№ 2770

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РФ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ «НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ ТЕХНОЛОГИЧЕСКИЙ УНИВЕРСИТЕТ «МИСиС»

Кафедра математики

В.А. Карасев
Г.Д. Лёвшина

Теория вероятностей и математическая статистика

Математическая статистика

Практикум

Рекомендовано редакционно-издательским советом университета

МИСиС

Москва 2016

УДК 519.2
К21

Р е ц е н з е н т
канд. экон. наук, проф. *В.Ф. Михин*

Карасев В.А.

К21 Теория вероятностей и математическая статистика : математическая статистика : практикум / В.А. Карасев, Г.Д. Лёвшина. – М. : Изд. Дом МИСиС, 2016. – 120 с.
ISBN 978-5-906846-01-3

Практикум предназначен для подготовки к занятиям, контрольным работам и к выполнению лабораторных работ (типовых расчетов) по курсу «Теория вероятностей и математическая статистика» разделу «Математическая статистика». Изложение теоретического материала сопровождается большим количеством примеров решения типовых задач, дается подробное описание выполнения лабораторных работ (типовых расчетов), разбирается решение типовых вариантов контрольных работ. В заключение приводятся варианты контрольных работ с ответами.

Предназначен для студентов всех направлений.

УДК 519.2

ISBN 978-5-906846-01-3

СОДЕРЖАНИЕ

Предисловие ... 5
1. Практические занятия по математической статистике ... 6
1.1. Первичная обработка результатов эксперимента и оценка основных параметров генеральной совокупности ... 6
1.2. Оценка математического ожидания по неравноточным измерениям ... 12
1.3. Оценка дисперсии по нескольким сериям экспериментов ... 14
1.4. Построение гистограммы распределения ... 16
1.5. Некоторые используемые в статистике распределения ... 17
1.6. Построение доверительных интервалов ... 20
1.6.1. Доверительный интервал для математического ожидания ... 21
1.6.2. Доверительный интервал для среднего квадратического отклонения ... 21
1.6.3. Доверительный интервал для дисперсии ... 22
1.7. Проверка статистических гипотез ... 25
1.7.1. Основные понятия ... 25
1.7.2. Проверка гипотез о дисперсии нормального распределения ... 29
1.7.3. Проверка гипотез о математических ожиданиях нормального распределения ... 35
1.7.4. Проверка гипотезы о виде распределения генеральной совокупности ... 43
1.8. Регрессионный анализ. Построение линейной и квадратичной регрессионных моделей ... 46
1.8.1. Оценка коэффициентов регрессии ... 46
1.8.2. Проверка гипотезы об адекватности регрессионной модели ... 49
1.8.3. Построение доверительных интервалов для коэффициентов регрессии ... 50
1.8.4. Задача регрессии для линейной функции ... 51
1.8.5. Задача регрессии для квадратичной функции ... 53

1.9. Линейный корреляционный анализ 62
1.9.1. Двумерный случайный вектор, его выборочные характеристики 62
1.9.2. Построение доверительного интервала для коэффициента корреляции. Проверка гипотезы о существовании линейной зависимости 68
2. Указания по выполнению типовых расчетов (лабораторных работ) «Обработка основных типов экспериментальных данных» 71
2.1. Типовой расчет 1. Сравнение двух случайных выборок (первичная обработка данных, проверка статистических гипотез) 71
2.1.1. Первичная обработка результатов измерений 72
2.1.2. Построение доверительных интервалов 73
2.1.3. Проверка гипотез о равенстве дисперсий, о равенстве математических ожиданий 73
2.1.4. Проверка гипотезы о нормальном распределении объединения двух случайных выборок 77
2.1.5. Построение гистограмм распределения объединения двух случайных выборок 81
2.2. Типовой расчет 2. Обработка данных методами регрессионного анализа 84
2.2.1. Задача 1 85
2.2.2. Задача 2 89
2.3. Типовой расчет 3. Обработка данных методами линейного корреляционного анализа 96
2.3.1. Нахождение оценок числовых характеристик двумерного случайного вектора. Расчет оценки коэффициента корреляции 96
2.3.2. Нахождение уравнений линейной регрессии 97
2.3.3. Построение доверительного интервала для коэффициента корреляции ρ. Проверка гипотезы о существовании линейной зависимости между величинами X и Y 97
2.4. Примерные варианты контрольной работы 97
Библиографический список 106
Приложение 107

ПРЕДИСЛОВИЕ

Пособие является продолжением практикума «Теория вероятностей и математическая статистика» раздел «Теория вероятностей» (библ. № 2454), вышедшего в 2015 г. В нем приведены необходимые сведения об основных современных методах математической статистики и обработки экспериментальных данных на примерах, заимствованных из области металлургии, металловедения и экономики.

Математическая статистика – это наука, которая методами теории вероятностей на основании результатов наблюдений изучает закономерности в массовых случайных явлениях. Математическая статистика (не путать со статистикой – разделом экономической теории) указывает способы сбора и группировки статических данных (результатов наблюдений или экспериментов), разрабатывает методы их обработки для оценки характеристик распределения, для установления зависимости случайной величины от других, для проверки статистических гипотез о виде распределения или значениях его параметров. Математическая статистика возникла и развивалась параллельно с теорией вероятностей.

Данный практикум адресуется студентам всех специальностей НИТУ «МИСиС». Практикум подготовлен в соответствии с требованиями государственных образовательных стандартов. Рассматриваются следующие разделы курса: основные выборочные характеристики, точечные и интервальные оценки, проверка статистических гипотез, регрессионный и корреляционный анализ.

Согласно учебному плану предусмотрено выполнение трех типовых расчетов (лабораторных работ). Исходя из этого сформирована структура практикума, состоящая из двух разделов. Первый содержит теоретический материал и методические указания по решению задач для подготовки к практическим занятиям. Приведены необходимые для решения задач понятия и формулы, рассмотрено большое количество примеров решения типовых задач, охватывающих основные темы курса математической статистики.

Во второй части практикума разбираются все три типовые расчеты (лабораторные работы), приводятся подробные примеры их выполнения. Затем даны типовые варианты контрольной работы с ответами. В приложении представлены статистические таблицы, необходимые для решения данных задач.

Студентам, желающим ознакомиться с более полным и строгим изложением основ математической статистики, рекомендуем литературу, приведенную в библиографическом списке.

1. ПРАКТИЧЕСКИЕ ЗАНЯТИЯ ПО МАТЕМАТИЧЕСКОЙ СТАТИСТИКЕ

1.1. Первичная обработка результатов эксперимента и оценка основных параметров генеральной совокупности

Математическая статистика позволяет с помощью математических методов обрабатывать, систематизировать и использовать численные результаты эксперимента для получения практических выводов.

Под *генеральной совокупностью* в математической статистике понимается множество (гипотетическое) всех возможных результатов измерения некоторой величины, которые могут быть получены в данных условиях. Тем же самым понятием в теории вероятностей является случайная величина X.

Реальная серия повторных измерений этой величины x_1, x_2,..., x_n трактуется как случайная выборка из генеральной совокупности, или просто *случайная выборка*. Число n называется *объемом* случайной выборки. Приведем примеры.

1. Проведена серия повторных измерений одной и той же физической величины в одних и тех же условиях. Разброс результатов обусловлен погрешностью измерительной аппаратуры.

2. Измеряется некоторая характеристика одинаковых изделий, изготовленных при поточном производстве. Разброс результатов обусловлен особенностями технологии производства.

3. Измеряется некоторая характеристика людей определенного пола и интервала возрастов, например рост. Разброс результатов обусловлен природными факторами.

В статистике принята следующая математическая модель подобных экспериментов. *Каждый элемент случайной выборки рассматривается как отдельная случайная величина.* Относительно этих случайных величин, которые в дальнейшем будем обозначать прописными буквами, известна некоторая априорная информация.

Случайная выборка называется *повторной*, если все входящие в нее случайные величины $X_1, X_2, ..., X_n$ взаимно независимы и имеют одинаковую функцию распределения $F(x)$, причем ту же, что и наблюдаемая случайная величина X. На практике это, в частности, означает, что измерения производятся *независимо* друг от друга (полученные результаты одних измерений не влияют на возможные результаты других). Величины имеют одинаковые математические ожидания $M(X_i)=a$, т.е., результаты измерений свободны от *систематических ошибок* (результаты в среднем не смещены относительно истинного значения $M(X)=a$), и одинаковые дисперсии $D(X_i)=D(X)=\sigma^2$, что называется *равноточностью* измерений (например, измерения физической величины проведены на одном и том же приборе при одинаковых условиях).

Если случайные величины $X_1, X_2, ..., X_n$, образующие повторную случайную выборку, имеют нормальное распределение с одинаковыми параметрами a, σ, т.е., $X_i \sim N(a; \sigma)$, то такая выборка называется *нормальной*, а соответствующая генеральная совокупность – *нормальной генеральной совокупностью.*

В математической статистике рассматривают и *неповторные* выборки, в которых нарушается хотя бы одно из указанных условий: взаимная независимость, одинаковость функции распределения. Слово «повторная» обычно опускается, и пишут просто «выборка». Для неповторной выборки обязательно пишут «неповторная выборка».

Распределение случайной величины X характеризуется рядом *параметров* (математическое ожидание, дисперсия и т.д.). Эти параметры называют *параметрами генеральной совокупности.* Важной задачей математической статистики является нахождение по случайной выборке приближенных значений каждого из параметров, называемых *точечными оценками параметров*, или просто *оценками*. Таким образом, *оценкой параметра* β называется функция $f(X_1, X_2, ..., X_n)$ от случайной выборки, значение которой принимается в качестве приближенного для данного параметра и обозначается $\tilde{\beta}$:

$$\beta \approx \tilde{\beta} = f(X_1, X_2, ..., X_n). \qquad (1.1)$$

Так как оценка зависит от случайной выборки, то она, в свою очередь, является случайной величиной. Для одного и того же параметра β по одной и той же выборке можно построить много различных оценок. Для сравнения оценок между собой введены специальные характеристики.

Оценка называется *несмещенной*, если ее математическое ожидание равно истинному значению параметра, т.е. $M(\tilde{\beta})=\beta$. Несмещенная оценка обеспечивает близость в среднем значений оценки к значению оцениваемого параметра, т.е. не дает систематической ошибки.

Оценка называется *состоятельной*, если при $n \to \infty$ она сходится по вероятности к истинному значению оцениваемого параметра, т.е.

$$\forall \varepsilon > 0 \ \lim_{n \to \infty} P(\left|\tilde{\beta}_n - \beta\right| > \varepsilon) = 0, \qquad (1.2)$$

где $\tilde{\beta}_n$ – оценка параметра β, найденная по выборке объема n.

Смысл понятия состоятельности заключается в том, что с увеличением объема выборки оценка стремится к истинному значению параметра.

Точностью оценки $\tilde{\beta}$ называется средний квадрат отклонения оценки от β: $q^2(\tilde{\beta}) = M[(\tilde{\beta} - \beta)^2]$. Для несмещенных оценок точность определяется величиной дисперсии оценки: $q^2(\tilde{\beta}) = D(\tilde{\beta})$. Чем меньше $q = \sqrt{q^2(\tilde{\beta})}$, тем оценка лучше (точнее). *Наилучшей линейной оценкой* параметра β называется такая его линейная несмещенная оценка, которая имеет наименьшую дисперсию в классе всех линейных несмещенных оценок.

Пусть задана повторная случайная выборка $X_1, X_2, \ldots, X_n$. За *оценку математического ожидания* a принимается среднее арифметическое элементов выборки:

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i . \qquad (1.3)$$

Оценкой дисперсии σ^2 при *известном математическом ожидании* a является величина S_0^2:

$$S_0^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - a)^2 . \qquad (1.4)$$

Оценкой дисперсии σ^2 при *неизвестном математическом ожидании* является величина S^2, которую называют *эмпирической дисперсией*:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 . \qquad (1.5)$$

Оценкой среднего квадратического отклонения σ при этом являются, соответственно, величины

$$S_0 = \sqrt{S_0^2}\ ; \quad S = \sqrt{S^2}\ . \tag{1.6}$$

Оценки математического ожидания и дисперсии, найденные по формулам (1.3) – (1.5), являются несмещенными и состоятельными. Среднее арифметическое (1.3) является наилучшей линейной оценкой математического ожидания для повторной случайной выборки.

Оценка параметра $\sigma = \sqrt{\sigma^2}$ с помощью значений (1.6) является состоятельной, но смещенной (ее смещение убывает с увеличением *n*). Число $k = n - 1$ в формуле (1.5) называется *числом степеней свободы* оценки S^2.

Для практических расчетов формулу (1.5) удобно преобразовать к виду

$$S^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - n\overline{X}^2\right). \tag{1.7}$$

Вычисление среднего значения $\overline{X}$ и оценки дисперсии S^2 упрощается, если отсчет значений X_i вести от подходящим образом выбранных начала отсчета *C* и масштабного коэффициента *h*, т.е. если сделать линейную замену:

$$X_i = C + hU_i \ \ (i = 1, 2, ..., n). \tag{1.8}$$

При такой замене формулы (1.3) и (1.5) – (1.6) принимают вид

$$\overline{X} = C + h\overline{U}, \qquad \overline{U} = \frac{1}{n}\sum_{i=1}^{n} U_i\ ; \tag{1.9}$$

$$S^2 = \frac{h^2}{n-1}\sum_{i=1}^{n}\left(U_i - \overline{U}\right)^2 = \frac{h^2}{n-1}\left(\sum_{i=1}^{n} U_i^2 - n\overline{U}^2\right). \tag{1.10}$$

Для контроля вычислений весь расчет повторяют с другим началом отсчета *C,* результаты должны совпадать с точностью до возможных ошибок округления.

Задача 1.1. В табл. 1.1 в первом столбце записаны результаты $n = 18$ независимых равноточных измерений заряда электрона $q = x \cdot 10^{-10}$, полученных Милликеном. Вычислить оценки математического ожидания, дисперсии и среднего квадратического отклонения величины *X*, провести контроль расчетов.

Решение

Выбираем $C = 4{,}780$ и, полагая $h = 10^{-3}$, подсчитываем значения $u_i = (x_i - C)/h = (x_i - 4{,}780)/10^{-3}$ и u_i^2. Суммы чисел второго и третьего столбцов дают возможность рассчитать $\overline{X}$ и S^2:

$$\overline{U} = 13/18 = 0{,}72, \quad \overline{X} = 4{,}780 + 0{,}72 \cdot 10^{-3} = 4{,}7807;$$

$$S^2 = 10^{-6}\left(6239 - 13^2/18\right)/17 = 3{,}66 \cdot 10^{-4},$$

откуда $S = \sqrt{S^2} = \sqrt{3{,}66 \cdot 10^{-4}} = 1{,}91 \cdot 10^{-2}$.

В последних двух столбцах приведены расчеты при другом начале отсчета $C_1 = 4{,}790$, т.е. при замене $V_i = (X_i - 4{,}790)/10^{-3}$. Эти расчеты дают те же значения $\overline{X}$ и S:

$$\overline{V} = 167/18 = -9{,}2, \quad \overline{X} = 4{,}790 - 9{,}28 \cdot 10^{-3} = 4{,}7809;$$

$$S^2 = 10^{-6}\left(7779 - 167^2/18\right)/17 = 3{,}66 \cdot 10^{-4}.$$

Таблица 1.1

Исходные данные и результаты расчета к задаче 1.1

Исходные данные X	Расчет		Контроль	
	U	U^2	V	V^2
4,761	–19	361	–29	841
4,792	12	144	2	4
4,758	–22	484	–32	1024
4,764	–16	256	–26	676
4,810	30	900	20	400
4,799	19	361	9	81
4,797	17	289	7	49
4,790	10	100	0	0
4,747	–33	1089	–43	1849
4,769	–11	121	–21	441
4,806	26	676	16	256
4,779	–1	1	–11	121
4,785	5	25	–5	25
4,790	10	100	0	0
4,777	–3	9	–13	169
4,749	–31	961	–41	1686
4,781	1	1	–9	81
4,799	19	361	9	81
Σ	13	6239	–167	7779

При большом числе исходных данных их предварительно группируют, т.е. весь диапазон значений X разбивают на l равных *интервалов*, подсчитывают число исходных данных, попавших в каждый j-й интервал, и относят это число (частоту m_j) к середине интервала X_j ($j = 1, 2, ..., l$). Затем середины этих интервалов кодируют по формуле (1.8), выбирая за новое начало отсчета C середину одного из интервалов, а за масштабный коэффициент h – длину интервала. При таком кодировании все значения U_j будут целыми числами, которые для соседних интервалов отличаются на единицу. Расчет по сгруппированным данным дает лишь приближенные значения оценки математического ожидания $\overline{X} \approx C + h\overline{U}$ и оценки дисперсии $S^2 \approx h^2 S_u^2$, где

$$\overline{U} = \frac{1}{n}\sum_{j=1}^{l} U_j m_j; \tag{1.11}$$

$$S_u^2 = \frac{1}{n-1}\sum_{j=1}^{l}(U_j - \overline{U})^2 m_j = \left(\sum_{j=1}^{l} U_j^2 m_j - n\overline{U}^2\right)\Big/(n-1). \tag{1.12}$$

Оценка дисперсия S^2, вычисленная по сгруппированным данным, оказывается меньше оценки дисперсии, найденной по несгруппированным результатам эксперимента на величину, приблизительно равную $h^2/12$. Это следует учитывать при округлении значения S, сохраняя лишь один сомнительный знак; значение среднего $\overline{X}$ округляют при этом до единиц того разряда, который сохранен в значении S. Подобными соображениями можно руководствоваться и в тех случаях, когда результаты измерений округлены с учетом цены деления шкалы измерительного прибора (при этом обычными методами через цену деления шкалы оценивают погрешность вычисления S^2).

Задача 1.2. Проведено 52 эксперимента, их результаты оказались в диапазоне 22,75–26,75. Этот диапазон разбит на 8 интервалов длины $h = 0{,}5$, и сгруппированные данные эксперимента приведены в первых столбцах табл. 1.2. Вычислить оценки математического ожидания, дисперсии и среднего квадратического отклонения, провести контроль расчетов.

Решение

Для расчета выбрано начало отсчета $C = 24{,}5$, а для контроля $C_1 = 25{,}0$. Кодировка имеет соответственно вид $U = (X - 24{,}5)/0{,}5$; $V = (X - 25{,}0)/0{,}5$.

Таблица 1.2

Исходные данные и результаты расчета к задаче 1.2

Интервал	Середина X	m	Расчет			Контроль		
			U	mU	mU^2	V	mV	mV^2
22,75–23,25	23,0	2	–3	–6	18	–4	–8	32
23,25–23,75	23,5	4	–2	–8	16	–3	–12	36
23,75–24,25	24,0	8	–1	–8	8	–2	–16	32
24,25–24,75	24,5	11	0	0	0	–1	–11	11
24,75–25,25	25,0	16	1	16	16	0	0	0
25,25–25,75	25,5	7	2	14	28	1	7	7
25,75–26,25	26,0	3	3	9	27	2	6	12
26,25–26,75	26,5	1	4	4	16	3	3	9
Σ	–	52	–	21	129	–	–31	139

Суммирование проводится только в столбцах, где учитывается частота m, т.е. в столбцах m, mU, mU^2 (mV, mV^2). По данным таблицы расчета получаем: $\overline{U} = 21/52 = 0,404$, $\overline{X} = 24,5 + 0,5 \cdot 0,404 = 24,702$; $S_U^2 = \left(129 - \left(21^2\right)/52\right)/51 = 2,363$, $S = 0,5\sqrt{2,363} = 0,769$. Контроль дает те же значения $\overline{X}$ и S_V^2, а значит, и S_V:

$$V = -31/52 = -0,596,\ \overline{X} = 25,0 + 0,5(-0,596) = 24,702;$$
$$S_V^2 = (139 - (-31)^2/52)/51 = 2,363,\ S_V = 0,769.$$

Погрешность вычисления S^2 приблизительно равна $(-0,5^2)/12$, что составляет примерно 3,5 %, погрешность вычисления $S = \sqrt{S^2}$ примерно 2 %, т.е. 0,015. Поэтому значения S и $\overline{x}$ округляем до двух десятичных знаков, что дает $\overline{X} = 24,70$, $S = 0,77$.

1.2. Оценка математического ожидания по неравноточным измерениям

Часто встречающимся на практике случаем неповторной выборки является выборка, в которой случайные величины $X_1, X_2, ..., X_n$ взаимно независимы, имеют одинаковые математические ожидания, но различные дисперсии. Такие измерения называют *неравноточными.* Как правило, дисперсии каждой величины X_i неизвестны, но известны отношения дисперсий. Числа, обратно пропорциональные дисперсиям, называют *весами измерений* и обозначают w_i:

$$D(X_1):D(X_2):\ldots:D(X_n)=(1/w_1):(1/w_2):\ldots:(1/w_n)$$

или

$$D(X_i)=\sigma^2/w_i, \quad (i=1,2,\ldots,n). \tag{1.13}$$

Коэффициент σ^2 в формуле (1.13) обычно неизвестен, он называется *дисперсией измерения с единичным весом*; веса w_i, как правило, известны.

Среднее арифметическое (1.3) для неравноточных измерений является несмещенной и состоятельной оценкой математического ожидания, но не является наилучшей линейной оценкой. Наилучшей линейной оценкой в этом случае будет *среднее взвешенное*:

$$\overline{X}_{\text{взв}}=\frac{\left(\sum_{i=1}^{n}X_iw_i\right)}{\sum_{i=1}^{n}w_i}, \tag{1.14}$$

эта оценка будет несмещенной и состоятельной. Она и используется на практике для неравноточных измерений.

Задача 1.3. В табл. 1.3 в первом столбце записаны результаты $n=5$ независимых случайных величин X_i, являющихся средними арифметическими пяти серий измерений. В каждой серии измерения независимы, имеют одинаковые математические ожидания и дисперсии, т.е. равноточны $D(X_{ij})=\sigma^2$, где i – номер серии; j – номер измерения в серии; n_i – число измерений в серии. Найти наилучшую оценку математического ожидания.

Таблица 1.3

Исходные данные и результаты расчета к задаче 1.3

i	X_i	n_i	X_in_i
1	2,41	5	12,05
2	2,83	2	5,66
3	2,62	4	10,48
4	2,49	6	14,94
5	2,75	3	8,25
$\sum$	13,10	20	51,38

Решение

Среднее арифметическое результатов X_i будет $\overline{X} = \frac{1}{5}\sum_{i=1}^{5} X_i$ = 13,10/5 = 2,62. Получена несмещенная оценка математического ожидания, но она не является наилучшей линейной оценкой, так как результаты измерений неравноточны. Среднее арифметическое по каждой серии $X_i = \sum_{j=1}^{n_i} X_{ij} / n_i$. Если измерения X_{ij} равноточны, то дисперсии средних арифметических равны $D(X_i) = \sigma^2 / n_i$ и величины x_i неравноточны. Сравнивая $D(X_i)$ с формулой (1.13), делаем вывод, что весами измерений в этом случае являются числа измерений n_i, т.е. $w_i = n_i$. Используя формулу (1.14), получаем $\overline{X}_{\text{взв}} = 51{,}38$ / 20 = 2,569. Это значение и будет наилучшей линейной оценкой, т.е., имеющей наименьшую погрешность.

1.3. Оценка дисперсии по нескольким сериям экспериментов

Пусть заданы L независимых повторных выборок – L серий измерений. Случайные величины различных выборок имеют в общем случае разные математические ожидания, но дисперсии всех величин во всех выборках одинаковы. Такая ситуация возникает, когда одним и тем же прибором производят измерения различных величин (например, измерения значений функции для различных значений аргумента).

В этом случае для оценки единой дисперсии можно использовать значения измерений всех серий. По каждой выборке находят эмпирическую дисперсию S_j^2 с числом степеней свободы k_j (j = 1, 2, ..., L). В качестве оценки единой дисперсии принимают сводную эмпирическую дисперсию

$$S_{\text{св}}^2 = \left(\sum_{j=1}^{L} S_j^2 k_j \right) \Big/ \sum_{j=1}^{L} k_j \qquad (1.15)$$

с числом степеней свободы $k_{\text{св}} = \sum_{j=1}^{L} k_j$. Сводная оценка дисперсии (1.15) является несмещенной, она более точная, чем каждая из эмпирических дисперсий S_j^2 .

Задача 1.4. При изучении зависимости *y* от *x* измерения производились независимо и предположительно с одинаковой точностью (табл. 1.4). Оценить эту точность, т.е. найти оценку дисперсии и оценку среднего квадратического отклонения при измерении величины *y*. Число измерений при разных значениях различно.

Таблица 1.4

Экспериментальные данные к задаче 1.4

x	*y*				
20	48,9	48,8	48,7	49,0	49,2
60	46,1	46,2	46,6	46,4	
110	43,8	44,0	44,2		
120	43,6	44,0	43,7	43,8	43,8
160	42,0	42,4	42,2		
200	41,2	41,3	41,4	41,1	41,6

Решение

Вначале вычислим эмпирические дисперсии результатов каждой серии измерений, т.е. при каждом значении *x*. Прежде всего заметим, что приведенные в табл. 1.4 значения *y* удобно уменьшить на 40, полученные данные обозначим через *Y* (табл. 1.5). Чтобы вести расчет с небольшими целыми числами, закодируем значения *Y* по формуле $U = 10(Y - C)$, где за начало отсчета *C* в каждой серии примем число, набранно полужирным курсивом в табл. 1.5, например при *x* = 20 примем *C* = 8,9, а при *x* = 60 примем *C* = 6,2.

Таблица 1.5

Результаты расчета к задаче 1.4

x	*Y*					*U* = 10(*Y* – C)					ΣU	ΣU^2	*n*	$kS^2 \cdot 10^2$	*k*	*S*
20	***8,9***	8,8	8,7	9,0	9,2	0	–1	–2	1	3	1	15	5	14,8	4	0,19
60	6,1	***6,2***	6,6	6,4		–1	0	4	2		5	21	4	14,75	3	0,22
110	3,8	***4,0***	4,2			–2	0	2			0	8	3	8	2	0,20
120	3,6	4,0	3,7	***3,8***	***3,8***	–2	2	–1	0	0	–1	9	5	8,8	4	0,15
160	2,0	2,4	***2,2***			–2	2	0			0	8	3	8	2	0,20
200	1,2	***1,3***	1,4	1,1	1,6	–1	0	1	–2	3	1	15	5	14,8	4	0,19
Σ											–	–	–	69,15	19	

В этой таблице все расчеты ведутся по строкам. $kS^2 \cdot 10^2 = kS_U^2 = = \sum u^2 - \left(\sum u\right)^2 \Big/ n$, где n – число замеров в серии. Суммы, подсчитанные в столбцах $kS^2 \cdot 10^2$ и k, позволяют получить оценки дисперсии и среднего квадратического отклонения (формула (1.15)):

$$S_{\text{св}}^2 = 0{,}1^2 \cdot 69{,}15/19 = 0{,}0364; \qquad S_{\text{св}} = 0{,}1\sqrt{3{,}64} = 0{,}191\,.$$

1.4. Построение гистограммы распределения

Для наглядного представления о выборке часто используют график, называемый *гистограммой*. Для построения гистограммы интервал, содержащий все элементы выборки, разбивают на l непересекающихся интервалов (как правило, равной длины). Подсчитывают числа n_i попаданий результатов экспериментов в каждый i-й интервал и строят столбиковую диаграмму, откладывая по оси ординат значения средней плотности $n_i/(nh_i)$, где h_i – длина i-го интервала. Площадь каждого столбика равна n_i/n, что соответствует относительной частоте попадания элементов выборки в i-й интервал. Площадь под всей ступенчатой фигурой равна единице.

При увеличении объема выборки и уменьшении интервалов группировки гистограмма приближается к функции плотности генеральной совокупности. Гистограмма является эмпирической функцией плотности, она дает приближенную функцию плотности генеральной совокупности (ее оценку) по случайной выборке.

Задача 1.5. Построить гистограмму для выборки, представленной первыми двумя строками табл. 1.6.

Таблица 1.6

Исходные данные и результаты расчета к задаче 1.5

Интервалы	10–20	20–30	30–40	40–50	50–60	60–70	70–80	80–90
Частоты n_i	1	4	6	13	10	8	6	2
n_i/nh_i	0,002	0,008	0,012	0,026	0,020	0,016	0,012	0,004

Решение

Согласно условию задачи $n = \sum_{i=1}^{8} n_i = 50$; $h_i = 10$; $i = 1,...,8$. Значения средней плотности $n_i\,/\,(nh_i)$, необходимые для построения гистограммы, рассчитаны в последней строке табл. 1.6. Гистограмма представлена на рис. 1.1.

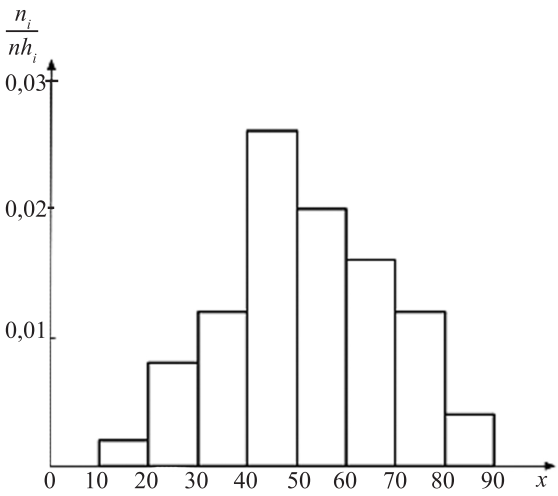

Рис. 1.1. Гистограмма

1.5. Некоторые используемые в статистике распределения

В статистике широко используются следующие законы распределения, связанные с обработкой экспериментов из нормальной генеральной совокупности:

1) стандартное нормальное распределение;
2) распределение Пирсона;
3) распределение Стьюдента;
4) распределение Фишера.

Стандартное нормальное распределение рассмотрено в практикуме [1]. Если случайная величина U имеет стандартное нормальное распределение, вероятность ее попадания в интервал (t_1, t_2) вычисляется по формуле

$$P(t_1 < U < t_2) = \Phi(t_2) - \Phi(t_1), \tag{1.16}$$

где $\Phi(x)$ – значение интеграла вероятностей. Если случайная величина X имеет нормальное распределение с параметрами a, σ ($X \sim N(a; \sigma)$), то вероятность ее попадания в интервал (x_1, x_2) вычисляется по формуле

$$P(x_1 < X < x_2) = \Phi\left(\frac{x_2 - a}{\sigma}\right) - \Phi\left(\frac{x_1 - a}{\sigma}\right). \tag{1.17}$$

Квантили стандартного нормального распределения приведены в табл. П1 приложения.

Распределением Пирсона с k степенями свободы называется распределение суммы квадратов k взаимно независимых случайных величин, каждая из которых имеет стандартное нормальное распределение. Можно также говорить, что это распределение квадрата длины случайного вектора, имеющего k координат, каждая из которых имеет стандартное нормальное распределение U_i. Случайная величина, имеющая распределение Пирсона, часто обозначается $\chi^2(k)$:

$$\chi^2(k) = U_1^2 + U_2^2 + ... + U_k^2$$

и само распределение называется χ^2-распределением.

Математическое ожидание и дисперсия величины $\chi^2(k)$ равны соответственно:

$$M(\chi^2) = k\,; \quad D(\chi^2) = 2k\,.$$

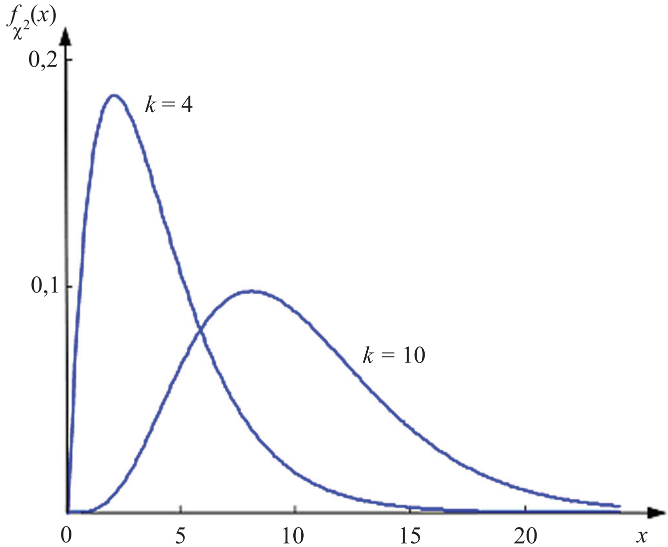

Рис. 1.2. График функции плотности вероятностей случайной величины, имеющей распределение Пирсона (χ^2) с k степенями свободы

График функции плотности вероятностей случайной величины, имеющей распределение Пирсона, представлен на рис. 1.2. Таблица квантилей $\chi_P^2(k)$ распределения Пирсона приведена в табл. П3 приложения.

Распределением Стьюдента (или ***t-распределением***) ***с k степенями свободы*** называется распределение отношения

$$T(k) = \frac{U}{\sqrt{(U_1^2 + U_2^2 + ... + U_k^2)/k}} = \frac{U}{\sqrt{\chi^2(k)/k}}, \qquad (1.18)$$

где все случайные величины U, U_1, $U_2, ...,$ U_k – взаимно независимы и имеют стандартное нормальное распределение. Распределение Стьюдента симметрично относительно центра $M\left(T(k)\right) = 0$. Кривая распределения Стьюдента (рис. 1.3) внешне похожа на кривую стандартного нормального распределения ([1], рис. 1.14).

