случайной величины, которые в принципе могут иметь место при данных условиях.

Содержательный смысл этого понятия состоит в том, что предполагается существование некоторых вполне определенных свойств, неслучайных закономерностей, присущих данной совокупности, тех свойств, которые и должны быть выяснены исследователем. Фактически эти свойства являются объективным отображением вероятностных свойств изучаемого объекта, которые могут быть охарактеризованы с помощью соответствующих законов распределения вероятностей или связанных с ними числовых параметров.

Считается, что указанные свойства не изменяются во времени и присущие генеральной совокупности неслучайные закономерности сохраняют постоянным свой характер, т.е. являются устойчивыми.

*Выборка* – это конечный набор значений случайной величины, полученный в результате наблюдений.

В гидрологии в качестве выборок можно рассматривать ряды наблюдений за среднегодовыми расходами воды, максимальными и минимальными расходами и уровнями воды, среднегодовой, максимальной и минимальной температурой воздуха и т.д.

Число элементов выборки называется ее *объемом*.

Выборка называется *репрезентативной*, если она достаточно полно характеризует генеральную совокупность.

Смысл статистических методов заключается в том, чтобы по выборке ограниченного объема  $n$ , т.е. по некоторой части генеральной совокупности, высказать обоснованное суждение о ее свойствах в целом.

Такое суждение можно составить, произведя специальную обработку результатов наблюдений. В частности, можно построить эмпирическую кривую распределения (см. п. 3.1) и вычислить числовые характеристики эмпирического распределения (выборки).

В силу случайности выборки ее числовые характеристики являются случайными величинами, отличаясь этим от достоверных числовых характеристик теоретического распределения, которому подчиняется генеральная совокупность.

Полученные по эмпирическим данным числовые характеристики случайной величины  $X$  (такие, как  $m_x^*$ ,  $\sigma_x^*$ ,  $D_x^*$  и т. д.) принято называть статистическими оценками числовых характеристик (или эмпирическими, или выборочными характеристиками)<sup>1</sup>.

При оценивании одного и того же параметра в зависимости от применяемой методики можно получить различные оценки. Для того чтобы выбрать наилучшую из них, необходимо сформулировать некоторые требования к свойствам оценок. Наиболее важными из этих свойств являются состоятельность, несмещенность и эффективность.

**Состоятельность.** Оценка  $G^* = f(x_1, x_2, x_3, \dots, x_n)$  неизвестного параметра  $G$  называется состоятельной, если по мере роста числа наблюдений  $n$  она стремится по вероятности к оцениваемому значению  $G$ , т.е.

$$\lim_{n \rightarrow \infty} P \{ |G - G^*| < \varepsilon \} = 1, \quad (3.1)$$

где  $\varepsilon$  — сколь угодно малое число.

Состоятельность оценки гарантирует исследователю увеличение точности оценивания с ростом  $n$  и то, что, хотя бы в пределе (при  $n \rightarrow \infty$ ) он может получить точное значение  $G$ .

**Несмещенность.** Оценка  $G^* = f(x_1, x_2, x_3, \dots, x_n)$  неизвестного параметра  $G$  называется несмещенной, если при любом объеме выборки  $n$  результат ее осреднения по всем возможным выборкам данного объема приводит к точному (истинному) значению оцениваемого параметра, т.е.  $M[G^*] = G$ .

Говоря другими словами, несмещенность означает отсутствие систематической погрешности при оценивании параметра.

Если оценка является смещенной, но причина смещенности известна, то ее можно устранить, введя поправку на смещенность в формулу оценки. Полученная на основе такой корректировки оценка будет сходиться к истинному значению оцениваемого параметра только при  $n \rightarrow \infty$ , т.е. будет являться асимптотически несмещенной.

Следует отметить, что асимптотически несмещенные оценки не обязательно более точные, чем смещенные, однако они являются более удобными с матема-

<sup>1</sup> Для обозначения статистических оценок будет использоваться символ (\*):  $\sigma_x^*$ ,  $D_x^*$  и т. д.

тической точки зрения (состоятельность, инвариантность относительно линейных преобразований и т.д.).

**Эффективность.** Оценка  $G^* = f(x_1, x_2, x_3, \dots, x_n)$  называется эффективной, если среди всех оценок параметра  $G$  она обладает наименьшей мерой случайного разброса относительно истинного значения оцениваемого параметра, т.е.  $D[G^*] = D_{\min}$ .

Эффективная оценка, следовательно, имеет минимальную случайную погрешность и в этом смысле является наиболее точной.

Конкретные методы расчета оценок числовых характеристик СВ рассматриваются в п. 3.2.

### 3.1. Эмпирические кривые обеспеченностей

В практике строительного, дорожного и водохозяйственного проектирования часто требуется определить расход воды (или иную гидрологическую характеристику) заданной обеспеченности, или, что то же самое, заданной вероятности ежегодного превышения.

Такая задача легко решается, если закон распределения случайной величины известен. Однако в большинстве практических ситуаций закон распределения исследуемой СВ не известен и представление о нем необходимо составить по эмпирическим данным.

Функция обеспеченностей, построенная по эмпирическим данным, называется *эмпирической кривой обеспеченностей*.

Возможны два способа построения эмпирической кривой обеспеченностей. Первый из них можно рекомендовать в том случае, когда имеется достаточно продолжительный ряд значений СВ.

Рассмотрим первый способ на примере данных наблюдений за среднегодовыми расходами на р. Днепре у пгт Лопманская Каменка с 1818 по 1962 г. (табл. 3.1).

Определяем амплитуду (размах  $R$ ) колебаний среднегодовых расходов как разницу между наибольшим ( $Q_{\max}$ ) и наименьшим ( $Q_{\min}$ ) значениями расходов:  $R = Q_{\max} - Q_{\min} = 3040 - 717 = 2323 \text{ м}^3/\text{с}$ .

Разбиваем амплитуду на  $k$  равных интервалов. Величину  $k$  можно приблизительно оценить по формуле

$$k \approx 5 \lg(n). \quad (3.2)$$

В данном случае  $n = 145$ , и следовательно,  $k = 5 \lg(145) = 11$ . Таким образом, примем в качестве первого приближения  $k = 11$ .

Определяем длину расчетного интервала:  $l = R/k = 2323/11 = 211$ . Для удобства расчета величину  $l$  можно округлить (но так, чтобы значение  $l$  изменилось не более чем на 10 – 15 %). Окончательно принимаем  $l = 200$ .

В качестве левой границы первого интервала примем значение, большее или равное  $Q_{\max}$ . С учетом того, что в данном случае значение  $l$  округлено до сотен, в качестве такой границы принимаем 3100. Тогда значение, соответствующее правой границе первого интервала, будет равно  $3100 - 200 = 2900$ .

Последующие интервалы будут иметь границы: 2900 – 2700; 2700 – 2500; 2500 – 2300 и т.д.

Таблица 3.1  
Среднегодовые расходы воды р. Днепра у пгт Лоцманская Каменка

Год	Q	Год	Q	Год	Q	Год	Q
1818	1480	1855	1900	1891	1470	1927	1730
1819	1600	1856	1500	1892	1120	1928	1840
1820	2400	1857	1300	1893	1820	1929	1910
1821	1740	1858	1400	1894	1460	1930	1180
1822	1370	1859	1000	1895	2420	1931	2480
1823	1200	1860	1800	1896	1990	1932	2620
1824	1150	1861	2400	1897	1590	1933	2630
1825	1800	1862	1400	1898	1190	1934	1860
1826	1200	1863	900	1899	1220	1935	1620
1827	1300	1864	1000	1900	1670	1936	1560
1828	2080	1865	1700	1901	1320	1937	1620
1829	2500	1866	1600	1902	1690	1938	1490
1830	2100	1867	2000	1903	1620	1939	1080
1831	2060	1868	2000	1904	1080	1940	1680
1832	1550	1869	1200	1905	1740	1941	2300
1833	1340	1870	1400	1906	2060	1942	2470
1834	1750	1871	2000	1907	2100	1943	1010
1835	1000	1872	1780	1908	2220	1944	1380
1836	1030	1873	1100	1909	1700	1945	1480
1837	1680	1874	1520	1910	1160	1946	1310
1838	1940	1875	900	1911	1310	1947	1530
1839	1540	1876	2100	1912	1890	1948	1660
1840	1930	1877	3040	1913	1800	1949	1290
1841	1930	1878	1960	1914	1610	1950	1060
1842	870	1879	2450	1915	1820	1951	1480
1843	1100	1880	1640	1916	2050	1952	1240
1844	2000	1881	1670	1917	2400	1953	1970
1845	3000	1882	1060	1918	1300	1954	874
1846	1000	1883	2070	1919	1910	1955	1520
1847	1400	1884	1340	1920	1480	1956	1740
1848	1200	1885	1160	1921	717	1957	1500
1849	2400	1886	1550	1922	1730	1958	2420
1850	1800	1887	1330	1923	1710	1959	1320
1851	1740	1888	2060	1924	1820	1960	1250
1852	1650	1889	1940	1925	874	1961	1330
1853	2100	1890	1250	1926	1890	1962	1510
1854	1800						

Для того чтобы границы интервалов не пересекались, проведем корректировку левых границ всех интервалов. В результате получим: 3099 – 2900; 2899 – 2700; 2699 – 2500 и т. д. Дальнейший расчет сведен в табл. 3.2.

Таблица 3.2

## Сгруппированные данные годового стока р. Днепра у пгт Лоцманская Каменка

Интервал расходов, $Q$ м <sup>3</sup> /с	Число случаев попадания в интервал (повторяемость, $m_i$ )	Относительная частота, $p^*$ %	Число превышений правой границы интервала	Эмпирическая обеспеченность, %
3099 - 2900	2	1.38	2	1.38
2899 - 2700	0	0	2	1.38
2699 - 2500	3	2.07	5	3.45
2499 - 2300	10	6.90	15	10.4
2299 - 2100	5	3.45	20	13.8
2099 - 1900	20	13.80	40	27.6
1899 - 1700	23	15.86	63	43.5
1699 - 1500	25	17.24	88	60.7
1499 - 1300	24	16.56	112	77.3
1299 - 1100	17	11.73	129	89.0
1099 - 900	12	8.28	141	97.3
899 - 700	4	2.76	145	100
сумма	145	100		

Из таблицы видно, что для покрытия амплитуды колебаний расходов воды в расчетном створе потребовалось (при  $l = 200$ ) не 11, а 12 интервалов (что не имеет принципиального значения).

Во второй графе табл. 3.2 приводим число расходов воды попавших в каждую градацию

В третьей графе даем относительную частоту попадания расходов воды в каждый интервал:  $p^* = (m_i / n) 100\%$ , где  $m_i$  – число значений расхода воды в  $i$ -том интервале (повторяемость).

Четвертую графу рассчитываем, последовательно суммируя данные второй графы. Первое из полученных чисел будет соответствовать числу случаев, когда расход воды равен или превышает 2900 м<sup>3</sup>/с, второе – когда расход равен или превышает 2700 м<sup>3</sup>/с, третье – когда расход равен или превышает 2500 м<sup>3</sup>/с и т. д.

В пятой графе число соответствующих превышений выразим в процентах от общего числа наблюдений.

По данным третьей графы можно построить график эмпирических частот, по данным пятой графы – график эмпирической кривой обеспеченностей (рис.3.1).

В гидрологии значения СВ принято откладывать по вертикальной оси, а значения вероятности  $p$  – по горизонтальной. Для гистограммы эмпирических частот  $p$  следует трактовать как вероятность попадания в заданный интервал, а для кривой обеспеченностей – как вероятность превышения соответствующего значения СВ.

В дополнение к сказанному можно сделать три замечания.

1. Если ступенчатый график эмпирических частот сгладить плавной линией, то мы получим эмпирическую функцию плотности вероятности.
2. Если суммирование эмпирических частот по интервалам вести не от больших значений к меньшим, а наоборот, то вместо кривой обеспеченностей мы получим интегральную кривую распределения. В зарубежной литературе эту кривую называют кривой накопленных частот.
3. В ряде случаев при построении кривых обеспеченностей по оси абсцисс удобнее откладывать не значения СВ, а ее нормированные значения или модульные коэффициенты.

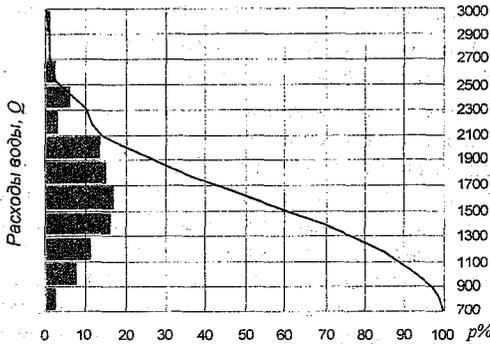


Рис. 3.1. Гистограмма эмпирических частот и эмпирическая кривая обеспеченностей годового стока р. Днепра у пгт Лоцманская Каменка.

Хотя изложенный способ построения эмпирической кривой обеспеченностей является довольно простым и наглядным, в практике гидрологических расчетов он применяется крайне редко, так как реальные ряды наблюдений за гидрологическими характеристиками, как правило, не превышают нескольких десятков лет.

При анализе рядов небольшого объема используется иной метод построения эмпирической кривой обеспеченности. При использовании этого метода члены эмпирического ряда ранжируются, т.е. располагаются в убывающем или возрастающем порядке. В гидрологии принято располагать члены ряда в убывающем порядке.

Допустим, имеется ряд величин какой-либо характеристики гидрологического режима, расположенных в убывающем порядке:

$x_1 > x_2 > x_3 > \dots > x_m > \dots > x_N$ . Тогда теоретическая вероятность превышения для  $m$ -го члена ряда может быть выражена формулой

$$P\{X \geq x_m\} = \lim_{N \rightarrow \infty} (m/N). \quad (3.3)$$

В то же время длина реальной выборки всегда конечна,  $n < \infty$ . Заменяя  $N$  на  $n$  и используя формулу (3.3), можно попытаться приблизительно оценить вероятность превышения для каждого члена имеющейся выборки

$$\hat{p}_m = P\{X \geq x_m\} \approx (m/n)100 \%, \quad (3.4)$$

где  $m$  – порядковый номер  $x_m$  в ранжированном ряду;  $\hat{p}_m$  – обеспеченность (в %)  $m$ -го члена ранжированного ряда.

В соответствии с формулой (3.4) обеспеченность первого (самого большого) члена ранжированного ряда будет равна  $\hat{p}_1 = (1/n)100 \%$ , второго –  $\hat{p}_2 = (2/n)100 \%$  и т. д. Обеспеченность последнего члена ряда будет равна  $\hat{p}_n = (n/n)100 = 100 \%$ . Таким образом, если принять данную формулу, получится, что последний член ранжированного ряда представляет собой абсолютный минимум и СВ  $X$  никогда не примет значение меньше  $x_n$ .

Данный парадокс возникает в связи с тем, что мы заменили  $N$  на  $n$ . В действительности можно получить бесконечное множество выборок из генеральной совокупности длиной  $n$ , каждая из которых будет иметь свой максимум и свой минимум. В этом смысле эмпирическая обеспеченность  $m$ -го члена ранжированного ряда сама будет являться случайной величиной, и в качестве расчетного значения разумно принять ее математическое ожидание, моду или иную устойчивую характеристику. С учетом сказанного в настоящее время разработано около десятка формул для расчета эмпирической обеспеченности:

Формула А. Хазена  $\hat{p}_m = [(m - 0.5)/n]100 \%; \quad (3.5)$

Формула С.Н. Крицкого – М.Ф. Менкеля (Вейбула)

$$\hat{p}_m = [m/(n+1)]100 \%; \quad (3.6)$$

Формула Н.Н. Чегодаева

$$\hat{p}_m = [(m - 0.3)/(n + 0.4)]100 \%. \quad (3.7)$$

Формула (3.5) предполагает замену ступенчатого графика эмпирических частот сглаженной кривой, проходящей через середины ступенек графика. Обеспеченность первого члена ряда по зависимости (3.5) составит  $(1/2n)100 \%$ . Такая оценка недостаточно обоснована, поскольку в этом случае повторяемость наблюдаемого максимума относится к периоду, вдвое превышающему период наблюдений.

Формула (3.6) соответствует математическому ожиданию эмпирической обеспеченности; а формула (3.7) – медианному значению эмпирической обеспеченности.

Среди перечисленных формул в определенном смысле наилучшей является формула (3.6), так как получаемая по ней оценка эмпирической обеспеченности является состоятельной, несмещенной и эффективной.

Однако при выводе формулы эмпирической обеспеченности допустимо рассуждать и по-иному. Значение СВ, расположенное в ранжированном ряду (длиной  $n$ ) на  $m$ -м месте, тоже является случайной величиной, и можно считать обеспеченность, соответствующую МО или моде  $m$ -го члена. Но в такой постановке задача разрешима лишь в том случае, когда закон распределения исходной СВ  $X$  известен. Приведем примеры таких формул.

Для распределения Гумбеля, а также экспоненциального распределения

$$\hat{p}_m = [(m - 0.44)/(n + 0.12)] 100 \% , \quad (3.8)$$

для нормального распределения

$$\hat{p}_m = [(m - 3/8)/(n + 1/4)] 100 \% . \quad (3.9)$$

Таким образом, различие этих двух подходов состоит в том, что в первом случае в качестве СВ мы рассматриваем обеспеченность  $m$ -го члена ранжированного ряда, а во втором в качестве СВ рассматривается значение  $m$ -го члена и затем вычисляется обеспеченность его МО или моды. Следует отметить, что оба подхода при выводе формул эмпирической обеспеченности дают одинаковые результаты лишь в том случае, если теоретическая кривая обеспеченности представляет собой линейную зависимость СВ  $X$  от обеспеченности, как это имеет место, например, для закона распределения равномерной плотности.

Учитывая сказанное, можно констатировать, что выбор формулы для расчета эмпирической обеспеченности содержит элемент субъективизма и зависит от традиций и специфики решаемых задач. Так, например, в гидротехнической практике за критерий надежности работы сооружений обычно принимается ежегодная вероятность нарушения нормальной работы установки. В этих условиях наиболее приемлемой при расчетах паводочного стока является формула (3.6), поскольку она дает большую надежность, особенно для малых значений обеспеченности. В США эта формула рекомендуется в качестве основной расчетной формулы.

В то же время, например, во Франции используется формула (3.5), а в Англии – формула (3.8).

Грингортен (1963 г.) предложил универсальную формулу для расчета эмпирической обеспеченности

$$\hat{p}_m = [(m - a)/(n + 1 - 2a)] 100 \% . \quad (3.10)$$

При  $a = 0.5$  мы получаем формулу (3.5), при  $a = 0.3$  – формулу (3.7), при  $a = 0$  – формулу (3.6), при  $a = 0.44$  – формулу (3.8) и при  $a = 3/8$  – формулу (3.9).

Сам же Грингортен рекомендовал устанавливать значение параметра  $a$  в зависимости от длины выборки по табл.3.3.

Таблица 3.3

Зависимость параметра  $a$  от длины выборки  $n$  по Грингортену

$n \dots$	10	20	30	40	50	80	90	100
$a \dots$	0.448	0.443	0.442	0.441	0.440	0.440	0.439	0.439

В России действующие в настоящее время нормативные документы рекомендуют в качестве основной расчетной формулы выражение (3.6).

**Пример 3.1.** Построить эмпирическую кривую обеспеченностей максимальных паводочных расходов для р. Луга – ст. Толмачево (см. табл. 1.2).

**Решение.**

1. Ранжируем исходный ряд, т.е. располагаем расходы в убывающем порядке (Табл. 3.4)
2. Для каждого расхода воды в ранжированном ряду рассчитываем эмпирическую обеспеченность по формуле (3.6). Так как длина ряда  $n = 32$ , обеспеченность первого члена ранжированного ряда равна  $\hat{p}_1 = (1/33) 100\% = 3.03\%$ ; второго –  $\hat{p}_2 = (2/33) 100\% = 6.06\%$  и т. д.
3. По данным граф 5 и 6 на клетчатке вероятности строим эмпирическую кривую обеспеченностей (Рис. 3.2). Масштаб по оси ординат рекомендуется выбирать так, чтобы точка, соответствующая максимуму эмпирической кривой, по оси ординат, располагалась не слишком близко к верхнему краю клетчатки.

Как видно из приведенного примера, эмпирическая кривая обеспеченностей позволяет составить представление о распределении исследуемой СВ в диапазоне обеспеченностей от  $100/(n+1)$  до  $(100n)/(n+1)\%$ , в данном случае от 3.03 до 97.0%. В то же время довольно часто требуется рассчитать значения расходов для обеспеченностей, которые гораздо меньше или гораздо больше указанных. Например, при проектировании гидротехнических сооружений первого класса требуется иметь расход обеспеченностью 0.01%. Для решения подобных задач необходимо уметь экстраполировать эмпирическую кривую в область больших и малых значений. Очевидно, что экстраполяция на глаз здесь не возможна, поскольку результаты такой экстраполяции будут носить слишком субъективный характер.

В практике гидротехнических расчетов для сглаживания и экстраполяции эмпирических кривых обеспеченностей используются аналитические кривые из числа перечисленных в главе 2. Выбранная аналитическая кривая по отношению к эмпирической кривой выполняет роль своего рода гидрологического лекала.

Расчет ординат эмпирической кривой обеспеченностей максимальных расходов дождевых паводков р. Луги – ст. Толмачево ( $\bar{Q} = 58.4 \text{ м}^3/\text{с}$ )

$m_i$	Годы	Расходы, $\text{м}^3/\text{с}$	Ранжированные		Обеспеченность, %
			расходы, $\text{м}^3/\text{с}$	модульные коэффициенты	
1	2	3	4	5	6
1	1954	71.4	145	2.48	3.03
2	1955	22.1	130	2.22	6.06
3	1956	32.7	130	2.22	9.09
4	1957	145	121	2.07	12.1
5	1958	46.0	104	1.78	15.2
6	1959	29.9	100	1.71	18.2
7	1960	28.4	84.5	1.45	21.2
8	1961	73.8	84.4	1.44	24.2
9	1962	121	78.1	1.34	27.3
10	1963	25.0	73.8	1.26	30.3
11	1964	31.4	71.4	1.22	33.3
12	1965	17.3	66.0	1.13	36.4
13	1966	54.8	58.7	1.00	39.4
14	1967	29.9	58.2	1.00	42.4
15	1968	104	54.8	0.94	45.5
16	1969	27.5	54.0	0.92	48.5
17	1970	84.5	46.0	0.79	51.5
18	1971	37.5	39.2	0.67	54.6
19	1972	24.0	37.5	0.64	57.6
20	1973	22.1	32.7	0.56	60.6
21	1974	130	31.4	0.54	63.6
22	1975	21.3	29.9	0.51	66.7
23	1976	58.2	29.9	0.51	69.7
24	1977	58.7	28.4	0.49	72.7
25	1978	130	27.5	0.47	75.8
26	1979	22.0	25.0	0.43	78.8
27	1980	54.0	24.0	0.41	81.8
28	1981	100	22.1	0.38	84.9
29	1982	78.1	22.1	0.38	87.9
30	1983	39.2	22.0	0.38	90.9
31	1984	66.0	21.3	0.36	93.9
32	1985	84.4	17.3	0.30	97.0

В такой постановке задача гидролога сводится к решению двух основных задач:

1. произвести выбор аналитической кривой;
2. оценить по эмпирическим данным параметры распределения (см. п. 3.2).  
Обычно требуется оценить  $MO$ ,  $SKO$  (или  $C_v$ ),  $C_s$  (или  $C_s/C_v$ ).

Если эта задача решена, то на клетчатку вероятности помимо эмпирической кривой обеспеченностей наносится и аналитическая кривая (см. рис. 3.2), что

позволяет проконтролировать степень согласованности аналитической кривой и эмпирических точек.

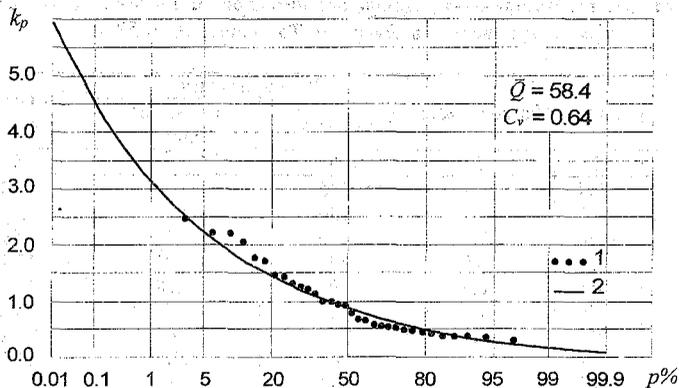


Рис.3.2. Эмпирическая кривая обеспеченностей (1) и аналитическая кривая обеспеченностей Кривцкого-Менкеля при  $C_s/C_v = 2.5$  (2) максимальных расходов дождевых паводков на р. Луге – ст. Толмачево.

При построении графика аналитической кривой на клетчатку вероятности наносят 20 – 25 точек (координаты точек определяют по соответствующим таблицам) и затем соединяют точки плавной линией.

Если аналитическая кривая хорошо согласуется с эмпирической кривой обеспеченностей, то в дальнейших расчетах используются ординаты аналитической кривой.

При массовых расчетах нет необходимости строить графики эмпирической и аналитической кривых обеспеченностей для каждого створа. Можно провести такой анализ только для нескольких опорных пунктов. И если вопрос о выборе типа аналитической кривой обеспеченностей решен, то для остальных створов достаточно по эмпирическим данным оценить параметры распределения (см. п. 3.2.) и выписать из таблиц ординаты аналитической кривой для тех обеспеченностей, которые диктуются условиями конкретной задачи.

### 3.2. Методы расчета оценок параметров распределения

При статистической обработке гидрологических данных используются несколько методов расчета оценок параметров распределения. Эти методы можно условно разделить на три группы: аналитические, графоаналитические и графические.

Графические и графоаналитические методы, как видно из названия, сочетают в себе элементы аналитического расчета и графические построения. В аналитических методах оценки параметров распределения представляют собой числа, полученные путем подстановки выборочных значений  $x_1, x_2, x_3, \dots, x_n$  СВ  $X$  в теоретическую формулу оцениваемого параметра. При этом желательно,

чтобы оценка удовлетворяла требованиям состоятельности, несмещенности и эффективности.

К числу аналитических методов относятся метод моментов и метод наибольшего правдоподобия.

### 3.2.1. Метод моментов

Метод основан на использовании моментов эмпирического распределения, которые являются состоятельными оценками соответствующих теоретических моментов (см. п. 1.4).

При замене теоретических моментов эмпирическими вместо  $N \rightarrow \infty$  используется конечное число значений случайной величины  $n$ , а вероятность ( $p_i$ ) того, что СВ  $X$  примет значение  $x_i$  заменяется частотой ( $p_i = 1/n$ ).

Эмпирические начальный ( $\alpha_s^*$ ) и центральный ( $\mu_s^*$ ) моменты  $S$ -го порядка определяются формулами:

$$\alpha_s^* = \frac{1}{n} \sum_{i=1}^n x_i^s, \quad (3.11)$$

$$\mu_s^* = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^s. \quad (3.12)$$

Как следует из (3.11), выборочную оценку математического ожидания (выборочное среднее  $\bar{x}$ ) можно вычислить по формуле

$$m_s^* = \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i. \quad (3.13)$$

Эта оценка является состоятельной и несмещенной.

Для обозначения выборочной дисперсии наряду с  $D^*$ , используется символ  $S^2$ . Выборочная дисперсия представляет собой второй центральный момент и, как следует из (3.12) определяется выражением

$$D^* = S^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2. \quad (3.14)$$

Так как математическое ожидание  $D^*$  равно не  $D$ , а величине  $[(n-1)/n]D$ , оценка (3.14) является состоятельной, но смещенной оценкой теоретической дисперсии. В связи с этим (особенно при  $n < 40$ ) рекомендуется применять несмещенную состоятельную оценку дисперсии

$$S_H^2 = \frac{n}{n-1} S_C^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2, \quad (3.15)$$

где  $S_H^2$  и  $S_C^2$  – соответственно несмещенная и смещенная оценки дисперсии;  $n/(n-1)$  – поправка на смещенность. Отсюда несмещенная оценка СКО и коэффициента вариации

$$\sigma^* = S = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}, \quad (3.16)$$

$$C_v^* = \frac{S}{\bar{x}} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (k_i - 1)^2}, \quad (3.17)$$

где  $k_i = x_i / \bar{x}$  – модульный коэффициент.

Оценка коэффициента асимметрии определяется выражением

$$C_s^* = \frac{\mu_3^*}{S^3} = \frac{\sum_{i=1}^n (x_i - \bar{x})^3}{nS^3} = \frac{\sum_{i=1}^n (k_i - 1)^3}{n(C_v^*)^3}. \quad (3.18)$$

Данная оценка так же является состоятельной, но смещенной. Это смещение можно устранить, вводя поправочный коэффициент  $n^2/[(n-1)(n-2)]$ :

$$C_{s,n}^* = \frac{n^2 C_{s,c}^*}{(n-1)(n-2)} = \frac{n \sum_{i=1}^n (x_i - \bar{x})^3}{(n-1)(n-2)S^3} = \frac{n \sum_{i=1}^n (k_i - 1)^3}{(n-1)(n-2)(C_v^*)^3}. \quad (3.19)$$

Оценки числовых характеристик, связанных с моментами более высоких порядков, в практике гидрологических расчетов обычно не используются, так как при существующей в настоящее время длине гидрологических рядов эти оценки не удовлетворяют необходимой точности.

Строго говоря, для достоверной оценки уже третьего центрального момента (а, следовательно, и  $C_s$ ) необходимы выборки в несколько сотен членов. Поэтому на практике вместо выборочного значения коэффициента асимметрии рекомендуется использовать районное соотношение  $C_s/C_v$ . Имея районное соотношение  $C_s/C_v$  и рассчитав выборочную оценку коэффициента вариации несложно получить и оценку  $C_s$ .

Для получения районного соотношения  $C_s/C_v$  в пределах гидрологически однородного района выделяется несколько десятков опорных створов с наиболее продолжительными рядами наблюдений. Для этих створов по формулам (3.17), (3.19) производится расчет  $C_v^*$  и  $C_s^*$  и вычисляется значение  $C_s^*/C_v^*$ . В качестве районного соотношения  $C_s/C_v$  принимается среднее значение по всем опорным створам.

При ограниченности исходных данных, когда невозможно получить районное соотношение  $C_s/C_v$ , подбор значения  $C_s/C_v$  для каждой выборки производят индивидуально, ориентируясь на наилучшее соответствие аналитической и эмпирической кривых обеспеченностей. Эту процедуру удобнее производить с использованием набора спрямляющих клетчаток вероятностей для различных сочетаний  $C_s/C_v$  (см. п. 3.2.4.).

К достоинствам метода моментов можно отнести то, что оценки параметров не зависят от закона распределения исследуемой случайной величины; а также то, что расчетные формулы достаточно просты и позволяют получить искомые параметры в явном виде. Поэтому метод моментов получил наибольшее распространение в практике гидрологических расчетов.

**Пример 3.2.** По ряду наблюдений за максимальными расходами дождевых паводков за период 1954 – 1985 гг. на р. Луге – ст. Толмачево (см. табл. 3.4) оценить параметры распределения методом моментов.

**Решение.** При оценке параметров этим методом используются формулы (3.13), (3.17), (3.19). В табл. 3.5 приведены необходимые для этого промежуточные расчеты.

Таблица 3.5

Вспомогательная таблица для расчета выборочных параметров распределения максимальных паводочных расходов на р. Луге – ст. Толмачево

№ п/п	$Q_i, \text{ м}^3/\text{с}$	$k_i = Q_i / \bar{Q}$	$k_i - 1$	$(k_i - 1)^2$	$(k_i - 1)^3$
1	71.4	1.22	0.22	0.0484	0.01065
2	22.1	0.38	-0.62	0.3844	-0.23833
3	32.7	0.56	-0.44	0.1936	-0.08518
...	...	...	...	...	...
30	39.2	0.67	-0.33	0.1089	-0.03594
31	66.0	1.13	0.13	0.0169	0.00219
32	84.4	1.45	0.45	0.2025	0.09112
Сумма	1868	32	0.00	12.698	6.857
Среднее	58.4	1	0		

По данным табл. 3.5 и вышеуказанным формулам рассчитываем оценки параметров распределения

$$\bar{Q} = \frac{\sum_{i=1}^n Q_i}{n} = \frac{1868}{32} = 58.4 \text{ м}^3/\text{с}, \quad C_v^* = \sqrt{\frac{\sum_{i=1}^n (k_i - 1)^2}{n - 1}} = \sqrt{\frac{12.698}{31}} = 0.64,$$

$$C_s^* = \frac{n \sum_{i=1}^n (k_i - 1)^3}{(n - 1)(n - 2)C_v^3} = \frac{32 \cdot 6.857}{31 \cdot 30 \cdot 0.64^3} = 0.90; \quad C_s/C_v = 1.41.$$

Нужно иметь в виду, что оценки дисперсии, коэффициента вариации и коэффициента асимметрии имеют отрицательную смещенность. Это приводит к тому, что при больших значениях коэффициента вариации ( $C_v > 0.5$ ) достовер-

ность моментных оценок ощутимо снижается, а введение поправочных коэффициентов становится неэффективным. Поэтому при  $C_v > 0.5$  для расчета оценок параметров распределения рекомендуется использовать метод наибольшего правдоподобия.

### 3.2.2. Метод наибольшего правдоподобия

Для нахождения оценки методом наибольшего правдоподобия необходимо прежде всего построить *функцию правдоподобия*. Чтобы понять, как это делается, рассмотрим выборку значений случайной величины  $X$  объемом в  $n$  членов. Предположим, что функция плотности вероятности СВ  $X$  имеет достаточно простой вид:  $f(x, G)$ , т. е. зависит от одного параметра  $G$ . Суть метода состоит в том, чтобы найти такое значение параметра  $G$ , при котором вероятность получить в результате  $n$  опытов именно данную выборку  $(x_1, x_2, x_3, \dots, x_n)$  являлась бы максимальной. С математической точки зрения эта задача сводится к нахождению максимума некоторой функции  $L(x_i, G)$ , которая и называется функцией правдоподобия. Функция  $L(x_i, G)$  представляет собой совместную плотность вероятности вектора  $X = (x_1, x_2, x_3, \dots, x_n)$  при данном  $G$ :

$$L(x_i, G) = f(x_1, G) f(x_2, G) f(x_3, G) \dots f(x_n, G) = \prod_{i=1}^n f(x_i, G). \quad (3.20)$$

Согласно правилам нахождения экстремумов, для определения максимума функции  $L(x_i, G)$  нужно решить уравнение

$$\frac{\partial L(x_i, G)}{\partial G} = 0. \quad (3.21)$$

Для упрощения решения обычно используют логарифмическую функцию правдоподобия  $\ln[L(x_i, G)]$ . Учитывая, что эта функция имеет максимум при том же значении  $G$  получаем

$$\frac{1}{L} \frac{\partial L}{\partial G} = \frac{\partial \ln L}{\partial G} = \sum_{i=1}^n \frac{\partial [\ln f(x_i, G)]}{\partial G} = 0. \quad (3.22)$$

Если функция распределения зависит от нескольких параметров, следует взять частные производные по каждому из них

$$\frac{\partial \ln L}{\partial G_j} = 0 \quad (j = 1, 2, 3, \dots, m).$$

В качестве примера рассмотрим оценку методом наибольшего правдоподобия параметров нормального распределения (см. п. 2.1) для которого функция плотности вероятности имеет вид

$$f(x) = [1/(\sigma_x \sqrt{2\pi})] \exp[-(x - m_x)^2 / (2\sigma_x^2)],$$

где  $m_x$  и  $\sigma_x$  — искомые параметры распределения.

Образует функцию правдоподобия (3.20), но предварительно найдем логарифм функции  $f(x)$ :

$$\ln f(x) = -\ln \sigma - 0.5 \ln(2\pi) - (1/2\sigma^2)(x_i - m_x)^2,$$

$$L = \sum_{i=1}^n \ln f(x) = -\sum_{i=1}^n \ln \sigma - 0.5 \sum_{i=1}^n \ln(2\pi) - (1/2\sigma^2) \sum_{i=1}^n (x_i - m_x)^2.$$

Далее находим уравнение правдоподобия применительно к оценке параметра  $m_x$

$$\frac{\partial L}{\partial m_x} = -\frac{1}{\sigma^2} \sum_{i=1}^n (x_i - m_x) = 0,$$

откуда, учитывая, что  $\sum_{i=1}^n m_x = nm_x$ , получаем  $m_x^* = \frac{1}{n} \sum_{i=1}^n x_i$ .

Таким образом, статистической оценкой параметра  $m_x$  (математического ожидания) в данном случае является среднее арифметическое значение ряда ( $x_1, x_2, x_3, \dots, x_n$ ).

Для получения оценки параметра  $\sigma_x$  образуем новое уравнение правдоподобия

$$\frac{\partial L}{\partial \sigma} = \frac{1}{\sigma^3} \sum_{i=1}^n (x_i - m_x)^2 - \frac{n}{\sigma} = 0.$$

Откуда получим

$$\sigma^* = \sqrt{\frac{\sum_{i=1}^n (x_i - m_x)^2}{n}},$$

т.е. выражение среднеквадратического отклонения.

Из приведенного анализа следует, что применительно к нормальному закону распределения оценки параметров, полученные методом наибольшего правдоподобия, совпадают с моментными оценками. Для других распределений такое совпадение не является обязательным.

Более сложная система уравнений правдоподобия получается для трехпараметрического гамма-распределения. На практике параметры этого распределения  $C_1$  и  $C_2$  определяются на основе приближенного метода наибольшего прав-

доподобия с помощью специально разработанных номограмм [4] как функция вспомогательных статистик  $\lambda_2$  и  $\lambda_3$ :

$$\lambda_2 = \frac{\sum_{i=1}^n \lg k_i}{n-1}, \quad \lambda_3 = \frac{\sum_{i=1}^n k_i \lg k_i}{n-1}, \quad (3.23)$$

где  $k_i = x_i / \bar{x}$  – модульные коэффициенты;  $\bar{x}$  – среднее арифметическое ряда. Пример такой номограммы приводится на рис. 3.3.

