Министерство образования и науки Российской Федерации
Волгоградский государственный архитектурно-строительный университет

Т. В. Ерещенко, Н. А. Михайлова

ПЛАНИРОВАНИЕ ЭКСПЕРИМЕНТА

Учебно-практическое пособие

Волгоград. ВолгГАСУ. 2014

УДК 519.86(075.8)
ББК 22.183.5я73
Е707

Р е ц е н з е н т ы:

кандидат технических наук *М. В. Филиппов*, доцент,
заведующий кафедрой информационных систем и технологий
НОУ ВПО «Волгоградский институт бизнеса»;
кандидат технических наук *И. В. Иванов*, доцент кафедры
прикладной математики и вычислительной техники
Волгоградского государственного архитектурно-строительного университета

Утверждено редакционно-издательским советом университета
в качестве учебно-практического пособия

Ерещенко, Т. В.

Е707 Планирование эксперимента [Электронный ресурс] : учебно-практическое пособие / Т. В. Ерещенко, Н. А. Михайлова ; М-во образования и науки Рос. Федерации, Волгогр. гос. архит.-строит. ун-т. — Электронные текстовые и графические данные (1,1 Мбайт). — Волгоград : ВолгГАСУ, 2014. — Учебное электронное издание сетевого распространения. — Систем. требования: PC 486 DX-33; Microsoft Windows XP; Internet Explorer 6.0; Adobe Reader 6.0. — Официальный сайт Волгоградского государственного архитектурно-строительного университета. Режим доступа: http://www.vgasu.ru/publishing/on-line/ — Загл. с титул. экрана.

ISBN 978-5-98276-728-8

Изложены основные идеи и методы статистического анализа экспериментальных данных, элементы регрессионного анализа, методы планирования однофакторных и многофакторных экспериментов. Содержит примеры и задачи для самостоятельного решения.

Пособие соответствует действующей программе дисциплины «Математическое моделирование» для студентов очной формы обучения направления «Строительство» по магистерским образовательным программам «Водоотведение и очистка сточных вод», «Проектирование, строительство и эксплуатация автомобильных дорог», «Теория и проектирование зданий и сооружений», «Теория и практика организационно-технологических и экономических решений в строительстве», «Судебная, строительно-техническая и стоимостная экспертиза объектов недвижимости».

Для удобства работы с изданием рекомендуется пользоваться функцией Bookmarks (Закладки) в боковом меню программы Adobe Reader.

УДК 519.86(075.8)
ББК 22.183.5я73

ISBN 978-5-98276-728-8

ОГЛАВЛЕНИЕ

ПРЕДИСЛОВИЕ

Высшей формой эмпирических методов познания окружающей действительности являются экспериментальные исследования. Процесс этот многоэтапный и включает различные формы: наблюдение, сравнение, контроль.

На начальной стадии эксперимента, наблюдая за поведением объекта или протеканием явления, исследователь делает предположения о наличии некоторых взаимосвязей и закономерностей их функционирования. В заключительной стадии формируется цель исследования, определяются величины и факторы, влияющие на свойства объекта и вид их взаимосвязи, т. е. выдвигается гипотеза о виде модели исследуемого объекта. В соответствии с ним строится план эксперимента. От правильного выбора плана проведения целенаправленного эксперимента (в особенности, если объект сложный) в первую очередь зависит успех дальнейших исследований. Правильно выбранный план позволяет не только уменьшить объем исследований, но и минимизировать влияние на их результат неучтенных или неконтролируемых факторов.

Дисциплина «Математическое моделирование» относится к базовой части общенаучного цикла.

Одним из разделов дисциплины является планирование эксперимента. Цель данного учебного пособия — дать фундаментальные знания по анализу и обработке результатов экспериментов, проводимых студентами при подготовке курсовых и дипломных проектов, научно-исследовательских разработках, выполнении выпускной квалификационной работы.

Теория планирования эксперимента основана на использовании методов математической статистики и формулирует приемы и способы оптимальной организации экспериментальных исследований, чтобы получить наиболее эффективный результат при наименьших затратах времени и средств.

Задачи для самостоятельного решения заимствованы из журналов контроля прочности бетонных образцов исследовательских лабораторий Волгоградской области, а также из [1, 2, 3].

ГЛАВА 1. СТАТИСТИЧЕСКИЙ АНАЛИЗ ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ

1.1. Основные понятия

Совокупность, из которой производится отбор, называется **генеральной,** а все ее обобщающие показатели — генеральными. Совокупность отобранных единиц именуют выборкой, а все ее обобщающие показатели — выборочными.

Выборка — необходимый для исследования минимум результатов (случаев, событий, образцов), отобранных с помощью определенной процедуры из генеральной совокупности. Элементы выборки называются **выборочными значениями** или **вариантами.** Количество проведенных наблюдений называется **объемом** выборки.

Поскольку некоторые значения выборки могут совпадать, то часто выборку представляют в виде **статистического ряда**, т. е. таблицы значений $(x_i^*; n_i)$, где $x_1^*, x_2^*, ..., x_k^*$ — различные элементы выборки, а n_i — частота появления значения x_i^* в выборке (сумма всех частот равна объему выборки: $\sum_{i=1}^{k} n_i = n)$:

x_i^*	x_1^*	x_2^*	…	x_k^*
n_i	n_1	n_2	…	n_k

Величины $\omega_i = \dfrac{n_i}{n}$ называются **относительными частотами**.

В случае выборки большого объема данные группируют и представляют в виде интервального статистического ряда.

Для графического изображения дискретных статистических рядов используется полигон. **Полигоном частот (**полигоном относительных частот) называется ломаная линия с вершинами в точках $(x_i^*; n_i)$ (соответственно в точках $\left(x_i^*; \dfrac{n_i}{n}\right)$).

Для графического изображения интервальных статистических рядов используется гистограмма. **Гистограмма частот (гистограмма относительных частот)** — ступенчатая фигура, состоящая из прямоугольников, построенных на интервалах группировки и имеющих высоты, равные $\frac{n_i}{h}$ (соответственно $\frac{n_i}{nh}$), где h — длина интервалов группировки. Гистограмма относительных частот обладает тем свойством, что ее площадь равна единице $\sum_{i=1}^{k}\frac{n_i}{nh}h=\sum_{i=1}^{k}\frac{n_i}{n}=1.$

При больших значениях n гистограмма относительных частот является хорошим приближением для графика плотности распределения наблюдаемой случайной величины. Поэтому по виду гистограммы можно выдвинуть предположение (гипотезу) о распределении изучаемой величины.

Эмпирическая функция распределения — это функция, которая определяется для каждого действительного x как относительная частота наблюдения значений, меньших x, т. е.

$$F_n^*(x)=\sum_{x_i^*<x}\frac{n_i}{n}.$$

Эмпирическая функция распределения обладает следующими свойствами:

1) принимает значения от 0 до 1: $0\le F_n^*(x)\le 1$;

2) является неубывающей;

3) если $x_{\min}$ и $x_{\max}$ — наименьшее и наибольшее выборочные значения, то $F_n^*(x)=0$ при $x\le x_{\min}$ и $F_n^*(x)=1$ при $x>x_{\max}$.

Замечание. При построении эмпирической функции распределения по интервальному статистическому ряду за случайную величину принимают середину интервала.

Эмпирическая функция распределения является приближением для теоретической функции распределения исследуемой случайной величины.

Примеры решения задач

Задача 1. Построить полигон частот и эмпирическую функцию распределения по выборке: 5; 2; 2; 1; 6; 3; 1; 2; 3; 5.

Решение. Подсчитав частоты наблюдения различных выборочных значений, составим статистический ряд:

x_i^*	1	2	3	5	6
n_i	2	3	2	2	1

Объем выборки равен сумме частот наблюдаемых значений n = 2 + 3 + 2 + + 2 + 1 = 10. Учитывая это, запишем эмпирическую функцию распределения.

Наименьшее наблюдаемое выборочное значение — 1, поэтому эмпирическая функция распределения равна 0 при $x\le 1$. Далее ее значение изменяется каждый раз при переходе x через x_i^*, увеличиваясь на величину относительной частоты $\omega_i = \dfrac{n_i}{n}$. Наибольшее выборочное значение — 6, поэтому эмпирическая функция распределения принимает значение 1 при $x > 6$.

Запишем эмпирическую функцию распределения:

$$F_n^*(x) = \begin{cases} 0,0 & \text{при } x \le 1, \\ 0,2 & \text{при } 1 < x \le 2, \\ 0,5 & \text{при } 2 < x \le 3, \\ 0,7 & \text{при } 3 < x \le 5, \\ 0,9 & \text{при } 5 < x \le 6, \\ 1,0 & \text{при } x > 6. \end{cases}$$

Построим полигон частот (рис. 1) и график эмпирической функции распределения (рис. 2).

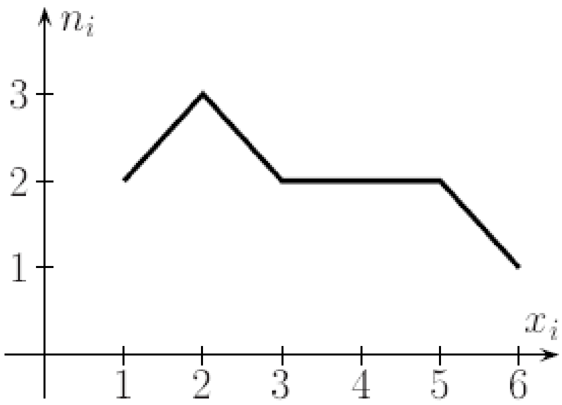


Рис. 1

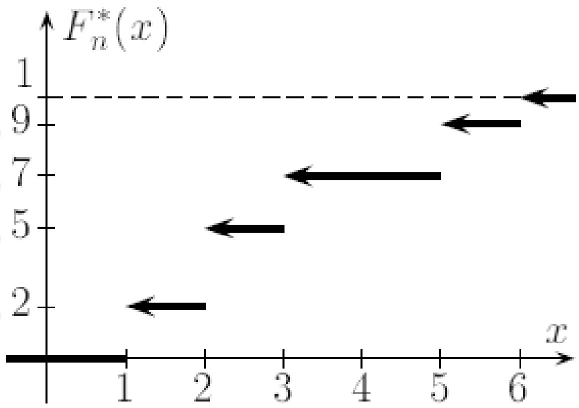


Рис. 2

Задача 2. Построить гистограмму относительных частот и эмпирическую функцию распределения по данному интервальному статистическому ряду:

$[x_{i-1}; x_i)$	[5; 10)	[10; 15)	[15; 20)	[20; 25)	[25; 30)	[30; 35)	[35; 40)
n_i	4	6	16	36	24	10	4

Решение. Объем выборки равен сумме наблюдаемых частот: $n = 4 + 6 + 16 + 36 + 24 + 10 + 4 = 100$. Длина каждого интервала $h = 5$. Для построения эмпирической функции распределения найдем середины интервалов x_i и относительные частоты $\dfrac{n_i}{n}$; для построения гистограммы относительных частот рассчитаем для каждого интервала значение $\dfrac{n_i}{nh}$. Заполним таблицу (табл. 1).

Таблица 1

$[x_{i-1}; x_i)$	[5; 10)	[10; 15)	[15; 20)	[20; 25)	[25; 30)	[30; 35)	[35; 40)
x_i^*	7,5	12,5	17,5	22,5	27,5	32,5	37,5
n_i	4	6	16	36	24	10	4
$\frac{n_i}{n}$	0,04	0,06	0,16	0,36	0,24	0,1	0,04
$\frac{n_i}{nh}$	0,008	0,012	0,032	0,072	0,048	0,02	0,008

Запишем эмпирическую функцию распределения:

$$F_n^*(x) = \begin{cases} 0,00 \text{ при } x \le 7,5, \\ 0,04 \text{ при } 7,5 < x \le 12,5, \\ 0,1 \text{ при } 12,5 < x \le 17,5, \\ 0,26 \text{ при } 17,5 < x \le 22,5, \\ 0,62 \text{ при } 22,5 < x \le 27,5, \\ 0,86 \text{ при } 27,5 < x \le 32,5, \\ 0,96 \text{ при } 32,5 < x \le 37,5, \\ 1,00 \text{ при } x > 37,5. \end{cases}$$

Построим график эмпирической функции распределения (рис. 3).

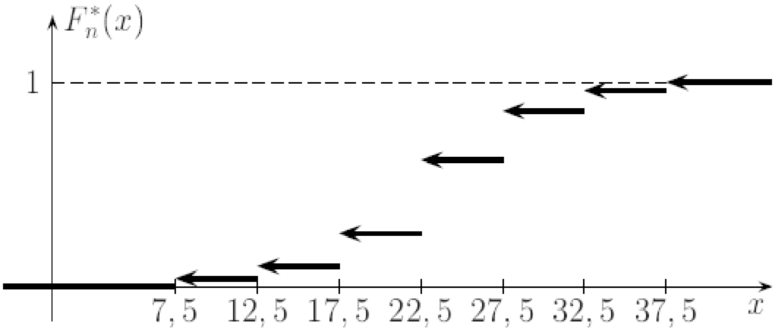


Рис. 3

Построим гистограмму относительных частот (рис. 4).

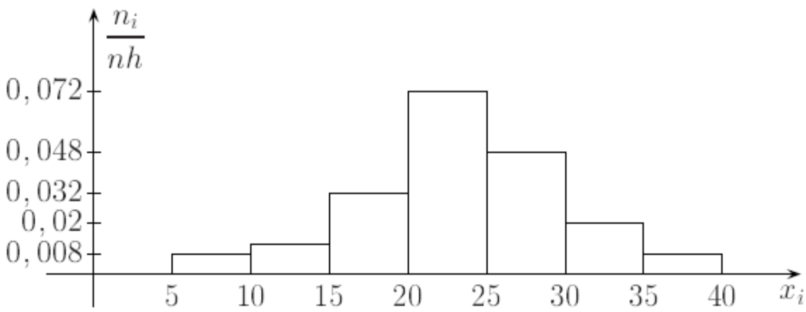


Рис. 4

По виду гистограммы можно предложить, что выборка взята из нормального распределения.

Задачи для самостоятельного решения

Задача 3. При проверке прочности бетонных образцов получены следующие данные по маркам:

M50	M75	M50	M75	M50	M50	M75	M100	M150	M75
2	1	2	3	3	2	1	2	2	2

Составить статистический ряд, построить полигон частот, найти эмпирическую функцию распределения и построить ее график.

Задача 4. Из большой партии по схеме случайной повторной выборки было проверено 150 проб грунта с целью определения процента влажности. Получены следующие результаты:

$[x_{i-1}; x_i)$	[11; 13)	[13; 15)	[15; 17)	[17; 19)	[19; 21)
n_i	8	42	51	37	12

Построить гистограмму относительных частот, найти эмпирическую функцию распределения и построить ее график. Можно ли по виду гистограммы предложить, что наблюдаемая случайная величина имеет нормальное распределение?

1.2. Точечное оценивание параметров распределения

Выборка представляет собой ряд наблюдений за одной и той же случайной величиной. Для содержательного статистического анализа экспериментальных данных необходимо знать распределение этой величины.

Во многих случаях можно считать, что наблюдаемая величина имеет нормальное распределение. Например, при измерениях одного и того же показателя на нескольких однотипных объектах колебания результатов будут вызваны незначительными случайными погрешностями в технологии изготовления или измерения. Если случайные колебания значений некоторой величины вызваны большим числом случайных причин, более или менее равноправных, то на основании центральной предельной теоремы теории вероятностей можно считать, что эта величина имеет нормальное распределение.

Нормальным называют распределение вероятностей непрерывной случайной величины *X*, плотность которого имеет вид

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}},$$

где μ — математическое ожидание, σ — среднее квадратическое отклонение *X*.

Нормальное распределение полностью определяется двумя параметрами: математическим ожиданием μ и дисперсией σ^2. Математическое ожидание и дисперсия являются основными числовыми характеристиками любой случайной величины. Поэтому при статистическом анализе выборки в первую очередь стремятся оценить математическое ожидание и дисперсию.

Пусть имеется выборка объема n: $x_1, x_2, \ldots, x_n$. По результатам этого ограниченного числа наблюдений мы не можем вычислить числовые характеристики и случайные величины, а только оценить их.

Любая функция $\hat{\theta}_n = \hat{\theta}_n(x_1, x_2, \ldots, x_n)$, зависящая от выборочных значений, называется **статистикой** или **выборочной функцией**.

Точечной оценкой параметра θ называется любая статистика $\hat{\theta}_n$, предназначенная для оценивания этого параметра и определяемая одним числом. Подчеркнем, что точечная оценка практически никогда не совпадает с истинным значением параметра, она может только оценивать его с большей или меньшей точностью.

Оценкой для математического ожидания является **выборочное среднее**, которое вычисляется по формуле

$$\overline{x} = \frac{1}{n}(x_1 + x_2 + \ldots + x_n)$$

или кратко:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i. \qquad (1)$$

Оценка для **дисперсии** рассчитывается по формуле

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2. \qquad (2)$$

Эту оценку называют **несмещенной оценкой дисперсии**. Число $f = n - 1$. Оно называется **числом степеней свободы** этой оценки дисперсии и равно количеству независимых наблюдений (объему выборки), по которым вычисляется данная оценка дисперсии, минус число параметров, которые оцениваются по этой выборке, кроме дисперсии. В данном случае одна степень свободы расходуется на вычисление оценки $\overline{x}$.

Иногда вместо (2) удобнее использовать формулу

$$s^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n}(x_i^2 - n(\overline{x})^2\right).$$

Замечание. В случае сгруппированного статистического ряда выборочные среднее и дисперсия рассчитываются по формулам

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{k} x_i^* n_i; \qquad (3)$$

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{k}(x_i^* - \overline{x})^2 n_i, \quad (4)$$

где k — количество различных вариант x_i^* выборки; n_i — частота наблюдения значения x_i^*.

Задача статистического оценивания параметров заключается в том, чтобы из всего множества оценок выбрать в некотором смысле наилучшую. Это означает, что распределение случайной величины $\hat{\theta}_n(x_1, x_2, ..., x_n)$ должно концентрироваться около истинного значения параметра θ.

Несмещенной называют точечную оценку, математическое ожидание которой равно оцениваемому параметру при любом объеме выборки. Требование несмещенности гарантирует отсутствие систематических ошибок при оценивании. Оно особенно важно при малом числе наблюдений (в случае выборок объема не более 30).

Смещенной оценкой генеральной дисперсии служит выборочная дисперсия

$$D_{\text{в}} = \left(\sum_{i=1}^{\kappa} n_i \left(x_i - \overline{x}\right)^2\right) / n.$$

Эта оценка является смещенной

Несмещенной оценкой генеральной дисперсии служит исправленная выборочная дисперсия

$$S^2 = \frac{n}{n-1} D_{\text{в}} = \frac{\sum_{i=1}^{k} n_i \left(x_i - \overline{x}\right)^2}{n-1}.$$

Оценка $\hat{\theta}_n$ называется **состоятельной**, если она сходится по вероятности к оцениваемому параметру θ, т. е. если для любого $\varepsilon > 0$

$$P(|\hat{\theta}_n - \theta| > \varepsilon) \to 0 \text{ при } n \to \infty.$$

Это означает, что при большом объеме выборки практически достоверно, что $\hat{\theta}_n \approx \theta$. Чем больше объем выборки, тем более точные оценки можно получить.

Выборочные среднее и дисперсия являются несмещенными состоятельными для математического ожидания и дисперсии соответственно.

Примеры решения задач

Задача 5. При проверке прочности бетонных кубиков-образцов были получены следующие результаты, МПа:

200; 220; 250; 210; 230.

Найти несмещенные оценки для математического ожидания и дисперсии, указать число степеней свободы оценки дисперсии.

Решение. Вычислим выборочное среднее по формуле (1).

$$\bar{x} = \frac{1}{5}(200 + 220 + 250 + 210 + 230) = 222.$$

Вычислим выборочную дисперсию по формуле (2).

$$s^2 = \frac{1}{5-1}((200-222)^2 + (220-222)^2 + (250-222)^2 +$$
$$+ (210-222)^2 + (230-222)^2) = 370.$$

Число степеней свободы выборочной дисперсии $f = 5 - 1 = 4$.

Задача 6. По выборке задачи 2 найти несмещенные оценки для математического ожидания и дисперсии, указать число степеней свободы оценки дисперсии.

Решение. Поскольку данные представлены в виде интервального статистического ряда, для расчета выборочных среднего и дисперсии следует воспользоваться формулами (3) и (4), где k — число интервалов; x_i^* — середина i-го интервала; n_i — частота наблюдения значений из i-го интервала. Несмещенные оценки для математического ожидания и дисперсии вычисляются следующим образом:

$$\bar{x} = \frac{1}{100}(7{,}5\cdot 4 + 12{,}5\cdot 6 + 17{,}5\cdot 16 + 22{,}5\cdot 36 + 27{,}5\cdot 24 + 32{,}5\cdot 10 + 37{,}5\cdot 4) = 23{,}3;$$

$$s^2 = \frac{1}{100-1}(4(7{,}5-23{,}3)^2 + 4(12{,}5-23{,}3)^2 + 4(17{,}5-23{,}3)^2 +$$
$$+ 4(22{,}5-23{,}3)^2 + 4(27{,}5-23{,}3)^2 + 4(32{,}5-23{,}3)^2 + 4(37{,}5-23{,}3)^2) = 43{,}8.$$

Число степеней свободы несмещенной оценки дисперсии $f = 100 - 1 = 99$.

Задачи для самостоятельного решения

Задача 7. По выборке задачи 3 найти несмещенные оценки для математического ожидания и дисперсии, указать число степеней свободы оценки дисперсии.

Задача 8. По выборке задачи 4 найти несмещенные оценки для математического ожидания и дисперсии, указать число степеней свободы оценки дисперсии.

Задача 9. Ниже дана выборка для прочности на разрыв (табл. 2). Определить среднее значение и дисперсию выборки.

Таблица 2

R, Н/м^2	Частота
18461	2
18466	12
18471	15
18476	10

Ответы к задачам

7. $\overline{x} = 73{,}75$; $s^2 = 952{,}30$; $f = 19$. **8.** $\overline{x} = 16{,}04$; $s^2 = 4{,}27$; $f = 149$ **9.** $\overline{x} = 18470{,}23$; $s^2 = 19{,}13$.

1.3. Выборочные распределения

В дальнейшем, в частности, при нахождении интервальных оценок параметров, нам нужно знать, каким распределениям подчиняются статистики $\overline{x}$ и s^2.

Большинство статистических методов разработаны в предположении, что результаты наблюдений являются независимыми, нормально распределенными случайными величинами. Запись $\xi \sim N(\mu;\sigma)$ означает, что случайная величина ξ имеет нормальное распределение с математическим ожиданием $M\xi = \mu$ и дисперсией $D\xi = \sigma^2$. С нормальным распределением связаны наиболее часто использующиеся в статистике распределения: χ^2 — распределение «хи-квадрат», t — распределение Стьюдента и F — распределение Фишера.

Утверждение. В случае выборки объема n из нормального распределения с известным математическим ожиданием и дисперсией

$$\overline{x} \sim N\left(\mu;\frac{\sigma}{\sqrt{n}}\right);\ \frac{\overline{x}-\mu}{\sigma/\sqrt{n}} \sim N(0,1).$$

Распределением χ^2 с k степенями свободы называется распределение суммы квадратов k независимых случайных величин, распределенных по нормальному закону, т. е. $\eta = \xi_1^{\ 2} + \xi_1^{\ 2} + \ \dots\ + \xi_1^{\ 2} \sim \chi_k^{\ 2}$.