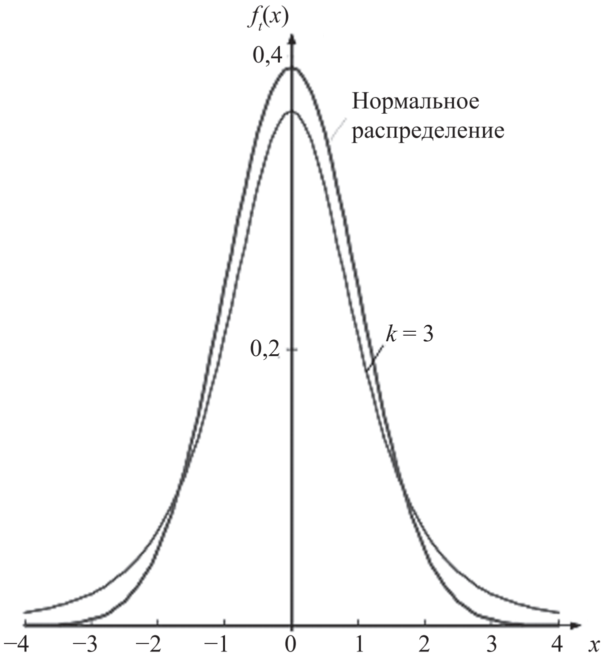

Рис. 1.3. Графики функций плотности вероятностей распределения случайных величин, имеющих стандартное нормальное распределение и распределение Стьюдента (t) с k степенями свободы

При $k \to \infty$ распределение Стьюдента стремится к стандартному нормальному распределению. Но при малых значениях k числовые значения функций плотности распределения вероятностей существенно различаются, что особенно сказывается на квантилях $t_P(k)$ при вероятностях P, близких к 0 и к 1. Квантили t-распределения Стьюдента $t_P(k)$ приведены в приложении (табл. П2).

Распределением Фишера (или ***F-распределением***) ***с k_1 и k_2 степенями свободы*** называется распределение отношения

$$F = \frac{\chi^2(k_1)/k_1}{\chi^2(k_2)/k_2}.$$

Кривая распределения Фишера внешне похожа на кривую распределения Пирсона (см. рис. 1.2). Квантили F-распределения Фишера $F_P(k_1, k_2)$ приведены в приложении (табл. П4).

1.6. Построение доверительных интервалов

В разделе 1.1 были рассмотрены точечные оценки параметров Напомним, что точечная оценка $\tilde{\beta}$ неизвестного параметра β для каждой случайной выборки дает лишь одно числовое значение, которое мы принимаем за приближенное значение этого параметра. Важная задача – определить точность полученного приближения, т.е. насколько точечная оценка может отклоняться от истинного значения параметра. Ответить на этот вопрос позволяют доверительные интервалы.

Доверительным интервалом параметра β называется интервал со случайными границами $(\tilde{\beta} - \varepsilon_1;\ \tilde{\beta} + \varepsilon_2)$, который накрывает истинное значение параметра β с заданной вероятностью P, которая называется *доверительной вероятностью.* Величина $\alpha = 1 - P$ называется *уровнем значимости*. При этом обычно требуют, чтобы вероятности выхода за границы доверительного интервала в обе стороны были равны между собой, а именно:

$$P(\beta < \tilde{\beta} - \varepsilon_1) = P(\beta > \tilde{\beta} + \varepsilon_2) = (1 - P)/2 = \alpha/2.$$

Это дополнительное требование обеспечивает единственность решения задачи.

Пусть задана повторная случайная выборка $X_1, X_2, ..., X_n$ из нормальной генеральной совокупности. Это означает, что результаты эксперимента независимы и подчиняются нормальному закону распределения с одинаковыми параметрами $X_i \sim N(a;\sigma)$. Доверительные интервалы для параметров нормального распределения находят следующим образом.

1.6.1. Доверительный интервал для математического ожидания

С вероятностью P математическое ожидание a принадлежит интервалу

$$a \in (\overline{X} - \varepsilon;\ \overline{X} + \varepsilon)\,, \tag{1.19}$$

– если σ известно, то

$$\varepsilon = u_{1-\alpha/2}\, \sigma \big/ \sqrt{n}\,, \tag{1.20}$$

– если σ не известно, то

$$\varepsilon = t_{1-\alpha/2}(k)\, S \big/ \sqrt{n}\,, \tag{1.21}$$

где $\overline{X}$ – оценка математического ожидания (1.3); $S = \sqrt{S^2}$ – оценка среднего квадратического отклонения σ (1.6) при неизвестном математическом ожидании; $u_{1-\alpha/2}$ – квантиль стандартного нормального распределения; $t_{1-\alpha/2}(k)$ – квантиль распределения Стьюдента с k степенями свободы; n – объем выборки; k – число степеней свободы при вычислении оценки S.

Часто доверительный интервал для математического ожидания записывают символически:

$$a = \overline{X} \pm \varepsilon\,. \tag{1.22}$$

1.6.2. Доверительный интервал для среднего квадратического отклонения

Доверительный интервал для среднего квадратического отклонения σ при доверительной вероятности $P = 1 - \alpha$ имеет вид:

– если математическое ожидание a известно, то

$$S_0 \sqrt{\frac{n}{\chi^2_{1-\alpha/2}(n)}} < \sigma < S_0 \sqrt{\frac{n}{\chi^2_{\alpha/2}(n)}}\,, \tag{1.23}$$

где S_0 – оценка среднего квадратического отклонения σ (1.6) при известном математическом ожидании;

– если математическое ожидание a неизвестно, то

$$S\sqrt{\frac{k}{\chi^2_{1-\alpha/2}(k)}} < \sigma < S\sqrt{\frac{k}{\chi^2_{\alpha/2}(k)}}, \tag{1.24}$$

где S – оценка среднего квадратического отклонения σ (1.6) при неизвестном математическом ожидании; $\chi^2_{1-\alpha/2}(k)$, $\chi^2_{\alpha/2}(k)$ – квантили распределения Пирсона с k степенями свободы; k – число степеней свободы оценки S.

1.6.3. Доверительный интервал для дисперсии

Доверительный интервал для дисперсии σ^2 при доверительной вероятности $P = 1-\alpha$, соответственно, находится по формулам:

– если математическое ожидание a известно, то

$$S_0^2 \frac{n}{\chi^2_{1-\alpha/2}(n)} < \sigma^2 < S_0^2 \frac{n}{\chi^2_{\alpha/2}(n)}, \tag{1.25}$$

– если математическое ожидание a неизвестно, то

$$S^2 \frac{k}{\chi^2_{1-\alpha/2}(k)} < \sigma^2 < S^2 \frac{k}{\chi^2_{\alpha/2}(k)} \tag{1.26}$$

Задача 1.6. Случайная выборка из нормальной генеральной совокупности состоит из одного элемента $X = 12{,}7$. Значение параметра σ известно, $\sigma = 0{,}3$. Найти доверительный интервал для математического ожидания с доверительной вероятностью $P = 0{,}95$.

Решение

По формуле (1.20) $\varepsilon = u_{1-\alpha/2}\sigma$, где $u_{1-\alpha/2}$ – квантиль стандартного нормального распределения, $\alpha = 1 - P = 0{,}05$; $1-\alpha/2 = 0{,}975$; $u_{0,975} = 1{,}96$. $\varepsilon = 1{,}96 \cdot 0{,}3 \approx 0.6$. По формуле (1.22) $a = 12{,}7 \pm 0{,}6$.

Замечание. Задача демонстрирует смысл коэффициента σ. Полуширина доверительного интервала для одного измерения с доверительной вероятностью $P = 0{,}95$ приблизительно равна 2σ (правило двух сигм).

Задача 1.7. Случайная выборка из нормальной генеральной совокупности состоит из 16 элементов. По ним найдено среднее арифметическое $\overline{X} = 12{,}56$ (1.3). Значение параметра σ известно, $\sigma = 0{,}3$. Найти доверительный интервал для математического ожидания с доверительной вероятностью $P = 0{,}95$.

Решение

По формуле (1.20) $\varepsilon = u_{1-\alpha/2}\,\sigma\big/\sqrt{n}$, где $u_{1-\alpha/2} = u_{0,975} = 1{,}96$. $\varepsilon = 1{,}96 \cdot 0{,}3/\sqrt{16} \approx 0{,}15$. По формуле (1.22) находим $a = 12{,}56 \pm 0{,}15$.

Задача 1.8. В задаче 1.1 для $n = 18$ результатов независимых измерений величины заряда электрона $q = x \cdot 10^{-10}$ были вычислены $\overline{X} = 4{,}7807$ и $S = 0{,}0191$. Предполагая, что величины ошибок измерений подчиняются нормальному закону распределения с нулевым математическим ожиданием (т.е. что систематические ошибки отсутствуют), построить доверительные интервалы для математического ожидания (истинного значения) заряда электрона $q_{\text{ист}} = a \cdot 10^{-10}$ и среднего квадратического отклонения с доверительной вероятностью $P = 0{,}95$.

Решение

По формуле (1.21) $\varepsilon = t_{1-\alpha/2}(k)\,S\big/\sqrt{n}$. Из табл. П2 приложения находим $t_{1-\alpha/2}(k) = t_{0,975}(17) = 2{,}11$.

$$\varepsilon = 2{,}11 \cdot 0{,}0191\big/\sqrt{18} \approx 0{,}0095 .$$

По формуле (1.22) $a = 4{,}7807 \pm 0{,}0095$, т.е. с вероятностью $P = 0{,}95$ выполняется неравенство $4{,}7712 < a < 4{,}7902$.

Из табл. П3 приложения находим

$$\chi^2_{\alpha/2}(k) = \chi^2_{0,025}(17) = 7{,}56;$$

$$\chi^2_{1-\alpha/2}(k) = \chi^2_{0,975}(17) = 30{,}2.$$

По формуле (1.24) получаем

$$0{,}0191\sqrt{\frac{17}{30{,}2}} < \sigma < 0{,}0191\sqrt{\frac{17}{7{,}56}},$$

откуда

$$0{,}0143 < \sigma < 0{,}0287,$$

т.е. среднее квадратическое отклонение заключено между 0,0143 и 0,0287 с вероятностью 0,95.

Задача 1.9. В задаче 1.2 для $n = 52$ измерений были вычислены приближенно $\overline{X} = 24{,}70$ и $S = 0{,}77$. Предполагая, что величины ошибок измерения подчиняются нормальному закону распределения с нулевым математическим ожиданием, построить доверительные интервалы для математического ожидания a и дисперсии σ^2 с доверительной вероятностью $P = 0{,}99$.

Решение

Из табл. П2 приложения находим (интерполируя): $t_{1-\alpha/2}(k) = t_{0,995}(51) = 2{,}68$. По формулам (1.21) и (1.22) получаем $\varepsilon = 2{,}68 \cdot 0{,}77 \big/ \sqrt{52} = 0{,}29$, откуда $a = 24{,}70 \pm 0{,}29$ или $24{,}41 < a < 24{,}99$.

Из табл. П3 приложения находим (интерполируя): $\chi^2_{\alpha/2}(k) = \chi^2_{0,005}(51) = 28{,}8$; $\chi^2_{1-\alpha/2}(k) = \chi^2_{0,995}(51) = 80{,}7$. Из формулы (1.26) следует $0{,}77^2 \cdot 51/80{,}7 < \sigma^2 < 0{,}77^2 \cdot 51/28{,}8$, откуда $0{,}375 < \sigma^2 < 1{,}051$. Заметим, что округлять границы следует в сторону увеличения доверительного интервала, чтобы заведомо обеспечить требуемую надежность.

Задача 1.10. Для оценки погрешности нового измерительного прибора на нем была проведена серия повторных измерений эталонной величины (т.е. математическое ожидание было известно). Число измерений $n = 40$. По формуле (1.4) была найдена оценка дисперсии $S_0^2 = 0{,}0216$. Предполагая, что измерения имеют нормальный закон распределения, найти доверительный интервал для σ с доверительной вероятностью $P = 0{,}95$.

Решение

Из табл. П3 приложения при числе степеней свободы $n = 40$ находим $\chi^2_{\alpha/2}(n) = \chi^2_{0,025}(40) = 24{,}4$; $\chi^2_{1-\alpha/2}(n) = \chi^2_{0,975}(40) = 59{,}3$. По формуле (1.24) находим $S_0 = \sqrt{0{,}0216} \approx 0{,}147$; $0{,}147 \cdot \sqrt{40 / 59{,}3} < \sigma < 0{,}147 \times \times \sqrt{40 / 24{,}4}$. Окончательно,

$$0{,}121 < \sigma < 0{,}188.$$

Задача 1.11. В задаче 1.4 в первой строке табл. 1.5 по $n = 5$ измерениям была вычислена оценка σ, равная $S_1 = 0{,}19$. Найти оценку математического ожидания Y_1 по этой строке. Построить доверительные интервалы для $M(y_1)$ и σ с доверительной вероятностью $P = 0{,}95$.

Решение

Используя формулу (1.7), находим оценку математического ожидания y_1 при $x = 20$: $\overline{Y_1} = C + h\overline{U}$ = 8,9 + 0,1·1/5 = 8,92. Предполагая, как и выше, что результаты измерения Y_{1i} подчиняются нормальному закону распределения, находим $t_{1-\alpha/2}(k)$ = $t_{0,975}(4) = 2{,}78$. По формулам (1.21) и (1.22) получаем $\varepsilon = 2{,}78 \cdot 0{,}19 \big/ \sqrt{5} = 0{,}24$; $M(y_1) = 8{,}92 \pm 0{,}24$. Из табл. П3 приложения находим: $\chi^2_{\alpha/2}(k)$ = $\chi^2_{0,025}(4) = 0{,}484$; $\chi^2_{1-\alpha/2}(k)$ = $= \chi^2_{0,975}(4) = 11{,}1$. Согласно формуле (1.24): $0{,}19 \cdot \sqrt{4/11{,}1} < \sigma < 0{,}19 \times \sqrt{4/0{,}484}$, откуда $0{,}114 < \sigma < 0{,}545$.

Задача 1.12. Решить задачу 1.11, используя для построения доверительных интервалов вместо эмпирической дисперсии S_1^2 подсчитанную в задаче 1.4 сводную оценку дисперсии $S^2_{\text{св}} = 0{,}0364$ с $k_{\text{св}} = 19$ степенями свободы.

Решение

Здесь $t_{0,975}(19) = 2{,}09$, $\chi^2_{0,025}(19) = 8{,}91$, $\chi^2_{0,975}(19) = 32{,}9$ и поэтому все доверительные интервалы значительно у́же:
$\varepsilon = 2{,}09 \cdot 0{,}191 \big/ \sqrt{5} = 0{,}18$, т.е. $M(y_1) = 8{,}92 \pm 0{,}18$.
$0{,}191 \cdot \sqrt{19/32{,}9} < \sigma < 0{,}191 \cdot \sqrt{19/8{,}91}$, т.е. $0{,}145 < \sigma < 0{,}279$.

1.7. Проверка статистических гипотез

1.7.1. Основные понятия

На практике часто встречаются задачи, в которых необходимо проверить предположение (гипотезу) о тех или иных свойствах распределения генеральной совокупности, имея в распоряжении случайную выборку (результаты экспериментов). Задачи такого типа возникают при сравнении параметров различных технологических процессов или характеристик материалов.

Пусть X – наблюдаемая случайная величина. *Статистической гипотезой H* называется предположение относительно параметров или вида распределения случайной величины X. Статистическая гипотеза называется *простой*, если она однозначно определяет распределение случайной величины X, в противном случае она называется *сложной*. Например, предположение о том, что случайная величина X распределена по нормальному закону с известными параметрами a, σ, является простой гипотезой. Если же высказывается предположение о том, что случайная величина X имеет нормальное распределение $N(a;\ \sigma)$, где $a \in (\alpha;\beta)$, то это сложная гипотеза. Если вид распределения случайной величины X известен, но не все его параметры известны, то гипотеза о значении этих параметров называется *параметрической*.

Проверяемая гипотеза называется *нуль-гипотезой* и обозначается H_0. При постановки нуль-гипотезы сразу ставится *альтернативная гипотеза* H_1, т.е. то предположение, которое следует принять, если нуль-гипотеза будет отвергнута. Например, если проверяется гипотеза о равенстве параметра β заданному значению β_0, т.е. H_0: $\beta = \beta_0$, то в качестве альтернативной гипотезы можно рассматривать одну из следующих гипотез: $H_1^{(1)}$: $\beta < \beta_0$; $H_1^{(2)}$: $\beta > \beta_0$; $H_1^{(3)}$: $\beta \neq \beta_0$; $H_1^{(4)}$: $\beta = \beta_1$, где β_1 – заданное значение $\beta_1 \neq \beta_0$. Выбор альтернативной гипотезы определяется конкретной задачей.

Правило, позволяющее принять или отвергнуть гипотезу H_0 в соответствии с данными эксперимента, называется *критерием*. Рассматривается *функция критерия* – некоторая функция результатов эксперимента $Q(X_1, X_2, ..., X_n)$, *распределение которой вполне определено при условии истинности гипотезы* H_0.

Проверка статистической гипотезы основывается на принципе, в соответствии с которым события с малой вероятностью считаются невозможными, а события с большой вероятностью – достоверными. Для этого фиксируется некоторая малая вероятность α, называемая *уровнем значимости*. Пусть V – множество значений функции критерия Q. Определяется *критическая область* $V_{\text{кр}}$ – подмножество множества V, в которое функция Q попадает с вероятностью α, и соответствует значениям функции Q, при которых альтернативная гипотеза выполняется наиболее явно. Множество значений V, не принадлежащих критической области, называется *областью принятия* гипотезы H_0 и обозначается $V_{\text{пр}}$.

Проверку статистической гипотезы проводят следующим образом. По заданной выборке вычисляют значение функции Q и рассматривают два случая:

– если вычисленное значение функции Q не принадлежит критической области, гипотеза H_0 принимается;

– если вычисленное значение функции Q принадлежит критической области, гипотеза H_0 отвергается, т.е. принимается гипотеза H_1.

Положение критической области на множестве значений функции критерия Q зависит от формулировки альтернативной гипотезы H_1.

Пусть проверяется параметрическая гипотеза H_0: $\beta = \beta_0$; $f(x)$ – плотность распределения функции критерия Q при условии, что верна гипотеза H_0, q_p – квантиль распределения функции Q.

Возможны различные варианты формулировки альтернативной гипотезы H_1:

1. H_1: $\beta < \beta_0$. Критическая область размещается в левом «хвосте» распределения функции Q, т.е. удовлетворяет неравенству $Q < q_\alpha$ (рис. 1.4, *а*).

2. H_1: $\beta > \beta_0$. Критическая область размещается в правом «хвосте» распределения функции Q, т.е. удовлетворяет неравенству $Q > q_{1-\alpha}$ (рис. 1.4, *б*).

3. H_1: $\beta \neq \beta_0$. Критическая область размещается в обоих «хвостах» распределения функции Q, т.е. удовлетворяет совокупности неравенств $Q < q_{\alpha/2}$ и $Q > q_{1-\alpha/2}$ (рис. 1.4, *в*).

В первых двух вариантах критерий называется *односторонним*, соответственно левосторонним и правосторонним. В случае третьего варианта критерий называется *двусторонним*. Результат проверки статистической гипотезы может быть ошибочен. При этом различают ошибки первого и второго рода. Если гипотеза отвергается, в то время как она верна, совершается *ошибка первого рода*. Вероятность ошибки первого рода равна вероятности попадания функции Q в критическую область при условии, что верна гипотеза H_0, т.е. равна уровню значимости α. Если же гипотеза H_0 принимается, а в действительности верна гипотеза H_1, то совершается *ошибка второго рода*. Вероятность ошибки второго рода, как правило, неизвестна, но известно, что при уменьшении ошибки первого рода α, ошибка второго рода, как правило, возрастает.

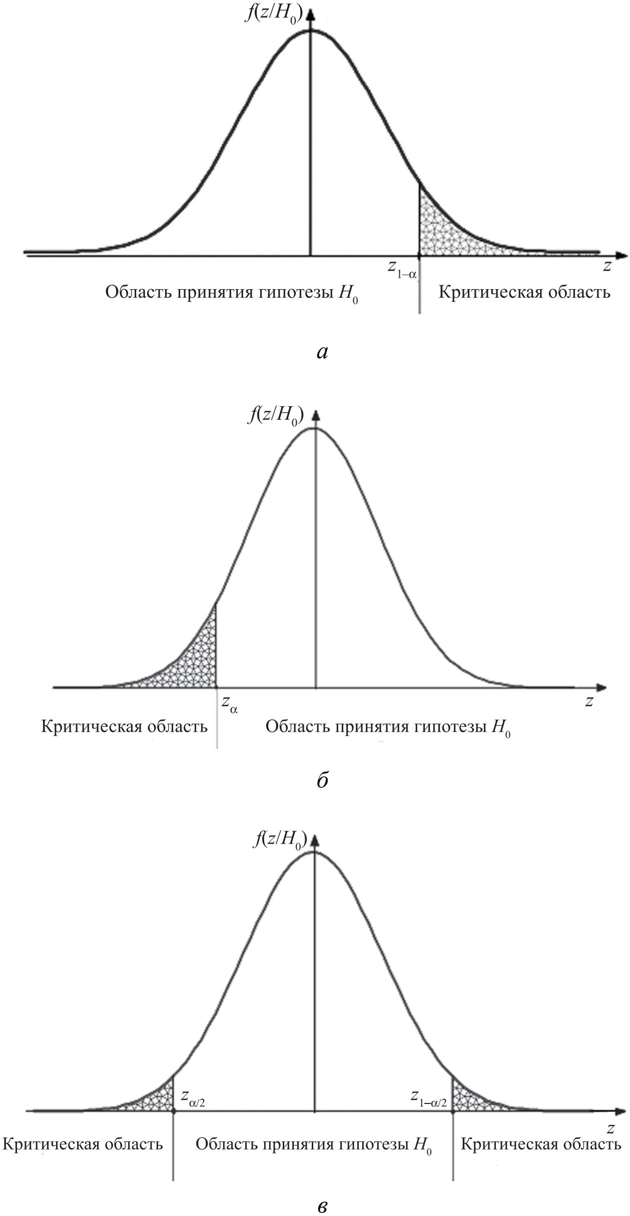

Рис. 1.4. Расположение области принятия гипотезы H_0 и критической области в случае правостороннего критерия (*а*), левостороннего критерия (*б*), двустороннего критерия (*в*)

1.7.2. Проверка гипотез о дисперсии нормального распределения

Пусть задана выборка X_1, X_2, ..., X_n из генеральной совокупности, имеющей нормальное распределение с математическим ожиданием a и дисперсией σ^2. Значение дисперсии неизвестно. Проверим *гипотезу о равенстве дисперсии заданному числу:* H_0: $\sigma^2 = \sigma_0^2$ (σ_0^2 – некоторое заданное число). Математическое ожидание неизвестно.

Гипотезу проверяют с помощью функции критерия

$$Z = k\, S^2 \;/\; \sigma_0^2, \tag{1.27}$$

которая имеет χ^2-распределение с числом степеней свободы k, т.е. $Z = \chi^2(k)$. S^2 – несмещенная оценка дисперсии σ^2 (1.5); $k = n-1$ – число степеней свободы оценки S^2. Рассмотрим проверку нуль-гипотезы при трех различных альтернативных гипотезах H_1.

1. H_1: $\sigma^2 \neq \sigma_0^2$. Критерий двусторонний. H_0-гипотезу принимают при выполнении неравенства

$$\chi^2_{\alpha/2}(k) < Z < \chi^2_{1-\alpha/2}(k), \tag{1.28}$$

в противном случае гипотезу отвергают.

2. H_1: $\sigma^2 < \sigma_0^2$. Критерий односторонний (левосторонний). H_0-гипотезу принимают при выполнении неравенства:

$$Z > \chi^2_{\alpha}(k), \tag{1.29}$$

в противном случае гипотезу отвергают.

3. H_1: $\sigma^2 > \sigma_0^2$. Критерий односторонний (правосторонний). H_0-гипотезу принимают при выполнении неравенства

$$Z < \chi^2_{1-\alpha}(k), \tag{1.30}$$

в противном случае гипотезу отвергают.

Замечание. Если проверку гипотезы H_0: $\sigma^2 = \sigma_0^2$ проводят при известном математическом ожидании, то вместо (1.27) используют функцию критерия

$$Z = n\, S_0^2 \;/\; \sigma_0^2, \tag{1.31}$$

где S_0^2 – оценка дисперсии σ^2 (1.4) при известном математическом ожидании; n – объем выборки. Гипотезу проверяют полностью аналогично алгоритму, описанному выше, только в формулах (1.28) – (1.30) число степеней свободы k заменяют на объем выборки n.

Пусть заданы две независимые выборки из двух нормальных генеральных совокупностей. Первая выборка имеет объем n_1, $X_i^{(1)} \sim N(a_1; \sigma_1)$, вторая – n_2, $X_i^{(2)} \sim N(a_2; \sigma_2)$. Проверить *гипотезу о равенстве дисперсий* этих двух генеральных совокупностей, т.е. H_0: $\sigma_1^2 = \sigma_2^2$. Математические ожидания a_1 и a_2 неизвестны.

В этом случае по каждой выборке находят несмещенные оценки дисперсий S_1^2 и S_2^2 с числами степеней свободы $k_1 = n_1 - 1$ и $k_2 = n_2 - 1$ соответственно. Гипотезу проверяют по критерию Фишера, функция критерия

$$F = S_1^2 \big/ S_2^2 \qquad (1.32)$$

имеет F-распределения Фишера с k_1 и k_2 степенями свободы, т.е. $F = F(k_1, k_2)$. Рассмотрим проверку нуль-гипотезы, как и в предыдущей задаче, при трех различных альтернативных гипотезах H_1.

1. H_1: $\sigma_1^2 \neq \sigma_2^2$. В этом случае критерий Фишера рассчитывают как отношение большей по величине оценки дисперсии к меньшей:

$$F = S_{\text{бол}}^2 \big/ S_{\text{мен}}^2 > 1. \qquad (1.33)$$

Критерий двусторонний. Гипотезу H_0 принимают при выполнении неравенства

$$F < F_{1-\alpha/2}(k_{S_{\text{бол}}}, k_{S_{\text{мен}}}), \qquad (1.34)$$

в противном случае гипотеза H_0 отвергается. Здесь $k_{S_{\text{бол}}}$ – число степеней свободы большей оценки дисперсии; $k_{S_{\text{мен}}}$ – число степеней свободы меньшей оценки дисперсии.

2. H_1: $\sigma_1^2 > \sigma_2^2$. Критерий Фишера равен $F = S_1^2 \big/ S_2^2$. Критерий односторонний (правосторонний). Гипотезу H_0 принимают при выполнении неравенства

$$F < F_{1-\alpha}(k_1, k_2) \qquad (1.35)$$

в противном случае гипотезу H_0 отвергают.

3. H_1: $\sigma_1^2 < \sigma_2^2$. В этом случае следует поменять нумерацию выборок и, соответственно, оценок дисперсии S_1^2 и S_2^2 и осуществить проверку, как указано в пункте 2.

Замечание. Если проверку гипотезы о равенстве дисперсий H_0: $\sigma_1^2 = \sigma_2^2$ проводят при известных математических ожиданиях, то вместо (1.32) используют функцию критерия

$$F = S_{01}^2 \big/ S_{02}^2 \,, \tag{1.36}$$

где S_{01}^2 и S_{02}^2 – оценки дисперсии σ^2 (1.4) при известном математическом ожидании. В этом случае критерий Фишера имеет F-распределение Фишера с n_1 и n_2 степенями свободы, т.е. $F = F(n_1, n_2)$. Гипотезу проверяют полностью аналогично алгоритму, описанному выше, только в формулах (1.33) – (1.35) числа степеней свободы k_1 и k_2 заменяют на объемы выборок n_1 и n_2.

Пусть заданы $n > 2$ независимых выборок из n нормальных генеральных совокупностей (n серий экспериментов). Для проверки гипотезы H_0 об однородности дисперсий: $\sigma_1^2 = \sigma_2^2 = ... = \sigma_n^2$ можно воспользоваться критерием Кохрена G. В случае равенства чисел степеней свободы $k_1 = k_2 = ... = k_n$ ($= k$) несмещенных оценок дисперсий S_1^2, S_2^2, ..., S_n^2 рассматривается отношение наибольшей из этих дисперсий $S_{\max}^2$ к сумме всех дисперсий:

$$G = S_{\max}^2 \big/ (S_1^2 + S_2^2 + ... + S_n^2)\,. \tag{1.37}$$

Согласно критерию Кохрена гипотезу H_0 принимают с уровнем значимости α, если $G < G_{1-\alpha}(k, n)$, и отвергают в противоположном случае. Квантили распределения Кохрена $G_p(k, n)$ приведены в табл. П8 приложения.

Задача 1.13. В двух сериях независимых экспериментов (двух выборках из нормальных генеральных совокупностей (ГС)) получены несмещенные оценки дисперсии S_1^2 = 1,95 с k_1 = 15 степенями свободы и S_2^2 = 0,75 с k_2 = 20 степенями свободы. Проверить гипотезу о равенстве дисперсий при альтернативной гипотезе $\sigma_1^2 \neq \sigma_2^2$ с уровнем значимости α = 0,05.

Решение

Вычислим значение критерия Фишера (1.33) F = 1,95/0,75 = 2,60 и сравним его с квантилью распределения Фишера по табл. П5 приложения (двусторонний критерий) $F_{0,975}(15; 20) = 2,57$. Так как $F = 2,60 > 2,57$, то гипотезу о равенстве дисперсий в двух сериях экспериментов следует отвергнуть.

Задача 1.14. Партия чугунных отливок принимается, если дисперсия контролируемого размера не превышает 0,15 с уровнем значимости 0,01. Из партии отливок произвели случайную выборку объемом $n = 46$. Оценка дисперсии получилась равной 0,23. Можно ли принять эту партию отливок, если генеральная совокупность имеет нормальное распределение?

Решение

Объем выборки $n = 46$; $S^2 = 0{,}23$; $\alpha = 0{,}01$; $\sigma_0^2 = 0{,}15$. Надо проверить гипотезу H_0: $\sigma^2 = \sigma_0^2$ при альтернативной гипотезе H_1: $\sigma^2 > \sigma_0^2$. Математическое ожидание ГС неизвестно. Критерий односторонний (правосторонний). Вычислим значение критерия Z по формуле (1.27):

$$Z = \frac{kS^2}{\sigma_0^2} = \frac{45 \cdot 0{,}23}{0{,}15} = 69$$

и сравним его с критическим значением $\chi^2_{1-\alpha}(k) = \chi^2_{0{,}99}(45) \approx 70$, который найден по таблице квантилей χ^2-распределения (табл. П3). Так как $Z = 69 < 70$, то гипотеза H_0 принимается с уровнем значимости $\alpha = 0{,}01$ (см. (1.30)) и данную партию отливок можно принять.

Задача 1.15. Сравнивается точность штамповки деталей на двух станках. Из продукции первого станка было отобрано 13 деталей, второго – 15 деталей. Оценки дисперсии этих выборок оказались равны $S_1^2 = 3{,}24$ и $S_2^2 = 1{,}2$. При уровне значимости $\alpha = 0{,}05$ проверяем гипотезу о том, что оба станка обеспечивают одинаковую точность, причем в качестве альтернативной выбираем две гипотезы: 1) станки обеспечивают неодинаковую точность; 2) второй станок обеспечивает более высокую точность.

Решение

Из условия задачи следует, что математические ожидания выборок не известны.

1. Вычислим значение критерия F по формуле (1.33):

$$F = S^2_{\text{бол}} \big/ S^2_{\text{мен}} = S_1^2 \,/\, S_2^2 = 3{,}24 / 1{,}2 = 2{,}7.$$

Критерий двусторонний, так как альтернативная гипотеза H_1: $\sigma_1^2 \neq \sigma_2^2$. Критическое значение критерия F

$$F_{1-\alpha/2}(k_1, k_2) = F_{0,975}(12;14) = 3,05.$$

$F = 2,7 < 3,05$, следовательно, гипотеза о равенстве дисперсий принимается с уровнем значимости α = 0,05. Принимаем утверждение, что станки обеспечивают одинаковую точность.

2. Альтернативная гипотеза H_1: $\sigma_1^2 > \sigma_2^2$, поэтому критерий односторонний, критическая область правосторонняя. Значение критерия Фишера $F = S_1^2 / S_2^2$ = 3,24/1,2 = 2,7. Критическое значение критерия Фишера

$$F_{1-\alpha}(k_1, k_2) = F_{0,95}(12;14) = 2,53.$$

$F = 2,7 > 2,53$, следовательно, гипотезу о равенстве дисперсий отвергаем с уровнем значимости α = 0,05 (1.35) и принимаем утверждение, что второй автомат обеспечивает более высокую точность.