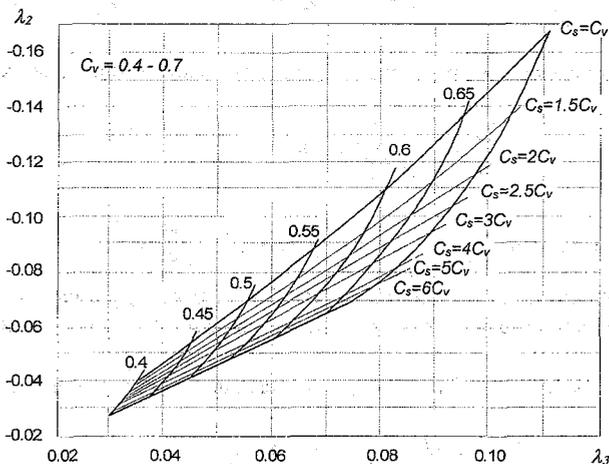


Рис. 3.3. Номограмма для вычисления параметров трехпараметрического гамма-распределения  $C_s$  и  $C_v$  методом приближенного наибольшего правдоподобия;  $C_v = 0.4 - 0.7$ .

Считается, что полученные таким образом оценки параметров трехпараметрического гамма-распределения являются состоятельными, эффективными и несмещенными.

В качестве особенностей метода наибольшего правдоподобия (МНП) можно отметить то, что в нем наибольший вес придается средним членам выборки, имеющим наибольшую вероятность, в отличие от метода моментов, где самый большой вклад вносят крайние члены. МНП дает состоятельные и наиболее эффективные оценки. Полученные по этому методу оценки могут быть незначительно смещены, но это смещение легко устраняется путем введения соответствующих поправок. Однако перечисленные свойства проявляются только в случае достаточно больших выборок, при малых  $n$  с оценкам МНП могут конкурировать другие оценки.

К недостаткам метода можно отнести то, что для его применения необходимо точно знать аналитическое выражение закона распределения, что в гидрологической практике не всегда возможно. К тому же не всегда удается получить решение в аналитическом виде и определение максимума функции правдоподобия приходится производить численными методами или строить вспомогательные номограммы.

В России номограммы для метода наибольшего правдоподобия разработаны применительно к распределению Крицкого – Менкеля. Однако на практике не редки ситуации, когда выборочные  $\lambda_2$  и  $\lambda_3$  приводят к выходу за пределы номограмм, т. е. решение отсутствует. В этом случае можно использовать сокращенный метод наибольшего правдоподобия.

При реализации сокращенного метода статистика  $\lambda_3$  не рассчитывается, а вместо нее используется районное соотношение  $C_s/C_v$ . Зная  $\lambda_2$  и  $C_s/C_v$ , по номограмме можно определить коэффициент вариации.

Сокращенный метод наибольшего правдоподобия используется и применительно к кривой Пирсона III типа. Для этого распределения зависимость коэффициента вариации от параметров  $\lambda_2$  и  $C_s/C_v$  представлена в виде таблицы (Прил.9).

**Пример 3.3.** По ряду наблюдений за максимальными расходами дождевых паводков за период 1954 – 1985 гг. на р. Луге – ст. Толмачево (см. табл. 3.4) оценить параметры распределения методом наибольшего правдоподобия.

**Решение.**

1. Для расчета среднего значения ряда и вспомогательных статистик  $\lambda_2, \lambda_3$  заполняем табл. 3.6.

Таблица 3.6

**Вспомогательная таблица для расчета оценок параметров распределения методом наибольшего правдоподобия**

№ n/n	Q, м <sup>3</sup> /с	k	lg(k)	k·lg(k)
1	71.4	1.22	0.086	0.1054
2	22.1	0.38	-0.4202	-0.1435
3	32.7	0.56	-0.2518	-0.1410
...	...	...	...	...
30	39.2	0.67	-0.1739	-0.1165
31	66.0	1.13	0.0531	0.0600
32	84.4	1.45	0.1614	0.2340
Сумма	1868		-2.697	2.573
Среднее	58.4			

2. По формулам (3.23) рассчитываем статистики  $\lambda_2$  и  $\lambda_3$ :

$$\lambda_2 = \frac{-2.697}{31} = -0.087, \quad \lambda_3 = \frac{2.573}{31} = 0.083.$$

3. Используя номограмму [4] (см. рис. 3.3) в зависимости от  $\lambda_2$  и  $\lambda_3$  определяем параметры распределения:  $C_v = 0.67, C_s/C_v = 2.8$ .

### 3.2.3. Графоаналитический метод (метод квантилей)

Графоаналитический метод расчета оценок параметров распределения включает в себя элементы графического и аналитического анализа.

При использовании этого метода необходимо заранее знать, какая из аналитических кривых будет применяться для аппроксимации распределения исследуемой СВ.

#### 3.2.3.1. Графоаналитический метод на основе кривой Пирсона III типа

Методика, применяемая в настоящее время в России, была разработана Г. А. Алексеевым (1960 г.) и предполагает использование кривой Пирсона III типа. Рассмотрим основные этапы расчета.

Сначала на клетчатку вероятности наносятся точки эмпирической кривой обеспеченностей (рис. 3.4). При этом по оси ординат следует откладывать значения исследуемой СВ, а не модульные коэффициенты. Затем в поле точек проводится сглаженная кривая.

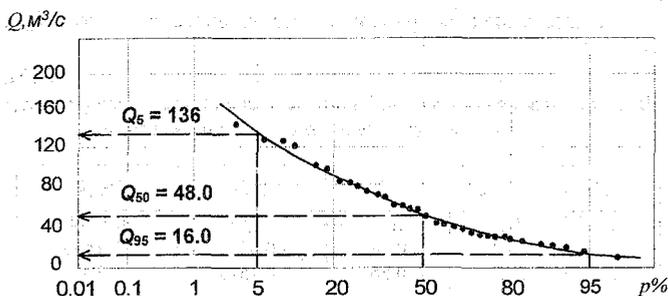


Рис. 3.4. Схема определения опорных ординат по сглаженной эмпирической кривой обеспеченностей ряда максимальных паводочных расходов на р. Луге – ст. Толмачево.

По сглаженной кривой определяются три опорные ординаты (для обеспеченностей – 5, 50 и 95 %):  $x_5, x_{50}, x_{95}$ .

Дальнейший расчет сводится к тому, чтобы найти такие значения параметров распределения ( $\bar{x}, C_s^*, C_s^*$ ), при которых аналитическая кривая будет проходить через точки  $x_5, x_{50}, x_{95}$ .

Для решения этой задачи трижды воспользуемся формулой (2.5)

$$\hat{x}_5 = \sigma_x^* \hat{t}_5 + \bar{x}, \quad (3.24)$$

$$\hat{x}_{50} = \sigma_x^* \hat{t}_{50} + \bar{x}, \quad (3.25)$$

$$\hat{x}_{95} = \sigma_x^* \hat{t}_{95} + \bar{x}, \quad (3.26)$$

где  $\hat{t}_5, \hat{t}_{50}, \hat{t}_{95}$  – нормированные ординаты кривой обеспеченностей Пирсона III типа (см. приложение 1). Таким образом, мы имеем три неизвестных параметра и три уравнения.

Для расчета оценки коэффициента асимметрии сначала рассчитаем так называемый коэффициент скошенности  $S$ :

$$S = (\hat{x}_p + \hat{x}_{100-p} - 2\hat{x}_{50}) / (\hat{x}_5 - \hat{x}_{95}). \quad (3.27)$$

В нашем случае обеспеченность первой ординаты  $\hat{p} = 5\%$ , следовательно:

$$S^* = (\hat{x}_5 + \hat{x}_{95} - 2\hat{x}_{50}) / (\hat{x}_5 - \hat{x}_{95}). \quad (3.27a)$$

Подставляя выражения (3.24) – (3.26) в формулу (3.27 а), получим

$$S^* = (\hat{t}_5 + \hat{t}_{95} - 2\hat{t}_{50}) / (\hat{t}_5 - \hat{t}_{95}), \quad (3.27б)$$

т. е. коэффициент скошенности не изменится при замене  $\hat{x}_5, \hat{x}_{50}, \hat{x}_{95}$  на нормированные ординаты. Но при фиксированном значении  $\hat{p}$  для нормированного распределения Пирсона III типа имеет место однозначная зависимость коэффициента скошенности от коэффициента асимметрии (см. приложение 1).

Следовательно, зная коэффициент скошенности (3.27а), по приложению 1 можно определить  $C_s^*$ .

Вычитая из уравнения (3.24) соответственно левую и правую части уравнения (3.26), получим выражение для оценки СКО

$$\sigma_x^* = (\hat{x}_5 - \hat{x}_{95}) / (\hat{t}_5 - \hat{t}_{95}). \quad (3.28)$$

Оценку МО находим из уравнения (3.25)

$$\bar{x} = \hat{x}_{50} - \sigma_x^* \hat{t}_{50}. \quad (3.29)$$

С учетом формулы (1.14) находим коэффициент вариации

$$C_v^* = \sigma_x^*/x. \quad (3.30)$$

Таким образом, аналитическая часть расчета сводится к последовательному использованию формул (3.27а), (3.28) – (3.30).

**Пример 3.4.** По ряду наблюдений за максимальными расходами дождевых паводков на р. Луге – ст. Толмачево (см. табл. 1.2) оценить параметры распределения графоаналитическим методом. В качестве аналитической кривой обеспеченностей использовать кривую Пирсона III типа.

**Решение.**

1. Ранжируем ряд и по формуле (3.6) рассчитываем ординаты эмпирической кривой обеспеченностей (см. табл. 3.4).
2. На клетчатке с умеренной асимметричностью строим эмпирическую кривую обеспеченностей (см. рис.3.4).
3. По сглаженной эмпирической кривой обеспеченностей определяем опорные ординаты  $Q_5^* = 136$ ,  $Q_{50}^* = 48$ ,  $Q_{95}^* = 16$ .
4. По формуле (3.27а) рассчитываем коэффициент скошенности  $S^* = (136 + 16 - 2 \cdot 48) / (136 - 16) = 0.47$ .
5. В зависимости от коэффициента скошенности по приложению 1 определяем три нормированные ординаты кривой Пирсона III типа и коэффициент асимметрии:  $f_5 = 1.98$ ;  $f_{50} = -0.26$ ;  $f_{95} = -1.07$ ;  $C_s^* = 1.67$ .
6. По формуле (3.28) вычисляем оценку СКО:  $\sigma_q^* = (136 - 16) / (1.98 + 1.07) = 39.3$ .
7. По формуле (3.29) вычисляем оценку МО:  $\bar{Q} = 48 - 39.3(-0.26) = 58.2$ .
8. По формуле (3.30) вычисляем оценку  $C_v$ :  $C_v^* = 39.3 / 58.2 = 0.68$ .
9. В результате получаем:  $\bar{Q} = 58.2$ ;  $C_v^* = 0.68$ ;  $C_s^* = 1.67$ ;  $C_s^* / C_v^* = 2.46$ .

Если на основе этих параметров требуется рассчитать ординаты аналитической кривой обеспеченностей, то расчет ведется по схеме, изложенной в п. 2.5 (см. пример 2.6).

### **3.2.3.2. Графоаналитический метод на основе трехпараметрической логнормальной кривой**

У трехпараметрического логнормального распределения третий параметр (а) не влияет на форму кривой обеспеченностей, а лишь приводит к сдвигу по

оси  $X$ . Поэтому для данного распределения имеет место однозначная зависимость коэффициента скошенности  $S$  от коэффициента асимметрии. И, следовательно, можно применить подход, аналогичный тому, который применялся для кривой Пирсона III типа. Методики отличаются только тем, что здесь требуется дополнительно рассчитать параметр сдвига  $a$ .

Как показал Г.А. Алексеев, параметр  $a$  связан с опорными ординатами формулой

$$a = \frac{\bar{x}_s \bar{x}_{95} - \bar{x}_{50}^2}{\bar{x}_s + \bar{x}_{95} - 2\bar{x}_{50}} \quad (3.31)$$

Дальнейший расчет ведется по стандартной схеме.

- По формуле (3.27а) рассчитывается коэффициент скошенности  $S^*$ .
- В зависимости от  $S^*$  по таблице логнормального распределения (табл.3.7) определяются три нормированные ординаты и коэффициент асимметрии.
- По формулам (3.28)–(3.30) рассчитываются  $\sigma_x^*$ ,  $\bar{x}$ ,  $C_{v,x}^*$ .

Таблица 3.7

Значения коэффициента скошенности  $S$ , вычисленного на основе опорных ординат  $t_5$ ,  $t_{50}$ ,  $t_{95}$  логнормального распределения в зависимости от коэффициента асимметрии  $C_s$

$C_s$	$S$	Нормированные ординаты $t_p$			$C_s$	$S$	Нормированные ординаты $t_p$		
		5 %	50 %	95 %			5 %	50 %	95 %
0	0	1.64	0	-1.64	1.9	0.41	1.89	-0.23	-1.12
0.1	0.03	1.67	-0.02	-1.62	2.0	0.42	1.89	-0.24	-1.10
0.2	0.06	1.70	-0.04	-1.59	2.1	0.44	1.89	-0.24	-1.08
0.3	0.09	1.72	-0.06	-1.56	2.2	0.45	1.89	-0.25	-1.06
0.4	0.11	1.75	-0.07	-1.53	2.3	0.46	1.88	-0.25	-1.04
0.5	0.14	1.77	-0.09	-1.49	2.4	0.48	1.88	-0.26	-1.02
0.6	0.16	1.79	-0.10	-1.46	2.5	0.49	1.88	-0.26	-1.00
0.7	0.19	1.81	-0.11	-1.43	2.6	0.50	1.87	-0.26	-0.99
0.8	0.21	1.82	-0.13	-1.40	2.7	0.51	1.87	-0.27	-0.97
0.9	0.23	1.84	-0.14	-1.37	2.8	0.51	1.86	-0.27	-0.96
1.0	0.25	1.85	-0.15	-1.34	2.9	0.52	1.86	-0.27	-0.95
1.1	0.27	1.86	-0.16	-1.31	3.0	0.53	1.85	-0.28	-0.93
1.2	0.29	1.87	-0.17	-1.29	3.2	0.55	1.84	-0.28	-0.90
1.3	0.31	1.88	-0.18	-1.26	3.4	0.56	1.83	-0.29	-0.88
1.4	0.33	1.88	-0.19	-1.23	3.6	0.57	1.81	-0.29	-0.86
1.5	0.35	1.89	-0.20	-1.21	3.8	0.58	1.80	-0.29	-0.84
1.6	0.37	1.89	-0.21	-1.18	4.0	0.59	1.78	-0.29	-0.82
1.7	0.38	1.89	-0.22	-1.16	4.5	0.62	1.75	-0.30	-0.78
1.8	0.39	1.89	-0.22	-1.14	5.0	0.64	1.71	-0.30	-0.74

**Пример 3.5.** По ряду расходов из предыдущего примера оценить параметры распределения графоаналитическим методом. В качестве аналитической кривой обеспеченностей использовать трехпараметрическую логнормальную кривую.

**Решение.**

Пункты 1–4 – аналогичны пунктам примера 3.4.

5. В зависимости от  $S$  по табл. 3.7 определяем три нормированные ординаты логнормального распределения и коэффициент асимметрии

$$\hat{i}_3 = 1.88; \hat{i}_{30} = -0.26; \hat{i}_{95} = -1.03; C_s^* = 2.35.$$

6. По формуле (3.28) вычисляем оценку СКО  $\sigma_Q^* = (136 - 16)/(1.88 + 1.03) = 41.2$ .

7. По формуле (3.29) вычисляем оценку МО:  $\bar{Q} = 48 - 41.2 \cdot (-0.26) = 58.7$ .

8. По формуле (3.30) вычисляем оценку  $C_v$ :  $C_v^* = 41.2/58.7 = 0.70$ .

В результате получаем  $\bar{Q} = 58.7$ ;  $C_v^* = 0.70$ ;  $C_s^* = 2.35$ ;  $C_s^*/C_v^* = 3.36$ .

Если на основе полученных параметров требуется рассчитать ординаты аналитической кривой обеспеченностей, то дополнительно по формуле (3.31) определяется параметр  $a$ :  $a = (136 \cdot 16 - 48^2)/(136 + 16 - 2 \cdot 48) = -2.29$ .

Дальнейший расчет ведется по схеме, изложенной в п. 2.3. (см. пример 2.4).

### 3.2.3.3. Графоаналитический метод на основе кривой Крицкого – Менкеля

Хотя кривая Крицкого–Менкеля рекомендуется нормативными документами как основная расчетная кривая, до последнего времени она не использовалась для оценки параметров распределения методом квантилей. Причиной этого является отсутствие для распределения Крицкого–Менкеля однозначной связи между коэффициентом асимметрии  $C_s$  и коэффициентом скошенности  $S$ .

В 1996 г. на кафедре гидрологии суши РГГМУ была предложена методика, учитывающая специфику данной кривой.

В отличие от стандартной схемы здесь наряду с обычным коэффициентом скошенности  $S$ , рассчитывается коэффициент  $S_2$

$$S_2 = 2x_{30}/(x_3 - x_{95}) = 2k_{50}/(k_3 - k_{95}). \quad (3.32)$$

Как показали исследования, при фиксированном значении  $C_s/C_v$  коэффициент  $S_2$  однозначно зависит от  $S$ . Данная зависимость представлена в виде номограммы на рис. 3.5.

Таким образом, рассчитав  $S$  и  $S_2$ , можно по номограмме оценить соотношение  $C_s/C_v$ . Значение  $C_s/C_v$ , определенное по номограмме, следует округлить до 0.5.

Большое значение коэффициента скошенности ( $S > 0.6$ ) указывает на то, что  $C_v > 1$ . В этой ситуации выборочная оценка соотношения  $C_s/C_v$  является крайне ненадежной и следует отказаться от ее расчета графоаналитическим методом (можно принять районное значение  $C_s/C_v$ ).

При фиксированном  $C_s/C_v$  имеет место однозначная зависимость между коэффициентом вариации  $C_v$  и коэффициентом скошенности  $S$ . Эта зависимость представлена в табл. 3.8. При известных значениях  $C_s/C_v$  и  $S$  по этой таблице можно определить коэффициент вариации.

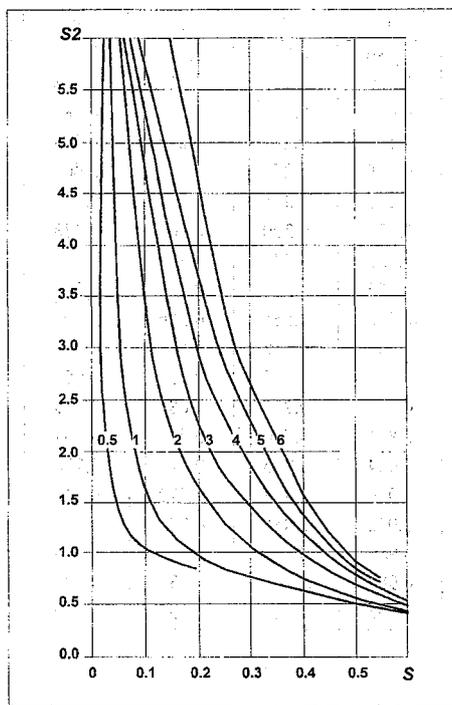


Рис.3.5. Номограмма для определения соотношения  $C_s/C_v$  в зависимости от коэффициента скошенности  $S$  и параметра  $S_2$  для кривой обеспеченностей Крицкого – Менкеля

Для расчета среднего значения необходимо по таблице ординат кривой обеспеченностей Крицкого – Менкеля (см. приложение 2) определить модульный коэффициент 50 %-ной обеспеченности ( $k_{50}$ ). Тогда

$$\bar{x} = x_{50} / k_{50}. \quad (3.33)$$

Таблица 3.8

Значения коэффициента скошенности  $S$  в зависимости от  $C_v$  и соотношения  $C_s/C_v$  для кривой обеспеченностей Крицкого–Менкеля

$C_v$	Соотношение $C_s/C_v$						
	0.5	1	2	3	4	5	6
0.1	0.027	0.036	0.055	0.061	0.068	0.091	0.152
0.2	0.030	0.056	0.113	0.163	0.198	0.238	0.270
0.3	0.041	0.084	0.169	0.229	0.286	0.326	0.356
0.4	0.050	0.113	0.226	0.298	0.348	0.378	0.402
0.5	0.074	0.156	0.279	0.357	0.403	0.427	0.443
0.6	0.114	0.206	0.335	0.409	0.450	0.464	0.481
0.7	0.197	0.270	0.390	0.461	0.488	0.489	0.508
0.8		0.358	0.445	0.504	0.527	0.538	0.539
0.9		0.456	0.506	0.546	0.561	0.563	0.571
1.0		0.562	0.565	0.586	0.595		
1.1		0.663	0.618	0.623	0.625		
1.2		0.754	0.667	0.658	0.652		
1.3		0.823	0.715	0.689	0.679		
1.4		0.871	0.760	0.719	0.704		
1.5		0.904	0.796	0.746	0.726		
1.6		0.930	0.823	0.772	0.748		
1.7		0.952	0.856	0.795	0.768		
1.8		0.969	0.882	0.817	0.786		
1.9		0.983	0.902	0.837	0.803		
2.0		0.996	0.920	0.855	0.819		

**Пример 3.6.** Оценить параметры распределения графоаналитическим методом, используя данные примера 3.4. В качестве аппроксимирующей кривой использовать кривую обеспеченностей Крицкого – Менкеля.

**Решение.**

Пункты 1 – 4 аналогичны пунктам Примера 3.4.

5. По формуле (3.32) определяем коэффициент  $S_2$ :  $S_2 = (2 \cdot 48) / (136 - 16) = 0.80$ .

6. Для значений  $S = 0.47$  и  $S_2 = 0.80$  по номограмме (см. рис. 3.5) определяем соотношение  $(C_s/C_v)$ . В данном случае (с округлением до целых)  $C_s/C_v = 3$ .
7. По табл. 3.8. для  $S = 0.47$  и  $C_s/C_v = 3$  определяем коэффициент вариации. В данном случае  $C_v = 0.72$ .
8. По таблицам кривой обеспеченностей Крицкого–Менкеля (см. приложение 2) при  $C_s/C_v = 3$  и  $C_v = 0.72$  определяем  $k_{50}$ :  $k_{50} = 0.815$ .
9. По формуле (3.33) определяем среднее значение  $\bar{x} = 48.0 / 0.815 = 58.9$ .

Расчет ординат аналитической кривой обеспеченностей производится по схеме, изложенной в п. 2.6. (см. пример 2.7).

### 3.2.3.4. Комментарии

В табл. 3.9 даны значения параметров распределения, рассчитанные графоаналитическим методом с использованием трех типов кривых распределения.

Таблица 3.9

Параметры распределения максимальных расходов дождевых паводков р. Луги – ст. Толмачево, рассчитанные графоаналитическим методом с использованием разных аналитических кривых

Кривая обеспеченностей	$\bar{x}$	$C_v^*$	$C_s^*$	$C_s^*/C_v^*$	Доп. параметры
Пирсона III типа	58.2	0.68	1.67	2.46	$S = 0.47$
Трехпараметрическая логнормальная	58.7	0.70	2.35	3.36	$S = 0.47$ $a = -2.29$
Крицкого–Менкеля	58.9	0.72	2.16	3.0	$S = 0.47$ $S_2 = 0.80$

Как видно из таблицы, все три варианта дают близкие значения *среднего и коэффициента вариации*. В то же время значения *коэффициента асимметрии* различаются весьма значительно, так как именно коэффициент асимметрии (который связан с коэффициентом скошенности  $S$ ) обеспечивает главным образом нужное соотношение опорных ординат, но для различных типов кривых это соотношение будет достигаться при различных  $C_s$ .

На аналитических кривых обеспеченностей различие в асимметрии будет проявляться прежде всего в области больших и малых значений, т. е. на концах кривых.

Хотя графоаналитический метод является достаточно простым и наглядным, его следует применять с осторожностью, так как проведение сглаженной эмпирической кривой всегда носит субъективный характер и требует определенных практических навыков.

### 3.2.4. Графический метод

При использовании графического метода для расчета оценок параметров распределения необходимо заранее знать, какая из аналитических кривых будет применяться для аппроксимации закона распределения исследуемой СВ.

Этот метод может применяться как для двухпараметрических, так и для трехпараметрических кривых распределения.

При использовании трехпараметрической кривой на первом этапе применяется набор спрямляющих клетчаток выбранного теоретического распределения для различных сочетаний  $C_s$  и  $C_v$ .

На каждой клетчатке строится эмпирическая кривая обеспеченностей в модульных коэффициентах. В качестве расчетного значения принимается такое соотношение  $C_s/C_v$ , при котором на соответствующей клетчатке эмпирическая кривая обеспеченностей превращается в прямую линию. Как отмечалось в п. 2.8, наклон прямой линии, в которую превращается кривая обеспеченностей на спрямляющей клетчатке, зависит от коэффициента вариации. С учетом этого на клетчатке вероятности можно построить шкалу  $C_v$ . Обычно для контроля такая шкала наносится в левом верхнем и правом нижнем углах клетчатки (рис.3.6). В качестве расчетного  $C_v^*$  принимается такое значение, при котором отсчеты по верхней и нижней шкалам совпадают и при этом аппроксимирующая прямая достаточно хорошо соответствует эмпирическим точкам.

Таким образом графический метод позволяет найти оценки коэффициента вариации и соотношения  $C_s/C_v$ . Оценка МО рассчитывается аналитически по формуле (3.13).

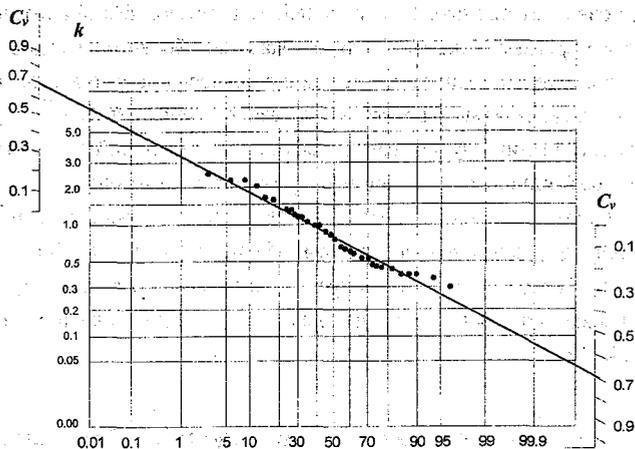


Рис.3.6. Кривая обеспеченностей максимальных паводочных расходов на р. Луге – ст. Толмачево, построенная на клетчатке, спрямляющей кривую Крицкого – Менкеля при соотношении  $C_s/C_v = 3$ .

**Пример 3.7.** По ряду наблюдений за максимальными дождевыми расходами воды на реке Луге – ст. Толмачево (см. табл. 1.2) оценить параметры распределения графическим методом. В качестве аналитической кривой обеспеченностей использовать кривую Крицкого – Менкеля.

**Решение.**

1. По формуле (3.13) рассчитываем оценку МО:  $\bar{Q} = 58.4$ .
2. По формуле  $k_i = Q_i/\bar{Q}$  рассчитываем модульные коэффициенты.
3. Ранжируем ряд модульных коэффициентов и по формуле (3.6) рассчитываем ординаты эмпирической кривой обеспеченностей (см. табл. 3.4).
4. На клетчатках, спрямляющих кривую Крицкого – Менкеля, пять раз строим эмпирическую кривую обеспеченностей (для  $C_s/C_v = 1; 1.5; 2.0; 3.0; 4.0$ ).
5. В качестве расчетного принимаем значение  $C_s/C_v = 3$ , так как на этой клетчатке точки эмпирической кривой группируются вдоль прямой линии (см. рис.3.6).
6. В поле точек проводим прямую линию и по угловой шкале определяем коэффициент вариации. В данном случае  $C_v^* = 0.70$ .

Таким образом, окончательно получаем  $\bar{Q} = 58.4$ ;  $C_v^* = 0.70$ ;  $(C_s/C_v)^* = 3$  и, следовательно,  $C_s^* = 2.1$ .

Метод позволяет оценить  $C_v$  с меньшей точностью, чем аналитические методы. В то же время оценка  $C_s/C_v$  может оказаться более надежной, поскольку при использовании аналитических методов в случае короткой выборки несколько «тяжелых» точек могут существенно исказить оценку отношения  $C_s/C_v$ .

При использовании двухпараметрических распределений  $C_s$  однозначно определяется коэффициентом вариации и требуется только одна клетчатка. Если эмпирические точки на такой клетчатке лягут в виде прямой линии, расчет  $C_v$  производится по изложенной выше схеме. Если точки будут образовывать вогнутую или выпуклую линии, то следует сделать вывод о том, что данная двухпараметрическая кривая не подходит для аппроксимации закона распределения исследуемой СВ.

### 3.2.5. Оценка параметров распределения Джонсона

Распределение Джонсона (см. п. 2.7) является одной из модификаций лог-нормального распределения. Вывод этого распределения базируется на том, что СВ  $Z$  имеет нормальное распределение, где  $Z = \ln[(X - a)/(b - X)]$ ;  $a$  и  $b$  — соответственно нижний и верхний пределы СВ  $X$ .

Данное распределение является четырехпараметрическим и требует в общем случае оценки четырех параметров:  $m_z$ ,  $\sigma_z$ ,  $a$  и  $b$ .

Если параметры  $a$  и  $b$  известны, то, преобразовав члены выборки по формуле  $z_i = \ln[(x_i - a)/(b - x_i)]$ , можно легко рассчитать оценки  $m_z$  и  $\sigma_z$  методом моментов по формулам (3.13) и (3.16).

Таким образом, основная проблема состоит в оценке параметров  $a$  и  $b$ . Оценка этих параметров на основе аналитического решения в зависимости от третьего и четвертого моментов не может дать надежного результата, так как эти числовые характеристики при небольшой длине выборок крайне неустойчивы. На практике параметры  $a$  и  $b$  находят методом последовательных приближений<sup>1</sup>.

На первом этапе устанавливаются диапазоны возможных значений параметров. Так, для положительных величин

$$0 \leq a < x_{\min}, \quad (3.34)$$

$$x_{\max} < b < M, \quad (3.35)$$

<sup>1</sup> Учитывая большой объем вычислений метод может быть реализован только на ЭВМ.

где  $x_{\min}$  и  $x_{\max}$  – соответственно минимальное и максимальное наблюдаемые значения. В качестве  $M$  принимается такое значение, которое заведомо является невероятным, т. е. никогда не будет превышено. Например, можно принять  $M = 5x_{\max}$ . Эта процедура необходима, чтобы заранее отсеять физически не обоснованные решения.

Затем для нескольких вариантов параметра  $a$ , например,  $a_0 = 0$ ,  $a_1 = 0.1x_{\min}$ ,  $a_2 = 0.2x_{\min}$ ,  $a_3 = 0.3x_{\min}$  и т. д., производится оптимизация параметра  $b$ . Т. е. для каждого  $a_i$  находят такое значение параметра  $b = b_i$ , при котором имеет место наибольшее соответствие эмпирической и аналитической кривых обеспеченностей. После чего выбирается наилучшая пара значений  $a$  и  $b$  (для которой степень соответствия эмпирической и аналитической кривых максимальна)

Степень соответствия эмпирической и аналитической кривых можно оценивать по-разному. Так, Г. Г. Сванидзе [9] рекомендует использовать для этой цели критерий согласия<sup>1</sup> *Хи-квадрат*, а В. А. Шелутко [11] – критерий согласия  $n\omega^2$ .

Мы предлагаем другой вариант. В качестве оптимального будем принимать такое значение параметра  $b$ , при котором коэффициент асимметрии СВ  $Z$  равен или близок нулю. Такой подход, на наш взгляд, является наиболее естественным, так как по условию СВ  $Z$  должна быть нормально распределенной, т. е. иметь нулевую асимметрию.

Однако как в первом случае (при использовании критериев согласия), так и во втором основная проблема оптимизации состоит в том, что можно получить несколько решений (несколько пар  $a$  и  $b$ ), при которых степень согласия эмпирической и аналитической кривых обеспеченностей будет достаточно высокой, и выбор окончательного решения носит субъективный характер. В то же время формальный подход на основе абсолютного минимума критерия качества<sup>2</sup> может вообще привести к абсурдным результатам из-за выборочного характера исходных данных. В силу указанных причин физический смысл параметров  $a$  и  $b$ , как абсолютного минимума и абсолютного максимума, размывается, и они превращаются в формальные параметры распределения. Именно поэтому четырехпараметрическое распределение Джонсона не нашло широкого распространения в практике гидрологических расчетов.

Если хотя бы один параметров ( $a$  или  $b$ ) известен, проблем с оптимизацией не возникает.

<sup>1</sup> О критериях согласия см. п. 4.7.

<sup>2</sup> Критерий качества – любой числовой показатель, характеризующий степень отклонения эмпирических данных от оптимизируемой аналитической кривой.

**Пример 3.8.** Для ряда максимальных расходов дождевых паводков  $p$ . Луги у ст. Толмачево (см. табл. 1.2) оценить параметры распределения. В качестве аппроксимирующей кривой использовать кривую обеспеченностей Джонсона.

**Решение.**

1. Ориентируясь на систему ограничений (3.34) – (3.35), приближенно определяем область возможных значений параметров  $a$  и  $b$ :  $0 \leq a < 17.3$ ;  
 $145 < b < (145 \cdot 5 = 725)$ .
2. Для различных значений параметра  $a$  ( $a_0 = 0, a_1 = 1, a_2 = 2, \dots, a_{17} = 17$ ) находим такие значения параметра  $b$  ( $b_0, b_1, b_2, \dots, b_{17}$ ), при которых для преобразованного ряда значений  $z_i = \ln [(x_i - a) / (b - x_i)]$  выборочный коэффициент асимметрии близок к нулю. Результаты такой оптимизации даны в табл. 3.10.

Таблица 3.10

Результаты оптимизации параметров  $a$  и  $b$  распределения Джонсона для максимальных паводочных расходов на р. Луге у ст. Толмачево

Оптимизируемый параметр		$C_{s,z}^*$	Оптимизируемый параметр		$C_{s,z}^*$
$a$	$b$		$a$	$b$	
0	Нет решения	–	11	490	0.00
5	Нет решения	–	12	365	0.00
10	770	0.00	15	202	0.00

3. В качестве расчетных параметров принимаем  $a = 11$ ;  $b = 490$ , так как это ближайшее к нулю значение  $a$ , при котором  $b < M$ .

*Замечание:* Поскольку значение  $M$  назначено достаточно произвольно, не исключено, что решением будет пара  $a = 10$ ;  $b = 770$  или пара  $a = 12$ ;  $b = 365$ . Таким образом, в ситуации, когда дополнительная информация о параметрах  $a$  и  $b$  отсутствует, выбор окончательного решения всегда будет достаточно субъективным.

4. Преобразуем исходный ряд по формуле  $z_i = \ln [(x_i - 11) / (490 - x_i)]$ .
5. Используя формулы (3.13) и (3.16), по ряду значений  $z_i$  вычисляем  $\bar{z}$  и  $\sigma_z$ .

В данном случае  $\bar{z} = -2.54$ ;  $\sigma_z = 0.95$ . Дальнейший расчет ведется по схеме, изложенной в п. 2.7 (см. пример 2.8).

### 3.3. Оценка погрешностей выборочных параметров распределения

Оценка параметра распределения некоторой СВ  $X$ , полученная по выборке тем или иным методом, сама представляет собой случайную величину, обладающую определенным разбросом. Имеется в виду, что можно получить сколько угодно выборку объемом  $n$  из одной и той же генеральной совокупности, и все они будут давать различные значения оцениваемого параметра. Поэтому точечная оценка параметра  $G$  в виде конкретного значения  $G^*$  не дает полного представления об искомом параметре без оценки погрешности ее вычисления. Мерой случайной погрешности для выборочного параметра  $G^*$  может служить среднеквадратическое отклонение  $\sigma_G$  (абсолютная погрешность) или относительное среднеквадратическое отклонение  $\varepsilon_G = (\sigma_G/G) \cdot 100\%$  (относительная погрешность)<sup>1</sup>. Формулы для расчета погрешности зависят от того, каким методом производилась оценка параметра.

**Метод моментов.** Абсолютная погрешность выборочного среднего определяется по формуле

$$\sigma_{\bar{x}} = \sigma/\sqrt{n}, \quad (3.36)$$

где  $\sigma$  – СКО выборки;  $n$  – объем выборки (длина ряда).

Относительная погрешность выборочного среднего:

$$\varepsilon_{\bar{x}} = \frac{\sigma_{\bar{x}}}{\bar{x}} \cdot 100 = \frac{\sigma}{x\sqrt{n}} \cdot 100 = \frac{C_v}{\sqrt{n}} \cdot 100\%, \quad (3.37)$$

где  $C_v$  – выборочный коэффициент вариации.

Для расчета абсолютной и относительной погрешностей коэффициента вариации используются формулы

$$\sigma_{c_v} = \frac{C_v \sqrt{1 + aC_v^2}}{\sqrt{2n}}, \quad (3.38)$$

$$\varepsilon_{c_v} = \frac{\sqrt{1 + aC_v^2}}{\sqrt{2n}} \cdot 100\%, \quad (3.39)$$

где  $a = 2$  для нормального распределения и  $a = 1$  для двухпараметрического гамма-распределения.

Напомним, что у двухпараметрического гамма-распределения соотношение  $C_s/C_v = 2$ , а у нормального распределения  $C_s = 0$ . Так как гидрологические ряды имеют, как правило, умеренную положительную асимметрию, эти формулы рекомендуется использовать при  $a = 1$  [7].

<sup>1</sup> Иногда вместо термина “погрешность” используется термин “ошибка”, а также термины “стандартная погрешность” и “стандартная ошибка”.

Помимо формулы (3.38) в российской гидрологической практике достаточно широкое распространение получила формула Е. Г. Блохинова:

$$\sigma_{c_v} = \frac{C_v}{n + 4C_v^2} \sqrt{\frac{n(1 + C_v^2)}{2}} \quad (3.40)$$

Данная формула также получена для двухпараметрического гамма распределения путем введения эмпирической поправки в формулу (3.38).

Для оценки погрешности коэффициента асимметрии предложены следующие формулы.

Формула С. Н. Крицкого и М. Ф. Менкеля

$$\sigma_{c_s} = \sqrt{(6/n)(1 + 6C_v^2 + 5C_v^4)}, \quad (3.41)$$

$$\varepsilon_{c_s} = \frac{1}{C_s} \sqrt{(6/n)(1 + 6C_v^2 + 5C_v^4)} \cdot 100\%; \quad (3.42)$$

формула А. Ш. Резниковского

$$\sigma_{c_s} = \sqrt{(6/n)(1 + C_v^2)}, \quad (3.43)$$

$$\varepsilon_{c_s} = \frac{1}{C_s} \sqrt{(6/n)(1 + C_v^2)} \cdot 100\%. \quad (3.44)$$

Формулы (3.41) и (3.43) разработаны для случая  $C_s = 2C_v$ , причем формула (3.41) получена теоретическим путем, а формула (3.43) на основании материалов статистического моделирования.