Утверждение. В случае выборки объема n из нормального распределения с неизвестным математическим ожиданием

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}.$$

Распределением Стьюдента с k степенями свободы называется распределение случайной величины $t = \dfrac{\xi}{\sqrt{\eta / k}} \sim t_k$, где случайные величины $\xi \sim N$ (0;1) и $\eta \sim \chi^2_k$ независимы.

Утверждение. В случае выборки объема n из нормального распределения и неизвестной дисперсией

$$\frac{\overline{x}-\mu}{\sqrt{s^2 / n}} \sim t_{n-1}.$$

***F*-распределением Фишера с числами степеней свободы k_1 и k_2** называется распределение случайной величины $F = \dfrac{\eta_1 / k_1}{n_2 / k_2} \sim F_{k_1 k_2}$, где случайные величины $\eta_1 \sim \chi^2_{k_2}$ и $\eta_2 \sim \chi^2_{k_2}$ независимы.

Существуют таблицы распределений χ^2, Стьюдента, Фишера, а также нормального распределения (прил. 1, 2, 3, 4, 5).

1.4. Интервальное оценивание параметров распределения

Точечные оценки не дают информации о степени близости оценки к оцениваемому параметру. Чтобы получить информацию о точности и надежности оценки, используют интервальные оценки.

Интервальной оценкой параметра θ называется интервал, границы которого $\hat{\theta}_1 = \hat{\theta}_1(x_1, x_2, \ ..., x_n)$ и $\hat{\theta}_2 = \hat{\theta}_2(x_1, x_2, \ ..., x_n)$ являются функциями выборочных значений и который с заданной вероятностью γ накрывает истинное значение оцениваемого параметра θ:

$$P(\hat{\theta}_1(x_1, x_2, \ ..., x_n) < \theta < \hat{\theta}_2(x_1, x_2, \ ..., x_n)) = \gamma.$$

Интервал $(\hat{\theta}_1; \hat{\theta}_2)$ называется **доверительным интервалом;** число γ — **доверительной вероятностью** или **надежностью** интервальной оценки; число $\alpha = 1 - \gamma$ — **уровнем значимости.**

Величина доверительного интервала существенно зависит от объема выборки (уменьшается с ростом *n*, т. е. чем больше объем выборки, тем более точную оценку можно получить) и от доверительной вероятности γ (величина доверительного интервала увеличивается с приближением γ к 1, т. е. чем более надежный вывод мы хотим получить, тем меньшую точность можем гарантировать).

Доверительную вероятность γ обычно выбирают равной 0,90; 0,95 или 0,99, чтобы получить интервал, который с большой вероятностью накроет истинное значение оцениваемого параметра.

Доверительный интервал для математического ожидания в случае выборки из нормального распределения с известной дисперсией σ^2 определяется соотношением

$$P\left(\overline{x} - u_\alpha \frac{\sigma}{\sqrt{n}} < \mu < \overline{x} + u_\alpha \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha, \tag{5}$$

где α — заданный уровень значимости; u_α — квантúль нормального распределения, задаваемый формулой $\Phi(u_\alpha) = \dfrac{1-\alpha}{2}$ и определяемый из таблицы функции Лапласа (прил. 1).

Доверительный интервал для математического ожидания в случае выборки из нормального распределения с **неизвестной дисперсией** определяется соотношением

$$P\left(\bar{x} - t_{\alpha;\, n-1}\frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{\alpha;\, n-1}\frac{s}{\sqrt{n}}\right) = 1 - \alpha, \qquad (6)$$

где α — заданный уровень значимости; $t_{\alpha,\, n-1}$ — квантиль распределения Стьюдента, удовлетворяющий соотношению $P\left(\left|t_{n-1}\right| \ge t_{\alpha;\, n-1}\right) = \alpha$ для случайной величины t_{n-1}, имеющей распределение Стьюдента с числом степеней свободы $\nu = n - 1$ (прил. 3).

Доверительный интервал для дисперсии в случае выборки из нормального распределения с **неизвестным математическим ожиданием** определяется соотношением

$$P\left(\frac{(n-1)s^2}{\chi^2_{\alpha/2;\, n-1}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{1-\alpha/2;\, n-1}}\right) = 1 - \alpha, \qquad (7)$$

где α — заданный уровень значимости; $\chi^2_{\alpha/2;\, n-1}$ и $\chi^2_{1-\alpha/2;\, n-1}$ — квантили распределения χ^2, задаваемые формулой $P\left(\chi^2_{n-1} \ge \chi^2_{\alpha;\, n-1}\right) = \alpha$ для случайной величины χ^2_{n-1}, имеющей распределение χ^2 с числом степеней свободы $\upsilon = n - 1$ (прил. 2).

Задача планирования эксперимента при построении доверительных интервалов заключается в том, чтобы определить объем выборки, необходимый для достижения заданной точности ε оценивания параметра при фиксированной доверительной вероятности γ. Так, в случае оценивания математического ожидания нормального распределения по неизвестной дисперсии для решения задачи достаточно

$$n_0 > \frac{s^2 t^2_{\alpha;\, n-1}}{\varepsilon^2} \qquad (8)$$

наблюдений, где n — объем выборки, по результатам которой делается вывод; s^2 — несмещенная оценка дисперсии, полученная по этой выборке.

Примеры решения задач

Задача 10. Оценить математическое ожидание нормального распределения с заданной надежностью γ, если:

1) среднеквадратическое отклонение $\sigma = 2$, по выборке объема 10 найдено выборочное среднее $\bar{x} = 5{,}4$ ($\gamma = 0{,}95$);

2) по выборке объема 9 найдены несмещенные оценки математического ожидания и дисперсии: $\bar{x} = 14{,}2$; $s^2 = 5{,}76$ ($\gamma = 0{,}99$).

Решение.

1. Поскольку дисперсия известна, построим доверительный интервал для математического ожидания нормального распределения с помощью формулы (5). Имеем:

$$n = 10;\ \overline{x} = 5,4;\ \sigma = 2;\ \alpha = 1 - \gamma = 0,05.$$

Квантиль $u_\alpha = u_{0,05}$ определим из соотношения $\Phi\left(u_{0,05}\right) = \dfrac{1 - 0,05}{2} = 0,475.$ Для этого по прил. 1 находим в столбце Φ(*x*) значение 0,475, тогда значение $u_{0,05}$ получаем в соответствующем столбце x: $u_{0,05} = 1,96.$

Запишем доверительный интервал по формуле (5):

$$P\left(5,4 - 1,96\frac{2}{\sqrt{10}} < \mu < 5,4 + 1,96\frac{2}{\sqrt{10}}\right) = 0,95;$$

$$P\left(4,16 < \mu < 6,64\right) = 0,95.$$

2. Поскольку дисперсия неизвестна и оценена по выборке, для построения доверительного интервала для математического ожидания нужно использовать формулу (6). Имеем:

$$n = 9;\quad \overline{x} = 14,2;\quad s = \sqrt{s^2} = \sqrt{5,76} = 2,4;\quad \alpha = 1 - \gamma = 0,01.$$

Квантиль $t_{\alpha;\, n-1} = t_{0,01;\ 8}$ определим по прил. 3. В соответствии с определением квантиля $t_{\alpha;\, n-1}$ находим в верхней части таблицы значение $\alpha = 0,01$, в столбце ν — значение ν = 8, на пересечении получаем $t_{0,01;\ 8} = 3,36.$

Запишем доверительный интервал по формуле (6):

$$P\left(14,2 - 3,36\frac{2,4}{\sqrt{9}} < \mu < 14,2 + 3,36\frac{2,4}{\sqrt{9}}\right) = 0,99;$$

$$P\left(11,512 < \mu < 16,888\right) = 0,99.$$

Задача 11. Для определения прочности бетона было испытано три бетонных кубика. Результаты испытаний — 19,8; 20,1; 20,4 МПа. Сколько надо провести таких испытаний, чтобы с надежностью 0,95 ошибка при определении средней прочности была в пределах 0,2 МПа, если считается, что ошибки прибора нормальны?

Решение. Задача заключается в определении объема выборки, необходимого для достижения заданной точности $\varepsilon = 0,2.$ Имеем

$$\gamma = 0,95;\ \alpha = 1 - 0,95 = 0,05;\ n = 3;\ t_{0,05;\ 2} = 4,3.$$

Рассчитаем оценки для математического ожидания и дисперсии:

$$\overline{x} = \frac{1}{3}\left(19,8 + 20,1 + 20,4\right) = 20,1;$$

$$s^2 = \frac{1}{3-1}\Big((19,8-20,1)^2 + (20,1-20,1)^2 + (20,4-20,1)^2\Big) = 0,09.$$

По формуле (8) определяем, что

$$n_0 > \frac{0,09 \cdot 4,3^2}{0,2^2} \approx 41,602.$$

Потребуется произвести 42 испытания.

Проверить, как изменится ответ, если принять надежность 0,90.

Задача 12. При взвешивании груза получены следующие результаты:

129; 125; 130; 122; 135; 125; 120; 130; 127.

Определить среднюю квадратическую ошибку взвешивания и построить для нее доверительный интервал с надежностью 0,90.

Решение. n = 9; γ = 0,9; α = 1 – 0,90 = 0,10. Рассчитаем оценки для математического ожидания и дисперсии:

$$\overline{x} = 127;$$

$$s^2 = 21; \ \ s = \sqrt{21} \approx 4,58.$$

Таким образом, средняя квадратическая ошибка взвешивания оценивается величиной $s \approx 4,58$. Чтобы получить для нее доверительный интервал, построим доверительный интервал для дисперсии по формуле (7). По прил. 2 определим квантили распределения χ^2:

$$\chi^2_{\alpha/2;\, n-1} = \chi_{0,05;\ 8} = 15,507; \ \ \chi_{1-\alpha/2;\, n-1} = \chi_{0,95;\ 8} = 2,7326.$$

Доверительный интервал для дисперсии определяется по формуле (7):

$$P\left(\frac{8 \cdot 21}{15,507} < \sigma^2 < \frac{8 \cdot 21}{2,7326}\right) = 0,90;$$

$$P\Big(10,83 < \sigma^2 < 61,48\Big) = 0,90 = 0,90.$$

Отсюда получаем доверительный интервал для среднего квадратического отклонения в виде

$$P\Big(\sqrt{10,83} < \sigma < \sqrt{61,48}\Big) = 0,90;$$

$$P\big(3,3 < \sigma < 7,8\big) = 0,9. = 0,90.$$

Задачи для самостоятельного решения

Задача 13. Найти доверительный интервал для математического ожидания нормального распределения с заданной надежностью γ, если:

1) среднее квадратическое отклонение известно: σ = 3, по результатам 25 независимых наблюдений найдено выборочное среднее $\overline{x}$ = 20,12 (γ = 0,99);

2) по выборке объема 12 найдены несмещенные оценки математического ожидания и дисперсии: $\overline{x} = 16,8$; $s^2 = 2,25 (\gamma = 0,95)$.

Задача 14. По данным 16 независимых равноточных измерений физической величины найдены несмещенные оценки математического ожидания и дисперсии: $\overline{x} = 23,161$; $s^2 = 0,16$. Требуется оценить истинное значение μ измеряемой величины и точность измерений σ с надежностью 0,95.

Задача 15. Даны результаты 5 независимых равноточных измерений толщины металлической пластинки: 2,015; 2,020; 2,025; 2,020; 2,015. Нужно:

1) оценить с помощью доверительного интервала истинную толщину пластинки при доверительной вероятности $\gamma = 0,95$;

2) найти минимальное число измерений, которое надо выполнить, чтобы с надежностью 0,95 можно было утверждать, что предельная погрешность точечной оценки истинной толщины металлической пластинки не превышает 0,003.

Ответы к задачам

13. 1) $18,57 < \mu < 21,67$; 2) $15,85 < \sigma < 17,75$. **14.** $22,948 < \mu < 23,374$; 0,224; 0,576. **15**. 1) 2,014; 2,024, 2) 16.

1.5. Проверка статистических гипотез

Основные задачи математической статистики разделяют на две категории, тесно связанные между собой:

1) оценивание параметров, т. е. получение по выборке оценок, наилучших в том или ином смысле;

2) проверка гипотез, когда по выборке требуется принять или отвергнуть некоторое предложение о распределении, из которого извлечена выборка.

Статистической гипотезой называется любое предположение о виде или параметрах неизвестного распределения, проще говоря, это предположение относительно свойств совокупности, из которой производится выборка. Правило, которое позволяет по выборке принять или отвергнуть проверяемую гипотезу, называется **статистическим критерием (статистикой).**

Замечание. Статистическими методами нельзя доказать правильность гипотезы. Если по результатам проверки статистическая гипотеза принимается, то говорят, что она согласуется с выборочными данными (не противоречит результатам наблюдений).

Проверяемую гипотезу обычно называют **нулевой** и обозначают H_0. Наряду с нулевой рассматривают **альтернативную (конкурирующую)** гипотезу. Нулевая и альтернативная гипотезы представляют собой две возможности выбора, осуществляемого в ходе проверки гипотез. Так как проверка гипотез основывается на полученных выборочных статистиках, вычисленных по n наблюдениям, то при принятии решений всегда возможны ошибки.

В задачах проверки гипотез возможны следующие четыре ситуации (табл. 3).

Таблица 3

Проверяемая гипотеза	Гипотеза принимается	Гипотеза отвергается
Объективно верна	Правильное решение	Ошибка 1-го рода
Объективно неверна	Ошибка 2-го рода	Правильное решение

Вероятность ошибки 1-го рода, т. е. непринятие верной, называется уровнем значимости статистического критерия и обозначается α. Вероятность ошибки 2-го рода, т. е. принятие неверной гипотезы, обозначается β. Пользуясь терминологией статистического контроля качества продукции, можно сказать, что α — это риск поставщика (выбраковка партии, удовлетворяющей стандарту), а β — риск потребителя (принятие партии, не удовлетворяющей стандарту).

Ясно, что при построении статистических критериев желательно, чтобы вероятности ошибок обоих родов были как можно меньше. Однако это требование противоречивое, т. к. невозможно одновременно уменьшить обе ошибки. Реально поступают следующим образом: задают уровень значимости α (как правило, равный 0,05; 0,01 или 0,10, а затем выбирают статистический критерий так, чтобы ошибка 2-го рода была наименьшей.

Этапы в задаче проверки гипотез:

1) определение гипотезы H_0 и ее альтернативы $\overline{H}$;

2) установление уровня значимости (ошибки первого рода) с учетом односторонности альтернативы (в случае односторонней альтернативы $\alpha = 2\alpha$);

3) выбор критерия значимости (статистики) для проверки гипотезы.

1.6. Критерии значимости

Статистические критерии, с помощью которых проверяются гипотезы о значениях параметров распределения или о соотношения между ними предположении, что вид распределения известен, называются **критериями значимости**.

Статистические критерии, с помощью которых проверяются гипотезы о виде распределения, называются **критериями согласия.** Наиболее известными критериями согласия являются критерий χ^2 Пирсона, критерий Колмогорова и пр.

Рассмотрим критерии значимости, предназначенные для проверки гипотез о величине параметров в случае выборок из нормального распределения, которое на практике встречается наиболее часто. Нормальное распределение имеет два параметра: математическое ожидание μ и дисперсию σ^2, которые оцениваются с помощью выборочного среднего и выборочной дисперсии соответственно.

Выборочное среднее является оценкой для среднего значения измеряемой величины и может служить характеристикой того или иного показателя качества. **Дисперсия** описывает разброс экспериментальных значений, а следовательно, служит мерой точности. Например, если произведено несколько измерений одной и той же величины, то дисперсия может дать представление о точности прибора, методах измерения и т. д.

1.6.1. Проверка гипотезы о равенстве математического ожидания заданному значению

Нулевая гипотеза H_0: $\mu = \mu_0$.

Альтернативная гипотеза $\bar{H}$: $\mu \neq \mu_0$.

Требуется по выборке объема n проверить гипотезу H_0 при заданном уровне значимости α. При этом полагается, что выборка взята из совокупности с нормальным распределением.

Если дисперсия σ^2 известна, то гипотеза H_0 принимается (т. е. согласуются с результатами наблюдений) при условии, что

$$u_{\text{расч}} = \frac{\left|\bar{x} - \mu\right|}{\sqrt{\sigma^2 / n}} < u_{\text{табл}} = u_\alpha, \tag{9}$$

где u_α удовлетворяет соотношению $\Phi(u_\alpha) = \dfrac{1-\alpha}{2}$ и определяется по прил. 1.

Если дисперсия σ^2 неизвестна, то гипотеза H_0 принимается при

$$t_{\text{расч}} = \frac{\left|\bar{x} - \mu_0\right|}{\sqrt{s^2 / n}} < t_{\text{табл}} = t_{\alpha;\, n-1}, \tag{10}$$

где $t_{\alpha;\, n-1}$ определяется по таблице распределения Стьюдента (прил. 3).

1.6.2. Проверка гипотезы о равенстве дисперсии нормального распределения заданному значению

Нулевая гипотеза H_0: $\sigma = \sigma_0^2$.

Альтернативная гипотеза $\bar{H}$: $\sigma^2 \neq \sigma_0^2$.

Гипотеза H_0 при заданном уровне значимости α принимается, если

$$\chi^2_{1-\alpha/2;\, n-1} < \chi^2_{\text{расч}} = \frac{(n-1)s^2}{\sigma_0^2} < \chi^2_{\alpha/2;\, n-1}, \tag{11}$$

где $\chi^2_{1-\alpha/2;\, n-1}$ и $\chi^2_{\alpha/2;\, n-1}$ определяются по таблице распределения χ^2 (прил. 2).

1.6.3. Сравнение двух дисперсий нормально распределенных признаков

Проводится при сравнении точности приборов, инструментов, методов измерения. Лучшим будет тот, который дает меньший разброс результатов, т. е. меньшую дисперсию.

Нулевая гипотеза H_0: $\sigma_1^2 \neq \sigma_2^2$.

Альтернативная гипотеза $\bar{H}$: $\sigma_1^2 \neq \sigma_2^2$.

Пусть для первой дисперсии по выборке объема n_1 найдена несмещенная оценка s_1^2, для второй — по выборке объема n_2 оценка s_2^2. Гипотеза H_0 при заданном уровне значимости α принимается, если

$$F_{\text{расч}} = \frac{s_{\max}^2}{s_{\min}^2} < F_{\text{табл}} = F_{\alpha/2,\, f_1,\, f_2}, \tag{12}$$

где $F_{\text{расч}}$ равно отношению большей несмещенной оценки дисперсии к меньшей; $F_{\alpha/2;\, f_1;\, f_2}$ определяются по таблице распределения Фишера (прил. 4), причем f_1 и f_2 — числа степеней свободы соответственно числителя и знаменателя, т. е. большей и меньшей оценок дисперсий.

1.6.4. Сравнение нескольких дисперсий нормально распределенных признаков

Пусть имеется N независимых выборок и требуется при заданном уровне и значимости α проверить гипотезу о равенстве дисперсий.

Нулевая гипотеза H_0: $\sigma_1^2 = \sigma_2^2 = \ \dots\ = \sigma_N^2$.

Альтернативная гипотеза $\bar{H}$: $\sigma_i^2 \neq \text{const}$.

Если гипотеза о равенстве дисперсий принимается, то эти дисперсии считаются **однородными.**

Пусть по выборкам объемов n_1 и n_2, …, n_N получены несмещенные оценки дисперсий $s_1^2, s_2^2,\ \dots, s_N^2$. Обозначим f_i — число степеней свободы дисперсии s_i^2.

В случае выборок одинакового объема, т. е. $n_1 = n_2 = \dots = n_N = n$, проверка проводится по критерию Кочрена. Гипотеза H_0 при заданном уровне значимости α принимается, если

$$G_{\text{расч}} = \frac{s_{\max}^2}{\sum_{i=1}^{N} s_i^2} < G_{\text{табл}} = G_{\alpha,\, f_1,\, f_2}, \tag{13}$$

где $G_{\text{расч}}$ равно отношению наибольшей оценки дисперсии к сумме всех сравниваемых дисперсий; $G_{\text{табл}} = G_{\alpha;\, f_1;\, f_2}$ определяются по таблице распределения Кочрена (прил. 5), причем $f_1 = n - 1$ — число степеней свободы числителя, $f_2 = N$ — количество сравниваемых дисперсий.

В случае выборок разных объемов для проверки однородности нескольких дисперсий используется более сложный критерий Бартлетта. Однако можно применить также и критерий Фишера, предназначенный для сравнения двух дисперсий. При этом проверяют гипотезу о равенстве наибольшей и наименьшей из сравниваемых дисперсий. Если они признаются равными, то можно принять, гипотезу об однородности всех сравниваемых дисперсий.

1.6.5. Сравнение двух средних в случае независимых нормально распределенных признаков

Нулевая гипотеза H_0: $\mu_1 = \mu_2$.

Альтернативная гипотеза $\overline{H}$: $\mu_1 \neq \mu_2$.

Требуется по выборкам объемов n_1 и n_2 проверить гипотезу H_0 при заданном уровне значимости α.

Если дисперсии σ_1^2 и σ_2^2 известны, то гипотеза H_0 принимается при условии, что

$$u_{\text{расч}} = \frac{\left|\overline{x}_1 - \overline{x}_2\right|}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} < u_{\text{табл}} = u_\alpha, \tag{14}$$

где u_α удовлетворяет соотношению $\Phi(u_\alpha) = \dfrac{1-\alpha}{2}$ и определяется по прил. 1.

Если дисперсии σ_1^2 и σ_2^2 неизвестны, но на основании проверки соответствующей гипотезы по критерию Фишера признаны однородными, то гипотеза H_0 принимается при

$$t_{\text{расч}} = \frac{\left|\overline{x}_1 - \overline{x}_2\right|}{\sqrt{s^2\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}} < t_{\text{табл}} = t_{\alpha,\, f}, \tag{15}$$

где общая средневзвешенная дисперсия s^2 вычисляется по формуле

$$s^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1 + n_2 - 2}$$

и имеет число степеней свободы $f = n_1 + n_2 - 2$; значение $t_{\alpha,\, f}$ определяется по прил. 3.

Если дисперсии σ_1^2 и σ_2^2 неизвестны и на основании проверки по критерию Фишера признаны неоднородными, то проверка проводится по критерию Стьюдента, однако этот критерий является приближенным. В этом случае гипотеза H_0 принимается, если

$$t_{\text{расч}} = \frac{\left|\overline{x}_1 - \overline{x}_2\right|}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}} < t_{\text{табл}} = t_{\alpha,\, f}, \tag{16}$$

где $t_{\alpha,\, f}$ определяется по прил. 3 при

$$f \approx \left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2 \Bigg/ \left(\frac{\left(\dfrac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\dfrac{s_2^2}{n_2}\right)^2}{n_2 - 1}\right).$$

Замечание. При сравнении двух средних в случае неизвестных дисперсий возникает необходимость проверки двух различных гипотез по одним и тем же данным. Сперва проверяют гипотезу о равенстве дисперсий, а затем гипотезу о равенстве средних.

Для сравнения нескольких средних в случае независимых нормально распределенных признаков используется специальная статистическая процедура, которая называется дисперсионным анализом.

1.6.6. Сравнение двух средних в случае зависимых нормально распределенных признаков

Такая задача возникает, если две выборки взаимосвязаны. Например, проводятся измерения одних и тех же величин на одних и тех же объектах двумя разными методами и требуется определить, одинаковые ли получаются результаты. Либо проводятся измерения какой-то характеристики для одних и тех же объектов до и после взаимодействия и требуется определить, влияет ли это воздействие на значения характеристики.

