Учреждение образования
«БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ
ТЕХНОЛОГИЧЕСКИЙ УНИВЕРСИТЕТ»

Е. И. Блинова

ПЛАНИРОВАНИЕ И ОРГАНИЗАЦИЯ ЭКСПЕРИМЕНТА

Рекомендовано
учебно-методическим объединением высших учебных заведений Республики Беларусь по химико-технологическому образованию в качестве учебно-методического пособия для студентов высших учебных заведений по специальности 1-54 01 03 «Физико-химические методы и приборы контроля качества продукции»

Минск 2010

УДК 519.242:001.891(075.4)
ББК 72ся73
Б69

Рецензенты:
кафедра теории вероятностей и математической статистики БГУ
(заведующий кафедрой доктор физико-математических наук,
профессор *Н. Н. Труш*);
кандидат физико-математических наук, доцент
кафедры высшей математики № 2 БНТУ *Л. И. Бородич*

Блинова, Е. И.

Б69 Планирование и организация эксперимента : учеб.-метод. пособие для студентов специальности 1-54 01 03 «Физико-химические методы и приборы контроля качества продукции» / Е. И. Блинова. — Минск : БГТУ, 2010. — 130 с.
ISBN 978-985-530-003-9.

В пособии излагаются основные процедуры статистической обработки результатов наблюдений и методы регрессионного анализа, рассмотрены вопросы планирования многофакторных регрессионных и оптимизационных экспериментов. Изложение материала сопровождается, где это возможно, примерами решения задач. Приведены задачи для самостоятельного решения с ответами, теоретические вопросы и контрольные задания по курсу.

УДК 519.242:001.891(075.4)
ББК 72ся73

ISBN 978-985-530-003-9

ПРЕДИСЛОВИЕ

«Все, что мы знаем о реальности, исходит из опыта и завершается им», — писал А. Эйнштейн. Организация и планирование экспериментальных исследований является важной частью деятельности ученых, инженеров и технологов. По выражению специалиста в области управления качеством Исикава Кафу, «инженер, который дает оценку на основе экспериментальных данных, должен знать статистические методы наизусть». Управление качеством — это прежде всего грамотное применение современных статистических методов. Специалист обязан понимать, в каких ситуациях применимы различные статистические методы и в каких эти методы применять нельзя, знать их свойства, уметь интерпретировать результаты статистического анализа.

Дисциплина «Планирование и организация эксперимента» относится к циклу общепрофессиональных и специальных дисциплин, которые создают теоретическую базу знаний и практических навыков по специальности. Она дает фундамент для анализа и обработки результатов экспериментов, проводимых студентами при подготовке курсовых и дипломных проектов.

Теория планирования эксперимента основана на использовании методов математической статистики и формулирует приемы и способы оптимальной организации экспериментальных исследований таким образом, чтобы получить максимум информации при наименьших затратах времени и средств.

В пособии изложены основные идеи и методы статистического анализа экспериментальных данных, элементы регрессионного анализа, методы планирования многофакторных регрессионных и экстремальных экспериментов. Изложение сопровождается большим количеством разобранных примеров, а также задач для самостоятельного решения с ответами и указаниями к решению. Примеры прикладного характера заимствованы из литературы, в частности, из [1], [2], [3], [8], [12].

Пособие написано в соответствии с действующей учебной программой курса «Планирование и организация эксперимента» для студентов специальности 1-54 01 03 «Физико-химические методы и приборы контроля качества продукции» высших учебных заведений. Оно предназначено для самостоятельного изучения дисциплины студентами заочной формы обучения. В конце пособия приведены вопросы по курсу и контрольные задания.

Раздел 1. СТАТИСТИЧЕСКИЙ АНАЛИЗ ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ

Выборка и ее описание

Выборка — множество значений результатов наблюдений над одной и той же случайной величиной. Элементы выборки называются **выборочными значениями** или **вариантами**. Количество проведенных наблюдений называется **объемом** выборки.

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

x_i^*	x_1^*	x_2^*	$\ldots$	x_k^*
n_i	n_1	n_2	$\ldots$	n_k

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

При изучении непрерывной случайной величины 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 $\dfrac{n_i}{h}$ (соответственно, $\dfrac{n_i}{nh}$), где h — длина интервалов группировки. Гистограмма относительных частот обладает тем свойством, что ее площадь равна $\sum\limits_{i=1}^{k} \dfrac{n_i}{nh} \cdot h = \sum\limits_{i=1}^{k} \dfrac{n_i}{n} = 1.$
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 случайной величины.

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

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

Эмпирическая функция распределения обладает следующими свойствами:

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

2) является неубывающей, непрерывной слева функцией;

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

Замечание. При построении эмпирической функции распределения по интервальному статистическому ряду ее относят к серединам интервалов.