Для распределений, у которых коэффициент асимметрии близок к нулю, формулы (3.41) – (3.44) не применимы, так как множитель  $1/C_s$  стремится к бесконечности. В этом случае относительная погрешность  $C_s$  вообще не вычисляется.

**Метод наибольшего правдоподобия.** Абсолютная и относительная погрешность среднего значения и коэффициента асимметрии определяется так же как и для метода моментов: по формулам (3.36) – (3.37); (3.41) – (3.42).

Для приближенной оценки погрешности коэффициента вариации в настоящее время рекомендуется применять формулы

$$\sigma_{c_v} = \frac{C_v}{\sqrt{2n}} \sqrt{\frac{3}{3 + C_v^2}}, \quad (3.45)$$

$$\varepsilon_{cv} = \sqrt{\frac{3}{2n(3 + C_v^2)}} 100\% \quad (3.46)$$

В гидрологической практике принято считать, что расчет параметров распределения выполнен надежно, если погрешность среднего значения не превышает 10%, а коэффициента вариации 15%.

Ошибка коэффициента асимметрии при имеющейся в настоящее время длине рядов довольно высока и может достигать нескольких десятков процентов, поэтому величина  $C_s$  назначается в соответствии с рекомендациями, изложенными в п. 3.2.1.

**Метод квантилей.** При использовании этого метода абсолютные и относительные погрешности параметров распределения не рассчитываются. Критерием правильности подбора аналитической кривой в этом случае служит условие выполнения неравенства

$$|\bar{x} - x_r| < 0.02\bar{x}, \quad (3.47)$$

где  $\bar{x}$  — среднее значение ряда, полученное по выборке;  $x_r$  — среднее значение, полученное путем расчета графоаналитическим методом.

При выполнении неравенства (3.47) предполагается, что оценка параметров аналитической кривой обеспеченностей произведена с достаточной точностью.

### 3.4. Экономические аспекты определения расчетных вероятностей

Как уже отмечалось в п. 3.1 одной из основных задач гидрологических расчетов при строительстве различного рода сооружений, связанных с использованием водных ресурсов, является определение той или иной гидрологической характеристики расчетной вероятности превышения (обеспеченности). В качестве примера рассмотрим в общих чертах схему определения расчетных обеспеченностей максимальных расходов паводков.

#### 3.4.1. Цели определения расчетных вероятностей

При проектировании любого гидротехнического сооружения необходимо учитывать возможность риска, связанного с прохождением паводков, так как всегда существует вероятность образования такого расхода воды, который превысит проектные возможности сооружения и причинит ущерб как самому сооружению, так и окружающей местности. Поэтому первым шагом при проектировании является изучение повторяемости паводков и получения кривой обеспеченностей максимальных расходов воды. Следующим шагом является определение оптимальной расчетной обеспеченности максимального расхода воды,

гарантирующего нормальную работу сооружения в период его строительства и эксплуатации.

При этом расчетный паводок, закладываемый в проект, должен удовлетворять следующим условиям.

1. Сооружение должно быть осуществимо с экономической точки зрения. Это означает, что затраты на строительство и эксплуатацию его не должны превышать прямые и косвенные затраты в случае возможного ущерба и каких-либо неблагоприятных воздействий для жизни людей и самого сооружения.
2. Средние годовые затраты на сооружение должны быть сведены к минимуму. К средним годовым затратам относятся: а) годовые расходы на строительство сооружения пропорционально продолжительности его эксплуатации; б) годовые расходы на эксплуатацию и ремонт сооружения; в) средний годовой ущерб, причиняемый паводками в проектных условиях (при строительстве сооружения или при эксплуатации), включая расходы на его восстановление в случае повреждения во время прохождения паводка.

### 3.4.2. Определение оптимальных расчетных вероятностей

В практике России средние годовые затраты  $U$  на строительство гидротехнических сооружений включают ряд отдельных элементов [5].

К первому элементу  $U_1$  относится годовая стоимость строительства новых сооружений, рассчитанных на срок службы  $T$  (лет)

$$U_1 = \frac{K_0 + \alpha Q_p}{T}, \quad (3.48)$$

где  $K_0$  – доля расходов на строительство сооружения независимо от расчетного максимального расхода воды;  $\alpha$  – коэффициент, представляющий собой увеличение расходов на строительство при каждом приращении расхода воды;  $Q_p$  – максимальный расход воды соответствующей ежегодной вероятности превышения  $p$ .

Второй элемент  $U_2$  представляет собой годовые затраты на эксплуатацию и текущий ремонт сооружения ( $U_2$  принимается пропорционально общим затратам на строительство,  $K_0 + \alpha Q_p$ )

$$U_2 = r(K_0 + \alpha Q_p), \quad (3.49)$$

где  $r$  – коэффициент пропорциональности между годовыми затратами на эксплуатацию сооружения и общей стоимостью строительства.

Третий элемент  $U_3$  выражает среднюю годовую стоимость ущерба в проектных условиях, включая повреждение сооружения,

$$U_3 = p \times [Y_0 + \eta(K_0 + \alpha Q_p)] \quad (3.50)$$

где  $p$  — вероятность превышения максимального расхода  $Q_p$ ;  $Y_0$  — годовая стоимость ущерба, не считая ущерба, наносимого сооружению при прохождении максимального расхода расчетной вероятности превышения;  $\eta$  — коэффициент пропорциональности между затратами на восстановление проектного сооружения (в случае его повреждения) и общими затратами на строительство.

В формулах (3.48) — (3.50) переменная  $Q_p$  обозначает максимальный расход воды соответствующей вероятности превышения  $p$ , определяемый по кривой обеспеченностей, построенной для места проектирования сооружения. Аналитически  $Q_p$  можно выразить в виде уравнения

$$Q_p = A\bar{Q}p^{-n}, \quad (3.51)$$

где  $\bar{Q}$  — средний многолетний максимальный расход воды;  $A$  и  $n$  — константы, значения которых зависят от характеристик распределения максимальных паводочных расходов.

Подставляя  $(A\bar{Q}p^{-n})$  вместо  $Q_p$  в уравнения (3.48) — (3.50) и суммируя  $U_1$ ,  $U_2$  и  $U_3$ , чтобы получить  $U$ , находим общую годовую сумму затрат, отнесенную к одному сооружению

$$U = K_0\left(\frac{1}{T} + r\right) + p(\eta K_0 + Y_0) + p^{-n}A\bar{Q}\alpha\left(\frac{1}{T} + r\right) + p^{1-n}A\bar{Q}\eta\alpha. \quad (3.52)$$

Минимальная общая годовая стоимость соответствует обеспеченности  $p$ , при которой производная  $dU/dp$  становится равной нулю:

$$\frac{dU}{dp} = \eta K_0 + Y_0 - np^{-(n+1)}A\bar{Q}\alpha\left(\frac{1}{T} + r\right) + (1-n)p^{-n}A\bar{Q}\alpha\eta = 0. \quad (3.53)$$

Уравнение (3.53) решается относительно  $p$ ; получаемое значение представляет собой оптимальную вероятность паводка, которую следует использовать при расчете.

Практика гидротехнического проектирования США в основном та же; главное отличие заключается в том, что величины  $U_1$ ,  $U_2$  и  $U_3$  не определяются по генерализованным связям, а рассчитываются на основе индивидуального изучения каждого отдельного проекта гидротехнического сооружения.

Не вдаваясь в детали определения элементов общих годовых затрат  $U$ , отметим, что решение задачи сводится к определению минимума функции, связывающей годовые затраты с вероятностью превышения максимального расхода паводка.

Расчет производится по следующей схеме.

1. Строится график, связывающий общие годовые затраты  $(U_1 + U_2 + U_3)$  с заданной вероятностью ежегодного превышения (рис. 3.7). При этом вероятность превышения откладывается на оси абсцисс по схеме, принятой в США, т. е. с уменьшением слева направо. Крайняя левая точка соответствует

- вероятности превышения наименьшего расхода, который вызовет разрушения гидротехнического сооружения в предпроектных условиях. Численное значение этой точки на оси ординат соответствует  $U_{3(0)}$ , или общему годовому ущербу в предпроектных условиях (когда сооружение еще не построено).
2. Нанесенные на график точки соединены плавной кривой. Оптимальная расчетная вероятность соответствует значению, при котором общие годовые затраты составляют минимум, т. е. точке на графике, в которой уклон кривой равен нулю. По форме кривой видно, что на довольно большом участке вероятностей разница в годовых затратах незначительная.

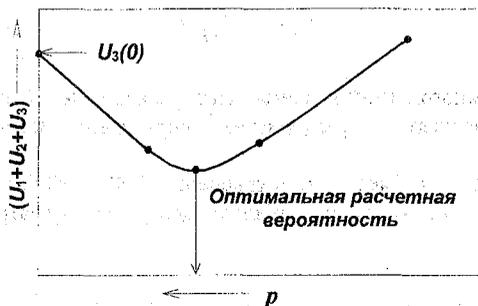


Рис.3.7. Графическое определение вероятности превышения максимальных расходов воды для использования в проектных расчетах.

Описанный анализ производится для каждого значительного сооружения.

Расчетные уравнения, о которых речь шла выше, включают экономические параметры, определение которых часто затруднительно, а порой вообще невозможно. В силу этого, как правило, технико-экономическое обоснование расчетных вероятностей превышения паводочного стока сводится к общим методическим построениям. В результате в ряде стран вошло в практику определение расчетных вероятностей превышения на основе: а) типа и значимости гидротехнических сооружений и б) степени экономического развития окружающей территории, которая бы потерпела потенциальный ущерб от паводка в случае недостаточной разработки проекта гидротехнического сооружения.

В табл. 3.11 приводятся используемые в Российской Федерации классы капитальности гидроэнергетических, навигационных и других гидротехнических сооружений.

В табл.3.12 приводится расчетная вероятность превышения, используемая для каждого класса капитальности гидротехнических сооружений.

Таблица 3.11

**Классы капитальности гидроэнергетических, навигационных и других гидротехнических сооружений**

Гидротехническое сооружение	Номер класса	
	Основные сооружения	Второстепенные сооружения
<i><b>Первая категория</b></i>		
Гидроэлектростанции мощностью более 1 млн. кВт	I	III
Гидротехнические сооружения на основных судоходных внутренних водных путях	II	III
Речные порты с годовым грузооборотом более 3 млн. т	II	III
<i><b>Вторая категория</b></i>		
Гидроэлектростанции мощностью от 301 тыс. до 1 млн. кВт	II	III
Гидротехнические сооружения на судоходных внутренних водных путях	III	IV
Речные порты с годовым грузооборотом от 701 тыс. до 3 млн. т	III	IV
<i><b>Третья категория</b></i>		
Гидроэлектростанции мощностью от 50 до 300 тыс. кВт	III	IV
Гидротехнические сооружения на внутренних водных путях местного значения	III	IV
Речные порты с годовым грузооборотом от 150 до 700 тыс. т	III	IV
<i><b>Четвертая категория</b></i>		
Гидроэлектростанции мощностью менее 50 тыс. кВт	IV	IV
Гидротехнические сооружения на малых внутренних водотоках местного значения	IV	IV
Речные порты и пристани с годовым грузооборотом менее 150 тыс. т	IV	IV

Таблица 3.12

**Связь расчетной вероятности превышения максимальных паводков с классом капитальности гидротехнического сооружения (практика России)**

Класс капитальности гидротехнического сооружения	Расчетная вероятность, %	Период повторяемости <sup>1</sup> , число лет
I	0.01	10 000
II	0.1	1 000
III	0.5	200
IV	1.0	100

<sup>1</sup> Период повторяемости в годах является эквивалентом вероятности превышения.

## 4. ИНТЕРВАЛЬНОЕ ОЦЕНИВАНИЕ ПАРАМЕТРОВ И ПРОВЕРКА СТАТИСТИЧЕСКИХ ГИПОТЕЗ

Для понимания материала, который излагается в этой главе, рассмотрим три теоремы математической статистики. Суть этих теорем состоит в определении закона распределения для случайной величины, которая является функцией других случайных величин.

### 4.1. Распределение $\chi^2$ (хи-квадрат)

**Теорема 4.1.** Если  $X_i$  – независимые случайные величины, подчиняющиеся нормальному распределению, у которых  $m_x$  равно нулю, а  $\sigma_x$  равно единице, то случайная величина  $\chi^2 = \sum_{i=1}^{\nu} X_i^2$  подчиняется распределению  $\chi^2$  (хи-квадрат) с  $\nu$  степенями свободы.

Распределение  $\chi^2$  определяется одним параметром  $\nu$ , который называется числом степеней свободы (значение  $\nu$  равно числу независимых случайных величин под знаком суммы).

Плотность вероятности равна

$$f(\chi^2) = \begin{cases} \left[ 2^{\frac{\nu}{2}} \Gamma(\nu/2) \right]^{-1} e^{-\frac{x}{2}} x^{\frac{\nu}{2}-1} & \text{при } x > 0 \\ 0 & \text{при } x \leq 0, \end{cases} \quad (4.1)$$

где  $\Gamma(\cdot)$  – гамма-функция,  $x$  – значение случайной величины  $\chi^2$ . График плотности вероятности распределения хи-квадрат приведен на рис. 4.1.

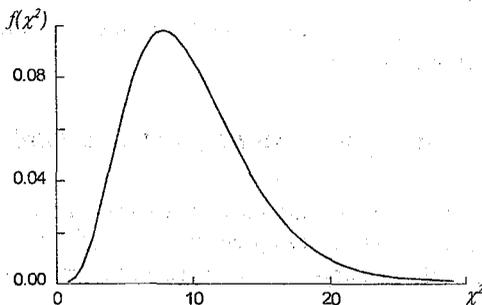


Рис. 4.1. Функция плотности вероятности распределения хи-квадрат при числе степеней свободы  $\nu = 10$ .

Математическое ожидание и дисперсия равны

$$m_x = \nu; \quad (4.2)$$

$$D_x = 2\nu. \quad (4.3)$$

Медиана может быть определена приближенным равенством

$$Me \approx \nu - 0.66. \quad (4.4)$$

Мода при  $\nu \geq 2$  равна

$$Mo = \nu - 2. \quad (4.5)$$

При  $\nu = 1$  мода отсутствует, так как  $f'_\nu = \infty$  при  $x = 0$ .

При увеличении значения  $\nu$  распределение  $\chi^2$  медленно приближается к нормальному.

В приложении 4 приводятся квантили распределения  $\chi^2$  для  $\nu$  от 1 до 30. При  $\nu > 30$  можно использовать аппроксимацию

$$\chi_p^2(\nu) \approx \frac{1}{2} \left( \sqrt{2\nu - 1} + t_p \right)^2, \quad (4.6)$$

где  $t_p$  – квантиль нормального распределения с  $m_t = 0$  и  $\sigma_t = 1$  (см. табл. 2.1);  $p$  – вероятность неперевышения.

Эта аппроксимация не годится при  $p$ , близких к 0 или 100 %. Лучшая аппроксимация в этом случае дается формулой

$$\chi_p^2(\nu) \approx \nu \left[ 1 - 2/(9\nu) + t_p \sqrt{2/(9\nu)} \right]^3. \quad (4.7)$$

Из теоремы 4.1, в частности, следует, что величина  $(n-1)[S_x^2/\sigma_x^2]$  имеет распределение  $\chi^2$  с  $(n-1)$  степенями свободы ( $S_x^2$  и  $\sigma_x^2$  – соответственно выборочная и теоретическая дисперсии).

## 4.2. *t*-распределение (Стьюдента)

**Теорема 4.2.** Если  $Z$  – нормированная нормально распределенная СВ, а  $U$  – независимая от  $Z$  СВ, подчиненная распределению  $\chi^2$  с  $\nu$  степенями свободы, тогда СВ  $t = Z\sqrt{\nu/U}$  подчиняется распределению Стьюдента (У. Госсета) с  $\nu$  степенями свободы.

Распределение Стьюдента называют также *t*-распределением (приложение 5). Плотность вероятности этого распределения определяется равенством

$$f_v(t) = c(\nu) [1 + t^2/\nu]^{-(\nu+1)/2}, \quad (4.8)$$

где  $c(\nu)$  – параметр, зависящий от числа степеней свободы:

$$c(\nu) = \Gamma[(\nu+1)/2] / [\sqrt{\nu\pi} \Gamma(\nu/2)]; \quad (4.9)$$

$\Gamma(\cdot)$  – гамма-функция;  $\pi$  – число пи.

Распределение Стьюдента симметрично. График функции плотности вероятности приведен на рис. 4.2.

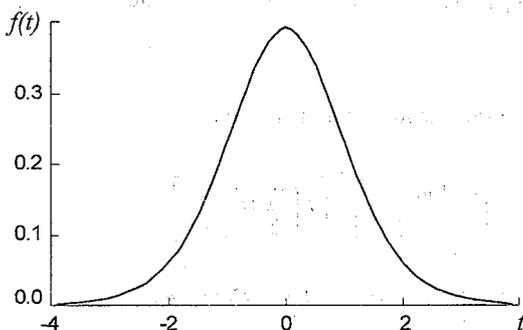


Рис.4.2. Функция плотности вероятности распределения Стьюдента при числе степеней свободы  $\nu = 10$ .

Математическое ожидание  $m_t$ , дисперсия  $D_t$  и среднее квадратическое отклонение  $\sigma_t$  равны

$$m_t = 0, \quad \nu > 1; \quad (4.10)$$

$$D_t = \sigma_t^2 = \nu/(\nu-2), \quad \nu > 2. \quad (4.11)$$

При  $\nu = 1$  распределение Стьюдента приводит к распределению Коши, у которого не существует математического ожидания и дисперсии. С увеличением  $\nu$  распределение Стьюдента асимптотически приближается к нормальному с параметрами  $m_t = 0$  и  $\sigma_t = 1$ .

Из теоремы 4.2, в частности, следует, что величина  $(\bar{x} - m_x)/(S/\sqrt{n})$  имеет распределение Стьюдента, где  $\bar{x}$  и  $S$  – выборочные среднее и СКО,  $n$  – длина выборки.

### 4.3. F-распределение (Фишера)

**Теорема 4.3.** Если  $Z$  и  $U$  независимые СВ, обладающие  $\chi^2$ -распределением с  $\nu_1$  и  $\nu_2$  степенями свободы, то случайная величина  $F = (Z/\nu_1) / (U/\nu_2)$  имеет распределение Фишера с  $\nu_1$  и  $\nu_2$  степенями свободы. Это распределение также называется  $F$ -распределением (приложение 6).

Плотность вероятности определяется равенством

$$f(F) = \begin{cases} c_1(\nu_1, \nu_2) \cdot F^{\frac{\nu_1-2}{2}} / \left(1 + \frac{\nu_1}{\nu_2} F\right)^{\frac{\nu_1+\nu_2}{2}} & \text{при } x > 0 \\ 0 & \text{при } x \leq 0 \end{cases} \quad (4.12)$$

где  $c_1(\nu_1, \nu_2)$  – параметр, зависящий от  $\nu_1$  и  $\nu_2$ :

$$c_1(\nu_1, \nu_2) = \left(\frac{\nu_1}{\nu_2}\right)^{\frac{\nu_1}{2}} \Gamma\left[\frac{\nu_1 + \nu_2}{2}\right] / \left[\Gamma\left(\frac{\nu_1}{2}\right) \Gamma\left(\frac{\nu_2}{2}\right)\right] \quad (4.13)$$

График плотности вероятности  $f(F)$  приведен на рис. 4.3

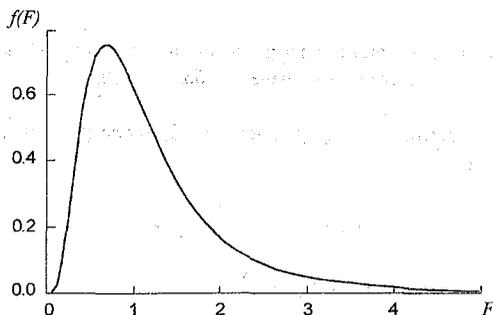


Рис. 4.3. График плотности вероятности  $F$ -распределения при числе степеней свободы  $\nu_1 = 10$  и  $\nu_2 = 10$

Математическое ожидание, дисперсия и мода соответственно равны

$$m_F = \nu_2 / (\nu_2 - 2), \quad \nu_2 > 2; \quad (4.14)$$

$$D_F = 2v_2^2(v_1 + v_2 - 2) / [v_1(v_2 - 2)^2(v_2 - 4)], \quad v_2 > 4; \quad (4.15)$$

$$Mo = v_2(v_1 - 2) / [v_1(v_2 + 2)], \quad v > 1. \quad (4.16)$$

Из теоремы, в частности, следует, что отношение выборочных дисперсий  $S_1^2 / S_2^2$  двух выборок длиной  $m$  и  $n$  будет иметь  $F$ -распределение с числом степеней свободы соответственно  $v_1 = (m - 1)$  и  $v_2 = (n - 1)$ .

#### 4.4. Интервальные оценки параметров распределения

Точечные оценки представляют собой числа, полученные путем подстановки выборочных значений СВ  $x_1, x_2, \dots, x_n$  в формулу для оцениваемого параметра. Например, точечной оценкой математического ожидания  $m_x$  является выборочное среднее  $\bar{x}$ , а точечной оценкой дисперсии  $D$  является выборочная дисперсия  $D^*$ . Однако точечные оценки параметров не дают информации о степени близости оценки  $G^*$  к оцениваемому параметру  $G$ . Подчеркнем, что несмещенная оценка  $G^*$  является случайной величиной, которая имеет математическое ожидание, равное  $G$ , но по отдельным выборкам из генеральной совокупности величина  $G^*$  может существенно отличаться от  $G$  как в большую, так и в меньшую сторону.

Более информативный способ оценивания неизвестных параметров состоит не в определении единичного точечного значения, а в построении интервала, в котором с заданной вероятностью окажется оцениваемый параметр, т.е. в построении так называемой интервальной оценки параметра  $G$ .

Интервальной оценкой параметра  $G$  называется интервал, границы которого  $l_1^*$  и  $l_2^*$  являются функциями выборочных значений  $x_1, x_2, \dots, x_n$  и который с заданной вероятностью  $p$  накрывает оцениваемый параметр  $G$ :

$$P\{l_1^* < G \leq l_2^*\} = p. \quad (4.17)$$

Интервал  $[l_1^*, l_2^*]$  называется доверительным интервалом, а величина  $p$  — доверительной вероятностью. В качестве  $p$  наиболее часто используются значения: 0.9; 0.95 и 0.99.

Используя функцию распределения выборочных значений параметра  $G$ , можно записать вероятность непревышения для  $l_1$  и  $l_2$

$$P\{G^* \leq l_1\} = F(l_1) = (1 - p)/2, \quad (4.18)$$

$$P\{G^* \leq l_2\} = F(l_2) = p + (1 - p)/2 = (1 + p)/2. \quad (4.19)$$

Например, если рассматривается 90%-ный доверительный интервал ( $p = 0.9$ ), то  $F(l_1) = 0.05$ ,  $F(l_2) = 0.95$  или соответственно 5 и 95 % (Рис. 4.4 а). На рис. 4.4

б показан тот же доверительный интервал на графике функции плотности вероятности. Не заштрихованная площадь на этом рисунке составляет 90 % от общей площади подграфика.

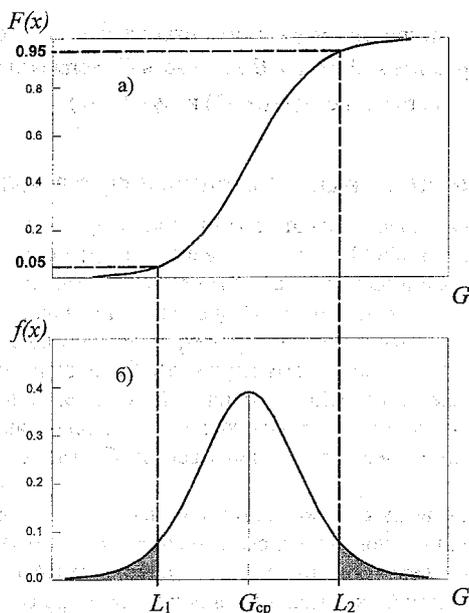


Рис. 4.4. 90 %-ный доверительный интервал  $(l_1, l_2)$  для параметра  $G$  относительно интегральной функции распределения (а) и функции плотности вероятности (б).

#### 4.4.1. Интервальная оценка математического ожидания

Для построения интервальной оценки математического ожидания воспользуемся следствием из теоремы 4.2. В соответствии с этой теоремой величина  $(\bar{x} - m_x)/(S/\sqrt{n})$  имеет распределение Стьюдента с числом степеней свободы  $\nu = n - 1$ . Таким образом, можно записать, что с вероятностью  $p$  она попадает в интервал:  $t_{(1-p)/2} \leq [(\bar{x} - m_x)/(S_x/\sqrt{n})] < t_{(1+p)/2}$ , где  $t_{(1-p)/2}$  и  $t_{(1+p)/2}$  — квантили распределения Стьюдента, соответствующие вероятностям непревышения  $(1-p)/2$  и  $(1+p)/2$ . Однако, учитывая, что распределение Стьюдента симметрично относительно нуля  $t_{(1-p)/2} = -t_{(1+p)/2}$  и, следовательно,

$$-t_{(1+p)/2} \leq [(\bar{x} - m_x)/(S_x/\sqrt{n})] < t_{(1+p)/2}.$$

После преобразования получаем

$$\bar{x} - \frac{t_{(1+p)/2} S_x}{\sqrt{n}} \leq m_x < \bar{x} + \frac{t_{(1+p)/2} S_x}{\sqrt{n}} \quad (4.20)$$

**Пример 4.1.** По выборке длиной  $n = 37$  лет вычислены среднее значение и среднеквадратическое отклонение  $\bar{x} = 12.0$ ;  $S_x = 6.00$ . Найти интервальную оценку математического ожидания ( $m_x$ ) в виде 90%-ного доверительного интервала.

**Решение.**

1. Для 90%-ного доверительного интервала  $p = 0.9$ , следовательно,  $t_{(1+p)/2} = t_{95}$ . По таблице  $t$ -распределения Стьюдента (см. приложение 5) при  $\nu = 36$  получаем  $t_{95} = 1.69$ .
2. Подставляем это значение в формулу (4.20):  

$$12 - (1.69 \cdot 6.00) / \sqrt{36} \leq m_x < 12 + (1.69 \cdot 6.00) / \sqrt{36}.$$
3. В окончательном виде 90%-ный доверительный интервал для  $m_x$  будет иметь вид  $10.31 \leq m_x < 13.69$ .

#### 4.4.2. Интервальная оценка дисперсии

Для построения интервальной оценки дисперсии воспользуемся теоремой 4.1. Из этой теоремы, в частности, следует, что величина  $[(n-1)S_x^2] / \sigma_x^2$  имеет распределение  $\chi^2$  с  $(n-1)$  степенями свободы, и следовательно, можно записать

$$\chi_{(1-p)/2}^2 \leq \frac{(n-1)S_x^2}{\sigma_x^2} < \chi_{(1+p)/2}^2 \quad (4.21)$$

где  $S_x^2 = D^*$  – выборочная дисперсия;  $\sigma_x^2 = D$  – фактическая дисперсия;  $n$  – длина выборки. После преобразований из (4.21) получаем

$$\frac{(n-1)S_x^2}{\chi_{(1+p)/2}^2} < D \leq \frac{(n-1)S_x^2}{\chi_{(1-p)/2}^2} \quad (4.22)$$

Из выражения (4.22) можно получить также интервальную оценку среднеквадратического отклонения

$$S_x \sqrt{(n-1) / \chi_{(1+p)/2}^2} < \sigma \leq S_x \sqrt{(n-1) / \chi_{(1-p)/2}^2} \quad (4.22a)$$

**Пример 4.2.** По выборке длиной  $n = 26$  лет установлено, что  $D^* = 147$ . Найти интервальную оценку дисперсии в виде 95 %-ного доверительного интервала.

### Решение.

1. Для 95 %-ного доверительного интервала  $p = 0.95$ , следовательно,  $\chi^2_{(1+p)/2} = \chi^2_{97.5}$ ;  $\chi^2_{(1-p)/2} = \chi^2_{2.5}$ . По таблице распределения  $\chi^2$  (см. приложение 4) при  $\nu = n - 1 = 25$  получаем  $\chi^2_{97.5} = 40.65$ ,  $\chi^2_{2.5} = 13.12$ .
2. Подставляем эти значения в формулу (4.22):  
 $(25 \cdot 147) / 40.65 < D \leq (25 \cdot 147) / 13.2$ , в окончательном виде:  $90.4 < D \leq 280$ .

**Пример 4.3.** Дано:  $n = 54$ ;  $D^* = 312$ . Требуется найти интервальную оценку дисперсии в виде 90 %-ного доверительного интервала.

### Решение.

1. При  $p = 0.9$   $\chi^2_{(1+p)/2} = \chi^2_{95}$ ,  $\chi^2_{(1-p)/2} = \chi^2_5$ . Так как в данном случае число степеней свободы  $\nu = n - 1 = 53$ , т.е. больше 30, вместо таблицы, приведенной в приложении 4, используем формулу (4.7):  
 $\chi^2_p \approx 53 \cdot \left[ 1 - 2 / (9 \cdot 53) + t_p \sqrt{2 / (9 \cdot 53)} \right]^3$ .
2. Значения квантилей нормального распределения определяем по табл. 2.1:  
 $t_{95} = 1.64$ ;  $t_5 = -1.64$ ; подставляя их в формулу, получаем:  $\chi^2_{95} = 70.9$ ;  
 $\chi^2_5 = 37.3$ .
3. Подставляем эти значения в формулу (4.22):  
 $(53 \cdot 312) / 70.9 < D \leq (53 \cdot 312) / 37.3$ ;  $233 < D \leq 443$ .

## 4.5. Статистическая гипотеза. Критерий статистической гипотезы

*Статистическая гипотеза* есть некоторое предположение относительно свойств генеральной совокупности, из которой извлечена выборка. Например, мы можем предположить, что математическое ожидание  $m_x$  генеральной совокупности равно 12. Тогда среднее значение выборки  $x_1, x_2, x_3, \dots, x_n$  (если  $n$  достаточно велико) должно мало отличаться от 12. В этом случае говорят: «Мы приняли нулевую гипотезу  $H_0: m_x = 12$ ». В качестве альтернативных гипотез ( $H_1$ ) здесь можно использовать следующие:

$$H_1: m_x \neq 12 \quad \text{или} \quad H_1: m_x > 12 \quad \text{или} \quad H_1 < 12.$$

Для проверки гипотез используются специальные тесты, которые называют *критериями*.

Критерий (тест) статистической гипотезы – это правило, позволяющее принять или отвергнуть данную гипотезу.

Для реализации теста используются определенные функции результатов  $g(x_1, x_2, x_3, \dots, x_n)$ , называемые *статистиками*.

Критерий и статистика часто имеют одинаковые названия, например критерий Стьюдента ( $t$ -критерий) базируется на использовании статистики Стьюдента, а критерий Фишера ( $F$ -критерий) – на статистике Фишера.

Продолжим рассмотрение примера. Итак, мы предполагаем, что  $m_x = 12$ . В то же время в нашем распоряжении имеется конкретная выборка с параметра-

ми  $\bar{x}$ ,  $S_x$ . Если наше предположение верно, то разница  $(\bar{x} - m_x)$  должна быть достаточно малой. Эту разницу можно рассматривать в качестве анализируемой статистики. Однако на практике используют другую статистику:

$t = (\bar{x} - m_x)/(S_x/\sqrt{n})$ . Статистика  $t$  является более удобной, так как заранее известно (теорема 4.2), что она подчинена распределению Стьюдента. На использовании этой статистики базируется критерий Стьюдента, который можно применить для проверки нашей нулевой гипотезы  $H_0: m_x = 12$ . С этой целью для конкретной реализации рассчитывают эмпирическое значение статистики Стьюдента:  $t^*$ . Если, например,  $n = 36$ ,  $\bar{x} = 11.0$ ,  $S = 5.00$ , то  $t^* = (\bar{x} - m_x)/(S_x/\sqrt{n}) = (11 - 12)/(5/\sqrt{36}) = 1.2$ .

Величина  $t$  является случайной величиной, и для различных выборок данной длины значение  $t^*$  будет различным. Область возможных значений (ОВЗ) этой статистики – вся числовая ось. ОВЗ делится на две части: область принятия гипотезы и критическую область (Рис. 4.5). Если  $t^*$  попадает в область принятия гипотезы,  $H_0$  не опровергается, если в критическую область,  $H_0$  опровергается.

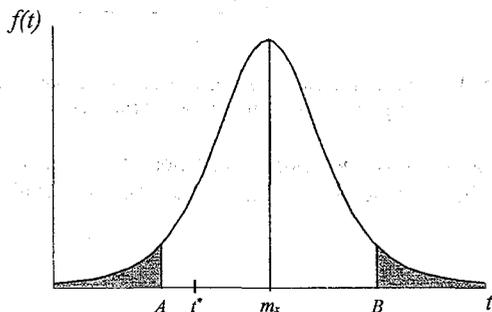


Рис. 4.5. Область принятия гипотезы  $[A, B]$  и критическая область  $(-\infty, A) + (B, +\infty)$  для статистики Стьюдента

Область принятия гипотезы называется доверительной областью (*доверительным интервалом*).

Вероятность по произвольной выборке получить  $t^*$ , которая попадает в доверительную область, геометрически равна незаштрихованной площади на рис. 4.5. Эта вероятность называется *доверительной вероятностью* ( $p_d$ ).

Вероятность попадания  $t^*$  в критическую область выражается равенством  $\alpha = 1 - p_d$ . Вероятность  $\alpha$  называют *уровнем значимости*.

Если критическая область состоит из двух частей, то вместо  $\alpha$  пишут  $2\alpha = 1 - p_d$ , где  $2\alpha$  указывает на то, что уровень значимости двусторонний. Геометрически уровень значимости равен заштрихованной площади (см. рис. 4.5). Различие между односторонним и двусторонним уровнями значимости показано на рис. 4.6.

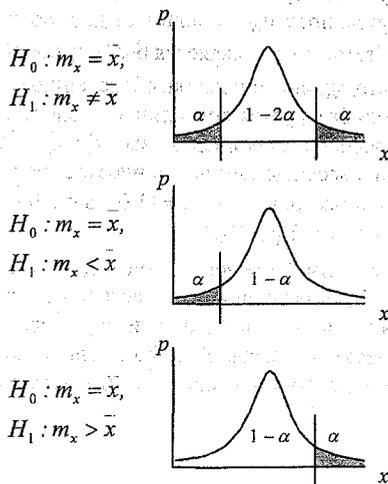


Рис. 4.6. Уровень значимости и доверительная вероятность для различных альтернативных гипотез

При использовании статистических критериев всегда есть возможность совершить ошибку. Типы возможных ошибок представлены в табл. 4.1.

Таблица 4.1

Типы возможных ошибок при проверке статистических гипотез

Гипотеза	Объективно верна	Объективно не верна
Принимается	Правильное решение	Ошибка II рода
Отвергается	Ошибка I рода	Правильное решение

С этой точки зрения *уровень значимости можно трактовать, как вероятность совершить ошибку I рода*. Поэтому на практике в качестве уровня значимости принимается такое достаточно малое значение вероятности, которое в данном случае характеризует практически невероятное (маловероятное) событие. Вообще говоря, назначение уровня значимости не является математической задачей и устанавливается обычно, исходя из тех последствий, которые возможны вследствие совершения ошибки при принятии или отклонении гипо-

тезы. В гидрологической практике наиболее часто используются уровни значимости 5 и 10 %.

Если для нашего примера принять уровень значимости  $2\alpha = 10\%$ , то, используя таблицу распределения Стьюдента (см. приложение 5) при  $\nu = n - 1 = 35$ , получаем теоретическое значение  $t$ -статистики:  $t = 1.69$ . Это означает, что доверительная область представлена отрезком  $[-1.69, +1.69]$ .

Так как  $t = -1.20$  и, значит, попадает в доверительную область, наша нулевая гипотеза не опровергается, т. е. с высокой долей вероятности можно считать, что  $m_x$  действительно равно 12, а разница между  $m_x$  и  $x$  является статистически не значимой.

Следует особо подчеркнуть, что при проверке статистических гипотез нужно избегать категорических формулировок типа: "гипотеза верна" или "гипотеза неверна". Если значение анализируемой статистики не попадает в критическую область, говорят: нулевая гипотеза  $H_0$  не опровергается при принятом уровне значимости  $\alpha$ ; если попадает, говорят:  $H_0$  опровергается при уровне значимости  $\alpha$ , т. е. расхождение эмпирических данных с нулевой гипотезой статистически значимо.

Хотя в данном случае речь шла о конкретном критерии (Стьюдента), основные принципы проверки нулевой гипотезы сохраняются и при использовании других критериев. Различие состоит лишь в том, как строится анализируемая статистика и какому распределению она подчиняется.

Методика проверки статистических гипотез, изложенная в данном разделе, широко применяется в практике гидрологических расчетов, однако это не единственный подход к решению задач такого типа.

Таблица 4.2

Традиционная интерпретация уровней значимости

$\alpha$	Значимость несоответствия эмпирических данных и $H_0$
$> 0.10$	Данные хорошо согласуются с $H_0$
0.05	Возможна значимость. Есть некоторые сомнения в истинности $H_0$
0.02	Значимость. Довольно сильный довод против $H_0$
0.01	Высокая значимость. Гипотеза $H_0$ почти наверняка опровергается

Второй подход заключается в том, что при анализе той или иной статистики границы доверительной области не фиксируются путем назначения уровня значимости. Вместо этого определяется вероятность, соответствующая данному значению выборочной статистики. Таким образом решается обратная задача: каким должен быть предельный уровень значимости, чтобы нулевая гипотеза не опровергалась. В этом случае говорят: гипотеза не опровергается на таком уровне значимости, при этом результат интерпретируется в соответствии с табл. 4.2 [10].

## 4.6. Критерии, используемые для проверки однородности гидрологических рядов

Если в качестве математической модели для описания статистической структуры гидрологических рядов рассматривать случайную величину, то статистические характеристики гидрологического ряда не должны изменяться, т.е. все элементы данной выборки должны относиться к одной генеральной совокупности. По сути, это эквивалентно предположению о неизменности условий формирования стока. Если в результате природных катаклизмов или антропогенной деятельности условия формирования стока изменятся, то это неминуемо приведет к изменению статистических характеристик ряда, например, среднего или дисперсии.