В этом случае имеются две выборки одинакового объема n:

$$\begin{array}{llll} x_{11}, & x_{12}, & \ldots, & x_{1n}; \\ y_{21}, & y_{22}, & \ldots, & y_{2n}. \end{array}$$

Поскольку значения в каждой паре $x_{1i},\ y_{2i}$ связаны (например, измерены на одном и том же объекте), то получим новую выборку с элементами $\Delta d_i = x_{1i} - y_{2i}$. Задача сводится к проверке гипотезы о равенстве нулю среднего значения новой выборки, т. е. $H_0\colon\ \mu_{\Delta d} = 0$.

Эта проверка проводится по критерию (10).

Вводим обозначения: $d_i = x_i - y_i$ — разности вариант с одинаковыми номерами; $\bar{d} = \frac{1}{n}\sum d_i$ — средняя разностей вариант с одинаковыми номерами; $s_d = \sqrt{\frac{\sum d_i^{\,2} - (\sum d_i)^2 / n}{n-1}}$ — исправленное среднее квадратическое отклонение.

Для того чтобы при заданном уровне значимости α проверить нулевую гипотезу $H_0\colon\ \mathrm{M}(x) = \mathrm{M}(y)$ о равенстве двух средних нормальных совокупностей с неизвестными дисперсиями (зависимая выборка) при конкурирующей гипотезе $H_1\colon\ \mathrm{M}(x) \neq \mathrm{M}(y)$, надо вычислить наблюдаемое значение критерия $t_{\text{набл}} = \bar{d}\,\frac{\sqrt{n}}{s_d}$ и по таблице критических точек распределения Стьюдента на заданном уровне α и числу степеней свободы ν найти критическую точку $t_{\text{двуст. кр.}}(\alpha;\nu)$. Если $\left|t_{\text{набл}}\right| < t_{\text{двуст. кр.}}$, то нет оснований отвергать нулевую гипотезу, в обратном случае гипотезу отвергают.

Замечание. Иногда возникает необходимость сравнения гипотезы H_0: $\theta = \theta_0$ с односторонней альтернативой $\bar{H}_1$: $\theta > \theta_0$ или $\bar{H}_2$: $\theta < \theta_0$. Например, при проверке гипотезы H_0: $\mu = \mu_0$ против альтернативы H_0: $\mu > \mu_0$ требуется выяснить, соответствует ли выборочное среднее значение норме или превосходит ее. Пусть дисперсия σ^2 известна. Оценкой для параметра μ является $\bar{x}$.

Ясно, что если $\bar{x} < \mu_0$, то гипотезу H_0 следует предпочесть альтернативе $\bar{H}_1$. В случае $\bar{x} > \mu_0$ гипотеза H_0 принимается на уровне значимости α, если выполняется условие (9) с $u_{\text{табл}} = u_{2\alpha}$, т. е. табличное значение определяется для удвоенного уровня значимости.

Аналогично с удвоенным уровнем значимости определяются табличные значения при использовании критериев (10), (12), (14—16) в случае односторонних альтернатив.

Критерий (11) используется следующим образом. В случае альтернативы $\bar{H}_1$: $\sigma^2 > \sigma_0^2$ гипотеза H_0: $\sigma^2 = \sigma_0^2$ при заданном уровне значимости α принимается, если

$$\chi^2_{\text{расч}} = \frac{(n-1)s^2}{\sigma_0^2} < \chi^2_{\alpha;\ n-1}.$$

В случае альтернативы $\bar{H}_2$: $\sigma^2 < \sigma_0^2$: гипотеза H_0 принимается, если

$$\chi^2_{\text{расч}} = \frac{(n-1)s^2}{\sigma_0^2} > \chi^2_{1-\alpha;\ n-1}.$$

Сделаем несколько замечаний о границах применимости указанных критериев значимости. Все рассмотренные выше критерии проверки гипотез о средних и дисперсиях предназначены для случая нормально распределенных совокупностей. Критерии сравнения дисперсий (11—13) весьма чувствительны к отклонениям распределений от нормального. В то же время критерии сравнения средних (9), (10), (14—16) устойчивы к умеренным отклонениям распределений от нормального; критерий (15) может использоваться при умеренном отклонении от равенства дисперсий, если объемы выборок приблизительно равны. Когда эти условия не выполняются, то для сравнения средних двух выборов используют критерии, не требующие предположения о нормальности распределения.

Примеры решения задач

Во всех примерах предполагается, что выборка взята из совокупности с нормальным распределением. В случае нарушения этого предположения необходимо использовать другие критерии.

Задача 16. Дисперсия предела прочности на разрыв некоторого волокна составляет 35,63 МПа. Ожидается, что внесенные в технологический процесс изменения снизят указанную дисперсию. Можно ли по выборке 151; 156; 147; 153; 155; 148; 160; 149; 160; 156 утверждать, что изменение процесса привело к уменьшению разброса значений предела прочности?

Решение. По условию задачи требуется выяснить, уменьшилась ли дисперсия или осталась прежней. Поскольку описанные выше критерии значимости предназначены для проверки нулевых гипотез о равенстве значений параметров, то следует проверить гипотезу $H_0\colon \sigma^2 = 35,63$ в сравнении с односторонней альтернативной $\overline{H}\colon \sigma^2 < 35,63$. Положим уровень значимости $\alpha = 0,05$.

Будем использовать для проверки гипотезы критерий (11). Рассчитаем выборочное среднее и дисперсию: $\overline{x} = 153,5$; $s^2 = 22,06$; $f = 10 - 1 = 9$. Учитывая вид альтернативы, сравним расчетное значение $\chi^2_{\text{расч}}$ с табличным $\chi^2_{\text{табл}} = \chi^2_{1-\alpha;\, f}$. Поскольку в прил. 2 отсутствуют значения для числа степени свободы $\nu = 9$, определим табличное значение приближенно:

$$\chi^2_{0,95;\ 9} \approx \frac{\chi^2_{0,95;\ 8} + \chi^2_{0,95;\ 10}}{2} = \frac{2,7326 + 3,9403}{2} = 3,33645.$$

Поскольку

$$\chi^2_{\text{расч}} = \frac{9 \cdot 22,06}{35,63} = 5,57 > \chi^2_{\text{табл}} = \chi^2_{0,95;\ 9} = 3,33645,$$

то, в соответствии замечанием относительно односторонних альтернатив, нулевая гипотеза $H_0\colon \sigma^2 = 35,63$ на уровне значимости $\alpha = 0,05$ не противоречит экспериментальным данным, т. е. сравниваемые дисперсии следует признать одинаковыми. Таким образом, изменение процесса не привело к уменьшению разброса наблюдаемых значений предела прочности на разрыв данного волокна.

Задача 17. Пяти лабораториям было поручено участвовать в проведении химического анализа образцов каменного угля с целью определения содержания в них золы. Один образец был расколот на 40 частей, в каждую из лабораторий отправили по 8 кусков. Дисперсии результатов измерений разных лабораторий получились следующими: 3,86; 4,27; 1,35; 3,9; 1,64.

Нужно проверить гипотезу об однородности дисперсий. Что характеризуют эти дисперсии?

Решение. Проверяем гипотезу $H_0\colon \sigma_i^2 = \text{const}$ в сравнении с альтернативой $\overline{H}\colon \sigma_i^2 \neq \text{const}$. Положим уровень значимости $\alpha = 0,05$.

Поскольку каждая из пяти дисперсий получена по результатам восьми измерений, мы имеем случай выборок одинакового объема, число степени свободы каждой дисперсии равно $f = 8 - 1 = 7$. Проверим гипотезу о равенстве этих пяти дисперсий по критерию Кочрена (13).

Вычислим расчетное значение критерия:

$$G_{\text{расч}} = \frac{4,27}{3,86 + 4,27 + 1,35 + 3,9 + 1,64} = \frac{4,27}{15,02} = 0,28$$

и сравним ее с табличным значением

$$G_{\text{табл}} = G_{0,05;\ 7;\ 5} \approx \frac{G_{0,05;\ 6;\ 5} + G_{0,05;\ 8;\ 5}}{2} = \frac{0,4783 + 0,4387}{2} = 0,4585.$$

Поскольку $G_{\text{расч}} < G_{\text{табл}}$, то при уровне значимости $\alpha = 0,05$ сравниваемые дисперсии следует признать однородными.

Дисперсии характеризуют разброс значений измерений, который зависит от содержания золы в куске (однородности образца) и методики проведения измерений, т. е. от качества проведения анализа данной лаборатории. Так как гипотеза об однородности дисперсии принимается, нет оснований считать, что качество анализа в какой-либо лаборатории хуже, чем в других.

Задача 18. Согласно паспортным данным, некий реактив должен содержать не менее 92 % основного вещества. Требуется проверить гипотезу статистической значимости различия между паспортными данными и результатами трех определений содержания основного вещества в реактиве: 92,3; 91,7; 90,8 %.

Решение. Вычислим $\bar{x} = 91,6$ %. Вычислим дисперсию по формуле (4): $s^2 = 0,57.$ Объем выборки $n = 3.$ Требуется определить, достаточно ли сильно номинальное значение 92 % превосходит средний результат измерений $\bar{x} = 91,6$ %, чтобы можно было утверждать, что реактивы не соответствуют паспортным данным. Для этого нужно проверить гипотезу H_0: $\mu = 92$ % при альтернативе H_1: $\mu < 92$ %. Положим уровень значимости $\alpha = 0,05.$

Мы имеем задачу сравнения выборочного среднего с заданным значением в случае неизвестной дисперсии. Такая задача решается с помощью критерия Стьюдента (10). Поскольку рассматривается односторонняя альтернатива, табличное значение должно быть определено при удвоенном уровне значимости: $t_{\text{табл}} = t_{2\alpha;\ n-1} = t_{0,1;\ 2} = 2,92.$

Учитывая, что

$$t_{\text{расч}} = \frac{|91,6 - 92|}{\sqrt{0,57/3}} = 0,918 < t_{\text{табл}} = 2,92,$$

делаем вывод: гипотеза H_0 при уровне значимости $\alpha = 0,05$ принимается. Следовательно, реактив следует признать доброкачественным.

Задача 19. Для того чтобы определить, сокращается ли время сварки на отливках, если при литье вместо сырой формовочной смеси использовать сухую смесь, было проведено специальное исследование. Стоимость литья в случае сухой формовочной смеси выше, но есть мнение, что это может быть оправдано, если время сварки значимо уменьшится. В табл. 4 приведены значения времени сварки в минутах при использовании формовочных смесей разных типов. Можно ли сказать, что имеет место значимое уменьшение времени сварки?

Таблица 4

Номер	Сырая смесь	Сухая смесь
1	19	21
2	28	15
3	14	11
4	29	12
5	15	21
$\bar{x}_i$	21	16
s_i^2	50,5	23

Решение. Требуется сравнить среднее время литья для двух типов смеси. Таким образом, мы имеем задачу сравнения двух средних в случае независимых выборок. Проверяем гипотезу $H_0: \mu_1 = \mu_2$ против альтернативы $\bar{H}_0: \mu_1 > \mu_2$. Положим уровень значимости $\alpha = 0,05$.

В задаче сравнения двух средних необходимо учитывать, однородны ли дисперсии двух выборок. Выдвигаем вспомогательную гипотезу о равенстве двух дисперсий: $H_0': \sigma_1^2 = \sigma_2^2$ и сравниваем ее с альтернативой $\bar{H}_0: \sigma \neq \sigma_2^2$ при заданном уровне значимости $\alpha = 0,05$. Проверка проводится по критерию Фишера (12), поскольку

$$F_{\text{расч}} = \frac{s_{\max}^2}{s_{\min}^2} = \frac{50,5}{23} = 2,2 < F_{\text{табл}} = F_{0,025;\ 4;\ 4} = 9,6,$$

заключаем, что данные дисперсии следует признать однородными.

Общая дисперсия $s^2 = (4 \cdot 50,5 + 4 \cdot 23) / (4 + 4) = 36,75,$ имеет число степеней $f = 4 + 4 = 8$.

Итак, для проверки гипотезы H_0 против альтернативы $\bar{H}$ используем критерий (15). Поскольку альтернатива $\bar{H}$ односторонняя, аналогично задаче 18 сравниваем $t_{\text{расч}}$ и $t_{\text{табл}}$:

$$t_{\text{расч}} = \frac{21 - 16}{\sqrt{36,75\left(\frac{1}{5} + \frac{1}{5}\right)}} = 1,3 < t_{\text{табл}}.$$

Следовательно, нет оснований утверждать, что при использовании сухой смеси время литья значимо уменьшается.

Задача 20. Двумя приборами в одном и том же порядке измерены шесть деталей и получены следующие результаты измерений:

$$x_1 = 2;\ x_2 = 3;\ x_3 = 5;\ x_4 = 6;\ x_5 = 8;\ x_6 = 10.$$

$$y_1 = 10;\ y_2 = 3;\ y_3 = 6;\ y_4 = 1;\ y_5 = 7;\ y_6 = 4.$$

При уровне значимости $\alpha = 0,05$ необходимо установить, значимо или незначимо различаются результаты измерений, если предположить, что они распределены нормально.

Решение. Задача сводится к проверке гипотезы о равенстве нулю среднего значения новой выборки, т. е. H_0: $\mu_{\Delta d} = 0.$ Эта проверка проводится по критерию (10).

Найдем разности:

$$d_i = x_i - y_i;$$

$$d_1 = –8;\ d_2 = 0;\ d_3 = –1;\ d_4 = 5;\ d_5 = 1;\ d_6 = 6.$$

Вычисляем выборочную среднюю: $\overline{d} = \dfrac{3}{6} = \dfrac{1}{2}.$

Найдем исправленное среднее квадратическое отклонение s_d, учитывая, что $\sum d_i^{\ 2} = 127,\ \sum d_i\ \ = 3$:

$$s_d = \sqrt{\frac{\sum d_i^{\ 2} - (\sum d_i)^2 / n}{n-1}} = \sqrt{\frac{127 - 9/6}{6-1}} = \sqrt{25{,}1}.$$

Найдем наблюдаемое значение критерия:

$$t_{\text{набл}} = \overline{d}\,\frac{\sqrt{n}}{s_d} = \frac{1}{2}\frac{\sqrt{6}}{\sqrt{25{,}1}} = 0{,}24.$$

По таблице критических точек распределения Стьюдента прил. 3 находим $t_{0{,}05;\ 5} = 2{,}57.$ Так как $\left|t_{\text{набл}}\right| < t_{\text{двуст. кр}}$, нет оснований отвергнуть нулевую гипотезу. Другими словами, средние результаты измерений различаются незначимо.

Задачи для самостоятельного решения

Задача 21. Семи лабораториям было поручено участвовать в проведении химического анализа образцов цементного клинкера с целью определения в нем трехкальциевого алюмината (C_3A). В каждую из лабораторий отправили по 8 кусков. Дисперсии результатов измерений разных лабораторий получились следующими: 4,86; 4,87; 4,85; 4,9; 4,88; 4,94; 4,91. Проверить гипотезу об однородности дисперсий.

Задача 22. На станке-автомате изготавливаются детали с номинальным контролируемым размером a = 12 мм. Известно, что распределение контролируемого размера является нормальным со среднеквадратическим отклонением $\sigma = 0{,}5$ мм. ОТК в течение смены произвел измерение 36 случайно отобранных деталей и подсчитал средний размер контролируемого параметра $\overline{x} = 11{,}7$ мм. Можно ли утверждать, что станок изготавливает детали уменьшенного размера и требуется произвести регулировку станка? Уровень значимости принять $\alpha = 0{,}05$.

Задача 23. В одном из цехов анализируется работа листопрокатного стана. Основным показателем качества является толщина готового листах, номинал которого 2,11 мм, n_i — частота наблюдения x_i. Можно ли на уровне значимости $\alpha = 0{,}05$ считать, что продукция соответствует номиналу?

x_i	2,04	2,06	2,08	2,09	2,1	2,11	2,12	2,14
n_i	1	1	2	4	10	8	3	1

Задача 24. Проверка скорости полимеризации проводится на нескольких образцах полимеров. Предполагаемая скорость составляет 24 %. В испытаниях получены следующие результаты: 23,6; 22,8; 25,7; 24,8; 26,4; 24,3; 23,9; 25.

Можно ли утверждать на уровне значимости $\alpha = 0,05$, что полученные результаты несовместимы с предполагаемым значением средней скорости?

Задача 25. На станке изготавливают цилиндрические болты определенного типа. Из одной партии болтов взята выборка объемом 20 и произведены измерения длины каждого болта, по которым рассчитаны статистические характеристики $x = 18$ мм, $s^2 = 0,000784$ мм^2. Известна точность (а тем самым и среднее квадратическое отклонение как мера рассеивания) станка при производстве болтов данного типа, которая приблизительно равна $\sigma_0 = 0,02$ мм. Можно ли на уровне значимости $\alpha = 0,05$ по выборочным результатам заключить, что станок дает допустимый разброс для данной партии или же расчетное значение s указывает на несоответствие точности станка предъявленным требованиям?

Задача 26. Оценивается качество работы двух стабилизаторов температуры: с усовершенствованием и без него. Эффективность стабилизаторов температуры измеряется даваемой ими дисперсией температур. Для оценки дисперсии первого стабилизатора проведено 4 опыта, второго — 6. Несмещенные оценки дисперсий оказались равными соответственно 0,016 и 0,07. Необходимо выяснить по результатам эксперимента, можно ли при уровне значимости $\alpha = 0,05$ считать усовершенствование эффективным.

Задача 27. Из текущей продукции горизонтальноковочной машины за семь смен ее работы отобрано семь проб, по одной пробе в смену, численностью 17 штамповок каждая. По данным каждой из этих проб подсчитаны выборочные дисперсии (в мм^2): 0,067; 0,136; 0,168; 0,068; 0,066; 0,102; 0,137.

Требуется при уровне значимости $\alpha = 0,05$ проверить гипотезу об отсутствии разладки машины по рассеиванию размеров штамповок за семь смен ее работы.

Задача 28. Исследовалась работа промышленного агрегата по процессу извлечению этана из природного газа. Целью является более полное (близкое к 100 %) отделение этана. Испытывались два технологических режима, чтобы выбрать более эффективный. Можно ли при уровне значимости $\alpha = 0,05$ считать, что обе технологии дают одинаковые результаты (табл. 5)?

Таблица 5

Технология	$\overline{x_i}$	s_i^2	n_i
Первая	99,2	0,14	110
Вторая	99,34	0,123	60

Задача 29.

Объектом исследования является крупноформатный пустотно-поризованный камень. Оборудование: пресс; ИПС МГ 4.03. Полученные результаты приведены в табл. 6.

Таблица 6

Камень	$R_{ср}$, кгс/см^2	
1	210,5	175,5
2	244	198,3
3	213,5	183
4	219	186
5	141	176
6	120,5	221,9

Таким образом, имеются две независимые выборки. На уровне значимости $\alpha = 0{,}05$ проверим гипотезу о равенстве средних. Цель исследования — сопоставить значения разрушающей нагрузки с результатами, полученными при помощи проборов неразрушающего контроля.

Ответы к задачам

21. 1) H_0: $\sigma_i = \text{const}$; $\bar{H}$: $\sigma_i \neq \text{const}$; $G_{расч} = 0{,}144$; $G_{табл} = 0{,}3645$; дисперсии однородны. **22.** H_0: $\mu = 12$; $\bar{H}$: $\mu < 12$; $u_{расч} = 3{,}6$; $u_{табл} = 1{,}645$; станок изготавливает детали уменьшенного размера, требуется произвести регулировку. **23**. H_0: $\mu = 12{,}11$; $\bar{H}$: $\mu \neq 12{,}11$; дисперсия неизвестна; $s^2 = 0{,}000345$; $t_{расч} = 2{,}95$; $t_{табл} = t_{0,05;\ 29} = 2{,}04$; гипотеза о соответствии продукции номиналу должна быть отвергнута. **24**. H_0: $\mu = 24$; $\bar{H}$: $\mu \neq 24$; нет оснований утверждать, что полученные результаты несовместимы с предполагаемым значением средней скорости. $t_{расч} = 1{,}37$; $t_{табл} = t_{0,05;\ 7} = 2{,}4$. **25.** $\chi^2_{расч} = 37{,}24$; $\chi^2_{табл} = \chi^2_{0,05;20} = 31{,}41$; расчетное значение s указывает на несоответствие точности станка предъявленным требованиям. **26**. H_0: $\sigma_1^2 = \sigma_2^2$; $\bar{H}$: $\sigma_1^2 < \sigma_2^2$; $F_{расч} = 4{,}375$; $F_{табл} = F_{0,05;\ 5;\ 3} = 12{,}745$; нет оснований считать дисперсии неоднородными, усовершенствование неэффективно. **27.** H_0: $\sigma_i^2 = \text{const}$; $\bar{H}$: $\sigma_i^2 \neq \text{const}$; $G_{расч} = 0{,}226$; $G_{табл} = G_{0,05;\ 16;\ 7} = 0{,}27985$, разладки машины нет. **28.** H_0: $\mu_1 = \mu_2$; $\bar{H}$: $\mu_1 \neq \mu_2$. Проверим вспомогательную гипотезу H_0': $\sigma_1^2 = \sigma_2^2$; H': $\sigma_1^2 \neq \sigma_2^2$; по критерию Фишера $F_{расч} = 1{,}14$; $F_{табл} = F_{0,025;\ 109;\ 59} = 1{,}48$; дисперсии однородны, используем критерий Стьюдента в форме (15); $t_{расч} = 2{,}38$; $t_{табл} = t_{0,05;\ 168} = 1{,}975$; гипотеза о том, что обе технологии дают одинаковые результаты не может быть принята. **29.** Задача сводится к проверке гипотезы о равенстве нулю среднего значения новой выборки, т. е. $H_0\colon \mu_{\Delta d} = 0.$ Эта проверка проводится по критерию (10). $t_{\text{расч}} = 0,055;$ Нет оснований отвергнуть нулевую гипотезу.

ГЛАВА 2. ЭЛЕМЕНТЫ РЕГРЕССИОННОГО АНАЛИЗА

2.1. Основные задачи корреляционного и регрессионного анализа

Статистические связи между переменными изучаются методами корреляционного и регрессионного анализа. **Основными задачами корреляционного анализа** являются выявление связи между переменными и оценка тесноты этой связи. **Основными задачами регрессионного анализа** являются установление формы зависимости, определение уравнения зависимости по экспериментальным данным, прогноз среднего значения зависимой переменной при заданном значении независимой переменной.

Пусть на основании экспериментальных данных изучается связь между двумя величинами. Тогда выборка объема N представляет собой N пар значений (x_i, y_i).

Две случайные величины могут быть: 1) независимы; 2) связаны функциональной зависимостью, когда каждому значению одной величины соответствует, строго определенное значение другой; 3) связаны статистической зависимостью, при которой каждому значению одной величины соответствует множество возможных значений другой. **Статистической (вероятностной, стохастической)** называется зависимость, при которой изменение значения одной величины влечет изменение распределения другой. Например, статистической будет зависимость урожайности зерновой культуры от количества вносимых удобрений, зависимость спроса на товар от его цены, надежности автомобиля от его возраста и т. д.

При изучении статистической зависимости обычно ограничиваются исследованием усредненной зависимости: как в среднем будет изменяться значение одной величины при изменении другой. Такая зависимость называется **регрессионной.** Более строго, регрессионная зависимость между двумя случайными величинами — это функциональная зависимость между значениями одной из них и условным математическим ожиданием другой.