Замечание. Как видно из рассмотренной выше задачи, *разные альтернативные гипотезы могут привести к разным выводам о* H_0*-гипотезе.*

Задача 1.16. Для оценки погрешности нового измерительного прибора на нем была проведена серия измерений эталонной величины (т.е. математическое ожидание было известно). Число измерений n = 45. Оценка дисперсии оказалась равной 0,21. Прибор считается годным, если дисперсия не превышает 0,15 при уровне значимости α = 0,05. Можно ли принять прибор в эксплуатацию, если генеральная совокупность нормальна?

Решение

Следует проверить гипотезу H_0: $\sigma^2 = \sigma_0^2$ при $\sigma_0^2 = 0,15$; $S_0^2 = 0,21$; n = 45 и альтернативной гипотезе H_1: $\sigma^2 > \sigma_0^2$. Математическое ожидание известно. Критерий односторонний (правосторонний).

Вычислим значение критерия Z по формуле (1.31):

$$Z = n\, S_0^2 \,/\, \sigma_0^2 = 45 \cdot 0,21 / 0,15 = 63$$

и сравним с критическим значением $\chi^2_{1-\alpha}(n) = \chi^2_{0,95}(45) = 61,7$. Расчетное значение Z = 63 больше критического, поэтому гипотезу H_0

отвергаем и делаем вывод, что прибор не может быть принят в эксплуатацию.

Задача 1.17. До наладки электронных весов на них была проведена серия повторных измерений эталонного образца, состоящая из 15 испытаний. Получена оценка дисперсии S_{01}^2 = 0,020. После наладки провели еще 20 испытаний того же образца. Получили новое значение оценки дисперсии S_{02}^2 = 0,012. Можно ли считать, что в результате наладки весов точность взвешивания увеличилась? Принять α = 0,05. Генеральная совокупность нормальна.

Решение

Надо проверить гипотезу H_0: $\sigma_1^2 = \sigma_2^2$ при альтернативной гипотезе H_1: $\sigma_1^2 > \sigma_2^2$. Математическое ожидание известно. S_{01}^2 = 0,020; n_1 = 15; S_{02}^2 = 0,012; n_2 = 20. Критерий односторонний (правосторонний).

Вычислим значение критерия Фишера по формуле (1.36):

$$F = S_{01}^2 \big/ S_{02}^2 = 0{,}020 / 0{,}012 = 1{,}67$$

и сравним эту величину с критическим значением

$$F_{1-\alpha}(n_1, n_2) = F_{0,95}(15;\, 20) = 2{,}20\,,$$

которое находим по таблице квантилей распределения Фишера (табл. П4). Поскольку вычисленное значение F меньше критического значения, гипотезу H_0 принимаем с уровнем значимости α = 0,05. Можно сделать вывод, что точность взвешивания не изменилась.

Задача 1.18. Проведено шесть серий независимых измерений значений функции в шести точках, по три измерения в каждой. В каждой точке результаты измерений предполагаются равноточными и распределенными по нормальному закону. При первичной обработке этих результатов получены следующие несмещенные оценки дисперсии: S_1^2 = 0,04; S_2^2 = 0,26; S_3^2 = 0,12; S_4^2 = 0,08; S_5^2 = 0,06; S_6^2 = 0,04. Можно ли считать все измерения равноточными при уровне значимости α = 0,05?

Решение

Вычисляем отношение (1.37):

$G = 0{,}26/(0{,}04 + 0{,}26 + 0{,}12 + 0{,}08 + 0{,}06 + 0{,}04) = 0{,}433$

и сравниваем его с квантилью распределения Кохрена (табл. П8): $G_{0,95}(2;\ 6) = 0{,}616$. Так как $0{,}433 < 0{,}616$, то нет оснований отвергать гипотезу о равноточности всех измерений.

1.7.3. Проверка гипотез о математических ожиданиях нормального распределения

Пусть задана выборка $X_1, X_2, ..., X_n$ из генеральной совокупности, имеющей нормальное распределение с математическим ожиданием a и дисперсией σ^2. Значение математического ожидания неизвестно. Найдена оценка математического ожидания $\overline{X}$. Проверяем *гипотезу о равенстве математического ожидания заданному числу.* $H_0: a = a_0$ (a_0 – некоторое заданное число). Рассмотрим два случая.

Первый случай: дисперсия σ^2 известна.

Гипотезу проверяется с помощью функции критерия

$$U = \frac{(\overline{X} - a_0)\sqrt{n}}{\sigma}, \qquad (1.38)$$

которая имеет стандартное нормальное распределение, т.е. $U \sim N(0;\ 1)$. u_p – квантиль стандартного нормального распределения. Рассмотрим проверку гипотезы H_0 при трех вариантах альтернативной гипотезы H_1.

1. H_1: $a \neq a_0$. Критерий двусторонний. Гипотеза H_0 принимается при выполнении неравенства

$$|U| < u_{1-\alpha/2}, \qquad (1.39)$$

в противном случае гипотеза H_0 отвергается.

2. H_1: $a < a_0$. Критерий односторонний (левосторонний). Гипотеза H_0 принимается при выполнении неравенства

$$U > u_\alpha, \qquad (1.40)$$

в противном случае гипотеза H_0 отвергается.

3. H_1: $a > a_0$. Критерий односторонний (правосторонний). Гипотеза H_0 принимается при выполнении неравенства

$$U < u_{1-\alpha}\,, \qquad (1.41)$$

в противном случае гипотеза H_0 отвергается.

Второй случай: дисперсия σ^2 неизвестна.

Гипотеза проверяется с помощью функции критерия

$$t = \frac{(\overline{X} - a_0)\sqrt{n}}{S}\,, \qquad (1.42)$$

которая имеет t-распределение Стьюдента с k степенями свободы, т.е. $t = t(k)$; S – оценка среднего квадратического отклонения σ (1.6); k – число степеней свободы при вычислении оценки S. $t_P(k)$ – квантиль распределения Стьюдента. Как и выше, рассматриваем проверку гипотезы H_0 при трех вариантах альтернативной гипотезы H_1.

1. H_1: $a \neq a_0$. Критерий двусторонний. Гипотеза H_0 принимается при выполнении неравенства

$$|\,t| < t_{1-\alpha/2}(k)\,, \qquad (1.43)$$

в противном случае гипотеза H_0 отвергается.

2. H_1: $a < a_0$. Критерий односторонний (левосторонний). Гипотеза H_0 принимается при выполнении неравенства

$$t > t_{\alpha}(k)\,, \qquad (1.44)$$

в противном случае гипотеза H_0 отвергается.

3. H_1: $a > a_0$. Критерий односторонний (правосторонний). Гипотеза H_0 принимается при выполнении неравенства

$$t < t_{1-\alpha}(k)\,, \qquad (1.45)$$

в противном случае гипотеза H_0 отвергается.

Пусть заданы две независимые выборки из двух нормальных генеральных совокупностей. Первая выборка имеет объем n_1, $X_i^{(1)} \sim N(a_1; \sigma_1)$, вторая – n_2, $X_i^{(2)} \sim N(a_2; \sigma_2)$. Математические ожидания a_1 и a_2 неизвестны.

Проверяем *гипотезу о равенстве математических ожиданий* этих двух генеральных совокупностей, т.е. H_0: $a_1 = a_2$. По каждой выборке находим оценки математических ожиданий $\overline{X}_1$ и $\overline{X}_2$. Рассмотрим два случая.

Первый случай: дисперсии σ_1 и σ_2 известны.

Гипотеза проверяется с помощью функции критерия

$$U = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}, \tag{1.46}$$

функция имеет стандартное нормальное распределение, т.е. $U \sim N(0; 1)$. Рассмотрим различные варианты альтернативной гипотезы H_1.

1. H_1: $a_1 \neq a_2$. Критерий двусторонний. Гипотеза H_0 принимается при выполнении неравенства

$$|U| < u_{1-\alpha/2}, \tag{1.47}$$

в противном случае гипотеза H_0 отвергается.

2. H_1: $a_1 < a_2$. Критерий односторонний (левосторонний). Гипотеза H_0 принимается при выполнении неравенства

$$U > u_{\alpha}, \tag{1.48}$$

в противном случае гипотеза H_0 отвергается.

3. H_1: $a_1 > a_2$. Критерий односторонний (правосторонний). Гипотеза H_0 принимается при выполнении неравенства

$$U < u_{1-\alpha}, \tag{1.49}$$

в противном случае гипотеза H_0 отвергается.

Второй случай: дисперсии σ_1^2 и σ_2^2 неизвестны, но гипотеза о равенстве дисперсий принимается. S_1^2 п S_2^2 – несмещенные оценки дисперсий первой и второй выборок. Находим сводную оценку дисперсии

$$S_{\text{св}}^2 = (S_1^2 k_1 + S_2^2 k_2) / (k_1 + k_2).$$

Гипотеза проверяется по критерию Стьюдента, функция критерия

$$t = \frac{\overline{X}_1 - \overline{X}_2}{S_{\text{св}} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \tag{1.50}$$

имеет t-распределение Стьюдента с $k_{св}$ степенями свободы, т.е. $t = t(k_{св})$. $k_{св} = k_1 + k_2 = n_1 + n_2 - 2$ – число степеней свободы при вычислении оценки $S_{св} = \sqrt{S_{св}^2}$. Рассмотрим различные варианты альтернативной гипотезы H_1.

1. H_1: $a_1 \neq a_2$. Критерий двусторонний. Гипотеза H_0 принимается при выполнении неравенства

$$|t| < t_{1-\alpha/2}(k_{св}), \tag{1.51}$$

в противном случае гипотеза H_0 отвергается.

2. H_1: $a_1 < a_2$. Критерий односторонний (левосторонний). Гипотеза H_0 принимается при выполнении неравенства

$$t > t_{\alpha}(k_{св}), \tag{1.52}$$

в противном случае гипотеза H_0 отвергается.

3. H_1: $a_1 > a_2$. Критерий односторонний (правосторонний). Гипотеза H_0 принимается при выполнении неравенства:

$$t < t_{1-\alpha}(k_{св}), \tag{1.53}$$

в противном случае гипотеза H_0 отвергается.

Рассмотрим примеры проверки статистических гипотез. Во всех нижеприведенных задачах предполагается, что рассматриваемые генеральные совокупности имеют нормальное распределение.

Задача 1.19. Из продукции цеха выбрано $n = 60$ отливок, средняя масса которых составила $\overline{X} = 2{,}87$. При уровне значимости $\alpha = 0{,}01$ проверить гипотезу о том, что математическое ожидание массы чугунной отливки составляет $a_0 = 3$, причем выдвинуты два варианта альтернативной гипотезы первая H_1: $a \neq a_0$ и вторая H_1: $a < a_0$. Решить задачу в двух случаях: 1) дисперсия известна: $\sigma^2 = 0{,}16$ и 2) дисперсия неизвестна, но оценка дисперсии, найденная по той же выборке, равна $S^2 = 0{,}16$.

Решение

1. По формуле (1.38) вычислим функцию критерия

$$U = \frac{(\bar{x} - a_0)\sqrt{n}}{\sigma} = \frac{(2{,}87 - 3)\sqrt{60}}{0{,}4} \approx -2{,}517\,.$$

• Критическая область двусторонняя. В таблице квантилей стандартного нормального распределения (табл. П1) найдем критическое значение $u_{1-\alpha/2} = u_{0,995} = 2,576$. Так как $|U|$ = 2,517 < 2,576 , то гипотезу H_0 принимается с уровнем значимости α = 0,01 (см. (1.39)).

• Критическая область левосторонняя. Найдем критическое значение $u_\alpha = u_{0,01} = -u_{0,99} = -2,326$. Так как U = –2,517 < –2,326, то гипотезу H_0 отклоняется с уровнем значимости α = 0,01 (см. (1.40)).

2. По формуле (1.42) вычислим функцию критерия

$$t = \frac{(\overline{X} - a_0)\sqrt{n}}{S} = \frac{(2,87-3)\sqrt{60}}{0,4} \approx -2,517 .$$

• Критическая область двусторонняя. В таблице квантилей распределения Стьюдента (табл. П2) найдем критическое значение $t_{1-\alpha/2}(k) = t_{0,995}(59) = 2,66$. Так как $|t|$ = 2,517 < 2,66, то гипотеза H_0 принимается с уровнем значимости α = 0,01 (см. (1.43)).

• Критическая область левосторонняя. Найдем критическое значение $t_\alpha(k) = t_{0,01}(59) = -t_{0,99}(59) = -2,39$. Так как t = –2,517 < –2,39, то гипотеза H_0 отклоняется с уровнем значимости α = 0,01 (см. (1.44)).

Задача 1.20. Номинальный диаметр шариков, изготовленных станком-автоматом, равен d_0 = 10 мм. В выборке из 36 шариков средний диаметр $\overline{d}$ оказался равным 10,4 мм. Используя односторонний критерий при α = 0,05, проверить гипотезу H_0: $d = d_0$, рассмотрев два случая: 1) дисперсия известна и равна σ^2 = 2,1 мм^2; 2) дисперсия неизвестна; оценка дисперсии, определенная по той же выборке, равна S^2 = 2,1 мм^2.

Решение

Альтернативная гипотеза H_1: $d > d_0$.

1. По формуле (1.38) вычислим функцию критерия

$$U = \frac{(\overline{d} - d_0)\sqrt{n}}{\sigma} = \frac{(10,4-10)\sqrt{36}}{\sqrt{2,1}} \approx 1,656 .$$

В таблице квантилей стандартного нормального распределения найдем критическое значение $u_{1-\alpha} = u_{0,95} = 1,645$. Так как

$U = 1{,}656 > 1{,}645$, то гипотеза H_0 отклоняется с уровнем значимости $\alpha = 0{,}05$ (см. (1.41)).

2. По формуле (1.42) вычислим функцию критерия

$$t = \frac{(\bar{d} - d_0)\sqrt{n}}{S} = \frac{(10{,}4 - 10)\sqrt{36}}{\sqrt{2{,}1}} \approx 1{,}656 .$$

В таблице квантилей распределения Стьюдента найдем критическое значение $t_{1-\alpha}(k) = t_{0{,}95}(35) = 1{,}69$. Так как $t = 1{,}656 < 1{,}69$, то гипотеза H_0 принимается с уровнем значимости $\alpha = 0{,}01$ (см. (1.45)).

Задача 1.21. Цех выплавляет по старой технологии в среднем 1000 кг за смену со средним квадратическим отклонением $\sigma = 60$ кг. Новая технология производства в среднем дает 1050 кг за смену с тем же средним квадратическим отклонением. Можно ли считать, что новая технология обеспечивает повышение производительности при уровне значимости $\alpha = 0{,}05$, имея в распоряжении данные о двадцати пяти плавках, проведенных по старой технологии ($n_1 = 25$), и данные о двадцати плавках, проведенных по новой технологии ($n_2 = 20$)?

Решение

Проверим гипотезу о равенстве математических ожиданий двух генеральных совокупностей. H_0: $a_1 = a_2$ при альтернативной гипотезе H_1: $a_1 < a_2$; дисперсии $\sigma_1^2 = \sigma_2^2$ – известны. Критерий односторонний (левосторонний). Вычислим функцию критерия (1.46):

$$U = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} = \frac{1000 - 1050}{60\sqrt{\dfrac{1}{25} + \dfrac{1}{20}}} \approx -2{,}778 .$$

Найдем критическое значение $u_\alpha = u_{0{,}05} = -u_{0{,}95} = -1{,}645$. Так как $U = -2{,}778 < -1{,}645$, то гипотеза H_0 отвергается с уровнем значимости $\alpha = 0{,}05$ (см. (1.48)). Следовательно, можно считать, что новая технология обеспечивает повышение производительности.

Задача 1.22. Проверить гипотезу о равенстве математических ожиданий контролируемой массы изделий, изготовленных на первом и втором станках, если по выборке объема $n_1 = 20$ найдена средняя масса $\overline{X}_1 = 198$ г изделий, изготовленных на первом станке; по выборке объема $n_2 = 42$ найдена средняя масса $\overline{X}_2 = 193$ г изделий, изготовленных на втором станке. Оценки дисперсий, найденные по тем же выборкам, равны соответственно 50 г^2 и 64 г^2. Гипотеза о равенстве дисперсий принята. Уровень значимости $\alpha = 0{,}01$. Рассмотреть два случая: 1) альтернативная гипотеза: масса изделий первого станка больше, чем второго; 2) альтернативная гипотеза: массы изделий первого и второго станков различны.

Решение

1. Проверим гипотезу о равенстве математических ожиданий двух генеральных совокупностей H_0: $a_1 = a_2$ при альтернативной гипотезе H_1: $a_1 > a_2$; дисперсии неизвестны. Критерий односторонний (правосторонний).

Найдем сводную оценку дисперсии

$$S^2_{\text{с в}} = (S_1^2 k_1 + S_2^2 k_2)\big/(k_1 + k_2) = (50 \cdot 19 + 64 \cdot 41) / (19 + 41) \approx 59{,}57\,,$$

где $k_1 = n_1 - 1 = 19$; $k_2 = n_2 - 1 = 41$. Вычислим функцию критерия (1.50):

$$t = \frac{\overline{X}_1 - \overline{X}_2}{S_{\text{св}}\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} = \frac{198 - 193}{\sqrt{59{,}57} \cdot \sqrt{\dfrac{1}{20} + \dfrac{1}{42}}} \approx 2{,}384\,. \qquad (1.54)$$

По таблице квантилей распределения Стьюдента найдем критическое значение $t_{1-\alpha}(k_{\text{св}}) = t_{0,99}(60) = 2{,}390$, где $k_{\text{св}} = k_1 + k_2 = 19 + 41 = 60$. Так как $t = 2{,}384 < 2{,}390$, то принимаем нуль-гипотезу (см. (1.53)) с уровнем значимости $\alpha = 0{,}01$.

2. Проверим гипотезу о равенстве математических ожиданий при альтернативной гипотезе H_1: $a_1 \neq a_2$; дисперсии неизвестны. Критическая область двусторонняя. Значение критерия Стьюдента вычисляется, как и в первой части задачи (1.54).

По таблице квантилей распределения Стьюдента найдем критическое значение $t_{1-\alpha/2}(k_{\text{св}}) = t_{0,995}(60) = 2{,}660$. Так как $|t| = 2{,}384 < 2{,}660$, то гипотеза принимается (см. (1.51)) с уровнем значимости $\alpha = 0{,}01$.

Задача 1.23. Из двух партий изделий, изготовленных на разных станках, извлечены две выборки с объемами $n_1 = 12$ и $n_2 = 15$. Результаты представлены в табл. 1.7 и 1.8. При уровне значимости $\alpha = 0{,}02$ проверить гипотезу о равенстве математических ожиданий контролируемых размеров изделий, изготовленных на первом и втором станках.

Таблица 1.7

Размер изделий, изготовленных на первом станке (к задаче 1.23)

Размер изделий первого станка, X_{1i}	5,4	5,5	5,5	5,7
Частота (число изделий), n_{1i}	2	5	4	1

Таблица 1.8

Размер изделий, изготовленных на втором станке (к задаче 1.23)

Размер изделий второго станка, X_{2i}	5,2	5,3	5,4
Частота (число изделий), n_{2i}	2	4	9

Решение

Сначала найдем оценки математических ожиданий и дисперсий для каждой выборки:

$$\overline{X}_1 = \frac{1}{n_1}\sum_{i=1}^{4} X_{1i}n_{1i} = \frac{1}{12}(5{,}4\cdot 2 + 5{,}5\cdot 5 + 5{,}6\cdot 4 + 5{,}7) \approx 5{,}53;$$

$$S_1^2 = \frac{1}{n_1 - 1}\left(\sum_{i=1}^{4} X_{1i}^2 n_{1i} - n_1\overline{X}_1^2\right) =$$

$$= \frac{1}{11}(5{,}4^2\cdot 2 + 5{,}5^2\cdot 5 + 5{,}6^2\cdot 4 + 5{,}7^2 - 12\cdot 5{,}53^2) = 0{,}0120.$$

Аналогично: $\overline{X}_2 = 5{,}35$; $S_2^2 = 0{,}0055$.

Проверим гипотезу о равенстве дисперсий H_0: $\sigma_1^2 = \sigma_2^2$ при альтернативной гипотезе H_1: $\sigma_1^2 \neq \sigma_2^2$. Найдем значение критерия Фишера (1.33) (отношение большей оценки дисперсии к меньшей):

$$F = S_1^2 \big/ S_2^2 \approx 2{,}18.$$

Критическое значение (квантиль распределения Фишера):

$$F_{1-\alpha/2}(n_1 - 1, n_2 - 1) = F_{0{,}99}(11{,}14) = 3{,}87.$$

Так как $F = 2{,}18 < 3{,}87$, то принимаем гипотезу о равенстве дисперсий. Найдем сводную оценку дисперсии

$$S_{\text{св}}^2 = (S_1^2 k_1 + S_2^2 k_2) \big/ (k_1 + k_2) = (0{,}012 \cdot 11 + 0{,}0055 \cdot 14) / (11 + 14) \approx$$

$\approx 0{,}0084$, где $k_1 = n_1 - 1 = 11$; $k_2 = n_2 - 1 = 14$.

Вычислим функцию критерия (1.50):

$$t = \frac{\overline{X}_1 - \overline{X}_2}{S_{\text{св}} \sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} = \frac{5{,}53 - 5{,}35}{\sqrt{0{,}0084} \cdot \sqrt{\dfrac{1}{12} + \dfrac{1}{15}}} \approx 4{,}901.$$

По таблице квантилей распределения Стьюдента найдем критическое значение $t_{1-\alpha/2}(k_{\text{св}}) = t_{0,99}(25) = 2{,}485$, где $k_{\text{св}} = k_1 + k_2 = 11 + 14 = 25$. Так как $|t| = 4{,}901 > 2{,}485$, то гипотеза H_0 отвергается (см. (1.51)) с уровнем значимости $\alpha = 0{,}02$. Делаем вывод, что математические ожидания размеров изделий различаются.

1.7.4. Проверка гипотезы о виде распределения генеральной совокупности

Если распределение случайной величины X неизвестно, можно рассмотреть гипотезу о том, что X имеет функцию распределения $F(x)$. Критерии значимости для проверки таких гипотез называют *критериями согласия*. Мы рассмотрим два критерия согласия – критерий χ^2 (или критерий Пирсона) и критерий ω^2.

Пусть X_1, X_2, ..., X_n – выборка наблюдений случайной величины X. Проверяется гипотеза H_0, утверждающая, что X имеет функцию распределения $F(x)$.

Проверку гипотезы H_0 с помощью критерия χ^2 проводят следующим образом. По выборке находят оценки неизвестных параметров предполагаемого закона распределения случайной величины X. Область возможных значений случайной величины X разбивают на l интервалов. Подсчитывают числа n_i попаданий результатов экспериментов в каждый i-й интервал. Используя предполагаемый закон распределения случайной величины X, находят вероятности p_i того, что значение X принадлежит i-му интервалу. Затем сравнивают полученные частоты n_i/n с вероятностями p_i. Критерий согласия Пирсона требует принять гипотезу о пригодности проверяемого распределения с уровнем значимости α, если значение *взвешенной суммы квадратов отклонений*

$$\chi^2 = n\sum_{i=1}^{l} \frac{(n_i / n - p_i)^2}{p_i} = \sum_{i=1}^{l} \frac{(n_i - np_i)^2}{np_i} \tag{1.55}$$

меньше квантиля распределения χ^2-распределения с $k = l - 1$ степенями свободы, т.е.

$$\chi^2 < \chi^2_{1-\alpha}(k),$$

в противном случае эту гипотезу отвергают, как противоречащую результатам эксперимента. Если при этом некоторые параметры распределения оценивают по результатам той же выборки, то квантиль χ^2-распределения следует брать для $k = l - 1 - m$ степеней свободы, где m – число оцениваемых параметров.

По критерию согласия ω^2 оценивают не частоты с вероятностями, а предполагаемую функцию распределения $F(x)$ с функцией эмпирического распределения $F_n(x)$, где

$$F(x) = \int_{-\infty}^{x} \varphi(x)dx, \quad F_n(x) = \begin{cases} 0, X \le X_1, \\ i / n, X_i < x \le X_{i+1}, \\ 1, x > X_n \end{cases}, \tag{1.56}$$

здесь X_1, X_2, ..., X_n – ранжированные результаты эксперимента (расположенные в порядке возрастания).

Критерий согласия ω^2 требует принять гипотезу о пригодности проверяемого распределения с уровнем значимости α, если значение величины

$$n\omega_n^2 = \frac{1}{12n} + \sum_{i=1}^{n} \left[F(x_i) - \frac{2i-1}{2n} \right]^2 \tag{1.57}$$

меньше квантиля предельного распределения величины $n\omega_n^2$ при вероятности $p = 1 - \alpha$, в противном случае гипотеза отвергается, как противоречащая результатам эксперимента. Упомянутые квантили можно найти в специальных таблицах [4], для P = 0,95 и 0,99 квантили равны соответственно 0,461 и 0,742. Предполагается, что проверяемое распределение не содержит неизвестных параметров.

Задача 1.24. Проведено $n = 100$ экспериментов, их результаты разнесены по восьми интервалам, как указано в табл. 1.9. Проверить пригодность нормального или логарифмически нормального закона распределения ($\alpha = 0{,}05$). Случайная величина Y имеет логарифмически нормальный закон распределения, если логарифм ее значений имеет нормальный закон распределения.

Таблица 1.9

Исходные данные задачи 1.24 и результаты расчета

$y_{i-1} - y_i$	n_i	t_i	$\Phi(t_i)$	np_i	$n_i - np_i$	$(n_i - np_i)^2/np_i$
1,0÷1,4	12	–1,131	–0,3710	12,90	–0,90	0,06
1,4÷1,8	11	–0,848	–0,3018	6,92	4,08	2,41
1,8÷2,6	24	–0,283	–0,1114	19,04	4,96	1,29
2,6÷3,4	23	0,283	0,1114	22,28	0,72	0,02
3,4÷4,2	12	0,848	0,3018	19,04	–7,04	2,60
4,2÷5,0	6	1,414	0,4213	11,95	–5,95	2,96
5,0÷5,8	7	1,980	0,4761	5,48	1,52	0,42
5,8÷7,4	5	∞	0,5000	2,39	2,61	2,85
Σ	100	–	–	100	0	12,61

Решение

Проведя первичную обработку результатов эксперимента, определили среднее значение $\overline{Y} = 3$ и эмпирическую дисперсию $S_y^2 = 2$. Вероятности для нормального распределения вычисляем по формуле (1.17): $p_i = \Phi(t_i) - \Phi(t_{i-1})$, где $t_i = (y_i - 3)\big/\sqrt{2}$, $\Phi(t)$ – интеграл вероятностей (см. табл. приложения [2]). Результаты расчета приведены в правой части табл. 1.9. Крайние интервалы для расчета нормального распределения следует считать бесконечными, поэтому полагаем $t_0 = -\infty$, $\Phi(t_0) = -0{,}5$, $t_8 = +\infty$, $\Phi(t_8) = 0{,}5$. Вычисленные значения p_i умножаем на 100. Сумма чисел последнего столбца, равная $\chi_n^2 = 12{,}61$, превосходит квантиль 11,07 для $P = 0{,}95$ и $k = 8 - 3 = 5$. Поэтому гипотезу о пригодности нормального закона отвергаем.

Для проверки пригодности логарифмически нормального распределения, т.е. нормального закона распределения логарифмов величины y, прологарифмируем полученные в эксперименте значения y, обозначив их $\ln y = x$, и проведем весь расчет со значениями величины X. Первичная обработка этих значений дает среднее значение $\overline{X} = 1$ и эмпирическую дисперсию $S_X^2 = 0{,}20$. Здесь полагаем $t_i = (x_i - 1)/0{,}45$, результаты расчета приведены в табл. 1.10.

Таблица 1.10

Результаты расчета (к задаче 1.24)

$x_{i-1} - x_i$	n_i	t_i	$\Phi(t_i)$	np_i	$n_i - np_i$	$(n_i - np_i)^2/np_i$
0÷0,336	12	–1,476	–0,4300	7,00	5,00	3,57
0,336÷0,588	11	–0,916	–0,3202	10,98	0,02	0,00
0,588÷0,956	24	–0,098	–0,0390	28,12	–4,12	0,60
0,956÷1,224	23	0,498	0,1908	22,98	0,02	0,00
1,224÷1,435	12	0,967	0,3332	14,24	–2,24	0,35
1,435÷1,609	6	1,353	0,4120	7,88	–1,88	0,45
1,609÷1,758	7	1,684	0,4539	4,19	2,81	1,88
1,758÷2,001	5	∞	0,5000	4,61	0,39	0,03
Σ	100	–	–	100	0	6,88

Здесь $\chi_n^2 = 6{,}88$ оказывается меньше соответствующей квантили 11,07, и поэтому нет оснований отвергать гипотезу о пригодности логарифмически нормального распределения для описания результатов эксперимента.

1.8. Регрессионный анализ. Построение линейной и квадратичной регрессионных моделей

1.8.1. Оценка коэффициентов регрессии

Важной задачей математической статистики является получение функциональной зависимости одной величины (y) от другой (x) по результатам эксперимента. Будем считать, что функциональная зависимость между величинами, называемая в дальнейшем *моделью*, известна из предварительных сведений с точностью до параметров $\beta_1, \beta_2, ..., \beta_m$ и имеет вид

$$y = f(x, \beta_1, \beta_2, ..., \beta_m)\,. \tag{1.58}$$

Для отыскания неизвестных параметров проведено n наблюдений (x_i, Y_i), где $i = 1, 2, ..., n$. Но так как результаты наблюдений не свободны от погрешностей измерений, которые мы будем рассматривать как случайные ошибки, то по ним нельзя точно найти искомые параметры. Поэтому приходится ставить задачу об отыскании не значений параметров, а их оценок по результатам эксперимента.

Будем предполагать, что значения аргументов x_i известны точно, а значения функции Y_i – взаимно независимые случайные величины, включающие случайные ошибки Z_i, т.е. $Y_i = f(x_i, \beta_1, \beta_2, ..., \beta_m) + Z_i$, где

$$M(Z_i) = 0; \quad D(Z_i) = D(Y_i) = \sigma^2. \tag{1.59}$$

Здесь предполагаем, что измерения *равноточны*. Для оценок параметров $\beta_1, \beta_2, ..., \beta_m$ используется *метод наименьших квадратов*. В качестве оценок этих параметров принимают значения $\tilde{\beta}_1, \tilde{\beta}_2, ..., \tilde{\beta}_m$, при которых имеет минимум функция (МНК-оценки)

$$Q(\beta_1, \beta_2, ..., \beta_m) = \sum_{i=1}^{n} (Y_i - f(x_i, \beta_1, \beta_2, ..., \beta_m))^2. \tag{1.60}$$

Уравнение (1.58) называют *уравнением регрессии*, а отыскание оценок параметров и исследование получаемых моделей – *регрессионным анализом*.

Будем рассматривать уравнения регрессии, линейные относительно оцениваемых параметров $\beta_1, \beta_2, ..., \beta_m$:

$$y = f(x, \beta_1, \beta_2, ..., \beta_m) = \beta_1 \varphi_1(x) + \beta_2 \varphi_2(x) + ... + \beta_m \varphi_m(x). \tag{1.61}$$

Функции $\varphi_1(x), \varphi_2(x), ..., \varphi_m(x)$ называют *базисными функциями*, их рассматривают на множестве точек $\{x_1, x_2, ..., x_n\}$, где n – число экспериментов. Функция Q (1.60) в этом случае запишется в виде

$$Q(\beta_1, \beta_2, ..., \beta_m) = \sum_{i=1}^{n} (Y_i - \beta_1 \varphi_1(x_i) - \beta_2 \varphi_2(x_i) - ... - \beta_m \varphi_m(x_i))^2. \tag{1.62}$$

Для нахождения минимума найдем частные производные функции $Q(\beta_1, \beta_2, ..., \beta_m)$ по переменным $\beta_1, \beta_2, ..., \beta_m$ и приравняем их к нулю (необходимые условия минимума функции). Получим систему уравнений

$$\begin{cases} \dfrac{\partial Q}{\partial \beta_1} = -2\sum_{i=1}^{n} (Y_i - \beta_1 \varphi_1(x_i) - \beta_2 \varphi_2(x_i) - ... - \beta_m \varphi_m(x_i)) \varphi_1(x_i) = 0, \\ \dfrac{\partial Q}{\partial \beta_2} = -2\sum_{i=1}^{n} (Y_i - \beta_1 \varphi_1(x_i) - \beta_2 \varphi_2(x_i) - ... - \beta_m \varphi_m(x_i)) \varphi_2(x_i) = 0, \\ \qquad\qquad\qquad\qquad .... \\ \dfrac{\partial Q}{\partial \beta_m} = -2\sum_{i=1}^{n} (Y_i - \beta_1 \varphi_1(x_i) - \beta_2 \varphi_2(x_i) - ... - \beta_m \varphi_m(x_i)) \varphi_m(x_i) = 0, \end{cases}$$

которую после преобразований можно записать в виде

$$\begin{cases} \beta_1 \sum_{i=1}^{n} \varphi_1{}^2(x_i) + \beta_2 \sum_{i=1}^{n} \varphi_1(x_i)\varphi_2(x_i) + \ldots + \beta_m \sum_{i=1}^{n} \varphi_1(x_i)\varphi_m(x_i) = \sum_{i=1}^{n} Y_i \varphi_1(x_i), \\ \beta_1 \sum_{i=1}^{n} \varphi_1(x_i)\varphi_2(x_i) + \beta_2 \sum_{i=1}^{n} \varphi_2{}^2(x_i) + \ldots + \beta_m \sum_{i=1}^{n} \varphi_2(x_i)\varphi_m(x_i) = \sum_{i=1}^{n} Y_i \varphi_2(x_i), \\ \ldots\ldots\ldots\ldots \\ \beta_1 \sum_{i=1}^{n} \varphi_1(x_i)\varphi_m(x_i) + \beta_2 \sum_{i=1}^{n} \varphi_2(x_i)\varphi_m(x_i) + \ldots + \beta_m \sum_{i=1}^{n} \varphi_m{}^2(x_i) = \sum_{i=1}^{n} Y_i \varphi_m(x_i). \end{cases} \quad (1.63)$$

Следовательно, оценки параметров $\tilde{\beta}_1, \tilde{\beta}_2, \ldots, \tilde{\beta}_m$ являются решениями линейной алгебраической системы m уравнений (1.63).