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

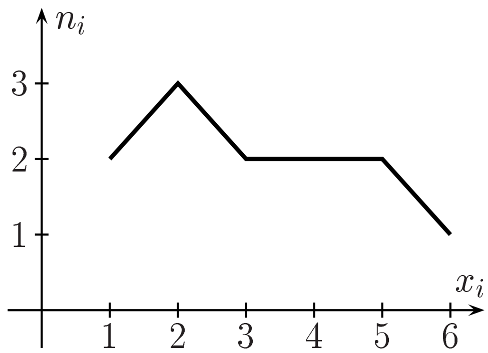
Объем выборки равен сумме частот наблюдаемых значений и равен $n = 2 + 3 + 2 + 2 + 1 = 10$. Учитывая это, запишем эмпирическую функцию распределения.

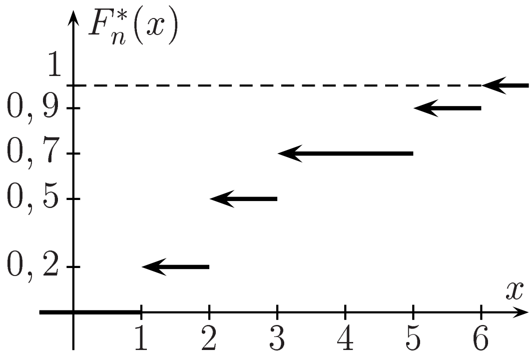
Наименьшее наблюдаемое выборочное значение 1, поэтому эмпирическая функция распределения равна 0 при $x \leqslant 1$. Далее ее значение изменяется каждый раз при переходе x через x_i^*: увеличивается на величину относительной частоты $w_i = \dfrac{n_i}{n}$. Наибольшее выборочное значение 6, поэтому эмпирическая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

Итак,

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

Построим полигон частот и график эмпирической функции распределения:





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

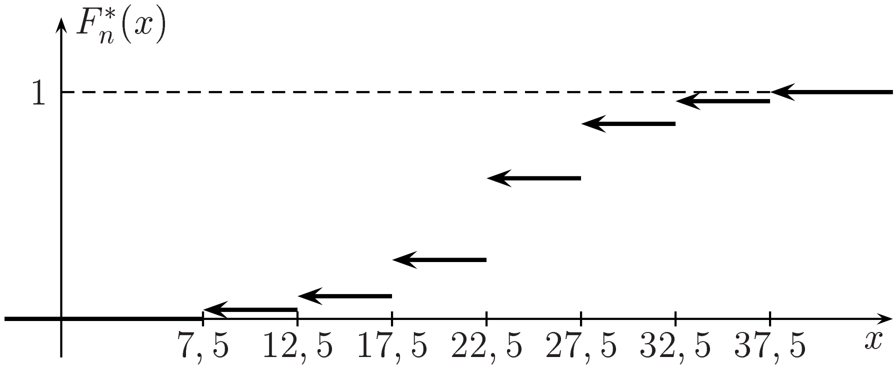
$[x_{i-1}; x_i)$	$[5; 10)$	$[10; 15)$	$[15; 20)$	$[20; 25)$	$[25; 30)$	$[30; 35)$	$[35; 40)$
n_i	4	6	16	36	24	10	4

Решение. Объем выборки равен сумме наблюдаемых частот: $n = 4 + 6 + 46 + 36 + 24 + 10 + 4 = 100$. Длина каждого интервала равна $h = 5$. 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 $\dfrac{n_i}{nh}$. Дополним этими значениями интервальный статистический ряд:

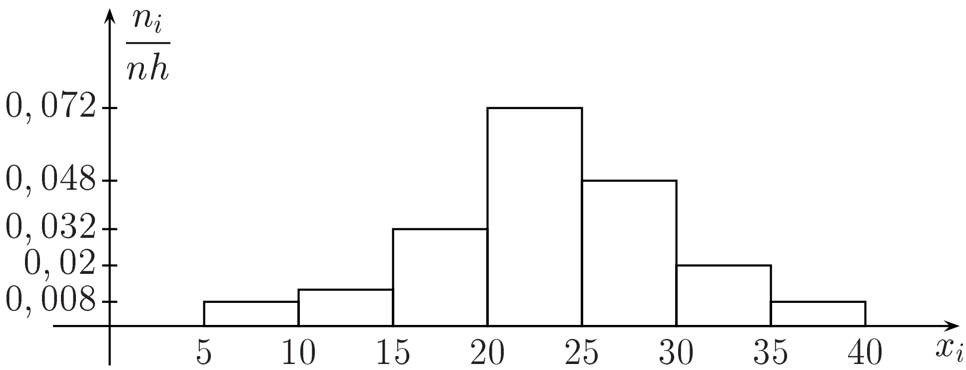
$[x_{i-1}; x_i)$	$[5; 10)$	$[10; 15)$	$[15; 20)$	$[20; 25)$	$[25; 30)$	$[30; 35)$	$[35; 40)$
x_i^*	7,5	12,5	17,5	22,5	27,5	32,5	37,5
n_i	4	6	16	36	24	10	4
$\dfrac{n_i}{n}$	0,04	0,06	0,16	0,36	0,24	0,1	0,04
$\dfrac{n_i}{nh}$	0,008	0,012	0,032	0,072	0,048	0,02	0,008