Если статистические характеристики отдельных частей ряда существенно отличаются, то такой ряд является *неоднородным* (нестационарным). При этом надо иметь в виду, что изменение условий формирования стока может сказаться на одних характеристиках стока и почти не отразиться на других. Например, строительство водохранилища многолетнего регулирования наверняка окажет влияние на характеристики максимального стока, но может практически не оказать влияния на среднегодовые расходы.

Статистические методы анализа гидрологических данных, которые были изложены в предыдущих главах, применимы только к однородным рядам, поэтому перед проведением любых статистических расчетов необходимо осуществить проверку однородности исходных гидрологических рядов.

Для проверки однородности гидрологических рядов используются критерии двух типов: *параметрические* и *непараметрические*. В параметрических критериях при построении анализируемой статистики используют выборочные оценки параметров распределения. При этом считается, что исходная выборка относится к генеральной совокупности с известным типом распределения (как правило, принимается нормальный закон распределения).

Непараметрические критерии базируются на использовании непараметрических статистик.

Статистика  $g(x_1, x_2, x_3, \dots, x_n)$  является непараметрической, если ее распределение не зависит от распределения  $X$ . Поэтому наряду с термином "непараметрический" используется более точный: "свободный от распределения". Хотя непараметрические критерии часто менее эффективны, чем стандартные (параметрические), потеря эффективности в этом случае компенсируется более широкими возможностями их применения.

Из рассматриваемых в этом разделе критериев к числу параметрических относятся приведенные ниже критерии Стьюдента и Фишера.

### 4.6.1. Критерий Стьюдента для проверки значимости различия средних значений двух выборок

Пусть  $(x_1, x_2, x_3, \dots, x_n)$  и  $(y_1, y_2, y_3, \dots, y_n)$  – выборки длиной  $m$  и  $n$  из нормальных распределений с неизвестными параметрами  $m_x, \sigma_x$  и  $m_y, \sigma_y$ , но при этом

известно, что  $\sigma_x = \sigma_y$ , т. е. они имеют одинаковое, хотя и неизвестное СКО, обозначим его буквой  $\sigma$  (без индекса).

Если предположить, что эти выборки относятся к одной генеральной совокупности, то разность должна быть близка к нулю, или говоря языком статистики, различие средних должно быть статистически незначимым. На основе этой разности построим статистику.

$$t = (\bar{x} - \bar{y}) / \sigma_{\bar{x} - \bar{y}}, \quad (4.23)$$

где  $\sigma_{\bar{x} - \bar{y}}$  — СКО разности  $(\bar{x} - \bar{y})$ . Как следует из теоремы 4.2, эта статистика подчиняется распределению Стьюдента при  $\nu = (m + n - 2)$ . В математической статистике доказано, что

$$\sigma_{\bar{x} - \bar{y}} = S \sqrt{(m - n) / mn}, \quad (4.24)$$

где  $S$  — эмпирическая оценка  $\sigma$ . Значение  $S$  определяется в зависимости от выборочных значений  $S_x$  и  $S_y$ .

$$S^2 = [(m - 1)S_x^2 + (n - 1)S_y^2] / (m + n - 2). \quad (4.25)$$

В окончательном виде выражение для статистики  $t$  имеет вид

$$t = [(\bar{x} - \bar{y}) / S] \sqrt{mn / (m + n)}. \quad (4.26)$$

В практике гидрологических расчетов эта статистика используется для проверки однородности гидрологических рядов по среднему значению. Исходный ряд делится на две части (две выборки). Если дата возможного нарушения стока не известна, ряд делится пополам. При этом предполагается, что если условия формирования стока не изменились, то разница двух выборочных средних должна быть незначимой. Уровень значимости обычно принимается  $2\alpha = 5\%$  или  $2\alpha = 10\%$ .

**Пример 4.4.** Дана выборка длиной 43 года. Требуется провести проверку ряда на однородность по среднему значению при уровне значимости  $2\alpha = 10\%$ .

**Решение.**

1. Так как дата возможного нарушения стока неизвестна, делим ряд примерно пополам:  $m = 22$ ,  $n = 21$ .
2. Определяем выборочное среднее и СКО для первой половины ряда:  
 $\bar{x} = 20.4$ ;  $S_x = 9.50$ .
3. Определяем выборочное среднее и СКО для второй половины ряда:  
 $\bar{y} = 32.7$ ;  $S_y = 11.4$ .
4. Вычисляем общую дисперсию по формуле (4.25) и СКО:

$$S^2 = [(22 - 1) \cdot (9.50)^2 + (21 - 1) \cdot (11.4)^2] / (22 + 21 - 2) = 109.6; \quad S = \sqrt{109.6} = 10.5.$$

5. Тогда выборочное значение статистики  $t$  с учетом (4.26) равно:

$$t^* = [(20.4 - 32.7) / 10.5] \cdot [(22 \cdot 21) / (22 + 21)] = -3.84$$

6. По приложению 5 определяем теоретическое значение  $t$  при уровне значимости  $2\alpha = 10\%$  и числе степеней свободы  $\nu = 43 - 2 = 41$ . Это значение равно  $t_{1-\alpha} = 1.68$ . Таким образом доверительная область для  $t^*$  определена отрезком  $[-1.68, +1.68]$ . Так как в данном случае  $t^*$  попадает в критическую область, т.е. выходит за пределы доверительной области, гипотеза об однородности ряда по среднему значению опровергается при уровне значимости  $2\alpha = 10\%$ .

Критерий Стьюдента является так называемым стандартным критерием и рекомендуется в большинстве нормативных документов в качестве одного из официальных тестов на однородность. Однако при его использовании следует учитывать три момента.

1. При построении критерия предполагалось, что анализируемые выборки относятся к нормальным совокупностям, а большинство гидрологических рядов, как правило, имеют небольшую положительную асимметрию, что повышает риск совершить ошибку (это замечание относится и к большинству других параметрических критериев).
2. При построении критерия предполагалось, что анализируемые выборки имеют одинаковую (хотя и неизвестную) дисперсию, поэтому перед использованием критерия Стьюдента следует проверить ряд на однородность по дисперсии.
3. В классической статистике длина выборок предполагается значительно большей, чем та, которую мы имеем на практике. Поэтому нельзя исключить ситуацию, когда гипотеза об однородности ряда опровергается из-за недостаточной длины этого ряда. Например, если ряд включает серию маловодных и серию многоводных лет, то при разрезке пополам, в одну выборку могут попасть расходы маловодной фазы, а в другую — многоводной.

#### 4.6.2. Критерий равенства двух дисперсий (Критерий Фишера)

Если  $x_1, x_2, x_3, \dots, x_m$  и  $y_1, y_2, y_3, \dots, y_n$  — выборки из нормальных совокупностей с параметрами  $m_x, \sigma_x$  и  $m_y, \sigma_y$  и если  $\sigma_x = \sigma_y = \sigma$ , то, как следует из теорем 4.1 и 4.3, отношение их выборочных дисперсий  $S_x^2 / S_y^2$  подчиняется распределению Фишера с числом степеней свободы  $\nu_1 = m - 1$  и  $\nu_2 = n - 1$ . Следовательно, при нулевой гипотезе  $H_0: S_x^2 = S_y^2$  и уровне значимости  $2\alpha$  доверительная область для отношения  $S_x^2 / S_y^2$  определяется выражением

$$F_\alpha(\nu_1, \nu_2) \leq (S_x^2 / S_y^2) < F_{1-\alpha}(\nu_1, \nu_2) \quad (4.27)$$

или

$$1 / F_{1-\alpha} \leq (S_x^2 / S_y^2) < F_{1-\alpha} \quad (4.28)$$

Распределение Фишера, как известно, несимметрично, и для того чтобы сократить объем таблицы, их составляют только для значений  $F > 1$ , а при сравнении  $S_x^2$  и  $S_y^2$  в числитель всегда подставляют большую дисперсию. В этом случае доверительная область при уровне значимости  $2\alpha$  определяется выражением

$$1 \leq (S_x^2/S_y^2) < F_{1-\alpha}. \quad (4.29)$$

Этот критерий используется для проверки однородности гидрологических рядов по дисперсии. Исходный ряд делится на две части, затем оцениваются дисперсии для каждой из частей ряда и вычисляется эмпирическое значение статистики Фишера  $F^* = S_1^2/S_2^2$ , где  $S_1^2 > S_2^2$ . Полученное значение  $F^*$  сравнивается с табличным значением  $F_{1-\alpha}$ . Если при принятом уровне значимости оказывается, что  $F^* < F_{1-\alpha}$ , то расхождение дисперсий считается незначимым и гипотеза об однородности ряда по дисперсии не опровергается.

**Пример 4.5.** Пусть дана выборка длиной 43 года. Проверить ряд на однородность при уровне значимости  $2\alpha = 10\%$ .

**Решение.**

1. Разбиваем ряд на две примерно равные части:  $m = 22, n = 21$ .
2. Находим выборочную дисперсию для первой части ряда:  $S_m^2 = 370$ .
3. Находим выборочную дисперсию для второй части ряда:  $S_n^2 = 422$ .
4. Вычисляем эмпирическое значение статистики  $F$  по формуле  $F^* = S_n^2/S_m^2$ ; так как  $S_n > S_m$  получаем  $F^* = 442/370 = 1.19$ .
5. По таблице  $F$ -распределения (см. приложение 6) для  $2\alpha = 10\%$  при  $\nu_1 = n - 1 = 20$  и  $\nu_2 = m - 1 = 21$  определяем теоретическое значение статистики Фишера  $F_{1-\alpha} = F_{95} = 2.09$ . Так как  $F^* < F_{1-\alpha}$ , гипотеза об однородности ряда по дисперсии не опровергается.

Критерий Фишера (так же как и критерий Стьюдента) относится к категории стандартных критериев и рекомендуется в большинстве нормативных документов в качестве официального теста на однородность.

### 4.6.3. Рангово-суммарный критерий Уилкоксона – Манна – Уитни

Данный критерий был предложен Уилкоксоном и в дальнейшем усовершенствован Манном и Уитни.

Критерий используется для проверки нулевой гипотезы о том, что две независимые выборки принадлежат к совокупностям, которые имеют идентичные функции распределения.

Критерий Уилкоксона – Манна – Уитни относится к категории непараметрических критериев и не подразумевает непосредственного расчета выборочных параметров функции распределения.

Достоинством критерия является то, что он не требует обязательной принадлежности выборок к нормальным совокупностям.

Рассмотрим схему построения статистики Уилкоксона. Пусть даны две выборки из совокупностей  $X$  и  $Y$  длиной  $m$  и  $n$  ( $m \leq n$ ). Объединим эти выборки в один ряд и расположим все значения в возрастающем порядке так, что  $y_2 < x_1$ ,  $x_1 < x_2$  и т. д. (табл. 4.3)

Таблица 4.3

Определение рангов (по Уилкоксону)

Ранжированный объединенный ряд ...	$y_1$	$y_2$	$x_1$	$x_2$	$x_3$	$y_3$	$x_4$	$y_4$	$x_5$	$y_5$	$y_6$
Ранг .....	1	2	3	4	5	6	7	8	9	10	11

Каждому значению нового ряда присвоим порядковый номер (ранг). Теперь выпишем значения полученных рангов отдельно для каждой из выборок (табл. 4.4.) и вычислим ранговые суммы  $\omega_1$  и  $\omega_2$ .

Таблица 4.4

Вычисление ранговых сумм (по Уилкоксону)

Выборка 1 ...	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$		Сумма рангов $\omega_1$
Ранг .....	3	4	5	7	9		28
Выборка 2 ...	$y_1$	$y_2$	$y_3$	$y_4$	$y_5$	$y_6$	Сумма рангов $\omega_2$
Ранг .....	1	2	6	8	10	11	38

Если расчет выполнен правильно; то должно выполняться равенство

$$\omega_1 + \omega_2 = N(N+1)/2, \quad (4.30)$$

где  $N = m + n$ . В данном случае  $\omega_1 + \omega_2 = 28 + 38 = 66$ ;  $N(N+1) = 66$ .

В качестве анализируемой статистики  $\omega^*$  Уилкоксон рассматривал ранговую сумму для меньшей выборки, в данном случае  $\omega^* = \omega_1$ . Символически формулу для статистики  $\omega$  можно записать в виде

$$\omega^* = r_1 + r_2 + r_3 + \dots + r_m, \quad (4.31)$$

где  $r_i$  — ранг  $x_i$ .

Для статистики Уилкоксона разработаны таблицы, позволяющие определить доверительный интервал для  $\omega$  (в зависимости от  $m$ ,  $n$  и уровня значимости  $2\alpha$ ). Однако в настоящее время более широкое распространение получила статистика, предложенная Манном и Уитни, так называемая  $U$ -статистика.

Для определения  $U$  вычислим

$$U_1 = mn + m(m+1)/2 - \omega_1, \quad (4.32)$$

$$U_2 = mn + m(m+1)/2 - \omega_2, \quad (4.33)$$

В качестве анализируемой статистики  $U^*$  можно использовать любое из полученных значений ( $U_1$  или  $U_2$ ). Обычно в качестве  $U^*$  принимается меньшее значение. При правильном расчете должно выполняться равенство

$$U_1 + U_2 = mn. \quad (4.34)$$

Отметим, что  $U_1$  представляет собой так называемое *число инверсий* для первой выборки; а  $U_2$  — *число инверсий* для второй выборки. Вообще говоря, их можно получить, не вычисляя ранговые суммы  $\omega_1$  и  $\omega_2$ .

Рассмотрим в качестве объединенного ряда ряд из табл.4.3. Если значение  $x_i$  предшествует значению  $y_j$ , то говорят, что  $x_i$  образует инверсию с  $y_j$ . В данном случае  $x_1$  образует инверсию с  $y_3, y_4, y_5$  и  $y_6$  — всего четыре инверсии. Еще по четыре инверсии образуют  $x_2$  и  $x_3$ ;  $x_4$  образует три инверсии: с  $y_4, y_5$  и  $y_6$  — и т. д. Если обозначить через  $U_1$  общее число инверсий, то для выборки из  $X$  получим:

$$U_1 = 4 + 4 + 4 + 3 + 2 = 17.$$

Аналогичный расчет для выборки из  $Y$  дает

$$U_2 = 5 + 5 + 2 + 1 = 13.$$

Если расчет провести по формулам (4.32) и (4.33), получим те же результаты:

$$U_1 = (5 \cdot 6) + (5 \cdot 6)/2 - 28 = 17,$$

$$U_2 = (5 \cdot 6) + (6 \cdot 7)/2 - 38 = 13.$$

При этом принимают  $U^* = U_2 = 13$ , так как  $U_2 < U_1$ .

Какому из методов расчета  $U$ -статистики отдать предпочтение — вопрос традиций и личного вкуса.

Распределение  $U$ -статистики Манна-Уитни является симметричным с МО и дисперсией:

$$m_U = (mn)/2, \quad (4.35)$$

$$D_U = \sigma_U^2 = mn(m+n+1)/12. \quad (4.36)$$

При  $m \geq 8$  и  $n \geq 8$  функция распределения нормированной величины статистики  $U$  может быть с достаточной точностью аппроксимирована стандартным нормальным распределением. Следовательно, доверительный интервал для статистики  $U$  при уровне значимости  $2\alpha$  имеет вид

$$m_U - t_{1-\alpha} \sigma_U \leq U < m_U + t_{1-\alpha} \sigma_U, \quad (4.37)$$

где  $t_{1-\alpha}$  – квантиль стандартного нормального распределения (см. табл. 2.1);  $m_U$  и  $\sigma_U$  – параметры, определяемые по формулам (4.35) и (4.36).

$U$ -критерий – один из наиболее строгих непараметрических критериев, так как статистика  $U$  – очень сложная функция от среднего значения, эксцесса и асимметрии. В то же время надо учитывать, что  $U$ -критерий наиболее чувствителен к различию выборок по характеристикам положения (например, моды или среднего), но весьма слабо реагирует на разницу дисперсий.

Критерий Уилкоксона – Манна – Уитни можно использовать для проверки однородности гидрологических рядов. В этом случае исходный ряд разбивается на две выборки длиной  $m$  и  $n$ .

**Пример 4.6.** Дан ряд максимальных наводочных расходов на р. Плюсса у дер. Брод за 1936 – 1955 гг.,  $N=17$  (табл. 4.5, графа 3). Проверить его однородность с использованием критерия Уилкоксона – Манна – Уитни при уровне значимости  $2\alpha = 10\%$ .

**Решение.**

1. Располагаем ряд в возрастающем порядке и каждому значению ранжированного ряда ставим в соответствии порядковый номер – ранг (см. табл. 4.5, графы 1, 2).
2. Для каждого члена исходного (неранжированного) ряда выписываем его ранг (см. табл. 4.5, графа 4).

Таблица 4.5

Расчет ранговых сумм (по Уилкоксону) для ряда максимальных наводочных расходов р. Плюсса – дер. Брод

Ранжированный ряд	Ранг	Исходный ряд	Ранг	Ранжированный ряд	Ранг	Исходный ряд	Ранг
1	2	3	4	1	2	3	4
11.5	1	53.4	6	67.4	9	38.4	4
27.0	2	99.2	14	69.0	10	67.4	9
30.5	3	27.0	2	78.0	11	162	17
38.4	4	11.5	1	91.6	12	64.0	8
49.6	5	95.6	13	95.6	13	30.5	3
53.4	6	49.6	5	99.2	14	122	15
60.0	7	69.0	10	122	15	137	16
64.0	8	78.0	11	137	16	91.6	12
				162	17	60.0	7

Сумма: 62

Сумма: 91.

3. Делим исходный ряд на две выборки  $m = 8$  и  $n = 9$  и для каждой выборки определяем ранговую сумму  $\omega_1$  и  $\omega_2$ :  $\omega_1 = 62$ ;  $\omega_2 = 91$ .

4. По формуле (4.30) проверяем правильность расчета:  
 $\omega_1 + \omega_2 = 62 + 91 = 153$ ;  $N(N+1)/2 = 17(17+1)/2 = 153$ .
5. По формулам (4.32) и (4.33) вычисляем  $U_m$  и  $U_n$ :  
 $U_m = 8 \cdot 9 + 8 \cdot (8+1)/2 - 62 = 46$ ;  $U_n = 8 \cdot 9 + 9 \cdot (9+1)/2 - 91 = 26$ .
6. Проверяем правильность расчета по формуле (4.34)  
 $U_m + U_n = 46 + 26 = 72$ ;  $mn = 72$ .
7. В качестве  $U^*$  принимаем  $U^* = U_n = 26$ , поскольку  $U_n < U_m$ .
8. По формулам (4.35) и (4.36) вычисляем МО и СКО:  
 $m_U = mn/2 = 36$ ;  $\sigma_U = \sqrt{8 \cdot 9 \cdot (8+9+1)/12} = 10.4$ .
9. Так как условие  $m \geq 8$  и  $n \geq 8$  выполнено, по табл. 2.1 определяем ординату стандартного нормального распределения при уровне значимости  $2\alpha = 10\%$ :  
 $t_{1-\alpha} = t_{95} = 1.64$ .
10. С учетом формулы (4.37) определяем доверительный интервал для статистики  $U$ :  $(36 - 1.64 \cdot 10.4) \leq U < (36 + 1.64 \cdot 10.4)$ ; округляя до целых, получаем  $19 \leq U < 53$ .
11. Так как  $U^* = 26$ , т.е. эмпирическое значение статистики  $U$  попадает в доверительный интервал, гипотеза об однородности ряда при уровне значимости  $2\alpha = 10\%$  не опровергается.

#### 4.6.4. Ранговый критерий рассеяния Зигеля – Тьюки

Поскольку критерий Уилкоксона со статистикой Манна–Уитни является слабочувствительным к различию выборок по характеристикам рассеяния, Зигель и Тьюки предложили другой ранговый критерий. Формально критерий Зигеля – Тьюки позволяет проверить ту же нулевую гипотезу: две независимые выборки относятся к одной и той же генеральной совокупности. Однако этот критерий является более чувствительным именно к различию дисперсий выборок.

Критерий Зигеля – Тьюки относится к категории непараметрических критериев, так как распределение его тестовой статистики не зависит от типа распределения исходных выборок.

Для построения этой статистики анализируемые выборки объединяются в один ряд длиной  $N = m + n$  и ранжируются в порядке возрастания (табл. 4.6), так что  $y_1 < y_2 < x_1 < x_2 < x_3 < y_3 < y_4$  и т. д.

Таблица 4.6

Определение рангов (по Зигелю – Тьюки)

Ранжированный объединенный ряд ...	$y_1$	$y_2$	$x_1$	$x_2$	$x_3$	$y_3$	$x_4$	$y_4$	$x_5$	$y_5$	$y_6$
Ранг .....	1	4	5	8	9	-	10	7	6	3	2

Таким образом, на первом этапе обработка данных аналогична той, что использовалась для критерия Уилкоксона – Манна – Уитни. Но присваивание ран-

гов происходит иначе. Первый ранг получает наименьшее значение (в данном случае –  $y_1$ ), второй и третий – два самых больших значения ( $y_6$  и  $y_5$ ), ранги 4 и 5 получают следующие наименьшие значения, 6 и 7 – следующие, наибольшие и т. д.

Если общее число наблюдений четное, то среднее значение получает высший ранг, если нечетное – оно не получает никакого ранга, а длину соответствующей выборки следует сократить на единицу. В данном случае  $y_3$  не получает ранга, а длина выборки  $Y$  будет  $n = 6 - 1 = 5$ .

При правильном расчете должно выполняться равенство

$$R_1 + R_2 = (m + n)(m + n + 1)/2, \quad (4.38)$$

где  $R_1$  и  $R_2$  – ранговые суммы для выборок из  $X$  и  $Y$ . В данном случае:  $R_1 + R_2 = 38 + 17 = 55$ ;  $(m + n)(m + n + 1)/2 = (5 + 5)(5 + 5 + 1)/2 = 55$ .

Если анализируемые выборки удовлетворяют условию:  $m$  и  $n > 9$  или  $m > 2$ ,  $n > 20$ , то в качестве тестовой статистики можно использовать величину

$$Z = \frac{2R_1 - m(m + n + 1) + 1}{\sqrt{m(m + n + 1)(n/3)}}, \quad (4.39)$$

где  $R_1$  – сумма рангов меньшей выборки;  $m$  и  $n$  – соответственно длины меньшей и большей выборок. При этом, если  $2R_1 > m(m + n + 1)$ , то в числителе выражения (4.39)  $+1$  заменяют на  $-1$ .

При соблюдении названных условий распределение статистики  $Z$  с достаточной точностью аппроксимируется стандартным нормальным распределением.

Критерий Зигеля – Тьюки можно использовать для проверки однородности гидрологических рядов в отношении характеристик рассеяния. В этом случае исходный гидрологический ряд разбивается на две выборки длиной  $m$  и  $n$ .

Данный критерий следует применять вместо критерия Фишера, если распределение исходного ряда существенно отличается от нормального.

**Пример 4.7.** Дан ряд расходов воды длиной  $N = 17$  лет (табл. 4.7, графа 3). Проверить однородность ряда с использованием критерия Зигеля – Тьюки при уровне значимости  $2\alpha = 10\%$ .

**Решение.**

1. Располагаем ряд в возрастающем порядке и каждому значению ранжированного ряда присваиваем ранг по Зигелю – Тьюки (см. табл. 4.7, графы 1, 2).
2. Для каждого члена исходного (неранжированного) ряда выписываем его ранг (см. табл. 4.7, графа 4).
3. Делим исходный ряд на две выборки. Длина первой выборки  $m = 8$ . Длина второй выборки  $n = 9 - 1 = 8$ , поскольку общее число членов объединенного ряда нечетное и значение 67.4 исключается (не получает ранга).

*Замечание.* Здесь мы немного отклонились от требования  $m$  и  $n > 9$ .

Таблица 4.7.

Расчет ранговых сумм (по Зигелю – Тьюки) для ряда максимальных паводочных расходов р. Плюссы – дер. Брод (1936 – 1955 гг.)

Ранжированный ряд	Ранг	Исходный ряд	Ранг	Ранжированный ряд	Ранг	Исходный ряд	Ранг
1	2	3	4	1	2	3	4
11.5	1	53.4	12	67.4	-	38.4	8
27.0	4	99.2	7	69.0	15	67.4	-
30.5	5	27.0	4	78.0	14	162	2
38.4	8	11.5	1	91.6	11	64.0	16
49.6	9	95.6	10	95.6	10	30.5	5
53.4	12	49.6	9	99.2	7	122	6
60.0	13	69.0	15	122	6	137	3
64.0	16	78.0	14	137	3	91.6	11
				162	2	60.0	13

Сумма: 72

Сумма: 64

- Для каждой выборки определяем ранговую сумму  $R_1 = 72, R_2 = 64$ .
- По формуле (4.38) проверяем правильность расчета рангов:  
 $R_1 + R_2 = 72 + 64 = 136; (m + n) \cdot (m + n + 1) / 2 = (8 + 8) \cdot (8 + 8 + 1) / 2 = 136$ .
- По формуле (4.39) вычисляем статистику  $Z'$ :

$$Z' = \frac{2 \cdot 72 - 8 \cdot (8 + 8 + 1) + 1}{\sqrt{8 \cdot (8 + 8 + 1) \cdot (8/3)}} = \frac{9.0}{19.0} = 0.47.$$

- По табл. 2.1 определяем ординату стандартного нормального распределения при уровне значимости  $2\alpha = 10\%$ :  $t_{1-\alpha} = t_{95} = 1.64$ , и строим доверительный интервал для  $Z$ :  $-1.64 \leq Z < +1.64$
- Так как  $Z' = 0.47$ , т. е. эмпирическое значение статистики  $Z$  попадает в доверительный интервал, гипотеза об однородности ряда при уровне значимости  $2\alpha = 10\%$  не опровергается.

#### 4.7. Критерии согласия

Одной из основных задач при определении расчетных гидрологических характеристик является выбор аналитического выражения для аппроксимации закона распределения исследуемой гидрологической величины.

Выбор считается обоснованным, если имеет место хорошее соответствие эмпирической и аналитической кривых обеспеченностей. На практике такой анализ нередко производится визуально, что, естественно, вносит элемент субъективизма. Более объективно решить поставленную задачу можно, используя

критерии согласия. При этом реализуется схема, аналогичная той, которая применялась при проверке гипотез об однородности гидрологических рядов: выдвигаются нулевая и альтернативная гипотезы; назначается уровень значимости; вычисляется эмпирическое значение тестовой статистики; по результатам расчетов принимается решение.

В данном случае в качестве нулевой принимается гипотеза о соответствии (согласии) аналитической и эмпирической функций распределения, а в качестве альтернативной – их несоответствие.

Степень согласия оценивается с помощью специальных статистик. Различные критерии (тесты) используют для оценки степени согласия разные статистики.

В гидрологической практике наиболее часто применяются критерий  $\chi^2$  (Пирсона), критерий Колмогорова и критерий  $\rho\omega^2$  (Крамера – Мизеса – Смирнова).

#### 4.7.1. Критерий $\chi^2$ (Пирсона)

Критерий  $\chi^2$  был предложен в начале XX в. Карлом Пирсоном и в настоящее время является наиболее распространенным критерием согласия. Для его применения область допустимых значений (ОДЗ) исследуемой СВ  $X$  разбивается на  $k$  интервалов. Число интервалов можно приблизительно определить по формуле (3.2) в зависимости от длины имеющейся выборки.

При назначении границ интервалов наиболее удобно использовать следующую схему.

1. Выбрать аналитическую кривую для аппроксимации закона распределения исследуемой случайной величины.
2. Оценить по имеющейся выборке параметры распределения.
3. Построить на клетчатке вероятности аналитическую кривую обеспеченностей.
4. Разбить ось обеспеченностей на  $k$  равных интервалов.
5. Используя аналитическую кривую, определить границы интервалов для СВ  $X$  в зависимости от границ интервалов на оси обеспеченностей. Например, если  $k = 5$ , то границы интервалов на оси обеспеченностей будут соответствовать значениям 0, 20, 40, 60, 80, 100 %; а ось  $X$  разобьется на интервалы:  $\bar{x}_{20} \leq x$ ;  $\bar{x}_{40} \leq x < \bar{x}_{20}$ ;  $\bar{x}_{60} \leq x < \bar{x}_{40}$ ;  $\bar{x}_{80} \leq x < \bar{x}_{60}$ ;  $x < \bar{x}_{80}$ .

При реализации данной схемы интервалы по оси  $X$  не будут равновеликими, но зато вероятность попадания значения СВ  $X$  в любой интервал будет одинаковой  $p = 1/k$ . Одинаковым будет и теоретическое число случаев попадания значения СВ  $X$  в каждый интервал  $m = n/k$  ( $n$  – длина выборки). Таким образом, принятая схема разбивки ОДЗ СВ  $X$  на интервалы обеспечивает следующие теоретические соотношения:

$$\begin{cases} m_1 = m_2 = m_3 \dots = m_k = m = n/k, \\ p_1 = p_2 = p_3 \dots = p_k = p = m/n. \end{cases} \quad (4.40)$$

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

$$\chi^2 = n \sum_{i=1}^k \frac{(p_i^* - p_i)^2}{p_i}, \quad (4.41)$$

где  $p_i^*$  и  $p_i$  — соответственно эмпирическая и теоретическая вероятности попадания значения СВ в  $i$ -й интервал;  $n$  — длина выборки;  $k$  — число интервалов.

К. Пирсон показал, что закон распределения данной статистики не зависит от вида исходного распределения и при достаточно большом  $n$  хорошо аппроксимируется распределением хи-квадрат (см приложение 4) с числом степеней свободы  $(\nu = k - r - 1)$ ; где  $r$  — число параметров исходного распределения, определяемых по эмпирическим данным.

*Замечание.* Следует отметить, что здесь мы сталкиваемся с некоторой терминологической путаницей. Хотя статистика Пирсона называется  $\chi^2$  (хи-квадрат), ее выборочное распределение лишь приближенно совпадает с распределением  $\chi^2$ .

Учитывая (4.40) и то, что  $p_i^* = m_i^*/n$ , выражение (4.41) можно привести к более удобному для расчетов виду

$$\begin{aligned} \chi^2 &= n \sum_{i=1}^k \frac{(p_i^* - p_i)^2}{p_i} = n \sum_{i=1}^k \left( \frac{m_i^*}{n} - \frac{m}{n} \right)^2 \frac{n}{m} = \sum_{i=1}^k \frac{(m_i^* - m)^2}{m} = \\ &= \frac{1}{m} \left[ \sum_{i=1}^k (m_i^*)^2 + \sum_{i=1}^k \left( \frac{n}{k} \right)^2 - 2 \sum_{i=1}^k m_i^* \frac{n}{k} \right] = \frac{1}{m} \left[ \sum_{i=1}^k (m_i^*)^2 + \frac{n^2}{k} - 2 \frac{n^2}{k} \right]. \end{aligned}$$

В итоге получаем

$$\chi^2 = \frac{1}{m} \sum_{i=1}^k (m_i^*)^2 - n, \quad (4.42)$$

где  $m_i^*$  и  $m$  — соответственно эмпирическое и теоретическое число случаев попадания значения СВ  $X$  в  $i$ -й интервал.

Из выражения (4.41) видно, что чем больше значение статистики  $\chi^2$ , тем больше расхождение между эмпирической и аналитической кривыми. Поэтому при использовании критерия  $\chi^2$  (Пирсона) назначают односторонний уровень значимости (обычно  $\alpha = 5\%$  или  $\alpha = 10\%$ ).

Гипотеза о соответствии (согласии) эмпирической и аналитической кривых обеспеченностей не опровергается, если эмпирическое значение статистики  $(\chi^2)^*$ , определенное по формуле (4.41) или (4.42), не превышает теоретическое значение  $\chi^2$ , соответствующее принятому уровню значимости ( $\alpha$ ), т.е. если

$$(\chi^2)^* \leq \chi_{1-\alpha}^2.$$

Критерий  $\chi^2$  может быть применен при выяснении вопроса о лучшем соответствии одной из нескольких аналитических кривых распределения одному и

тому же эмпирическому ряду. При этом меньшее значение  $\chi^2$  будет свидетельствовать о лучшем соответствии данной функции распределения эмпирическим данным.

Следует подчеркнуть, что использование критерия Пирсона для выборок небольшого объема может приводить к неправильным выводам, а ряды наблюдений за гидрологическими характеристиками обычно не превышают нескольких десятков лет. Это накладывает определенные ограничения на использование данного критерия. Принято считать, что критерий согласия  $\chi^2$  допустимо применять при  $n > 50$ . При этом желательно, чтобы число разрядов (интервалов) было равно 8 – 12 и в каждом разряде было не менее пяти элементов.

**Пример 4.8.** На рис. 4.7 приведены эмпирическая кривая обеспеченностей средних годовых расходов р. Тихвинки - дер. Горелуха ( $n = 98$ ) и аналитическая кривая трехпараметрического гамма-распределения (Крицкого – Менкеля) с параметрами  $\bar{Q} = 19.5 \text{ м}^3/\text{с}$ ,  $C_v = 0.25$ ,  $C_r/C_v = 2$ . Оценить степень соответствия эмпирической и аналитической кривой с помощью критерия согласия  $\chi^2$  (Пирсона) при уровне значимости  $\alpha = 5\%$ .

**Решение.**

1. Для решения этой задачи разбиваем всю ось обеспеченностей на 10 равных интервалов  $k \approx 5 \lg(n) \approx 10$ .
2. Находим значения ординат аналитической кривой обеспеченностей на границах интервалов. Для первого интервала (0 – 10 %) верхняя граница будет равна бесконечности, поскольку функция распределения не ограничена сверху. Для второго интервала (10–20%) верхняя граница равна нижней первого интервала и т.д.; для последнего интервала нижняя граница равна нулю, так как функция распределения в качестве нижнего предела имеет нуль (табл. 4.8):

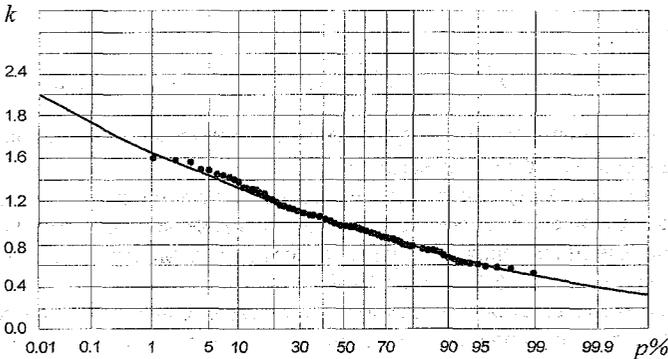


Рис. 4.7. Эмпирическая и аналитическая (Крицкого – Менкеля) кривые обеспеченностей среднегодовых расходов воды для створа р. Тихвинка – дер. Горелуха в модульных коэффициентах. ( $n = 98$ ;  $\bar{Q} = 19,5$ ;  $C_v = 0,25$ ;  $C_r = 2 C_v$ ).

Исходные данные для расчета критерия согласия  $\chi^2$ 

интервал $k$	Обеспечен- ность, %	Границы интервалов	Число случаев попадания расхода воды в интервал $m_i^*$	$(m_i^*)^2$
1	2	3	4	5
1	0 – 10	$\infty - 1.33$	11	121
2	10 – 20	1.33 – 1.20	8	64
3	20 – 30	1.20 – 1.11	10	100
4	30 – 40	1.11 – 1.05	8	64
5	40 – 50	1.05 – 0.98	13	169
6	50 – 60	0.98 – 0.92	11	121
7	60 – 70	0.92 – 0.85	10	100
8	70 – 80	0.85 – 0.79	6	36
9	80 – 90	0.79 – 0.70	11	121
10	90 – 100	0.70 – 0.00	10	100
			Сумма:	996

- Для каждого интервала подсчитываем число случаев попадания в него расхода воды ( $m_i^*$ ) (Табл. 4.9). Если значение расхода приходится на границу интервала, ее относят к верхнему интервалу. Значение  $m_i^*$  заносим в графу 4 табл. 4.8.
- Теоретическое число случаев попадания расхода воды в каждый интервал определяется по выражению  $m = n/k$  и в данном случае  $m = 98/10 = 9.8$ . Эмпирические данные, приведенные в графе 4 табл. 4.8, отличаются от теоретических. Расхождения эти обусловлены либо случайными колебаниями выборочных частот либо несоответствием эмпирической и аналитической кривых обеспеченностей (что и предстоит выяснить).
- Используя формулу (4.42), рассчитаем эмпирическое значение  $\chi^2$ :

$$(\chi^2)^* = \frac{1}{m} \sum_{i=1}^{10} (m_i^*)^2 - n = \frac{1}{9.8} \cdot 996 - 98 = 3.63.$$

- По приложению 4 находим теоретическое значение  $\chi^2$  при числе степеней свободы  $\nu = k - r - 1 = 10 - 3 - 1 = 6$  и уровне значимости  $\alpha = 5\%$ . В данном случае  $\chi_{95}^2 = 12.6$ .
- Поскольку  $(\chi^2)^* < \chi_{95}^2$ , можно сделать вывод о том, что гипотеза о соответствии эмпирической и аналитической кривых обеспеченностей при уровне значимости  $\alpha = 5\%$  не опровергается.

Ранжированный ряд модульных коэффициентов годового стока  
для створа р. Тихвинка - д. Горелуха

№ п/п	№ в группе	$k$	№ п/п	№ в группе	$k$
1	1	1.61	78	1	0.78
2	2	1.58	79	2	0.78
3	3	1.57	80	3	0.78
4	4	1.51	81	4	0.78
5	5	1.48	82	5	0.76
6	6	1.45	83	6	0.75
7	7	1.44	84	7	0.75
8	8	1.42	85	8	0.75
9	9	1.4	86	9	0.75
10	10	1.38	87	10	0.73
11	11	1.34	88	11	0.70
12	1	1.32	89	1	0.68
13	2	1.32	90	2	0.67
14	3	1.31	91	3	0.64
15	4	1.29	92	4	0.64
16	5	1.29	93	5	0.62
17	6	1.27	94	6	0.62
18	7	1.22	95	7	0.59
19	8	1.22	96	8	0.59
			97	9	0.58
20	1	1.18	98	10	0.53
...	...	...			

#### 4.7.2. Критерий Колмогорова

В качестве меры отличия эмпирических данных от аналитической функции распределения А. М. Колмогоров принимает наибольшее по абсолютной величине расхождение между эмпирической  $P^*(x)$  и аналитической  $P(x)$  функциями обеспеченностей

$$\Delta = \max |P^*(x) - P(x)|. \quad (4.43)$$

Вычисления по критерию Колмогорова производятся следующим образом. Для каждого значения СВ  $X$ , полученного по выборке, вычисляются  $P^*(x)$ ,  $P(x)$  и их разности. Затем выбирается наибольшая по модулю разность  $\Delta$  и рассчитывается статистика

$$\lambda^* = \Delta \sqrt{n}, \quad (4.44)$$

где  $n$  — объем выборки. Колмогоров доказал, что функция обеспеченностей статистики  $\lambda$  при достаточно больших  $n$  ( $n > 40$ ) может быть аппроксимирована выражением

$$P\{\lambda^* > \lambda\} \approx P(\lambda) = 2 \sum_{k=1}^n (-1)^{k-1} \exp(-2k^2 \lambda^2). \quad (4.45)$$

Координаты функции (4.45) представлены в приложении 7.