Введем обозначения: $\sum_{i=1}^{n} \varphi_j(x_i)\varphi_k(x_i) = (\varphi_j, \varphi_k)$; $\sum_{i=1}^{n} Y_i \varphi_k(x_i) = (Y, \varphi_k)$, тогда система (1.63) запишется в виде

$$\begin{cases} \beta_1(\varphi_1, \varphi_1) + \beta_2(\varphi_1, \varphi_2) + \ldots + \beta_m(\varphi_1, \varphi_m) = (Y, \varphi_1), \\ \beta_1(\varphi_1, \varphi_2) + \beta_2(\varphi_2, \varphi_2) + \ldots + \beta_m(\varphi_2, \varphi_m) = (Y, \varphi_2), \\ \ldots\ldots\ldots\ldots \\ \beta_1(\varphi_1, \varphi_m) + \beta_2(\varphi_2, \varphi_m) + \ldots + \beta_m(\varphi_m, \varphi_m) = (Y, \varphi_m). \end{cases}$$

С использованием следующих матричных обозначений:

$A = \begin{pmatrix} (\varphi_1, \varphi_1) & (\varphi_1, \varphi_2) & \ldots & (\varphi_1, \varphi_m) \\ (\varphi_1, \varphi_2) & (\varphi_2, \varphi_2) & \ldots & (\varphi_2, \varphi_m) \\ \ldots & \ldots & \ldots & \ldots \\ (\varphi_1, \varphi_m) & (\varphi_2, \varphi_m) & \ldots & (\varphi_m, \varphi_m) \end{pmatrix}$ – матрица коэффициентов при неизвестных,

$Y = \begin{pmatrix} (Y, \varphi_1) \\ (Y, \varphi_2) \\ \ldots \\ (Y, \varphi_m) \end{pmatrix}$ – вектор правых частей, $\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \ldots \\ \beta_m \end{pmatrix}$ – вектор параметров,

система (1.63) принимает вид

$$A\beta = Y. \quad (1.64)$$

При условии, что A – невырожденная матрица, решение системы (1.64) можно записать в виде

$$\tilde{\beta} = A^{-1} Y \, , \tag{1.65}$$

где $\tilde{\beta} = \begin{pmatrix} \tilde{\beta}_1 \\ \tilde{\beta}_2 \\ \dots \\ \tilde{\beta}_m \end{pmatrix}$ – вектор МНК-оценок параметров регрессионной модели (1.61).

Оценки параметров линейной регрессии, получаемые по методу наименьших квадратов, имеют следующие свойства:

1) они являются линейными функциями результатов наблюдений $Y_i, i = 1, 2, \dots, n$, и несмещенными оценками параметров, т.е. $M(\tilde{\beta}_j) = \beta_j, \quad j = 1, 2, \dots, m;$

2) они имеют минимальные дисперсии в классе несмещенных оценок, являющихся линейными функциями результатов наблюдений.

1.8.2. Проверка гипотезы об адекватности регрессионной модели

Регрессионная модель называется *адекватной*, если предсказанные по ней значения переменной Y согласуются с результатами эксперимента. Если модель адекватна, то отклонения результатов эксперимента от полученной функции регрессии $\Delta Y_i = Y_i - \tilde{Y}(x_i)$ являются реализациями случайных ошибок эксперимента Z_i, которые, в силу предположений (1.58), должны быть независимыми нормально распределенными случайными величинами с нулевыми средними и одинаковыми дисперсиями σ^2. Проверка выполнения этих предположений осуществляется статистическими методами и лежит в основе оценки адекватности модели регрессии.

Для проверки адекватности регрессионной модели вычисляют остаточную дисперсию (так называемую дисперсию адекватности) по формуле

$$S^2_{\text{ад}} = \frac{\sum_{i=1}^{n} (\Delta Y_i)^2}{k_{\text{ад}}}; \quad k_{\text{ад}} = n - m \, , \tag{1.66}$$

где ΔY_i – отклонения результатов экперимента Y_i от проверяемой модели регрессии; $k_{\text{ад}}$ – число степеней свободы дисперсии адекватности; n – число точек, в которых проводился эксперимент; m – число оцениваемых параметров β_j в проверяемой модели.

Если истинная функция регрессии имеет тот же вид, что и рассматриваемая модель (например, так же, как и модель, представляет собой квадратичную функцию), то дисперсия адекватности служит несмещенной оценкой истинной дисперсии эксперимента и ее можно сравнивать с другими подобными оценками. В частности, может быть проведена независимая серия измерений для получения оценки дисперсии эксперимента $S^2_{\text{эксп}}$. В этом случае $S^2_{\text{эксп}}$ оценивает дисперсию эксперимента $D_{\text{эксп}}$, $S^2_{\text{ад}}$ характеризует степень отклонения экспериментальных точек от регрессионной модели, т.е. оценивает некую дисперсию адекватности $D_{\text{ад}}$. Проверка адекватности модели заключается в проверке гипотезы H_0: $D_{\text{ад}} = D_{\text{эксп}}$ при альтернативной гипотезе H_1: $D_{\text{ад}} > D_{\text{эксп}}$ (если модель неадекватна, отклонения экспериментальных точек от модели будут больше погрешностей эксперимента). Таким образом, задача сводится к проверке гипотезы о равенстве дисперсий, которая решается с помощью критерия Фишера. Вычисляем отношение

$$F = S^2_{\text{ад}} / S^2_{\text{эксп}}. \tag{1.67}$$

Если при заданном уровне значимости α отношение F окажется меньше квантили $F_{1-\alpha}(k_1, k_2)$, где $k_1 = k_{\text{ад}}$, $k_2 = k_{\text{эксп}}$, то рассматриваемая модель не противоречит результатам эксперимента и принимается; в противоположном случае модель отвергается с уровнем значимости α, как противоречащая результатам эксперимента.

1.8.3. Построение доверительных интервалов для коэффициентов регрессии

Оценки параметров регрессии $\tilde{\beta}_j$, определяемые формулами (1.65), являются точечными оценками истинных значений параметров β_j. Если результаты экспериментов независимы и подчиняются нормальному закону распределения с дисперсией σ^2, то доверительный интервал с доверительной вероятностью $P = 1 - \alpha$ для каждого параметра $\tilde{\beta}_j$ можно определить неравенством

$$\left|\beta_j - \tilde{\beta}_j\right| < \varepsilon_j \text{ или } \beta_j = \tilde{\beta}_j \pm \varepsilon_j,$$

где $$\varepsilon_j = t_{1-\alpha/2}(k) S \sqrt{a_{jj}} \quad (j = 1, 2, ..., m), \tag{1.68}$$

здесь S^2 – несмещенная оценка дисперсии σ^2 с числом степеней свободы k; $t_{1-\alpha/2}(k)$ – квантиль распределения Стьюдента; a_{jj} – диагональный элемент матрицы A^{-1}, $\alpha = 1 - P$, P – доверительная вероятность.

Как было указано выше, если рассматриваемая модель регрессии адекватна, то дисперсия адекватности служит несмещенной оценкой истинной дисперсии эксперимента. Следовательно, в случае адекватности модели в формулу (1.68) в качестве оценки среднего квадратического отклонения S можно подставить корень из дисперсии адекватности $S_{\text{ад}} = \sqrt{S_{\text{ад}}^2}$.

В построенной регрессионной модели (1.61) некоторые коэффициенты могут быть незначимы, т.е. может выполняться гипотеза H_0: $\beta_j = 0$. Для проверки этой гипотезы можно найти доверительный интервал для коэффициента β_j с уровнем значимости α. Если этот интервал «накрывает» значение $\beta_j = 0$, гипотеза H_0 принимается и коэффициент β_j признается незначимым, в противоположном случае коэффициент β_j значим.

1.8.4. Задача регрессии для линейной функции

Рассмотрим случай, когда уравнение регрессии (1.61) является линейной функцией

$$y = \beta_1 + \beta_2 x, \tag{1.69}$$

т.е. базисные функции $\varphi_1(x) = 1$, $\varphi_2(x) = x$. В этом случае система (1.63) имеет вид

$$\begin{cases} \beta_1 n + \beta_2 \sum\limits_{i=1}^{n} x_i = \sum\limits_{i=1}^{n} Y_i, \\ \beta_1 \sum\limits_{i=1}^{n} x_i + \beta_2 \sum\limits_{i=1}^{n} x_i^2 = \sum\limits_{i=1}^{n} Y_i x_i. \end{cases} \tag{1.70}$$

Расчет упростится, если ввести замену $X = \dfrac{x - \bar{x}}{h}$ и рассматривать уравнение

$$y = B_1 + B_2 X = B_1 + B_2 \frac{x - \bar{x}}{h}, \qquad (1.71)$$

где $\bar{x} = \dfrac{1}{n}\sum_{i=1}^{n} x_i$ – среднее арифметическое аргументов x; h выбирается из условия, чтобы значения X были целыми не имеющими общего множителя. Уравнение (1.71) будем называть уравнением с *кодированным переменным*, в отличие от уравнения (1.69) с *реальным переменным*. В этом случае $\sum_{i=1}^{n} X_i = 0$ и система (1.70) будет иметь вид

$$\begin{cases} B_1 n = \sum_{i=1}^{n} Y_i, \\ B_2 \sum_{i=1}^{n} x_i^{\,2} = \sum_{i=1}^{n} Y_i x_i. \end{cases}$$

Откуда имеем формулы для оценок коэффициентов регрессии уравнения с кодированным переменным:

$$\tilde{B}_1 = \frac{\sum_{i=1}^{n} Y_i}{n}, \; \tilde{B}_2 = \frac{\sum_{i=1}^{n} Y_i X_i}{\sum_{i=1}^{n} X_i^{\,2}}. \qquad (1.72)$$

Для контроля расчетов удобно воспользоваться свойством отклонений $\Delta Y_i = Y_i - \tilde{Y}(x_i)$ экспериментальных результатов Y_i от рассчитанных по оценкам (1.72) значений функции регрессии $\tilde{Y}(x_i) = \tilde{B}_1 + \tilde{B}_2 x_i$:

$$\sum_{i=1}^{n} \Delta Y_i = 0 . \qquad (1.73)$$

Дисперсия адекватности (1.66) для проверки адекватности линейной регрессионной модели вычисляется по формуле

$$S^2_{\text{ад}} = \frac{\sum_{i=1}^{n}(\Delta Y_i)^2}{k_{\text{ад}}}; \quad k_{\text{ад}} = n-2\,. \tag{1.74}$$

Границы доверительных интервалов для параметров линейной функции регрессии с кодированным переменным (1.71) имеют вид

$$\tilde{B}_1 \pm \varepsilon_1; \quad \varepsilon_1 = t_{1-\frac{\alpha}{2}}(k)\frac{S}{\sqrt{n}}; \quad \tilde{B}_2 \pm \varepsilon_2; \quad \varepsilon_2 = t_{1-\frac{\alpha}{2}}(k)\frac{S}{\sqrt{\sum_{i=1}^{n} X_i^2}}. \tag{1.75}$$

Оценки коэффициентов регрессии линейной функции (1.69) с реальным переменным при этом могут быть найдены по формулам

$$\tilde{\beta}_1 = \tilde{B}_1 - \tilde{B}_2\frac{\overline{x}}{h}; \quad \tilde{\beta}_2 = \frac{\tilde{B}_2}{h}. \tag{1.76}$$

Границы доверительных интервалов для коэффициентов линейной функции с реальным переменным (1.69) имеют вид

$$\tilde{\beta}_1 \pm \hat{\varepsilon}_1; \qquad \hat{\varepsilon}_1 = t_{1-\frac{\alpha}{2}}(k) S\sqrt{\frac{1}{n} + \frac{\overline{x}^2}{h^2\sum_{i=1}^{n} X_i^2}};$$

$$\tilde{\beta}_2 \pm \hat{\varepsilon}_2; \qquad \hat{\varepsilon}_2 = t_{1-\frac{\alpha}{2}}(k)\frac{S}{h\sqrt{\sum_{i=1}^{n} X_i^2}}. \tag{1.77}$$

1.8.5. Задача регрессии для квадратичной функции

Рассмотрим случай, когда уравнение регрессии (1.61) является квадратичной функцией

$$y = \beta_1 + \beta_2 x + \beta_3 x^2, \tag{1.78}$$

т.е. базисные функции $\varphi_1(x) = 1$, $\varphi_2(x) = x$, $\varphi_3(x) = x^2$. В этом случае система (1.63) имеет вид

$$
\begin{cases}
\beta_1 n + \beta_2 \sum_{i=1}^{n} x_i + \beta_3 \sum_{i=1}^{n} x_i^{\,2} = \sum_{i=1}^{n} Y_i, \\
\beta_1 \sum_{i=1}^{n} x_i + \beta_2 \sum_{i=1}^{n} x_i^{\,2} + \beta_3 \sum_{i=1}^{n} x_i^{\,3} = \sum_{i=1}^{n} Y_i x_i, \\
\beta_1 \sum_{i=1}^{n} x_i^{\,2} + \beta_2 \sum_{i=1}^{n} x_i^{\,3} + \beta_3 \sum_{i=1}^{n} x_i^{\,4} = \sum_{i=1}^{n} Y_i x_i^{\,2}.
\end{cases}
\tag{1.79}
$$

Как и в предыдущем параграфе, сделаем замену $X = \dfrac{x - \overline{x}}{h}$, где $\overline{x} = \dfrac{1}{n}\sum_{i=1}^{n} x_i$ – среднее арифметическое аргументов x; h выбираем из условия, чтобы значения X были целыми, не имеющими общего множителя, т.е. преобразуем уравнение (1.78) к виду

$$
y = B_1 + B_2 X + B_3 X^2 = B_1 + B_2 \frac{x - \overline{x}}{h} + B_3 \left(\frac{x - \overline{x}}{h} \right)^2 . \tag{1.80}
$$

В этом случае $\sum_{i=1}^{n} X_i = 0$. Кроме того, введем условие $\sum_{i=1}^{n} X_i^{\,3} = 0$. Оно будет выполняться, например, в том случае, когда переменная x в исходных данных меняется с постоянным шагом. Тогда система (1.79) существенно упрощается:

$$
\begin{cases}
B_1 n + B_3 \sum_{i=1}^{n} x_i^{\,2} = \sum_{i=1}^{n} Y_i, \\
B_2 \sum_{i=1}^{n} x_i^{\,2} = \sum_{i=1}^{n} Y_i x_i, \\
B_1 \sum_{i=1}^{n} x_i^{\,2} + B_3 \sum_{i=1}^{n} x_i^{\,4} = \sum_{i=1}^{n} Y_i x_i^{\,2}.
\end{cases}
\tag{1.81}
$$

Решая эту систему, получаем формулы для оценок коэффициентов регрессии уравнения с кодированным переменным (1.80):

$$\tilde{B}_1 = \frac{\sum\limits_{i=1}^{n} X_i^{\,4} \cdot \sum\limits_{i=1}^{n} Y_i - \sum\limits_{i=1}^{n} X_i^{\,2} \cdot \sum\limits_{i=1}^{n} Y_i X_i^2}{n\sum\limits_{i=1}^{n} X_i^{\,4} - \left(\sum\limits_{i=1}^{n} X_i^{\,2}\right)^2}; \; \tilde{B}_2 = \frac{\sum\limits_{i=1}^{n} Y_i X_i}{\sum\limits_{i=1}^{n} X_i^{\,2}};$$

$$\tilde{B}_3 = \frac{n\sum\limits_{i=1}^{n} Y_i X_i^2 - \sum\limits_{i=1}^{n} X_i^{\,2} \cdot \sum\limits_{i=1}^{n} Y_i}{n\sum\limits_{i=1}^{n} X_i^{\,4} - \left(\sum\limits_{i=1}^{n} X_i^{\,2}\right)^2}. \qquad (1.82)$$

Формула (1.73) для квадратичной модели регрессии также имеет место, и ее удобно использовать для контроля расчетов. Дисперсия адекватности (1.66) для проверки адекватности квадратичной функции регрессии вычисляется по формуле

$$S^2_{\text{ад}} = \frac{\sum\limits_{i=1}^{n} (\Delta Y_i)^2}{k_{\text{ад}}}; \quad k_{\text{ад}} = n - 3. \qquad (1.83)$$

Границы доверительных интервалов для параметров квадратичной функции регрессии с кодированным переменным (1.80) имеют вид

$$\tilde{B}_1 \pm \varepsilon_1; \qquad \varepsilon_1 = t_{1-\frac{\alpha}{2}}(k) S \sqrt{\frac{\sum\limits_{i=1}^{n} X_i^{\,4}}{n\sum\limits_{i=1}^{n} X_i^{\,4} - \left(\sum\limits_{i=1}^{n} X_i^{\,2}\right)^2}};$$

$$\tilde{B}_2 \pm \varepsilon_2; \qquad \varepsilon_2 = t_{1-\frac{\alpha}{2}}(k) \frac{S}{\sqrt{\sum\limits_{i=1}^{n} X_i^{\,2}}}; \qquad (1.84)$$

$$\tilde{B}_3 \pm \varepsilon_3; \qquad \varepsilon_3 = t_{1-\frac{\alpha}{2}}(k) S \sqrt{\frac{n}{n\sum\limits_{i=1}^{n} X_i^{\,4} - \left(\sum\limits_{i=1}^{n} X_i^{\,2}\right)^2}}.$$

Оценки коэффициентов регрессии квадратичной функции с реальным переменным (1.78) при этом будут

$$\tilde{\beta}_1 = \tilde{B}_1 - \tilde{B}_2 \frac{\overline{x}}{h} + \tilde{B}_3 \left(\frac{\overline{x}}{h}\right)^2; \ \tilde{\beta}_2 = \frac{\tilde{B}_2}{h} - \frac{2\tilde{B}_3 \overline{x}}{h^2}; \ \tilde{\beta}_3 = \frac{\tilde{B}_3}{h^2}. \qquad (1.85)$$

Границы доверительных интервалов для параметров квадратичной функции с реальным переменным (1.78) имеют вид

$$\tilde{\beta}_1 \pm \hat{\varepsilon}_1; \qquad \hat{\varepsilon}_1 = t_{1-\frac{\alpha}{2}}(k)\, S \sqrt{\frac{\sum\limits_{i=1}^{n} X_i^4 + \frac{\overline{x}^4}{h^4} n}{n\sum\limits_{i=1}^{n} X_i^4 - \left(\sum\limits_{i=1}^{n} X_i^2\right)^2} + \frac{\overline{x}^2}{h^2 \sum\limits_{i=1}^{n} X_i^2}};$$

$$\tilde{\beta}_2 \pm \hat{\varepsilon}_2; \qquad \hat{\varepsilon}_2 = t_{1-\frac{\alpha}{2}}(k)\, \frac{S}{h} \sqrt{\frac{1}{\sum\limits_{i=1}^{n} X_i^2} + \frac{4n\overline{x}^2}{h^2 \left(n\sum\limits_{i=1}^{n} X_i^4 - \left(\sum\limits_{i=1}^{n} X_i^2\right)^2\right)}}; \quad (1.86)$$

$$\tilde{\beta}_3 \pm \hat{\varepsilon}_3; \qquad \hat{\varepsilon}_3 = t_{1-\frac{\alpha}{2}}(k)\, \frac{S}{h^2} \sqrt{\frac{n}{n\sum\limits_{i=1}^{n} X_i^4 - \left(\sum\limits_{i=1}^{n} X_i^2\right)^2}}.$$

Задача 1.25. Результаты экспериментов представлены в первых двух столбцах табл. 1.11. Экспериментальные значения Y являются независимыми и равноточными. Построить линейную и квадратичную регрессионные модели.

Таблица 1.11

Условие задачи 1.25 и результаты расчета

	x	Y	$X = \dfrac{x - 0{,}6}{0{,}2}$	$X{\cdot}Y$	X^2	$Y_{\text{лин}}$	$\Delta Y_{\text{лин}}$	$\Delta Y_{\text{лин}}^2$
	0,2	4,5	–2	–9,0	4	5,3	–0,8	0,64
	0,4	7,0	–1	–7,0	1	6,25	0,75	0,56
	0,6	8,0	0	0,0	0	7,2	0,8	0,64
	0,8	7,5	1	7,5	1	8,15	–0,65	0,42
	1,0	9,0	2	18,0	4	9,1	–0,1	0,04
Σ	3,0	36,0	0	9,5	10		0	2,30

Решение

Сначала найдем решение задачи регрессии в кодированных значениях переменной x (1.71). Введем новую переменную по формуле $X = \dfrac{x - \overline{x}}{h}$, где $\overline{x} = \dfrac{1}{n}\sum_{i=1}^{n} x_i = \dfrac{1}{5}\cdot 3 = 0,6$. Если значения величины $x - 0,6$ поделить на число $h = 0,2$, то получатся целые значения, не имеющие общего множителя. Поэтому $X = \dfrac{x - 0,6}{0,2}$.

По формулам (1.72) находим оценки коэффициентов линейной регрессии

$$\tilde{B}_1 = \frac{\sum_{i=1}^{n} Y_i}{n} = \frac{36}{5} = 7,2; \quad \tilde{B}_2 = \frac{\sum_{i=1}^{n} Y_i X_i}{\sum_{i=1}^{n} X_i^2} = \frac{9,5}{10} = 0,95.$$

Получили линейную модель регрессии

$$\widehat{Y}_{\text{лин}} = 7,2 + 0,95X.$$

По полученной формуле вычисляем значения линейной функции регрессии $\widehat{Y}_{\text{лин}}$ при всех значениях аргумента X, а затем рассчитываем $\Delta Y_i = \widehat{Y}_i - \widehat{Y}_{i\,\text{лин}}$ отклонения экспериментальных значений Y_i от значений $\widehat{Y}_{i\,\text{лин}}$, полученных по функции регрессии. Контроль, согласно формуле $\sum_{i=1}^{n} \Delta Y_i = 0$, выполнен. Все расчеты приведены в табл. 1.11.

Уравнение линейной регрессии Y от реального переменного x найдем, сделав преобразование

$$Y_{\text{лин}} = \beta_1 + \beta_2 x = 7,2 + 0,95\frac{x - 0,6}{0,2} = 4,35 + 4,75x.$$

Построим квадратичную модель регрессии (1.80):

$$y = B_1 + B_2 X + B_3 X^2 = B_1 + B_2\frac{x - 0,6}{0,2} + B_3\left(\frac{x - 0,6}{0,2}\right)^2.$$

Заметим, что в нашем случае $\sum_{i=1}^{n} X_i^3 = 0$, поэтому для оценок коэффициентов регрессии можно воспользоваться формулами (1.82). Все расчеты приведены в табл. 1.12.

Таблица 1.12

Результаты расчета квадратичной модели регрессии (к задаче 1.25)

	X^3	YX^2	X^4	$Y_{кв}$	$\Delta Y_{кв}$	$\Delta Y_{кв}^{\ 2}$
	–8	18	16	4,8	–0,3	0,09
	–1	7	1	6,5	0,5	0,25
	0	0	0	7,7	0,3	0,09
	1	7,5	1	8,4	–0,9	0,81
	8	36	16	8,6	0,4	0,16
Σ	0	68,5	34		0	1,4

$$\tilde{B}_1 = \frac{\sum_{i=1}^{n} X_i^4 \cdot \sum_{i=1}^{n} Y_i - \sum_{i=1}^{n} X_i^2 \cdot \sum_{i=1}^{n} Y_i X_i^2}{n\sum_{i=1}^{n} X_i^4 - \left(\sum_{i=1}^{n} X_i^2\right)^2} = \frac{34 \cdot 36 - 10 \cdot 68,5}{5 \cdot 34 - 100} = 7,7;$$

$$\tilde{B}_2 = \frac{\sum_{i=1}^{n} Y_i X_i}{\sum_{i=1}^{n} X_i^2} = \frac{9,5}{10} = 0,95;$$

$$\tilde{B}_3 = \frac{n\sum_{i=1}^{n} Y_i X_i^2 \cdot - \sum_{i=1}^{n} X_i^2 \cdot \sum_{i=1}^{n} Y_i}{n\sum_{i=1}^{n} X_i^4 - \left(\sum_{i=1}^{n} X_i^2\right)^2} = \frac{5 \cdot 68,5 - 10 \cdot 36}{5 \cdot 34 - 100} = -0,25\,.$$

Получили квадратичную модель регрессии

$$\widehat{Y}_{кв} = 7,7 + 0,95X - 0,25X^2.$$

Условие $\sum_{i=1}^{n} \Delta Y_i = 0$ выполняется.

Уравнение квадратичной регрессии Y от реального переменного x найдем, сделав преобразование

$$Y_{\text{кв}} = \beta_1 + \beta_2 x + \beta_3 x^2 = 7,7 + 0,95\frac{x-0,6}{0,2} - 0,25\left(\frac{x-0,6}{0,2}\right)^2 =$$

$$= 2,6 + 12,25x - 6,25x^2.$$

Пригодность каждой из полученных регрессионных моделей будет оценена ниже.

Задача 1.26. В задаче 1.25 рассчитаны линейная и квадратичная модели регрессии. По отдельной независимой серии измерений получена несмещенная оценка дисперсии $S^2 = 0,32$ с числом степеней свободы k, равным 20. Проверить адекватность линейной и квадратичной моделей регрессии с уровнем значимости $\alpha = 0,05$.

Решение

Для нахождения дисперсии адекватности (1.66) необходимо вычислить сумму квадратов отклонений результатов эксперимента от функции регрессии $\sum \Delta Y^2$ для каждой модели регрессии. Для линейной модели эта сумма равна 2,3 (см. последний столбец табл. 1.11). Число точек, в которых проводился эксперимент, $n = 5$; число оцениваемых параметров $m = 2$, тогда $k_{\text{ад}} = 5 - 2 = 3$. Дисперсия адекватности (1.74) равна $S^2_{\text{ад лин}} = 2,3/3 = 0,767$. Для проверки гипотезы об адекватности линейной модели вычисляем критерий Фишера (1.67):

$F = S^2_{\text{ад}} / S^2_{\text{эксп}} = 0,767/0,32 = 2,40$. Квантиль распределения Фишера $F_{0,95}(3; 20) = 3,10$. Так как $2,4 < 3,1$, то гипотеза об адекватности линейной модели принимается с уровнем значимости $\alpha = 0,05$.

Адекватность квадратичной модели проверяется аналогично. Сумма квадратов отклонений $\sum \Delta Y^2$ (см. последний столбец табл. 1.12) равна 1,4. $n = 5$; $m = 3$; $k_{\text{ад}} = 5 - 3 = 2$; $S^2_{\text{ад кв}} = 1,4/2 = 0,7$, $F = S^2_{\text{ад}} / S^2_{\text{эксп}} = 0,7/0,32 = 2,19$. $F_{0,95}(2; 20) = 3,49$. Так как $2,19 < 3,49$, гипотеза об адекватности квадратичной модели принимается с уровнем значимости $\alpha = 0,05$.

Квадратичную модель в данной задаче строить нецелесообразно. Так как линейная модель оказалась адекватной, ее уточнять не было необходимости.

Задача 1.27. В задаче 1.25 получено уравнение линейной и квадратичной модели регрессии. По отдельной независимой серии измерений получена несмещенная оценка дисперсии $S^2 = 0{,}32$ с числом степеней свободы k равным 20. Считая результаты экспериментов независимыми, равноточными и подчиняющимися нормальному закону распределения, построить доверительные интервалы для истинных значений коэффициентов обеих моделей с надежностью $P = 0{,}95$.

Решение

В задаче 1.25 получена линейная модель регрессии от кодированной переменной

$$y = \widehat{B}_1 + \widehat{B}_2 X = 7{,}2 + 0{,}95X.$$

Границы доверительных интервалов (1.75) для коэффициентов B_1 и B_2 при этом будут:

$$B_1 = 7{,}2 \pm \varepsilon_1;$$

$$\varepsilon_1 = t_{1-\frac{\alpha}{2}}(k)\frac{S}{\sqrt{n}} = t_{0{,}975}(20)\frac{\sqrt{0{,}32}}{\sqrt{5}} = 2{,}086\frac{\sqrt{0{,}32}}{\sqrt{5}} = 0{,}52;$$

$$B_2 = 0{,}95 \pm \varepsilon_2; \qquad \varepsilon_2 = t_{1-\frac{\alpha}{2}}(k)\frac{S}{\sqrt{\sum\limits_{i=1}^{n} X_i^2}} = 2{,}086\frac{\sqrt{0{,}32}}{\sqrt{10}} = 0{,}38.$$

Уравнение линейной регрессии от реального переменного x

$$y = \beta_1 + \beta_2 x = 4{,}35 + 4{,}75x.$$

Границы доверительных интервалов для коэффициентов β_1 и β_2 находим по формулам (1.77):

$$\beta_1 = 4{,}35 \pm \hat{\varepsilon}_1;$$

$$\hat{\varepsilon}_1 = t_{1-\frac{\alpha}{2}}(k)\, S\sqrt{\frac{1}{n} + \frac{\overline{x}^2}{h^2\sum\limits_{i=1}^{n} X_i^2}} = 2{,}086\sqrt{0{,}32}\cdot\sqrt{\frac{1}{5} + \frac{0{,}36}{0{,}04\cdot 10}} = 1{,}24;$$

$$\beta_2 = 4{,}75 \pm \hat{\varepsilon}_2; \qquad \hat{\varepsilon}_2 = t_{1-\frac{\alpha}{2}}(k)\frac{S}{h\sqrt{\sum\limits_{i=1}^{n} X_i^2}} = 2{,}086\frac{\sqrt{0{,}32}}{0{,}2\sqrt{10}} = 1{,}9.$$

Квадратичная модель регрессии с кодированным переменным

$$y = B_1 + B_2 X + B_3 X^2 = 7,7 + 0,95X - 0,25X^2.$$

Границы доверительных интервалов (1.84) для коэффициентов B_1, B_2 и B_3:

$$B_1 = 7,7 \pm \varepsilon_1; \qquad \varepsilon_1 = t_{1-\frac{\alpha}{2}}(k) S \sqrt{\frac{\sum_{i=1}^{n} X_i^4}{n\sum_{i=1}^{n} X_i^4 - \left(\sum_{i=1}^{n} X_i^2\right)^2}} =$$

$$= 2,086\sqrt{0,32} \cdot \sqrt{\frac{34}{5 \cdot 34 - 100}} = 0,822;$$

$$B_2 = 0,95 \pm \varepsilon_2; \qquad \varepsilon_2 = t_{1-\frac{\alpha}{2}}(k) S \frac{1}{\sqrt{\sum_{i=1}^{n} X_i^2}} = 2,086\sqrt{0,32} \cdot \frac{1}{\sqrt{10}} = 0,373;$$

$$B_3 = -0,1774 \pm \varepsilon_3; \qquad \varepsilon_3 = t_{1-\frac{\alpha}{2}}(k) S \sqrt{\frac{n}{n\sum_{i=1}^{n} X_i^4 - \left(\sum_{i=1}^{n} X_i^2\right)^2}} =$$

$$= 2,086\sqrt{0,32} \cdot \sqrt{\frac{5}{5 \cdot 34 - 100}} = 0,315.$$

Полуширина доверительного интервала для коэффициента B_3 оказалась больше абсолютной величины этого коэффициента, т.е. доверительный интервал для этого коэффициента накрывает значение $B_3 = 0$. В этом случае говорят, что коэффициент $\tilde{B}_3$ незначим, т.е. переход от линейной модели к квадратичной необоснован и следовало остановиться на линейной модели. Аналогичный вывод был получен при проверке адекватности полученных моделей регрессии.

1.9. Линейный корреляционный анализ

1.9.1. Двумерный случайный вектор, его выборочные характеристики

Рассмотрим систему двух случайных величин или *двумерный случайный вектор* $(X,Y)^T$ с центром распределения $\begin{pmatrix} M(X) \\ M(Y) \end{pmatrix} = \begin{pmatrix} a_x \\ a_y \end{pmatrix}$ и ковариационной матрицей

$$K = \begin{pmatrix} D(X) & K_{xy} \\ K_{xy} & D(Y) \end{pmatrix}, \qquad (1.87)$$

где a_x и a_y – математические ожидания; $D(X) = \sigma_x^2$ и $D(Y) = \sigma_y^2$ – дисперсии случайных величин X и Y соответственно. K_{xy} – ковариация между величинами X и Y, определяется следующим образом:

$$K_{xy} = \operatorname{cov}(X,Y) = M[(X - a_x)(Y - a_y)], \qquad (1.88)$$

В качестве нормированной ковариации вводится коэффициент корреляции

$$\rho_{xy} = \frac{K_{xy}}{\sigma_x \sigma_y}, \qquad (1.89)$$

который характеризует степень *линейной зависимости* между случайными величинами X и Y.