Запишем эмпирическую функцию распределения и построим ее график:

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



Построим гистограмму относительных частот:



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

Задачи для самостоятельного решения

Задача 3. При проверке партии изделий получены следующие данные по сортам:

$$\begin{array}{cccccccccc} 1 & 2 & 1 & 2 & 1 & 1 & 2 & 3 & 4 & 2 \\ 1 & 1 & 2 & 1 & 3 & 2 & 1 & 4 & 1 & 2 \end{array}$$

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 изделий с целью определения процента влажности древесины, из которой изготовлены эти изделия. 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 предположить, что наблюдаемая случайная величина имеет нормальное распределение?

Ответы к задачам 3–4. **3.**

x_i^*	1	2	3	4
n_i	9	7	2	2

$$F_n^*(x) = \begin{cases} 0 & \text{при } x \leqslant 1, \\ 0,45 & \text{при } 1 < x \leqslant 2, \\ 0,8 & \text{при } 2 < x \leqslant 3, \\ 0,9 & \text{при } 3 < x \leqslant 4, \\ 1 & \text{при } x > 4. \end{cases}$$

4. $$F_n^*(x) = \begin{cases} 0 & \text{при } x \leqslant 12, \\ 0,053 & \text{при } 12 < x \leqslant 14, \\ 0,333 & \text{при } 14 < x \leqslant 16, \\ 0,673 & \text{при } 16 < x \leqslant 18, \\ 0,92 & \text{при } 18 < x \leqslant 20, \\ 1 & \text{при } x > 20. \end{cases}$$

Точечное оценивание параметров распределения

Выборка представляет собой ряд наблюдений над 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.

Нормальное распределение полностью определяется двумя параметрами — математическим ожиданием μ и дисперсией σ^2. Математическое ожидание и дисперсия являются основными числовыми характеристиками любой случайной величины. Математическое ожидание — это в некотором смысле (т. е. с учетом большей и меньшей вероятности различных значений) среднее значение случайной величины. Дисперсия характеризует разброс значений случайной величины относительно ее математического ожидания. 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 в ы ч и с л и т ь числовые характеристики случайной величины, а только о ц е н и т ь их.

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

Точечной оценкой параметра θ называется любая статистика $\widehat{\theta}_n$, 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

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

или кратко:

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

Оценка для дисперсии рассчитывается по формуле

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

Эту оценку называют **несмещенной оценкой дисперсии**. Число $f = n - 1$ называется **числом степеней свободы** этой оценки дисперсии, оно 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 (1) удобнее использовать формулу

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

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

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{k} x_i^* n_i, \qquad s^2 = \frac{1}{n-1}\sum_{i=1}^{k}\left(x_i^* - \overline{x}\right)^2 n_i, \qquad (3)$$

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

Для любого параметра можно предложить разные оценки. Так, в качестве оценки для математического ожидания можно использовать не только $\overline{x}$, но и среднее арифметическое наибольшего и наименьшего элементов выборки $\dfrac{x_{\max} + x_{\min}}{2}$, первый элемент выборки x_1 и т. д.

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

Если математическое ожидание оценки равно оцениваемому параметру, то такая оценка называется **несмещенной**. Требование несмещенности гарантирует отсутствие систематических ошибок при оценивании. Оно особенно важно при малом числе наблюдений (в случае выборок объема не более 30).

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

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

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

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

Примеры решения задач

Задача 5. При 5 проверках октанового числа одного и того же сорта бензина получены следующие результаты:

43; 44; 40; 39; 44.

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

Решение. Вычислим выборочное среднее

$$\overline{x} = \frac{1}{5}(43 + 44 + 40 + 39 + 44) = 42$$

и выборочную дисперсию

$$s^2 = \frac{1}{5-1}\Big((43-42)^2 + (44-42)^2 + (40-42)^2 + \\ + (39-42)^2 + (44-42)^2\Big) = 5,5.$$

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

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), где k — число интервалов, x_i^* — середина i-го интервала, n_i — частота наблюдения значений из i-го интервала. Итак, несмещенные оценки для математического ожидания и дисперсии вычисляются следующим образом:

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

$$s^2 = \frac{1}{100-1}\Big((7,5-23,3)^2\cdot 4 + (12,5-23,3)^2\cdot 6 + \\ + (17,5-23,3)^2\cdot 16 + (22,5-23,3)^2\cdot 36 + (27,5-23,3)^2\cdot 24 + \\ + (32,5-23,3)^2\cdot 10 + (37,5-23,3)^2\cdot 4\Big) \approx 43,8.$$

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

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.