Если значение  $P(\lambda)$  больше принятого уровня значимости, то гипотеза о соответствии эмпирической и аналитической кривых обеспеченностей не опровергается.

Расчетная схема для критерия согласия Колмогорова достаточно проста. Вместе с тем этот критерий имеет определенные недостатки, в частности, при его применении учитывается лишь наибольшее расхождение между эмпирической и аналитической функциями распределения, т. е. используется далеко не вся информация, заключенная в рядах наблюдений. Оценка согласия по одной точке (особенно при небольшой длине рядов) может в отдельных случаях плохо отражать соответствие эмпирических данных теоретическому закону распределения. К тому же наибольшая разность  $\Delta$  обычно отмечается в средней части кривой распределения, в то время как в гидрологической практике нас чаще интересуют ее левая и правая части. Кроме того, данный критерий не учитывает числа параметров, входящих в теоретическую функцию распределения, хотя на практике оценку каждого из них приходится производить по эмпирическим данным.

Отмеченные недостатки не исключают возможности применения критерия Колмогорова в гидрологических расчетах. Статистика  $\lambda$  может быть применена и для сравнения степени соответствия нескольких аналитических функций распределения применительно к одному эмпирическому материалу аналогично тому, как это было указано при использовании критерия  $\chi^2$ .

**Пример 4.9.** В качестве исходных будем использовать данные из примера 4.8. Оценить с помощью критерия Колмогорова степень согласия эмпирических данных с аналитической кривой Крицког — Менкеля (см. рис. 4.7) при уровне значимости  $\alpha = 5\%$ .

**Решение.**

1. Для каждого из 98 значений исходного ряда рассчитываем абсолютное значение разницы между эмпирической и теоретической обеспеченностью. Фрагмент такого расчета представлен в табл. 4.10.
2. Как видно из таблицы, наибольшая разница между  $P^*$  и  $P$  составила 3.75 %, следовательно  $\Delta = 0.0375$ .
3. Рассчитываем эмпирическое значение статистики Колмогорова

$$\lambda^* = \Delta \sqrt{n} = 0.0375 \sqrt{98} = 0.367.$$

4. По приложению 7 определяем ординату функции обеспеченностей Колмогорова:  $P(\lambda^*) = 0.999$ .

5. Так как  $0.999 \gg 0.05$ , т.е.  $P(\lambda^*)$  намного больше принятого уровня значимости, то гипотеза о соответствии эмпирических точек кривой Крицкого – Менкеля не опровергается.

Таблица 4.10

Расчет разностей между эмпирической и теоретической обеспеченностями

№ п/п	$k_i$	$P^*$ %	$P$ %	$ P^* - P $
...	...	...	...	...
40	1.03	40.40	42.14	1.74
41	1.02	41.41	43.72	2.31
42	1.01	42.42	45.22	2.80
43	1.00	43.43	46.70	3.27
44	0.99	44.44	48.19	3.75
45	0.99	45.45	48.19	2.74
46	0.98	46.46	49.68	3.22
47	0.98	47.47	49.68	2.21
48	0.98	48.48	49.68	1.20
49	0.98	49.49	49.68	0.19
...	...	...	...	...

### 4.7.3 Критерий $n\omega^2$ (Крамера-Мизеса-Смирнова)

Тестовой статистикой данного критерия является средний квадрат отклонений между аналитической  $P(x)$  и эмпирической  $P^*(x)$  функциями обеспеченностей по всем значениям случайной величины  $X$ :

$$n\omega^2 = \sum_{i=1}^n [P(x) - P^*(x)]^2 + \frac{1}{12n}, \quad (4.46)$$

где  $P(x)$  рассчитывается по формуле (3.5).

В работе [2] приводится модификация этого критерия, учитывающая особенности построения эмпирических функций обеспеченностей гидрологических характеристик. Суть этого предложения состоит в том, что для расчета эмпирической обеспеченности вместо формулы (3.5) используется формула (3.6). В этом случае выражение для тестовой статистики имеет вид<sup>1</sup>:

<sup>1</sup> Хотя в левой части формулы (4.47) стоит выражение  $(n+1)\omega^2$ , мы будем по-прежнему в качестве названия критерия использовать термин  $n\omega^2$ .

$$(n+1)\omega^2 = \sum_{i=1}^n \left[ P(x) - \frac{m}{n+1} \right]^2 + \frac{n}{12(n+1)^2} \quad (4.47)$$

При  $n > 40$  распределение статистики  $n\omega^2$  не зависит от вида исходного теоретического распределения и близко к некоторому предельному распределению (табл. 4.11).

Таблица 4.11

Значения верхних доверительных границ статистики  $n\omega^2$   
в зависимости от уровня значимости  $\alpha$  %

$\alpha$ , % .....	20	10	5	3	2	1	0.1
$n\omega^2$ .....	0.241	0.347	0.461	0.549	0.620	0.744	1.168

Если эмпирическое значение тестовой статистики, вычисленное с использованием выражения (4.47), оказывается больше теоретического значения  $n\omega^2$  при уровне значимости  $\alpha$  %, то гипотеза о соответствии эмпирической и аналитической функций обеспеченностей опровергается.

Следует отметить, что по сравнению с предыдущими использование данного критерия более трудоемко, но это компенсируется наиболее полным использованием информации, заключенной в фактических рядах наблюдений.

Применение этого критерия, строго говоря, предполагает знание не только типа распределения, но и параметров, обычно оцениваемых по выборке, что может приводить к признанию нуль-гипотезы даже в случае недостаточного согласования аналитических и эмпирических функций обеспеченностей.

**Пример 4.10.** Имеется ряд наблюдений за средними годовыми расходами на р. Сабе – дер. Райково с 1945 по 1988 г. ( $n = 44$ ). Расчет оценок параметров распределения выполнен методом моментов:  $\bar{Q} = 9.09$ ,  $C_s^* = 0.29$ ,  $C_s/C_v = 2$ . Оценить степень соответствия эмпирических данных с аналитической кривой трехпараметрического гамма-распределения (Крицкого – Менкеля) (рис. 4.8) с помощью критерия  $n\omega^2$  при  $\alpha = 5$  %.

**Решение.**

1. В табл. 4.12 приведены результаты расчета квадратов разностей  $\Delta p$  между эмпирической  $p^*$  и теоретической  $p$  обеспеченностями для всех членов выборки.
2. Эмпирическое значение тестовой статистики  $[(n+1)\omega^2]^*$ , рассчитанное по формуле (4.47) и равное 0.04468, сравнивается с табличным (табл. 4.11) значением  $n\omega^2$  при  $\alpha = 5$  %:  $(n\omega^2)_{5\%} = 0.461$ .
3. Так как  $[(n+1)\omega^2]^* \ll (n\omega^2)_{5\%}$ , гипотеза о соответствии эмпирических данных аналитической кривой обеспеченностей Крицкого – Менкеля не опровергается.

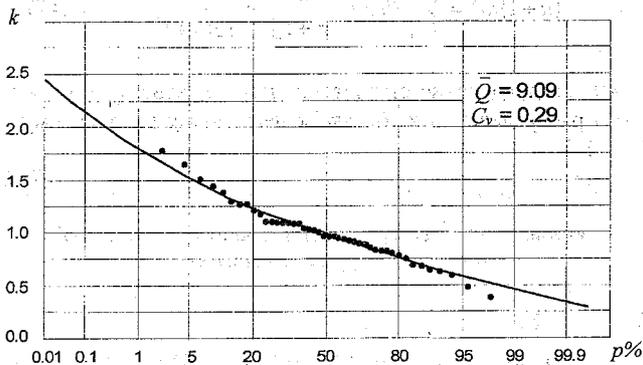


Рис. 4.8. Эмпирическая и аналитическая (Крицкого – Менкеля) кривые обеспеченностей среднегодовых расходов на р.Сабе – дер. Райково.

Таблица 4.12

Расчет квадратов разностей  $\Delta p$  между эмпирическими  $p^*$  и теоретическими  $p$  обеспеченностями (значения  $(\Delta p)^2$  в долях от единицы, т. е.  $(\Delta p)^2 = [(p^* - p) / 100]^2$ )

№ п/п	$k_i$	$p^* \%$	$p \%$	$(\Delta p)^2$	№ п/п	$k_i$	$p^* \%$	$p \%$	$(\Delta p)^2$
1	1.78	2.22	1.73	0.00002	23	0.97	51.11	60.00	0.00012
2	1.65	4.44	2.91	0.00023	24	0.96	53.33	51.43	0.00036
3	1.51	6.67	6.07	0.00004	25	0.94	55.56	54.29	0.00016
...	...	...	...	...	...	...	...	...	...
12	1.11	26.67	32.50	0.00340	34	0.83	75.56	68.75	0.00776
13	1.10	28.89	33.75	0.00236	35	0.80	77.78	72.00	0.00334
...	...	...	...	...	...	...	...	...	...
20	1.02	44.44	43.75	0.00005	42	0.60	93.33	92.48	0.00007
21	1.00	46.67	46.25	0.00002	43	0.49	95.56	97.75	0.00048
22	0.97	48.89	50.00	0.00012	44	0.38	97.78	99.60	0.00033
Сумма:									0.04287

#### 4.8. Критерии случайности

Изложенные ранее 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 случай-

ность 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.

#### 4.8.1. Критерии серий

Под *серией* в данном случае будем понимать всякий участок последовательности, состоящий из элементов одного и того же рода.

К серии из элементов  $a$  будем относить члены последовательности, значения которых превышают выборочное среднее (или медианное) значение, а к серии из элементов  $b$  – значения которых меньше среднего. Длинной серии назовем число элементов, входящих в серию. Обработанная таким образом выборка будет представлять собой последовательность серий различной продолжительности.

##### 4.8.1.1. Критерий общего числа серий

Для проверки гипотезы о том, что данная совокупность сформирована из независимых значений случайной величины, используется статистика  $R$ , представляющая собой сумму серий из элементов  $a$   $r_a$  и элементов  $b$   $r_b$  (длина серий  $i$  при этом значения не имеет).

Для случайных совокупностей статистика  $R = r_a + r_b$  имеет нормальное распределение с параметрами

$$m_R = \frac{n+1}{2}; \quad \sigma_R = \frac{\sqrt{n-1}}{2}. \quad (4.48)$$

Исходя из этого, задавшись уровнем значимости  $\alpha$ , можно построить для  $R$  доверительный интервал (см. п. 4.4):

$$m_R - t_{1-\alpha} \sigma_R < R \leq m_R + t_{1-\alpha} \sigma_R, \quad (4.49)$$

где  $t_{1-\alpha}$  – квантиль стандартного нормального распределения.

Если значение  $R$ , определенное по выборке, попадает в этот интервал, то гипотеза о случайности данной совокупности не опровергается.

**Пример 4.11.** Имеется ряд максимальных в году расходов на р. Систе – дер. Среднее Райково ( $F = 573 \text{ км}^2$ ) за период 1946 – 1984 гг. ( $n = 39$ ). Проверить гипотезу о том, что вероятностная структура гидрологического ряда соответствует модели *случайной величины* с помощью критерия общего числа серий при уровне значимости  $2\alpha = 5\%$ .

### Решение.

1. Рассчитываем средний расход:  $Q_{\max} = 57.3 \text{ м}^3/\text{с}$ . Присвоим значению расхода, имеющему значение, большее  $57.3 \text{ м}^3/\text{с}$ , индекс  $a$ , а меньшему – индекс  $b$  (табл. 4.13 ).

Таблица 4.13

Расчет общего числа серий  $R$  для ряда максимальных расходов воды на р. Систе – дер. Среднее Райково за период 1946 – 1984 гг.

Год	$Q$ м <sup>3</sup> /с	Ин- декс	Число лет $i$	Год	$Q$ м <sup>3</sup> /с	Ин- декс	Число лет $i$	Год	$Q$ м <sup>3</sup> /с	Ин- декс	Число лет $i$
1946	66.0	$a$	1	1959	77.4	$a$	2	1972	29.4	$b$	4
1947	46.1	$b$	1	1960	81.5	$a$		1973	19.4	$b$	
1948	61.1	$a$	1	1961	29.6	$b$	1	1974	26.6	$b$	
1949	40.1	$b$	2	1962	70.2	$a$	1	1975	23.6	$b$	
1950	26.5	$b$		1963	40.8	$b$	1	1976	72.7	$a$	
1951	72.7	$a$	1	1964	84.6	$a$	1	1977	39.9	$b$	5
1952	35.5	$b$	3	1965	49.4	$b$	1	1978	32.5	$b$	
1953	34.1	$b$		1966	79.0	$a$	1	1979	29.4	$b$	
1954	18.0	$b$		1967	36.4	$b$	1	1980	32.5	$b$	
1955	79.6	$a$	3	1968	64.0	$a$	2	1981	35.7	$b$	
1956	68.3	$a$		1969	54.1	$a$		1982	58.2	$a$	
1957	68.7	$a$	1	1970	33.0	$b$	1	1983	67.5	$a$	3
1958	45.1	$b$	1	1971	50.4	$a$	1	1984	67.4	$a$	

2. По данным табл. 4.13 подсчитаем число серий (независимо от их длины  $l$ ), состоящих из элементов  $a$  и  $b$ . В данном случае  $r_a = 12$ ,  $r_b = 11$ . Таким образом, общее число серий в данной выборке равно 23 ( $R = 23$ ).
3. По формулам (4.48) получаем:  $m_R = 20$ ,  $\sigma_R = 3.08$ . По табл. 2.1 определяем квантиль стандартного нормального распределения, соответствующий уровню значимости  $2\alpha = 5\%$ :  $t_{1-\alpha} = t_{97.5} = 1.96$ .
4. Подставив полученные значения в (4.49), рассчитаем доверительный интервал для  $R$ :  $14 \leq R < 26$ .
5. Так как общее число серий, полученное по выборке ( $R^* = 23$ ), попадает в доверительный интервал, можно сделать заключение, что гипотеза о случайности ряда не опровергается.

#### 4.8.1.2. Критерий наибольшей длины серий

Критерий *наибольшей длины серий* использует в качестве тестовой статистики наибольшую длину серии из элементов  $a$  или  $b$ :  $K = l_{\max}$ . Теоретически доказано, что для случайных независимых совокупностей значение  $K$  выражается формулой

$$K_{\alpha} = - \frac{\lg \left[ - \frac{n}{\ln(1-\alpha)} \right]}{\lg 2} - 1, \quad (4.50)$$

где  $\alpha$  представляет собой вероятность (в долях единицы), с которой в выборке объемом  $n$  можно встретить хотя бы одну серию из элементов  $a$  или  $b$  длиной  $K$  и более.

При проверке нулевой гипотезы о случайности выборки эмпирическое значение статистики  $K^*$  сравнивается с теоретическим, рассчитанным по формуле (4.50) при уровне значимости  $\alpha$ . Гипотеза не опровергается если  $K^* < K_{\alpha}$ .

**Пример 4.12.** По данным примера 4.11 проверить гипотезу о том, что вероятностная структура гидрологического ряда соответствует модели случайной величины с помощью критерия наибольшей длины серий  $K$  при уровне значимости  $2\alpha = 5\%$ .

**Решение.**

1. Как видно из табл. 4.13, наибольшую длину имеет серия, состоящая из элементов  $b$  и приходящаяся на период 1977 – 1981 гг., т. е.  $K = 5$ .
2. По формуле (4.50) рассчитаем значение  $K_{\alpha} = 8.6$ .
3. Так как  $K^* < K_{\alpha}$ , гипотеза о случайности ряда при данном уровне значимости не опровергается.

#### 4.8.2. Критерий числа повышений и понижений

Пусть имеется выборка СВ  $X$ :  $x_1, x_2, x_3, \dots, x_n$ . Если  $x_{i-1} < x_i$  будем называть переход от  $x_{i-1}$  к  $x_i$  повышением и обозначать его знаком “+”, если  $x_{i-1} > x_i$  будем называть переход от  $x_{i-1}$  к  $x_i$  понижением и обозначать знаком “-”.

Для случайных последовательностей число повышений (или понижений) распределяется асимптотически нормально с параметрами

$$m_+ = m_- = n/2, \quad D_+ = D_- = (n+1)/12. \quad (9.51)$$

Отсюда, задавшись уровнем значимости  $\alpha$  и учитывая, что математические ожидания и дисперсии числа повышений и понижений равны, можно построить доверительный интервал

$$m - t_{1-\alpha} \sqrt{D} < k + t_{1-\alpha} \sqrt{D}, \quad (9.52)$$

где  $k$  – число повышений ( $k_+$ ) или понижений ( $k_-$ ) в исследуемой выборке;  $t_{1-\alpha}$  – квантиль стандартного нормального распределения.

Если эмпирические значения  $k_+$  и  $k_-$  попадают в доверительный интервал, гипотеза о случайности выборки не опровергается.

**Пример 4.13.** Имеется ряд максимальных в году расходов воды на р. Систе – дер. Среднее Райково за период 1946 – 1984 гг. ( $n = 39$ ). Проверить гипотезу о

случайности ряда с помощью критерия числа повышений и понижений при уровне значимости  $2\alpha = 5\%$ .

**Решение.**

1. В табл. 4.14 приведен исследуемый ряд и в графе 3 индексами "+" и "-" отмечены повышения и понижения.

Таблица 4.14

**Расчет числа повышений "+", понижений "-" и экстремумов "э" для ряда максимальных в году расходов воды в створе р. Сиса - дер. Среднее Райково за период 1946 – 1984 гг.**

Год	$Q, \text{ м}^3/\text{с}$	+/-	э	Год	$Q, \text{ м}^3/\text{с}$	+/-	э	Год	$Q, \text{ м}^3/\text{с}$	+/-	э
1946	66.0	-	э	1959	77.4	+		1972	29.4	-	
1947	46.1	+	э	1960	85.1	+	э	1973	19.4	-	э
1948	61.1	+	э	1961	29.6	-	э	1974	26.6	+	э
1949	40.1	-		1962	70.2	+	э	1975	23.6	-	э
1950	26.5	-	э	1963	40.8	-	э	1976	72.7	+	э
1951	72.7	+	э	1964	84.6	+	э	1977	39.9	-	
1952	35.5	-		1965	49.4	-	э	1978	32.5	-	
1953	34.1	-		1966	79.0	+	э	1979	29.4	-	э
1954	18.0	-	э	1967	36.4	-	э	1980	32.5	+	
1955	79.6	+	э	1968	64.0	+	э	1981	35.7	+	
1956	68.3	-	э	1969	54.1	-		1982	58.2	+	
1957	68.7	+	э	1970	33.0	-	э	1983	67.5	+	э
1958	45.1	-	э	1971	50.4	+	э	1984	67.4	-	

2. Из табл.4.14 видно, что число повышений в данной выборке  $k_+^* = 17$ , число понижений  $k_-^* = 21$ .
3. По формулам (4.51) рассчитываем математическое ожидание и дисперсию для числа повышений и понижений. В данном случае  $m_+ = m_- = 19.5$ ;  $D_+ = D_- = 3.33$ ,  $\sigma_+ = \sigma_- = 1.82$ .
4. По табл. 2.1 определяем квантиль нормированного нормального распределения, соответствующий уровню значимости  $2\alpha = 5\%$ :  $t_{1-\alpha} = t_{97.5} = 1.96$ .
5. Подставив полученные значения в (4.52), рассчитаем доверительный интервал для  $k$ :  $16 \leq k < 23$ .
6. Так как определенные по выборке  $k_+^*$  и  $k_-^*$  попадают в этот интервал, проверяемая гипотеза не отвергается.

**4.8.3. Критерий числа экстремумов**

Экстремум – элемент последовательности  $x_1, x_2, x_3, \dots, x_n$ , для которого выполняется одно из неравенств:

$$x_{i-1} < x_i > x_{i+1} \quad - \text{ максимум,}$$

$$K_{\alpha} = \frac{\lg \left[ -\frac{n}{\ln(1-\alpha)} \right]}{\lg 2} - 1, \quad (4.50)$$

где  $\alpha$  представляет собой вероятность (в долях единицы), с которой в выборке объемом  $n$  можно встретить хотя бы одну серию из элементов  $a$  или  $b$  длиной  $K$  и более.

При проверке нулевой гипотезы о случайности выборки эмпирическое значение статистики  $K^*$  сравнивается с теоретическим, рассчитанным по формуле (4.50) при уровне значимости  $\alpha$ . Гипотеза не опровергается если  $K^* < K_{\alpha}$ .

**Пример 4.12.** По данным примера 4.11 проверить гипотезу о том, что вероятностная структура гидрологического ряда соответствует модели случайной величины с помощью критерия наибольшей длины серий  $K$  при уровне значимости  $2\alpha = 5\%$ .

#### Решение.

1. Как видно из табл. 4.13, наибольшую длину имеет серия, состоящая из элементов  $b$  и приходящаяся на период 1977 – 1981 гг., т. е.  $K = 5$ .
2. По формуле (4.50) рассчитаем значение  $K_{\alpha} = 8.6$ .
3. Так как  $K^* < K_{\alpha}$ , гипотеза о случайности ряда при данном уровне значимости не опровергается.

### 4.8.2. Критерий числа повышений и понижений

Пусть имеется выборка СВ  $X$ :  $x_1, x_2, x_3, \dots, x_n$ . Если  $x_{i-1} < x_i$  будем называть переход от  $x_{i-1}$  к  $x_i$  повышением и обозначать его знаком “+”, если  $x_{i-1} > x_i$  будем называть переход от  $x_{i-1}$  к  $x_i$  понижением и обозначать знаком “-”.

Для случайных последовательностей число повышений (или понижений) распределяется асимптотически нормально с параметрами

$$m_+ = m_- = n/2, \quad D_+ = D_- = (n+1)/12. \quad (9.51)$$

Отсюда, задавшись уровнем значимости  $\alpha$  и учитывая, что математические ожидания и дисперсии числа повышений и понижений равны, можно построить доверительный интервал

$$m - t_{1-\alpha} \sqrt{D} < k + t_{1-\alpha} \sqrt{D}, \quad (9.52)$$

где  $k$  – число повышений ( $k_+$ ) или понижений ( $k_-$ ) в исследуемой выборке;  $t_{1-\alpha}$  – квантиль стандартного нормального распределения.

Если эмпирические значения  $k_+$  и  $k_-$  попадают в доверительный интервал, гипотеза о случайности выборки не опровергается.

**Пример 4.13.** Имеется ряд максимальных в году расходов воды на р. Систе – дер. Среднее Райково за период 1946 – 1984 гг. ( $n = 39$ ). Проверить гипотезу о

случайности ряда с помощью критерия числа повышений и понижений при уровне значимости  $2\alpha = 5\%$ .

**Решение.**

1. В табл. 4.14 приведен исследуемый ряд и в графе 3 индексами "+" и "-" отмечены повышения и понижения.

Таблица 4.14

**Расчет числа повышений "+", понижений "-" и экстремумов "э" для ряда максимальных в году расходов воды в створе р. Сиса - дер. Среднее Райково за период 1946 - 1984 гг.**

Год	Q, м³/с	+/-	э	Год	Q, м³/с	+/-	э	Год	Q, м³/с	+/-	э
1946	66.0			1959	77.4	+		1972	29.4	-	
1947	46.1	-	э	1960	85.1	+	э	1973	19.4	-	э
1948	61.1	+	э	1961	29.6	-	э	1974	26.6	+	э
1949	40.1	-		1962	70.2	+	э	1975	23.6	-	э
1950	26.5	-	э	1963	40.8	-	э	1976	72.7	+	э
1951	72.7	+	э	1964	84.6	+	э	1977	39.9	-	
1952	35.5	-		1965	49.4	-	э	1978	32.5	-	
1953	34.1	-		1966	79.0	+	э	1979	29.4	-	э
1954	18.0	-	э	1967	36.4	-	э	1980	32.5	+	
1955	79.6	+	э	1968	64.0	+	э	1981	35.7	+	
1956	68.3	-	э	1969	54.1	-		1982	58.2	+	
1957	68.7	+	э	1970	33.0	-	э	1983	67.5	+	э
1958	45.1	-	э	1971	50.4	+	э	1984	67.4	-	

2. Из табл. 4.14 видно, что число повышений в данной выборке  $k_+^* = 17$ , число понижений  $k_-^* = 21$ .
3. По формулам (4.51) рассчитываем математическое ожидание и дисперсию для числа повышений и понижений. В данном случае  $m_+ = m_- = 19.5$ ;  $D_+ = D_- = 3.33$ ,  $\sigma_+ = \sigma_- = 1.82$ .
4. По табл. 2.1 определяем квантиль нормированного нормального распределения, соответствующий уровню значимости  $2\alpha = 5\%$ :  $t_{1-\alpha} = t_{97.5} = 1.96$ .
5. Подставив полученные значения в (4.52), рассчитаем доверительный интервал для  $k$ :  $16 \leq k < 23$ .
6. Так как определенные по выборке  $k_+^*$  и  $k_-^*$  попадают в этот интервал, проверяемая гипотеза не опровергается.

### 4.8.3. Критерий числа экстремумов

Экстремум – элемент последовательности  $x_1, x_2, x_3, \dots, x_n$ , для которого выполняется одно из неравенств:

$$x_{i-1} < x_i > x_{i+1} \quad \text{— максимум,}$$

$$x_{i-1} > x_i < x_{i+1} - \text{МИНИМУМ.}$$

Для выборок, представляющих собой последовательности независимых значений случайной величины, число экстремумов распределено асимптотически нормально с параметрами:

$$m_3 = \frac{2n}{3}; \quad D_3 = \frac{16n-29}{90}. \quad (4.53)$$

Проверка гипотезы о случайности ряда производится так же, как в предыдущем случае (см. п. 4.8.2).

**Пример 4.14.** По данным примера 4.13 проверить гипотезу о случайности ряда с помощью критерия числа экстремумов при уровне значимости  $2\alpha = 5\%$ .

**Решение.**

1. Как видно из табл. 4.14, число экстремумов для данной выборки равно 26 ( $k_3^* = 26$ ).
2. По формулам (4.53) рассчитываем значения  $m_3$  и  $D_3$ . В данном случае  $m_3 = 26$ ;  $D_3 = 6.6$ ;  $\sigma_3 = 2.57$ .
3. По табл. 2.1 определяем квантиль нормированного нормального распределения, соответствующий уровню значимости  $2\alpha = 5\%$ :  $t_{1-\alpha} = t_{97.5} = 1.96$ .
4. Рассчитываем доверительный интервал для числа экстремумов:  
 $m_3 - \sigma_3 t_{1-\alpha} < k_3 < m_3 + \sigma_3 t_{1-\alpha} \Rightarrow 21 \leq k_3 < 31$ .
5. Так как  $k_3^* = 26$  и попадает в доверительный интервал, нулевая гипотеза не опровергается.

## 5. Статистический анализ зависимостей между гидрологическими переменными

При изучении природных явлений часто встречаются ситуации, когда СВ  $Y$  зависит от других величин ( $X_1, X_2, X_3, \dots, X_k$ ), и по эмпирическим данным требуется найти аналитическое выражение зависимости  $y = f(x_1, x_2, x_3, \dots, x_k)$ .

Величину  $Y$  называют зависимой переменной, или предиктантом, а ( $X_1, X_2, X_3, \dots, X_k$ ) – независимыми переменными, или предикторами. В специальной литературе, особенно западной, зависимую переменную иногда называют откликом, а независимые переменные – факторами.

Сначала рассмотрим ситуацию, когда  $Y$  зависит только от одного предиктора, т. е.  $y = f(x)$ . В этом случае данные наблюдений можно представить в виде простой таблицы (табл. 5.1) или графика (Рис. 5.1).

Таблица 5.1

Минимальные 30-суточные ( $x$ ) и минимальные суточные ( $y$ ) расходы воды за период открытого русла ( $m^3/c$ ) на р. Вороне у г. Воронежа ( $n = 13$ )

$x \dots$	8.53	10.5	17.0	16.1	26.0	13.4	15.3	13.0	14.6	12.9	11.4	29.7	19.0
$y \dots$	9.53	11.0	17.6	17.9	27.9	14.4	16.0	14.0	15.4	13.3	12.3	30.3	19.9

Обычно точки на графике не лежат на одной линии, даже если  $X$  и  $Y$  связаны функциональной зависимостью, а образуют так называемое поле рассеяния (или корреляционное поле). Причиной рассеяния при наличии функциональной зависимости является неточность измерений. Однако для гидрологической практики более характерна другая ситуация. Гидрологические явления, как правило, обусловлены весьма большим числом факторов, полный учет которых практически не возможен, а в большинстве случаев и не целесообразен, так как только некоторые из них (например, один или два) оказывают решающее влияние на зависимую переменную. Они-то и рассматриваются в качестве предикторов. В такой ситуации при любой точности измерений точки на графике будут образовывать поле рассеяния.

В обоих рассмотренных случаях мы имеем дело с так называемыми стохастическими (вероятностными) зависимостями, когда одному и тому же значению  $X$  соответствует множество значений  $Y$ . Но в первом случае исследователь должен по эмпирическим данным выявить функциональную зависимость  $y = f(x)$ , а во втором случае требуется найти некоторую осредненную кривую  $\bar{y}(x)$ , относительно которой будет наблюдаться минимальный разброс точек.

Линию, относительно которой наблюдается минимальный разброс точек, называют линией регрессии, а соответствующее ей аналитическое выражение – уравнением регрессии<sup>1</sup>.

<sup>1</sup> Строго говоря, регрессией  $Y$  на  $X$  называют условное МО СВ  $Y$  при фиксированном значении СВ  $X = x$ .

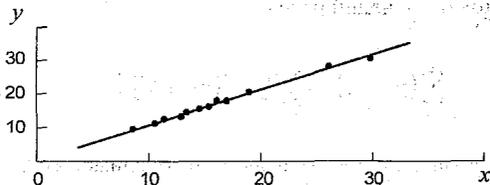


Рис. 5.1. График зависимости минимальных 30-суточных расходов ( $y$ ) от минимальных суточных расходов ( $x$ ) за период открытого русла на р. Вороне у г. Воронежа.

Вид зависимости  $\bar{y}(x)$  может быть известен заранее – из физических или иных соображений. Тогда задача исследователя сводится к оценке параметров аналитического выражения по эмпирическим данным. Например, если зависимость является линейной, т.е.  $\bar{y}(x) = ax + b$ , то достаточно найти такие  $a$  и  $b$ , при которых разброс точек относительно этой прямой будет наименьшим.

Если же вид  $\bar{y}(x)$  не известен, то до определения параметров необходимо выдвинуть гипотезу о виде этой зависимости.

Параметры известной или предполагаемой зависимости оцениваются с помощью излагаемого ниже метода наименьших квадратов.

### 5.1. Метод наименьших квадратов. Уравнение линейной регрессии для двух переменных

Рассмотрим ситуацию, когда зависимость  $\bar{y}(x)$  является линейной. Линейные зависимости довольно часто используются в гидрологической практике, так как аппарат для них хорошо разработан. В пользу линейных моделей говорят и следующие факты.

1. Если вид зависимости  $\bar{y}(x)$  не известен, то в качестве первой модели логично испытать линейную, как наиболее простую.
2. Если зависимость является слабо нелинейной, то область допустимых значений (ОДЗ) переменной  $X$  можно разбить на несколько интервалов, в пределах которых зависимость можно считать линейной.
3. Почти всегда переменные  $Y$  и  $X$  можно преобразовать так, что зависимость между новыми переменными станет линейной.

Искомым уравнением регрессии в случае линейной зависимости является следующее:

$$y_i = ax_i + b. \quad (5.1)$$

Суть метода наименьших квадратов (МНК) состоит в том, чтобы определить такие расчетные параметры  $a$  и  $b$ , при которых сумма квадратов отклонений наблюдаемых значений  $y_i$  от рассчитанных по формуле (5.1) будет иметь минимальное значение.

Сумма квадратов отклонений равна

$$\sum_{i=1}^n (y_i - \bar{y}_i)^2 = \sum_{i=1}^n [y_i - (ax_i + b)]^2. \quad (5.2)$$

Для того чтобы сумма была минимальной, необходимо, чтобы частные производные этого выражения по параметрам  $a$  и  $b$  равнялись нулю:

$$\frac{\partial}{\partial a} \left[ \sum_{i=1}^n (y_i - ax_i - b)^2 \right] = 0, \quad (5.3)$$

$$\frac{\partial}{\partial b} \left[ \sum_{i=1}^n (y_i - ax_i - b)^2 \right] = 0. \quad (5.4)$$

Продифференцировав выражения (5.3) и (5.4), приходим к двум уравнениям, содержащим параметры  $a$  и  $b$  (к так называемым нормальным уравнениям):

$$b \sum_{i=1}^n x_i + a \sum_{i=1}^n x_i^2 = \sum_{i=1}^n x_i y_i, \quad (5.5)$$

$$bn + a \sum_{i=1}^n x_i = \sum_{i=1}^n y_i. \quad (5.6)$$

Решая уравнения (5.5), (5.6) относительно параметров  $a$  и  $b$ , получаем

$$a = \frac{\sum_{i=1}^n [(x_i - \bar{x})(y_i - \bar{y})]}{\sum_{i=1}^n (x_i - \bar{x})^2}, \quad (5.7)$$

$$b = \bar{y} - a\bar{x}, \quad (5.8)$$

где  $\bar{x}$  и  $\bar{y}$  — средние значения соответственно для выборок из  $X$  и  $Y$ .

Коэффициент  $a$ , который характеризует тангенс угла наклона искомой прямой, принято называть коэффициентом регрессии зависимости  $\hat{y}(x)$ . Из выражения (5.7) следует, что коэффициент регрессии можно представить в виде

$$a_{y/x} = r \frac{\sigma_y}{\sigma_x}, \quad (5.9)$$

где  $\sigma_y$  и  $\sigma_x$  – среднеквадратические отклонения выборок из  $Y$  и  $X$ ;  $r$  – выборочный коэффициент парной корреляции, определяемый по формуле

$$r = \frac{\sum_{i=1}^n [(x_i - \bar{x})(y_i - \bar{y})]}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (5.10)$$

Коэффициент  $r$  представляет собой эмпирическую меру линейной зависимости между  $Y$  и  $X$ . Он может принимать значения от  $-1$  до  $+1$ . Если коэффициент корреляции по модулю равен единице, зависимость является функциональной (точки лежат на прямой без разброса). Знак «плюс» указывает на то, что зависимость прямая, знак «минус» – зависимость обратная. Чем ближе абсолютное значение  $r$  к нулю, тем слабее линейная зависимость между  $Y$  и  $X$  (рис. 5.2).

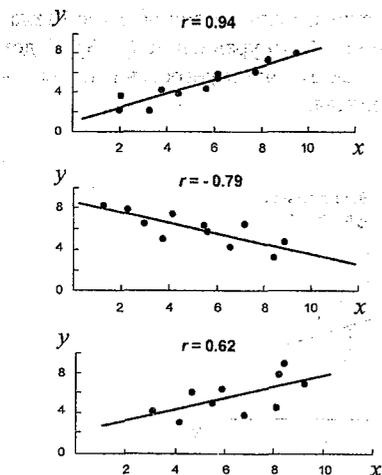


Рис.5.2. Разброс точек на графиках зависимостей вида  $\hat{y}_i = ax_i + b$  при различных значениях коэффициента корреляции.

Приведем еще одну формулу для коэффициента парной корреляции

$$r = \frac{\text{cov}(x, y)}{\sigma_x \sigma_y} \quad (5.11)$$

В этой формуле  $\text{cov}(x, y)$  – ковариация (второй смешанный момент),

$$\text{cov}(x, y) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}). \quad (5.12)$$

Следует отметить, что если при практических расчетах вместо  $\sigma_x$  и  $\sigma_y$  используются их несмещенные эмпирические оценки (см. п.3.2), то при расчете  $r$  величину  $1/n$  в формуле (5.12) необходимо заменить на  $1/(n-1)$ .

С учетом равенств (5.8) и (5.9) уравнение линейной регрессии можно представить в виде

$$(y_i - \bar{y}) = r \frac{\sigma_y}{\sigma_x} (x_i - \bar{x}). \quad (5.13)$$

Полученное равенство есть уравнение линейной регрессии  $y$  по  $x$ . Уравнение линейной регрессии  $x$  по  $y$  получается аналогично и имеет вид:

$$(x_i - \bar{x}) = r \frac{\sigma_x}{\sigma_y} (y_i - \bar{y}). \quad (5.14)$$

Уравнения (5.13) и (5.14) не эквивалентны и не выводятся одно из другого. Проиллюстрируем этот факт с помощью рисунка. Если по оси абсцисс откладывать значения  $X$ , а по оси ординат – значения  $Y$ , то уравнению (5.13) будет соответствовать минимальное отклонение точек от линии регрессии по вертикали (рис. 5.3), а уравнению (5.14) – по горизонтали.

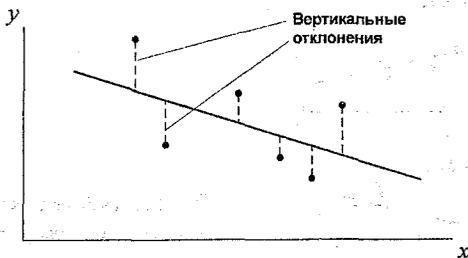


Рис. 5.3. Прямая регрессии  $y$  по  $x$  и вертикальные отклонения, минимизирующие сумму квадратов в МНК.

Нетрудно показать, что прямые, соответствующие уравнениям (5.13) и (5.14), будут пересекаться (рис. 5.4) в точке с координатами  $(\bar{x}, \bar{y})$ . Угол между этими прямыми  $\varphi$  зависит от коэффициента корреляции. При слабой зависимости  $\varphi$  будет большим, при сильной – малым. Угол  $\varphi$  будет максимальным ( $\varphi = 90^\circ$ ), когда  $r = 0$ . Если  $|r| = 1$  (функциональная зависимость), то угол  $\varphi$  равен нулю, т. е. линии сольются в одну.

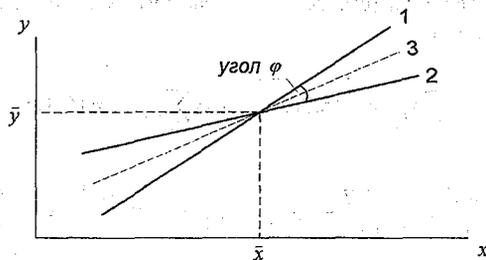


Рис.5.4. Линии линейной регрессии. 1)  $y = f(x)$ ; 2)  $x = f(y)$ ; 3) средняя линия.