Свойства коэффициента корреляции:

1. Коэффициент корреляции является безразмерным коэффициентом, не зависящим от начала отсчета величин X и Y.

2. $-1 \le \rho_{xy} \le 1$.

3. Если $|\rho_{xy}| = 1$, случайные величины X и Y *связаны линейной функциональной зависимостью.*

4. Если $\rho_{xy} = 0$, случайные величины X и Y *некоррелированы*, т.е. между ними *отсутствует линейная зависимость*.

5. Чем ближе $|\rho_{xy}|$ к единице, тем сильнее линейная зависимость между X и Y. Чем ближе $|\rho_{xy}|$ к нулю, тем слабее линейная зависимость между X и Y.

6. Если $\rho_{xy} > 0$, то с увеличением одной случайной величины математическое ожидание (среднее значение) другой увеличивается, если $\rho_{xy} < 0$, то с увеличением одной случайной величины математическое ожидание (среднее значение) другой уменьшается.

Случайный вектор $(U_1, U_2)^T$ имеет *стандартное нормальное распределение*, если его координаты U_1 и U_2 взаимно независимы и имеют стандартное нормальное распределение. Его центр распределения совпадает с началом координат, а матрица ковариаций является единичной матрицей.

Случайный вектор $(X, Y)^T$ имеет *двумерное нормальное распределение*, если его можно представить в виде

$$\begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} a_x \\ a_y \end{pmatrix} + B \begin{pmatrix} U_1 \\ U_2 \end{pmatrix},$$

где $\begin{pmatrix} a_x \\ a_y \end{pmatrix}$ – числовой вектор, $\begin{pmatrix} U_1 \\ U_2 \end{pmatrix}$ – случайный вектор, имеющий стандартное нормальное распределение; B – невырожденная матрица второго порядка. Центром распределения вектора $\begin{pmatrix} X \\ Y \end{pmatrix}$ является вектор $\begin{pmatrix} a_x \\ a_y \end{pmatrix}$, матрица ковариаций равна $K = B \cdot B^T$.

Для нормального закона распределения выполняется следующее свойство: если компоненты двумерного нормального вектора некоррелированы ($\rho_{xy} = 0$), то они и независимы. Для других типов распределения это утверждение может и не выполняться, т.е. из некоррелированности случайных величин, в общем случае, не следует их независимость.

Для случайного вектора $\begin{pmatrix} X \\ Y \end{pmatrix}$ вводятся *условные математические ожидания* $M(X / Y = y)$ и $M(Y / X = x)$. $M(X / Y = y)$ – это математическое ожидание случайной величины X при условии, что Y приняло одно из своих возможных значений y. Аналогично

$M(Y / X = x)$ – это математическое ожидание случайной величины Y при условии, что X приняло одно из своих возможных значений x.

Функцией регрессии Y на X называется зависимость величины $M(Y / X = x)$ от аргумента x. Она характеризует зависимость математического ожидания величины Y от значения, принимаемого величиной X. Аналогично *функцией регрессии* X на Y называется зависимость величины $M(X / Y = y)$ от аргумента y. Она характеризует зависимость математического ожидания величины X от значения, принимаемого величиной Y. Если обе функции регрессии Y на X и X на Y являются линейными, *корреляционная зависимость* между случайными величинами X и Y называется *линейной*. В случае линейной корреляционной зависимости уравнения регрессии Y на X и X на Y называются *уравнениями линейной регрессии*.

Уравнение линейной регрессии Y на X имеет вид

$$y = a_y + \rho_{xy} \frac{\sigma_y}{\sigma_x}(x - a_x)\,, \tag{1.90}$$

а уравнение линейной регрессии X на Y –

$$y = a_y + \frac{1}{\rho_{xy}} \frac{\sigma_y}{\sigma_x}(x - a_x)\,. \tag{1.91}$$

Если случайные величины X и Y имеют двумерное нормальное распределение, корреляционная зависимость между ними может быть только линейной. Для других типов распределения корреляционные зависимости могут быть нелинейными.

Пусть (X_i, Y_i), $i = 1, 2, \dots n$ – выборка объема n из наблюдений случайного двумерного вектора $(X, Y)^T$. Определим оценки числовых характеристик этого вектора. За оценку математических ожиданий a_x и a_y принимаются средние арифметические $\overline{X}$ и $\overline{Y}$ (1.3), за оценку дисперсий σ_x^2 и σ_y^2 – соответствующие эмпирические дисперсии S_x^2 и S_y^2, вычисленные по формуле (1.5). Несмещенной оценкой ковариации K_{xy} является величина

$$\tilde{K}_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})\,. \tag{1.92}$$

Для практических расчетов формулу (1.92) удобно преобразовать к виду

$$\tilde{K}_{xy} = \frac{1}{n-1}\left(\sum_{i=1}^{n} X_i Y_i - n\overline{X}\,\overline{Y}\right). \qquad (1.93)$$

Расчет упрощается, если, как и при нахождении оценок параметров одномерной случайной величины, ввести линейную замену (1.8):

$$X_i = C_1 + h_1 U_i;\ Y_i = C_2 + h_2 V_i. \qquad (1.94)$$

При такой замене формула (1.93) принимает вид

$$\tilde{K}_{xy} = \frac{h_1 h_2}{n-1}\left(\sum_{i=1}^{n} U_i V_i - n\overline{U}\,\overline{V}\right). \qquad (1.95)$$

Оценку коэффициента корреляции ρ_{xy} находят по формуле

$$\rho_{xy} \approx r = \frac{\tilde{K}_{xy}}{S_x S_y}. \qquad (1.96)$$

Уравнения оценочных (выборочных) прямых регрессии получают по следующим формулам.

Уравнение линейной регрессии Y на X:

$$\frac{y-\overline{Y}}{S_y} = r\frac{x-\overline{X}}{S_x}. \qquad (1.97)$$

Уравнение линейной регрессии X на Y:

$$\frac{y-\overline{Y}}{S_y} = \frac{1}{r}\frac{x-\overline{X}}{S_x}. \qquad (1.98)$$

Выборочные уравнения прямых регрессии используют для предсказания среднего значения одной переменной по значению другой.

Задача 1.28. Во втором столбце табл. 1.13 записаны измеренные значения величины X – изменения содержания азота в стали при выпуске из конвертера по сравнению с начальным содержанием [$\times 10^{-4}$, %]; во втором – величины Y (значения начальной концентрации углерода в этой же стали [%]). Найти оценку коэффициента корреляции по этой двумерной выборке. Вычислить выборочные параметры линейной регрессии Y на X и X на Y.

Таблица 1.13

Исходные данные и результаты расчетов к задаче 1.28

Номер эксперимента	X	Y	U	V	U^2	V^2	UV
1	–2,0	0,11	–4	1	16	1	–4
2	0,5	0,09	1	–1	1	1	–1
3	–1,5	0,13	–3	3	9	9	–9
4	–5,5	0,11	–11	1	121	1	–11
5	3,5	0,06	7	–4	49	16	–28
6	–1,0	0,12	–2	2	4	4	–4
7	2,0	0,08	4	–2	16	4	–8
8	0,0	0,11	0	1	0	1	0
9	1,5	0,07	3	–3	9	9	–9
Σ	–	–	–5	–2	225	46	–74

Решение

Вводим линейную замену (1.94), выбирая $C_1 = 0$, $h_1 = 0{,}5$; $C_2 = 0{,}10$, $h_2 = 10^{-2}$. Вычисляем оценки математических ожиданий (1.9):

$$\overline{U} = -\frac{5}{9} \approx 0{,}556; \quad \overline{X} = -0{,}5\cdot 0{,}556 = -0{,}278;$$

$$\overline{V} = -\frac{2}{9} \approx -0{,}22; \quad \overline{Y} = 0{,}10 - 0{,}22\cdot 10^{-2} = 0{,}0978.$$

Несмещенные оценки дисперсий находим по формуле (1.10):

$$S_x^2 = \frac{(0{,}5)^2}{8}\left(225 - 9\left(-\frac{5}{9}\right)^2\right) \approx 6{,}94; \quad S_x \approx 2{,}63;$$

$$S_y^2 = \frac{10^{-4}}{8}\left(46 - 9\left(-\frac{2}{9}\right)^2\right) \approx 5{,}69\cdot 10^{-4}; \quad S_y = 2{,}39\cdot 10^{-2}.$$

Расчет оценки ковариации проводим по формуле (1.95):

$$\tilde{K}_{xy} = \frac{0{,}5\cdot 10^{-2}}{8}\left(-74 - 9\left(-\frac{5}{9}\right)\left(-\frac{2}{9}\right)\right) \approx -4{,}69\cdot 10^{-2}\,.$$

Оценку коэффициента корреляции находим по формуле (1.96):

$$r = \frac{-4,69 \cdot 10^{-2}}{2,69 \cdot 2,39 \cdot 10^{-2}} \approx -0,746\,.$$

Выборочное уравнение линейной регрессии Y на X:

$$\frac{y - 0,0978}{2,39 \cdot 10^{-2}} = -0,746 \frac{x + 0,278}{2,63}$$

или

$$y - 0,0978 = -0,00678\left(x + 0,278\right).$$

Выборочное уравнение линейной регрессии X на Y:

$$\frac{y - 0,0978}{2,39 \cdot 10^{-2}} = -\frac{1}{0,746} \frac{x + 0,278}{2,63}\,;$$

или

$$y - 0,0978 = -0,0122\left(x + 0,278\right).$$

Прямые регрессии представлены на рис. 1.5, там же приведены экспериментальные точки.

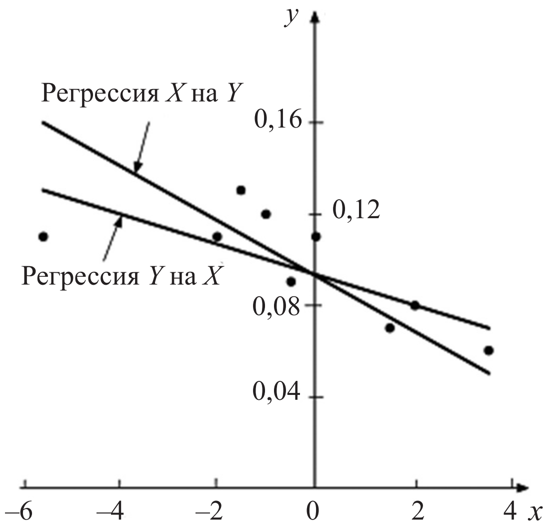

Рис. 1.5. Зависимость изменения концентрации азота в стали (y) при выпуске из конвертера от начальной концентрации углерода (x)

1.9.2. Построение доверительного интервала для коэффициента корреляции. Проверка гипотезы о существовании линейной зависимости

Будем предполагать, что заданная двумерная выборка имеет двумерное нормальное распределение. Тогда доверительный интервал для коэффициента корреляции можно найти по номограммам. В приложении приведены такие номограммы (рис. П1) для доверительной вероятности $P = 0{,}95$ и $P = 0{,}99$. По горизонтальной оси номограммы отложены значения выборочного коэффициента корреляции r, по вертикальной оси – значения истинного коэффициента корреляции ρ_{xy}, числа над кривыми указывают объемы выборок n. Отложив на горизонтальной оси вычисленное значение выборочного коэффициента корреляции, следует подняться над этой точкой вертикально вверх и найти две точки пересечения с кривыми, соответствующими объему заданной выборки. Ординаты этих двух точек являются границами доверительного интервала истинного коэффициента корреляции.

Эти же графики можно использовать для проверки гипотезы H_0 об отсутствии линейной зависимости между величинами X и Y, т.е. о том, что истинный коэффициент корреляции $\rho_{xy} = 0$ при альтернативной гипотезе H_1: $\rho_{xy} \neq 0$. Гипотеза H_0 принимается, т.е. линейная зависимость между величинами не существует (с уровнем значимости $\alpha = 1 - P$), если значение $\rho_{xy} = 0$ принадлежит найденному доверительному интервалу. Здесь P – доверительная вероятность при определении доверительного интервала. Гипотеза H_0 отвергается, т.е. принимается альтернативная гипотеза H_1 (линейная зависимость между величинами существует), если значение $\rho_{xy} = 0$ не принадлежит найденному доверительному интервалу.

Для проверки гипотезы H_0: $\rho_{xy} = 0$ при альтернативной гипотезе H_1: $\rho_{xy} \neq 0$ можно использовать другой критерий. Гипотеза H_0 принимается с уровнем значимости α, т.е. линейная зависимость между величинами не существует, если $|r| < T$, где T – значение критерия:

$$T = \frac{t_{1-\alpha/2}(n-2)}{\sqrt{n - 2 + t^2_{1-\alpha/2}(n-2)}}, \tag{1.99}$$

в противном случае принимается гипотеза H_1, т.е. предполагается, что линейная зависимость между величинами существует. $t_{1-\alpha/2}(n-2)$ – квантиль распределения Стьюдента с числом степеней свободы $k=n-2$.

Если принята гипотеза о существовании линейной зависимости между случайными величинами, то, зная доверительный интервал для коэффициента корреляции, можно сделать вывод о силе взаимосвязи между X и Y. Если доверительный интервал примыкает к единице или минус единице, то говорят, что связь сильная. Если доверительный интервал примыкает к нулю, то говорят, что связь слабая. Если доверительный интервал расположен примерно посередине интервала (–1; 0) или (0; 1), то говорят, что связь средней величины.

Задача 1.29. Найти доверительный интервал для коэффициента корреляции (P = 0,95) и проверить гипотезу об отсутствии линейной зависимости с уровнем значимости α = 0.05 между величинами X и Y для данных задачи 1.28.

Решение

По номограммам приложения для значения $r=-$ 0,746; $n=9$ находим $-0{,}95<\rho<-0{,}14$. Так как значение ρ = 0 не принадлежит найденному доверительному интервалу, гипотеза о существовании линейной зависимости не противоречит экспериментальным данным с уровнем значимости α = 0,05.

Проверим гипотезу об отсутствии линейной зависимости между величинами X и Y с помощью критерия T (1.99). По таблице квантилей распределения Стьюдента находим $t_{0{,}975}(7)=2{,}365$.

$$T=\frac{t_{1-\alpha/2}(n-2)}{\sqrt{n-2+t^2_{1-\alpha/2}(n-2)}}=\frac{2{,}365}{\sqrt{7+2{,}365^2}}=0{,}666.$$

Так как $|r|=0{,}746>0{,}666$, принимаем гипотезу о существовании линейной зависимости между величинами X и Y.

Полученные результаты позволяют сделать вывод, что с увеличением одной из величин среднее значение другой величины уменьшается. Так как коэффициент корреляции значим, можно пользоваться уравнениями выборочных прямых регрессии для предсказания среднего значения одной переменной по значению другой. При этом следует иметь в виду, что предсказанное значение переменной может иметь значительную погрешность. Это следует из очень широкого доверительного интервала для ρ_{xy}. Для получения более точных результатов надо иметь большее число экспериментов.

Задача 1.30. При изучении зависимости величин X и Y по двумерной нормальной случайной выборке объема $n = 25$ получена выборочная матрица ковариаций $\tilde{K} = \begin{pmatrix} 5,20 & 1,82 \\ 1,82 & 1,30 \end{pmatrix}$. Найти выборочный коэффициент корреляции, доверительный интервал для коэффициента корреляции с доверительной вероятностью $P = 0{,}95$; проверить гипотезу об отсутствии линейной зависимости между величинами X и Y с уровнем значимости $\alpha = 0{,}05$. Сделать вывод о силе и направлении взаимосвязи между величинами X и Y.

Решение

На главной диагонали выборочной матрицы ковариаций стоят оценки дисперсии $S_x^2 = 5{,}20$ и $S_y^2 = 1{,}30$, на побочной диагонали – оценка ковариации $\tilde{K}_{xy} = 1{,}82$. По формуле (1.96) находим выборочный коэффициент корреляции $r = \dfrac{1,82}{\sqrt{5,2 \cdot 1,3}} = 0,7$.

По номограммам приложения (рис. П1) для значения $r = 0{,}7$; $n = 25$ находим $0{,}4 < \rho < 0{,}85$. Так как значение $\rho = 0$ не принадлежит найденному доверительному интервалу, гипотеза о существовании линейной зависимости не противоречит экспериментальным данным (с уровнем значимости $\alpha = 0{,}05$).

Проверим гипотезу об отсутствии линейной зависимости между величинами X и Y с помощью критерия T (1.99). По таблице квантилей распределения Стьюдента находим $t_{0,975}(23) = 2{,}07$.

$$T = \frac{t_{1-\alpha/2}(n-2)}{\sqrt{n-2+t_{1-\alpha/2}^2(n-2)}} = \frac{2,07}{\sqrt{23+2,07^2}} = 0{,}396.$$

Так как $|r| = 0{,}7 > 0{,}396$, принимаем гипотезу о существовании линейной зависимости между величинами X и Y.

Полученные результаты позволяют сделать вывод, что линейная зависимости между величинами X и Y сильная, с увеличением одной из величин среднее значение другой величины возрастает.

2. УКАЗАНИЯ ПО ВЫПОЛНЕНИЮ ТИПОВЫХ РАСЧЕТОВ (ЛАБОРАТОРНЫХ РАБОТ) «ОБРАБОТКА ОСНОВНЫХ ТИПОВ ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ»

2.1. Типовой расчет 1. Сравнение двух случайных выборок (первичная обработка данных, проверка статистических гипотез)

Заданы результаты двух серий измерений – две случайные выборки разного объема из двух нормальных генеральных совокупностей. Необходимо для каждой выборки найти оценку математического ожидания и дисперсии, доверительные интервалы для полученных оценок, а затем сравнить полученные результаты, т.е. сделать вывод, можно ли считать в обеих выборках одинаковыми математические ожидания и дисперсии. В заключение проверяется гипотеза о нормальности распределения объединения данных из обеих выборок в одну выборку.

Подобные задачи часто возникают при обработке экспериментальных данных. Приведем примеры.

1. Первая выборка представляет собой повторные измерения некоторой физической величины. Вторая выборка – это повторные измерения той же физической величины, но в иных условиях. Математическое ожидание является здесь истинным значением измеряемой величины, а дисперсия характеризует погрешность измерений. Нахождение доверительных интервалов для математического ожидания и среднего квадратического отклонения необходимо для определения степени близости полученных оценок к истинным значениям соответствующих величин. Сравнивая дисперсии двух выборок, мы определяем, одинакова ли погрешность измерения в каждой из серий. Сравнивая оценки математического ожидания, мы можем узнать, изменяется ли измеряемая физическая величина при изменении условий эксперимента.

2. Предлагается новая технология производства некоторых изделий для улучшения качества, при этом значения числового параметра, характеризующего качество изделия увеличиваются, например прочность металла после термической обработки. Следует экспериментально обосновать преимущество новой технологии, указав, что она действительно увеличивает значение рассматриваемого параметра. Если применение новой технологии увеличивает значение параметра в несколько раз и этот результат повторяется от изделия к изделию, то нет необходимости в применении статистических методов, преимущество новой технологии очевидно. Но если наблюдается не очень большое увеличение значений параметра, к тому же существует заметный разброс в результатах от изделия к изделию, необходимо применить статистические методы. Для этого измеряют значение параметра у ряда изделий, изготовленных по старой и по новой технологии. Таким образом получают две случайные выборки. Выборочное среднее и выборочная дисперсия по каждой из выборок характеризуют среднее значение параметра и степень его разброса у изделий, изготовленных по старой и новой технологии. Доверительные интервалы выборочного среднего и выборочной дисперсии характеризует степень близости полученных оценок соответственно к генеральному среднему и генеральной дисперсии, т.е. к среднему значению параметра и его дисперсии по всем изделиям, изготовленным по данной технологии. Сравнивая выборочные дисперсии, выясняют, изменяется ли степень разброса параметра при изменении технологии; сравнивая выборочные средние, определяют, действительно ли новая технология улучшает качество изделий.

3. Для характеристики месторождения некоторого металла взяты пробы руды. Измеряя концентрацию ценного компонента в этих пробах, получают случайную выборку. Выборочная средняя здесь характеризует концентрацию ценного компонента во всем месторождении, выборочная дисперсия – разброс концентрации компонента в различных районах месторождения. Доверительные интервалы выборочного среднего и выборочной дисперсии характеризуют степень близости полученных оценок к соответствующим параметрам месторождения в целом. Если одновременно обрабатываются пробы руды с другого месторождения, то, сравнивая выборочные дисперсии и выборочные средние, можно определить, одинаков ли разброс концентрации ценного компонента в этих месторождениях и одинакова ли в них его концентрация.

Этапы выполнения

2.1.1. Первичная обработка результатов измерений

Рассчитать для каждой выборки оценку математического ожидания, несмещенную оценку дисперсии, оценку среднего квадратического отклонения.

Для упрощения расчетов и последующего контроля результатов вычислений следует провести кодировку данных по формуле (1.8) и найти оценки по формулам (1.9), (1.10), (1.6). Для контроля вычислений весь расчет необходимо повторить с другим началом отсчета. Результаты расчета должны совпасть с точностью до возможных ошибок округления. Пример расчета приведен в задаче 1.1.

Внимание! Для упрощения последующих расчетов при выборе параметров кодировки $C_1^{(1)}$, $C_2^{(1)}$ первой серии измерений и $C_1^{(2)}$, $C_2^{(2)}$ второй серии измерений следует одно из начал отсчета $C_j^{(1)}$, $C_j^{(2)}$ сделать одинаковым.

2.1.2. Построение доверительных интервалов

По условию в каждой выборке результаты измерений независимы и имеют нормальное распределение с одинаковыми параметрами. Для каждой выборки необходимо найти доверительные интервалы для математического ожидания и среднего квадратического отклонения с доверительной вероятностью $P = 0{,}95$. Построение доверительных интервалов для математического ожидания m и стандартного отклонения σ рассмотрено в разд. 1.6 и в задаче 1.8.

2.1.3. Проверка гипотез о равенстве дисперсий, о равенстве математических ожиданий

Проверить гипотезу о равенстве дисперсий в двух сериях измерений H_0: $\sigma_1^2 = \sigma_2^2$ (математические ожидания a_1 и a_2 неизвестны) с уровнем значимости $\alpha = 0{,}05$ при альтернативной гипотезе H_1: $\sigma_1^2 \neq \sigma_2^2$ (см. задачу 1.13).

Если гипотеза о равенстве дисперсий отвергается, перейти к проверке гипотезы о нормальном распределении объединения двух случайных выборок, рассмотренной в следующем пункте. Если же гипотеза о равенстве дисперсий не противоречит экспериментальным данным, найти сводную оценку дисперсии (1.15) и построить доверительный интервал для σ, используя сводную оценку дисперсии.

Затем проверить гипотезу о равенстве математических ожиданий в двух сериях измерений H_0: $a_1 = a_2$ при альтернативной гипотезе H_1: $a_1 \neq a_2$ с уровнем значимости α. Если гипотеза о равенстве математических ожиданий отвергается, перейти к выполнению следующего пункта. Если же гипотеза о равенстве математических ожиданий не противоречит экспериментальным данным, найти сводную оценку математического ожидания и доверительный интервал для математического ожидания, используя сводную оценку дисперсии.

Сводная оценка математического ожидания определяется как средняя по обеим сериям:

$$\overline{X}_{\text{св}} = \frac{n_1 \overline{X}_1 + n_2 \overline{X}_2}{n_1 + n_2}. \tag{2.1}$$

Доверительный интервал для математического ожидания при этом можно пересчитать, заменив в формулах (1.19) – (1.22) $\overline{X}$ на $\overline{X}_{\text{св}}$, S на $S_{\text{св}}$ и k на $k_{\text{св}} = k_1 + k_2$, n на $n_1 + n_2$.

Рассмотрим примеры.

Задача 2.1. По случайной нормальной выборке объема $n_1 = 11$ найдена выборочная дисперсия $S_1^2 = 0{,}373$. По другой случайной нормальной выборке объема $n_2 = 9$ также найдена выборочная дисперсия $S_2^2 = 0{,}0955$. Проверить гипотезу о равенстве дисперсий H_0: $\sigma_1^2 = \sigma_2^2$ при альтернативной гипотезе H_1: $\sigma_1^2 \neq \sigma_2^2$ с уровнем значимости $\alpha = 0{,}05$. Если гипотеза о равенстве дисперсий не противоречит экспериментальным данным, найти сводную оценку дисперсии и построить доверительный интервал для σ, используя сводную оценку дисперсии, с доверительной вероятностью $P = 0{,}95$.

Решение

Найдем отношение большей дисперсии S_1^2 к меньшей S_2^2 (1.33): $F = \dfrac{0{,}373}{0{,}0955} = 3{,}91$.

По таблице квантилей распределения Фишера $F_{0,975}(10; 8) = 4{,}30$. Так как отношение выборочных дисперсий $F = 3{,}91$ меньше значения квантиля 4,30, гипотеза о равенстве дисперсий принимается (1.34), как не противоречащая результатам эксперимента с уровнем значимости 0,05.

Можно вычислить сводную оценку дисперсии (1.15):

$$S_{\text{св}}^2 = \frac{10 \cdot 0{,}373 + 8 \cdot 0{,}0955}{10 + 8} = 0{,}250$$

с числом степеней свободы $k_{\text{св}} = k_1 + k_2 = 10 + 8 = 18$.

Рассчитаем доверительный интервал для σ. Из табл. П3 находим $\chi^2_{\alpha/2}(k) = \chi^2_{0,025}(18) = 8{,}23$; $\chi^2_{1-\alpha/2}(k) = \chi^2_{0,975}(18) = 31{,}5$. По формуле (1.24) получаем $0{,}5\sqrt{\frac{18}{31{,}5}} < \sigma < 0{,}5\sqrt{\frac{18}{8{,}23}}$, откуда $0{,}378 < \sigma < 0{,}739$.

Задача 2.2. Рассмотрим задачу, аналогичную задаче 2.1, но с другими числовыми данными: $n_1 = 26$; $S_1^2 = 0{,}395$; $n_2 = 19$; $S_2^2 = 1{,}67$.

Решение

Найдем отношение большей дисперсии S_2^2 к меньшей S_1^2: $F = \frac{1{,}67}{0{,}395} = 4{,}23$. По таблице квантилей распределения Фишера (табл. П5) следует найти значение $F_{0,975}(18;\,25)$. Однако в этой таблице квантиля с требуемыми параметрами нет. В этом случае находят приближенное значение квантиля, принимая допущение о линейном характере изменения величины F. Сначала найдем значение $F_{0,975}(15;\,25)$ как среднее арифметическое между значениями $F_{0,975}(15;\,24) = 2{,}44$ и $F_{0,975}(15;\,26) = 2{,}39$:

$$F_{0,975}(15;\,25) \approx \frac{2{,}44 + 2{,}39}{2} = 2{,}415\,.$$

Аналогично

$$F_{0,975}(20;\,25) \approx \frac{2{,}33 + 2{,}28}{2} = 2{,}305\,.$$

Затем оценим значение $F(0{,}975;\,18;\,25)$ методом линейной интерполяции (рис. 2.1):

$$F_{0,975}(18;\,25) \approx 2{,}305 + (2{,}415 - 2{,}305)\frac{20-18}{20-15} = 2{,}305 + \frac{2}{5}\cdot 0{,}11 \approx 2{,}35\,.$$

Так как отношение выборочных дисперсий $F = 4{,}23$ больше значения квантиля 2,35, гипотезу о равенстве дисперсий отвергаем с уровнем значимости $\alpha = 0{,}05$.

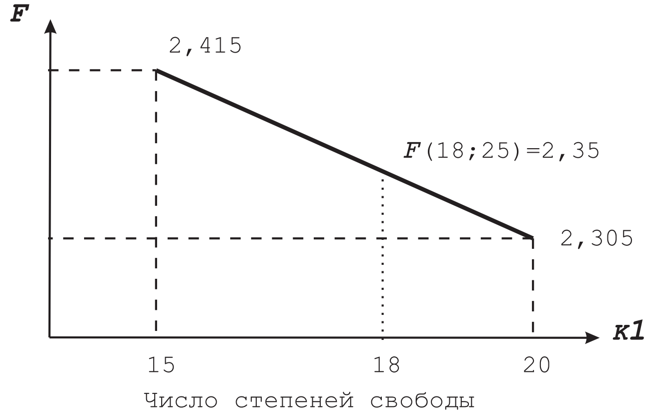

Рис. 2.1. Иллюстрация метода линейной интерполяции

Задача 2.3. По двум случайным нормальным выборкам получены выборочные средние $\overline{X}_1 = 12{,}95$ и $\overline{X}_2 = 12{,}13$. Объемы выборок равны соответственно $n_1 = 12$ и $n_2 = 15$. Дисперсии обеих выборок одинаковы. Сводная оценка среднего квадратического отклонения $S_{св} = 0{,}872$ с числом степеней свободы $k_{св} = 25$. Проверить гипотезу H_0 о равенстве математических ожиданий при альтернативной гипотезе $H_1 : a_1 \neq a_2$ с уровнем значимости $\alpha = 0{,}05$. Если гипотеза о равенстве математических ожиданий не противоречит экспериментальным данным, найти сводную оценку математического ожидания и доверительный интервал для математического ожидания, используя сводную оценку математического ожидания и сводную оценку дисперсии с доверительной вероятностью $P = 0{,}95$.

Решение

Найдем значение критерия Стьюдента t по формуле (1.50):

$$t = \frac{12{,}95 - 12{,}13}{0{,}872\sqrt{1/12 + 1/15}} = 2{,}43; \qquad |t| = 2{,}43.$$

Из таблицы квантилей распределения Стьюдента $t_{0,975}(25) = 2{,}06$. Так как вычисленное значение $|t| = 2{,}43$ больше этого значения, гипотеза о равенстве математических ожиданий отвергается с уровнем значимости $\alpha = 0{,}05$ (1.51).

Задача 2.4. Рассмотрим задачу, аналогичную разобранной в задаче 2.3, но с другими числовыми данными: $\overline{X}_1 = 27{,}43$; $n_1 = 16$; $\overline{X}_2 = 28{,}76$; $n_2 = 21$; $S_{св} = 2{,}35$; $k_{св} = 35$.

Решение

Найдем значение критерия Стьюдента t по формуле (1.50):

$$t = \frac{27{,}43 - 28{,}76}{2{,}35\sqrt{1/16 + 1/32}} = -1{,}71; \quad \left| t \right| = 1{,}71 .$$

Из таблицы квантилей распределения Стьюдента $t_{0,975}(35) = 2{,}03$. Так как вычисленное значение $|t| = 1{,}71$ меньше табличного, гипотеза о равенстве математических ожиданий принимается как не противоречащая выборочным данным с уровнем значимости 0,05. Сводная оценка математического ожидания (2.1) равна

$$\overline{X}_{св} = \frac{16 \cdot 27{,}43 + 21 \cdot 28{,}76}{16 + 21} = 28{,}18 ,$$

доверительный интервал для математического ожидания (1.21) и (1.22):

$$\varepsilon = t_{1-\alpha/2}(k)\, S_{св} \Big/ \sqrt{n_{св}} = t_{0,975}(35)\; 2{,}35 / \sqrt{16 + 21} = 0{,}78;$$

$$a = \overline{X}_{св} \pm \varepsilon = 28{,}18 \pm 0{,}78.$$

2.1.4. Проверка гипотезы о нормальном распределении объединения двух случайных выборок

Проверить гипотезу о нормальном распределении объединения двух заданных случайных выборок, величину уровня значимости α выбрать ту же, что и в разд. 2.1.3.

Для проверки гипотезы о нормальном распределении случайной величины X по критерию согласия Пирсона (разд. 1.7.4) весь диапазон возможных значений этой величины разбивают на l интервалов (значение l задано в условии типового расчета), вычисляют p_i – вероятность попадания в каждый из интервалов ($i = 1, 2, ..., l$). Затем вычисляют величину χ^2 по формуле (1.55) и сравнивают ее с квантилем $\chi^2_{1-\alpha}(k)$ распределения Пирсона. Так как для вычисления вероятностей p_i параметры нормального распределения оценивают по той же выборке, по которой строят критерий согласия, то число степеней свободы $k = l - 3$. Если $\chi^2 > \chi^2_{1-\alpha}(k)$, гипотеза отвергается при заданном уровне значимости $\alpha = 1 - P$. Если $\chi^2 < \chi^2_{1-\alpha}(k)$, гипотеза принимается, как не противоречащая результатам эксперимента.

В общем случае вероятности p_i находятся с помощью интеграла вероятностей $\Phi(t)$. При этом оценками параметров нормального распределения являются $\overline{X}$ и S, найденные по объединению данных двух выборок. За оценку математического ожидания принимаем среднее арифметическое по объединению двух выборок (2.1). Оценка среднего квадратического отклонения σ объединения выборок зависит от результата проверки гипотезы о равенстве математических ожиданий этих выборок. Если гипотеза о равенстве математических ожиданий принимается, то за оценку σ может быть взята $S = \sqrt{S^2_{\text{св}}}$, полученная по сводной оценке дисперсии (1.15):

$$S^2_{\text{св}} = \frac{k_1 S_1^2 + k_2 S_2^2}{k_1 + k_2}; \qquad k_j = n_j - 1; \qquad j = 1,\ 2, \tag{2.2}$$

где S_1^2, S_2^2 – несмещенные оценки дисперсии первой и второй выборок.