Ответы к задачам 7–8. **7.** $\overline{x} = 1,85$; $s^2 = 0,976$; $f = 19$. **8.** $\overline{x} = 16,04$; $s^2 = 4,27$; $f = 149$.

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 — независимые, нормально распределенные случайные величины. Нормальное распределение — самое важное в прикладной статистике, поскольку многие наблюдаемые данные можно успешно описать нормальным распределением. В дальнейшем запись $\xi \sim \mathcal{N}(\mu;\sigma)$ будет означать, что случайная величина ξ имеет нормальное распределение с математическим ожиданием $\mathbf{M}\xi = \mu$ и дисперсией $\mathbf{D}\xi = \sigma^2$.

Утверждение 1. *Пусть имеется выборка объема n из нормального распределения с математическим ожиданием μ и дисперсией σ^2, т. е. $x_1, x_2, \ldots, x_n \sim \mathcal{N}(\mu;\sigma)$. Тогда*

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

С нормальным распределением связаны также следующие три наиболее часто использующиеся в статистике распределения: χ^2-распределение, t-распределение Стьюдента и F-распределение Фишера.

Распределением χ^2 с k степенями свободы называется распределение суммы квадратов k независимых случайных величин, распределенных по нормальному закону с параметрами 0 и 1, т. е. если $\xi_1, \xi_2, \ldots, \xi_k \sim \mathcal{N}(0;1)$, то $\eta = \xi_1^2 + \xi_2^2 + \ldots + \xi_k^2 \sim \chi_k^2$.

Утверждение 2. *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 $\xi \sim \mathcal{N}(0;1)$ и $\eta \sim \chi^2_k$ независимы.

Утверждение 3. *В случае выборки объема n из нормального распределения с 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 $F = \dfrac{\eta_1/k_1}{\eta_2/k_2} \sim F_{k_1;\,k_2}$, где случайные величины $\eta_1 \sim \chi^2_{k_1}$ и $\eta_2 \sim \chi^2_{k_2}$ независимы.

Замечание. Существуют таблицы распределений χ^2, Стьюдента, Фишера, а также нормального распределения. Однако при использовании статистических таблиц необходимо обращать пристальное внимание на то, какие величины затабулированы и какие нужны при вычислениях.

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 $\widehat{\theta}_1 = \widehat{\theta}_1(x_1, x_2, \ldots, x_n)$ и $\widehat{\theta}_2 = \widehat{\theta}_2(x_1, x_2, \ldots, x_n)$ являются функциями выборочных значений и который с заданной вероятностью γ накрывает истинное значение оцениваемого параметра θ:

$$\mathbf{P}\left(\widehat{\theta}_1(x_1, x_2, \ldots, x_n) < \theta < \widehat{\theta}_2(x_1, x_2, \ldots, x_n)\right) = \gamma.$$

Интервал $\left(\widehat{\theta}_1; \widehat{\theta}_2\right)$ 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 мы можем гарантировать.)

Доверительную вероятность γ обычно выбирают равной $0,9$; $0,95$ или $0,99$, т. е. так, 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 строится на основании утверждения 1 и определяется соотношением

$$\mathbf{P}\left(\overline{x} - u_\alpha \cdot \frac{\sigma}{\sqrt{n}} < \mu < \overline{x} + u_\alpha \cdot \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha, \tag{4}$$

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

Доверительный интервал для математического ожидания в случае выборки из нормального распределения с неизвестной дисперсией строится на основании утверждения 3 и определяется соотношением

$$\mathbf{P}\left(\overline{x} - t_{\alpha;\,n-1} \cdot \frac{s}{\sqrt{n}} < \mu < \overline{x} + t_{\alpha;\,n-1} \cdot \frac{s}{\sqrt{n}}\right) = 1 - \alpha, \tag{5}$$

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

Доверительный интервал для дисперсии в случае выборки из нормального распределения с неизвестным математическим ожиданием строится на основании утверждения 2 и определяется соотношением

$$\mathbf{P}\left(\frac{(n-1)s^2}{\chi^2_{\alpha/2;\,n-1}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{1-\alpha/2;\,n-1}}\right) = 1-\alpha, \qquad (6)$$