Помимо уравнения регрессии в гидрологической практике для аппроксимации линейной зависимости между  $X$  и  $Y$  используется уравнение средней линии, которое соответствует биссектрисе угла  $\varphi$  (см. рис. 5.4). Уравнение средней линии можно получить как из (5.13), так и из (5.14), приравняв  $r$  единице.

**Пример 5.1.** По данным о минимальных 30-суточных ( $x$ ) и минимальных суточных ( $y$ ) расходах воды на р. Вороне у г. Воронежа за 13 лет (см. табл. 5.1) рассчитать параметры уравнения линейной регрессии.

**Решение.**

1. Производим расчет, используя вспомогательную табл. 5.2

Таблица 5.2

Вспомогательная таблица для расчета параметров уравнения линейной регрессии для двух переменных

№ п/п	$x_i$	$y_i$	$\Delta x = (x_i - \bar{x})$	$\Delta y = (y_i - \bar{y})$	$(\Delta x)^2$	$(\Delta y)^2$	$\Delta x \Delta y$	$\Delta x + \Delta y$	$(\Delta x + \Delta y)^2$
1	8.53	9.53	-7.43	-7.36	55.20	54.17	54.68	-14.79	218.74
2	10.5	11.0	-5.46	-5.85	29.81	34.22	31.94	-11.31	127.92
3	17.0	17.6	1.04	0.71	1.08	0.50	0.74	1.75	3.06
4	16.1	17.9	0.14	1.01	0.02	1.02	0.14	1.15	1.32
5	26.0	27.9	10.04	11.01	100.8	121.2	110.5	21.05	443.10
6	13.4	14.4	-2.56	-2.45	6.55	6.00	6.27	-5.01	25.10
7	15.3	16.0	-0.66	-0.89	0.44	0.79	0.59	-1.55	2.44
8	13.0	14.0	-2.96	-2.89	8.76	8.35	8.55	-5.85	34.22
9	14.6	15.4	-1.36	-1.49	1.85	2.22	2.03	-2.85	8.12
10	12.9	13.3	-3.06	-3.59	9.36	12.89	10.99	-6.65	44.22
11	11.4	12.3	-4.56	-4.59	20.79	21.07	20.93	-9.15	83.72
12	29.7	30.3	13.74	13.41	188.8	179.8	184.2	27.15	737.12
13	19.0	19.0	3.04	3.01	9.24	9.06	9.15	6.05	36.60
Сумма	207.5	219.5	0.0	0.0	423.7	451.3	440.8	0.0	1765.6
Средн.	15.96	16.89							

$$\text{Контроль: } \sum (\Delta x)^2 + \sum (\Delta y)^2 + 2 \cdot \sum \Delta x \Delta y = \sum (\Delta x + \Delta y)^2 \rightarrow 432.7 + 451.3 + 2 \cdot 440.8 = 1765.6 \rightarrow 1765.6 = 1765.6.$$

2. На основании таблицы (см. табл. 5.2) рассчитываем параметры уравнения линейной регрессии.

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = 15.96; \quad \bar{y} = \frac{\sum_{i=1}^n y_i}{n} = 16.89;$$

$$\sigma_x^* = \sqrt{\frac{\sum (\Delta x)^2}{n-1}} = \sqrt{\frac{427.7}{12}} = 6.00; \quad \sigma_y^* = \sqrt{\frac{\sum (\Delta y)^2}{n-1}} = \sqrt{\frac{451.3}{12}} = 6.13;$$

$$r = \frac{\sum (\Delta x \Delta y)}{\sqrt{\sum (\Delta x)^2 \sum (\Delta y)^2}} = \frac{440.8}{\sqrt{432.7 \cdot 451.3}} = 0.997;$$

$$a_{y/x} = r \frac{\sigma_y^*}{\sigma_x^*} = 0.997 \frac{6.13}{6.00} = 1.02; \quad b_{y/x} = \bar{y} - a\bar{x} = 16.89 - 1.02 \cdot 15.96 = 0.61;$$

$$y(x) = 1.02x + 0.61.$$

Следует отметить, что аппроксимация по методу наименьших квадратов может сопровождаться дополнительными условиями. Например, можно потребовать, чтобы линия линейной регрессии проходила через начало координат. В этом случае свободный член  $b$  равен нулю, а коэффициент регрессии  $a$  в соответствии с МНК будет определяться формулой

$$a_{y/x} = \frac{1}{n} \sum_{i=1}^n \left( \frac{y_i}{x_i} \right). \quad (5.15)$$

При этом, в связи с потерей одного параметра, сумма квадратов отклонений для уравнения  $y(x) = ax$  будет больше, чем для уравнения  $y(x) = ax + b$ .

## 5.2. Линеаризация нелинейных зависимостей

Если в процессе анализа выясняется, что коэффициент корреляции между переменными  $Y$  и  $X$  близок к нулю, то это указывает на отсутствие между ними линейной зависимости. Однако не исключено, что зависимость между  $Y$  и  $X$  все-таки существует, но носит нелинейный характер. Чтобы это проверить, по эмпирическим данным строится график  $y = f(x)$ . И если точки на графике лежат с небольшим разбросом относительно некоторой кривой, то принимается нелинейная модель. Например, если точки на графике лежат так, как это показано на рис. 5.5а, можно попытаться использовать для аппроксимации зависимости  $y = f(x)$  уравнение экспоненты

$$y = c \cdot e^{ax}, \quad (5.16)$$

где  $a$  и  $c$  – параметры, которые должны определяться по эмпирическим данным.

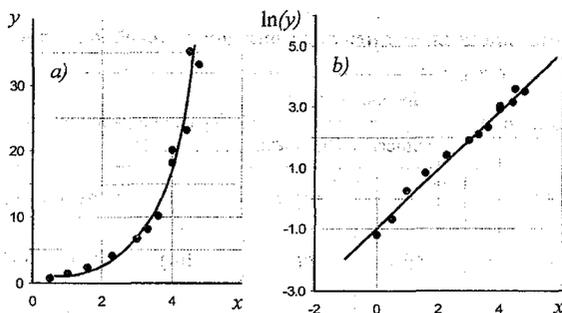


Рис.5.5. График зависимости  $y_i = c \cdot \exp(ax_i)$  без линеаризации (а) и линеаризованный на основе логарифмического преобразования (б).

Метод наименьших квадратов позволяет найти параметры и в случае нелинейной модели, однако можно существенно упростить расчеты, проведя линеаризацию исходного выражения.

Прологарифмировав обе части выражения (5.16), получим

$$\ln(y) = \ln(c) + ax. \quad (5.17)$$

Вместо переменных  $y$  и  $x$  введем новые:  $y^1 = \ln(y)$ ;  $x^1 = x$ . Выражение (5.17) перепишем в виде

$$y^1 = ax^1 + b, \quad (5.18)$$

где  $b = \ln(c)$ .

Как видно из (5.18), новые переменные  $y^1$  и  $x^1$  связаны линейной зависимостью (рис. 5.5 б), и для оценки параметров  $a$  и  $b$  можно применить методикку, изложенную ранее (см. п. 5.1.).

После того как параметры будут найдены, проводят обратное преобразование. В данном случае параметры выражения (5.16) следующие:  $c = e^b$ , а показатель степени  $a$  равен коэффициенту регрессии в выражении (5.18).

В табл. 5.3 приводятся примеры преобразований, применяемых при линеаризации нелинейных зависимостей.

При использовании перечисленных преобразований следует иметь в виду, что параметры исходного выражения, полученные методом наименьших квадратов, будут смещенными, так как в данном случае минимизируется сумма квадратов отклонений преобразованных величин. Однако в большинстве практических ситуаций такой подход оказывается вполне приемлемым, так как ошибка аппроксимации, обусловленная смещенностью параметров, находится в пределах точности исходных данных

## Преобразования, применяемые при линеаризации нелинейных зависимостей

Исходная зависимость	Преобразование		Уравнение в линейной форме
	Абсцисса	Ордината	
$y = c e^{ax}$	$x$	$\ln(y)$	$\ln(y) = ax + \ln(c)$
$y = c x^a$	$\ln(x)$	$\ln(y)$	$\ln(y) = a \ln(x) + \ln(c)$
$y = (a/x) + b$	$1/x$	$y$	$y = a(1/x) + b$
$y = x/(ax + b)$	$x$	$x/y$	$(x/y) = ax + b$
$y = c/(kx + m)$	$x$	$1/y$	$(1/y) = (k/c)x + (m/c)$

### 5.3. Оценка точности уравнения линейной регрессии для двух переменных

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.

Ранее отмечалось, что в качестве меры линейной зависимости между двумя переменными используется коэффициент парной корреляции. В соответствии с действующими в гидрологии нормативными документами зависимость может использоваться для практических расчетов, если  $|r| \geq 0.7$ .

Смысл этого неравенства будет более понятен если рассмотреть величину  $r^2$ , которую называют *коэффициентом детерминации*. Коэффициент детерминации характеризует долю общего разброса относительно среднего значения  $y$ , объясняемую регрессией. Если зависимость функциональная и точки относительно линии регрессии лежат без разброса, то весь разброс относительно  $y$  объясняется регрессией, т. е.  $r^2 = 1$ . При  $r = 0.7$  коэффициент детерминации  $r^2 = 0.49 \approx 0.5$ . Таким образом, в гидрологии зависимость может использоваться для практических расчетов, если регрессия объясняет 50 % или более разброса относительно  $y$ .

Коэффициент корреляции – важный, но не единственный показатель надежности уравнения регрессии. Рассмотрим еще ряд статистических характеристик, позволяющих судить о точности полученного уравнения.

- $\sigma_{y(x)}$  – **стандартная ошибка уравнения линейной регрессии**. По сути эта величина характеризует среднеквадратическое отклонение точек от принятой линии регрессии.

$$\sigma_{y(x)} = \sqrt{\sum_{i=1}^n (y_i - \hat{y}_i)^2 / (n-2)}, \quad (5.19)$$

где  $y_i$  – наблюдаемая величина;  $\hat{y}_i$  – величина, рассчитанная по уравнению регрессии (в зависимости от  $x_i$ ). Параметр  $(n-2)$  называют числом степеней свободы.

Число степеней свободы равно числу наблюдений минус число параметров, определяемых по эмпирическим данным. В данном случае таких параметров два: коэффициент регрессии  $a$  и свободный член  $b$ .

Используя коэффициент корреляции, выражение (5.19) можно представить в виде

$$\sigma_{y(x)} = \sigma_y^* \sqrt{\frac{(1-r^2)(n-1)}{(n-2)}}, \quad (5.20)$$

где  $\sigma_y^*$  – несмещенная оценка СКО для ряда  $Y$ .

Иногда при практических расчетах пренебрегают величиной  $\sqrt{(n-1)/(n-2)}$  и используют более простую формулу

$$\sigma_{y(x)} = \sigma_y^* \sqrt{1-r^2}. \quad (5.21)$$

- $\sigma_r$  – **стандартная ошибка коэффициента парной корреляции**

$$\sigma_r = (1-r^2) / \sqrt{n-1}. \quad (5.22)$$

При малой длине выборок ( $n < 25$ ) В.Н. Романовский рекомендует вводить в формулу (5.22) поправочный коэффициент

$$\sigma_r = \frac{1-r^2}{\sqrt{n-1}} \sqrt{1 + \frac{11r^2}{2n} + \frac{75r^2-13}{2n^2}}. \quad (5.23)$$

- $\sigma_a$  – **стандартная ошибка коэффициента регрессии**

$$\sigma_a = \sigma_{y(x)} / \sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}. \quad (5.24)$$

С учетом выражения (5.20) формулу (5.24) можно представить в виде

$$\sigma_a = \sigma_y^* \sqrt{\frac{1-r^2}{n-2}}, \quad (5.25)$$

где  $\sigma_x^*$  и  $\sigma_y^*$  – несмещенные оценки СКО соответственно для ряда  $X$  и  $Y$ .

- $\sigma_b$  – стандартная ошибка свободного члена

$$\sigma_b = \sigma_{y(x)} \times \sqrt{\frac{\sum_{i=1}^n (x_i)^2}{n \sum_{i=1}^n (x_i - \bar{x})^2}}, \quad (5.26)$$

или с учетом формулы (5.20)

$$\sigma_b = \sigma_y^* \times \sqrt{\frac{1-r^2}{n-2}} \times \sqrt{1 + \left(\frac{\bar{x}}{\sigma_x^*}\right)^2}. \quad (5.27)$$

Сформулируем теперь полный набор требований, предъявляемых в гидрологии к уравнению линейной регрессии.

Уравнение можно рекомендовать для практических расчетов, если выполнены следующие условия:

$$n \geq 10; \quad |r| \geq 0,7; \quad |r|/\sigma_r \geq 2; \quad |a|/\sigma_a \geq 2. \quad (5.28)$$

Желательным (но не обязательным) является условие  $(|b|/\sigma_b) \geq 2$ . При малых значениях параметра  $b$  это требование почти никогда не выполняется.

Помимо комплекса условий (5.28) при анализе уравнения линейной регрессии следует учитывать, что истинное значение  $y_i$ , соответствующее аргументу  $x_i$ , будет отличаться от  $\hat{y}_i$ , полученного по уравнению линейной регрессии (5.1), на некоторую величину  $\varepsilon_i$ . При этом предполагается, что  $\varepsilon_i$  является нормально распределенной случайной величиной с нулевым средним значением и среднеквадратическим отклонением  $\sigma_{y(x)}$ .

Рассмотренная система оценок позволяет, в большинстве случаев, составить правильное представление о надежности полученного уравнения регрессии и принять решение о возможности его применения для практических расчетов. Однако иногда такого анализа оказывается недостаточно. Проиллюстрируем это с помощью рис. 5.6.

Как видно на рисунке, в данном случае эмпирические точки расположены неравномерно. Точки основной группы образуют поле рассеяния в диапазоне значений  $X$  от 3 до 15, а две точки расположены в области больших значений  $X$ . Первая имеет координаты  $x = 25, y = 35$ , вторая –  $x = 35, y = 17$ .

Значение коэффициента корреляции, полученное для точек основной группы значительно меньше 0.7 ( $r = 0.45$ ). Но если к основной группе точек добавить точку № 1, коэффициент корреляции резко возрастет ( $r = 0.82$ ). Использование всех данных, включая и точку № 2, вновь приводит к значительному из-

менению коэффициента корреляции ( $r = 0.66$ ). Заметно изменится и положение линии регрессии.

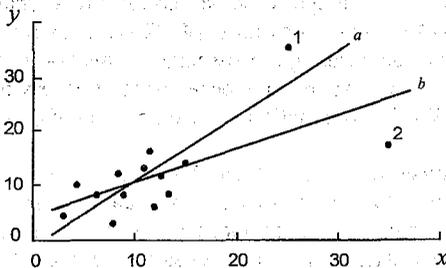


Рис. 5.6. Линии линейной регрессии  $\hat{y}(x)$ :  $a$  – без учета точки №2 ( $r = 0.83$ );  
 $b$  – с учетом всех точек ( $r = 0.66$ )

Таким образом, здесь мы сталкиваемся с ситуацией, когда добавление одной или двух точек может существенно повлиять на параметры уравнения регрессии. Если в данном случае ориентироваться только на систему ограничений (5.28), можно принять ошибочное решение.

Действительно, как видно из табл. 5.4, во втором варианте, когда из совокупности данных исключена точка № 2, условие (5.28) выполняется.

Таблица 5.4

Параметры уравнения линейной регрессии для зависимости, представленной на рис. 5.6

Параметр	Вариант № 1 (Основная группа без точек №1 и 2)	Вариант № 2 (Основная группа плюс точка № 1)	Вариант № 3 (Все точки)
$n$	12	13	14
$r$	0.45	0.82	0.66
$a$	0.48	1.18	0.61
$b$	4.86	-1.34	4.10
$\sigma_{y(x)}$	3.63	4.82	6.13
$\sigma_r$	0.24	0.096	0.17
$\sigma_a$	0.30	0.25	0.20
$\sigma_b$	3.06	3.19	3.05
$r/\sigma_r$	1.88	8.54	3.88
$a/\sigma_a$	1.60	4.72	3.05
$b/\sigma_b$	1.59	0.42	1.34
Выполнение условий (5.28)	Не выполняются	Выполняются	Не выполняются

Основная причина возникших здесь трудностей состоит в том, что выборочные точки группируются крайне неравномерно. В подобных ситуациях следует провести расчет параметров уравнения регрессии с учетом и без учета «тяжелой» точки, и, если они существенно различаются, уравнение не следует использовать в качестве расчетного по крайней мере до тех пор, пока не появятся дополнительные данные. При этом надо понимать, что дополнительные данные могут как подтвердить, так и опровергнуть гипотезу о наличии высокой корреляции между  $X$  и  $Y$ , но в любом случае решение будет более обоснованным.

Ситуации, подобные той, которую мы рассмотрели, не так уж редки в гидрологической практике. Можно назвать по крайней мере две причины, по которым эмпирические точки могут группироваться очень неравномерно.

1. Метод наименьших квадратов подразумевает, что ряды  $X$  и  $Y$  являются выборками из нормальных совокупностей, что обеспечивает примерно равное количество больших и малых значений, а выборочная асимметрия гидрологических величин (особенно при небольших  $n$ ) может быть весьма значительной.
2. Гидрологи, как правило, работают в условиях так называемого пассивного эксперимента и не могут по своему усмотрению изменять независимую переменную. В результате всегда есть риск получить нерепрезентативные выборки, особенно если длина выборки  $n$  недостаточно велика.

## **5.4. Интервальная оценка и оценка значимости параметров линейной регрессии для двух переменных**

### **5.4.1. Интервальная оценка коэффициента парной корреляции. Z-преобразование Фишера**

Коэффициент парной корреляции, определяемый по выборкам конечной длины, является случайной величиной. Если значение  $r$  не очень велико и длина выборки не превышает 40 лет, то распределение выборочных коэффициентов корреляции хорошо аппроксимируется нормальным законом со среднеквадратическим отклонением  $\sigma_r$ , определяемым по формуле (5.22). В этом случае доверительный интервал для истинного коэффициента корреляции можно представить в виде

$$r^* - t_{1-\alpha} \sigma_r^* \leq r < r^* + t_{1-\alpha} \sigma_r^*, \quad (5.29)$$

где  $r^*$  – выборочный коэффициент парной корреляции;  $t_{1-\alpha}$  – квантиль стандартного нормального распределения (см. табл. 2.1), соответствующий двустороннему уровню значимости  $2\alpha$ .

**Пример 5.2.** При  $r^* = 0.82$ ;  $\sigma_r^* = 0.096$ ;  $n = 13$  (см. табл. 5.4, вариант № 2) найти интервальную оценку коэффициента корреляции в виде 90%-ного доверительного интервала.

**Решение.**

1. Определяем, какой двусторонний уровень значимости соответствует принятому доверительному интервалу:  $2\alpha = 1 - 0.90 = 0.10$  (или 10 %).
2. Определяем  $\alpha$ :  $\alpha = 0.10/2 = 0.05$  (или 5 %).
3. По табл. 2.1 находим квантиль  $t_{1-\alpha}$ :  $t_{1-\alpha} = t_{95} = 1.64$ .
4. Строим доверительный интервал для  $r$ :  
 $0.82 - 1.64 \cdot 0.096 \leq r < 0.82 + 1.64 \cdot 0.096 \rightarrow 0.66 \leq r < 0.98$ .

При больших значениях  $r$  и малой длине выборок ( $r > 0.4$ ;  $n < 40$ ), распределение выборочных значений коэффициента корреляции при уменьшении  $n$  начинает все больше отклоняться от нормального, приобретая отрицательную асимметрию.

В этом случае при построении доверительного интервала рекомендуется применять  $Z$ -преобразование Фишера, основанное на использовании специальной статистики  $Z$ , функционально связанной с  $r$  выражением

$$Z = \frac{1}{2} \ln \left( \frac{1+r}{1-r} \right). \quad (5.30)$$

В отличие от  $r$  статистика  $Z$  имеет нормальное распределение даже при малых  $n$ . СКО  $Z$  определяется формулой

$$\sigma_z = 1/\sqrt{n-3}. \quad (5.31)$$

Схема построения интервальной оценки  $r$  при использовании преобразования Фишера следующая.

1. В зависимости от  $r$  по формуле (5.30) рассчитывается эмпирическое значение  $Z$ .
2. По формуле (5.31) вычисляется  $\sigma_z^*$ .
3. Строится доверительный интервал для  $Z$ :

$$Z^* - t_{1-\alpha} \sigma_z^* \leq Z < Z^* + t_{1-\alpha} \sigma_z^*. \quad (5.32)$$

4. Строится доверительный интервал для коэффициента корреляции путем обратного перехода от  $Z$  к  $r$ . Формулу обратного перехода от  $Z$  к  $r$  можно получить из (5.30):

$$r = (e^{2z} - 1)/(e^{2z} + 1). \quad (5.33)$$

**Пример 5.3.** По данным примера 5.2 найти интервальную оценку коэффициента корреляции в виде 90%-ного доверительного интервала (применить  $Z$ -преобразование Фишера).

**Решение.** Последовательно реализуем изложенную выше схему.

$$1. Z^* = \frac{1}{2} \ln \left( \frac{1+0.82}{1-0.82} \right) = 1.157.$$

$$2. \sigma_z^* = 1/\sqrt{13-3} = 0.316.$$

$$3. (1.57 - 1.64 \cdot 0.316) \leq Z < (1.57 + 1.64 \cdot 0.316) \rightarrow 0.639 \leq Z < 1.675.$$

$$4. \frac{\exp(2 \cdot 0.639) - 1}{\exp(2 \cdot 0.639) + 1} \leq r < \frac{\exp(2 \cdot 1.675) - 1}{\exp(2 \cdot 1.675) + 1} \rightarrow 0.56 \leq r < 0.93.$$

Так как в данном примере величина  $r^*$  существенно больше 0.4, а  $n < 40$ , интервальные оценки  $r$ , полученные с учетом и без учета  $Z$ -преобразования Фишера, заметно различаются. При этом, последняя оценка является более корректной.

Следует также отметить, что при  $r < 0.4$  и  $n > 40$  интервальная оценка  $r$ , полученная с учетом  $Z$ -преобразований Фишера, будет совпадать с оценкой, полученной без него. Поэтому  $Z$ -преобразование Фишера можно рекомендовать для любых сочетаний  $r$  и  $n$ .

Выборочный коэффициент корреляции можно использовать для проверки значимости линейной зависимости между  $X$  и  $Y$ . В этом случае выдвигается нулевая гипотеза  $H_0: r = 0$ . Гипотеза опровергается, если

$$\frac{|r^*|}{\sigma_r^*} > t_{1-\alpha}, \quad (5.34)$$

и связь считается статистически значимой. Если условие (5.34) не выполняется, связь считается статистически незначимой, т. е. линейная зависимость между  $X$  и  $Y$  отсутствует.

Заметим, что при уровне значимости  $2\alpha = 5\%$  получаем  $t_{1-\alpha} = t_{97.5} = 1.96 \approx 2$  и выражение (5.34) в этом случае идентично третьему выражению системы ограничений (5.28).

#### 5.4.2. Интервальная оценка коэффициента регрессии

Если предположить, что разброс наблюдений относительно линии регрессии нормален, то доверительный интервал для коэффициента регрессии имеет вид

$$a^* - t_{1-\alpha} \sigma_a < a \leq a^* + t_{1-\alpha} \sigma_a, \quad (5.35)$$

где  $a^*$  – эмпирическое значение коэффициента регрессии, определенное по формуле (5.7) или (5.9);  $\sigma_a$  – стандартная ошибка коэффициента регрессии, определенная по формуле (5.24) или (5.25);  $t_{1-\alpha}$  – квантиль распределения Стьюдента (см. приложение 5), соответствующий двустороннему уровню значимости  $2\alpha$  при числе степеней свободы  $\nu = n - 2$ .

**Пример 5.4.** При  $a = 1.18$ ;  $\sigma_a = 0.25$ ;  $n = 13$  (см. табл. 5.4, вариант № 2) найти интервальную оценку коэффициента регрессии в виде 90%-ного доверительного интервала.

**Решение.**

1. Определяем, какой двусторонний уровень значимости соответствует принятому доверительному интервалу:  $2\alpha = 1 - 0.90 = 0.10$  (или 10 %).
2. Определяем  $\alpha$ :  $\alpha = 0.10/2 = 0.05$  (или 5 %).
3. По таблице распределения Стьюдента (см. приложение 5) при  $\nu = 13 - 2 = 11$  находим квантиль  $t_{1-\alpha}$ :  $t_{1-\alpha} = t_{0.95} = 1.80$ .
4. Строим доверительный интервал для  $a$ :  
 $1.18 - 1.80 \cdot 0.25 \leq a < 1.18 + 1.80 \cdot 0.25 \rightarrow 0.75 \leq a < 1.63$ .

При проверке значимости коэффициента регрессии выдвигается нулевая гипотеза  $H_0: a = 0$ . Гипотеза опровергается если

$$|t_a^*| > t_{1-\alpha}, \quad (5.36)$$

где  $t_a^*$  – эмпирическое значение статистики Стьюдента, определяемое по формуле

$$t_a^* = a^* / \sigma_a. \quad (5.37)$$

Если равенство (5.36) выполняется коэффициент регрессии считается статистически значимым, в противном случае коэффициент  $a^*$  является статистически незначимым и линейная связь между  $X$  и  $Y$  отсутствует.

Анализируя выражение (5.37), можно заключить, что при имеющейся длине рядов (20 – 60 членов) четвертое выражение в системе ограничений (5.28) приблизительно соответствует уровню значимости  $2\alpha = 5\%$ .

### 5.4.3. Интервальная оценка свободного члена

Если предположить, что разброс наблюдений относительно линии регрессии нормален, то доверительный интервал для свободного члена имеет вид

$$b^* - t_{1-\alpha} \sigma_b < b \leq b^* + t_{1-\alpha} \sigma_b, \quad (5.38)$$

где  $b^*$  – эмпирическое значение свободного члена, определенное по формуле (5.8);  $\sigma_b$  – стандартная ошибка свободного члена, определенная по формуле (5.26) или (5.27);  $t_{1-\alpha}$  – квантиль распределения Стьюдента (см. приложение 5),

соответствующий двустороннему уровню значимости  $2\alpha$  при числе степеней свободы  $\nu = n - 2$ .

При проверке значимости свободного члена выдвигается нулевая гипотеза  $H_0: b = 0$ . Гипотеза отвергается если

$$|t_b^*| > t_{1-\alpha}, \quad (5.39)$$

где  $t_b^*$  – эмпирическое значение статистики Стьюдента, определяемое по формуле

$$t_b^* = b^* / \sigma_b. \quad (5.40)$$

Если равенство (5.39) выполняется, свободный член считается статистически значимым, в противном случае коэффициент  $b^*$  является статистически незначимым и для аппроксимации зависимости между  $X$  и  $Y$  вместо выражения  $\hat{y}_i = ax_i + b$  следует использовать выражение  $\hat{y}_i = ax_i$ .

**Пример 5.5.** При  $b = -1.34$ ;  $\sigma_b = 3.19$ ;  $n = 13$  (см. табл. 5.4, вариант № 2) оценить значимость свободного члена; уровень значимости принять  $2\alpha = 10\%$ .

**Решение.**

1. По формуле (5.40) рассчитываем эмпирическое значение статистики Стьюдента  $t_b^* = |-1.34/3.19| = 0.42$ .
2. По таблице распределения Стьюдента (см. приложение 5) при  $\nu = 13 - 2 = 11$  находим квантиль  $t_{1-\alpha}$ :  $t_{1-\alpha} = t_{95} = 1.80$ .
3. Так как  $0.42 < 1.80$ , т. е.  $t_b^* < t_{1-\alpha}$  нулевая гипотеза  $H_0: b = 0$  не отвергается – коэффициент  $b$  статистически незначим и для аппроксимации зависимости между  $X$  и  $Y$  можно рекомендовать уравнение  $\hat{y}_i = ax_i$ .

#### 5.4.4. F-критерий значимости регрессии

Довольно часто для проверки значимости линейной регрессии используется критерий, представляющий собой отношение среднего квадрата, обусловленного регрессией (с числом степеней свободы  $\nu = 1$ ), к среднему квадрату относительно регрессии (с числом степеней свободы  $\nu = n - 2$ ).

$$F_{y(x)}^* = \frac{\sum_{i=1}^n (\hat{y}_i - y)^2}{\sum_{i=1}^n (y_i - \hat{y}_i)^2 / (n-2)} = \frac{(n-1)S_y^2 r^2}{\sigma_{y(x)}^2}, \quad (5.41)$$

где  $\sigma_{y(x)}$  – стандартная ошибка уравнения линейной регрессии, определяемая по формуле (5.19) или (5.20);  $S_y$  – несмещенная оценка СКО для ряда  $Y$ ;  $r$  – коэффициент корреляции;  $n$  – длина рядов.

Доказано, что это отношение имеет распределение Фишера с числом степеней свободы  $\nu_1 = 1$  и  $\nu_2 = n - 2$ . Связь считается значимой, если

$$F_{y(x)}^* > F_{1-\alpha}, \quad (5.42)$$

где  $F_{1-\alpha}$  – теоретическое значение статистики Фишера (см. приложение 6) при уровне значимости  $\alpha$ .

Принято считать, что связь можно использовать для практических расчетов, если значение  $F_{y(x)}^*$  превосходит  $F_{1-\alpha}$  по крайней мере в 4 – 5 раз (при  $\alpha = 5 - 10\%$ ).

#### 5.4.5. Построение доверительного интервала 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  $MO = y_k$  и стандартным отклонением

$$\sigma_{\text{рег}} = \sigma_{y(x)} \sqrt{\frac{1}{n} + \frac{(x_k - \bar{x})^2}{\sum (x_i - \bar{x})^2}} = \sigma_{y(x)} \sqrt{\frac{1}{n} + \frac{(x_k - \bar{x})^2}{S_x^2(n-1)}}, \quad (5.43)$$

где  $\sigma_{y(x)}$  – стандартная ошибка уравнения линейной регрессии, определяемая по формуле (5.19) или (5.20);  $S_x$  – несмещенная оценка СКО для ряда  $X$ ;  $\bar{x}$  – среднее значение для ряда  $X$ ;  $n$  – длина рядов.

Как видно из формулы (5.43), величина  $\sigma_{\text{рег}}$  достигает минимума, когда  $x_k = \bar{x}$ , и возрастает по мере того, как мы “удаляем”  $x_k$  от  $\bar{x}$  в любом направлении. Другими словами, чем больше разность между  $x_k$  и средним значением, тем больше ошибка, с которой мы будем предсказывать  $y_{k,0}$  по данному  $x_k$ .

С помощью формулы (5.43) можно построить доверительные пределы для уравнения регрессии (рис. 5.7):

$$y_{k,0} = y_k \pm t_{1-\alpha} \sigma_{\text{рег}}, \quad (5.44)$$

где  $y_k = ax_k + b$ ;  $t_{1-\alpha}$  – квантиль распределения Стьюдента, соответствующий уровню значимости  $2\alpha$  при числе степеней свободы  $\nu = n - 2$  (в данном случае  $\nu$  – это число степеней свободы, на котором основана оценка  $\sigma_{y(x)}$ );

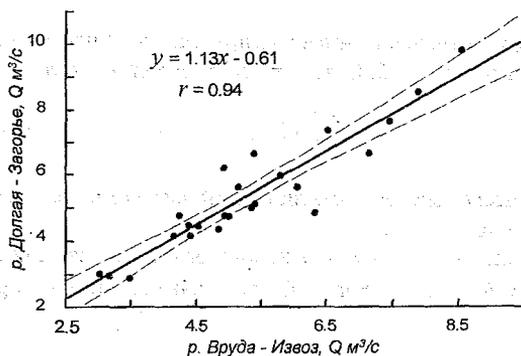


Рис. 5.7. График зависимости среднегодовых расходов воды на р. Долгой – дер. Загорье от среднегодовых расходов на р. Вруде – дер. Извоз. Штриховыми линиями обозначены 95%-ные доверительные пределы для уравнения регрессии.

### 5.5. Множественная линейная корреляция

При изучении многих гидрологических процессов требуется установить вид линейной зависимости между несколькими переменными. Для решения этой задачи привлекается аппарат множественной линейной корреляции. Сущность этого подхода состоит в распространении основных положений метода линейной корреляции двух переменных на случай зависимости интересующей нас переменной  $Y$  от произвольного числа аргументов. Основой для поиска такой зависимости служат материалы наблюдений за величиной  $Y$  и определяющими ее величинами  $X_1, X_2, \dots, X_m$ . Опираясь на данные указанных наблюдений, требуется найти параметры уравнения множественной линейной регрессии, которое согласно принципу наименьших квадратов, будет наилучшим образом описывать связь между зависимой переменной и предикторами. Уравнение линейной регрессии в этом случае будет иметь вид

$$(\bar{y} - \bar{y}) = a_1(x_1 - \bar{x}_1) + a_2(x_2 - \bar{x}_2) + \dots + a_m(x_m - \bar{x}_m), \quad (5.45)$$

где  $a_1, a_2, \dots, a_m$  – коэффициенты регрессии;  $\bar{y}, \bar{x}_1, \bar{x}_2, \dots, \bar{x}_m$  – средние значения соответственно независимой переменной и предикторов;  $m$  – число предикторов.

Для решения поставленной задачи нужно решить систему нормальных уравнений относительно коэффициентов  $a_1, a_2, \dots, a_m$ . Обычно эта задача решается на компьютере с использованием численных методов линейной алгебры. При небольшом числе предикторов можно получить аналитическое решение задачи методом Крамера. В этом случае для определения коэффициентов  $a_i$  необходимо рассчитать главный определитель системы  $D$

$$D = \begin{vmatrix} 1 & r_{0,1} & r_{0,2} & r_{0,3} & \dots & r_{0,j} & \dots & r_{0,m} \\ r_{1,0} & 1 & r_{1,2} & r_{1,3} & \dots & r_{1,j} & \dots & r_{1,m} \\ r_{2,0} & r_{2,1} & 1 & r_{2,3} & \dots & r_{2,j} & \dots & r_{2,m} \\ r_{3,0} & r_{3,1} & r_{3,2} & 1 & \dots & r_{3,j} & \dots & r_{3,m} \\ \vdots & \vdots \\ r_{i,0} & r_{i,1} & r_{i,2} & r_{i,3} & \dots & r_{i,j} & \dots & r_{i,m} \\ \vdots & \vdots \\ r_{m,0} & r_{m,1} & r_{m,2} & r_{m,3} & \dots & r_{m,j} & \dots & 1 \end{vmatrix}, \quad (5.46)$$

где  $r_{i,j}$  – парные коэффициенты корреляции между переменными. Например,  $r_{0,3}$  – это коэффициент корреляции между зависимой переменной и третьим предиктором, а  $r_{2,1}$  – коэффициент корреляции между вторым и первым предикторами. При этом  $r_{0,3} = r_{3,0}$ , а  $r_{2,1} = r_{1,2}$ . На главной диагонали этого определителя стоят единицы, так как  $r_{0,0} = r_{1,1} = r_{2,2} = \dots = r_{m,m} = 1$ .

Коэффициенты регрессии будут определяться выражением

$$a_j = -(-1)^j \frac{\sigma_y D_{0,j}}{\sigma_{x,j} D_{0,0}}, \quad (5.47)$$

где  $\sigma_y$  – среднеквадратическое отклонение независимой переменной;  $\sigma_{x,j}$  – среднеквадратическое отклонение  $j$ -го предиктора;  $D_{0,0}$  – определитель, получаемый из главного определителя системы (5.46) путем вычеркивания из него нулевой строки и нулевого столбца;  $D_{0,j}$  – определитель, получаемый из главного определителя путем вычеркивания из него нулевой строки и  $j$ -го столбца<sup>1</sup>.

Рассмотрим частный случай, когда число переменных равно трем (зависимая переменная и два предиктора). В этом случае

$$D = \begin{vmatrix} 1 & r_{0,1} & r_{0,2} \\ r_{1,0} & 1 & r_{1,2} \\ r_{2,0} & r_{2,1} & 1 \end{vmatrix}, \quad (4.48)$$

$$D_{0,0} = \begin{vmatrix} 1 & r_{1,2} \\ r_{2,1} & 1 \end{vmatrix} = 1 - r_{2,1} \cdot r_{1,2} = 1 - r_{1,2}^2, \quad (5.49)$$

$$D_{0,1} = \begin{vmatrix} r_{1,0} & r_{1,2} \\ r_{2,0} & 1 \end{vmatrix} = r_{1,0} - r_{2,0} \cdot r_{1,2}, \quad (5.50)$$

<sup>1</sup> В линейной алгебре определитель  $D_{ij}$  называется минором определителя  $D$ .

$$D_{0,2} = \begin{vmatrix} r_{1,0} & 1 \\ r_{2,0} & r_{2,1} \end{vmatrix} = r_{1,0} \cdot r_{2,1} - r_{2,0} \quad (5.51)$$

Подставляя выражения (5.49) – (5.51) в формулу (5.47), получаем

$$a_1 = + \frac{\sigma_y (r_{1,0} - r_{2,0} \cdot r_{1,2})}{\sigma_{x,1} (1 - r_{1,2}^2)}; \quad a_2 = - \frac{\sigma_y (r_{1,0} \cdot r_{2,1} - r_{2,0})}{\sigma_{x,2} (1 - r_{1,2}^2)}$$

При увеличении числа предикторов выражения для коэффициентов регрессии становятся все более громоздкими, поэтому метод Крамера обычно используют при числе предикторов  $m \leq 3$ .

В настоящее время разработано большое количество компьютерных программ, обеспечивающих быстрый и качественный расчет параметров множественной регрессии. Такая программа, в частности, входит в "пакет анализа" электронных таблиц Microsoft Excel.

### 5.5.1. Оценка точности уравнения множественной линейной регрессии

Так же как и в случае парной линейной корреляции, теснота связи при множественной линейной корреляции оценивается коэффициентом корреляции, который называют *общим*, или *сводным*, коэффициентом корреляции:

$$R = \sqrt{1 - \frac{D}{D_{0,0}}} \quad (5.52)$$

Сводный коэффициент корреляции можно также рассчитать как парный коэффициент корреляции между фактическими значениями независимой переменной и значениями независимой переменной, полученными по уравнению множественной линейной регрессии при тех же значениях аргументов.