На практике по экспериментальным данным можно найти только оценку (приближенное выражение) функции регрессии. Говорят о том, что по выборке определяют **выборочное (эмпирическое) уравнение регрессии.** Как правило, до проведения эксперимента выбирают вид выборочной функции регрессии с точностью до нескольких параметров, а значения параметров определяют по выборке. Вид эмпирической функции регрессии определяют исходя из: 1) теоретических соображений о физической сущности исследуемой зависимости; 2) опыта предыдущих исследований; 3) характера расположения точек на **корреляционном поле,** которое получается, если отметить на плоскости все точки с координатами (x_i, y_i), соответствующие наблюдениям. Наибольший интерес представляет линейное эмпирическое уравнение регрессии, поскольку: 1) это наиболее простой случай для расчетов и анализа; 2) при нормальном распределении модельная функция регрессии является линейной.

2.2. Выборочный коэффициент корреляции и его свойства

Количественной мерой линейной связи между двумя наблюдаемыми величинами служит **выборочный коэффициент корреляции:**

$$r_{x,\,y} = \frac{\sum_{i=1}^{N} x_i y_i - N\overline{xy}}{\sqrt{(\sum_{i=1}^{N} x_i^2 - N(\overline{x})^2)(\sum_{i=1}^{N} y_i^2 - N(\overline{y})^2)}}, \tag{17}$$

где $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$; $\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$; N — объем выборки.

Выборочный коэффициент корреляции обладает следующими свойствами:

$\left|r_{x,\,y}\right| \le 1,$ при этом если $r_{x,\,y}$ близок к единице, то наблюдаемые величины x и y связаны линейной зависимостью $y = b_0 + b_1 x$;

если $r_{x,\,y} = 0,$ то наблюдаемые величины независимы.

Часто теснота связи величин оценивается по таблице Чеддока.

Для корреляционного поля с выборочным коэффициентом корреляции $r_{x,\,y} \approx 0{,}95$ т. е. близком к 1, экспериментальные точки расположены вдоль прямой. Если корреляционное поле менее сконцентрировано около прямой линии, коэффициент корреляции уменьшается (по модулю). В практических задачах даже значение $r_{x,\,y} \approx 0{,}7$ может быть признано довольно высоким, свидетельствующим о линейной зависимости между переменными. Если разбросаны, экспериментальные точки зависимость между увеличением значений одной переменной и увеличением или уменьшением значений второй не прослеживается, коэффициент корреляции близок к нулю.

Подчеркнем, что коэффициент корреляции является мерой именно линейной зависимости. Когда точки образуют кривые линии, т. е. зависимость между переменными хотя и явно присутствует, но отличается от линейной, коэффициент корреляции не равен единице по абсолютной величине. В случае нелинейной зависимости связь между величиной коэффициента корреляции и близостью точек корреляционного поля к некоторой линии не прослеживается. Поэтому в практических задачах при выборе вида эмпирической функции регрессии обязательно учитывают характер расположения точек на корреляционном поле.

Проверка значимости коэффициента корреляции (т. е., значимо ли отличается выборочный коэффициент корреляции от нуля) производится в случае выборки из нормального распределения по критерию Стьюдента.

Нулевая гипотеза H_0: $r = 0$.

Альтернативная гипотеза $\overline{H}$: $r \neq 0$.

Гипотезу H_0 при заданном уровне значимости α принимают, если

$$t_{\text{расч}} = \left| r_{x,\,y} \right| \sqrt{\frac{N-2}{1 - r_{x,\,y}^2}} < t_{\text{табл}} = t_{\alpha;\,N-2}, \tag{18}$$

где N — количество наблюдений; значение $t_{\text{табл}}$ определяется по прил. 3.

Если $t_{\text{расч}} > t_{\text{табл}}$, то считают, что коэффициент корреляции значимо отличается от нуля, а следовательно, связь между наблюдаемыми величинами является статистически значимой.

2.3. Определение коэффициентов эмпирического уравнения регрессии в случае линейной однофакторной зависимости

Пусть имеется выборка объема N наблюдений над двумя величинами x и y и принята гипотеза о линейной регрессионной зависимости между ними, т. е. оценкой для модельного уравнения регрессии y на x является эмпирическое линейное уравнение регрессии $\hat{y} = b_0 + b_1 x$.

Далее для нахождения коэффициентов b_0 и b_1 используем **метод наименьших квадратов (МНК)**. Идея МНК заключается в следующем: значения коэффициентов b_1 и b_0 выбирают так, чтобы сумма квадратов отклонений наблюдаемых значений y_i от предсказываемых по уравнению регрессии $\hat{y} = b_0 + b_1 x_i$ была наименьшей, т. е.

$$S = \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{N} (y_i - (b_0 + b_1 x_i))^2 \to \min_{b_0,\, b_1}. \tag{19}$$

Удовлетворяющие этому условию значения параметров b_0 и b_1 находят из системы, которая называется системой нормальных уравнений метода наименьших квадратов:

$$\begin{cases} N b_0 + b_1 \sum\limits_{i=1}^{N} x_i = \sum\limits_{i=1}^{N} y_i, \\ b_0 \sum\limits_{i=1}^{N} x_i + b_1 \sum\limits_{i=1}^{N} x_i^2 = \sum\limits_{i=1}^{N} x_i y_i. \end{cases} \tag{20}$$

Пример решения задачи

Задача 30. При контроле качества жидкой химической добавки для определения концентрации тех или иных веществ находят эмпирические линейные уравнения зависимостей. Ниже приведены данные для определения концентрации формальдегида в нафталинформальдегидной добавке.

Концентрация раствора, мг/кг	0,02	0,04	0,06	0,08	0,10
Оптическая плотность раствора	0,035	0,070	0,150	0,140	0,175

По имеющимся данным требуется:

1) построить корреляционное поле;

2) найти выборочный коэффициент корреляции и проверить его значимость при $\alpha = 0{,}05$;

3) определить коэффициенты эмпирического линейного уравнения регрессии, построить полученную прямую на корреляционном поле.

Решение. 1. Построим корреляционное поле, отмечая по оси Ox данные концентрации, а по оси Oy — соответствующие им оптические плотности (рис. 5).

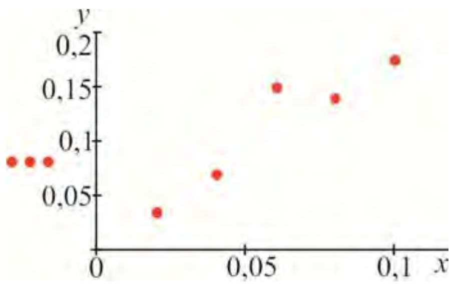


Рис. 5

2. По виду корреляционного поля можно предположить, что выборочный коэффициент корреляции положителен и значимо отличается от 0. Рассчитаем его по формуле (17). Число опытов $N = 5$; вычислим необходимые суммы. Для расчета составим вспомогательную таблицу (табл. 6).

Таблица 6

Номер п/п	Концентрация x_i	Оптическая плотность y_i	$x_i y_i$	x_i^2	y_i^2
1	0,02	0,035	0,0007	0,0004	0,001225
2	0,04	0,070	0,0028	0,0016	0,004900
3	0,06	0,150	0,0090	0,0036	0,022500
4	0,08	0,140	0,0112	0,0064	0,019600
5	0,10	0,175	0,0175	0,0100	0,030625
$\sum$	0,30	0,570	0,0412	0,0220	0,078850

Подставляем полученные значения в (16), получим:

$$r_{x,\,y} = \frac{0{,}0412 - 5\dfrac{0{,}30}{5}\dfrac{0{,}570}{5}}{\sqrt{\left(0{,}0220 - 5\left(\dfrac{0{,}30}{5}\right)^2\right)\left(0{,}078850 - 5\left(\dfrac{0{,}570}{5}\right)^2\right)}} \approx 0{,}94.$$

Поскольку

$$t_{\text{расч}} = 0,94\sqrt{\frac{5-2}{1-0,94^2}} = 4,77 > t_{\text{табл}} = t_{0,05;\ 5-2} = 3,18,$$

то при уровне значимости α = 0,05 следует признать, что выборочный коэффициент корреляции значимо отличается от 0. Учитывая расположение точек на корреляционном поле, можно считать, что величины x и y связаны линейной зависимостью, т. е. $\hat{y} = b_0 + b_1 x$.

Для того чтобы найти коэффициенты b_0 и b_1 эмпирического линейного уравнения регрессии $\hat{y} = b_0 + b_1 x$ по методу наименьших квадратов, составим систему нормальных уравнений по формуле (20):

$$\begin{cases} 5b_0 + 0,3b_1 = 0,57, \\ 0,3b_0 + 0,022b_1 = 0,0412. \end{cases}$$

Решим эту систему линейных уравнений в системе Mathcad:

given

$$5b0 + \frac{3}{10}\cdot b1 = \frac{57}{100}$$

$$\frac{3}{10}\cdot b0 + \frac{22}{1000}\cdot b1 = \frac{412}{1000}$$

$$\text{Find}(b0, b1) \rightarrow \begin{pmatrix} \frac{9}{1000} \\ \frac{7}{4} \end{pmatrix}$$

Чтобы построить прямую на корреляционном поле, подставим в это уравнение два значения переменной x. Например, при x = 0 получим y = 0,009; при x = 0,1 получим y = 0,184. Согласно методу наименьших квадратов, построенная прямая приближает экспериментальные данные наилучшим образом т. е. сумма квадратов длин вертикальных отрезков, показанных на рис. 6, будет наименьшей.

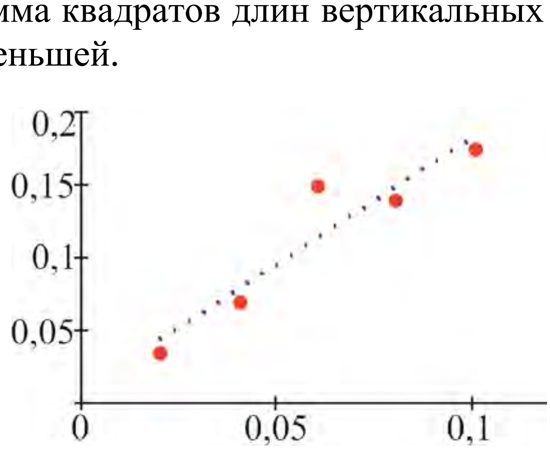


Рис. 6

Задачи для самостоятельного решения

Задача 31. Для разработки методики прогнозирования морозостойкости керамического кирпича на предприятии «Волгоградский кирпичный завод» проводился сбор статистической информации. Измеряемые свойства и их величины приведены в табл. 8.

Таблица 8

Плотность x_1, %	Водопоглощение под вакуумом x_2, %	Прочность на сжатие x_3, %	Морозостойкость y, цикл
19,5	18,4	15,4	30
18,5	17,9	15,4	30
37,9	18,9	13,7	76
25,8	19	15,1	41
31,7	18,7	15,4	60
33,1	18,2	14,2	54
34	18,9	14,1	65
27,2	19,2	15,1	40

По имеющимся данным для переменных x и y требуется:

1) построить корреляционное поле;

2) найти выборочный коэффициент корреляции и проверить его значимость при $\alpha = 0,05$;

3) определить коэффициенты эмпирического линейного уравнения регрессии, если это целесообразно;

4) построить прямую на корреляционном поле.

Вспомогательные величины: $\sum x_{1\,i} = 227,7$; $\sum x_{1\,i}^2 = 6820,89$; $\sum y_i = 396$; $\sum y^2 = 21598$; $\sum x_{1\,i} y_i = 12065,6$.

Задача 32. Решить задачу 31 для переменных x_2 и y. Вспомогательные величины: $\sum x_{2\,i} = 149,2$; $\sum x_{2\,i}^2 = 2783,96$; $\sum x_{2\,i} y_i = 7405,7$.

Задача 33. Решить задачу 31 для переменных x_3 и y. Вспомогательные величины: $\sum x_{3\,i} = 118,4$; $\sum x_{3\,i}^2 = 1755,64$; $\sum x_{3\,i} y_i = 5795,6$.

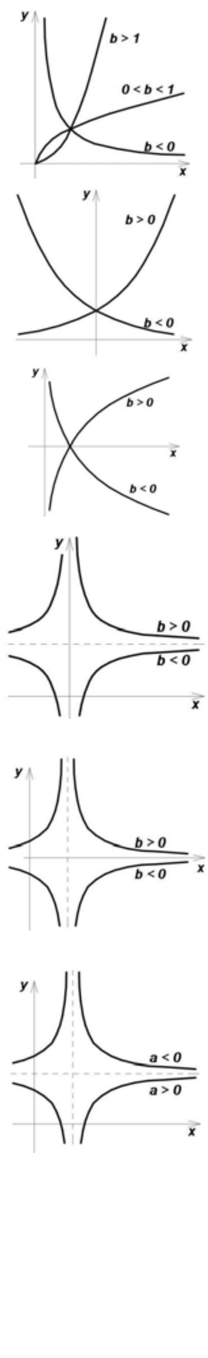
Ответы к задачам

31. 2) $r_{x,\,y} = 0,964$; $t_{\text{табл}} = t_{0,05;\,6} \approx \frac{1}{8-5}(1t_{0,05;\,8} + 2t_{0,05;\,5}) = 2,48$; $t_{\text{расч}} = 8,93 > t_{\text{табл}} = t_{0,05;\,6}$. Следовательно, при уровне значимости $\alpha = 0,05$ можно принять гипотезу о линейной зависимости величины y от x_1; 3) $\hat{y} = -17,01 + 2,34x_1$. **32.** 2) $r_{x,\,y} = 0,387$; $t_{\text{расч}} = 1,03 < t_{\text{табл}}$. При уровне значимости $\alpha = 0,05$ коэффициент корреляции следует признать незначимым, определение линейного уравнения зависимости величины y от x_2 нецелесообразно, прогнозирование морозостойкости изделий по водопоглощению под вакуумом невозможно. **33.** 2) $r_{x,\,y} = -0,8$. Коэффициент корреляции значимо отличается от нуля, принимается гипотеза о линейной зависимости между переменными; 3) $\hat{y} = 340{,}15 - 19{,}64x_3$.

2.4. Криволинейная регрессия

На практике зависимость между двумя величинами далеко не всегда можно выразить линейной функцией. По расположению точек на кор-реляционном поле исследователем определеятеся вид зависимости с точностью до нескольких параметров. Предварительно с помощью замены переменных функцию стараются свести к линейной (если это возможно), а далее для определения искомых параметров может быть использован МНК. следует иметь в виду, что МНК может быть использован для функции любого типа, но в данном случае исследователь сталкивается с решением достаточно сложных систем нелинейных уравнений.

Ниже приведены некоторые примеры замены элементарных функций линейными (рис. 7).



Зависимость вида $y = ax^b$
сводится к линейной $Y = b_0 + b_1X$
заменой $Y = \ln(y)$; $X = \ln(x)$, где $a = e^{b_0}$; $b = b_1$.

Зависимость вида $y = a e^{bx}$
сводится к линейной $Y = b_0 + b_1X$
заменой $Y = \ln(y)$; $X = \ln(x)$, где $a = e^{b_0}$; $b = b_1$.

Зависимость вида $y = a + b\ln(x)$
сводится к линейной $Y = b_0 + b_1X$
заменой $Y = y$; $X = \ln(x)$, где $a = b_0$; $b = b_1$.

Зависимость вида $y = a + \dfrac{b}{x}$
сводится к линейной $Y = b_0 + b_1X$
заменой $Y = y$; $X = \dfrac{1}{x}$, где $a = b_0$; $b = b_1$.

Зависимость вида $y = \dfrac{1}{a + bx}$
Сводится к линейной $Y = b_0 + b_1X$
Заменой $Y = \dfrac{1}{y}$; $X = x$, где $a = b_0$; $b = b_1$.

Зависимость вида $y = \dfrac{x}{a + bx}$
сводится к линейной $Y = b_0 + b_1X$
заменой $Y = \dfrac{x}{y}$; $X = x$, где $a = b_0$; $b = b_1$.

Рис. 7

Пример решения задачи

Задача 34. Установить зависимость твердения бетона от времени:

t, сут.	1	3	6	10	16	21	28
V, %	12	32	51	72	93	97	100

Решение. Построим корреляционное поле, отмечая на горизонтальной оси время в сут., а по вертикальной — % твердения бетона (рис. 8).

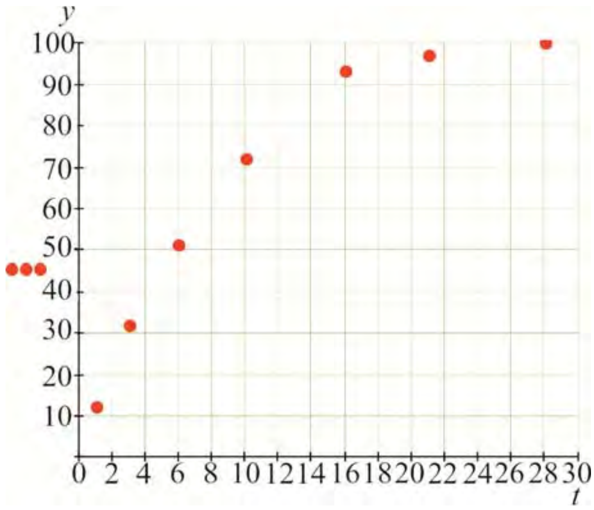


Рис. 8

По виду зависимости предполагаем логарифмическую зависимость. $V = a + b\ln(t)$. Сделаем замену переменных и построим новое корреляционное поле, где $X = \ln(t)$, $Y = V$ (рис. 9).

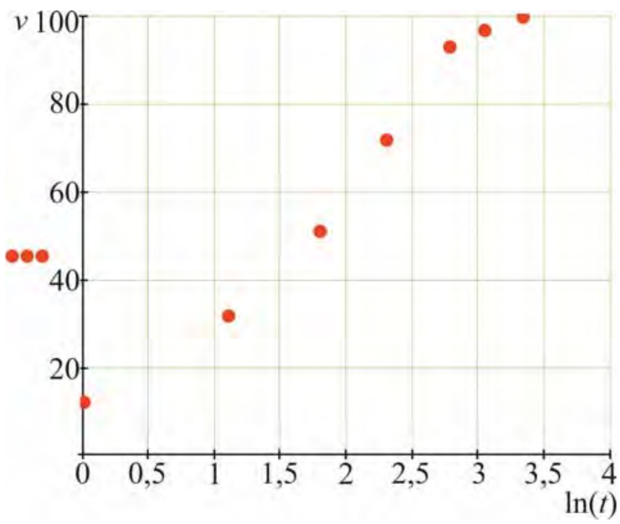


Рис. 9

Заметим, что точки группируются вдоль некоторой кривой линии. Рассмотрим еще одну зависимость, пытаясь получить линейный график. Сделаем замену $X = t$ на $Y = \dfrac{t}{V}$. Точки выстраиваются вдоль прямой (рис. 10).

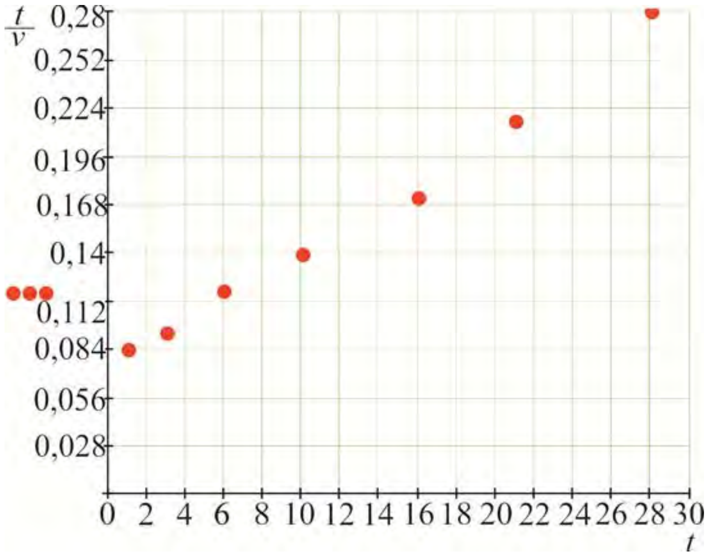


Рис. 10

Найдем коэффициенты зависимости $V = \dfrac{t}{a + bt}$, для этого определяем коэффициенты линейного уравнения регрессии $Y = b_0 + b_1 X$, где $X = t$ $Y = \dfrac{t}{V}$, $b_0 = a,\ b_1 = b.$

Составляем систему нормальных уравнений по формулам (20). Необходимые вычисления производим в дополнительной таблице (табл. 9):

Таблица 9

Номер п/п	X_i	Y_i	$X_i Y_i$	X_i^2
1	1	0,083	0,083	1
2	3	0,094	0,282	9
3	6	0,118	0,708	36
4	10	0,139	1,39	100
5	16	0,172	2,752	256
6	21	0,216	4,536	441
7	28	0,28	7,84	784
Σ	85	1,102	17,591	1627

$$\begin{cases} 7b_0 + 85b_1 = 1{,}102, \\ 85b_0 + 1627b_1 = 17{,}591. \end{cases}$$

Проводим вычисления в системе Mathcad:

Given
7·b0 + 85·b1 = 1.102
85·b0 + 1627·b1 = 17.591
$\mathrm{Find}(b0,b1) \rightarrow \begin{pmatrix} 0{,}071498318924111431316 \\ 0{,}0070766090297790585975 \end{pmatrix}$

Получаем, $b_0 = 0{,}071$ $b_1 = 0{,}0071$. Тогда линейное уравнение регрессии имеет вид $Y = 0{,}071 + 0{,}0071X$, а искомая зависимость запишется как $V = \dfrac{t}{0{,}071 + 0{,}0071t}$ (рис. 11).

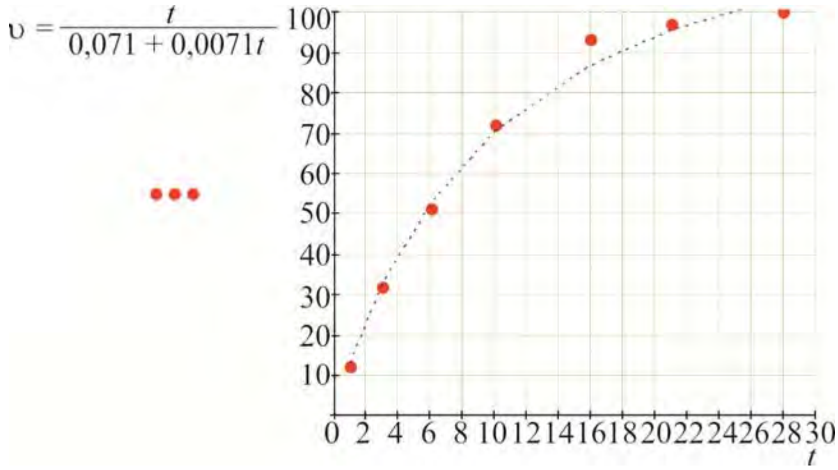


Рис. 11

Задачи для самостоятельного решения

Задача 35. Установить вид зависимости по данным:

t	2	4	12	25	30	60	80
V	3	7	12	16	17	20	22

Задача 36. Установить вид зависимости по данным:

t	0	5	10	15	20	25	30
Y	42	54	74	100	136	182	254

Задача 37. Установить вид зависимости по данным:

t	1	1,2	1,4	1,7	1,9	2	2,3
Z	2,0	3,6	5,7	10,5	14,8	17,3	26,7

Ответы к задачам

35. $V = 5\ln(t)$; **36.** $Y = 40{,}8e^{0{,}06t}$; **37.** $Z = 2{,}02x^{3{,}1}$**.**

2.5. Множественная регрессия

Во многих случаях необходимо исследовать зависимость величины y от нескольких переменных $X_1, X_2, ..., X_n$. Эти переменные $X_1, X_2, ..., X_n$ называются **факторами,** а зависимая переменная y — **откликом (параметром оптимизации).** Зависимость отклика от изучаемых факторов $y = f(X_1, X_2, ..., X_n)$ называется **функцией отклика.**

Поскольку вид функции отклика, как правило, неизвестен, ее представляют в виде полинома (многочлена) и, находя по экспериментальным данным оценки коэффициентов полинома, получают эмпирическое уравнение регрессии в виде

$$\hat{y} = b_0 + \sum_{j=1}^{l} b_j X_j + \sum_{j<t}^{l} b_{j,t} X_j X_t + \sum_{j=1}^{l} b_{j,j} X_j^2 + \ldots \tag{21}$$

Коэффициент b_0 называется оценкой свободного члена уравнения регрессии, коэффициенты b_j — оценками линейных эффектов, $b_{j,t}$ — оценками эффектов взаимодействия, $b_{i,i}$ — оценками квадратичных эффектов.