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 $\mathbf{P}\left(\chi^2_{n-1} \geqslant \chi^2_{\alpha;\,n-1}\right) = \alpha$ для случайной величины χ^2_{n-1}, имеющей распределение χ^2 с числом степеней свободы $\nu = n-1$ (см. приложение 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 при неизвестной дисперсии для решения этой задачи достаточно

$$n_0 > \frac{s^2 t^2_{\alpha;\,n-1}}{\varepsilon^2} \qquad (7)$$

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

Примеры решения задач

Задача 9. Оценить математическое ожидание нормального распределения с заданной надежностью γ, если:
1) среднеквадратическое отклонение известно: $\sigma = 2$ и по выборке объема 10 найдено выборочное среднее $\overline{x} = 5,4$ ($\gamma = 0,95$);
2) по выборке объема 9 найдены несмещенные оценки математического ожидания и дисперсии: $\overline{x} = 14,2$; $s^2 = 5,76$ ($\gamma = 0,99$).

Решение. 1) Поскольку дисперсия известна, построим доверительный интервал для математического ожидания нормального распределения с помощью формулы (4). Имеем:

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

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

Запишем доверительный интервал по формуле (4):

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

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

2) Поскольку дисперсия неизвестна и оценена по выборке, для построения доверительного интервала для математического ожидания нужно использовать формулу (5). Имеем:

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

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

Запишем доверительный интервал по формуле (5):

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

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

Задача 10. Для определения высоты полета на самолете установлено три высотомера, которые в некоторый момент времени дали результаты 3680 м; 3740 м; 3710 м. Сколько надо иметь таких приборов на самолете, чтобы с надежностью 0,95 ошибка при определении средней высоты была в пределах 50 м, если считается, что ошибки высотомеров нормальны?

Решение. Задача заключается в определении объема выборки, необходимого для достижения заданной точности $\varepsilon = 50$ м. 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:

$$\overline{x} = \frac{1}{3}\left(3680 + 3740 + 3710\right) = 3710;$$

$$s^2 = \frac{1}{3-1}\left((3680 - 3710)^2 + (3740 - 3710)^2 + (3710 - 3710)^2\right) = 900.$$

По формуле (7) определяем, что

$$n_0 > \frac{900 \cdot 4,3^2}{50^2} \approx 6,7.$$

Таким образом, потребуется установить 7 приборов.

Задача 11. При взвешивании груза получены следующие результаты:

$$129;\quad 125;\quad 130;\quad 122;\quad 135;\quad 125;\quad 120;\quad 130;\quad 127.$$

Определить среднюю квадратическую ошибку взвешивания и построить для нее доверительный интервал с надежностью $0,9$.

Решение. Имеем: $n = 9$; $\gamma = 0,9$; $\alpha = 1 - 0,9 = 0,1$. Рассчитаем оценки для математического ожидания и дисперсии: $\overline{x} = 127$; $s^2 = 21$; $s = \sqrt{21} \approx 4,58$.

Таким образом, среднеквадратическая ошибка взвешивания оценивается величиной $s \approx 4,58$. Чтобы получить для нее доверительный интервал, построим доверительный интервал для дисперсии по формуле (6). По таблице приложения 2 определим квантили распределения χ^2:

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

Доверительный интервал для дисперсии определяется по формуле (6):

$$\mathbf{P}\left(\frac{8\cdot 21}{15,507} < \sigma^2 < \frac{8\cdot 21}{2,7326}\right) = 0,9;$$

$$\mathbf{P}\left(10,83 < \sigma^2 < 61,48\right) = 0,9.$$

Отсюда получаем доверительный интервал для среднего квадратического отклонения в виде:

$$\mathbf{P}\left(\sqrt{10,83} < \sigma < \sqrt{61,48}\right) = 0,9;$$

$$\mathbf{P}\,(3,3 < \sigma < 7,8) = 0,9.$$

Задачи для самостоятельного решения

Задача 12. Найти доверительный интервал для математического ожидания нормального распределения с заданной надежностью γ, если:
1) среднеквадратическое отклонение известно: $\sigma = 3$ и по результатам 25 независимых наблюдений найдено выборочное среднее $\overline{x} = 20,12$ ($\gamma = 0,99$);
2) по выборке объема 12 найдены несмещенные оценки математического ожидания и дисперсии: $\overline{x} = 16,8$; $s^2 = 2,25$ ($\gamma = 0,95$).

Задача 13. 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 14. Даны результаты 5 независимых равноточных измерений толщины металлической пластинки:

$$2,015; \quad 2,020; \quad 2,025; \quad 2,020; \quad 2,015.$$

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 12–14. **12.** 1) $18,57 < \mu < 21,67$; 2) $15,85 < \mu < 17,75$. **13.** $22,948 < \mu < 23,374$; $0,224 < \sigma < 0,576$. **14.** 1) $2,014 < \mu < 2,024$; 2) 16.