В отличие от парного коэффициента корреляции, который может изменяться от  $-1$  до  $+1$ , коэффициент множественной корреляции всегда положителен и имеет пределы изменения от  $0$  до  $+1$ .

Стандартная ошибка коэффициента множественной корреляции определяется по формуле

$$\sigma_R = \frac{1 - R^2}{\sqrt{n - m - 1}}, \quad (5.53)$$

где  $n$  – длина анализируемых рядов;  $m$  – число предикторов.

Стандартную ошибку уравнения множественной линейной регрессии можно оценить по формуле

$$\sigma_{Y(X)} \approx \sigma_y^* \sqrt{1 - R^2}. \quad (5.54)$$

Стандартная ошибка  $j$ -го коэффициента регрессии определяется выражением

$$\sigma_{a_j} = \frac{\sigma_y^*}{\sigma_{x_j}} \sqrt{\frac{(1 - R^2) \Delta_{j,j}}{(n - m - 1) D_{0,0}}}, \quad (5.55)$$

где  $\Delta_{j,j}$  – минор определителя  $D_{0,0}$ .

В соответствии с действующими нормативными документами уравнение множественной линейной регрессии должно удовлетворять требованиям, аналогичным (5.28), т. е.

$$R \geq 0.7; \quad (R/\sigma_R) \geq 2; \quad (a_j/\sigma_{a_j}) \geq 2; \quad n \geq 10. \quad (5.56)$$

При этом следует обратить внимание, что третье условие считается выполненным, если оно выполняется для каждого коэффициента регрессии в отдельности. Что касается четвертого условия, то оно в данном случае является слишком слабым. Практика показывает, что при использовании одного предиктора длина рядов  $n$  должна быть не менее 10, при двух предикторах минимальная длина рядов должна составлять 25 – 30, при четырех 50 – 60, при пяти 100 – 120 и т. д. Только в этом случае можно получить более или менее надежные оценки параметров уравнения регрессии. Поскольку продолжительность гидрологических рядов относительно невелика, в практике гидрологических расчетов редко используются регрессионные модели, содержащие более четырех предикторов.

### 5.5.2. Выбор "наилучшего" уравнения регрессии

В расчетах по уравнению множественной регрессии необходимо использовать только те предикторы, которые дают объем информации, превышающий ошибку расчета. Неэффективные предикторы исключаются. В противном случае привлечение дополнительных данных может привести не к уменьшению ошибки, а к ее увеличению. Назовем несколько причин такой неэффективности.

- Связь между зависимой переменной и предиктором отсутствует или выражена очень слабо.
- Связь между зависимой переменной и предиктором существует, но в силу небольшой продолжительности рядов привлечение дополнительного предиктора приводит к тому, что коэффициенты регрессии становятся незначи-

мыми. В частности, перестает выполняться требование  $(a_j/\sigma_{a_j}) \geq 2$  для одного или нескольких коэффициентов регрессии.

- Наличие дублирующих предикторов. Эта ситуация возникает, когда несколько предикторов имеют высокий коэффициент взаимной корреляции. В этом случае определитель  $D_{0,0}$  близок к нулю и, как следует из формулы (5.47), система нормальных уравнений будет неопределенной либо будет иметь неустойчивое решение.

Однако даже если неэффективные предикторы отсеяны, может оказаться, что из имеющегося набора предикторов можно получить несколько уравнений, удовлетворяющих набору требований (5.56). В этом случае необходимо выбрать "наилучшее" уравнение регрессии. Для решения этой задачи разработано несколько методов.

**Метод всех возможных регрессий.** При использовании этого метода рассчитываются все возможные уравнения регрессии, после чего исключаются те, которые не удовлетворяют системе ограничений (5.56). Из оставшихся уравнений наилучшим будет то, которое имеет самый высокий сводный коэффициент корреляции.

**Метод исключения.** Метод исключения более экономичен, чем метод всех регрессий, поскольку в нем делается попытка использовать только наилучшие регрессионные уравнения. Основные шаги этого метода сводятся к следующему.

1. Рассчитывается регрессионное уравнение, включающее все переменные.
2. Вычисляется значение частного  $F$ -критерия<sup>1</sup> для каждого предиктора в предположении, как будто это – последняя переменная, введенная в регрессионное уравнение.
3. Наименьшее значение частного  $F$ -критерия, обозначаемое, скажем, как  $F_L$ , сравнивается с заранее выбранным критическим значением, например  $F_0$ , далее возможны два варианта:
  - если  $F_L < F_0$ , то предиктор  $X_L$  исключается и производится пересчет уравнения регрессии с учетом оставшихся переменных; затем переходят к следующему шагу;
  - если  $F_L > F_0$ , то уравнение регрессии оставляют без изменений и процедура поиска "наилучшего" уравнения завершается.

**Шаговый регрессионный метод (метод включения).** На первом этапе выбирается предиктор  $X$ , наиболее сильно скоррелированный с  $Y$  (предположим, что это  $X_1$ ), и находят линейное регрессионное уравнение первого порядка  $Y = f(X_1)$ . Затем проверяется, значима ли эта переменная. Если это не так, то мы должны согласиться с выводом, что наилучшая модель выражается уравнением  $Y = \bar{Y}$ . В противном случае ищем второй предиктор для включения в модель. Обычно в модель включают тот предиктор, который имеет самый высокий коэффициент частной корреляции с зависимой переменной. Частный коэффициент корреляции характеризует индивидуальный вклад предиктора в об-

<sup>1</sup> Частный  $F$ -критерий связан с проверкой гипотезы  $H_0: a_j = 0$  против альтернативной  $H_1: a_j \neq 0$  для любого отдельного коэффициента регрессии.

щую регрессию в дополнение к тем вкладам, которые дают другие, включенные в уравнение предикторы. Так, если в уравнение множественной регрессии включается дублирующий предиктор, то его частный коэффициент корреляции с зависимой переменной будет низким. Частный коэффициент корреляции между зависимой переменной и  $j$ -м предиктором определяется по формуле

$$\eta_{0,j} = -\frac{A_{0,j}}{\sqrt{A_{0,0}A_{j,j}}}, \quad (5.57)$$

где  $A_{i,j}$  – алгебраическое дополнение главного определителя (5.46) системы

$$A_{i,j} = (-1)^{i+j} D_{i,j}, \quad (5.58)$$

$D_{i,j}$  – минор главного определителя системы, т. е. определитель получаемый из определителя  $D$  путем вычеркивания из него  $i$ -й строки и  $j$ -го столбца.

После включения в модель второго предиктора находится второе регрессионное уравнение  $Y = f(X_1, X_2)$ . Полное уравнение проверяется на значимость. Отмечается улучшение сводного коэффициента корреляции и исследуются частные  $F$ -критерии для обеих переменных. Наименьшее из этих частных  $F$ -критериев ( $F_L$ ) сравнивается с табличным значением  $F$ -критерия ( $F_0$ ) при заданном уровне значимости. Если  $F_L < F_0$ , то соответствующий предиктор исключается из уравнения. Такая проверка "наименее полезного предиктора" проводится на каждом шаге. Может оказаться, что предиктор, который на предыдущем шаге был наилучшим кандидатом на включение в уравнение, на более позднем шаге оказывается ненужным. Это может быть вызвано теми связями, которые существуют между этой и другими переменными, содержащимися теперь в уравнении.

Процесс прекращается, если никакую из переменных, содержащихся в текущем уравнении, не удастся исключить из него, а ближайший наилучший предиктор-претендент не в состоянии занять место в уравнении.

### 5.5.3. Ложная корреляция

На практике при построении регрессионных моделей в число предикторов иногда включают не только переменные  $Y, X_1, X_2, \dots, X_m$ , но и их комбинации, например  $(X_1 \cdot X_2); (X_1/X_2); (1/X_1)$  и т. д. Во многих случаях такой подход оправдан и позволяет повысить сводный коэффициент корреляции. Однако если в число предикторов включить комбинацию зависимой ( $Y$ ) переменной с одной из независимых переменных ( $X_i$ ) может возникнуть эффект *ложной корреляции*. Говоря другими словами, мы можем получить высокий коэффициент корреляции, хотя на самом деле никакой связи между  $Y$  и переменной  $X_i$  не существует.

Действительно, если дисперсия  $Y$  в несколько раз превышает дисперсию  $X$ , то при построении зависимостей типа  $Y = f(YX), Y = f(Y/X), Y = f(Y+X)$  мы по

сути коррелируем зависимую переменную с самой собой, а  $X$  выполняет роль слабого шума.

В частности, если коэффициент вариации  $Y$  превышает коэффициент вариации  $X$ , то даже при полном отсутствии связи между  $Y$  и  $X$ , коэффициент корреляции для зависимости  $Y = f(X)$  будет не менее 0.7 (см. приложение 8).

Не учет эффектов ложной корреляции при построении прогностических зависимостей может привести к серьезным ошибкам.

## Список литературы

1. *Вентцель Е. С., Овчаров Л. А.* Теория вероятностей и ее инженерные приложения. – М.: Наука, 1988. – 480 с.
2. *Международное руководство по методам расчета основных гидрологических характеристик.* – Л.: Гидрометеоздат, 1984. – 247 с.
3. *Митропольский А. К.* Техника статистических вычислений. – М.: Наука, 1971. – 576 с.
4. *Пособие по определению расчетных гидрологических характеристик.* – Л.: Гидрометеоздат, 1984. – 447 с.
5. *Расчеты паводочного стока. Методика расчета на основе мирового опыта/* Под ред. А. А. Соколова, С. Е. Рантца и др. – Л.: Гидрометеоздат, 1978. – 304 с.
6. *Рождественский А. В., Ежов А. В., Сахарюк А. В.* Оценка точности гидрологических расчетов. – Л.: Гидрометеоздат, 1990. – 273 с.
7. *Рождественский А. В.* Оценка точности кривых распределения гидрологических характеристик. – Л.: Гидрометеоздат, 1977. – 268 с.
8. *Рождественский А. В., Чеботарев А. И.* Статистические методы в гидрологии. – Л.: Гидрометеоздат, 1974. – 424 с.
9. *Сванидзе Г. Г.* Математическое моделирование гидрологических рядов. – Л.: Гидрометеоздат, 1977. – 296 с.
10. *Справочник по прикладной статистике. Т. 1, 2.* Под ред. Э. Ллойда, У. Линдермана. – М.: Финансы и статистика, 1990.
11. *Шелутко В. А.* Численные методы в гидрологии. – Л.: Гидрометеоздат, 1991. – 238 с.

Нормированные ординаты кривой обеспеченностей Пирсона III типа  
(биномиального непрерывного распределения);

$$\hat{t}_p = f(C_s, p), \text{ где } \hat{t}_p = (\hat{x}_p - \bar{x}) / \sigma = (k_p - 1) / C_s.$$

$C_s$	Обеспеченность, $p$ %									$S$
	0.01	0.1	1.0	5.0	10	20	25	30	50	
-1.8	1.11	1.11	1.09	1.02	0.94	0.80	0.72	0.64	0.28	-0.51
-1.6	1.26	1.24	1.20	1.10	0.99	0.81	0.73	0.64	0.25	-0.45
-1.4	1.41	1.39	1.32	1.17	1.04	0.83	0.73	0.64	0.22	-0.39
-1.2	1.68	1.58	1.45	1.24	1.08	0.84	0.74	0.63	0.19	-0.34
-1.0	1.92	1.79	1.59	1.32	1.13	0.85	0.73	0.62	0.16	-0.27
-0.8	2.23	2.02	1.74	1.38	1.17	0.86	0.73	0.60	0.13	-0.22
-0.6	2.57	2.27	1.88	1.45	1.20	0.85	0.72	0.59	0.10	-0.17
-0.4	2.98	2.54	2.03	1.52	1.23	0.85	0.71	0.57	0.07	-0.11
-0.2	3.37	2.81	2.18	1.58	1.26	0.85	0.69	0.55	0.03	-0.05
0.0	3.72	3.09	2.33	1.64	1.28	0.84	0.67	0.52	0.00	0.00
0.2	4.16	3.38	2.47	1.70	1.30	0.83	0.65	0.50	-0.03	0.06
0.4	4.61	3.66	2.61	1.75	1.32	0.82	0.63	0.47	-0.07	0.11
0.6	5.05	3.96	2.75	1.80	1.33	0.80	0.61	0.44	-0.10	0.17
0.8	5.50	4.24	2.89	1.84	1.34	0.78	0.58	0.41	-0.13	0.22
1.0	5.96	4.53	3.02	1.88	1.34	0.76	0.55	0.38	-0.16	0.28
1.2	6.41	4.81	3.15	1.92	1.34	0.73	0.52	0.35	-0.19	0.34
1.4	6.87	5.09	3.27	1.95	1.34	0.71	0.49	0.31	-0.22	0.39
1.6	7.31	5.37	3.39	1.97	1.33	0.68	0.46	0.28	-0.25	0.45
1.8	7.76	5.64	3.50	1.99	1.32	0.64	0.42	0.24	-0.28	0.51
2.0	8.21	5.91	3.60	2.00	1.30	0.61	0.39	0.20	-0.31	0.57
2.2	8.63	6.14	3.68	2.02	1.27	0.57	0.35	0.16	-0.33	0.62
2.4	9.00	6.37	3.78	2.00	1.25	0.52	0.29	0.12	-0.35	0.67
2.6	9.39	6.54	3.86	2.00	1.21	0.48	0.25	0.085	-0.37	0.72
2.8	9.77	6.86	3.96	2.00	1.18	0.44	0.22	0.057	-0.39	0.76
3.0	10.16	7.10	4.05	1.97	1.13	0.39	0.19	0.027	-0.40	0.80

Приложение 1(продолжение)

Нормированные ординаты кривой обеспеченностей Пирсона III типа  
(биномиального непрерывного распределения);

$$i_p = f(C_s, p), \text{ где } i_p = (\hat{x}_p - \bar{x}) / \sigma = (k_p - 1) / C_s.$$

$C_s$	Обеспеченность, $p$ %								$S$
	70	75	80	90	95	97	99	99.9	
-1.8	-0.24	-0.42	-0.64	-1.32	-1.99	-2.46	-3.50	-5.64	-0.51
-1.6	-0.28	-0.46	-0.68	-1.33	-1.97	-2.42	-3.39	-5.37	-0.45
-1.4	-0.31	-0.49	-0.71	-1.34	-1.95	-2.37	-3.27	-5.09	-0.39
-1.2	-0.35	-0.52	-0.73	-1.34	-1.92	-2.31	-3.15	-4.81	-0.34
-1.0	-0.38	-0.55	-0.76	-1.34	-1.88	-2.25	-3.02	-4.53	-0.27
-0.8	-0.41	-0.58	-0.79	-1.34	-1.84	-2.18	-2.89	-4.24	-0.22
-0.6	-0.44	-0.61	-0.80	-1.33	-1.80	-2.12	-2.75	-3.96	-0.17
-0.4	-0.47	-0.63	-0.82	-1.32	-1.75	-2.04	-2.61	-3.66	-0.11
-0.2	-0.50	-0.65	-0.83	-1.30	-1.70	-1.96	-2.47	-3.38	-0.05
0.0	-0.52	-0.67	-0.84	-1.28	-1.64	-1.88	-2.33	-3.09	0.00
0.2	-0.55	-0.69	-0.85	-1.26	-1.58	-1.79	-2.21	-2.81	0.06
0.4	-0.57	-0.71	-0.85	-1.23	-1.52	-1.70	-2.03	-2.54	0.11
0.6	-0.59	-0.72	-0.85	-1.20	-1.45	-1.61	-1.88	-2.27	0.17
0.8	-0.60	-0.73	-0.86	-1.17	-1.38	-1.52	-1.74	-2.02	0.22
1.0	-0.62	-0.73	-0.85	-1.13	-1.32	-1.42	-1.59	-1.79	0.28
1.2	-0.63	-0.74	-0.84	-1.08	-1.24	-1.33	-1.45	-1.58	0.34
1.4	-0.64	-0.73	-0.83	-1.04	-1.17	-1.23	-1.32	-1.39	0.39
1.6	-0.64	-0.73	-0.81	-0.99	-1.10	-1.14	-1.20	-1.24	0.45
1.8	-0.64	-0.72	-0.80	-0.94	-1.02	-1.06	-1.09	-1.11	0.51
2.0	-0.64	-0.71	-0.78	-0.90	-0.95	-0.97	-0.99	-1.00	0.57
2.2	-0.64	-0.69	-0.75	-0.84	-0.88	-0.90	-0.90	-0.91	0.62
2.4	-0.62	-0.67	-0.72	-0.79	-0.82	-0.83	-0.83	-0.83	0.67
2.6	-0.61	-0.66	-0.70	-0.75	-0.76	-0.77	-0.77	-0.77	0.72
2.8	-0.60	-0.64	-0.67	-0.70	-0.71	-0.71	-0.72	-0.72	0.76
3.0	-0.59	-0.62	-0.64	-0.66	-0.66	-0.67	-0.67	-0.67	0.80

Приложение 2

Ординаты кривых обеспеченностей Крицкого – Менкеля (трехпараметрического гамма-распределения) в модульных коэффициентах  $k_p = f(C_s/C_v, C_v, p)$

$$C_s/C_v = 0.5$$

p %	Коэффициент вариации, $C_v$							
	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7
0.01	1.00	1.38	1.76	2.13	2.47	2.74	2.91	2.89
0.1	1.00	1.31	1.63	1.95	2.25	2.50	2.69	2.74
0.3	1.00	1.28	1.56	1.85	2.12	2.36	2.55	2.64
0.5	1.00	1.26	1.53	1.79	2.05	2.28	2.48	2.59
1.0	1.00	1.24	1.48	1.72	1.95	2.17	2.37	2.50
3	1.00	1.19	1.38	1.58	1.78	1.97	2.16	2.33
5	1.00	1.17	1.33	1.51	1.68	1.86	2.03	2.22
10	1.00	1.13	1.26	1.39	1.53	1.67	1.83	2.01
20	1.00	1.08	1.17	1.25	1.35	1.44	1.56	1.70
25	1.00	1.07	1.13	1.20	1.27	1.35	1.45	1.56
30	1.00	1.05	1.10	1.16	1.21	1.27	1.34	1.42
40	1.00	1.02	1.05	1.07	1.10	1.12	1.15	1.16
50	1.00	0.999	0.997	0.993	0.988	0.980	0.962	0.920
60	1.00	0.974	0.946	0.915	0.881	0.839	0.780	0.690
70	1.00	0.947	0.882	0.834	0.769	0.693	0.596	0.476
75	1.00	0.932	0.862	0.789	0.709	0.615	0.503	0.376
80	1.00	0.915	0.829	0.740	0.643	0.533	0.409	0.282
90	1.00	0.872	0.744	0.615	0.480	0.343	0.215	0.115
95	1.00	0.837	0.676	0.517	0.362	0.221	0.113	0.047
97	1.00	0.814	0.633	0.458	0.295	0.160	0.070	0.024
99	1.00	0.772	0.554	0.354	0.189	0.080	0.025	0.006
99.5	1.00	0.748	0.511	0.302	0.144	0.051	0.013	0.002
99.7	1.00	0.732	0.482	0.269	0.117	0.037	0.008	0.001
99.9	1.00	0.700	0.428	0.210	0.076	0.019	0.003	0.000

Приложение 2 (продолжение)

$$C_s/C_v = 1.0$$

p %	Коэффициент вариации, C <sub>v</sub>									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0.01	1.38	1.81	2.26	2.70	3.15	3.57	3.95	4.31	4.64	4.92
0.1	1.32	1.67	2.03	2.40	2.77	3.13	3.48	3.82	4.13	4.42
0.3	1.28	1.59	1.91	2.23	2.56	2.89	3.21	3.53	3.84	4.14
0.5	1.27	1.55	1.84	2.15	2.46	2.77	3.08	3.38	3.69	3.99
1.0	1.24	1.49	1.76	2.03	2.30	2.59	2.88	3.16	3.46	3.75
3	1.19	1.39	1.60	1.82	2.04	2.27	2.50	2.75	3.01	3.29
5	1.17	1.34	1.52	1.70	1.90	2.10	2.30	2.53	2.76	3.02
10	1.13	1.26	1.40	1.54	1.68	1.83	1.99	2.16	2.35	2.55
20	1.08	1.17	1.25	1.34	1.42	1.51	1.60	1.70	1.80	1.90
25	1.07	1.13	1.20	1.26	1.33	1.39	1.46	1.52	1.59	1.64
30	1.05	1.10	1.15	1.20	1.24	1.29	1.33	1.37	1.39	1.40
40	1.02	1.04	1.06	1.08	1.09	1.10	1.10	1.08	1.05	0.995
50	0.998	0.993	0.985	0.972	0.954	0.928	0.891	0.836	0.76	0.665
60	0.973	0.943	0.909	0.870	0.824	0.768	0.698	0.613	0.512	0.406
70	0.946	0.890	0.830	0.764	0.692	0.609	0.515	0.413	0.309	0.215
75	0.932	0.861	0.787	0.708	0.622	0.528	0.426	0.321	0.224	0.144
80	0.915	0.829	0.740	0.648	0.549	0.445	0.338	0.237	0.151	0.088
90	0.873	0.748	0.623	0.500	0.378	0.264	0.165	0.092	0.045	0.019
95	0.838	0.683	0.533	0.392	0.263	0.157	0.081	0.036	0.013	0.004
97	0.816	0.642	0.478	0.329	0.202	0.107	0.048	0.018	0.005	0.001
99	0.775	0.568	0.383	0.229	0.115	0.047	0.015	0.004	0.001	0.000
99.5	0.752	0.528	0.335	0.182	0.081	0.028	0.008	0.002	0.000	0.000
99.7	0.737	0.502	0.303	0.154	0.062	0.019	0.004	0.001	0.000	0.000
99.9	0.707	0.451	0.247	0.108	0.036	0.008	0.001	0.000	0.000	0.000

p %	Коэффициент вариации, C <sub>v</sub>									
	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
1.0	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
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	0.00
70	0.141	0.085	0.06	0.045	0.035	0.025	0.015	0.010	0.000	0.00
75	0.086	0.046	0.025	0.020	0.010	0.005	0.000	0.000	0.000	0.00
80	0.047	0.023	0.015	0.005	0.000	0.000	0.000	0.000	0.000	0.00
90	0.007	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.00
95	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.00
97	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.00
99	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.00
99.5	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.00
99.7	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.00
99.9	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.00

Приложение 2 (продолжение)

$$C_s/C_v = 2$$

p %	Коэффициент вариации, $C_v$									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0.01	1.42	1.92	2.52	3.20	3.98	4.85	5.81	6.85	7.98	9.21
0.1	1.34	1.73	2.19	2.70	3.27	3.87	4.56	5.30	6.08	6.91
0.3	1.30	1.64	2.02	2.45	2.91	3.42	3.96	4.55	5.16	5.81
0.5	1.28	1.59	1.94	2.32	2.74	3.20	3.68	4.19	4.74	5.30
1.0	1.25	1.52	1.82	2.16	2.51	2.89	3.29	3.71	4.15	4.60
3	1.20	1.41	1.64	1.87	2.13	2.39	2.66	2.94	3.21	3.51
5	1.17	1.35	1.54	1.74	1.94	2.15	2.36	2.57	2.78	3.00
10	1.13	1.26	1.40	1.54	1.67	1.80	1.94	2.06	2.19	2.30
20	1.08	1.16	1.24	1.31	1.38	1.44	1.50	1.54	1.58	1.61
25	1.06	1.13	1.18	1.23	1.28	1.31	1.34	1.37	1.38	1.39
30	1.05	1.09	1.13	1.16	1.19	1.21	1.22	1.22	1.22	1.20
40	1.02	1.04	1.05	1.05	1.04	1.03	1.01	0.984	0.955	0.916
50	0.997	0.986	0.970	0.948	0.918	0.886	0.846	0.800	0.748	0.693
60	0.972	0.938	0.898	0.852	0.803	0.748	0.692	0.632	0.568	0.511
70	0.945	0.886	0.823	0.760	0.691	0.622	0.552	0.488	0.424	0.357
75	0.931	0.858	0.784	0.708	0.634	0.556	0.489	0.416	0.352	0.288
80	0.915	0.830	0.745	0.656	0.574	0.496	0.419	0.352	0.280	0.223
90	0.873	0.754	0.640	0.532	0.436	0.352	0.272	0.208	0.154	0.105
95	0.842	0.696	0.565	0.448	0.342	0.256	0.181	0.120	0.082	0.051
97	0.821	0.660	0.517	0.392	0.288	0.202	0.139	0.088	0.046	0.030
99	0.782	0.594	0.436	0.304	0.206	0.130	0.076	0.040	0.019	0.010
99.5	0.761	0.560	0.394	0.269	0.166	0.099	0.054	0.027	0.012	0.005
99.7	0.748	0.537	0.374	0.240	0.144	0.082	0.042	0.019	0.008	0.003
99.9	0.719	0.492	0.319	0.192	0.107	0.052	0.027	0.008	0.004	0.001

p %	Коэффициент вариации, $C_v$									
	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2.0
0.01	10.5	11.8	13.2	14.7	16.4	18.2	20.2	22.2	24.4	26.6
0.1	7.75	8.65	9.60	10.6	11.6	12.5	13.5	14.6	15.8	17.0
0.3	6.47	7.10	7.98	8.70	9.50	10.5	11.0	11.9	12.7	13.6
0.5	5.90	6.50	7.13	7.80	8.42	9.0	9.5	10.1	10.8	11.4
1.0	5.05	5.53	6.02	6.55	7.08	7.5	8.0	8.6	9.2	9.8
3	3.80	4.12	4.42	4.71	4.98	5.2	5.5	5.8	6.2	6.5
5	3.22	3.40	3.60	3.80	3.96	4.0	4.3	4.5	4.7	5.0
10	2.40	2.50	2.57	2.64	2.70	2.7	2.6	2.6	2.6	2.6
20	1.62	1.63	1.62	1.61	1.59	1.6	1.6	1.56	1.5	1.5
25	1.39	1.35	1.33	1.31	1.28	1.26	1.24	1.22	1.2	1.18
30	1.18	1.14	1.11	1.08	1.04	1.015	0.98	0.95	0.92	0.89
40	0.870	0.830	0.770	0.725	0.670	0.625	0.58	0.53	0.48	0.44
50	0.640	0.580	0.520	0.460	0.405	0.355	0.310	0.265	0.23	0.20
60	0.450	0.390	0.334	0.283	0.234	0.190	0.160	0.130	0.105	0.085
70	0.300	0.250	0.203	0.155	0.120	0.090	0.070	0.060	0.05	0.045
75	0.241	0.193	0.146	0.106	0.077	0.060	0.050	0.040	0.03	0.025
80	0.175	0.130	0.094	0.065	0.046	0.035	0.027	0.020	0.015	0.01
90	0.074	0.049	0.030	0.016	0.009	0.005	0.004	0.003	0.002	0.001
95	0.030	0.016	0.009	0.004	0.002	0.001	0.000	0.000	0.000	0.000
97	0.016	0.008	0.004	0.002	0.001	0.000	0.000	0.000	0.000	0.000
99	0.005	0.002	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000
99.5	0.002	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
99.7	0.001	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 2 (продолжение)

$$C_v/C_v = 3$$

p %	Коэффициент вариации, $C_v$									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0.01	1.46	2.05	2.83	3.80	4.94	6.26	7.70	9.30	11.0	12.8
0.1	1.36	1.81	2.35	3.01	3.74	4.56	5.44	6.38	7.37	8.41
0.3	1.31	1.69	2.12	2.65	3.21	3.82	4.48	5.17	5.88	6.61
0.5	1.28	1.63	2.03	2.48	2.97	3.50	4.06	4.64	5.24	5.84
1.0	1.25	1.55	1.90	2.26	2.66	3.07	3.50	3.96	4.41	4.87
3	1.20	1.42	1.66	1.91	2.17	2.43	2.69	2.95	3.21	3.47
5	1.17	1.36	1.55	1.75	1.95	2.14	2.34	2.52	2.70	2.88
10	1.13	1.26	1.40	1.52	1.65	1.76	1.87	1.97	2.06	2.15
20	1.08	1.16	1.23	1.29	1.34	1.38	1.42	1.45	1.47	1.49
25	1.07	1.12	1.17	1.21	1.24	1.26	1.28	1.28	1.29	1.29
30	1.05	1.09	1.12	1.14	1.15	1.16	1.16	1.15	1.14	1.13
40	1.02	1.03	1.03	1.03	1.01	0.995	0.972	0.946	0.915	0.883
50	0.997	0.981	0.959	0.930	0.898	0.862	0.823	0.783	0.741	0.699
60	0.972	0.933	0.890	0.843	0.794	0.745	0.695	0.646	0.597	0.549
70	0.945	0.884	0.822	0.758	0.696	0.636	0.578	0.523	0.471	0.422
75	0.931	0.858	0.786	0.715	0.647	0.583	0.522	0.465	0.412	0.363
80	0.915	0.830	0.748	0.669	0.596	0.528	0.465	0.407	0.354	0.306
90	0.876	0.761	0.656	0.563	0.479	0.406	0.341	0.284	0.235	0.193
95	0.844	0.708	0.588	0.487	0.400	0.326	0.263	0.210	0.166	0.129
97	0.825	0.675	0.548	0.443	0.355	0.282	0.221	0.171	0.131	0.099
99	0.786	0.618	0.484	0.369	0.283	0.213	0.158	0.116	0.083	0.058
99.5	0.769	0.588	0.446	0.334	0.249	0.182	0.131	0.092	0.064	0.043
99.7	0.756	0.568	0.422	0.312	0.228	0.163	0.114	0.079	0.053	0.034
99.9	0.732	0.531	0.381	0.273	0.192	0.131	0.088	0.057	0.036	0.022

p %	Коэффициент вариации, $C_v$									
	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2.0
0.01	14.8	16.8	19.0	21.2	23.5	25.9	28.4	31.0	33.7	36.5
0.1	9.49	10.6	11.8	13.0	14.2	15.4	16.7	18.0	19.4	20.8
0.3	7.37	8.15	8.94	9.75	10.6	11.4	12.3	13.1	14.0	14.8
0.5	6.47	7.10	7.75	8.41	9.07	9.74	10.4	11.1	11.8	12.4
1.0	5.33	5.79	6.26	6.74	7.21	7.68	8.14	8.61	9.07	9.53
3	3.73	3.98	4.20	4.44	4.67	4.89	5.10	5.31	5.51	5.70
5	3.05	3.22	3.37	3.52	3.66	3.80	3.92	4.04	4.15	4.26
10	2.23	2.30	2.36	2.42	2.47	2.51	2.55	2.58	2.60	2.62
20	1.50	1.50	1.50	1.49	1.48	1.46	1.45	1.42	1.40	1.37
25	1.28	1.27	1.25	1.23	1.20	1.18	1.15	1.12	1.08	1.05
30	1.11	1.08	1.06	1.03	0.997	0.964	0.929	0.892	0.855	0.818
40	0.848	0.812	0.775	0.736	0.697	0.659	0.620	0.581	0.544	0.507
50	0.656	0.614	0.572	0.531	0.491	0.452	0.415	0.379	0.345	0.313
60	0.503	0.459	0.417	0.377	0.339	0.304	0.271	0.240	0.212	0.186
70	0.375	0.333	0.293	0.257	0.224	0.194	0.166	0.142	0.121	0.102
75	0.318	0.277	0.239	0.206	0.176	0.149	0.125	0.105	0.087	0.071
80	0.263	0.224	0.190	0.160	0.133	0.110	0.090	0.073	0.059	0.047
90	0.156	0.126	0.100	0.078	0.061	0.047	0.035	0.026	0.019	0.014
95	0.100	0.076	0.057	0.042	0.030	0.022	0.015	0.010	0.007	0.004
97	0.073	0.054	0.038	0.027	0.018	0.012	0.008	0.005	0.003	0.002
99	0.040	0.027	0.017	0.011	0.007	0.004	0.002	0.001	0.001	0.000
99.5	0.028	0.018	0.011	0.006	0.004	0.002	0.001	0.000	0.000	0.000
99.7	0.022	0.014	0.008	0.004	0.002	0.001	0.000	0.000	0.000	0.000
99.9	0.013	0.007	0.004	0.002	0.001	0.000	0.000	0.000	0.000	0.000

Приложение 2 (продолжение)

$$C_s/C_v = 4$$

p %	Коэффициент вариации, C <sub>v</sub>									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0.01	1.50	2.18	3.17	4.43	5.91	7.58	9.41	11.4	13.4	15.5
0.1	1.38	1.88	2.53	3.29	4.15	5.07	6.05	7.08	8.15	9.26
0.3	1.32	1.74	2.24	2.82	3.44	4.09	4.79	5.50	6.22	6.96
0.5	1.29	1.67	2.12	2.61	3.13	3.68	4.26	4.85	5.43	6.03
1.0	1.25	1.58	1.94	2.31	2.75	3.17	3.59	4.03	4.47	4.91
3	1.20	1.44	1.68	1.93	2.18	2.43	2.68	2.92	3.16	3.39
5	1.17	1.36	1.56	1.75	1.94	2.12	2.29	2.46	2.62	2.78
10	1.13	1.26	1.39	1.51	1.62	1.72	1.81	1.90	1.98	2.05
20	1.08	1.15	1.22	1.27	1.31	1.34	1.37	1.40	1.41	1.42
25	1.07	1.12	1.16	1.19	1.21	1.23	1.24	1.24	1.24	1.24
30	1.05	1.08	1.11	1.12	1.13	1.13	1.13	1.12	1.11	1.10
40	1.02	1.02	1.02	1.01	0.996	0.976	0.954	0.929	0.902	0.873
50	0.997	0.976	0.950	0.920	0.888	0.853	0.818	0.781	0.744	0.707
60	0.972	0.929	0.885	0.839	0.793	0.747	0.702	0.658	0.614	0.572
70	0.945	0.883	0.821	0.761	0.704	0.649	0.597	0.548	0.501	0.457
75	0.931	0.858	0.788	0.722	0.660	0.601	0.546	0.495	0.448	0.403
80	0.915	0.832	0.754	0.681	0.614	0.553	0.496	0.443	0.395	0.351
90	0.877	0.767	0.671	0.586	0.511	0.444	0.384	0.331	0.284	0.243
95	0.846	0.719	0.611	0.519	0.440	0.372	0.312	0.261	0.217	0.180
97	0.829	0.690	0.576	0.481	0.400	0.332	0.274	0.224	0.182	0.147
99	0.790	0.638	0.516	0.417	0.336	0.269	0.214	0.168	0.132	0.102
99.5	0.776	0.612	0.485	0.386	0.305	0.239	0.186	0.144	0.110	0.083
99.7	0.762	0.594	0.466	0.366	0.286	0.221	0.170	0.129	0.097	0.072
99.9	0.742	0.561	0.430	0.331	0.252	0.189	0.141	0.104	0.075	0.054

p %	Коэффициент вариации, C <sub>v</sub>									
	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2.0
0.01	17.9	20.3	22.8	25.4	28.0	30.8	33.6	36.5	39.4	42.4
0.1	10.4	11.6	12.8	14.0	15.3	16.6	17.9	19.2	20.6	21.9
0.3	7.73	8.53	9.31	10.1	10.9	11.7	12.5	13.3	14.2	15.0
0.5	6.65	7.29	7.91	8.53	9.16	9.79	10.4	11.0	11.7	12.3
1.0	5.34	5.79	6.22	6.66	7.09	7.52	7.95	8.73	8.78	9.19
3	3.62	3.83	4.04	4.25	4.45	4.64	4.83	5.01	5.18	5.34
5	2.93	3.07	3.21	3.34	3.46	3.57	3.68	3.78	3.87	3.96
10	2.12	2.18	2.24	2.28	2.32	2.36	2.39	2.42	2.44	2.45
20	1.43	1.44	1.43	1.43	1.42	1.41	1.39	1.38	1.36	1.33
25	1.23	1.22	1.21	1.19	1.17	1.15	1.13	1.10	1.08	1.05
30	1.08	1.06	1.04	1.01	0.985	0.958	0.929	0.90	0.871	0.841
40	0.843	0.812	0.781	0.748	0.716	0.684	0.652	0.62	0.588	0.558
50	0.670	0.634	0.598	0.562	0.529	0.495	0.464	0.433	0.403	0.375
60	0.532	0.494	0.457	0.421	0.388	0.356	0.327	0.299	0.273	0.249
70	0.416	0.377	0.341	0.308	0.277	0.248	0.223	0.199	0.177	0.157
75	0.362	0.325	0.290	0.258	0.230	0.203	0.179	0.158	0.139	0.121
80	0.311	0.274	0.242	0.212	0.185	0.162	0.140	0.122	0.105	0.090
90	0.207	0.176	0.148	0.125	0.104	0.087	0.072	0.060	0.049	0.040
95	0.148	0.121	0.098	0.080	0.064	0.051	0.041	0.032	0.025	0.019
97	0.119	0.095	0.075	0.059	0.046	0.036	0.028	0.021	0.016	0.012
99	0.078	0.060	0.045	0.034	0.025	0.018	0.013	0.009	0.006	0.004
99.5	0.062	0.046	0.034	0.024	0.017	0.012	0.008	0.006	0.004	0.003
99.7	0.053	0.038	0.027	0.019	0.013	0.009	0.006	0.004	0.003	0.002
99.9	0.038	0.026	0.018	0.012	0.008	0.005	0.003	0.002	0.001	0.001

Приложение 2 (продолжение)

$C_s/C_v = 5$

$p\%$	Коэффициент вариации, $C_v$								
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.01	1.54	2.34	3.43	4.91	6.65	8.70	10.70	12.71	15.05
0.1	1.40	1.95	2.66	3.51	4.44	5.40	6.43	7.54	8.64
0.3	1.34	1.78	2.31	2.92	3.52	4.22	4.91	5.69	6.41
0.5	1.31	1.70	2.16	2.69	3.21	3.77	4.34	4.94	5.52
1.0	1.27	1.61	1.98	2.38	2.79	3.21	3.65	4.06	4.50
3	1.20	1.44	1.67	1.93	2.17	2.42	2.62	2.88	3.10
5	1.17	1.36	1.55	1.74	1.90	2.08	2.22	2.41	2.54
10	1.13	1.26	1.37	1.49	1.60	1.70	1.79	1.86	1.94
20	1.08	1.15	1.21	1.25	1.30	1.32	1.34	1.36	1.36
25	1.06	1.11	1.15	1.17	1.20	1.20	1.20	1.22	1.22
30	1.05	1.08	1.09	1.10	1.10	1.11	1.10	1.10	1.09
40	1.02	1.02	1.01	1.00	0.98	0.97	0.94	0.92	0.90
50	0.99	0.97	0.94	0.92	0.88	0.85	0.82	0.78	0.75
60	0.97	0.93	0.88	0.84	0.79	0.75	0.71	0.67	0.63
70	0.94	0.88	0.82	0.77	0.71	0.66	0.61	0.56	0.52
75	0.93	0.86	0.79	0.73	0.67	0.62	0.56	0.51	0.47
80	0.91	0.83	0.75	0.69	0.63	0.57	0.52	0.47	0.42
90	0.88	0.77	0.68	0.61	0.53	0.47	0.41	0.36	0.32
95	0.84	0.73	0.63	0.55	0.47	0.40	0.34	0.29	0.25
97	0.82	0.70	0.60	0.51	0.43	0.36	0.31	0.26	0.22
99	0.78	0.66	0.55	0.45	0.37	0.31	0.25	0.20	0.16
99.5	0.76	0.63	0.52	0.42	0.34	0.28	0.23	0.18	0.14
99.7	0.75	0.62	0.51	0.41	0.32	0.26	0.21	0.16	0.12
99.9	0.73	0.59	0.47	0.37	0.29	0.23	0.18	0.14	0.10