На практике обычно ограничиваются рассмотрением задачи определения коэффициентов линейного

$$\hat{y} = b_0 + \sum_{j=1}^{l} b_j X_j + b_2 X_2 + \ldots + b_1 X_2 + \ldots + b_l X_l \tag{22}$$

или квадратичного

$$\hat{y} = b_0 + \sum_{j=1}^{l} b_j X_j + \sum_{j<t} b_{j,t} X_j X_t + \sum_{j=1}^{l} b_{j,j} X_j^2 \tag{23}$$

уравнения регрессии. В некоторых случаях получают уравнение вида

$$\hat{y} = b_0 + \sum_{j=1}^{l} b_j X_j + \sum_{j<t} b_{j,t} X_j X_t, \tag{24}$$

которое называется неполным квадратичным или уравнением с парными взаимодействиями.

Постановка задачи. Пусть выбран вид функции регрессии и по результатам N наблюдений (опытов) над откликом y и факторами $X_1, X_2, \ldots, X_n$. Требуется оценить коэффициенты эмпирического уравнения регрессии (21).

Поскольку уравнение (21) линейно относительно параметров $b_j,\ b_{j,t},\ b_{j,j},\ \ldots,$ то для оценки коэффициентов можно использовать МНК. Идея обобщения метода наименьших квадратов на случай регрессионной модели вида (21) заключается в том, что любое произведение факторов или их степень можно рассматривать в качестве нового фактора.

Упростим систему обозначений — заменим члены второго и более высоких порядков линейными:

$$X_{1+1} = X_1 X_2, \quad X_{1+2} = X_1 X_3, \ \ldots \tag{25}$$

Кроме того, введем фиктивную переменную $X_0 = 1$, которая всегда принимает значение 1.

Тогда уравнение (21) будет записываться как однородное линейное уравнение:

$$\hat{y} = b_0 X_0 + b_1 X_1 + b_2 X_2 + \ldots + b_k X_k. \tag{26}$$

Аналогично случаю одного фактора, коэффициенты $b_0, b_1, b_2, ..., b_k$ определяются из системы нормальных уравнений:

$$\begin{cases} b_0 \sum_{i=1}^{N} X_{0\,i}^2 + b_1 \sum_{i=1}^{N} X_{0\,i} X_{2\,i} + b_2 \sum_{i=1}^{N} X_{0\,i} X_{2\,i} + \dots + b_k \sum_{i=1}^{N} X_{0\,i} X_{k,\,i} = \sum_{i=1}^{N} X_{0\,i} y_i, \\ b_0 \sum_{i=1}^{N} X_{0\,i} X_{1\,i} + b_1 \sum_{i=1}^{N} X_{1\,i}^2 + b_k \sum_{i=1}^{N} X_{1\,i} X_{2\,i} + \dots + b_k \sum_{i=1}^{N} X_{1\,i} X_{k,\,i} = \sum_{i=1}^{N} X_{1\,i} y_i, \\ \dots \\ b_0 \sum_{i=1}^{N} X_{0\,i} X_{k,\,i} + b_1 \sum_{i=1}^{N} X_{1\,i} X_{k,\,i} + b_2 \sum_{i=1}^{N} X_{2\,i} X_{k,\,i} + \dots + b_k \sum_{i=1}^{N} X_{k,\,i}^2 = \sum_{i=1}^{N} X_{k,\,i} y_i. \end{cases} \quad (27)$$

Здесь $X_{j,\,i}$ означает значение фактора X_j в i-м опыте. Заметим, что в системе нормальных уравнений количество уравнений всегда равно количеству неизвестных. В данном случае неизвестными являются $b_0, b_1, b_2, ..., b_k$, следовательно, система имеет $k+1$ уравнение.

Запомнить систему нормальных уравнений можно следующим образом. Запишем уравнение (26) N раз, подставив в него результаты каждого из N опытов. Если каждое из полученных соотношений умножить на соответствующее значение фактора X_0, а потом просуммировать все эти соотношения, то получим первое уравнение системы. Если умножать на значения фактора X_1, то получим второе уравнение системы и т. д. При умножении на значения фактора X_k получим последнее уравнение системы.

Итак, МНК можно использовать для определения коэффициентов уравнения регрессии не только в случае линейной зависимости параметра y от факторов. Для применения МНК важно, чтобы коэффициенты $b_0, b_1, b_2, ..., b_k$ входили в уравнение регрессии линейно и чтобы число оцениваемых коэффициентов было не больше числа различных опытов. При вычислении коэффициентов нелинейного уравнения регрессии вводят замену переменных: обозначают через $X_0, X_1, X_2, ..., X_k$ те факторы или функции от исходных факторов, при которых стоят неизвестные искомые коэффициенты $b_0, b_1, b_2, ..., b_k$. При этом необходимо, чтобы новые переменные были линейно независимы (т. е. их значения не могут быть пропорциональными, линейно выражаться друг через друга).

Пример решения задачи

Задача 38. По имеющимся результатам эксперимента (табл. 10) получить (если это возможно):

1) линейное уравнение регрессии;

2) уравнение вида $\hat{y} = b_0 + b_{12} X_1 X_2 + b_{22} X_2^2$;

3) квадратичное уравнение регрессии;

4) уравнение $\hat{y} = b_0 + b_1 X_1 + b_2 X_2 + b_{11} X_1^2$.

Таблица 10

Номер опыта	X_{1i}	X_{2i}	y_i
1	1	0	6
2	0	2	1
3	1	–1	8
4	0	1	2

Решение. 1. Линейное уравнение регрессии в случае двух факторов имеет вид $\hat{y} = b_0 + b_1 X_1 + b_2 X_2$.

Для нахождения его коэффициентов составим систему нормальных уравнений по формуле (27). Запишем систему в общем виде, учитывая, что $X_{0i} = 1$ для всех опытов, а число опытов $N = 4$:

$$\begin{cases} 4b_0 + b_1 \sum_{i=1}^{N} X_{1\,i} + b_2 \sum_{i=1}^{N} X_{2\,i} = \sum_{i=1}^{N} y_i, \\ b_0 \sum_{i=1}^{N} X_{1\,i} + b_1 \sum_{i=1}^{N} X_{1\,i}^2 + b_2 \sum_{i=1}^{N} X_{1\,i} X_{2\,i} = \sum_{i=1}^{N} X_{1\,i} y_i, \\ b_0 \sum_{i=1}^{N} X_{2\,i} + b_1 \sum_{i=1}^{N} X_{1\,i} X_{2\,i} + b_2 \sum_{i=1}^{N} X_{2\,i}^2 = \sum_{i=1}^{N} X_{2\,i} y_i. \end{cases}$$

Для вычисления нужных сумм составим таблицу (табл. 11).

Таблица 11

Номер опыта	$X_{1\,i}$	$X_{2\,i}$	y_i	$X_{1\,i}^2$	$X_{1\,i}X_{2\,i}$	$X_{1\,i}y_i$	$X_{2\,i}^2$	$X_{2\,i}y_i$
1	1	0	6	1	0	6	0	0
2	0	2	1	0	0	0	4	2
3	1	–1	8	1	–1	8	1	–8
4	0	1	2	0	0	0	1	2
Σ	2	2	17	2	–1	14	6	–4

Запишем систему нормальных уравнений:

$$\begin{cases} 4b_0 + 2b_1 + 2b_2 = 17, \\ 2b_0 + 2b_0 - b_2 = 14, \\ 2b_0 - b_1 + 6b_2 = -4. \end{cases}$$

Для ее решения выразим b_2 из второго уравнения и подставим в первое и третье:

$$\begin{cases} 4b_0 + 2b_1 + 2(2b_0 + 2b_1 - 14) = 17, \\ b_2 = 2b_0 + 2b_1 - 14, \\ 2b_0 - b_1 + 6(2b_0 + 2b_1 - 14) = -4. \end{cases} \Rightarrow \begin{cases} 8b_0 + 6b_1 = 45, \\ b_2 = 2b_0 + 2b_1 - 14, \\ 14b_0 + 11b_1 = 80. \end{cases}$$

Выразим из первого уравнения $b_1 = \dfrac{45 - 8b_0}{6}$ и подставим в третье:

$$14b_0 + 11\frac{45 - 8b_0}{6} = 80;$$

$$6 \cdot 14b_0 + 11(45 - 8b_0) = 6 \cdot 80;$$

$$-4b_0 = -15;$$

$$b_0 = 3{,}75.$$

Следовательно

$$b_1 = \frac{45 - 8 \cdot 3{,}75}{6} = 2{,}5; \; b_2 = 2 \cdot 3{,}75 + 2 \cdot 2{,}5 - 14 = -1{,}5.$$

Таким образом, искомая линейная зависимость имеет вид

$$\hat{y} = 3{,}75 + 2{,}5X_1 - 1{,}5X_2.$$

2. Поскольку указанное уравнение имеет три неизвестных коэффициента, система нормальных уравнений будет содержать три уравнения. Запишем систему в общем виде с помощью формулы (27):

$$\begin{cases} 4b_0 + b_{1,\,2}\sum\limits_{i=1}^{N} X_{1\,i}X_{2\,i} + b_{2,\,2}\sum\limits_{i=1}^{N} X_{2\,i}^2 = \sum\limits_{i=1}^{N} y_i, \\ b_0\sum\limits_{i=1}^{N} X_{1\,i}X_{2\,i} + b_{1,2}\sum\limits_{i=1}^{N} X_{1\,i}^2X_{2\,i}^2 + b_{2,\,2}\sum\limits_{i=1}^{N} X_{1\,i}X_{2i}^3 = \sum\limits_{i=1}^{N} X_{1\,i}X_{2\,i}y_i, \\ b_0\sum\limits_{i=1}^{N} X_{2\,i}^2 + b_{1,\,2}\sum\limits_{i=1}^{N} X_{1\,i}X_{2\,i}^3 + b_{2,\,2}\sum\limits_{i=1}^{N} X_{2\,i}^4 = \sum\limits_{i=1}^{N} X_{2\,i}^2y_i. \end{cases}$$

Для вычисления нужных сумм составим таблицу (табл. 12).

Таблица 12

Номер опыта	$X_{1\,i}$	$X_{2\,i}$	y_i	$X_{1\,i}X_{2\,i}$	$X_{2\,i}^2$	$X_{1\,i}^2X_{2\,i}^2$	$X_{1\,i}X_{2\,i}^3$	$X_{1\,i}X_{2\,i}y_i$	$X_{2\,i}^4$	$X_{2\,i}^2y_i$
1	1	0	6	0	0	0	0	0	0	0
2	0	2	1	0	4	0	0	0	16	4
3	1	–1	8	–1	1	1	–1	–8	1	8
4	0	1	2	0	1	0	0	0	1	2
Σ	2	2	17	–1	6	1	–1	–8	18	14

Запишем систему нормальных уравнений:

$$\begin{cases} 4b_0 - b_{1,\,2} + 6b_{2,\,2} = 17, \\ -b_0 + b_{1,\,2} - b_{2,\,2} = -8, \\ 6b_0 - b_{1,\,2} + 18b_{2,\,2} = 14. \end{cases}$$

Решив систему, получим $b_0 \approx 4{,}73$; $b_{1,\,2} \approx -4{,}31$; $b_{2,\,2} \approx -1{,}04$.

Итак, искомая зависимость имеет вид $\hat{y} = 4{,}73 - 4{,}31X_1X_2 - 1{,}04X_2^2$.

3. Квадратичное уравнение регрессии в случае двух факторов имеет вид $\hat{y} = b_0 + b_1X_1 + b_2X_2 + b_{1,\,2}X_2 + b_{1,\,1}X_1^2 + b_{2,\,2}X_2^2$, т. е. содержит 6 неизвестных коэффициентов. Для их определения необходимо не менее 6 опытов. По условию даны результаты всего четырех опытов. Таким образом, по имеющимся данным получить квадратичное уравнение невозможно.

4. Уравнение $\hat{y} = b_0 + b_1X_1 + b_2X_2 + b_{1,\,2}X_2 + b_{1,\,1}X_1^2$ содержит 4 коэффициента, однако по результатам данного эксперимента рассчитать значения этих коэффициентов невозможно, поскольку переменные $X_0 = 1$, X_1, X_2, $X_3 = X_1^2$ линейно зависимы (значения X_1 и $X_3 = X_1^2$ во всех опытах совпадают).

Задачи для самостоятельного решения

Задача 39. По результатам эксперимента (табл. 13) получить (если это возможно):

1) линейное уравнение регрессии;

2) уравнение вида $\hat{y} = b_0 + b_{1,\,2}X_1X_2 + b_{2,\,2}X_2^2$;

3) квадратичное уравнение регрессии;

4) уравнение $\hat{y} = b_0 + b_1X_1 + b_2X_2 + b_{1,\,1}X_1^2$.

Таблица 13

Номер опыта	$X_{1\,i}$	$X_{2\,i}$	y_i
1	–1	0	2
2	1	0	3
3	0	–1	4
4	0	–1	5

Задача 40.

По результатам эксперимента (табл. 14) получить (если это возможно): линейное уравнение регрессии.

Таблица 14

$X_{1\,i}$	–1	1	1	–1	–1	–1	1	1
$X_{2\,i}$	–1	–1	–1	1	1	1	1	1
y_i	1	2	3	4	6	8	10	12

Ответы к задачам

39. 1) $\hat{y} = 3{,}5 + 0{,}5X_1 + 0{,}5X_2$; 2) уравнение получить невозможно, так как переменная $X_3 = X_1X_2$ принимает одни и те же значения; 3) невозможно, так как количество опытов меньше количества неизвестных коэффициентов; 4) $\hat{y} = 4{,}5 + 0{,}5X_1 + 0{,}5X_2 - 2X_1^2$; **40.** $\hat{y} = 4{,}875 + 1{,}875X_1 + 3{,}5X_2$.

ГЛАВА 3. ПЛАНИРОВАНИЕ ЭКСПЕРИМЕНТА

3.1. Цель и этапы эксперимента

Цель планирования эксперимента — получить большее количество информации при меньших затратах. Часто экспериментаторы стремятся лишь к тому, чтобы собранные данные подогнать под какую-либо компактную формулировку. Например, требуется проверить однородность выборки путем сравнения дисперсий по t-критерию или получить уравнение регрессии. При этом остается нераскрытым вопрос, каким образом были собраны данные.

Основными этапами исследовательской работы являются планирование, эксперимент и анализ. Эксперимент включает в себя постановку задачи, которая будет решаться. На этом этапе выбирается независимая переменная (или несколько переменных). Решается вопрос, насколько точно можно измерить исследуемые величины. Устанавливается, какие независимые переменные (факторы) могут влиять на зависимую величину или отклик и каким распределением они могут быть описаны. На этапе планирования определяется размер выборки, порядок рандомизации, строится математическая модель для описания эксперимента. Окончательным этапом является анализ. К нему относятся процессы сбора и упорядочения данных вычисления некоторых статистик, необходимых для принятия решений, а также выбор соответствующих правил принятия решений для проверки математических гипотез относительно математической модели [1].

Этот эксперимент состоит:

1) из постановки задачи;
2) выбора отклика (зависимой переменной);
3) выбора варьируемых факторов;
4) выбора уровней факторов.

Этап планирования определяет:

1) необходимое число наблюдений;
2) порядок проведения эксперимента;
3) используемый метод рандомизации;
4) математическую модель для описания эксперимента.

Этап анализа — это:

1) сбор и обработка данных;
2) вычисление статистик для проверки гипотез;
3) интерпретация результатов.

3.2. Выбор факторов

При выборе области эксперимента прежде всего надо оценить границы областей определения факторов. При этом должны учитываться ограничения нескольких типов. Первый тип — принципиальные ограничения для значения факторов, которые не могут быть нарушены ни при каких обстоятельствах. Например, если фактор — температура, то нижним пределом будет абсолютный ноль. Второй тип — ограничения, связанные с технико-экономическими соображениями, например, со стоимостью сырья, дефицитностью отдельных компонентов, временем ведения процесса. Третий тип ограничений, с которым чаще всего приходится иметь дело, определяется конкретными условиями проведения процесса, например, существующей аппаратурой, технологией, организацией. В реакторе, изготовленном из некоторого материала, температуру нельзя поднять выше температуры имеющегося катализатора. Область совместного существования факторов приведена на рис. 12.

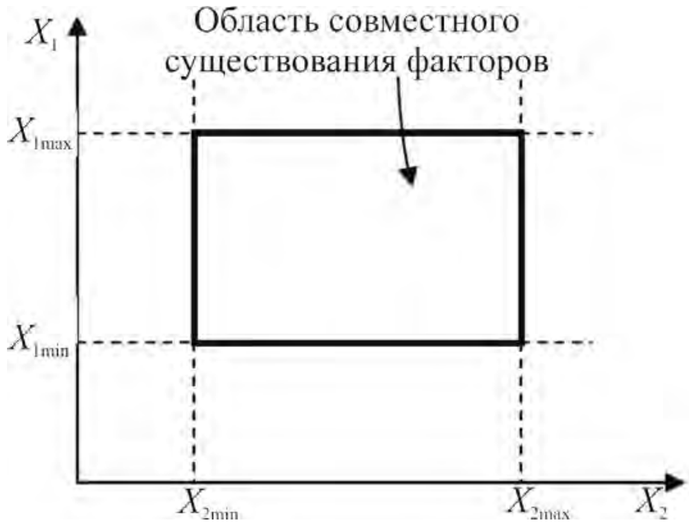


Рис. 12

Информацию о границах областей определения факторов будем называть **априорной**.

П р и м е р. Требуется провести ионообменное разделение смесей группы редкоземельных элементов растворами иминодиуксусной кислоты. Параметр оптимизации — содержание неодима в выходном растворе (элюате) в процентах. Рассматривается всего два фактора: концентрация элюанта (входного раствора), % от веса ($\tilde{x}_1$), и pH элюанта ($\tilde{x}_2$). Строим область определения факторов. Начнем с $\tilde{x}_1$. Известно, что при $\tilde{x}_1 > 3$ работать нельзя, так как это является пределом растворимости данного вещества при нормальной температуре. Значит, верхний предел $\tilde{x}_1 = 3$. При $\tilde{x}_1 = 0{,}5$ pH < 3, т. е. кислота находится в недиссоциированном состоянии, а при pH > 8 оба соединения разрушаются. Следовательно, $\tilde{x}_2$ может изменяться от 3 до 8 [4].

3.3. Выбор основного уровня и интервалов варьирования

Наилучшим условиям, определенным при помощи анализа априорной информации, соответствует комбинация (или несколько комбинаций) уровней факторов. Каждая комбинация является многомерной точкой в факторном пространстве, которую можно рассматривать как исходную точку для построения плана эксперимента. Назовем ее основным (нулевым) уровнем. Построение плана эксперимента сводится к выбору экспериментальных точек, симметричных относительно нулевого уровня.

В разных случаях мы располагаем различными сведениями об области наилучших условий. Если имеются сведения о координатах одной наилучшей точки и нет информации о границах определения факторов, то остается рассматривать эту точку в качестве основного уровня. Аналогичное решение принимается, если границы известны и наилучшие условия лежат внутри области.

Положение усложняется, если эта точка лежит на границе области или весьма близко к границе. Тогда основной уровень нужно выбирать с некоторым сдвигом от наилучших условий.

Когда отсутствуют дополнительные данные (технологического, экономического характера и т. д.), выбор произволен.

Пусть из априорной информации отмечены наилучшие условия. Известно также, что есть возможность дальнейшего улучшения параметра оптимизации, а данное значение нас не удовлетворяет. Эту точку нельзя рассматривать в качестве основного уровня, так как она расположена на границе области определения. Но требование симметрии экспериментальных точек относительно нулевого уровня привело бы в этом случае к выходу за границы области определения, чего делать также нельзя. В общем случае выбор основного уровня выполняется в соответствии с блок-схемой, приведенной на рис. 13.


Выбор основного уровня
Известна наилучшая точка и область определения
Известна наилучшая точка
Известно несколько наилучших точек
Известна подобласть, в которой процесс протекает достаточно хорошо
Точка лежит внутри области
Точка лежит на границе области
Имеются специальные соображения по выбору одной из точек
Ни одной из точек нельзя отдать предпочтения
Точка принимается за основной уровень
Нулевой уровень перемещается внутрь области
Выбирается наилучшая точка
Выбирается случайная точка
Ставится несколько планов для разных точек
Выбирается центр подобласти
Выбирается случайная точка в подобласти

Рис. 13

Теперь цель состоит в том, чтобы для каждого фактора выбрать два уровня, на которых он будет варьироваться в эксперименте.

Интервалом варьирования факторов называется некоторое число (свое для каждого фактора), прибавление которого к основному уровню дает верхний, а вычитание — нижний уровни фактора. Другими словами, интервал варьирования — это расстояние на координатной оси между основным и верхним (или нижним) уровнем. Графическое представление кодирования факторов приведено на рис. 14.

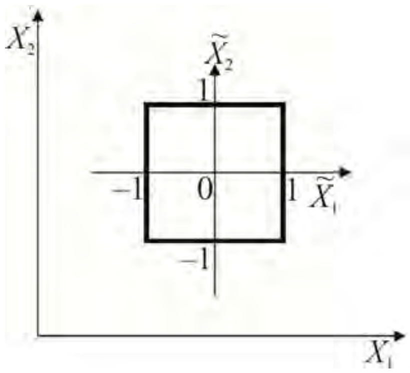


Рис. 14

Таким образом, задача выбора уровней сводится к более простой задаче выбора интервала варьирования. Для упрощения записи условий эксперимента и обработки экспериментальных данных масштабы по осям выбираются так, чтобы верхний уровень соответствовал +1, нижний — –1, а основной — нулю. Для факторов с непрерывной областью определения это всегда можно сделать с помощью преобразования

$$x_j = \frac{\tilde{x}_j - \tilde{x}_{j\,0}}{I_j}, \tag{28}$$

где x_j — кодированное значение фактора; $\tilde{x}_j$ — натуральное значение фактора; $\tilde{x}_{j\,0}$ — натуральное значение основного уровня; I_j — интервал варьирования; 0 — номер фактора.

Для качественных факторов, имеющих два уровня, один уровень обозначается +1, а другой –1, порядок уровней не имеет значения.

Пусть прогресс определяется четырьмя факторами. Основной уровень и интервалы варьирования приведены в табл. 15.