Проверка статистических гипотез

Основные задачи математической статистики разделяют на две категории, тесно связанные между собой, но отличающиеся постановкой задач: оценивание параметров и проверка статистических гипотез. Основной задачей оценивания параметров является получение по выборке оценок, наилучших в том или ином смысле. При проверке гипотез задача ставится иначе: требуется по выборке принять или отвергнуть некоторое предположение о распределении, из которого извлечена выборка.

Статистической гипотезой называется любое предположение о виде или параметрах неизвестного распределения. Правило, которое позволяет по выборке принять или отвергнуть проверяемую гипотезу, называется **статистическим критерием**.

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

Проверяемую гипотезу обычно называют **нулевой** и обозначают H_0. Наряду с нулевой рассматривают **альтернативную**, или **конкурирующую**, гипотезу H_a $(\overline{H})$. Нулевая и альтернативная гипотезы представляют собой две возможности выбора, осуществляемого в задачах проверки гипотез.

В задачах проверки гипотез возможны следующие четыре ситуации.

Проверяемая гипотеза H_0	H_0 принимается	H_0 отвергается
объективно верна	правильное решение	ошибка 1-го рода
объективно неверна	ошибка 2-го рода	правильное решение

Вероятность ошибки 1-го рода, т. е. вероятность отвергнуть нулевую гипотезу, когда она верна, называется **уровнем значимости** статистического критерия и обозначается α. Вероятность ошибки 2-го рода, т. е. вероятность ошибочно принять нулевую гипотезу, обозначается β. Пользуясь терминологией статистического контроля качества продукции, можно сказать, что α — это риск поставщика (забраковка 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. Невозможно одновременно уменьшить обе ошибки. Реально поступают следующим образом: задают уровень значимости α (как правило, равный $0,05$; $0,01$ или $0,1$), а затем выбирают статистический критерий так, чтобы ошибка 2-го рода была наименьшей.

Критерии значимости

Статистические критерии, с помощью которых проверяются гипотезы о значениях параметров распределения или о соотношениях между ними в предположении, что тип 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 и критерий Колмогорова.

Рассмотрим критерии значимости, предназначенные для проверки гипотез о значениях 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 оценкой того или иного показателя качества. Дисперсия характеризует разброс экспериментальных значений, а следовательно, служит мерой точности. Например, если произведено несколько измерений одной и той же величины, то дисперсия может характеризовать точность прибора, метода измерения и т. д.

1. Проверка гипотезы о равенстве математического ожидания заданному значению.

Нулевая гипотеза $H_0: \ \mu = \mu_0$.

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

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

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

$$u_{\text{расч}} = \frac{|\overline{x} - \mu_0|}{\sqrt{\sigma^2/n}} < u_{\text{табл}} = u_\alpha, \tag{8}$$

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

Если дисперсия σ^2 не известна, то гипотеза H_0 принимается при

$$t_{\text{расч}} = \frac{|\overline{x} - \mu_0|}{\sqrt{s^2/n}} < t_{\text{табл}} = t_{\alpha;\,n-1}, \tag{9}$$

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

2. Проверка гипотезы о равенстве заданному значению дисперсии нормального распределения.

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

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

Гипотеза H_0 при заданном уровне значимости α принимается, если

$$\chi^2_{1-\alpha/2;\,n-1} < \chi^2_{\text{расч}} = \frac{(n-1)s^2}{\sigma_0^2} < \chi^2_{\alpha/2;\,n-1}, \tag{10}$$

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

3. Сравнение двух дисперсий нормально распределенных признаков. Такая задача возникает, если требуется сравнить точность приборов, инструментов, методов измерения. Лучшим будет тот прибор, инструмент, метод, который дает меньший разброс результатов, т. е. меньшую дисперсию.

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

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

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

$$F_{\text{расч}} = \frac{s_{\max}^2}{s_{\min}^2} < F_{\text{табл}} = F_{\alpha/2;\, f_1;\, f_2}. \tag{11}$$

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

4. Сравнение нескольких дисперсий нормально распределенных признаков. Пусть имеется N независимых выборок и требуется при заданном уровне значимости α проверить гипотезу о равенстве дисперсий.

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

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

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

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

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

$$G_{\text{расч}} = \frac{s_{\max}^2}{\sum\limits_{i=1}^{N} s_i^2} < G_{\text{табл}} = G_{\alpha;\, f_1;\, f_2}. \tag{12}$$

Здесь $G_{\text{расч}}$ равно отношению наибольшей оценки дисперсии к сумме всех сравниваемых дисперсий, $G_{\alpha;\, f_1;\, f_2}$ определяются по таблице распределения Кохрена (приложение 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 применять 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.