Если гипотеза о равенстве математических ожиданий отвергнута, необходимо рассчитать несмещенную оценку дисперсии объединения выборок по основной формуле (1.5)

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - x\overline{X})^2, n = n_1 + n_2. \tag{2.3}$$

Чтобы не повторять заново расчет оценок дисперсии, можно использовать следующий прием. Для упрощения расчетов и организации контроля при нахождении оценок математического ожидания и дисперсии по выборкам исходные данные рекомендуется кодировать (1.8) $U_i = \dfrac{X_i - C}{h}$ и расчет S_j^2 вести по формуле (1.10):

$$S_j^2 = \frac{h^2}{k_j}\left(\sum_{i=1}^{n_j} U_{ij}^2 - n_j \overline{U}_j^2\right), \ j = 1, 2\,.$$

Если заранее выбрать h и C одинаковые для обеих выборок, расчет оценки дисперсии объединения выборок может быть выполнен по формуле

$$S^2 = \frac{h^2}{n-1}\left(\sum_{i=1}^{n_1}{}' U_i^2 + \sum_{i=1}^{n_2}{}'' U_i^2 - n\overline{U}^2\right),\ \overline{U} = \frac{n_1\overline{U}_1 + n_2\overline{U}_2}{n_1 + n_2}\ , \qquad (2.4)$$

где первая сумма квадратов $\sum'$ кодированных данных относится к первой выборке, вторая $\sum''$ – ко второй.

При расчетах будем использовать *интервалы равной вероятности*, т.е. весь диапазон возможных значений случайной величины разобьем на такие интервалы, чтобы вероятности p_i попадания в каждый из них были бы одинаковы, т.е. $p_i = 1/l$. Для нахождения границ таких интервалов x_i необходимо с помощью интеграла вероятностей $\Phi(t)$ найти границы интервалов равной вероятности u_i для случайной величины U, имеющей стандартное нормальное распределение (1.16). Тогда оценочные границы интервалов равной вероятности x_i случайной величины X могут быть найдены по формуле

$$x_i = \overline{X} + Su_i\ . \qquad (2.5)$$

Затем для вычисления χ^2 подсчитаем числа n_i значений величины X, принадлежащих каждому полученному интервалу, и произведем вычисления по формуле (1.54).

Задача 2.5. Проверить гипотезу о нормальности распределения величины X по случайной выборке объема $n = 51$. Выбрано $l = 5$, $p_i = 1/l = 0{,}2$. Оценки параметров нормального распределения $\overline{X} = 28{,}23$, $S = 2{,}37$.

Решение

Найдем границы интервалов равной вероятности u_i для случайной величины со стандартным нормальным распределением. Если разбить область под функцией плотности стандартного нормального распределения f_N на криволинейные трапеции, равной площади $p = 0{,}2$ (рис. 2.2), то границы оснований этих трапеций будут искомыми числами u_i. Следовательно, u_i – соответствующие квантили стандартного нормального распределения. Для их нахождения в первом столбце расчетной таблицы (табл. 2.1) записываем значения функции распределения $F(u)$ в искомых точках u_i (вероятность случайной величине u попасть левее этой точки). Это будут числа, кратные $p = 0{,}2$, начиная с нуля (что соответствует левой бесконечной границе) и кончая единицей (соответствует правой бесконечной границе). Во втором столбце записываем значения интеграла вероятности $\Phi(u)$ в искомых точках: $\Phi(u) = F(u) - 0{,}5$. Используя таблицу значений функции $\Phi(u)$ [2], находим для положительных значений функции $\Phi(u)$ значения u_i как обратной к функции $\Phi(u)$, т.е. по известным значениям функции определяем соответствующие значения аргументов. Для отрицательных значений функции $\Phi(u)$ используем свойство нечетности этой функции: $\Phi(-u) = -\Phi(u)$.

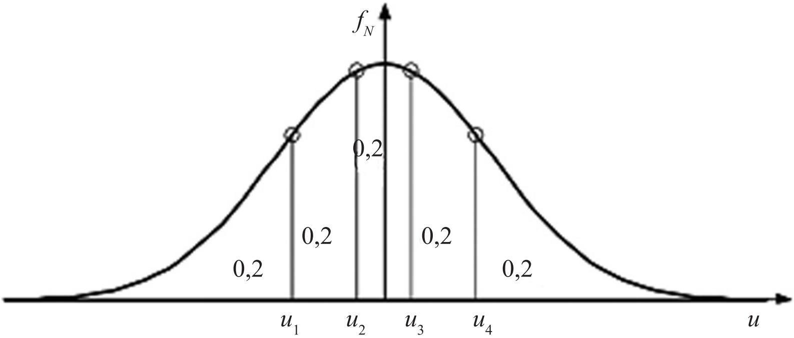

Рис. 2.2. Пример разбиения области под графиком функции плотности вероятностей на криволинейные трапеции равной площади

При нахождении u_i по значению $\Phi(u_i)$ используем метод линейной интерполяции (см. задачу 2.3). Например, значения $\Phi(u) = 0{,}1$ в таблице из [2] нет. Выписываем ближайшие значения $\Phi(u)$: 0,0987 и 0,1026. Им соответствуют значения аргумента 0,25 и 0,26. Тогда

$$u_i \approx 0{,}25 + (0{,}26 - 0{,}25)\frac{0{,}1 - 0{,}0987}{0{,}1026 - 0{,}0987} = 0{,}253\,.$$

Границы интервалов равной вероятности x_i для рассматриваемой величины X находятся по формуле (2.5). Значения x_i желательно вычислять с запасным знаком по сравнению с элементами случайной выборки. Это делается для того, чтобы элементы выборки, как правило, не попадали на границы интервалов. Результаты расчетов приведены в табл. 2.1.

Таблица 2.1

Результаты расчетов (к задаче 2.5)

$F(u_i)$	$\Phi(u_i)$	u_i	x_i
0	–0,5	$-\infty$	$-\infty$
0,2	–0,3	–0,842	26,235
0,4	–0,1	–0,253	27,631
0,6	0,1	0,253	28,829
0,8	0,3	0,842	30,225
1,0	0,5	$+\infty$	$+\infty$

Найдя границы интервалов равной вероятности, подсчитываем числа n_i попадания элементов случайной выборки в каждый из l интервалов: $np_i = 51/5 = 10{,}2$. Подсчет величины χ^2 по формуле (1.54) дает значение $\chi^2 = 46{,}8/10{,}2 = 4{,}59$. Результаты расчетов приведены в табл. 2.2.

Таблица 2.2

Результаты расчетов (к задаче 2.5)

i	(x_i, x_{i+1})	n_i	$n_i - np_i$	$(n_i - np_i)^2$
1	$(-\infty; 26{,}235)$	12	1,8	3,24
2	(26,235; 27,631)	9	–1,2	1,44
3	(27,631; 28,829)	5	–5,2	27,04
4	(28,829; 30,225)	14	3,8	14,44
5	$(30{,}225; +\infty)$	11	0,8	0,64
Σ		51	0	46,80

Сравнивая найденное значение χ^2 с квантилем $\chi^2_{1-\alpha}(k)$ при уровне значимости $\alpha = 0{,}05$ и при $k = 5 - 3 = 2$, т.е. $\chi^2_{0,95}(2) = 5{,}99$, замечаем, что $4{,}59 < 5{,}99$. Следовательно, можно считать, что при заданном уровне значимости $\alpha = 0{,}05$ гипотеза о нормальном распределении величины X не противоречит результатам эксперимента.

2.1.5. Построение гистограмм распределения объединения двух случайных выборок

На этом этапе расчета следует построить гистограмму распределения объединения двух заданных выборок (разд. 1.4), разбив диапазон изменения значений элементов выборок на l интервалов равной длины. В качестве l для определенности взять число l интервалов равной вероятности, использованных для проверки гипотезы о нормальности распределения (предыдущий пункт).

Для построения гистограммы диапазон изменения значений элементов выборки накрывают отрезком чуть более широким, чтобы наибольшее и наименьшее значения выборки являлись внутренними точками этого отрезка. Полученный отрезок разбивают на l интервалов равной длины, подсчитывают число n_i попаданий элементов выборки в каждый i-й интервал. При этом желательно, чтобы элементы выборки не попадали на границы интервалов. Затем строят столбиковую диаграмму, откладывая по оси ординат величины, пропорциональные n_i (можно откладывать значения n_i).

Задача 2.6. Построить гистограмму по случайной выборке объема $n = 50$. Наибольший элемент выборки 12,73, наименьший 9,51, значения всех элементов выборки записаны с двумя знаками после запятой. Выбрано число интервалов $l = 5$.

Решение

Длина диапазона изменения элементов выборки равна 12,73 – 9,51 = 3,22. Увеличим эту длину до ближайшего числа, которое при делении на l дает частное с числом знаков после запятой не более, чем элементы выборки. В данном примере таким числом является 3,25, т.е. длину диапазона необходимо увеличить на 0,03. При этом левую границу сдвинем влево, например на 0,01, правую границу сдвинем вправо на 0,02. Получили отрезок [9,50; 12,75], для которого все элементы выборки являются внутренними точками. Делим отрезок на 5 интервалов равной длины, длина каждого отрезка равна $h = 3{,}25/5 = 0{,}65$.

Полученное разбиение на интервалы обладает тем недостатком, что элементы выборки могут попасть на границы интервалов. Этого можно избежать, если сдвинуть полученный отрезок на половину последнего разряда значений элементов выборки, т.е. на величину 0,005. Сдвинем отрезок на эту величину, например влево. Получим отрезок [9,495; 12,745] с шагом интервалов разбиения $h = 0{,}65$. Затем подсчитаем число n_i попаданий элементов выборки в каждый интервал (табл. 2.3) и строим гистограмму (рис. 2.3).

Таблица 2.3

Результаты расчетов (к задаче 2.6)

Номер	Граница интервала	n_i
1	(9,495; 10,145)	3
2	(10,145; 10,795)	10
3	(10,795; 11,445)	18
4	(11,445; 12,095)	13
5	(12,095; 12,745)	6
Σ		50

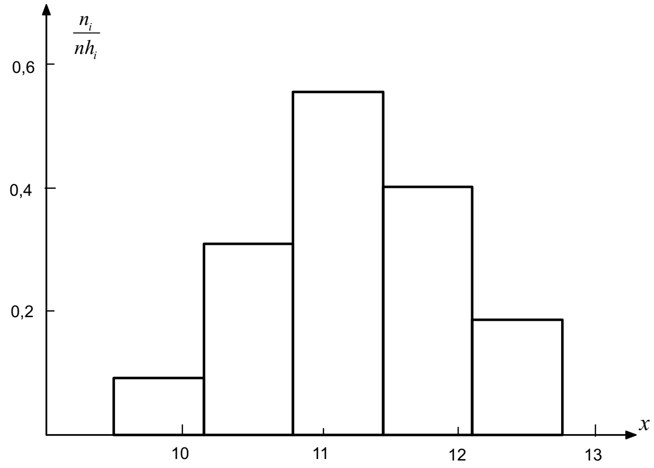

Рис. 2.3. Гистограмма

Выводы по работе

В конце выполнения типового расчета необходимо сделать выводы о результатах проверок статистических гипотез и привести окончательные числовые результаты.

1. **Результаты проверки гипотезы о равенстве дисперсий.** Если гипотеза отвергнута, привести оценки дисперсии и среднеквадратического отклонения по каждой выборке с соответствующими доверительными интервалами для σ. Если гипотеза принята, привести сводные оценки дисперсии и среднеквадратического отклонения, а также уточненный интервал для σ.

2. **Результат проверки гипотезы о равенстве математических ожиданий или информации о том, что эта гипотеза не проверялась.** Если гипотеза отвергнута, привести оценки математического ожидания и соответствующие доверительные интервалы для каждой выборки отдельно. Если гипотеза принята, привести сводную оценку математического ожидания и соответствующий уточненный доверительный интервал.

3. **Результат проверки гипотезы о нормальном распределении объединении двух серий измерений.**

Задача 2.7. Пример выводов по первому типовому расчету.

1. Гипотеза о равенстве дисперсий принята как не противоречащая экспериментальным данным с уровнем значимости $\alpha = 0{,}05$; $S^2_{\text{св}} = 0{,}760$; $S_{\text{св}} = 0{,}872$; $\sigma \in (0{,}684;\ 1{,}203)$.

2. Гипотеза о равенстве математических ожиданий отвергнута как противоречащая экспериментальным данным с уровнем значимости $\alpha = 0{,}05$, $m_1 = 12{,}95 \pm 0{,}55$; $m_2 = 12{,}13 \pm 0{,}48$.

3. Гипотеза о нормальном распределении объединения двух выборок отвергнута с уровнем значимости $\alpha = 0{,}05$ как противоречащая экспериментальным данным.

2.2. Типовой расчет 2. Обработка данных методами регрессионного анализа

Типовой расчет состоит из двух задач. В каждой задаче приведены результаты независимых равноточных экспериментов по изучению зависимости одной величины (y) от другой (x). В каждой серии проведено n наблюдений (x_i, Y_i), $i = 1, 2, ..., n$. Будем предполагать, что значения аргументов x_i известны точно, а значения функции Y_i – взаимно независимые случайные величины $Y_i = f(x_i) + Z_i$, где погрешности эксперимента Z_i – независимые равноточные случайные величины, имеющие нормальное распределение

$$M(Z_i) = 0;\ \ D(Y_i) = \sigma^2.$$

По отдельной серии измерений найдена несмещенная оценка дисперсии S^2.

Задача 1. По приведенным исходным данным требуется:

– найти оценки для параметров линейной регрессии Y на x, проверить адекватность построенной модели, приняв уровень значимости $\alpha = 0{,}05$;

– найти границы доверительных интервалов для параметров линейной модели с доверительной вероятностью $P = 0{,}95$;

– построить график полученной модели регрессии и график отклонений функции регрессии от экспериментальных данных.

Задача 2. По приведенным исходным данным требуется:

– найти оценки для параметров линейной регрессии Y на x; проверить адекватность построенной модели, приняв уровень значимости $\alpha = 0{,}05$;

– если линейная модель адекватна, найти границы доверительных интервалов для параметров полученной линейной модели с доверительной вероятностью $P = 0{,}95$;

– если линейная модель неадекватна, найти оценки для параметров квадратичной модели регрессии, проверить адекватность полученной квадратичной модели с тем же уровнем значимости;

– найти границы доверительных интервалов для параметров квадратичной модели;

– построить графики полученных моделей регрессии и графики отклонений функции регрессии от экспериментальных данных.

2.2.1. Задача 1

Результаты эксперимента представлены в первых двух столбцах табл. 2.4. По отдельной серии из $n_1 = 19$ экспериментов найдена оценка дисперсии $S^2 = 0{,}96$. Найдем оценки параметров линейной регрессии Y на x.

Таблица 2.4

Исходные данные и результаты расчета (к задаче 1)

x	Y	$x-\overline{x}$	X	X^2	YX	$Y_{\text{лин}}$	ΔY	ΔY^2
5	19,1	–45	–9	81	–171,9	19,65	–0,55	0,3025
20	25,4	–30	–6	36	–152,4	24,6	0,8	0,64
40	33,0	–10	–2	4	–66,0	31,2	1,8	3,24
50	34,1	0	0	0	0	34,5	–0,4	0,16
55	35,0	5	1	1	35,0	36,15	–1,15	1,3225
65	38,0	15	3	9	114,0	39,45	–1,45	2,1025
75	42,4	25	5	25	212,0	42,75	–0,35	0,1225
90	49,0	40	8	64	392,0	47,7	1,3	1,69
$\sum$400	276	0	0	220	362,7		0	9,58

Сначала найдем решение задачи регрессии в кодированных значениях переменной x (1.71). Введем новую переменную по формуле $X = \dfrac{x-\overline{x}}{h}$, где $\overline{x} = \dfrac{1}{n}\sum_{i=1}^{n} x_i = \dfrac{1}{8}\cdot 400 = 50.$ Значения величины $x-\overline{x}$

являются целыми числами с общим множителем $h = 5$. Поэтому переменная $X = \dfrac{x-50}{5}$ принимает целые значения, не имеющие общего множителя.

По формулам (1.72) находим оценки коэффициентов линейной регрессии

$$\tilde{B}_1 = \frac{\sum_{i=1}^{n} Y_i}{n} = \frac{276}{8} = 34,5; \;\; \tilde{B}_2 = \frac{\sum_{i=1}^{n} Y_i X_i}{\sum_{i=1}^{n} X_i^2} = \frac{362,7}{220} = 1,6484 \approx 1,65.$$

Получили линейную модель регрессии

$$Y_{\text{лин}} = 34,5 + 1,65X. \tag{2.6}$$

Вычисляем значения линейной функции регрессии $Y_{\text{лин}}$ по формуле (2.6) при всех значениях аргумента X, а затем рассчитываем $\Delta Y_i = Y_i - Y_{i\,\text{лин}}$ отклонения экспериментальных значений Y_i от значений $Y_{i\,\text{лин}}$, полученных по функции регрессии, и для контроля суммируем значения ΔY_i. Условие (1.73) выполняется.

Далее суммируем значения ΔY_i^2 для нахождения дисперсии адекватности (1.74):

$$S_{\text{ад}}^2 = \frac{\sum_{i=1}^{n} (\Delta Y_i)^2}{n-2} = \frac{9,58}{6} = 1,597.$$

Все расчеты приведены в табл. 2.4.

Проверяем адекватность линейной модели регрессии, используя критерий Фишера (1.67):

$$F = \frac{S_{\text{ад}}^2}{S_{\text{эксп}}^2} = \frac{1,597}{0,96} = 1,664. \;\; k_{\text{ад}} = n-2.$$

Квантиль распределения Фишера $F_{1-\alpha}(k_{\text{ад}}, k_{\text{эксп}}) = F_{0,95}(6;18) = 2,66.$ Так как

$$F = 1,664 < 2,66 = F_{1-\alpha}(k_{\text{ад}}, k_{\text{эксп}}),$$

то гипотеза об адекватности линейной модели регрессии принимается.

Найдем доверительные интервалы для коэффициентов регрессии. Границы доверительных интервалов для коэффициентов B_1 и B_2 найдем по формулам (1.75). Так как линейная модель регрессии адекватна, в качестве оценки дисперсии возьмем дисперсию адекватности.

$$B_1 = 34{,}5 \pm \varepsilon_1; \qquad \varepsilon_1 = t_{1-\frac{\alpha}{2}}(k_{\text{ад}})\frac{S_{\text{ад}}}{\sqrt{n}} = 2{,}447\frac{\sqrt{1{,}597}}{\sqrt{8}} = 1{,}093;$$

$$B_2 = 1{,}65 \pm \varepsilon_2; \qquad \varepsilon_2 = t_{1-\frac{\alpha}{2}}(k_{\text{ад}})\frac{S_{\text{ад}}}{\sqrt{\sum\limits_{i=1}^{n} X_i^2}} = 2{,}447\frac{\sqrt{1{,}597}}{\sqrt{220}} = 0{,}208.$$

Уравнение линейной регрессии Y от исходного переменного x найдем, сделав преобразование:

$$Y_{\text{лин}} = \beta_1 + \beta_2 x = 34{,}5 + 1{,}65\frac{x-50}{5} = 18{,}0 + 0{,}33x. \qquad (2.7)$$

Границы доверительных интервалов для коэффициентов β_1 и β_2 (1.77):

$$\beta_1 = 18{,}0 \pm \hat{\varepsilon}_1;$$

$$\hat{\varepsilon}_1 = t_{1-\frac{\alpha}{2}}(k_{\text{ад}})\, S_{\text{ад}} \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{h^2 \sum\limits_{i=1}^{n} X_i^2}} = 2{,}447\sqrt{1{,}597} \cdot \sqrt{\frac{1}{8} + \frac{2500}{25 \cdot 220}} = 2{,}355;$$

$$\beta_2 = 0{,}33 \pm \hat{\varepsilon}_2; \qquad \hat{\varepsilon}_2 = t_{1-\frac{\alpha}{2}}(k_{\text{ад}})\frac{S_{\text{ад}}}{h\sqrt{\sum\limits_{i=1}^{n} X_i^2}} = 2{,}447\frac{\sqrt{1{,}597}}{5\sqrt{220}} = 0{,}0416.$$

На рис. 2.4 построен график линейной модели регрессии. Точками помечены результаты эксперимента. На рис. 2.5 приведен график отклонений экспериментальных данных от функции регрессии. Отклонения носят хаотический характер, что подтверждает адекватность полученной модели.

Выводы. По приведенным исходным данным методом наименьших квадратов построена адекватная линейная модель регрессии с уровнем значимости $\alpha = 0{,}05$; $S^2_{\text{ад}} = 1{,}597$.

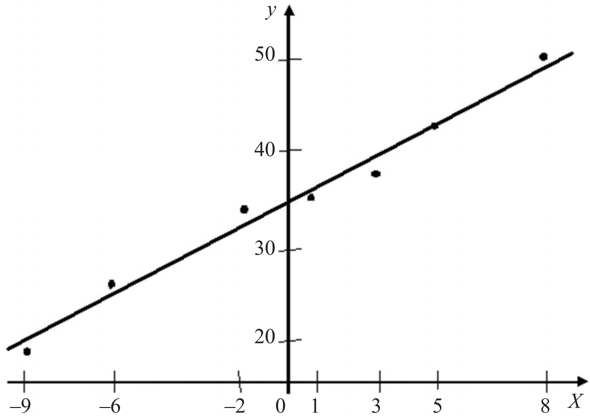

Рис. 2.4. График линейной модели регрессии

Уравнение регрессии в кодированных значениях переменной $X = \dfrac{x-50}{5}$:

$$y_{\text{лин}} = B_1 + B_2 X,$$

где $B_1 = 34{,}5 \pm 1{,}093$; $B_2 = 1{,}65 \pm 0{,}208$.

Уравнение регрессии в исходной переменной *x*: $y_{\text{лин}} = \beta_1 + \beta_2 x$, где $\beta_1 = 18{,}0 \pm 2{,}355$; $\beta_2 = 0{,}33 \pm 0{,}0416$.

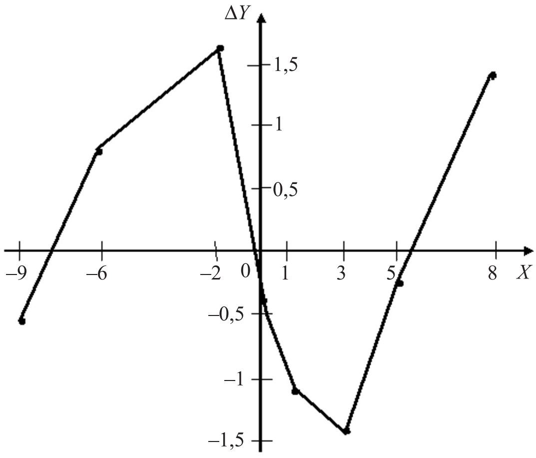

Рис. 2.5. График отклонений экспериментальных данных от линейной функции регрессии

2.2.2. Задача 2

Результаты эксперимента представлены в первых двух столбцах табл. 2.5. По отдельной серии из $n_2 = 20$ экспериментов найдена оценка дисперсии $S^2 = 0{,}176$. Найдем оценки параметров линейной регрессии Y на x.

Таблица 2.5

Исходные данные и результаты расчета линейной модели регрессии (к задаче 2)

x	Y	$x-\bar{x}$	X	X^2	YX	$Y_{\text{лин}}$	$\Delta Y_{\text{лин}}$	$\Delta Y^2_{\text{лин}}$
0	12,28	–0,5	–5	25	–61,40	14,923	–2,643	6,9854
0,2	18,66	–0,3	–3	9	–55,98	17,785	0,875	0,7656
0,4	22,80	–0,1	–1	1	–22,80	20,647	2,153	4,6354
0,6	24,97	0,1	1	1	24,97	23,509	1,461	2,1345
0,8	26,70	0,3	3	9	80,10	26,371	0,329	0,1082
1,0	27,06	0,5	5	25	135,30	29,233	–2,173	4,7219
$\sum 3{,}0$	132,67	0	0	70	100,19		–0,002	19,3510

Найдем решение задачи линейной регрессии в кодированных значениях переменной x (1.71). Введем новую переменную по формуле $X = \dfrac{x-\bar{x}}{h}$, где $\bar{x} = \dfrac{1}{n}\sum_{i=1}^{n} x_i = \dfrac{1}{6}\cdot 3{,}0 = 0{,}5$. Значения величины $x-\bar{x}$ будут целые числа, не имеющие общего множителя, если принять $h = 0{,}1$. Поэтому введем переменную $X = \dfrac{x-0{,}5}{0{,}1}$.

По формулам (1.72) находим оценки коэффициентов линейной регрессии

$$\tilde{B}_1 = \frac{\sum_{i=1}^{n} Y_i}{n} = \frac{132{,}47}{6} = 22{,}078;,\ \tilde{B}_2 = \frac{\sum_{i=1}^{n} Y_i X_i}{\sum_{i=1}^{n} X_i^2} = \frac{100{,}19}{70} = 1{,}431.$$

Получили линейную модель регрессии

$$\widehat{Y}_{\text{лин}} = 22{,}078 + 1{,}431X. \qquad (2.8)$$

Вычисляем значения линейной модели регрессии $Y_{\text{лин}}$ по формуле (2.8) при всех значениях аргумента X, а затем рассчитываем $\Delta Y_{i\,\text{лин}} = Y_i - Y_{i\,\text{лин}}$ отклонения экспериментальных значений Y_i от значений $Y_{i\,\text{лин}}$, полученных по функции регрессии. Сумма $\sum_{i=1}^{n} \Delta Y_{i\,\text{лин}}$ не равна нулю в силу того, что при вычислении коэффициентов регрессии результаты округлялись с точностью трех знаков после запятой и при вычислении суммы отклонений набежала ошибка округления. Далее суммируем значения ΔY_i^2 для нахождения дисперсии адекватности (1.74):

$$S^2_{\text{ад. лин}} = \frac{\sum_{i=1}^{n} (\Delta Y_i)^2}{n-2} = \frac{19{,}351}{4} = 4{,}838.$$

Все расчеты приведены в табл. 2.5.

Проверяем адекватность линейной модели регрессии, используя критерий Фишера (1.67):

$$F_{\text{лин}} = \frac{S^2_{\text{ад лин}}}{S^2_{\text{эксп}}} = \frac{4{,}838}{0{,}176} = 27{,}487.$$

Квантиль распределения Фишера $F_{1-\alpha}(k_{\text{ад. лин}}, k_{\text{эксп}}) = F_{0{,}95}(4;19) = 2{,}90.$ Так как

$$F_{\text{лин}} = 27{,}487 > 2{,}90 = F_{1-\alpha}(k_{\text{ад. лин}}, k_{\text{эксп}}),$$

то гипотеза об адекватности линейной модели регрессии отвергается.

Построим квадратичную модель регрессии с кодированной переменной (1.80):

$$y = B_1 + B_2 X + B_3 X^2 = B_1 + B_2 \frac{x - 0{,}5}{0{,}1} + B_3 \left(\frac{x - 0{,}5}{0{,}1} \right)^2. \qquad (2.9)$$

Заметим, что в нашем случае $\sum_{i=1}^{n} X_i^3 = 0$, поэтому для оценок коэффициентов регрессии можно воспользоваться формулами (1.82). Все расчеты приведены в табл. 2.6.

Таблица 2.6

Результаты расчета квадратичной модели регрессии (к задаче 2)

X^3	YX^2	X^4	$Y_{кв}$	$\Delta Y_{кв}$	$\Delta Y_{кв}{}^2$
–125	307,00	626	12,558	–0,278	0,07728
–27	167,94	81	18,258	0,402	0,16160
–1	22,80	1	22,540	0,26	0,06760
1	24,97	1	25,402	–0,432	0,18662
27	240,30	81	26,844	–0,144	0,02074
125	676,50	625	26,878	0,182	0,03312
0	1439,51	1414		–0,010	0, 54996

$$\tilde{B}_1 = \frac{\sum_{i=1}^{n} X_i^4 \cdot \sum_{i=1}^{n} Y_i - \sum_{i=1}^{n} X_i^2 \cdot \sum_{i=1}^{n} Y_i X_i^2}{n\sum_{i=1}^{n} X_i^4 - \left(\sum_{i=1}^{n} X_i^2\right)^2} = \frac{1414 \cdot 132,47 - 70 \cdot 1439,51}{6 \cdot 1414 - 70^2} = 24,148;$$

$$\tilde{B}_2 = \frac{\sum_{i=1}^{n} Y_i X_i}{\sum_{i=1}^{n} X_i^2} = \frac{100,19}{70} = 1,431;$$

$$\tilde{B}_3 = \frac{n\sum_{i=1}^{n} Y_i X_i^2 - \sum_{i=1}^{n} X_i^2 \cdot \sum_{i=1}^{n} Y_i}{n\sum_{i=1}^{n} X_i^4 - \left(\sum_{i=1}^{n} X_i^2\right)^2} = \frac{6 \cdot 1439,51 - 70 \cdot 132,47}{6 \cdot 1414 - 70^2} = -0,1774 .$$

Получили квадратичную модель регрессии

$$Y_{кв} = 24,148 + 1,431X - 0,1774X^2. \qquad (2.10)$$

Для расчета дисперсии адекватности (1.83) вычисляем $\Delta Y_{i\,кв} = Y_i - Y_{i\,кв}$ отклонения экспериментальных значений Y_i от значений $Y_{i\,кв}$, полученных по функции регрессии. Отличие $\sum_{i=1}^{n} Y_{i\,кв}$ от нуля объясняется ошибками округления. Проверяем адекватность квадратичной модели регрессии:

$$S^2_{\text{ад. кв}} = \frac{\sum\limits_{i=1}^{n} (\Delta Y_{i\,\text{кв}})^2}{n-3} = 0{,}1833. \quad F_{\text{кв}} = \frac{S^2_{\text{ад. кв}}}{S^2_{\text{эксп}}} = \frac{0{,}1833}{0{,}176} = 1{,}042.$$

Квантиль распределения Фишера $F_{1-\alpha}(k_{\text{ад. кв}}, k_{\text{эксп}}) = F_{0,95}(3;19) = 3{,}13$. Так как

$$F_{\text{кв}} = 1{,}042 < 3{,}13 = F_{1-\alpha}(k_{\text{ад. кв}}, k_{\text{эксп}}),$$

то гипотеза об адекватности квадратичной модели регрессии принимается.