Таблица 15

Факторы	$\tilde{x}_1$	$\tilde{x}_2$	$\tilde{x}_3$	$\tilde{x}_4$
Основной уровень	3	30	1,5	15
Интервал варьирования	2	10	1	10

Остановимся на первом факторе. Отметим на координатной оси три уровня: нижний, основной и верхний. Нужно найти кодированное значение для $\tilde{x} = 2,0.$ Это значение лежит между 1,0 и 3,0, т. е. между –1 и 0 в кодированном масштабе. Так как в натуральном масштабе 2,0 лежит посередине, то ему соответствует –0,5 в кодированном масштабе (для x_1 = 2,5 будет x_1 = –0,25, для x_1 = 1,5 будет x_1 = –0,75 и т. д.) (см. ниже).

Натуральные значения	1	2	3	5	$\tilde{x}_1$
Кодированные значения	–1	X	0	1	x_1

На выбор интервалов варьирования накладываются естественные ограничения сверху и снизу. Интервал варьирования не может быть меньше той ошибки, с которой экспериментатор фиксирует уровень фактора. Иначе верхний и нижний уровни окажутся неразличимыми. С другой стороны, интервал не может быть настолько большим, чтобы верхний или нижний уровни оказались за пределами области определения. Внутри этих ограничений обычно еще остается значительная неопределенность выбора, которая устраняется с помощью интуитивных решений.

Размер интервала варьирования составляет некоторую долю от области определения фактора. Можно, например, условиться о следующем: если интервал составляет не более 10 % от области определения, то считать его узкими, не более 30 % — средним, в остальных случаях — широким.

Продолжим рассмотрение примера из п. 3.2. Область определения факторов была выбрана следующим образом: для $\tilde{x}_1$ — 0,5…3, для $\tilde{x}_2$ — 3…8. Основной уровень: $\tilde{x}_1 = 1,5;$ $\tilde{x}_2 = 0,7.$

Экспериментатор имел такую априорную информацию: точность фиксирования факторов средняя, поверхность отклика линейная, диапазон изменения параметра оптимизации довольно узок. Принимаемое решение — широкий интервал варьирования. Экспериментатор выбрал такие интервалы: $I_1 = 0,5;$ $I_2 = 1,0,$ что составляет 20 % от области определения факторов.

3.4. Пример решения задачи (матрица эксперимента)

Задача 41. Исходные параметры технологического процесса составляют:
толщина пленки 55 мкм;
время экспозиции 30 с.
Составить матрицу эксперимента и получить линейную модель процесса.
Решение:

$$x_{1\,c} = 55; \;\; x_{2\,c} = 30.$$

Возьмем верхние и нижние значения обоих факторов так, чтобы они располагались симметрично относительно текущего значения, например:

$$x_{1\,\text{в}} = 60; \quad x_{2\,\text{в}} = 35;$$
$$x_{1\,\text{н}} = 50. \quad x_{2\,\text{н}} = 25.$$

Расположение экспериментальных точек в двумерном факторном пространстве приведено на рис. 15.

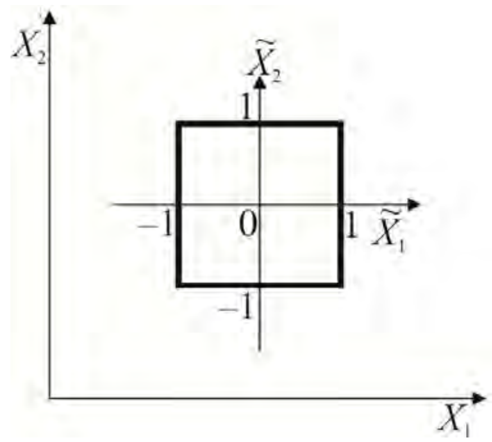


Рис. 15

Составим таблицу, в которой значения обоих факторов находятся во всех возможных сочетаниях, и проведем измерения в этих точках (значения отклика приводим условно) (табл. 16).

Таблица 16

x_1	x_2	y
50	25	140
50	35	210
60	25	170
60	35	220

Полагая, что линейная модель процесса имеет вид $y = a_0 + a_1x_1 + a_2x_2$, на основании полученных результатов составляем систему четырех уравнений с двумя переменными. Ниже показана эта система, а также ее сокращенная запись в виде матрицы. Матрицу данного вида назовем матрицей эксперимента.

$$\begin{cases} a_0 + 50a_1 + 25a_2 = 140, \\ a_0 + 50a_1 + 35a_2 = 210, \\ a_0 + 60a_1 + 25a_2 = 170, \\ a_0 + 60a_1 + 35a_2 = 220. \end{cases} \qquad \begin{pmatrix} x_0 & x_1 & x_2 & y \\ 1 & 50 & 25 & 140 \\ 1 & 50 & 35 & 210 \\ 1 & 60 & 25 & 170 \\ 1 & 60 & 35 & 220 \end{pmatrix}.$$

В матрице эксперимента второй и третий столбцы представляют собой значения факторов, четвертый столбец — значения отклика системы, а первый содержит единицы, соответствующие единичным коэффициентам свободного члена модели a_0. Будем считать этот столбец некоторым виртуальным фактором x_0, который всегда принимает единичные значения.

Решим систему, переходя к нормированным координатам. Чтобы облегчить решение системы, проведем нормировку факторов. Верхним значениям факторов присвоим нормированное значение +1, нижним значениям — нормированное значение –1, среднему значению – нормированное 0. В общем виде нормировка фактора выражается формулой

$$\tilde{x} = \frac{2(x_i - x_{i\ \text{с}})}{x_{i\ \text{в}} - x_{i\ \text{н}}} = \frac{2x_i - x_{i\ \text{в}} - x_{i\ \text{н}}}{x_{i\ \text{в}} - x_{i\ \text{н}}} = \frac{x_i - x_{i\ \text{с}}}{\Delta}, \tag{29}$$

где $\Delta = \dfrac{x_{i\ \text{в}} - x_{i\ \text{н}}}{2}$.

С учетом нормировки факторов система уравнений и матрица эксперимента примут следующий вид:

$$\begin{cases} \tilde{a}_0 - \tilde{a}_1 - \tilde{a}_2 = 140, \\ \tilde{a}_0 - \tilde{a}_1 + \tilde{a}_1 = 210, \\ \tilde{a}_0 + \tilde{a}_1 - \tilde{a}_1 = 170, \\ \tilde{a}_0 + \tilde{a}_1 + \tilde{a}_1 = 220. \end{cases} \quad \begin{pmatrix} \tilde{x}_0 & \tilde{x}_1 & \tilde{x}_2 & y \\ +1 & -1 & -1 & 140 \\ +1 & -1 & +1 & 210 \\ +1 & +1 & -1 & 170 \\ +1 & +1 & +1 & 220 \end{pmatrix}.$$

Поскольку сумма членов во втором и третьем столбце матрицы равны нулю, свободный член модели можно найти, сложив все четыре уравнения:

$$4\tilde{a}_0 = 140 + 210 + 170 + 220 = 740;$$

$$\tilde{a}_0 = 185.$$

Чтобы найти какой-либо другой коэффициент модели, нужно изменить знаки в уравнениях таким образом, чтобы в соответствующем столбце оказались одни единицы, после чего сложить все четыре уравнения.

В общем случае решение системы будет выглядеть как

$$\tilde{a}_k = \frac{1}{2^n} \sum_{i=1}^{2^n} y_i \tilde{x}_{k,\ i}. \tag{30}$$

Тогда остальные параметры будут равны

$$4\tilde{a}_1 = -140 - 210 + 170 + 220 = 40;$$

$$\tilde{a}_1 = 10.$$

$$4\tilde{a}_2 = -140 + 210 - 170 + 220 = 120;$$

$$\tilde{a}_2 = 30.$$

Таким образом, линейная модель технологического процесса в окрестностях точки (55, 30) имеет вид

$$y = 185 + 10\tilde{x}_1 + 30\tilde{x}_2.$$

Возвращаемся к ненормированным факторам. Переход от нормированных к ненормированным факторам осуществляется обратным преобразованием:

$$x_i = \tilde{x}_i \Delta + x_{i\ \text{с}} = \tilde{x}_i \Delta + \frac{x_{i\ \text{в}} + x_{i\ \text{н}}}{2}. \qquad (31)$$

Чтобы найти параметры модели для ненормированных координат, подставим выражения для нормированных координат в уравнение модели:

$$y = \tilde{a}_0 + \tilde{a}_1 \tilde{x}_1 + \tilde{a}_2 \tilde{x}_2 = \tilde{a}_0 + \tilde{a}_1 \frac{2(x_1 - x_{1\text{с}})}{x_{1\ \text{в}} - x_{1\ \text{н}}} + \tilde{a}_1 \frac{2(x_2 - x_{2\ \text{с}})}{x_{2\ \text{в}} - x_{2\ \text{н}}} =$$

$$= \tilde{a}_0 - \tilde{a}_1 \frac{2x_{1\ \text{с}}}{x_{1\ \text{в}} - x_{1\ \text{н}}} - \tilde{a}_2 \frac{2x_{2\ \text{с}}}{x_{2\ \text{в}} - x_{2\ \text{н}}} + \frac{2\tilde{a}_1}{x_{1\ \text{в}} - x_{1\ \text{н}}} x_1 + \frac{2\tilde{a}_2}{x_{2\ \text{в}} - x_{2\ \text{н}}} =$$

$$= \tilde{a}_0 - \tilde{a}_1 \frac{x_{1\ \text{в}} + x_{1\ \text{н}}}{x_{1\ \text{в}} - x_{1\ \text{н}}} - \tilde{a}_2 \frac{x_{2\ \text{н}} + x_{2\ \text{н}}}{x_{2\ \text{н}} - x_{2\ \text{н}}} + \frac{2\tilde{a}_1}{x_{1\ \text{в}} - x_{1\ \text{н}}} x_1 + \frac{2\tilde{a}_2}{x_{2\ \text{в}} - x_{2\ \text{н}}} x_2.$$

Сравнивая последнее выражение с выражением для линейной модели в ненормированных координатах $y = a_0 + a_1 x_1 + a_2 x_2$, получим выражение для параметров модели:

$$a_0 = \tilde{a}_0 - \tilde{a}_1 \frac{x_{1\ \text{в}} + x_{1\ \text{н}}}{x_{1\ \text{в}} - x_{1\ \text{н}}} - \tilde{a}_2 \frac{x_{2\ \text{в}} + x_{2\ \text{н}}}{x_{2\ \text{в}} - x_{2\ \text{н}}};$$

$$a_1 = \frac{2\tilde{a}_1}{x_{2\ \text{в}} - x_{1\ \text{н}}};$$

$$a_2 = \frac{2\tilde{a}_2}{x_{2\ \text{в}} - x_{2\ \text{н}}}.$$

В общем случае:

$$a_0 = \tilde{a}_0 - \sum_{i=1}^{n} \tilde{a}_i \frac{x_{i\ \text{в}} + x_{i\ \text{н}}}{x_{i\ \text{в}} - x_{i\ \text{н}}}; \qquad (32)$$

$$a_k = \frac{2\tilde{a}_k}{x_{k\ \text{в}} - x_{k\ \text{н}}}. \qquad (33)$$

Для приведенного выше примера:

$$a_0 = 185 - 10\frac{60+50}{60-50} - 30\frac{35+25}{35-25} = -105;$$

$$a_1 = \frac{2 \cdot 10}{60-50} = 2;$$

$$a_2 = \frac{2 \cdot 30}{35-25} = 6.$$

Окончательно получаем модель в естественных координатах:

$$y = -105 + 2x_1 + 6x_2.$$

3.5. Полный факторный эксперимент типа 2^k

3.5.1. Матрица полного факторного эксперимента в общем виде

В общем виде матрица полного факторного эксперимента с n факторами имеет вид

$$\begin{pmatrix} \tilde{x}_0 & \tilde{x}_1 & \dots & \tilde{x}_n & y \\ +1 & -1 & \dots & -1 & y_1 \\ +1 & +1 & \dots & -1 & y_2 \\ \dots & \dots & \dots & \dots & \dots \\ +1 & +1 & \dots & +1 & y_{2^n} \end{pmatrix}.$$

Первый этап планирования эксперимента для получения линейной модели основан на варьировании факторов на двух уровнях. В этом случае, если число факторов известно, можно сразу найти число опытов, необходимое для реализации всех возможных сочетаний уровней факторов. Простая формула, которая для этого используется, уже приводилась в п. 1, и мы ее напомним: $N = 2^k$, где N — число опытов; k — число факторов; 2 — число уровней. В общем случае эксперимент, в котором реализуются все возможные сочетания уровней факторов, называется **полным факторным экспериментом**. Если число уровней каждого фактора равно двум, то имеем **полный факторный эксперимент типа 2^k**.

3.5.2. Матрица планирования эксперимента 2^2

Записываем все сочетания уровней в эксперименте с двумя факторами. Напомним, что в планировании эксперимента используются кодированные значения факторов: +1 и –1 (часто для простоты записи единицы опускают). Условия эксперимента можно оформить в виде таблицы, где строки соответствуют различным опытам, а столбцы — значениям факторов. Будем называть такие таблицы **матрицами планирования эксперимента**.

Матрица планирования для двух факторов приведена в табл. 17. Каждый столбец в матрице планирования называют вектор-столбцом, а каждую строку — вектор-строкой. Таким образом, в табл. 17 мы имеем два вектора-столбца независимых переменных и один вектор-столбец параметра оптимизации.

Таблица 17

Номер опыта	x_1	x_2	y	Номер опыта	x_1	x_2	y
1	–1	–1	y_1	3	–1	+1	y_1
2	+1	–1	y_2	4	+1	+1	y_2

То, что записано в этой таблице в алгебраической форме, можно изобразить геометрически.

Номера вершин квадрата соответствуют номерам опытов в матрице планирования. Площадь, ограниченная квадратом, называется **областью эксперимента**. Иногда удобнее считать областью эксперимента площадь, ограниченную окружностью, описывающей квадрат. В задачах интерполяции область эксперимента есть область предсказываемых значений y.

Запись матрицы планирования, особенно для многих факторов, громоздка. Для ее сокращения удобно ввести условные буквенные обозначения строк.

Это делается следующим образом. Порядковый номер фактора ставится в соответствие строчной букве латинского алфавита: $x_1 - a$, $x_2 - b$ и т. д. Если для строки матрицы планирования выписать латинские буквы только для факторов, находящихся на верхних уровнях, то условия опыта будут заданы однозначно.

3.5.3. Проведение эксперимента

Без ограничения общности можно считать, что кодированные значения x_i принимают значения –1 и +1 соответственно (принято обозначать – или +). Множество всех точек в k-мерном пространстве, координаты которых являются комбинациями + и –, называется **полным факторным планом или планом полного факторного эксперимента** типа 2^k (ПФЭ). Количество точек в этом плане $N = 2^k$.

Для примера возьмем полный факторный эксперимент с тремя независимыми переменными x_1, x_2 и x_3, (табл. 18).

Второй, третий и четвертый столбцы таблицы соответствуют собственно плану экспериментов, с пятого по восьмой столбцы содержат значения произведений независимых переменных. Фиктивная переменная $x_0 = 1$ (первый столбец) введена для единообразия записи расчетных формул коэффициентов полинома. Строки соответствуют опытам, например, первая строка характеризует эксперимент, в котором все независимые переменные находятся на нижнем уровне.

Таблица 18

Матрица планирования								Вектор результатов
x_0	x_1	x_2	x_3	x_1x_2	x_1x_3	x_2x_3	$x_1x_2x_3$	y
+	–	–	–	+	+	+	–	y_1
+	+	–	–	+	–	–	+	y_2
+	–	+	–	–	+	–	+	y_3
+	+	+	–	–	–	+	–	y_4
+	–	–	+	–	–	+	+	y_5
+	+	–	+	–	+	–	–	y_6
+	–	+	+	+	–	–	–	y_7
+	+	+	+	+	+	+	+	y_8

Существует несколько способов построения подобных матриц планирования. В частности, можно воспользоваться приемом, характерным для записи последовательности двоичных чисел. В столбце последней переменной x_3 знаки меняются поочередно, в столбце предпоследней переменной x_2 — чередуются через два элемента, третьей справа переменной x_1 — через четыре элемента. Аналогично строится матрица для любого количества переменных, порядок перечисления переменных не играет роли. Столбцы с произведениями переменных вычисляются путем умножения значений элементов в соответствующих столбцах простых переменных.

Из анализа матрицы планирования легко видеть, что полный факторный эксперимент обладает свойствами:

ортогональности. Сумма парных произведений элементов любых двух различных столбцов равна нулю. В частности, для простых переменных $\sum_{u=1}^{N} x_{i,\,u} x_{j,\,u} = 0;\ i \neq j;\ \ i, j = \overline{0,k};$

симметричности. Сумма всех элементов любого столбца, за исключением первого, равна нулю. Например, $\sum_{u=1}^{N} x_{i,\,u} = 0;\ i = \overline{1,k};$

нормированности. Сумма квадратов элементов любого столбца равна числу опытов. Так, для i-й переменной $\sum_{u=1}^{N} x_{i,\,u}^2 = N;\ i = \overline{0,k}$.

Первые два свойства обеспечивают независимость оценок коэффициентов модели и допустимость их физической интерпретации. Нарушение этих свойств приводит к взаимной зависимости оценок и невозможности придания смысла коэффициентам.

Включение в матрицу планирования переменных вида x_i^2 приведет к появлению единичных столбцов, совпадающих друг с другом и со столбцом x_0. Следовательно, нельзя будет определить, за счет чего получено значение. Поэтому планы ПФЭ 2^k не применимы для построения функции отклика в виде полного полинома второй степени.

Примеры решения задач

Задача 42. Проводится ПФЭ 2^3 для установления зависимости отклика от указанных факторов (табл. 19). Требуется получить линейное уравнение регрессии.

Таблица 19

X_1	60	40	60	40	60	40	60	40
X_2	20	20	80	80	20	20	80	80
X_3	0	0	0	0	10	10	10	10
y	2	3	4	5	4	5	8	7

Решение. Проведено N различных опытов. Имеем ПФЭ 2^3 в натуральных переменных без дублирования опытов. Определим центр плана и интервалы варьирования факторов.

$$X_1^{\max} = 60;\ X_1^{\min} = 40;\ \ X_1^0 = \frac{60+40}{2} = 50;\ \Delta X_1^0 = \frac{60-40}{2} = 10.$$

$$X_2^{\max} = 80;\ X_2^{\min} = 20;\ \ X_2^0 = \frac{80+20}{2} = 50;\ \Delta X_2^0 = \frac{80-40}{2} = 20.$$

$$X_3^{\max} = 10;\ X_3^{\min} = 0;\ \ X_3^0 = \frac{10+0}{2} = 5;\ \Delta X_3^0 = \frac{10-0}{2} = 5.$$

Переходим от натуральных переменных к кодированным.

$$x_1 = \frac{X_1 - 50}{10};\ x_2 = \frac{X_2 - 50}{30};\ x_3 = \frac{X_3 - 5}{5}.$$

Запишем результаты эксперимента в табл. 20.

Таблица 20

Номер опыта	x_0	x_1	x_2	x_3	y
1	+	+	–	–	2
2	+	–	–	–	3
3	+	+	+	–	4
4	+	–	+	–	5
5	+	+	–	+	4
6	+	–	–	+	5
7	+	+	+	+	8
8	+	–	+	+	7

Определим коэффициенты линейного уравнения регрессии в кодированных переменных $\hat{y} = b_0 + b_1x_1 + b_2x_2 + b_3x_3$ по формуле (30).

$$b_0 = \frac{1}{8}(2+3+4+5+4+5+8+7);$$

$$b_1 = \frac{1}{8}(2-3+4-5+4-5+8-7);$$

$$b_2 = \frac{1}{8}(-2-3+4+5-4-5+8+7);$$

$$b_3 = \frac{1}{8}(-2-3-4-5+4+5+8+7).$$

Итак, линейное уравнение регрессии в кодированных переменных принимает вид

$$\hat{y} = 4,75 - 0,25x_1 + 1,25x_2 + 1,25x_3.$$

В натуральных переменных

$$\hat{y} = 4,75 - 0,25\frac{X_1 - 50}{10} + 1,25\frac{X_2 - 50}{30} + 1,25\frac{X_3 - 5}{5}.$$

Задача 43.

Получить линейное уравнение регрессии по данным результатам эксперимента (табл. 21).

Таблица 21

x_1	–1	+1	–1	+1	–1	+1
x_2	–1	–1	+1	+1	+1	–1
y	2	3	4	5	5	2

Решение. Эксперимент двухфакторный. Проведено 6 опытов, из них 4 различных. Это результаты ПФЭ типа 2^2 в кодированных переменных с неравномерным дублированием опытов. В этом случае составляем систему нормальных уравнений. Для линейного уравнения регрессии с двумя факторами $\hat{y} = b_0 + b_1x_1 + b_2x_2$ эта система имеет вид

$$\begin{cases} Nb_0 + b_1\sum_{i=1}^{N} x_{1\,i} + b_2\sum_{i=1}^{N} x_{2\,i} = \sum_{i=1}^{N} y_i, \\ b_0\sum_{i=1}^{N} x_{1\,i} + b_1\sum_{i=1}^{N} x_{1\,i}^2 + b_2\sum_{i=1}^{N} x_{1\,i}x_{2\,i} = \sum_{i=1}^{N} x_{1\,i}y_i, \\ b_0\sum_{i=1}^{N} x_{2\,i} + b_1\sum_{i=1}^{N} x_{1\,i}x_{2\,i} + b_2\sum_{i=1}^{N} x_{2\,i}^2 = \sum_{i=1}^{N} x_{2\,i}y_i. \end{cases}$$

Для вычисления необходимых сумм составим таблицу, в которой будут учтены все 6 опытов (табл. 22).

Таблица 22

Номер опыта	$x_{1\,i}$	$x_{2\,i}$	y_i	$x_{1\,i}^2$	$x_{1\,i}x_{2\,i}$	$x_{1\,i}y_i$	$x_{2\,i}^2$	$x_{2\,i}y_i$
1	–1	–1	2	1	+1	–2	1	–2
2	+1	–1	3	1	–1	3	1	–3
3	–1	+1	4	1	–1	–4	1	4
4	+1	+1	5	1	+1	5	1	5
5	–1	+1	5	1	–1	-5	1	5
6	+1	–1	2	1	–1	2	1	–2
Σ	0	0	21	6	–2	–1	6	7

Запишем систему нормальных уравнений

$$\begin{cases} 6b_0 = 21, \\ 6b_1 - 2b_2 = -1, \\ -2b_1 + 6b_2 = 7. \end{cases} \Rightarrow \begin{cases} b_0 = 3{,}5, \\ b_1 = 0{,}25, \\ b_2 = 1{,}25. \end{cases}$$

Следовательно, линейное уравнение регрессии имеет вид

$$\hat{y} = 3{,}5 + 0{,}25x_1 + 1{,}25x_2.$$

Задачи для самостоятельного решения

Задача 44. Проводится ПФЭ 2^3 для установления зависимости отклика от указанных факторов (табл. 23). Требуется получить линейное уравнение регрессии.

Таблица 23

x_1	60	40	60	40	60	40	60	40
x_2	20	20	80	80	20	20	80	80
x_3	20	20	20	20	10	10	10	10
y	2	3	4	5	4	5	8	7

В задачах 45, 46, 47 требуется:

1) определить, являются ли приведенные данные результатами ПФЭ в кодированных или натуральных переменных (табл. 24, 25, 26);

2) получить линейное уравнение регрессии в кодированных и натуральных переменных.