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

$$u_{\text{расч}} = \frac{|\overline{x}_1 - \overline{x}_2|}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} < u_{\text{табл}} = u_\alpha, \tag{13}$$

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

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

$$t_{\text{расч}} = \frac{|\overline{x}_1 - \overline{x}_2|}{\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} < t_{\text{табл}} = t_{\alpha;\, f}, \tag{14}$$

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 таблице приложения 3.

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

$$t_{\text{расч}} = \frac{|\overline{x}_1 - \overline{x}_2|}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} < t_{\text{табл}} = t_{\alpha;\, f}, \tag{15}$$

где $t_{\alpha;\, f}$ определяется по таблице приложения 3 при

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

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.

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, одинаковы ли результаты использования двух методов измерения. Либо если проводятся измерения какой-то характеристики для одних и тех же объектов до и после воздействия, и требуется определить, влияет ли это воздействие на значение характеристики.

В этом случае имеются две выборки одинакового объема n:

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

Поскольку значения в каждой паре x_{1i}, x_{2i} связаны (например, измерены на одном и том же объекте), то получим новую выборку с элементами $\Delta x_i = x_{1i} - x_{2i}$. Задача сводится к проверке гипотезы о равенстве нулю среднего значения новой выборки, т. е. $H_0 : \mu_{\Delta x} = 0$. Эта проверка проводится по критерию (9).

Замечание. Иногда возникает необходимость сравнения гипотезы $H_0 : \ \theta = \theta_0$ с **односторонней** альтернативой $\overline{H}_1 : \ \theta > \theta_0$ или $\overline{H}_2 : \ \theta < \theta_0$.

Например, при проверке гипотезы $H_0 : \ \mu = \mu_0$ против альтернативы $\overline{H}_1 : \mu > \mu_0$ требуется выяснить, соответствует ли выборочное среднее значение норме или превосходит ее. Пусть дисперсия σ^2 известна. Оценкой для параметра μ является $\overline{x}$. Ясно, что если $\overline{x} < \mu_0$, то гипотезу H_0 следует предпочесть альтернативе $\overline{H}_1$. В случае $\overline{x} > \mu_0$ гипотеза H_0 принимается на уровне значимости α, если выполняется условие (8) 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 (9), (11), (13)–(15) в случае односторонних альтернатив.

Критерий (10) используется следующим образом. В случае альтернативы $\overline{H}_1 : \sigma^2 > \sigma_0^2$ 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 $\overline{H}_2 : \sigma^2 < \sigma_0^2$ гипотеза H_0 принимается, если

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

Сделаем теперь 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 (10)–(12) весьма чувствительны к отклонениям распределений от нормального. В то же время критерии сравнения средних (8), (9), (13)–(15) устойчивы к умеренным отклонениям распределений от нормального; критерий (14) может использоваться при умеренном отклонении от выполнения требования о равенстве дисперсий, если объемы выборок приблизительно равны. В случае, когда эти условия не выполняются, для сравнения средних двух выборок используют критерий Манна – Уитни, который не требует 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 15. Дисперсия предела прочности на разрыв некоторого волокна составляет 35,63 фунт^2. 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 $H_0 : \sigma^2 = 35,63$ в сравнении с односторонней альтернативой $\overline{H} :\ \sigma^2 < 35,63$. Положим уровень значимости $\alpha = 0,05$.

Будем использовать для проверки гипотезы критерий (10). 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 таблице приложения 2 отсутствуют значения для числа степеней 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 с замечанием относительно односторонних альтернатив, нулевая гипотеза $H_0 : \sigma^2 = 35,63$ 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 16. 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, и в каждую из лабораторий отправили по 8 кусков. Дисперсии результатов измерений в разных лабораториях получились следующими:

3,86; 4,27; 1,35; 3,9; 1,64.

Проверить гипотезу об однородности дисперсий. Что характеризуют эти дисперсии?

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

Поскольку каждая из $N = 5$ дисперсий получена по результатам 8 измерений, мы имеем случай выборок одинакового объема, число степеней свободы каждой дисперсии равно $f = 8 - 1 = 7$. Проверим гипотезу о равенстве этих пяти дисперсий по критерию Кохрена (12).

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 его 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. Разброс значений измерений зависит от содержания золы в куске (однородности образца) и от методики проведения измерений, т. е. от качества проведения анализа в данной лаборатории. Поскольку гипотеза об однородности дисперсий принимается, нет оснований считать, что качество анализа в какой-либо лаборатории хуже, чем в других.

Задача 17. Согласно паспортным данным, диаммонийфосфат марки ЧДА должен содержать не менее 99% основного вещества. Требуется проверить гипотезу статистической значимости различия между паспортными данными и следующими результатами трех определений содержания диаммонийфосфата в реактиве: 98,3; 97,3; 97,8%.