Найдем доверительные интервалы для коэффициентов регрессии. Границы доверительных интервалов (1.84) для коэффициентов B_1, B_2 и B_3:

$$B_1 = 24{,}148 \pm \varepsilon_1;$$

$$\varepsilon_1 = t_{1-\frac{\alpha}{2}}(n-3)\, S_{\text{ад. кв}} \sqrt{\frac{\sum\limits_{i=1}^{n} X_i^4}{n\sum\limits_{i=1}^{n} X_i^4 - \left(\sum\limits_{i=1}^{n} X_i^2\right)^2}} =$$

$$= 3{,}182\sqrt{0{,}1833} \cdot \sqrt{\frac{1414}{6 \cdot 1414 - 70^2}} = 0{,}856;$$

$$B_2 = 1{,}431 \pm \varepsilon_2;$$

$$\varepsilon_2 = t_{1-\frac{\alpha}{2}}(n-3)\, S_{\text{ад. кв}} \frac{1}{\sqrt{\sum\limits_{i=1}^{n} X_i^2}} = 3{,}182\sqrt{0{,}1833} \cdot \frac{1}{\sqrt{70}} = 0{,}1628;$$

$$B_3 = -0{,}1774 \pm \varepsilon_3;$$

$$\varepsilon_3 = t_{1-\frac{\alpha}{2}}(n-3)\, S_{\text{ад. кв}} \sqrt{\frac{n}{n\sum\limits_{i=1}^{n} X_i^4 - \left(\sum\limits_{i=1}^{n} X_i^2\right)^2}} =$$

$$= 3{,}182\sqrt{0{,}1833} \cdot \sqrt{\frac{6}{6 \cdot 1414 - 70^2}} = 0{,}05574.$$

Уравнение квадратичной регрессии Y от исходного переменного x найдем, сделав преобразование

$$Y_{\text{кв}} = \beta_1 + \beta_2 x + \beta_3 x^2 = 24{,}148 + 1{,}431\frac{x-0{,}5}{0{,}1} - 0{,}1774\left(\frac{x-0{,}5}{0{,}1}\right)^2 =$$

$$= 12{,}558 + 32{,}05x - 17{,}74x^2.$$

Границы доверительных интервалов для коэффициентов β_1, β_2 и β_3 (1.86):

$$\beta_1 = 12{,}558 \pm \hat{\varepsilon}_1;$$

$$\hat{\varepsilon}_1 = t_{1-\frac{\alpha}{2}}(n-3)\,S_{\text{ад. кв}}\sqrt{\frac{\sum\limits_{i=1}^{n} X_i^4 + \frac{\overline{x}^4}{h^4}n}{n\sum\limits_{i=1}^{n} X_i^4 - \left(\sum\limits_{i=1}^{n} X_i^2\right)^2} + \frac{\overline{x}^2}{h^2\sum\limits_{i=1}^{n} X_i^2}} =$$

$$= 3{,}182\sqrt{0{,}1833}\cdot\sqrt{\frac{1414 + 6\cdot 625}{6\cdot 1414 - 70^2} + \frac{25}{70}} = 1{,}8267\,;$$

$$\beta_2 = 32{,}05 \pm \hat{\varepsilon}_2;$$

$$\hat{\varepsilon}_2 = t_{1-\frac{\alpha}{2}}(n-3)\frac{S_{\text{ад. кв}}}{h}\sqrt{\frac{1}{\sum\limits_{i=1}^{n} X_i^2} + \frac{4n\overline{x}^2}{h^2\left(n\sum\limits_{i=1}^{n} X_i^4 - \left(\sum\limits_{i=1}^{n} X_i^2\right)^2\right)}} =$$

$$= 3{,}182\frac{\sqrt{0{,}1833}}{0{,}1}\cdot\sqrt{\frac{1}{70} + \frac{24\cdot 25}{6\cdot 1414 - 70^2}} = 5{,}8069\,;$$

$$\beta_3 = -17{,}74 \pm \hat{\varepsilon}_3;$$

$$\hat{\varepsilon}_3 = t_{1-\frac{\alpha}{2}}(n-3)\frac{S_{\text{ад. кв}}}{h^2}\sqrt{\frac{n}{n\sum\limits_{i=1}^{n} X_i^4 - \left(\sum\limits_{i=1}^{n} X_i^2\right)^2}} =$$

$$= 3{,}182\frac{\sqrt{0{,}1833}}{0{,}01}\cdot\sqrt{\frac{6}{6\cdot 1414 - 70^2}} = 5{,}5740\,.$$

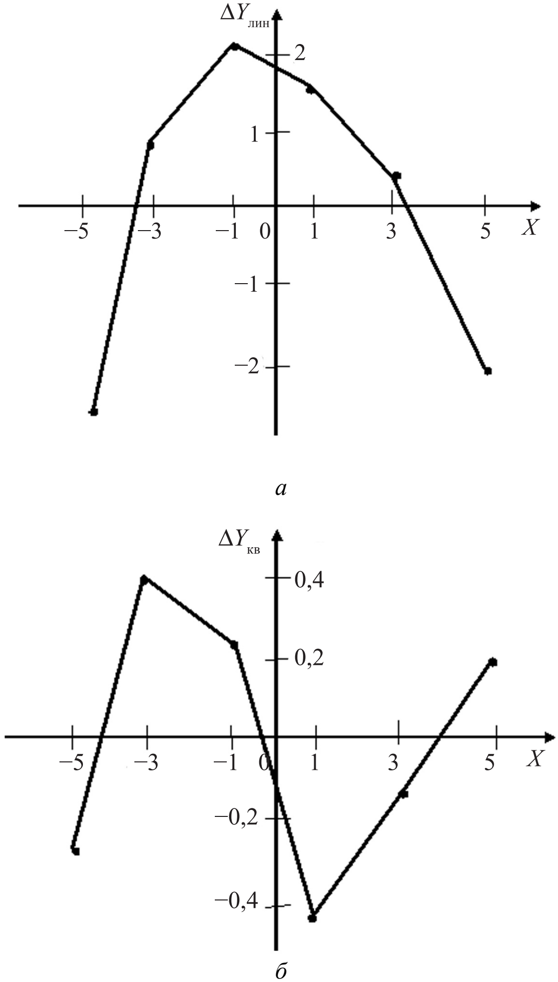

Рис. 2.6. Графики отклонения экспериментальных данных от линейной модели (*а*) и от квадратичной модели (*б*)

Для графического анализа результатов расчета построены графики отклонений линейной и квадратичной моделей регрессии от экспериментальных данных.

На рис. 2.6, *а* представлен график отклонений $\Delta Y_{лин}$ по приведенной выше табл. 2.5. Из рис. 2.6, *а* видна не только непригодность линейной модели (что следует уже из больших значений отклонений $\Delta Y_{лин}$, явно превышающих погрешность эксперимента, но и целесообразность расчета квадратичной модели, так как расположение точек наводит на мысль о параболе.

На рис. 2.6, *б* представлен график отклонений $\Delta Y_{кв}$ по табл. 2.6, но построенный уже в другом масштабе с увеличением в 5 раз. Из рис. 2.6, *б* видно, что отклонения от параболы, т.е. $|\Delta Y_{кв}|$ малы (имеют порядок ошибок эксперимента); это свидетельствует о соответствии квадратичной модели регрессии результатам эксперимента.

Надо также графически сравнить линейную и квадратичную модели с экспериментальными точками. На рис. 2.7 такое сравнение проведено для рассматриваемого примера; оно показывает соответствие квадратичной модели с экспериментом.

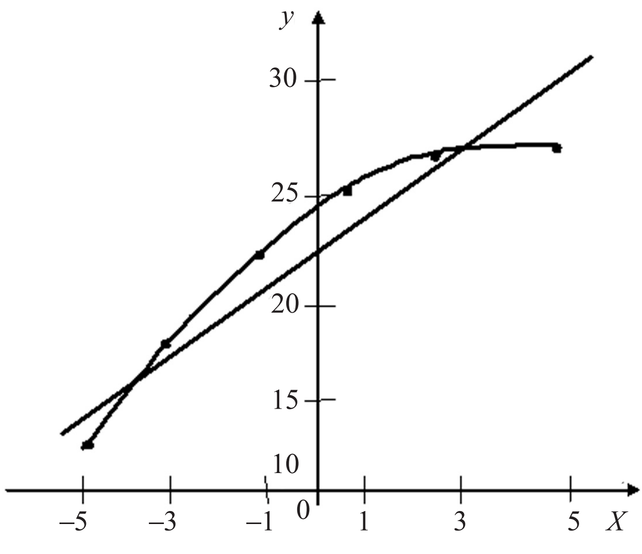

Рис. 2.7. Сравнение линейной и квадратичной моделей с данными эксперимента

Выводы. По приведенным исходным данным построенная методом наименьших квадратов линейная модель регрессии неадекватна, квадратичная модель адекватна (уровень значимости α = 0,05).

$$S^2_{ад.\,кв} = 0{,}1833\,.$$

Уравнение регрессии в кодированных значениях переменной $X = \dfrac{x - 0,5}{0,1}$:

$$y_{кв} = B_1 + B_2 X + B_3 X^2,$$

где $B_1 = 24,148 \pm 0,856$; $B_2 = 1,431 \pm 0,1628$; $B_3 = -0,1774 \pm 0,05574$.

Уравнение регрессии в исходной переменной x:

$$y_{кв} = \beta_1 + \beta_2 x + \beta_3 x^2,$$

где $\beta_1 = 12,558 \pm 1,8267$; $\beta_2 = 32,05 \pm 5,8069$; $\beta_3 = -17,74 \pm 5,5740$.

2.3. Типовой расчет 3. Обработка данных методами линейного корреляционного анализа

Заданы результаты n экспериментов, в каждом из которых измерены значения двух величин x и y, т.е. выборка объема n, извлеченная из двумерной нормальной генеральной совокупности (X, Y).

Требуется найти:

– оценки характеристик наблюдаемого двумерного случайного вектора;

– оценку коэффициента корреляции;

– эмпирические уравнения линейной регрессии;

– проверить гипотезу об отсутствии линейной зависимости между величинами X и Y;

– сделать вывод о силе и характере связи между величинами X и Y.

2.3.1. Нахождение оценок числовых характеристик двумерного случайного вектора. Расчет оценки коэффициента корреляции

Определим оценки числовых характеристик двумерного случайного вектора. За оценку математических ожиданий a_x и a_y принимают средние арифметические $\overline{X}$ и $\overline{Y}$ (1.3), за оценку дисперсий σ_x^2 и σ_y^2 – соответствующие эмпирические дисперсии S_x^2 и S_y^2, вычисленные по формуле (1.5). Несмещенная оценка ковариации $\tilde{K}_{xy}$ находится по формуле (1.92).

Для упрощения расчетов и последующего контроля результатов вычислений следует провести кодировку данных по формуле (1.94). Оценки находятся по формулам (1.9), (1.10), (1.6), (1.95).

Для контроля вычислений весь расчет необходимо повторить с другим началом отсчета. Результаты должны совпадать с точностью до возможных ошибок округления. Если результаты расчетов совпадают, находится оценка коэффициента корреляции по формуле (1.96).

Пример расчета приведен в задаче 1.28.

2.3.2. Нахождение уравнений линейной регрессии

Записать выборочные уравнения линейной регрессии Y на X и X на Y. На одном чертеже построить прямые регрессии и нанести все экспериментальные точки.

Выборочные уравнения линейной регрессии получают по формулам (1.97) – (1.98). Пример расчета и построения уравнений линейной регрессии приведен в задаче 1.28.

2.3.3. Построение доверительного интервала для коэффициента корреляции ρ. Проверка гипотезы о существовании линейной зависимости между величинами X и Y

Найти доверительный интервал для коэффициента корреляции и проверить гипотезу об отсутствии линейной зависимости между величинами X и Y. Сделать вывод о силе и характере связи между величинами X и Y. Уровень значимости α при проверке гипотезы задает преподаватель. Доверительная вероятность $P = 1 - \alpha$.

Методика нахождения доверительного интервала для коэффициента корреляции и проверки гипотезы об отсутствии линейной зависимости между величинами X и Y разобраны в разд. 1.9.2. Пример расчета и выводов по работе приведен в задаче 1.29.

2.4. Примерные варианты контрольной работы

Во всех вариантах предполагается, что соответствующая генеральная совокупность имеет нормальное распределение.

Вариант 1

1. Задана случайная выборка: {0,786; 0,782; 0,779; 0,791; 0,785; 0,784} – результаты независимых равноточных измерений. Найти оценку математического ожидания, дисперсии, среднего квадратического отклонения.

2. По результатам задачи 1 построить доверительные интервалы для математического ожидания и среднего квадратического отклонения с доверительной вероятностью 0,95.

3. Согласно протоколу на поставку партия листового проката может быть сдана заказчику, если дисперсия толщины листов не превышает значения 0,15 с уровнем значимости $\alpha = 0{,}05$. Было проведено 76 измерений толщины прокатанных листов. Оценка дисперсии оказалась равной 0,18. Может ли быть сдана данная партия листового проката?

4. В двух сериях независимых измерений с числом экспериментов соответственно $n_1 = 13$ и $n_2 = 9$ получены оценки математического ожидания $\overline{x}_1 = 10{,}7$; $\overline{x}_2 = 9{,}8$ и оценки дисперсий $S_1^2 = 0{,}45$; $S_2^2 = 1{,}44$. Проверить с уровнем значимости $\alpha = 0{,}05$:

1) гипотезу о равенстве дисперсий;

2) гипотезу о равенстве математических ожиданий.

В качестве альтернативных гипотез рассмотреть:

1) а) $H_1:\ \sigma_1^2 \neq \sigma_2^2$; б) $H_1:\ \sigma_1^2 < \sigma_2^2$.

2) а) $H_1:\ a_1 \neq a_2$; б) $H_1:\ a_1 > a_2$.

5. В табл. 2.7 представлены экспериментальные данные зависимости y от x. Результаты измерения величины y являются независимыми, равноточными, имеют нормальный закон распределения.

Таблица 2.7

x	0,2	0,3	0,4	0,7	0,9
Y	3,4	3,0	3,0	1,7	1,4

По отдельной серии из $n = 11$ повторных измерений получена оценка дисперсии $S^2 = 0{,}012$. Построить линейную модель регрессии. Проверить адекватность полученной модели с уровнем значимости $\alpha = 0{,}05$. Построить график модели регрессии.

6. При изучении зависимости величин X и Y по серии из $n = 50$ измерений получена эмпирическая матрица ковариаций:

$$\begin{pmatrix} 1{,}80 & 0{,}72 \\ 0{,}72 & 0{,}80 \end{pmatrix}.$$

Найти выборочный коэффициент корреляции. Проверить гипотезу о существовании линейной зависимости между X и Y с уровнем значимости $\alpha = 0{,}05$. Если гипотеза принимается, сделать вывод о силе и направлении взаимосвязи между X и Y.

Вариант 2

1. Случайная выборка (результаты независимых равноточных измерений) задана табл. 2.8.

Таблица 2.8

x_i	3	5	6	7
Частота n_i	8	9	10	3

Найти оценку математического ожидания, дисперсии, среднего квадратического отклонения. Построить полигон и гистограмму.

2. По результатам задачи 1 найти доверительные интервалы для математического ожидания и среднего квадратического отклонения с доверительной вероятностью 0,99.

3. Для выполнения заказа на поставку стальных прутков диаметром 45 мм необходимо объединить продукцию двух станов. Это возможно лишь в том случае, если оба прокатных стана обеспечивают одинаковую точность проката. Для принятия решения об объединении двух партий в одну проконтролировали размеры случайно выбранных прутков: 31 пруток с первого стана и 28 прутков со второго. Оценки дисперсий полученных выборок получились равными $S_1^2 = 0{,}27$ и $S_2^2 = 0{,}12$ соответственно. С уровнем значимости $\alpha = 0{,}01$ принять решение о возможности объединения продукции в одну партию. В качестве альтернативной гипотезы рассмотреть два варианта: а) второй стан обеспечивает более высокую точность; б) станы производят продукцию с различной точностью.

4. В табл. 2.9 приведены результаты 10 экспериментов по измерению изменения состава металла при выпуске из конвертера: X – изменение содержания азота, %·10 000; Y – начальная концентрация углерода, %.

Таблица 2.9

X	–3	–2,5	3,5	–1	–1,5	1,5	–4	7	–2	0
Y	0,12	0,10	0,08	0,06	0,08	0,10	0,14	0,03	0,12	0,08

Найти: а) эмпирическую матрицу ковариаций; б) выборочный коэффициент корреляции.

5. В условиях задачи 4 проверить гипотезу о существовании линейной зависимости между X и Y с уровнем значимости $\alpha = 0{,}05$. Если гипотеза принимается, сделать вывод о направлении взаимосвязи между X и Y.

6. В табл. 2.9 представлены экспериментальные данные зависимости Y от x. Построить линейную модель регрессии. На чертеж нанести экспериментальные точки и график полученной модели.

Вариант 3

1. Дана случайная выборка {0,127; 0,129; 0,133; 0,131; 0,136} – результаты независимых равноточных измерений. Найти оценку математического ожидания, дисперсии, среднего квадратического отклонения.

2. По результатам задачи 1 построить доверительные интервалы для математического ожидания и среднего квадратического отклонения с доверительной вероятностью 0,9.

3. По выборке объема $n = 30$ рассчитана оценка дисперсии $S^2 = 2{,}5$. Считая, что математическое ожидание известно, проверить гипотезу о том, что дисперсия ГС равна $\sigma_0^2 = 2$ с уровнем значимости $\alpha = 0{,}1$. Альтернативная гипотеза $H_1: \quad \sigma^2 \neq \sigma_0^2$.

4. По двум независимым выборкам объемов $n_1 = 20$ и $n_2 = 45$ найдены оценки математического ожидания $\overline{x}_1 = 16$; $\overline{x}_2 = 19$. Дисперсии генеральной совокупности известны: $\sigma_1^2 = 2$; $\sigma_2^2 = 4$. При уровне значимости $\alpha = 0{,}1$ проверить гипотезу о равенстве математических ожиданий. Альтернативная гипотеза $H_1: \quad a_1 \neq a_2$.

5. В табл. 2.10 представлены экспериментальные данные зависимости Y от x. Результаты измерения величины y являются независимыми, равноточными, имеют нормальный закон распределения.

Таблица 2.10

x	1,0	1,2	1,3	1,7	1,8	2,0
Y	–8	–1	–1	11	11	18

По отдельной серии из $n = 19$ повторных измерений получена оценка дисперсии $S^2 = 0{,}65$. Построить линейную модель регрессии.

Проверить адекватность полученной модели с уровнем значимости $\alpha = 0{,}05$. Построить график модели регрессии.

6. При изучении зависимости величин X_1 и X_2 по серии из $n = 20$ измерений получены оценки математического ожидания $\overline{X}_1 = 2$; $\overline{X}_2 = 2{,}5$; оценки дисперсий $S_1^2 = 0{,}25$; $S_2^2 = 0{,}49$ и оценка ковариации $\tilde{K}_{1,2} = 0{,}28$. Найти выборочный коэффициент корреляции r. Построить доверительный интервал для коэффициента корреляции с доверительной вероятностью $P = 0{,}95$, уровнем значимости $\alpha = 0{,}05$. Проверить гипотезу о существовании линейной зависимости между X_1 и X_2 с уровнем значимости $\alpha = 0{,}05$. Написать выборочные уравнения прямых регрессии.

Вариант 4

1. Задана случайная выборка {3,23; 3,18; 3,25; 3,22; 3,20; 3,26; 3,27} – результаты независимых равноточных измерений. Найти оценку математического ожидания, дисперсии, среднего квадратического отклонения.

2. По результатам задачи 1 построить доверительный интервал для математического ожидания и дисперсии с доверительной вероятностью 0,95.

3. В двух сериях независимых экспериментов с числом измерений, соответственно, $n_1 = 15$ и $n_2 = 10$ получены оценки математического ожидания $\overline{X}_1 = 20{,}5$; $\overline{X}_2 = 18{,}7$ и оценки дисперсии $S_1^2 = 2{,}2$; $S_2^2 = 5{,}5$. Известно, что результаты измерений в каждой серии имеют нормальный закон распределения. Используя двусторонние критерии, проверить с уровнем значимости $\alpha = 0{,}05$ гипотезу о равенстве дисперсий и гипотезу о равенстве математических ожиданий.

4. По случайной нормальной выборке объема $n = 18$ найдена оценка дисперсии $S^2 = 14{,}6$. Проверить гипотезу H_0 о равенстве дисперсии выборки σ^2 значению $\sigma_0^2 = 20$ при альтернативной гипотезе H_1: $\sigma^2 < \sigma_0^2$ с уровнем значимости $\alpha = 0{,}05$.

5. В табл. 2.11 представлены экспериментальные данные двумерной выборки (X, Y). Найти матрицу ковариаций и коэффициент корреляции случайной величины (X, Y).

Таблица 2.11

X	0,5	–4,0	2,5	4,5	–0,5	–0,5	–2,0	0,0
Y	0,22	0,12	0,24	0,24	0,17	0,18	0,14	0,18

6. По результатам задачи 5 проверить гипотезу о существовании линейной зависимости между X и Y с уровнем значимости $\alpha = 0{,}05$. Написать уравнения прямых регрессии, построить их графики.

Вариант 5

1. Задана случайная выборка {71,3; 71,8; 71,1; 71,4; 71,6} – результаты независимых равноточных измерений. Найти оценку математического ожидания, дисперсии, среднего квадратического отклонения.

2. По результатам задачи 1 построить доверительный интервал для для математического ожидания и дисперсии с доверительной вероятностью 0,95.

3. По случайной нормальной выборке объема $n = 16$ найдена оценка математического ожидания $\overline{X} = 22{,}5$. Проверить гипотезу H_0 о равенстве математического ожидания выборки a значению $a_0 = 24{,}0$ при альтернативной гипотезе H_1: $a < a_0$ с уровнем значимости $a = 0{,}05$, если оценка σ, найденная по той же выборке, равна $S = 3{,}2$.

4. Номинальный диаметр шариков, изготовляемых станком-автоматом, равен $d_0 = 20$ мм. В выборке из 16 шариков средний диаметр оказался равным 20,5 мм. Используя двусторонний критерий при уровне значимости $\alpha = 0{,}01$, проверить гипотезу о соответствии партии стандарту, если дисперсия известна и равна $\sigma^2 = 0{,}49$.

5. При изучении зависимости величин X и Y по серии из $n = 20$ измерений получена эмпирическая матрица ковариаций $\begin{pmatrix} 2{,}00 & 2{,}25 \\ 2{,}25 & 4{,}50 \end{pmatrix}$. $\overline{X} = 0{,}5$; $\overline{Y} = 1{,}5$. Проверить гипотезу о существовании линейной зависимости между X и Y с уровнем значимости $\alpha = 0{,}05$. Сделать вывод о силе и направлении взаимосвязи между X и Y. Написать уравнения прямых регрессии, построить их графики.

6. В табл. 2.12 представлены экспериментальные данные зависимости Y от X. Результаты измерения величины Y являются независимыми, равноточными, имеют нормальный закон распределения.

Таблица 2.12

X	0,1	0,2	0,3	0,6	0,8
Y	0,1	0,5	0,9	1,8	2,2

По отдельной серии из n = 13 повторных измерений получена оценка дисперсии S^2 = 0,003. Построить линейную модель регрессии. Проверить адекватность полученной модели с уровнем значимости a = 0,05. Построить график полученной модели.

Ответы к вариантам контрольной работы

Вариант 1

1. $\overline{x} = 0{,}7845$; $S^2 = 16{,}3\cdot 10^{-6}$; $S = 4{,}04\cdot 10^{-3}$.

2. $a = 0{,}7845 \pm 4{,}24\cdot 10^{-3}$, или $a \in (0{,}7803; 0{,}7887)$; $\sigma \in (2{,}52\cdot 10^{-3}; 9{,}90\cdot 10^{-3})$.

3. Да, так как $H_0: \quad \sigma^2 = \sigma_0^2$ при $H_1: \quad \sigma^2 > \sigma_0^2$.
$Z = 90 < \chi^2_{0,05}(75) = 96{,}2$.

4. 1а) $F = 3{,}2 < F_{\text{кр}} = F_{0,975}(8; 12) = 3{,}51 \Rightarrow H_0: \quad \sigma_1^2 = \sigma_2^2$ принимается $S_{\text{св}}$ = 0,920.

1б) $F = 3{,}2 > F_{\text{кр}} = F_{0,95}(8;\ 12) = 2{,}85 \Rightarrow H_0$ отвергается, принимается альтернативная гипотеза $H_1: \quad \sigma_1^2 < \sigma_2^2$.

2а) $t = 2{,}256 > t_{\text{кр}} = t_{0,975}(20) = 2{,}09 \Rightarrow H_0: \quad a_1 = a_2$ отвергается, принимается гипотеза $H_1: \quad a_1 \neq a_2$.

2б) $t = 2{,}256 > t_{\text{кр}} = t_{0,95}(20) = 1{,}73 \Rightarrow H_0$ отвергается, принимается альтернативная гипотеза $H_1: \quad a_1 > a_2$.

5. $y = 2{,}5 - 0{,}3X$, где $X = 10(x - 0{,}5)$, $\Rightarrow$ $y = -3x + 4$; $S^2_{\text{ад}} = 0{,}033$.

$F = 2{,}75 < F_{\text{кр}} = F_{0,95}(3;\ 10) = 3{,}71 \Rightarrow$ линейная модель адекватна.

6. $r = 0{,}6$; доверительный интервал $\rho \in (0{,}4; 0{,}75)$; T = 0,279. Гипотеза о существовании линейной зависимости между X и Y не противоречит экспериментальным данным. С увеличением X среднее значение Y увеличивается. Связь между X и Y достаточно сильная.

Вариант 2

1. $\overline{x}$ = 5; S^2 = 1,862; S = 1,365.

2. $a = 5 \pm 0{,}69$ или $a \in (4{,}31; 5{,}69)$; $\sigma \in (1{,}044; 1{,}944)$.

3. а) $F = 2,25 < F_{\text{кр}} = F_{0,99}(30;\ 27) = 2,47 \ \Rightarrow$ гипотеза о равенстве дисперсий принимается;

б) $F = 2,25 < F_{\text{кр}} = F_{0,995}(30;\ 27) = 2,73 \ \Rightarrow$ гипотеза о равенстве дисперсий принимается.

4. а) $\tilde{K} = \begin{pmatrix} 11,29 & -8,53\cdot 10^{-2} \\ -8,53\cdot 10^{-2} & 10,32\cdot 10^{-4} \end{pmatrix}$; б) $r = -0,79$.

5. $\rho \in (-0,85; -0,3)$. Так как значение $\rho = 0$ не принадлежит найденному доверительному интервалу, гипотеза о существовании линейной зависимости не противоречит экспериментальным данным. $T = 0,632 \ < |r|$, проверка с помощью критерия подтверждает принятие гипотезы о существовании линейной зависимости. С увеличением значения одной величины среднее значение другой уменьшается.

6. $y_{\text{лин}} = 9,1\cdot 10^{-2} - 0,75\cdot 10^{-2}(x - 0,2)$.

Вариант 3

1. $\bar{x} = 0,1312$; $S^2 = 12,2\cdot 10^{-6}$; $S = 4,35\cdot 10^{-3}$.

2. $a = 0,1312 \pm 0,0044$ или $a \in (0,1268; 0,1356)$; $\sigma \in (2,095\cdot 10^{-3}; 10,03\cdot 10^{-3})$.

3. $Z = 37,5 \in (18,49; 43,77) \ \Rightarrow$ гипотеза о равенстве дисперсий принимается с уровнем значимости α.

4. $|U| \approx 6,9 > u_{\text{кр}} = 1,645 \ \Rightarrow \ H_0$ гипотеза отвергается с уровнем значимости α, принимается гипотеза H_1.

5. $y = 5,0 + 2,5\dfrac{x - 1,5}{0,1} = 5,0 + 25(x - 1,5)$. $S^2_{ад} = 1,75$.

$F = 2,69 < F_{\text{кр}} = F_{0,95}(4;\ 18) = 2,93 \ \Rightarrow$ линейная модель адекватна.

6. $r = 0,8$; $\rho \in (0,54; 0,96)$; $T = 0,444$. Гипотеза о существовании линейной зависимости между X_1 и X_2 принимается. С увеличением X_1 среднее значение X_1 увеличивается. Связь между X и Y достаточно сильная.

Регрессия X_2 на X_1: $\dfrac{x_2 - 2{,}5}{0{,}7} = 0{,}8\dfrac{x_1 - 2}{0{,}5} \Leftrightarrow x_2 = 1{,}12x_1 + 0{,}26$.

Регрессия X_1 на X_2: $\dfrac{x_2 - 2{,}5}{0{,}7} = \dfrac{1}{0{,}8} \cdot \dfrac{x_1 - 2}{0{,}5} \Leftrightarrow x_2 = 1{,}75x_1 - 1{,}0$.

Вариант 4

1. $\overline{x} = 3{,}230$; $S^2 = 10{,}7 \cdot 10^{-4}$; $S = 3{,}27 \cdot 10^{-2}$.

2. $a = 3{,}230 \pm 0{,}0302$, или $a \in (3{,}200; 3{,}260)$; $\sigma^2 \in (4{,}445 \cdot 10^{-4}; 5{,}161 \cdot 10^{-3})$.

3. $F = 2{,}5$; $F_{кр} = F_{0,975}(9;\ 14) = 3{,}21$. Гипотеза о равенстве дисперсий принимается.

$S^2_{св} = 3{,}491$; $S_{св} = 1{,}8685$; $t_{кр} = t_{0,975} = 2{,}069$. Гипотеза о равенстве математических ожиданий отвергается.

4. $Z = \dfrac{kS^2}{\sigma_0^2} = 12{,}41$; $\chi^2_{0,005}(17) = 8{,}67$; $\Rightarrow H_0$ принимается.

5. $S^2_x = 6{,}747$; $S^2_y = 19{,}7\ 10^{-4}$; $S_{xy} = 0{,}1088$; $r = 0{,}9442$.

6. $T = 0{,}707$. Регрессия Y на X: $y - 0{,}186 = 0{,}0161(x - 0{,}0625)$.

Регрессия X на Y: $y - 0{,}186 = 0{,}0181(x - 0{,}0625)$.

Вариант 5

1. $\overline{x} = 71{,}44$; $S^2 = 0{,}073$; $S = 0{,}27$.

2. $a = 71{,}44 \pm 0{,}335$, или $a \in (71{,}105; 71{,}775)$; $\sigma^2 \in (0{,}0262; 0{,}6021)$.

3. $t = -1{,}875$; $t_{кр} = t_{0,05}(15) = -1{,}753$. $t < t_{кр} \Rightarrow H_0$ отвергается.

4. $u = 2{,}86$; $u_{кр} = u_{0,975} = 2{,}576$. $u > u_{кр} \Rightarrow H_0$ отвергается.

5. $r = 0{,}75$; $\rho \in (0{,}48;\ 0{,}9)$; $T = 0{,}444 < r \Rightarrow$ Гипотеза о существовании линейной зависимости между X и Y принимается. Регрессия Y на X: $y - 1{,}5 = 1{,}125(x - 0{,}5)$. Регрессия X на Y: $y - 1{,}5 = 2x - 1$.

6. $y = 1{,}1 + 0{,}3\dfrac{x - 0{,}4}{0{,}1} = 3x - 0{,}1$. $S^2_{ад} = 0{,}0133$; $F = 4{,}44$; $F_{0,95}(3;\ 12) = 3{,}39$. Линейная модель регрессии неадекватна.

Библиографический список

1. *Карасев В.А., Лёвшина Г.Д.* Теория вероятностей и математическая статистика. Теория вероятностей: Практикум. М.: Изд. Дом МИСиС, 2015.

2. *Гмурман В.Е.* Руководство к решению задач по теории вероятностей и математической статистике. М.: Высш. шк., 1998.

3. *Гмурман В.Е.* Теория вероятностей и математическая статистика. М.: Юрайт, 2014.

4. *Кибзун А.И., Горяинова Е.Р., Наумов А.В.* Теория вероятностей и математическая статистика. Базовый курс с примерами и задачами. М.: Физматлит, 2014. 232 с.

5. Сборник задач по математике для втузов Ч. 4. Теория вероятностей и математическая статистика/ Под ред. А.В. Ефимова и А.С. Поспелова. М.: Физматлит, 2004.

6. Математика для экономистов: Учеб. / О.В. Татарников, Р.В. Сагитов, А.С. Чуйко и др. М.: Юрайт, 2015. Сер. 58 Бакалавр. Академический курс.

7. Математика для экономистов: Практикум / О.В. Татарников, Л.Г. Бирюкова, Г.И. Бобрик и др.: М.: Юрайт, 2015. Сер. 58 Бакалавр. Академический курс.

8. *Большев Л.Н., Смирнов Н.В.* Таблицы математической статистики. Изд. 3-е. М.: Наука, 1983.