Задача 45

Таблица 24

x_1	–1	+1	–1	+1
x_2	–1	+1	+1	–1
y	2	3	4	5

Задача 46

Таблица 25

x_1	20	20	40	40	20	20	40	40
x_2	10	30	10	30	10	30	10	30
y	4	6	8	10	4	4	8	8

Задача 47

Таблица 26

x_1	20	40	20	40	20	40
x_2	0	0	0	20	20	20
y	2	3	4	5	6	7

Ответы к задачам

44. ПФЭ 2^2 в кодированных переменных $\hat{y} = 4,75 - 0,25x_1 + 1,25x_3$. **45.** 1) ПФЭ 2^2 в кодированных переменных без дублирования опытов; 2) $y = 3,5 + 0,5x_1$. **46**. 1) ПФЭ 2^2 в натуральных переменных с равномерным дублированием опытов, каждый опыт повторяется m = 2 раза; 2) $\hat{y} = 6,5 + +2x_1 + 0,5x_2$; $\hat{y} = 6,5 + 2 \cdot \dfrac{x_1 - 30}{10} + 0,5 \cdot \dfrac{x_2 - 20}{10}$. **47.** 1) ПФЭ 2^2 в натуральных переменных с неравномерным дублированием опытов; для определения коэффициентов линейного уравнения регрессии необходимо составить систему нормальных уравнений, причем для упрощения расчетов следует перейти к кодированным переменным; 2) $\hat{y} = 4,5 + 1,5x_2$; $\hat{y} = 4,5 + 1,5\dfrac{x_2 - 10}{10}$.

3.6. Модели со взаимодействиями

В ПФЭ типа 2^k число различных опытов значительно превосходит число коэффициентов линейной модели: $N = 2^k > k + 1$, поэтому по результатам этого эксперимента можно получить более сложную модель.

Для того чтобы выяснить, какие члены можно включить в уравнение регрессии, вспомним требование линейной независимости столбцов матрицы базисных функций *X*. Если мы включаем в уравнение регрессии квадратичный член x_j^2, элементы соответствующего столбца $x_j^2 = (\pm 1)^2 = +1$ совпадают с элементами столбца для нулевого (фиктивного) фактора; элементы столбца $x_j^3 = x_j^2 x_j = x_j$ совпадают с элементами столбца для фактора x_j. Поэтому в уравнение можно включить только члены вида $x_j x_t$ (парные взаимодействия), $x_j x_t x_s$ (тройные взаимодействия) и т. д. Итак, ПФЭ типа 2^k позволяет получить не только линейную, но и более сложную модель со взаимодействиями, в которой учитывается ситуация, когда эффект влияния одного фактора зависит от того, на каком уровне поддерживаются другие факторы:

$$\hat{y} = b_0 + \sum_{j=1}^{k} b_j x_j + \sum_{j<t} b_{j,\,t} x_j x_t + \sum_{j<t<s} b_{j,\,t,\,s} x_j x_t x_s + \ldots \tag{34}$$

Так, по результатам ПФЭ 2^2, может быть определено уравнение с парным взаимодействием

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + b_{1,2} x_1 x_2. \tag{35}$$

Наиболее полная модель, которая может быть получена по результатам ПФЭ 2^3, включает парные и тройные взаимодействия, т. е.

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + b_{1,2} x_1 x_2 + b_{1,3} x_1 x_3 + b_{2,3} x_2 x_3 + b_{1,2,3} x_1 x_2 x_3. \tag{36}$$

Коэффициенты в кодированных переменных для экспериментов со взаимодействиями вычисляются по формулам:

$$b_j = \frac{1}{N} \sum_{i=1}^{N} x_{j,\,i} y_i; \quad b_{j,\,t} = \frac{1}{N} \sum_{i=1}^{N} x_{j,\,t} x_{t,\,i} y_i; \quad b_{j,\,t,\,s} = \frac{1}{N} \sum_{i=1}^{N} x_{j,\,t} x_{t,\,i} x_{s,\,i} y_i. \tag{37}$$

Замечание. В случае с равномерным дублированием для расчета коэффициентов подставляют средние значения отклика *y*. При неравномерном дублировании необходимо составлять систему нормальных уравнений.

Пример решения задачи

Задача 48. По результатам эксперимента задачи 42 получить уравнение регрессии с парными взаимодействиями.

Решение. Имеем результаты ПФЭ типа 2^3 в натуральных переменных без дублирования опытов.

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + b_{1,\,2} x_1 x_2 + b_{1,\,3} x_1 x_3 + b_{2,\,3} x_2 x_3.$$

Составим матрицу планирования (табл. 27).

Таблица 27

Номер опыта	x_0	x_1	x_2	x_3	$x_1 x_2$	$x_1 x_3$	$x_2 x_3$	y
1	+	+	–	–	–	–	+	2
2	+	–	–	–	+	+	+	3
3	+	+	+	–	+	–	–	4

Окончание табл. 27

Номер опыта	x_0	x_1	x_2	x_3	$x_1 x_2$	$x_1 x_3$	$x_2 x_3$	y
4	+	–	+	–	–	+	–	5
5	+	+	–	+	–	+	–	4
6	+	–	–	+	+	–	–	5
7	+	+	+	+	+	+	+	8
8	+	–	+	+	–	–	+	7
b_j	4,75	–0,25	1,25	1,25	0,25	0,25	0,25	—

Коэффициенты уравнения регрессии вычислим по формулам (37). Свободный член и линейные эффекты возьмем из задачи 42. Эффекты парных взаимодействий запишем как

$$b_{12} = \frac{1}{8}(-2+3+4-5-4+5+8-7) = 0,25;$$

$$b_{13} = \frac{1}{8}(-2+3-4+5+4-5+8-7) = 0,25;$$

$$b_{23} = \frac{1}{8}(2+3-4-5-4-5+8+7) = 0,25;$$

$$\hat{y} = 4,75 - 0,25x_1 + 1,25x_2 + 1,25x_3 + 0,25x_1x_2 + 0,25x_1x_3 + 0,25x_2x_3.$$

Можно перейти при помощи формул (33) к натуральным переменным.

Задачи для самостоятельного решения

Задача 49 По данным задачи 45 получить уравнение с парными взаимодействиями.

Задача 50. По данным задачи 46 получить уравнение с парными взаимодействиями.

Ответы к задачам

49. $\hat{y} = 3,5 + 0,5x_1 - x_1x_2$; **50.** $b_{1,\,2} = 0$. В кодированных переменных $\hat{y} = 6,5 + 2x_1 + 0,5x_2$; В натуральных переменных $\hat{y} = 6,5 + 2\dfrac{X_1 - 30}{10} + 0,5\dfrac{X_2 - 20}{10}$.

3.7. Расчет дисперсии воспроизводимости

Для статистического анализа уравнения регрессии необходимо иметь количественную оценку ошибок эксперимента в целом, т. е. общую для всех опытов дисперсию отклика y. Назовем ее **дисперсией воспроизводимости** и обозначим $s^2\{y\}$.

Для расчета дисперсии воспроизводимости необходимы повторные опыты, т. е. несколько наблюдений над величиной y, проведенных при неизменных значениях основных факторов. При этом если различные опыты повторяются по несколько раз, нужно рассчитать дисперсию для каждого опыта, а затем вычислить общую средневзвешенную дисперсию воспроизводимости. Объединять все наблюдаемые значения y в одну выборку нельзя, поскольку значения отклика в различных опытах могут существенно отличаться из-за разных условий проведения опытов, т. е. разных значений основных факторов.

Будем проводить дублирование опытов, т. е. каждый из N различных опытов проведем соответственно $k_1, k_2, \ldots, k_N$ раз. Обозначим $y_{i,\,n}$ наблюдаемое значение отклика при n-м повторении i-го опыта. Для каждого опыта нужно рассчитать

$$\bar{y}_i = \frac{1}{k_i}\sum_{n=1}^{k_i} y_{i,\,n};\;\; s^2\{y_i\} = \frac{1}{k_i - 1}\sum_{n=1}^{k_i}\left(y_{i,\,n} - \bar{y}_i\right)^2;\;\; f_i = k_i - 1. \qquad (38)$$

Здесь $\bar{y}_i$ — среднее значение отклика, наблюдаемое в i-м опыте (при i-м комплексе условий), дисперсия воспроизводимости $s^2\{y\}$ характеризует степень разброса значений отклика в i-м опыте, число ее степеней свободы f_i тем больше, чем больше наблюдений проведено при i-м комплексе условий.

Одним из условий применимости регрессионного анализа является требование воспроизводимости отклика во всех опытах с одинаковой точностью. Это означает, что дисперсии во всех опытах должны быть примерно одинаковы.

Для проверки воспроизводимости эксперимента проверяют гипотезу об однородности дисперсий в различных опытах:

$$H_0\colon \sigma_1^2 = \sigma_2^2 = \ldots = \sigma_N^2.$$

В случае равномерного дублирования опытов (все опыты повторяются одинаковое количество раз) эта проверка проводится по критерию Кочрена (12); в случае неравномерного дублирования — по критерию Бартлетта (при дополнительном условии, что объем каждой выборки больше 4). Можно также в обоих случаях использовать критерий Фишера, проверяя гипотезу об однородности наибольшей и наименьшей из сравниваемых дисперсий.

Замечание. Существенным недостатком критерия Фишера является игнорирование всех оценок дисперсии воспроизводимости, кроме максимального и минимального значения.

Если гипотеза об однородности дисперсий принимается, то говорят, что эксперимент **воспроизводим**. Это означает, что во всех опытах параметр y наблюдается с одинаковой точностью. Для количественной характеристики точности определения значений отклика рассчитывают дисперсию воспроизводимости как взвешенное среднее отдельных дисперсий, т. е.

$$s^2\{y\} = \frac{f_1 s^2\{y_1\} + f_2 s^2\{y_2\} + \ldots + f_N s^2\{yN\}}{f_1 + f_2 + \ldots + f_N} = \frac{\sum_{i=1}^{N} f_i s^2\{y_i\}}{\sum_{i=1}^{N} f_i}, \qquad (39)$$

а ее число степеней свободы равно

$$f_{\text{воспр}} = f_1 + f_2 + \ \dots \ + f_N. \tag{40}$$

Как правило, планируется равномерное дублирование опытов $k_1 = k_2 = k_3 = \dots = k_N = k.$ В этом случае формула (39) для расчета дисперсии воспроизводимости упрощается:

$$s^2\{y\} = \frac{1}{N}\sum_{i=1}^{N} s^2\{y_i\}; \quad f_{\text{воспр}} = N(k-1). \tag{41}$$

Замечание. В некоторых случаях дублирование опытов не проводится, а для оценки ошибок наблюдения отклика y ставится отдельная серия опытов при одних и тех же условиях (при одних и тех же значениях изучаемых факторов). В этом случае за дисперсию воспроизводимости принимается дисперсия, рассчитанная по полученной выборке с помощью формулы (1), ее число степеней свободы равно объему выборки минус 1.

Пример решения задачи

Задача 51. Оценить воспроизводимость эксперимента, результаты которого приведены в задаче 40.

Решение. Рассмотрим, при каких условиях (при каких значениях факторов X_1, X_2) проводились опыты в данном эксперименте. При $X_1 = -1$, $X_2 = -1$ опыт проводился один раз, получено значение $y = 1$; при $X_1 = 1$, $X_2 = -1$ проведено два опыта с результатами 2 и 3; опыт при условиях $X_1 = -1$, $X_2 = 1$ был повторен три раза и дал результаты 4, 6 и 8; при $X_1 = 1$, $X_2 = 1$ получены значения 10 и 12. Таким образом, имеется 4 различных опыта, которые продублированы неравномерно.

Для удобства расчета дисперсий отклика в каждом из опытов и проверки воспроизводимости эксперимента занесем все данные в таблицу. Четыре строки таблицы соответствуют четырем различным опытам. Результаты 1-го опыта, который не дублировался, при проверке воспроизводимости эксперимента и расчете дисперсии воспроизводимости учитываться не будут. В остальных опытах получаем:

$$\overline{y_2} = \frac{1}{2}(2+3) = 2{,}5;$$

$$s^2\{y_2\} = \frac{1}{2-1}\left(0{,}5^2 + 0{,}5^2\right) = 0{,}5; \quad f_2 = 2 - 1 = 1;$$

$$\overline{y_3} = \frac{1}{3}(4+6+8) = 6; \; s^2\{y_3\} = \frac{1}{3-1}\left(2^2 + 0^2 + 2^2\right) = 4; \quad f_3 = 3 - 1 = 2;$$

$$\overline{y_4} = \frac{1}{2}(10+12) = 11;$$

$$s^2\{y_4\} = \frac{1}{2-1}\left(1^2+1^2\right) = 11; \quad f_4 = 2-1 = 1.$$

Для проверки однородности полученных трех дисперсий применим критерий Фишера (11). Положим уровень значимости равным $\alpha = 0{,}05$. Поскольку $F_{\text{расч}} = \frac{s^2_{\max}}{s^2_{\min}} = \frac{4}{0{,}5} = 8 < F_{\text{табл}} = F_{0{,}025;\ 2;\ 1} = 799{,}5$, то эксперимент на уровне значимости $\alpha = 0{,}05$ следует признать воспроизводимым. Рассчитаем дисперсию воспроизводимости по формуле (39):

$$s^2\{y\} = \frac{1\cdot 0{,}5 + 2\cdot 4 + 1\cdot 2}{1+2+1} = \frac{10{,}5}{4} = 2{,}625.$$

Ее число степеней свободы равно $f_{\text{воспр}} = 1+2+1 = 4$.

Задачи для самостоятельного решения

В задачах требуется:

1) проверить гипотезу о воспроизводимости эксперимента при уровне значимости $\alpha = 0{,}05$;

2) вычислить дисперсию воспроизводимости и найти число степеней свободы (табл. 28—31).

Задача 52

Таблица 28

Номер опыта	$y_{i\,1}$	$y_{i\,2}$	$y_{i\,3}$	$y_{i\,4}$	$y_{i\,5}$
1	11	8	8	5	–
2	3	5	7	1	4
3	4	8	12	–	–

Задача 53

Таблица 29

Номер опыта	$y_{i\,1}$	$y_{i\,2}$	$y_{i\,3}$
1	15	17	22
2	10	12	14
3	12	16	14
4	18	22	14
5	10	6	14

Задача 54

Таблица 30

Номер опыта	$s^2\{y_i\}$	f_i
1	120	8
2	55	8
3	85	8
4	90	8
5	65	8

Задача 55

Таблица 31

Номер опыта	$s^2\{y_i\}$	f_i
1	230	20
2	140	12
3	180	10
4	212	8
5	80	15

Ответы к задачам

52. 1) эксперимент воспроизводим, $F_{\text{расч}} = 3,2 < F_{0,025;\ 2;\ 4} = 10,7$; 2) $s^2\{y\} = 7,78$, $f_{\text{воспр}} = 9$. **53.** 1) воспроизводимость эксперимента может быть проверена по критерию Кочрена: $G_{\text{расч}} = 0,3 < G_{0,05;\ 2;\ 5} = 0,6838$ или по критерию Фишера: $F_{\text{расч}} = 4 < F_{0,025;\ 2;\ 2} = 39$; эксперимент воспроизводим; 2) $s^2\{y\} = 10,6$, $f_{\text{воспр}} = 10$. **54.** 1) воспроизводимость эксперимента может быть проверена по критерию Кочрена: $G_{\text{расч}} = 0,29 < G_{0,05;\ 8;\ 5} = 0,4387$ или по критерию Фишера: $F_{\text{расч}} = 2,18 < F_{0,025;\ 8;\ 8} = 4,43$; эксперимент воспроизводим; 2) $s^2\{y\} = 83$, $f_{\text{воспр}} = 40$. **55.** 1) $F_{0,025;\ 20;\ 15} = 2,76$; эксперимент не воспроизводим, расчет дисперсии воспроизводимости нецелесообразен; 2) $s^2\{y\} = 168,87$, $f_{\text{воспр}} = 65$.

3.8. Проверка адекватности эмпирического уравнения регрессии

После определения по результатам эксперимента коэффициентов эмпирического уравнения регрессии нужно проверить соответствие полученного уравнения данным эксперимента. Цель этой проверки — выяснить, соответствует ли полученная регрессионная модель реальной функции отклика, описывающей данное явление; будет ли эмпирическое уравнение регрессии предсказывать значения отклика с той же точностью, что и результаты эксперимента.

Пусть по результатам N различных опытов, каждый из которых проведен соответственно $k_1, k_2, \ldots, k_N$ раз, получено уравнение регрессии

$$\hat{y} = b_0 X_0 + b_1 X_1 + \ldots + b_k X_k.$$

Обозначим $\hat{y}_i = b_0 X_{0\,i} + b_1 X_{1\,i} + \ \ldots\ + b_k X_{k\,i}$, значение отклика, предсказываемое по полученному уравнению регрессии для i-го опыта; $\bar{y}_i$ — среднее наблюдаемых значений отклика в i-м опыте.

Дисперсия адекватности характеризует расхождение между результатами эксперимента и значениями $\hat{y}$ и вычисляется по формуле

$$s_{\text{ад}}^2 = \frac{1}{N-d}\sum_{i=1}^{N} k_i(\hat{y}_i - \overline{y}_i)^2, \tag{42}$$

где $N - d = f_{\text{ад}}$ — число степеней свободы дисперсии адекватности; N — число различных опытов; d — число коэффициентов проверяемого уравнения регрессии, которые определялись по результатам эксперимента; k_i — число повторений i-го опыта.

Модель признается адекватной, т. е. удовлетворительно описывает исследуемую зависимость отклика от факторов, если расхождение между результатами эксперимента и значениями, полученными по уравнению регрессии, вызвано только ошибками эксперимента, а не связано, например, с неудачным выбором вида математической модели. Поэтому для проверки адекватности модели сравнивают дисперсию адекватности и дисперсию воспроизводимости, которая характеризует точность экспериментального определения значений отклика.

Проверка адекватности модели производится по критерию Фишера. Поскольку в случае неадекватной модели можно ожидать значительных расхождения между $\hat{y}_i$ и результатами эксперимента, то дисперсия адекватности, как правило, значительно больше, чем дисперсия воспроизводимости. Гипотеза об адекватности модели при заданном уровне значимости α принимается, если

$$F_{\text{расч}} = \frac{s_{\text{ад}}^2}{s^2\{y\}} < F_{\text{табл}} = F_{\alpha;\, f_{\text{ад}};\, f_{\text{воспр}}}. \tag{43}$$

В случае неадекватности модели принимают одно из следующих решений:

1) переход к более сложной модели;

2) уменьшение диапазона изменения факторов, т. е. сужение области исследования.

Замечание. Проверка адекватности возможна только для ненасыщенного плана, т. е. когда $N > d$.

Примеры решения задач

Задача 56. Для условия задачи 40 проверить адекватность линейного уравнения регрессии при уровне значимости $\alpha = 0,05$.

Решение. В условии задачи 40 даны результаты эксперимента, состоящего из 4 различных опытов, продублированных неравномерно. В задаче 41 оценена воспроизводимость этого эксперимента и рассчитана дисперсия воспроизводимости: $s^2\{y\} = 2,625$; $f_{\text{воспр}} = 4$.

В задаче 40 получено линейное уравнение регрессии:

$$\hat{y} = 4,875 + 1,875x_1 + 3,5x_2.$$

Рассчитаем по этому уравнению значения отклика для каждого опыта, подставляя соответствующие значения факторов:

$$\hat{y} = 4,875 + 1,875(-1) + 3,5(-1) = -0,5;$$

$$\hat{y} = 4,875 + 1,875(1) + 3,5(-1) = 3,25;$$

$$\hat{y} = 4,875 + 1,875(-1) + 3,5(1) = 6,5;$$

$$\hat{y} = 4,875 + 1,875(1) + 3,5(1) = 10,25.$$

Данные эксперимента и результаты расчетов удобно оформить в виде следующей таблицы, в которой количество строк равно количеству различных опытов (табл. 32).

Таблица 32

Номер опыта	$x_{1\,i}$	$x_{2\,i}$	$y_{i\,1}$	$y_{i\,2}$	$y_{i\,3}$	$\bar{y}_i$	$\hat{y}_i$	$(\hat{y}_i - \bar{y}_i)^2$	k_i
1	–1	–1	1	—	—	1	–0,50	2,2500	1
2	1	–1	2	3	—	2,5	3,25	0,5625	2
3	–1	1	4	6	8	6	6,50	0,2500	3
4	1	1	10	12	—	11	10,25	0,5625	2

Вычислим дисперсию адекватности по формуле (42):

$$s_{\text{ад}}^2 = \frac{1}{4-3}(2,25 \cdot 1 + 0,5625 \cdot 2 + 0,25 \cdot 3 + 0,5625 \cdot 2) = 5,25.$$

Ее число степеней свободы $f_{\text{ад}} = 4 - 3 = 1.$

Проверим гипотезу об адекватности модели по критерию Фишера. Поскольку

$$F_{\text{расч}} = \frac{5,25}{2,625} = 2 < F_{\text{табл}} = F_{0,05;\ 1;\ 4} = 7,71,$$

то полученное уравнение регрессии на уровне значимости $\alpha = 0,05$ можно признать адекватным.

Задача 57. В пункте 2 задачи 38 проверить адекватность уравнения регрессии при уровне значимости $\alpha = 0,05,$ считая, что каждое значение y_i есть среднее из 3 параллельных опытов и дисперсия воспроизводимости $s^2\{y\} = 2.$

Решение. В задаче 38 даны результаты эксперимента, состоящего из 4 различных опытов. В условии указано, что каждый из этих опытов повторен 3 раза, т. е. имеет место равномерное дублирование опытов. В этом случае дисперсия воспроизводимости рассчитывается по формуле (37). Следовательно, дисперсия $s^2\{y\} = 2$ имеет число степеней свободы $f_{\text{воспр}} = 4(3-1) = 8.$

В пункте 2 задачи 36 получено следующее уравнение регрессии:

$$\hat{y} = 4,73 - 4,13x_1x_2 - 1,04x_2^2.$$

Рассчитаем по этому уравнению значения отклика для каждого опыта, подставляя соответствующие значения факторов:

$$\hat{y} = 4,73 - 4,31 \cdot 1 \cdot 0 - 1,04 \cdot 0^2 = 4,73;$$

$$\hat{y} = 4,73 - 4,13 \cdot 0 \cdot 2 - 1,04 \cdot 2^2 = 0,57;$$

$$\hat{y} = 4,73 - 4,13 \cdot 1 \cdot (-1) - 1,04 \cdot (-1)^2 = 8;$$

$$\hat{y} = 4,73 - 4,13 \cdot 0 \cdot 1 - 1,04 \cdot 1^2 = 3,69.$$

Данные эксперимента и результаты расчетов занесем в таблицу (табл. 33).

Таблица 33

Номер опыта	$x_{1\,i}$	$x_{2\,i}$	$\overline{y}_i$	$\hat{y}_i$	$(\hat{y}_i - \overline{y}_i)^2$
1	1	0	6	4,73	1,6129
2	0	2	1	0,57	0,1849
3	1	–1	8	8,00	0,0000
4	0	1	2	3,69	2,8591

Вычислим дисперсию адекватности по формуле (42), учитывая, что число повторений всех опытов равно $k_i = 3$.

$$s^2_{\text{ад}} = \frac{3}{4-3}(1,6129 + 0,1849 + 0 + 2,8591) = 13,9617.$$

Число степеней свободы дисперсии адекватности равно $f_{\text{ад}} = 4 - 3 = 1$.