$C_s/C_v = 6$

$p\%$	Коэффициент вариации, $C_v$								
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.01	1.60	2.48	3.75	5.48	7.30	9.39	11.55	13.80	16.40
0.1	1.41	2.02	2.80	3.68	4.58	5.54	6.57	7.63	8.79
0.3	1.35	1.83	2.38	2.98	3.64	4.31	5.00	5.66	6.38
0.5	1.32	1.74	2.22	2.73	3.26	3.82	4.38	4.93	5.51
1.0	1.29	1.63	2.01	2.40	2.81	3.22	3.63	4.03	4.44
3	1.21	1.45	1.68	1.92	2.14	2.38	2.60	2.82	3.04
5	1.18	1.37	1.55	1.73	1.89	2.05	2.20	2.36	2.51
10	1.14	1.26	1.37	1.47	1.56	1.66	1.73	1.82	1.90
20	1.08	1.14	1.19	1.23	1.27	1.30	1.32	1.34	1.36
25	1.07	1.10	1.13	1.16	1.18	1.19	1.20	1.21	1.20
30	1.04	1.07	1.08	1.10	1.10	1.10	1.10	1.09	1.08
40	1.02	1.02	1.01	0.99	0.98	0.96	0.94	0.92	0.89
50	0.99	0.97	0.94	0.91	0.88	0.85	0.82	0.79	0.75
60	0.96	0.92	0.88	0.84	0.80	0.76	0.72	0.68	0.64
70	0.94	0.88	0.83	0.77	0.72	0.67	0.63	0.58	0.54
75	0.93	0.86	0.80	0.74	0.68	0.63	0.58	0.53	0.49
80	0.91	0.84	0.77	0.70	0.64	0.58	0.53	0.48	0.44
90	0.88	0.78	0.70	0.62	0.55	0.49	0.43	0.38	0.33
95	0.85	0.74	0.65	0.56	0.49	0.43	0.37	0.32	0.27
97	0.83	0.72	0.62	0.53	0.46	0.39	0.33	0.28	0.24
99	0.80	0.67	0.57	0.48	0.40	0.33	0.28	0.23	0.19
99.5	0.78	0.65	0.55	0.45	0.37	0.31	0.25	0.20	0.17
99.7	0.76	0.64	0.53	0.43	0.36	0.29	0.24	0.19	0.15
99.9	0.75	0.61	0.50	0.40	0.33	0.26	0.21	0.16	0.12

Ординаты кривых обеспеченностей Н-распределения (А.В. Ежова)  
в модульных коэффициентах  $k_p = f(C_s, C_v, C_v, p)$

$p \%$	Коэффициент вариации, $C_v$									
	0.1	0.2	0.3	0.4	0.1	0.2	0.3	0.4	0.5	
	$C_s / C_v = 0.3$					$C_s / C_v = 0.5$				
0.01	1.38	1.76	2.15	2.52	1.38	1.78	2.17	2.57	2.76	
0.05	1.34	1.67	2.01	2.35	1.34	1.68	2.04	2.38	2.55	
0.1	1.31	1.63	1.95	2.26	1.32	1.64	1.97	2.30	2.46	
0.5	1.26	1.52	1.79	2.05	1.26	1.53	1.80	2.08	2.21	
1	1.24	1.47	1.71	1.95	1.24	1.48	1.72	1.97	2.09	
3	1.19	1.38	1.57	1.77	1.19	1.38	1.58	1.78	1.88	
5	1.17	1.33	1.50	1.67	1.17	1.34	1.51	1.68	1.77	
10	1.13	1.26	1.39	1.52	1.13	1.26	1.39	1.52	1.59	
20	1.08	1.17	1.25	1.34	1.08	1.17	1.25	1.34	1.38	
30	1.05	1.10	1.16	1.21	1.05	1.10	1.15	1.21	1.23	
40	1.02	1.05	1.07	1.10	1.02	1.05	1.07	1.09	1.10	
50	1.00	1.00	1.00	0.99	1.00	1.00	0.99	0.99	0.99	
60	0.97	0.95	0.92	0.89	0.97	0.95	0.92	0.88	0.87	
70	0.95	0.89	0.84	0.78	0.95	0.89	0.84	0.78	0.74	
80	0.92	0.83	0.74	0.65	0.92	0.83	0.74	0.65	0.60	
90	0.87	0.74	0.62	0.48	0.87	0.74	0.62	0.48	0.42	
95	0.84	0.67	0.51	0.34	0.84	0.68	0.52	0.36	0.28	
97	0.81	0.63	0.45	0.26	0.81	0.63	0.46	0.28	0.20	
99	0.77	0.54	0.33	0.13	0.77	0.55	0.34	0.16	0.09	
99.5	0.74	0.50	0.26	0.08	0.75	0.51	0.29	0.11	0.05	
99.9	0.70	0.41	0.15	0.02	0.70	0.42	0.18	0.04	0.01	

Приложение 3 (продолжение)

$$C_s/C_v = 1.0$$

p %	Коэффициент вариации, $C_v$					
	0.1	0.2	0.3	0.4	0.5	0.6
0.01	1.39	1.81	2.26	2.72	3.18	3.65
0.05	1.34	1.71	2.10	2.50	2.91	3.32
0.1	1.32	1.67	2.03	2.40	2.79	3.17
0.5	1.27	1.55	1.84	2.15	2.46	2.78
1	1.24	1.49	1.76	2.03	2.31	2.60
3	1.19	1.39	1.60	1.82	2.04	2.26
5	1.17	1.34	1.52	1.70	1.89	2.09
10	1.13	1.26	1.40	1.53	1.68	1.82
20	1.08	1.17	1.25	1.33	1.42	1.51
30	1.05	1.10	1.15	1.19	1.24	1.28
50	1.00	0.99	0.98	0.97	0.96	0.93
70	0.95	0.89	0.83	0.76	0.69	0.62
80	0.92	0.83	0.74	0.65	0.55	0.45
90	0.87	0.75	0.62	0.50	0.38	0.26
95	0.84	0.68	0.53	0.39	0.26	0.15
97	0.82	0.64	0.48	0.33	0.20	0.10
99	0.77	0.57	0.38	0.22	0.11	0.04
99.5	0.75	0.53	0.33	0.18	0.07	0.02
99.9	0.70	0.45	0.24	0.10	0.03	0.0004

$$C_s/C_v = 1.5$$

p %	Коэффициент вариации, $C_v$							
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
0.01	1.40	1.86	2.37	2.92	3.51	4.14	4.80	5.50
0.05	1.35	1.75	2.18	2.66	3.16	3.37	4.27	4.86
0.1	1.33	1.70	2.10	2.54	3.00	3.50	4.02	4.57
0.5	1.27	1.57	1.89	2.23	2.60	2.99	3.40	3.83
1	1.24	1.51	1.79	2.09	2.41	2.75	3.11	3.48
3	1.20	1.40	1.62	1.85	2.09	2.34	2.61	2.88
5	1.17	1.35	1.53	1.72	1.92	2.13	2.35	2.57
10	1.13	1.26	1.40	1.54	1.68	1.82	1.97	2.12
20	1.08	1.17	1.25	1.32	1.40	1.48	1.55	1.61
30	1.05	1.10	1.14	1.18	1.21	1.24	1.26	1.28
50	1.00	0.99	0.98	0.96	0.93	0.90	0.86	0.81
70	0.95	0.89	0.83	0.76	0.69	0.61	0.54	0.46
80	0.92	0.83	0.74	0.65	0.56	0.47	0.38	0.30
90	0.87	0.75	0.63	0.52	0.41	0.31	0.22	0.15
95	0.84	0.69	0.54	0.42	0.31	0.21	0.13	0.08
97	0.82	0.65	0.50	0.36	0.25	0.16	0.09	0.05
99	0.78	0.58	0.41	0.27	0.16	0.09	0.045	0.019
99.5	0.76	0.54	0.37	0.23	0.13	0.06	0.028	0.011
99.9	0.71	0.47	0.29	0.15	0.07	0.03	0.006	0.002

$\chi^2$  - распределение.

Значения квантилей даны в зависимости от числа степеней свободы  $\nu$  и уровня значимости  $\alpha$  %.

$F(x)$	0.5	2.5	5	10	90	95	97.5	99.5	
$\alpha$	99.5	97.5	95	90	10	5	2.5	0.5	
Число степеней свободы, $\nu$	3	0.071	0.216	0.352	0.584	6.251	7.815	9.348	12.838
	4	0.207	0.484	0.711	1.064	7.779	9.488	11.143	14.830
	5	0.412	0.831	1.145	1.610	9.236	11.070	12.832	16.750
	6	0.676	1.237	1.635	2.204	10.645	12.592	14.449	18.475
	7	0.989	1.690	2.167	2.833	12.017	14.067	16.013	20.278
	8	1.314	2.180	2.733	3.490	13.362	15.507	17.535	21.955
	9	1.735	2.700	3.325	4.168	14.684	16.919	19.023	23.589
	10	2.156	3.247	3.940	4.865	15.987	18.307	20.484	25.188
	11	2.603	3.816	4.575	5.578	17.275	19.675	21.920	26.757
	12	3.074	4.404	5.226	6.304	18.549	21.026	23.336	28.300
	13	3.565	5.009	5.892	7.042	19.812	22.362	24.736	29.819
	14	4.075	5.629	6.571	7.790	21.064	23.685	26.119	31.319
	15	4.601	6.262	7.261	8.547	22.307	24.996	27.488	32.804
	16	5.142	6.908	7.962	9.312	23.542	26.296	28.845	34.267
	17	5.697	7.564	8.672	10.085	24.769	27.587	30.191	35.713
	18	6.265	8.231	9.390	10.865	25.989	28.869	31.526	37.156
	19	6.844	8.907	10.117	11.651	27.204	30.144	32.852	38.582
	20	7.434	9.591	10.851	12.443	28.412	31.410	34.170	39.997
	21	8.034	10.283	11.591	13.240	29.615	32.671	35.479	41.401
	22	8.643	10.982	12.338	14.041	30.813	33.924	36.781	42.796
	23	9.260	11.688	13.091	14.848	32.007	35.172	38.076	44.181
	24	9.886	12.401	13.848	15.659	33.196	36.415	39.364	45.558
	25	10.520	13.120	14.671	16.473	34.382	37.652	40.646	46.928
	26	11.760	13.844	15.379	17.292	35.563	38.885	41.923	48.290
	27	11.808	14.573	16.151	18.114	36.741	40.113	42.194	49.645
	28	12.461	15.308	16.928	18.939	37.916	41.337	44.461	50.993
	29	13.121	16.047	17.708	19.768	39.087	42.557	45.722	52.336
	30	13.787	16.791	18.498	20.599	40.256	43.773	46.979	53.672

**Распределение Стьюдента.**

Абсолютные значения квантилей даны в зависимости от числа степеней свободы  $\nu$  для одностороннего ( $\alpha$  %) и двухстороннего ( $2\alpha$  %) уровней значимости.

$F(t)$	90	95	97.5	99.0	99.5
$2\alpha$	20	10	5	2	1
$\alpha$	10	5	2.5	1	0.5
1	3.077	6.313	12.71	31.82	63.66
2	1.885	2.920	4.302	6.964	9.924
3	1.638	2.353	3.182	4.540	5.840
4	1.533	2.132	2.776	3.746	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.439	1.943	2.446	3.142	3.707
7	1.415	1.895	2.365	2.998	3.500
8	1.397	1.859	2.306	2.897	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.813	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.105
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.625	2.976
15	1.341	1.753	2.131	2.603	2.947
16	1.336	1.746	2.119	2.583	2.920
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.551	2.878
19	1.328	1.729	2.093	2.540	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.079	2.517	2.831
22	1.321	1.717	2.074	2.508	2.819
24	1.318	1.711	2.064	2.492	2.797
26	1.315	1.706	2.055	2.478	2.778
28	1.313	1.701	2.048	2.467	2.763
30	1.310	1.697	2.042	2.457	2.750
32	1.308	1.693	2.036	2.448	2.738
34	1.307	1.691	2.032	2.441	2.728
36	1.305	1.688	2.028	2.435	2.720
38	1.304	1.686	2.024	2.429	2.712
40	1.303	1.684	2.021	2.423	2.705
50	1.299	1.676	2.009	2.403	2.678
60	1.296	1.671	2.000	2.390	2.660
80	1.292	1.664	1.990	2.373	2.638
100	1.290	1.660	1.984	2.364	2.626
200	1.286	1.653	1.972	2.345	2.601
300	1.284	1.650	1.968	2.339	2.592
400	1.284	1.649	1.966	2.336	2.588
500	1.283	1.647	1.964	2.333	2.585

Число степеней свободы  $\nu$

**F – распределение (Фишера).**

Значения квантилей даны в зависимости от числа степеней свободы  $\nu_1$  (для большей дисперсии) и  $\nu_2$  (для меньшей дисперсии) при уровне значимости  $2\alpha$  1, 5 и 10 %.

$2\alpha = 10\%$

$\nu_2$	Число степеней свободы $\nu_1$										
	1	4	6	8	10	15	20	30	60	120	$\infty$
1	161	225	234	239	242	246	248	250	252	253	254
2	18.5	19.2	19.3	19.4	19.4	19.4	19.4	19.5	19.5	19.5	19.5
3	10.1	9.12	8.94	8.84	8.78	8.70	8.66	8.62	8.57	8.56	8.53
4	7.71	6.39	6.16	6.04	5.96	5.86	5.80	5.74	5.68	5.66	5.63
5	6.61	5.19	4.95	4.82	4.47	4.62	4.56	4.50	4.42	4.40	4.63
6	5.99	4.53	4.28	4.15	4.06	3.94	3.87	3.81	3.72	3.71	3.67
7	5.59	4.12	3.87	3.73	3.63	3.51	3.44	3.38	3.28	3.28	3.23
8	5.32	3.84	3.58	3.44	3.34	3.22	3.15	3.08	3.00	2.98	2.93
9	5.12	3.63	3.37	3.23	3.13	3.01	2.93	2.86	2.77	2.76	2.71
10	4.96	3.48	3.22	3.07	2.97	2.85	2.77	2.70	2.61	2.59	2.54
11	4.84	3.36	3.09	2.95	2.86	2.72	2.65	2.57	2.47	2.45	2.40
12	4.75	3.26	3.00	2.85	2.76	2.62	2.54	2.46	2.36	2.35	2.30
13	4.67	3.18	2.92	2.77	2.67	2.53	2.46	2.38	2.28	2.26	2.21
14	4.60	3.11	2.85	2.70	2.60	2.46	2.39	2.31	2.21	2.19	2.13
15	4.54	3.06	2.79	2.64	2.55	2.40	2.33	2.25	2.15	2.12	2.07
16	4.49	3.01	2.74	2.59	2.49	2.35	2.28	2.20	2.09	2.07	2.01
17	4.45	2.96	2.70	2.55	2.45	2.31	2.23	2.15	2.04	2.02	1.96
18	4.41	2.93	2.66	2.51	2.41	2.27	2.19	2.11	2.00	1.98	1.92
19	4.38	2.90	2.63	2.48	2.38	2.23	2.15	2.07	1.96	1.94	1.88
20	4.35	2.87	2.60	2.45	2.35	2.20	2.12	2.04	1.92	1.90	1.84
21	4.32	2.84	2.57	2.42	2.32	2.18	2.09	2.00	1.89	1.87	1.81
22	4.30	2.82	2.55	2.40	2.30	2.15	2.07	1.98	1.87	1.84	1.78
23	4.28	2.80	2.53	2.38	2.28	2.13	2.04	1.96	1.84	1.82	1.76
24	4.26	2.78	2.51	2.36	2.26	2.11	2.02	1.94	1.82	1.80	1.73
25	4.24	2.76	2.49	2.34	2.24	2.09	2.00	1.92	1.80	1.77	1.71
26	4.23	2.74	2.47	2.32	2.22	2.07	1.99	1.90	1.78	1.76	1.69
27	4.21	2.73	2.46	2.30	2.20	2.06	1.97	1.88	1.76	1.74	1.67
28	4.20	2.71	2.44	2.29	2.19	2.04	1.96	1.87	1.75	1.72	1.65
29	4.18	2.70	2.43	2.28	2.18	2.03	1.94	1.85	1.73	1.71	1.64
30	4.17	2.69	2.42	2.27	2.16	2.01	1.93	1.84	1.72	1.69	1.62
40	4.08	2.61	2.34	2.18	2.07	1.92	1.84	1.74	1.61	1.59	1.51
60	4.00	2.52	2.25	2.10	1.99	1.84	1.75	1.65	1.50	1.48	1.39
120	3.92	2.44	2.17	2.01	1.90	1.75	1.65	1.65	1.49	1.36	1.25
$\infty$	3.84	2.37	2.09	1.94	1.83	1.67	1.57	1.46	1.28	1.24	1.00

Приложение 6 (продолжение)

F – распределение (Фишера).

$2\alpha = 5\%$

$\nu_2$	Число степеней свободы $\nu_1$										
	1	4	6	8	10	15	20	30	60	120	$\infty$
1	648	899	937	957	969	985	993	1001	1010	1014	1018
2	38.5	39.2	39.3	39.4	39.4	39.4	39.4	39.5	39.5	39.5	39.5
3	17.4	15.1	14.7	14.6	14.4	14.3	14.2	14.1	14.0	13.9	13.9
4	12.2	9.60	9.20	8.98	8.84	8.66	8.56	8.46	8.36	8.31	8.26
5	10.0	7.39	6.98	6.76	6.62	6.42	6.33	6.23	6.12	6.07	6.02
6	8.81	6.23	5.82	5.60	5.46	5.27	5.17	5.07	4.96	4.90	4.85
7	8.07	5.52	5.12	4.90	4.76	4.57	4.47	4.36	4.25	4.20	4.14
8	7.57	5.05	4.65	4.43	4.30	4.10	4.00	3.89	3.78	3.73	3.67
9	7.21	4.72	4.32	4.10	3.96	3.77	3.67	3.56	3.45	3.39	3.33
10	6.94	4.47	4.07	3.85	3.72	3.52	3.42	3.31	3.20	3.14	3.08
11	6.72	4.28	3.88	3.66	3.53	3.33	3.23	3.12	3.00	2.94	2.88
12	6.55	4.12	3.73	3.51	3.37	3.18	3.07	2.96	2.85	2.79	2.72
13	6.41	4.00	3.61	3.39	3.25	3.05	2.95	2.84	2.72	2.66	2.60
14	6.30	3.89	3.50	3.29	3.15	2.95	2.84	2.73	2.61	2.55	2.49
15	6.20	3.80	3.41	3.20	3.05	2.86	2.76	2.64	2.52	2.46	2.40
16	6.16	3.73	3.34	3.12	2.99	2.79	2.68	2.57	2.45	2.38	2.32
17	6.04	3.66	3.28	3.06	2.92	2.72	2.62	2.50	2.38	2.32	2.25
18	5.98	3.61	3.22	3.01	2.87	2.67	2.56	2.44	2.32	2.26	2.19
19	5.92	3.56	3.17	2.96	2.82	2.62	2.51	2.38	2.27	2.20	2.13
20	5.87	3.51	3.13	2.91	2.77	2.57	2.46	2.35	2.22	2.16	2.09
21	5.83	3.47	3.09	2.87	2.73	2.53	2.42	2.31	2.18	2.11	2.04
22	5.79	3.44	3.05	2.84	2.70	2.50	2.39	2.27	2.14	2.08	2.03
23	5.75	3.41	3.02	2.81	2.67	2.47	2.36	2.24	2.11	2.04	1.97
24	5.72	3.38	2.99	2.78	2.64	2.44	2.33	2.21	2.08	2.01	1.94
25	5.69	3.35	2.97	2.75	2.61	2.41	2.30	2.18	2.05	1.98	1.91
26	5.66	3.33	2.94	2.72	2.59	2.39	2.28	2.16	2.03	1.95	1.88
27	5.63	3.31	2.92	2.71	2.57	2.34	2.25	2.13	2.00	1.93	1.85
28	5.61	3.29	2.90	2.69	2.55	2.34	2.23	2.11	1.98	1.91	1.83
29	5.59	3.27	2.88	2.67	2.53	2.32	2.21	2.09	1.96	1.89	1.81
30	5.57	3.25	2.87	2.65	2.51	2.31	2.20	2.07	1.94	1.87	1.77
40	5.42	3.13	2.74	2.53	2.39	2.18	2.07	1.94	1.80	1.72	1.74
60	5.28	3.01	2.63	2.41	2.27	2.06	1.94	1.82	1.67	1.58	1.48
120	5.15	2.89	2.52	2.30	2.16	1.95	1.82	1.69	1.53	1.43	1.31
$\infty$	5.02	2.79	2.41	2.19	2.05	1.83	1.71	1.57	1.39	1.27	1.00

Приложение 6 (продолжение)

F – распределение (Фишера).

$$2\alpha = 1\%$$

v <sub>2</sub>	Число степеней свободы v <sub>1</sub>										
	1	4	6	8	10	15	20	30	60	120	∞
2	198	199	199	199	199	199	199	199	199	199	199
3	55.6	46.2	44.8	44.1	43.7	43.1	42.8	42.5	42.1	42.0	41.8
4	31.3	23.2	22.0	21.4	21.0	20.4	20.2	19.9	19.6	19.5	19.3
5	22.8	15.6	14.5	14.0	13.6	13.1	12.9	12.7	12.4	12.3	12.1
6	18.6	12.1	11.1	10.6	10.3	9.81	9.59	9.36	9.12	9.00	8.88
7	16.2	10.1	9.16	8.68	8.38	7.97	7.75	7.53	7.31	7.19	7.08
8	14.7	8.81	7.95	7.50	7.21	6.81	6.61	6.40	6.18	6.06	5.95
9	13.6	7.96	7.13	6.69	6.42	6.03	5.83	5.52	5.41	5.30	5.19
10	12.8	7.34	6.51	6.12	5.85	5.47	5.27	5.07	4.86	4.75	4.64
11	12.2	6.88	6.10	5.68	5.42	5.05	4.86	4.65	4.45	4.34	4.23
12	11.7	6.52	5.76	5.35	5.09	4.72	4.53	4.33	4.12	4.01	3.90
13	11.4	6.23	5.48	5.08	4.82	4.46	4.27	3.97	3.76	3.76	3.65
14	11.1	6.00	5.26	4.86	4.60	4.25	4.06	3.86	3.66	3.55	3.44
15	10.8	6.80	5.08	4.67	4.42	4.07	3.88	3.69	3.48	3.37	3.26
16	10.6	5.54	4.91	4.52	4.27	3.96	3.73	3.54	3.33	3.22	3.11
17	10.4	5.50	4.78	4.39	4.14	3.79	3.61	3.41	3.21	3.10	2.98
18	10.2	5.37	4.66	4.28	4.03	3.68	3.50	3.30	3.10	2.99	2.87
19	10.1	5.27	4.56	4.18	3.93	3.59	3.40	3.21	3.00	2.89	2.78
20	9.94	5.17	4.47	4.09	3.85	3.50	3.32	3.12	2.92	2.81	2.69
21	9.83	5.09	4.39	4.01	3.77	3.43	3.24	3.05	2.84	2.73	2.61
22	9.73	5.02	4.32	3.94	3.70	3.36	3.18	2.98	2.77	2.66	2.55
23	9.63	4.95	4.26	3.88	3.64	3.30	3.12	2.92	2.71	2.60	2.48
24	9.55	4.89	4.20	3.83	3.59	3.25	3.06	2.87	2.66	2.55	2.43
25	9.48	4.84	4.15	3.76	3.54	3.20	3.01	2.82	2.61	2.50	2.38
26	9.41	4.79	4.10	3.73	3.49	3.15	2.97	2.77	2.56	2.45	2.33
27	9.34	4.74	4.06	3.69	3.44	3.11	2.93	2.73	2.52	2.41	2.29
28	9.28	4.70	4.02	3.65	3.41	3.07	2.89	2.69	2.48	2.37	2.25
29	9.23	4.66	3.98	3.61	3.38	3.04	2.86	2.66	2.45	2.33	2.21
30	9.18	4.62	3.95	3.58	3.34	3.01	2.82	2.63	2.42	2.30	2.18
40	8.83	4.37	3.71	3.35	3.12	2.78	2.59	2.40	2.18	2.06	1.93
60	8.48	4.14	3.49	3.13	2.90	2.57	2.39	2.19	1.96	1.83	1.69
120	8.18	3.92	3.25	2.93	2.71	2.37	2.19	1.98	1.75	1.61	1.43
∞	7.88	3.72	3.09	2.74	2.52	2.19	2.00	1.79	1.53	1.36	1.00

# Содержание

<b>ПРЕДИСЛОВИЕ</b> .....	9
<b>СПИСОК УСЛОВНЫХ ОБОЗНАЧЕНИЙ И СОКРАЩЕНИЙ</b> ...	11
<b>ВВЕДЕНИЕ</b> .....	12
<b>1. НЕКОТОРЫЕ СВЕДЕНИЯ ИЗ ТЕОРИИ ВЕРОЯТНОСТЕЙ</b> .....	13
1.1. Случайные величины и функции распределения .....	13
1.2. Дискретные и непрерывные случайные величины .....	15
1.3. Случайная величина и время .....	17
1.4. Числовые характеристики случайных величин .....	18
1.4.1. Мода. Медиана. Математическое ожидание.....	18
1.4.2. Моменты случайной величины .....	20
1.4.3. Дисперсия. Среднеквадратическое отклонение. Коэффициент вариации.....	21
1.4.4. Асимметрия и эксцесс .....	22
1.4.5. Свойства числовых характеристик случайных величин .....	22
1.4.6. Стандартные преобразования случайной величины .....	24
1.4.7. Квантили распределения .....	24
<b>2. АНАЛИТИЧЕСКИЕ ФУНКЦИИ РАСПРЕДЕЛЕНИЯ, ИСПОЛЗУЕМЫЕ В ГИДРОЛОГИИ</b> .....	26
2.1. Нормальное распределение .....	26
2.2. Закон равномерной плотности .....	29
2.3. Логарифмически нормальное распределение .....	30
2.4. Закон распределения крайних членов выборки (распределение Гумбеля) .....	34
2.5. Распределение Пирсона III типа (непрерывное биномиальное распределение) .....	37
2.6. Распределение Крицкого – Менкеля (трехпараметрическое гамма-распределение) .....	40
2.7. Распределение Джонсона .....	43
2.8. Графическое представление функций распределения на клетчатке вероятности .....	44
2.9. Комментарии к разделу .....	46
<b>3. ПОСТРОЕНИЕ КРИВЫХ ОБЕСПЕЧЕННОСТЕЙ И ОЦЕНКА ПАРАМЕТРОВ РАСПРЕДЕЛЕНИЯ ПО ЭМПИРИЧЕСКИМ ДАННЫМ</b> .....	50
3.1. Эмпирические кривые обеспеченностей.....	52

3.2.	Методы расчета оценок параметров распределения .....	60
3.2.1.	Метод моментов .....	61
3.2.2.	Метод наибольшего правдоподобия .....	64
3.2.3.	Графоаналитический метод (метод квантилей) .....	68
3.2.3.1.	Графоаналитический метод на основе кривой Пирсона III типа .....	68
3.2.3.2.	Графоаналитический метод на основе трехпараметрической логнормальной кривой .....	70
3.2.3.3.	Графоаналитический метод на основе кривой Крицкого – Менкеля .....	72
3.2.3.4.	Комментарии .....	75
3.2.4.	Графический метод .....	76
3.2.5.	Оценка параметров распределения Джонсона .....	78
3.3.	Оценка погрешностей выборочных параметров распределения .....	81
3.4.	Экономические аспекты определения расчетных вероятностей .....	83
3.4.1.	Цели определения расчетных вероятностей .....	83
3.4.2.	Определение оптимальных расчетных вероятностей .....	84
<b>4.</b>	<b>ИНТЕРВАЛЬНОЕ ОЦЕНИВАНИЕ ПАРАМЕТРОВ И ПРОВЕРКА СТАТИСТИЧЕСКИХ ГИПОТЕЗ .....</b>	<b>89</b>
4.1.	Распределение $\chi^2$ (хи - квадрат) .....	89
4.2.	t - распределение (Стьюдента) .....	90
4.3.	F - распределение (Фишера) .....	92
4.4.	Интервальные оценки параметров распределения .....	93
4.4.1.	Интервальная оценка математического ожидания .....	94
4.4.2.	Интервальная оценка дисперсии .....	95
4.5.	Статистическая гипотеза. Критерий статистической гипотезы .....	96
4.6.	Критерии, используемые для проверки однородности гидрологических рядов .....	100
4.6.1.	Критерий Стьюдента для проверки значимости различия средних значений двух выборок .....	100
4.6.2.	Критерий равенства двух дисперсий (Критерий Фишера) .....	102
4.6.3.	Рангово-суммарный критерий Уилкоксона-Манна-Уитни .....	103
4.6.4.	Ранговый критерий рассеяния Зигеля-Тьюки .....	107
4.7.	Критерии согласия .....	109
4.7.1.	Критерий $\chi^2$ (Пирсона) .....	110
4.7.2.	Критерий Колмогорова .....	114
4.7.3.	Критерий $n\omega^2$ (Крамера-Мизеса-Смирнова) .....	116
4.8.	Критерии случайности .....	118
4.8.1.	Критерии серий .....	119
4.8.1.1.	Критерий общего числа серий .....	119
4.8.1.2.	Критерий наибольшей длины серий .....	120
4.8.2.	Критерий числа повышений и понижений .....	121
4.8.3.	Критерий числа экстремумов .....	122

## Приложение 9

Таблица для определения коэффициента вариации  $C_v$  методом приближенного наибольшего правдоподобия в зависимости от статистики -  $\lambda_2$  при фиксированном соотношении  $C_s/C_v$  (для кривой обеспеченностей Пирсона III типа).

$$C_s = C_v$$

$C_v$	0	1	2	3	4	5	6	7	8	9
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	0.0360	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	0.435
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	-	-	-	-	-	-	-	-	-

$$C_s = 2C_v$$

$C_v$	0	1	2	3	4	5	6	7	8	9
0.1	0.0022	0.0026	0.0031	0.0037	0.0043	0.0049	0.0056	0.0063	0.0071	0.0079
0.2	0.0087	0.0096	0.0106	0.0116	0.0126	0.0137	0.0148	0.0160	0.0172	0.0185
0.3	0.0198	0.0212	0.0296	0.0241	0.0256	0.0271	0.0287	0.0304	0.0321	0.0339
0.4	0.0357	0.0375	0.0394	0.0414	0.0434	0.0454	0.0476	0.0497	0.0519	0.0542
0.5	0.0565	0.0589	0.0613	0.0638	0.0664	0.0690	0.0716	0.0743	0.0771	0.0799
0.6	0.0828	0.0858	0.0887	0.0918	0.0949	0.0981	0.101	0.105	0.108	0.111
0.7	0.115	0.118	0.122	0.126	0.129	0.133	0.137	0.141	0.145	0.149
0.8	0.153	0.157	0.162	0.166	0.170	0.175	0.180	0.184	0.189	0.194
0.9	0.198	0.203	0.208	0.213	0.218	0.224	0.229	0.234	0.240	0.245
1.0	0.251	0.256	0.262	0.268	0.274	0.280	0.286	0.292	0.298	0.304
1.1	0.311	0.317	0.323	0.330	0.337	0.343	0.350	0.357	0.364	0.371
1.2	0.378	0.386	0.393	0.400	0.408	0.415	0.423	0.431	0.438	0.446
1.3	0.454	0.462	0.471	0.479	0.487	0.496	0.504	0.513	0.521	0.530
1.4	0.539	0.548	0.557	0.566	0.575	0.584	0.594	0.603	0.613	0.623
1.5	0.632	0.642	0.652	0.662	0.672	0.682	0.692	0.703	0.713	0.724
1.6	0.734	0.745	0.756	0.767	0.778	0.789	0.800	0.811	0.823	0.824
1.7	0.846	0.857	0.860	0.881	0.893	0.905	0.917	0.929	0.911	0.953
1.8	0.966	0.978	0.991	1.004	1.016	1.029	1.042	1.055	1.069	1.082
1.9	1.095	1.109	1.122	1.136	1.150	1.163	1.177	1.191	1.205	1.220

## Приложение 9 (продолжение)

$$C_s = 3C_v$$

$C_v$	0	1	2	3	4	5	6	7	8	9
0.1	0.0022	0.0026	0.0031	0.0037	0.0043	0.0048	0.0055	0.0062	0.0070	0.0078
0.2	0.0085	0.0094	0.0104	0.0113	0.0122	0.0132	0.0143	0.0154	0.0165	0.0176
0.3	0.0188	0.0201	0.0214	0.0227	0.0240	0.0253	0.0267	0.0282	0.0297	0.0311
0.4	0.0326	0.0342	0.0359	0.0375	0.0391	0.0408	0.0425	0.0443	0.0461	0.0479
0.5	0.0497	0.0526	0.0553	0.0555	0.0574	0.0594	0.0614	0.0635	0.0656	0.0677
0.6	0.0698	0.0720	0.0742	0.0764	0.0786	0.0808	0.0832	0.0856	0.0879	0.0903
0.7	0.0926	0.0951	0.0976	0.100	0.103	0.105	0.108	0.110	0.113	0.116
0.8	0.118	0.121	0.124	0.127	0.129	0.132	0.135	0.138	0.141	0.144
0.9	0.147	0.150	0.153	0.156	0.159	0.162	0.165	0.168	0.171	0.174
1.0	0.177	0.181	0.184	0.187	0.191	0.194	0.197	0.201	0.204	0.208
1.1	0.211	0.215	0.218	0.222	0.225	0.229	0.232	0.236	0.240	0.243
1.2	0.247	0.251	0.255	0.258	0.262	0.266	0.270	0.274	0.277	0.281
1.3	0.286	0.289	0.294	0.298	0.302	0.306	0.311	0.315	0.319	0.323
1.4	0.327	0.332	0.336	0.340	0.345	0.349	0.354	0.358	0.363	0.367
1.5	0.372	0.376	0.381	0.386	0.390	0.395	0.400	0.404	0.409	0.414
1.6	0.419	0.424	0.428	0.433	0.438	0.443	0.448	0.453	0.458	0.463
1.7	0.468	0.474	0.479	0.484	0.489	0.495	0.500	0.505	0.511	0.516
1.8	0.521	0.527	0.532	0.538	0.543	0.549	0.555	0.560	0.566	0.572
1.9	0.578	0.584	0.589	0.595	0.600	0.606	0.612	0.618	0.624	0.630

$$C_s = 4C_v$$

$C_v$	0	1	2	3	4	5	6	7	8	9
0.1	0.0022	0.0026	0.0031	0.0037	0.0043	0.0048	0.0055	0.0062	0.0069	0.0076
0.2	0.0083	0.0092	0.0101	0.0110	0.0119	0.0128	0.0138	0.0148	0.0159	0.0169
0.3	0.0180	0.0192	0.0203	0.0215	0.0227	0.0239	0.0252	0.0265	0.0279	0.0292
0.4	0.0305	0.0319	0.0334	0.0348	0.0363	0.0377	0.0392	0.0408	0.0423	0.0439
0.5	0.0454	0.0471	0.0488	0.0505	0.0522	0.0538	0.0556	0.0574	0.0592	0.0609
0.6	0.0627	0.0646	0.0665	0.0683	0.0702	0.0721	0.0741	0.0760	0.0780	0.0800
0.7	0.0820	0.0840	0.0861	0.0882	0.0903	0.0923	0.0946	0.0966	0.0989	0.101
0.8	0.103	0.105	0.108	0.110	0.112	0.114	0.117	0.119	0.121	0.124
0.9	0.126	0.128	0.131	0.133	0.136	0.138	0.141	0.143	0.146	0.148
1.0	0.151	0.153	0.156	0.158	0.161	0.163	0.166	0.169	0.171	0.174
1.1	0.177	0.179	0.182	0.185	0.188	0.190	0.193	0.196	0.199	0.201
1.2	0.204	0.207	0.210	0.213	0.216	0.219	0.222	0.224	0.227	0.230
1.3	0.233	0.236	0.239	0.242	0.245	0.248	0.251	0.254	0.258	0.261
1.4	0.264	0.267	0.270	0.273	0.276	0.279	0.283	0.286	0.289	0.292
1.5	0.296	0.299	0.302	0.305	0.309	0.312	0.315	0.319	0.322	0.325
1.6	0.329	0.332	0.336	0.339	0.342	0.346	0.349	0.353	0.356	0.360
1.7	0.363	0.367	0.370	0.374	0.377	0.381	0.384	0.388	0.392	0.395
1.8	0.399	0.403	0.406	0.410	0.414	0.417	0.421	0.425	0.429	0.432
1.9	0.436	0.440	0.444	0.447	0.451	0.455	0.459	0.463	0.467	0.471

Учебное издание

Владислав Сергеевич Дружинин  
Александр Владимирович Сикан

Методы статистической обработки  
гидрометеорологической информации

Учебное пособие

Редактор И.Г. Максимова

ЛР № 020309 от 30.12.96

---

Подписано в печать 26.09.01	Формат 60x90 <sup>1</sup> / <sub>16</sub>	Бумага офсетная
Печать офсетная	Печ. л. 11,0	Тираж 500 экз.

---

РГГМУ, 195196, СПб, Малоохтинский пр. 98  
Отпечатано ООО «АСпринт»

1900

State of Tennessee  
County of \_\_\_\_\_

Know all men by these presents that \_\_\_\_\_  
of the County of \_\_\_\_\_ State of Tennessee

do hereby certify that

\_\_\_\_\_

\_\_\_\_\_

Witness my hand and seal of office this \_\_\_\_\_ day of \_\_\_\_\_ 1900.  
\_\_\_\_\_