ПРИЛОЖЕНИЕ

Таблица П1

Квантили u_p нормального распределения $N(0,1)$

p	0,90	0,95	0,975	0,99	0,995	0,999	0,9995
u_p	1,282	1,645	1,960	2,326	2,576	3,090	3,291

Таблица П2

Квантили распределения Стьюдента $t_p(k)$

Число степеней свободы k	Доверительная вероятность p				
	0,900	0,950	0,975	0,990	0,995
1	3,078	6,314	12,708	31,821	63,657
2	1,886	2,920	4,303	6,965	9,925
3	1,638	2,353	3,182	4,541	5,841
4	1,533	2,132	2,776	3,747	4,604
5	1,476	2,015	2,571	3,365	4,032
6	1,440	1,943	2,447	3,143	3,707
7	1,415	1,895	2,365	2,998	3,499
8	1,397	1,860	2,306	2,896	3,355
9	1,383	1,833	2,262	2,821	3,250
10	1,372	1,812	2,228	2,764	3,169
11	1,363	1,796	2,201	2,718	3,106
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	2,624	2,997
15	1,341	1,753	2,131	2,602	2,947
16	1,337	1,746	2,120	2,583	2,921
17	1,333	1,740	2,110	2,567	2,898
18	1,330	1,734	2,101	2,552	2,878
19	1,328	1,729	2,093	2,539	2,861
20	1,325	1,725	2,086	2,528	2,845
21	1,323	1,721	2,080	2,518	2,831
22	1,321	1,717	2,074	2,508	2,819
23	1,319	1,714	2,069	2,500	2,807
24	1,318	1,711	2,064	2,492	2,797
25	1,316	1,708	2,060	2,485	2,787
26	1,315	1,706	2,056	2,479	2,779
27	1,314	1,703	2,052	2,473	2,771
28	1,313	1,701	2,048	2,467	2,763
29	1,311	1,699	2,045	2,462	2,756
30	1,310	1,697	2,042	2,457	2,750
40	1,303	1,684	2,021	2,423	2,704
60	1,296	1,671	2,000	2,390	2,660
80	1,289	1,658	1,980	2,358	2,617
∞	1,282	1,645	1,960	2,326	2,576

Таблица П3

Квантили хи-квадрат распределения $\chi_p^2(k)$

k \ p	0,005	0,100	0,025	0,05	0,10	0,20	0,30
1	0,0000393	0,000157	0,000982	0,00393	0,0158	0,0642	0,148
2	0,0100	0,0201	0,0506	0,103	0,211	0,446	0,713
3	0,717	0,115	0,216	0,352	0,584	1,00	1,42
4	0,207	0,297	0,484	0,711	1,06	1,65	2,19
5	0,412	0,554	0,831	1,15	1,61	2,34	3,00
6	0,676	0,872	1,24	1,64	2,20	3,07	3,83
7	0,989	1,24	1,69	2,17	2,83	3,82	4,67
8	1,34	1,65	2,18	2,73	3,49	4,59	5,53
9	1,73	2,09	2,70	3,33	4,17	5,38	6,39
10	2,16	2,56	3,25	3,94	4,87	6,18	7,27
11	2,60	3,05	3,82	4,57	5,58	6,99	8,15
12	3,07	3,57	4,40	5,23	6,30	7,81	9,03
13	3,57	4,11	5,01	5,89	7,04	8,63	9,93
14	4,07	4,66	5,63	6,57	7,79	9,47	10,8
15	4,60	5,23	6,26	7,26	8,55	10,3	11,7
16	5,14	5,81	6,91	7,96	9,31	11,2	12,6
17	5,70	6,41	7,56	8,67	10,1	12,0	13,5
18	6,26	7,01	8,23	9,39	10,9	12,9	14,4
19	6,84	7,63	8,91	10,1	11,7	13,7	15,4
20	7,43	8,26	9,59	10,9	12,4	14,6	16,3
21	8,03	8,90	10,3	11,6	13,2	15,4	17,2
22	8,64	9,54	11,0	12,3	14,0	16,3	18,1
23	9,26	10,2	11,7	13,1	14,8	17,2	19,0
24	9,89	10,9	12,4	13,8	15,7	18,1	19,9
25	10,5	11,5	13,1	14,6	16,5	18,9	20,9
26	11,2	12,2	13,8	15,4	17,3	19,8	21,8
27	11,8	12,9	14,6	16,2	18,1	20,7	22,7
28	12,5	13,6	15,3	16,9	18,9	21,6	23,6
29	13,1	14,3	16,0	17,7	19,8	22,5	24,6
30	13,8	15,0	16,8	18,5	20,6	23,4	25,5
35	17,2	18,5	20,6	22,5	24,8	27,8	30,2
40	20,7	22,2	24,4	26,5	29,1	32,3	34,9
40	20,7	22,2	24,4	26,5	29,1	32,3	34,9
45	24,3	25,9	28,4	30,6	33,4	36,9	39,6
50	28,0	29,7	32,4	34,8	37,7	41,4	44,3
75	47,2	49,5	52,9	56,1	59,8	64,5	68,1
100	67,3	70,1	74,2	77,9	82,4	87,9	92,1

p / k	0,70	0,80	0,90	0,95	0,975	0,990	0,995	0,999
1	1,07	1,64	2,71	3,84	5,02	6,63	7,88	10,8
2	2,41	3,22	4,61	5,99	7,38	9,21	10,6	13,8
3	3,67	4,64	6,25	7,81	9,35	11,3	12,8	16,3
4	4,88	5,99	7,78	9,49	11,1	13,3	14,9	18,5
5	6,06	7,29	9,24	11,1	12,8	15,1	16,7	20,5
6	7,23	8,56	10,6	12,6	14,4	16,8	18,5	22,5
7	8,38	9,8	12,0	14,1	16,0	18,5	20,3	24,3
8	9,52	11,0	13,4	15,5	17,5	20,1	22,0	26,1
9	10,7	12,2	14,7	16,9	19,0	21,7	23,6	27,9
10	11,8	13,4	16,0	18,3	20,5	23,2	25,2	29,6
11	12,9	14,6	17,3	19,7	21,9	24,7	26,8	31,3
12	14,0	15,8	18,5	21,0	23,3	26,2	28,3	32,9
13	15,1	17,0	19,8	22,4	24,7	27,7	29,8	34,5
14	16,2	18,2	21,1	23,7	26,1	29,1	31,3	36,1
15	17,3	19,3	22,3	25,0	27,5	30,6	32,8	37,7
16	18,4	20,5	23,5	26,3	28,8	32,0	34,3	39,3
17	19,5	21,6	24,8	27,6	30,2	33,4	35,7	40,8
18	20,6	22,8	26,0	28,9	31,5	34,8	37,2	42,3
19	21,7	23,9	27,2	30,1	32,9	36,2	38,6	43,8
20	22,8	25,0	28,4	31,4	34,2	37,6	40,0	45,3
21	23,9	26,9	29,6	32,7	35,5	38,9	41,4	46,8
22	24,9	27,3	30,8	33,9	36,8	40,3	42,8	48,3
23	26,0	28,4	32,0	35,2	38,1	41,6	44,2	49,7
24	27,1	29,6	33,2	36,4	39,4	43,0	45,6	51,2
25	28,2	30,7	34,4	37,7	40,6	44,3	46,9	52,6
26	29,2	31,8	35,6	38,9	41,9	45,6	48,3	54,1
27	30,3	32,9	36,7	40,1	43,2	47,0	49,6	55,6
28	31,4	34,0	37,9	41,3	44,5	48,3	51,0	56,9
29	32,5	35,1	39,1	42,6	45,7	49,6	52,3	58,3
30	33,5	36,3	40,3	43,8	47,0	50,9	53,7	59,7
35	38,9	41,8	46,1	49,8	53,2	57,3	60,3	66,6
40	44,2	47,3	51,8	55,8	59,3	63,7	66,8	73,1
45	49,5	52,7	57,5	61,7	65,4	70,0	73,2	80,1
50	54,7	58,2	63,2	67,5	71,4	76,2	79,5	86,7
75	80,9	85,1	91,1	96,2	100,8	106,4	110,3	118,6
100	106,9	111,7	118,5	124,3	129,6	135,6	140,2	149,4

Таблица П4

Квантили распределения Фишера $F_p(k_1, k_2)$, $p = 0{,}95$

Число степеней свободы k_2	Число степеней свободы k_1								
	1	2	3	4	5	6	7	8	9
1	161,4	199,5	199,5	224,6	230,2	234,0	236,8	238,9	240,5
2	18,51	19,00	19,0	19,25	19,30	19,33	19,35	19,37	19,37
3	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,85
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,04
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,82
6	5,99	5,14	4,75	4,53	4,39	4,28	4,21	4,15	4,15
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,73
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,44
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,23
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,07
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,95
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,85
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,77
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,70
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,64
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,59
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,55
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,51
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,48
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,45
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,42
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,40
23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,37
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,36
25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,34
26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,32
27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,31	2,31
28	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,29
29	4,18	3,33	2,93	2,70	2,55	2,43	2,35	2,28	2,28
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,27
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,18
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,10
120	3,92	3,07	2,68	2,45	2,29	2,17	2,09	2,02	2,02
∞	3,04	3,00	2,60	2,37	2,21	2,10	2,01	1,94	1,94

Окончание табл. П4

Число степеней свободы k_2	Число степеней свободы k_1								
	10	12	15	20	24	30	40	60	120
1	241,9	243,9	245,9	248,0	249,1	250,1	251,1	252,2	253,3
2	19,40	19,41	19,43	19,45	19,45	19,46	19,47	19,48	19,49
3	8,79	8,74	8,70	8,66	8,64	8,62	8,59	8,57	8,55
4	5,96	5,91	5,86	5,80	5,77	5,75	5,72	5,69	5,66
5	4,74	4,68	4,62	4,56	4,53	4,50	4,46	4,43	4,40
6	4,06	4,00	3,94	3,87	3,84	3,81	3,77	3,74	3,70
7	3,64	3,57	3,51	3,44	3,41	3,38	3,34	3,30	3,27
8	3,35	3,28	3,22	3,15	3,12	3,08	3,04	3,01	2,97
9	3,14	3,07	3,01	2,94	2,90	2,86	2,83	2,79	2,75
10	2,98	2,91	2,85	2,77	2,74	2,70	2,66	2,62	2,53
11	2,85	2,79	2,72	2,65	2,61	2,57	2,53	2,49	2,45
12	2,75	2,69	2,62	2,54	2,51	2,47	2,43	2,38	2,34
13	2,67	2,60	2,53	2,46	2,42	2,38	2,34	2,30	2,25
14	2,60	2,53	2,46	2,39	2,35	2,31	2,27	2,22	2,18
15	2,54	2,43	2,40	2,33	2,29	2,25	2,20	2,16	2,11
16	2,49	2,42	2,35	2,28	2,24	2,19	2,15	2,11	2,06
17	2,45	2,38	2,31	2,23	2,19	2,15	2,10	2,06	2,01
18	2,41	2,34	2,27	2,19	2,15	2,11	2,06	2,02	1,97
19	2,38	2,31	2,23	2,16	2,11	2,07	2,03	1,98	1,93
20	2,35	2,28	2,20	2,12	2,08	2,04	1,99	1,95	1,90
21	2,32	2,25	2,18	2,10	2,05	2,01	1,96	1,92	1,87
22	2,30	2,23	2,15	2,07	2,03	1,98	1,94	1,89	1,84
23	2,27	2,20	2,13	2,05	2,01	1,96	1,91	1,86	1,81
24	2,25	2,18	2,11	2,03	1,98	1,94	1,89	1,84	1,79
25	2,24	2,16	2,09	2,01	1,96	1,92	7,87	1,82	1,77
26	2,22	2,15	2,07	1,99	1,95	1,90	1,85	1,80	1,75
27	2,20	2,13	2,06	1,97	1,93	1,88	1,84	1,79	1,73
28	2,19	2,12	2,04	1,96	1,91	1,87	1,82	1,77	1,71
29	2,18	2,10	2,03	1,94	1,90	1,85	1,81	1,75	1,70
30	2,16	2,09	2,01	1,93	1,89	1,84	1,79	1,74	1,68
40	2,08	2,00	1,92	1,84	1,79	1,74	1,69	1,64	1,58
60	1,99	1,92	1,84	1,75	1,70	1,65	1,59	1,53	1,47
120	1,91	1,83	1,75	1,66	1,61	1,55	1,50	1,43	1,35
∞	1,83	1,75	1,67	1,57	1,52	1,46	1,39	1,32	1,22

Таблица П5

Квантили распределения Фишера $F_p(k_1,k_2)$ **,** $p = 0{,}975$

Число степеней свободы k_2	Число степеней свободы k_1								
	1	2	3	4	5	6	7	8	9
1	647,8	799,5	864,2	899,6	921,8	937,1	948,2	956,7	963,3
2	38,51	39,00	39,17	39,25	39,30	39,33	39,36	39,37	39,39
3	17,44	16,04	15,44	15,10	14,88	14,73	14,62	14,54	14,47
4	12,22	10,65	9,98	9,60	9,36	9,20	9,07	8,98	8,90
5	10,01	8,43	7,76	7,39	7,15	6,98	6,85	6,76	6,68
6	8,81	7,26	6,60	6,23	5,99	5,82	5,70	5,60	5,52
7	8,07	6,54	5,89	5,52	5,29	5,12	4,99	4,90	4,82
8	7,57	6,06	5,42	5,05	4,82	4,65	4,53	4,43	4,36
9	7,21	5,71	5,08	4,72	4,48	4,32	4,20	4,10	4,03
10	6,94	5,46	4,83	4,47	4,24	4,07	3,95	3,85	3,78
11	6,72	5,26	4,63	4,28	4,04	3,88	3,76	3,66	3,59
12	6,55	5,10	4,47	4,12	3,89	3,73	3,61	3,51	3,44
13	6,41	4,97	4,35	4,00	3,77	3,60	3,48	3,39	3,31
14	6,30	4,86	4,24	3,89	3,66	3,50	3,38	3,29	3,21
15	6,20	4,77	4,15	3,80	3,58	3,41	3,29	3,20	3,12
16	6,12	4,69	4,08	3,73	3,50	3,34	3,22	3,12	3,05
17	6,04	4,62	4,01	3,66	3,44	3,28	3,16	3,06	2,98
18	5,98	4,56	3,95	3,61	3,38	3,22	3,10	3,01	2,93
19	5,92	4,51	3,90	3,56	3,33	3,17	3,05	2,96	2,88
20	5,87	4,46	3,86	3,51	3,29	3,13	3,01	2,91	2,84
21	5,83	4,42	3,82	3,48	3,25	3,09	2,97	2,87	2,80
22	5,79	4,38	3,78	3,44	3,22	3,05	2,93	2,84	2,76
23	5,75	4,35	3,75	3,41	3,18	3,02	2,90	2,81	2,73
24	5,72	4,32	3,72	3,38	3,15	2,99	2,87	2,78	2,70
25	5,69	4,29	3,69	3,35	3,13	2,97	2,85	2,75	2,68
26	5,66	4,27	3,67	3,33	3,10	2,94	2,82	2,73	2,65
27	5,63	4,24	3,65	3,31	3,08	2,92	2,80	2,71	2,63
28	5,61	4,22	3,63	3,29	3,06	2,90	2,78	2,69	2,61
29	5,59	4,20	3,61	3,27	3,04	2,88	2,76	2,67	2,59
30	5,57	4,18	3,59	3,25	3,03	2,87	2,75	2,65	2,57
40	5,42	4,05	3,46	3,13	2,90	2,74	2,62	2,53	2,45
60	5,29	3,93	3,34	3,01	2,79	2,63	2,51	2,41	2,33
120	5,15	3,80	3,23	2,89	2,67	2,52	2,39	2,30	2,22
∞	5,02	3,69	3,12	2,79	2,57	2,41	2,29	2,19	2,11

Число степеней свободы k_2	Число степеней свободы k_1								
	10	12	15	20	24	30	40	60	120
1	968,6	976,7	984,9	993,1	997,2	1001	1006	1010	1014
2	39,40	39,41	39,43	39,45	39,46	39,46	39,47	39,48	39,49
3	14,42	14,34	14,25	14,17	14,12	14,08	14,04	13,99	13,95
4	8,84	8,75	8,66	8,56	8,51	8,46	8,41	8,36	8,31
5	6,62	6,52	6,43	6,33	6,28	6,23	6,18	6,12	6,07
6	5,46	5,37	5,27	5,17	5,12	5,07	5,01	4,96	4,90
7	4,76	4,67	4,57	4,47	4,42	4,36	4,31	4,25	4,20
8	4,30	4,20	4,10	4,00	3,95	3,89	3,84	3,78	3,73
9	3,96	3,87	3,77	3,67	3,61	3,56	3,51	3,45	3,39
10	3,72	3,62	3,52	3,42	3,37	3,31	3,26	3,20	3,14
11	3,53	3,43	3,33	3,23	3,17	3,12	3,06	3,00	2,94
12	3,37	3,28	3,18	3,07	3,02	2,96	2,91	2,85	2,79
13	3,25	3,15	3,05	2,95	2,89	2,84	2,78	2,72	2,66
14	3,15	3,05	2,95	2,84	2,79	2,73	2,67	2,61	2,55
15	3,06	2,96	2,86	2,76	2,70	2,64	2,59	2,52	2,46
16	2,99	2,89	2,79	2,68	2,63	2,57	2,51	2,45	2,38
17	2,92	2,82	2,72	2,62	2,56	2,50	2,44	2,38	2,32
18	2,87	2,77	2,67	2,56	2,50	2,44	2,38	2,32	2,26
19	2,82	2,72	2,62	2,51	2,45	2,39	2,33	2,27	2,20
20	2,77	2,68	2,57	2,46	2,41	2,35	2,29	2,22	2,16
21	2,73	2,64	2,53	2,42	2,37	2,31	2,25	2,18	2,11
22	2,70	2,60	2,50	2,39	2,33	2,27	2,21	2,14	2,08
23	2,67	2,57	2,47	2,36	2,30	2,24	2,18	2,11	2,04
24	2,64	2,54	2,44	2,33	2,27	2,21	2,15	2,08	2,01
25	2,61	2,51	2,41	2,30	2,24	2,18	2,12	2,05	1,98
26	2,59	2,49	2,39	2,28	2,22	2,16	2,09	2,03	1,95
27	2,57	2,47	2,36	2,25	2,19	2,13	2,07	2,00	1,93
28	2,55	2,45	2,34	2,23	2,17	2,11	2,05	1,98	1,91
29	2,53	2,43	2,32	2,21	2,15	2,09	2,03	1,96	1,89
30	2,51	2,41	2,31	2,20	2,14	2,07	2,01	1,94	1,87
40	2,39	2,29	2,18	2,07	2,01	1,94	1,88	1,80	1,72
60	2,27	2,17	2,06	1,94	1,88	1,82	1,74	1,67	1,58
120	2,16	2,05	1,94	1,82	1,76	1,69	1,61	1,53	1,43
∞	2,05	1,94	1,83	1,71	1,64	1,57	1,48	1,39	1,27

Таблица П6

Квантили распределения Фишера $F_p(k_1,k_2)$, $p = 0,99$

Число степеней свободы k_2	Число степеней свободы k_1								
	1	2	3	4	5	6	7	8	9
1	4052	4999,5	5403	5,625	5,764	5,859	5928	5982	6022
2	98,50	99,00	99,17	99,25	99,30	99,33	99,36	99,37	99,39
3	34,12	30,82	29,46	28,71	28,24	27,91	27,67	27,49	27,35
4	21,20	18,00	16,99	15,98	15,52	15,21	14,98	14,80	14,66
5	16,26	13,27	12,06	11,39	10,97	10,67	10,46	10,29	10,16
6	13,75	10,92	9,78	9,15	8,75	8,47	8,26	8,10	7,98
7	12,25	9,55	8,45	7,85	7,46	7,19	6,99	6,84	6,72
8	11,26	8,65	7,59	7,01	6,63	6,37	6,18	6,03	5,91
9	10,56	8,02	6,99	6,42	6,06	5,80	5,61	5,47	5,35
10	10,04	7,56	6,55	5,99	5,64	5,39	5,20	5,06	4,94
11	9,65	7,21	6,22	5,67	5,32	5,07	4,89	4,74	4,63
12	9,33	6,93	5,95	5,41	5,06	4,82	4,64	4,50	4,39
13	9,07	6,70	5,74	5,21	4,86	4,62	4,44	4,30	4,19
14	8,86	6,51	5,56	5,04	4,69	4,46	4,28	4,14	4,03
15	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,89
16	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89	3,78
17	8,40	6,11	5,18	4,67	4,34	4,10	3,93	3,79	3,68
18	8,29	6,01	5,09	4,58	4,25	4,01	3,84	3,71	3,60
19	8,18	5,93	5,01	4,50	4,17	3,94	3,77	3,63	3,52
20	8,10	5,85	4,94	4,43	4,10	3,87	3,70	3,56	3,46
21	8,02	5,78	4,87	4,37	4,04	3,81	3,64	3,51	3,40
22	7,95	5,72	4,82	4,31	3,99	3,76	3,59	3,45	3,35
23	7,88	5,68	4,76	4,26	3,94	3,71	3,54	3,41	3,30
24	7,82	5,61	4,72	4,22	3,90	3,67	3,50	3,36	3,26
25	7,77	5,57	4,68	4,18	3,85	3,63	3,46	3,32	3,22
26	7,72	5,53	4,64	4,14	3,82	3,59	3,42	3,29	3,18
27	7,68	5,49	4,60	4,11	3,78	3,56	3,39	3,26	3,15
28	7,64	5,45	4,57	4,07	3,75	3,53	3,36	3,23	3,12
29	7,60	5,42	4,54	4,04	3,73	3,50	3,33	3,20	3,09
30	7,56	5,39	4,51	4,02	3,70	3,47	3,30	3,17	3,07
40	7,31	5,18	4,31	3,83	3,51	3,29	3,12	2,99	2,89
60	7,08	4,98	4,13	3,65	3,34	3,12	2,95	2,82	2,72
120	6,85	4,79	3,95	3,48	3,17	2,96	2,79	2,66	2,56
∞	6,63	4,61	3,78	3,32	3,02	2,80	2,64	2,51	2,41

Число степеней свободы k_2	Число степеней свободы k_1								
	10	12	15	20	24	30	40	60	120
1	6056	6106	6157	6209	6235	6261	6287	6313	6339
2	99,40	99,42	99,43	99,45	99,46	99,47	99,47	99,48	99,49
3	27,23	27,05	26,87	26,69	26,60	26,50	26,41	26,32	26,22
4	14,55	14,37	14,20	14,02	13,93	13,84	13,75	13,65	13,56
5	10,05	9,89	9,72	9,55	9,47	9,38	9,29	9,20	9,11
6	7,87	7,72	7,56	7,40	7,31	7,23	7,14	7,06	6,97
7	6,62	6,47	6,31	6,16	6,07	5,99	5,91	5,82	5,74
8	5,81	5,67	5,52	5,36	5,28	5,20	5,12	5,03	4,95
9	5,26	5,11	4,96	4,81	4,73	4,65	4,57	4,48	4,40
10	4,85	4,71	4,56	4,41	4,33	4,25	4,17	4,08	4,00
11	4,54	4,40	4,25	4,10	4,02	3,94	3,86	3,78	3,69
12	4,30	4,16	4,01	3,86	3,78	3,70	3,62	3,54	3,45
13	4,10	3,96	3,82	3,66	3,59	3,51	3,43	3,34	3,25
14	3,94	3,80	3,66	3,51	3,43	3,35	3,27	3,18	3,09
15	3,80	3,67	3,52	3,37	3,29	3,21	3,13	3,05	2,96
16	3,69	3,55	3,41	3,26	3,18	3,10	3,02	2,93	2,84
17	3,59	3,46	3,31	3,16	3,08	3,00	2,92	2,83	2,75
18	3,51	3,37	3,23	3,08	3,00	2,92	2,84	2,75	2,66
19	3,43	3,30	3,15	3,00	2,92	2,84	2,76	2,67	2,58
20	3,37	3,23	3,09	2,94	2,86	2,78	2,69	2,61	2,52
21	3,31	3,17	3,03	2,88	2,80	2,72	2,64	2,55	2,46
22	3,26	3,12	2,98	2,83	2,75	2,67	2,58	2,50	2,40
23	3,21	3,07	2,93	2,78	2,70	2,62	2,54	2,45	2,35
24	3,17	3,03	2,89	2,74	2,66	2,58	2,46	2,40	2,31
25	3,13	2,99	2,85	2,70	2,62	2,54	2,45	2,36	2,27
26	3,09	2,96	2,81	2,66	2,58	2,50	2,42	2,33	2,23
27	3,06	2,93	2,78	2,63	2,55	2,47	2,38	2,29	2,20
28	3,03	2,90	2,75	2,60	2,52	2,44	2,35	2,26	2,17
29	3,00	2,87	2,73	2,57	2,49	2,41	2,33	2,23	2,14
30	2,98	2,84	2,70	2,55	2,47	2,39	2,30	2,21	2,11
40	2,80	2,66	2,52	2,37	2,29	2,20	2,11	2,02	1,92
60	2,63	2,50	2,35	2,20	2,12	2,03	1,94	1,84	1,73
120	2,47	2,34	2,19	2,03	1,95	1,86	1,76	1,66	1,53
∞	2,32	2,18	2,04	1,88	1,79	1,70	1,59	1,47	1,32

Таблица П7

Квантили распределения Фишера $F_p(k_1, k_2)$, $p = 0{,}995$

Число степеней свободы k_2	Число степеней свободы k_1								
	1	2	3	4	5	6	7	8	9
1	16211	20000	21615	22500	23056	23437	23715	23925	24091
2	198,5	199,0	199,2	199,2	199,3	199,3	199,4	199,4	199,4
3	55,55	49,80	47,47	46,19	45,39	44,84	44,43	44,13	43,88
4	31,33	26,28	24,26	23,15	22,46	21,97	21,62	21,35	21,14
5	22,78	18,31	16,53	15,56	14,94	14,51	14,20	13,96	13,77
6	18,63	14,54	12,92	12,03	11,46	11,07	10,79	10,57	10,39
7	16,24	12,40	10,88	10,05	9,52	9,16	8,89	8,68	8,51
8	14,69	11,04	9,60	8,81	8,30	7,95	7,69	7,50	7,34
9	13,61	10,11	8,72	7,96	7,47	7,13	6,88	6,69	6,54
10	12,83	9,43	8,08	7,34	6,87	6,54	6,30	6,12	5,97
11	12,23	8,91	7,60	6,88	6,42	6,10	5,86	5,68	5,54
12	11,75	8,51	7,23	6,52	6,07	5,76	5,52	5,35	5,20
13	11,37	8,19	6,93	6,23	5,79	5,48	5,25	5,08	4,94
14	11,06	7,92	6,68	6,00	5,56	5,26	5,03	4,86	4,72
15	10,80	7,70	6,48	5,80	5,37	5,07	4,85	4,67	4,54
16	10,58	7,51	6,30	5,64	5,21	4,91	4,69	4,52	4,38
17	10,38	7,35	6,16	5,50	5,07	4,70	4,56	4,39	4,25
18	10,22	7,21	6,03	5,37	4,96	4,66	4,44	4,28	4,14
19	10,07	7,09	5,92	5,27	4,85	4,56	4,34	4,18	4,04
20	9,94	6,99	5,82	5,17	4,76	4,47	4,26	4,09	3,96
21	9,83	6,89	5,73	5,09	4,68	4,39	4,18	4,01	3,88
22	9,73	6,81	5,65	5,02	4,61	4,32	4,11	3,94	3,81
23	9,63	6,73	5,58	4,95	4,54	4,26	4,05	3,88	3,75
24	9,55	6,65	5,52	4,89	4,49	4,20	3,99	3,83	3,69
25	9,48	6,60	5,46	4,84	4,43	4,15	3,94	3,78	3,64
26	9,41	6,54	5,41	4,79	4,38	4,10	3,89	3,73	3,60
27	9,34	6,49	5,36	4,74	4,34	4,06	3,85	3,69	3,56
28	9,28	6,44	5,32	4,70	4,30	4,02	3,81	3,65	3,52
29	9,23	6,40	5,28	4,66	4,26	3,98	3,77	3,61	3,48
30	9,18	6,35	5,24	4,62	4,23	3,95	3,74	3,58	3,45
40	8,83	6,07	4,98	4,37	3,99	3,71	3,51	3,35	3,22
60	8,49	5,79	4,73	4,14	3,76	3,49	3,29	3,13	3,01
120	8,18	5,54	4,50	3,92	3,55	3,28	3,09	2,93	2,81
∞	7,88	5,30	4,28	3,72	3,35	3,09	2,90	2,74	2,62

Окончание таблица П7

Число степеней свободы k_2	Число степеней свободы k_1								
	10	12	15	20	24	30	40	60	120
1	24224	24426	24630	24836	24940	25044	25148	25253	25359
2	199,4	199,4	199,4	199,4	199,5	199,5	199,5	199,5	199,5
3	43,69	43,39	43,08	42,78	42,62	42,47	42,31	42,15	41,99
4	20,97	20,70	20,44	20,17	20,03	19,89	19,75	19,61	19,47
5	13,62	13,38	13,15	12,90	12,78	12,66	12,53	12,40	12,27
6	10,25	10,03	9,81	9,59	9,47	9,36	9,24	9,12	9,00
7	8,38	8,18	7,97	7,75	7,65	7,53	7,42	7,31	7,19
8	7,21	7,01	6,81	6,61	6,50	6,40	6,29	6,13	6,06
9	6,42	6,23	6,03	5,83	5,73	5,62	5,52	5,41	5,30
10	5,85	5,66	5,47	5,27	5,17	5,07	4,97	4,86	4,73
11	5,42	5,24	5,05	4,86	4,76	4,65	4,55	4,44	4,34
12	5,09	4,91	4,72	4,53	4,43	4,33	4,23	4,12	4,01
13	4,82	4,64	4,46	4,27	4,17	4,07	3,97	3,87	3,76
14	4,60	4,43	4,25	4,06	3,96	3,86	3,76	3,66	3,55
15	4,42	4,25	4,07	3,88	3,79	3,69	3,58	3,48	3,37
16	4,27	4,10	3,92	3,73	3,64	3,54	3,44	3,33	3,22
17	4,14	3,97	3,79	3,61	3,51	3,41	3,31	3,21	3,10
18	4,03	3,86	3,68	3,50	3,40	3,30	3,20	3,10	2,99
19	3,93	3,76	3,59	3,40	3,31	3,21	3,11	3,00	2,89
20	3,85	3,68	3,50	3,32	3,22	3,12	3,02	2,92	2,81
21	3,77	3,60	3,43	3,24	3,15	3,05	2,95	2,84	2,73
22	3,70	3,54	3,36	3,18	3,08	2,98	2,88	2,77	2,66
23	3,64	3,47	3,30	3,12	3,02	2,92	2,82	2,71	2,60
24	3,59	3,42	3,25	3,06	2,97	2,87	2,77	2,66	2,55
25	3,54	3,37	3,20	3,01	2,92	2,82	2,72	2,61	2,50
26	3,40	3,33	3,15	2,97	2,87	2,77	2,67	2,56	2,45
27	3,45	3,28	3,11	2,93	2,83	2,73	2,63	2,52	2,41
28	3,41	3,25	3,07	2,89	2,79	2,69	2,59	2,48	2,37
29	3,38	3,21	3,04	2,86	2,76	2,66	2,56	2,45	2,33
30	3,34	3,18	3,01	2,82	2,73	2,63	2,52	2,42	2,30
40	3,12	2,95	2,78	2,60	2,50	2,40	2,30	2,18	2,06
60	2,90	2,74	2,57	2,39	2,29	2,19	2,08	1,96	1,83
120	2,71	2,54	2,37	2,19	2,09	1,98	1,87	1,75	1,61
∞	2,52	2,36	2,19	2,00	1,90	1,79	1,67	1,53	1,36

Таблица П8

Квантили критерия Кохрена *G*
при числе степеней свободы *k* и количестве серий измерения *n* (*P* = 0,95)

n	*k*						
	2	4	6	8	10	16	36
2	0,975	0,906	0,853	0,812	0,788	0,734	0,660
3	0,871	0,746	0,677	0,633	0,602	0,547	0,475
4	0,768	0,629	0,560	0,517	0,488	0,437	0,372
5	0,684	0,544	0,478	0,439	0,412	0,364	0,307
6	0,616	0,480	0,418	0,382	0,357	0,313	0,261
7	0,561	431	0,373	0,338	0,315	0,276	0,228
8	0,516	391	0,336	0,304	0,283	0,246	0,202
9	0,477	358	0,329	0,279	0,257	0,223	0,182
10	0,445	0,331	0,282	0,254	0,235	0,203	0,165
12	0,392	0,288	0,244	0,219	0,202	0,174	0,140
15	0,335	0,242	0,203	0,181	0,167	0,143	0,114
20	0,270	0,192	0,160	0,142	0,130	0,111	0,088
24	0,235	0,166	0,137	0,122	0,111	0,094	0,074
30	0,198	0,138	0,114	0,100	0,092	0,077	0,060
40	0,158	0,108	0,089	0,078	0,071	0,059	0,046
60	0,113	0,076	0,062	0,055	0,050	0,041	0,032

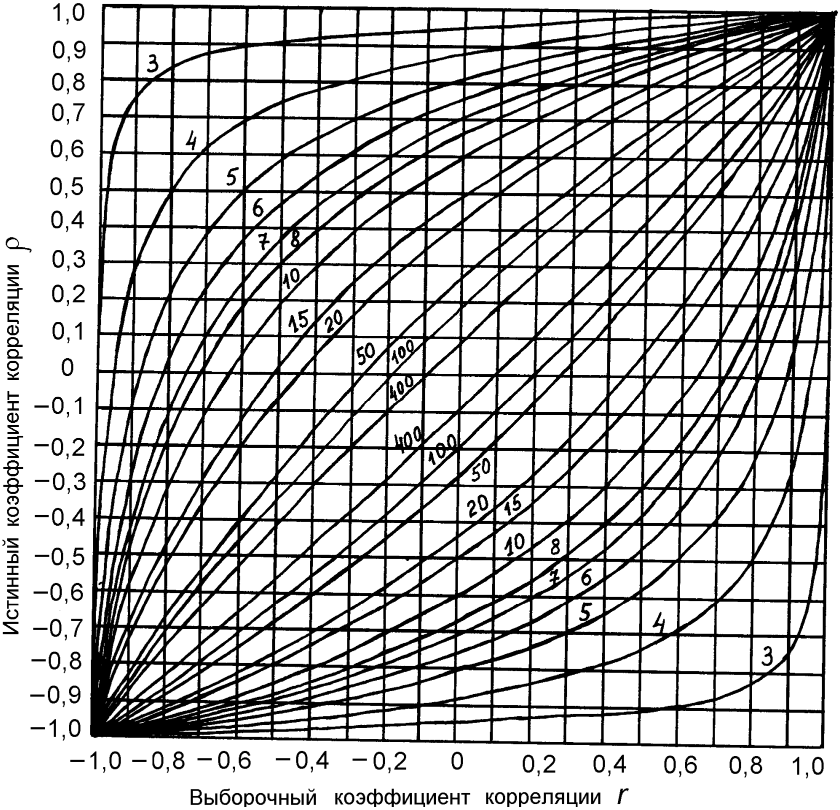

Рис. П1. Номограммы для нахождения доверительного интервала коэффициента корреляции при доверительной вероятности $P = 0{,}95$

Учебное издание

Карасев Владимир Анатольевич
Лёвшина Галина Дмитриевна

ТЕОРИЯ ВЕРОЯТНОСТЕЙ
И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Математическая статистика

Практикум

Редактор *И.Е. Оратовская*

Компьютерная верстка *А.В. Калинкина*

Подписано в печать 08.06.16	Рег. № 699	Уч.-изд. л. 7,5
Формат $60 \times 90\ ^1/_{16}$	Электронная версия	Заказ 5128

Национальный исследовательский
технологический университет «МИСиС»,
119049, Москва, Ленинский пр-т, 4

Издательский Дом МИСиС,
119049, Москва, Ленинский пр-т, 4
Тел. (495) 638-45-22

Отпечатано в типографии Издательского Дома МИСиС
119049, Москва, Ленинский пр-т, 4
тел. (499) 236-76-17, тел./факс (499) 236-76-35

ISBN 978-5-906846-01-3

9 785906 846013