Проверим гипотезу об адекватности модели по критерию Фишера. Поскольку $F_{\text{расч}} = \dfrac{13,9617}{2} = 6,98 > F_{\text{табл}} = F_{0,05;\ 1;\ 8} = 5,32$, то полученное уравнение регрессии следует признать неадекватным.

Задачи для самостоятельного решения

Задача 58. В пункте 1 задачи 38 проверить адекватность уравнения регрессии при уровне значимости $\alpha = 0,05$, считая, что дисперсия воспроизводимости $s^2\{y\} = 0,1$ оценена по отдельной серии из 5 опытов.

Задача 59. В пункте 1 задачи 39 проверить адекватность уравнения регрессии при уровне значимости $\alpha = 0,05$, считая, что каждое значение y_i есть среднее из 4 параллельных опытов и дисперсия воспроизводимости $s^2\{y\} = 1$.

Задача 60. Проверьте при уровне значимости $\alpha = 0,05$ адекватность уравнения $\hat{y} = 0,45 - 0,09x + 0,8x^2$, полученного по данным результатам эксперимента:

x_i	–2	–1	–1	–1	0	0	1	1	1	2
y_i	3	1,5	2,5	2	0,3	–0,3	1	1	0,7	4

Ответы к задачам

58. $s^2_{ад} = 0,25$; $f_{ад} = 4 - 3 = 1$; $s^2\{y\} = 0,1$; $f_{воспр} = 5 - 1 = 4$; $F_{расч} = 2,5 < F_{0,05;\ 1;\ 4} = 7,71$. Модель адекватна. **59.** $s^2_{ад} = 16$; $f_{ад} = 4 - 3 = 1$; $s^2\{y\} = 1$; $f_{воспр} = 4(4 - 1) = 12$; $F_{расч} = 16 > F_{0,05;\ 1;\ 12} = 4,75$. Модель неадекватна. **60.** $s^2_{ад} = 1,44$; $f_{ад} = 5 - 3 = 2$; $s^2\{y\} = 0,148$; $f_{воспр} = 2 + 1 + 2 = 5$; $F_{расч} = 9,73 > F_{0,05;\ 2;\ 5} = 6,04$. Модель неадекватна.

Список использованной литературы

1. *Хикс Ч.* Основные принципы планирования эксперимента. — М. : Мир, 1967. — 406 с.
2. *Ахназарова С. Л., Кафаров В. В.* Методы оптимизации эксперимента в химической технологии. — М. : Высшая школа, 1985. — 327 с.
3. *Красовский Г. И.* Планирование эксперимента. — Минск : БГУ, 1982. — 302 с.
4. *Адлер Ю. П.* Планирование эксперимента при поиске оптимальных условий. — М. : Наука, 1976. — 320 с.

Список рекомендованной литературы

1. *Казаков, В. Ю.* Планирование и организация эксперимента / В. Ю. Казаков. — Томск : Изд-во ТПУ, 2009. — 127 с.
2. *Володарский, Е. Т.* Планирование и организация измерительного эксперимента / Т. Е. Володарский, Б. Н. Малиновский, Ю. М. Туз. — Киев : Высшая школа, 1987. — 280 с.
3. *Гмурман, В. Е.* Теория вероятностей и математическая статистика : учеб. пособие для вузов / В. Е. Гмурман. — М. : Юрайт, 2012. — 479 с.
4. *Вентцель, Е. С.* Теория вероятностей и ее инженерные приложения : учеб. пособие для втузов / Е. С. Вентцель, Л. А. Овчаров. — М. : Высшая школа, 2000. — 480 с.

Приложение 1

ФУНКЦИЯ ЛАПЛАСА

Значения $\Phi(x)=\frac{1}{\sqrt{2\pi}}\int\limits_0^x e^{-\frac{t^2}{2}}dt$

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
0,00	0,00	0,40	0,1554	0,80	0,2881	1,20	0,3849	1,60	0,4452
0,01	0,0040	0,41	0,1591	0,81	0,2910	1,21	0,3869	1,61	0,4463
0,02	0,0080	0,42	0,1628	0,82	0,2939	1,22	0,3888	1,62	0,4474
0,03	0,0120	0,43	0,1664	0,83	0,2967	1,23	0,3907	1,63	0,4484
0,04	0,0160	0,44	0,1700	0,84	0,2995	1,24	0,3925	1,64	0,4495
0,05	0,0199	0,45	0,1736	0,85	0,3023	1,25	0,3944	1,65	0,4505
0,06	0,0239	0,46	0,1772	0,86	0,3051	1,26	0,3962	1,66	0,4515
0,07	0,0279	0,47	0,1808	0,87	0,3078	1,27	0,3980	1,67	0,4525
0,08	0,0319	0,48	0,1844	0,88	0,3106	1,28	0,3997	1,68	0,4535
0,09	0,0359	0,49	0,1879	0,89	0,3133	1,29	0,4015	1,69	0,4545
0,10	0,0398	0,50	0,1915	0,90	0,3159	1,30	0,4032	1,70	0,4554
0,11	0,0438	0,51	0,1950	0,91	0,3186	1,31	0,4049	1,71	0,4564
0,12	0,0478	0,52	0,1985	0,92	0,3212	1,32	0,4066	1,72	0,4573
0,13	0,0517	0,53	0,2019	0,93	0,3238	1,33	0,4082	1,73	0,4582
0,14	0,0557	0,54	0,2054	0,94	0,3264	1,34	0,4099	1,74	0,4591
0,15	0,0596	0,55	0,2088	0,95	0,3289	1,35	0,4115	1,75	0,4599
0,16	0,0636	0,56	0,2123	0,96	0,3315	1,36	0,4131	1,76	0,4608
0,17	0,0675	0,57	0,2157	0,97	0,3340	1,37	0,4147	1,77	0,4616
0,18	0,0714	0,58	0,2190	0,98	0,3365	1,38	0,4162	1,78	0,4625
0,19	0,0753	0,59	0,2224	0,99	0,3389	1,39	0,4177	1,79	0,4633
0,20	0,0793	0,60	0,2257	1,00	0,3413	1,40	0,4192	1,80	0,4641
0,21	0,0832	0,61	0,2291	1,01	0,3438	1,41	0,4207	1,81	0,4649
0,22	0,0871	0,62	0,2324	1,02	0,3461	1,42	0,4222	1,82	0,4656
0,23	0,0910	0,63	0,2357	1,03	0,3485	1,43	0,4236	1,83	0,4664
0,24	0,0948	0,64	0,2389	1,04	0,3508	1,44	0,4251	1,84	0,4671
0,25	0,0987	0,65	0,2422	1,05	0,3531	1,45	0,4265	1,85	0,4678
0,26	0,1026	0,66	0,2454	1,06	0,3554	1,46	0,4279	1,86	0,4686
0,27	0,1064	0,67	0,2486	1,07	0,3577	1,47	0,4292	1,87	0,4693
0,28	0,1103	0,68	0,2517	1,08	0,3599	1,48	0,4306	1,88	0,4699
0,29	0,1141	0,69	0,2549	1,09	0,3621	1,49	0,4319	1,89	0,4706
0,30	0,1179	0,70	0,2580	1,10	0,3643	1,50	0,4332	1,90	0,4713
0,31	0,1217	0,71	0,2611	1,11	0,3665	1,51	0,4345	1,91	0,4719
0,32	0,1255	0,72	0,2642	1,12	0,3686	1,52	0,4357	1,92	0,4726
0,33	0,1293	0,73	0,2673	1,13	0,3708	1,53	0,4370	1,93	0,4732
0,34	0,1331	0,74	0,2704	1,14	0,3729	1,54	0,4382	1,94	0,4738
0,35	0,1368	0,75	0,2734	1,15	0,3749	1,55	0,4394	1,95	0,4744
0,36	0,1406	0,76	0,2764	1,16	0,3770	1,56	0,4406	1,96	0,4750
0,37	0,1443	0,77	0,2794	1,17	0,3790	1,57	0,4418	1,97	0,4756
0,38	0,1480	0,78	0,2823	1,18	0,3810	1,58	0,4429	1,98	0,4761
0,39	0,1517	0,79	0,2852	1,19	0,3830	1,59	0,4441	1,99	0,4767
2,00	0,4772	2,30	0,4893	2,60	0,4953	2,90	0,4981	3,20	0,499313
2,01	0,4778	2,31	0,4896	2,61	0,4955	2,91	0,4982	3,21	0,499336

Окончание прил. 1

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
2,02	0,4783	2,32	0,4898	2,62	0,4956	2,92	0,4982	3,22	0,499359
2,03	0,4788	2,33	0,4901	2,63	0,4957	2,93	0,4983	3,23	0,499381
2,04	0,4793	2,34	0,4904	2,64	0,4959	2,94	0,4984	3,24	0,499402
2,05	0,4798	2,35	0,4906	2,65	0,4960	2,95	0,4984	3,25	0,499423
2,06	0,4803	2,36	0,4909	2,66	0,4961	2,96	0,4985	3,26	0,499443
2,07	0,4808	2,37	0,4911	2,67	0,4962	2,97	0,4985	3,27	0,499462
2,08	0,4812	2,38	0,4913	2,68	0,4963	2,98	0,4986	3,28	0,499481
2,09	0,4817	2,39	0,4916	2,69	0,4964	2,99	0,4986	3,29	0,499499
2,10	0,4821	2,40	0,4918	2,70	0,4965	3,00	0,4987	3,30	0,499517
2,11	0,4826	2,41	0,4920	2,71	0,4966	3,01	0,4987	3,31	0,499533
2,12	0,4830	2,42	0,4922	2,72	0,4967	3,02	0,4987	3,32	0,499550
2,13	0,4834	2,43	0,4925	2,73	0,4968	3,03	0,4988	3,33	0,499566
2,14	0,4838	2,44	0,4927	2,74	0,4969	3,04	0,4988	3,34	0,499581
2,15	0,4842	2,45	0,4929	2,75	0,4970	3,05	0,4989	3,35	0,499596
2,16	0,4846	2,46	0,4931	2,76	0,4971	3,06	0,4989	3,36	0,499610
2,17	0,4850	2,47	0,4932	2,77	0,4972	3,07	0,4989	3,37	0,499624
2,18	0,4854	2,48	0,4934	2,78	0,4973	3,08	0,4990	3,38	0,499638
2,19	0,4857	2,49	0,4936	2,79	0,4974	3,09	0,4990	3,39	0,499650
2,20	0,4861	2,50	0,4938	2,80	0,4974	3,10	0,4990	3,40	0,499663
2,21	0,4864	2,51	0,4940	2,81	0,4975	3,11	0,4991	3,50	0,499767
2,22	0,4868	2,52	0,4941	2,82	0,4976	3,12	0,4991	3,60	0,499841
2,23	0,4871	2,53	0,4943	2,83	0,4977	3,13	0,4991	3,70	0,499892
2,24	0,4875	2,54	0,4945	2,84	0,4977	3,14	0,4992	3,80	0,499928
2,25	0,4878	2,55	0,4946	2,85	0,4978	3,15	0,4992	3,90	0,499952
2,26	0,4881	2,56	0,4948	2,86	0,4979	3,16	0,4992	4,00	0,499968
2,27	0,4884	2,57	0,4949	2,87	0,4979	3,17	0,4992	4,50	0,499997
2,28	0,4887	2,58	0,4951	2,88	0,4980	3,18	0,4993	5,00	0,4999997
2,29	0,4890	2,59	0,4952	2,89	0,4981	3,19	0,4993	$x > 5$	0,5

Приложение 2

КВАНТИЛИ χ^2-РАСПРЕДЕЛЕНИЯ

Значения $\chi^2_{\alpha,\nu}$ в зависимости от числа степеней свободы ν

при заданной вероятности α: $P(\chi^2_\nu > \chi^2_{\alpha,\nu}) = \alpha$

ν	α							
	0,01	0,025	0,05	0,1	0,9	0,95	0,975	0,99
1	6,6349	5,0239	3,8415	2,7055	0,0158	0,0039	0,0010	0,0002
2	9,2103	7,3778	5,9915	4,6052	0,2107	0,1026	0,0506	0,0201
3	11,345	9,3484	7,8147	6,2514	0,5844	0,3518	0,2158	0,1148
4	13,277	11,143	9,4877	7,7794	1,0636	0,7107	0,4844	0,2971
8	20,090	17,535	15,507	13,362	3,4895	2,7326	2,1797	1,6465
10	23,209	20,483	18,307	15,987	4,8652	3,9403	3,247	2,5582
15	30,578	27,488	24,996	22,307	8,5468	7,2609	6,2621	5,2293
20	37,566	34,170	31,410	28,412	12,443	10,851	9,5908	8,2604
30	50,892	46,979	43,773	40,256	20,599	18,493	16,791	14,953
40	63,691	59,342	55,758	51,805	29,051	26,509	24,433	22,164
50	76,154	71,42	67,505	63,167	37,689	34,764	32,357	29,707
100	135,81	129,56	124,34	118,5	82,358	77,929	74,222	70,065
200	249,45	241,06	233,99	226,02	174,84	168,28	162,73	156,43

Приложение 3

КВАНТИЛИ *t*-РАСПРЕДЕЛЕНИЯ СТЬЮДЕНТА

Значения $t_{\alpha,\nu}$ в зависимости от числа степеней свободы ν

при заданной вероятности α: $P(|t_\nu| > t_{\alpha,\nu}) = \alpha$

ν	α							
	0,2	0,1	0,05	0,025	0,01	0,005	0,001	0,0005
1	3,08	6,31	12,7	25,5	63,7	127	637	1273
2	1,89	2,92	4,30	6,21	9,92	14,1	31,6	44,7
3	1,64	2,35	3,18	4,18	5,84	7,45	12,9	16,3
4	1,53	2,13	2,78	3,50	4,60	5,60	8,61	10,3
5	1,48	2,02	2,57	3,16	4,03	4,77	6,87	7,98
8	1,40	1,86	2,31	2,75	3,36	3,83	5,04	5,62
10	1,37	1,81	2,23	2,63	3,17	3,58	4,59	5,05
12	1,36	1,78	2,18	2,56	3,05	3,43	4,32	4,72
15	1,34	1,75	2,13	2,49	2,95	3,29	4,07	4,42
20	1,33	1,72	2,09	2,42	2,85	3,15	3,85	4,15
30	1,31	1,70	2,04	2,36	2,75	3,03	3,65	3,90
40	1,30	1,68	2,02	2,33	2,70	2,97	3,55	3,79
50	1,30	1,68	2,01	2,31	2,68	2,94	3,50	3,72
60	1,30	1,67	2,00	2,30	2,66	2,91	3,46	3,68
100	1,29	1,66	1,98	2,28	2,63	2,87	3,39	3,60
200	1,29	1,65	1,97	2,26	2,60	2,84	3,34	3,54
∞	1,28	1,64	1,96	2,24	2,58	2,81	3,29	3,48

Приложение 4

КВАНТИЛИ *F*-РАСПРЕДЕЛЕНИЯ ФИШЕРА

Значения $F_{\alpha,\,\nu_1,\,\nu_2}$ в зависимости от числа степеней свободы ν_1 и ν_2

при заданной вероятности α: $p(F_{\nu_1,\,\nu_2} > F_{\alpha,\,\nu_1,\,\nu_2}) = \alpha$

$\alpha = 0,05$

ν_2	ν_1											
	1	2	4	6	8	10	12	15	20	40	60	∞
1	161	200	225	234	239	242	244	246	248	251	252	254
2	18,5	19,0	19,2	19,3	19,4	19,4	19,4	19,4	19,4	19,5	19,5	19,5
4	7,71	6,94	6,39	6,16	6,04	5,96	5,91	5,86	5,80	5,72	5,69	5,63
6	5,99	5,14	4,53	4,28	4,15	4,06	4,00	3,94	3,87	3,77	3,74	3,67
8	5,32	4,46	3,84	3,58	3,44	3,35	3,28	3,22	3,15	3,04	3,01	2,93
10	4,96	4,10	3,48	3,22	3,07	2,98	2,91	2,85	2,77	2,66	2,62	2,54
12	4,75	3,88	3,26	3,00	2,85	2,75	2,69	2,62	2,54	2,43	2,38	2,30
15	4,54	3,68	3,06	2,79	2,64	2.54	2,48	2,40	2,33	2,20	2,16	2,07
20	4,35	3,49	2,87	2,60	2,45	2,35	2,28	2,20	2,12	1,99	1,95	1,84
40	4,08	3,23	2,61	2,34	2,18	2,08	2,00	1,92	1,84	1,69	1,64	1,51
60	4,00	3,15	2,53	2,25	2,10	1,99	1,92	1,84	1,75	1,59	1,53	1,39
80	3,96	3,11	2,49	2,21	2,06	1,95	1,88	1,79	1,70	1,54	1,48	1,32
100	3,94	3,09	2,46	2,19	2,03	1,93	1,85	1,77	1,68	1,52	1,45	1,28
∞	3,84	3,00	2,37	2,10	1,94	1,83	1,75	1,67	1,57	1,39	1,32	1,00

$\alpha = 0,025$

ν_2	ν_1											
	1	2	4	6	8	10	12	15	20	40	60	∞
1	648	800	900	937	957	969	977	985	993	1006	1010	1018
2	38,5	39	39,3	39,3	39,7	39,4	39,4	39,4	39,5	39,5	39,5	39,5
4	12,2	10,7	9,6	9,2	8,98	8,84	8,75	8,66	8,56	8,41	8,36	8,26
6	8,81	7,26	6,23	5,82	5,60	5,46	5,37	5,27	5,17	5,01	4,96	4,85
8	7,57	6,06	5,05	4,65	4,43	4,30	4,20	4,10	4,00	3,84	3,78	3,67
10	6,94	5,46	4,47	4,07	3,85	3,72	3,62	3,52	3,42	3,26	3,20	3,08
12	6,55	5,10	4,12	3,73	3,51	3,37	3,28	3,18	3,07	2,91	2,85	2,72
15	6,20	4,77	3,80	3,41	3,20	3.06	2,96	2,86	2,76	2,59	2,52	2,40
20	5,87	4,46	3,51	3,13	2,91	2,77	2,68	2,57	2,46	2,29	2,22	2,09
40	5,42	4,05	3,13	2,74	2,53	2,39	2,29	2,18	2,07	1,88	1,80	1,64
60	5,29	3,93	3,01	2,63	2,41	2,27	2,17	2,06	1,94	1,74	1,67	1,48
80	5,22	3,86	2,95	2,57	2,35	2,21	2,11	2,00	1,88	1,68	1,60	1,40
100	5,18	3,83	2,92	2,54	2,32	2,18	2,08	1,97	1,85	1,64	1,56	1,35
∞	5,02	3,69	2,79	2,41	2,19	2,05	1,94	1,83	1,71	1,48	1,39	1,00

Приложение 5

КВАНТИЛИ *G*-РАСПРЕДЕЛЕНИЯ КОЧРЕНА

Таблица П.5.1

Значения $G_{\alpha, v, N}$ в зависимости от числа степеней свободы ν и *N* при заданной вероятности $\alpha = 0,054$; $P(G_{v,N} > G_{0,05;\, v;\, N}) = 0,05$

N	*v*								
	1	2	3	4	6	8	10	16	36
2	0,9985	0,9750	0,9392	0,9057	0,8534	0,8159	0,7880	0,7341	0,6602
3	0,9969	0,8709	0,7977	0,7457	0,6771	0,6333	0,6025	0,5466	0,4748
4	0,9065	0,7679	0,6841	0,6287	0,5598	0,5175	0,4884	0,4366	0,3720
5	0,8412	0,6838	0,5981	0,5440	0,4783	0,4387	0,4118	0,3645	0,3066
6	0,7808	0,6161	0,5321	0,4803	0,4184	0,3817	0,3568	0,3135	0,2612
8	0,6798	0,5157	0,4377	0,3910	0,3362	0,3043	0,2829	0,2462	0,2022
10	0,6020	0,4450	0,3733	0,3311	0,2823	0,2541	0,2353	0,2032	0,1655
15	0,4709	0,3346	0,2758	0,2419	0,2034	0,1815	0,1671	0,1429	0,1144
20	0,3894	0,2705	0,2205	0,1921	0,1602	0,1422	0,1303	0,1108	0,0879
24	0,3434	0,2354	0,1907	0,1656	0,1374	0,1216	0,1113	0,0942	0,0743
30	0,2929	0,1980	0,1593	0,1377	0,1137	0,1002	0,0921	0,0771	0,0604
40	0,2370	0,1576	0,1259	0,1082	0,0887	0,0780	0,0713	0,0595	0,0462
60	0,1737	0,1131	0,0895	0,0766	0,0623	0,0552	0,0497	0,0411	0,0316
120	0,0998	0,0632	0,0495	0,0419	0,0337	0,0292	0,0266	0,0218	0,0165

Таблица П.5.2

Значения $G_{\alpha, v, N}$ в зависимости от числа степеней свободы *v* и *N* при заданной вероятности $\alpha = 0,01; P(G_{v,\, N} > G_{0,05;\, v;\, N}) = 0,01$

N	*v*								
	1	2	3	4	6	8	10	16	36
2	0,9999	0,9950	0,9794	0,9586	0,9172	0,8823	0,8539	0,7946	0,7067
3	0,9933	0,9423	0,8831	0,8335	0,7606	0,7107	0,6743	0,6059	0,5153
4	0,9676	0,8643	0,7814	0,7212	0,6410	0,5897	0,5536	0,4884	0,4057
5	0,9279	0,7885	0,6957	0,6329	0,5531	0,5037	0,4697	0,4094	0,3351
6	0,8828	0,7218	0,6258	0,5635	0,4866	0,4401	0,4084	0,3529	0,2858
8	0,7945	0,6152	0,5209	0,4627	0,3932	0,3522	0,3248	0,2779	0,2214
10	0,7175	0,5358	0,4469	0,3934	0,3308	0,2945	0,2704	0,2297	0,1811
15	0,5747	0,4069	0,3317	0,2882	0,2386	0,2104	0,1918	0,1612	0,1251
20	0,4799	0,3297	0,2654	0,2288	0,1877	0,1646	0,1501	0,1248	0,0960
24	0,4247	0,2871	0,2295	0,1970	0,1608	0,1406	0,1283	0,1283	0,0810
30	0,3632	0,2412	0,1913	0,1635	0,1327	0,1157	0,1054	0,0867	0,0658
40	0,2940	0,1915	0,1508	0,1281	0,1033	0,0898	0,0816	0,0668	0,0503
60	0,2151	0,1371	0,1069	0,0902	0,0722	0,0625	0,0567	0,0461	0,0344
120	0,1225	0,0759	0,0585	0,0489	0,0387	0,0334	0,0302	0,0242	0,0178

Учебное электронное издание

Ерещенко Татьяна Владимировна
Михайлова Наталия Анатольевна

ПЛАНИРОВАНИЕ ЭКСПЕРИМЕНТА

Учебно-практическое пособие

Начальник РИО *М. Л. Песчаная*
Редактор *И. Б. Чижикова*
Компьютерная правка и верстка *А. Г. Вишняков*

Минимальные систем. требования:
PC 486 DX-33; Microsoft Windows XP; Internet Explorer 6.0; Adobe Reader 6.0.

Подписано в свет 17.12.2014.
Гарнитура «Times New Roman». Уч.-изд. л. 3,3. Объем данных 1,1 Мбайт.

Федеральное государственное бюджетное образовательное учреждение
высшего профессионального образования
«Волгоградский государственный архитектурно-строительный университет»
Редакционно-издательский отдел
400074, Волгоград, ул. Академическая, 1
http://www.vgasu.ru, info@vgasu.ru