Решение. Вычислим $\overline{x} = 97,8\%$, $s^2 = 0,25$. Объем выборки $n = 3$. Требуется определить, достаточно ли сильно номинальное значение 99% превосходит средний результат измерений $\overline{x} = 97,8\%$, чтобы можно было утверждать, что реактивы не соответствуют паспортным данным. Для этого нужно проверить гипотезу $H_0 : \mu = 99\%$ при альтернативе $H_1 : \mu < 99\%$. 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 (9). 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

$$t_{\text{расч}} = \frac{|97,8 - 99|}{\sqrt{0,25/3}} = 4,16 > t_{\text{табл}} = 2,92,$$

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

Задача 18. 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 таблице приведены значения времени сварки в минутах при использовании формовочных смесей разных типов. Можно ли сказать, что имеет место значимое уменьшение времени сварки?

№ п/п	Сырая смесь	Сухая смесь
1	19	21
2	28	15
3	14	11
4	29	12
5	15	21
$\overline{x_i}$	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 $\overline{H} : \mu_1 > \mu_2$. 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 $\overline{H'} : \sigma_1^2 \neq \sigma_2^2$ при заданном уровне значимости $\alpha = 0,05$. Проверка проводится по критерию Фишера (11): поскольку

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

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 свободы $f = 4 + 4 = 8$.

Итак, для проверки гипотезы H_0 против альтернативы $\overline{H}$ используем критерий (14). Поскольку альтернатива $\overline{H}$ односторонняя, аналогично задаче 17 сравниваем

$$t_{\text{расч}} = \frac{21 - 16}{\sqrt{36,75 \cdot \left(\frac{1}{5} + \frac{1}{5}\right)}} = 1,3 < t_{\text{табл}} = t_{0,1;8} = 1,86.$$

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

Задача 19. Проводилось исследование предела прочности на разрыв различных по химической структуре твердых смол. При этом требовалось выяснить, вносят ли действия оператора какое-нибудь смещение в результаты наблюдений. Было взято по два образца каждой смолы. Двум операторам A и B предложили испытать по одному образцу смол каждого типа. Можно ли говорить, что наблюдаются различия между результатами наблюдений двух операторов?

Смола	127	135	138	139	146	152
Оператор A	5250	4975	5050	5075	4795	5190
Оператор B	5230	4980	5020	5085	4750	5120
Δx_i	20	-5	30	-10	45	70

Решение. Мы имеем задачу сравнения средних в случае парных (зависимых) выборок, т. к. значения предела прочности на разрыв различных смол могут отличаться и нужно выявить различие между измерениями, проведенными операторами A и B. Поскольку значения в каждой паре x_{1i}, x_{2i} связаны, получим новую выборку с элементами $\Delta x_i = x_{1i} - x_{2i}$ и проверим по критерию (9) гипотезу о равенстве нулю среднего значения новой выборки, т. е. $H_0 : \mu_{\Delta x} = 0$ при альтернативе $\overline{H} : \mu_{\Delta x} \neq 0$.

Положим уровень значимости $\alpha = 0,05$. Вычислим $\overline{\Delta x} = 25$, $s^2 = 920$, объем выборки $n = 6$. Сравнивая

$$t_{\text{расч}} = \frac{25}{\sqrt{920/6}} = 2,02 < t_{\text{табл}} = t_{0,05;5} = 2,57,$$

заключаем, что при уровне значимости $\alpha = 0,05$ нет оснований для отклонения гипотезы H_0, а значит, различия между результатами измерений у двух операторов следует признать незначительными.

Задачи для самостоятельного решения

Во всех задачах предполагается, что рассматриваемые признаки подчиняются нормальному распределению.

Задача 20. 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. Отдел технического контроля 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-автомат изготавливает детали уменьшенного размера и поэтому требуется произвести подналадку станка? (Принять уровень значимости равным $\alpha = 0,05$.)

Задача 21. В одном из цехов анализируется работа листопрокатного стана по результатам контроля качества продукции. Основным показателем качества является толщина готового стального листа, номинал которой 2,11 мм. 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

Вспомогательные величины: $\sum x_i = 16,74$, $\sum x_i^2 = 35,0358$, $\sum x_i n_i = 63$, $\sum x_i^2 n_i = 132,31$.

Задача 22. Проверка скорости полимеризации проводится на нескольких образцах полимеров. Предполагаемая скорость полимеризации составляет 24% в час. В восьми опытах получены следующие результаты:

$$23,6;\quad 22,8;\quad 25,7;\quad 24,8;\quad 26,4;\quad 24,3;\quad 23,9;\quad 25.$$

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

Задача 23. Станок в автоматическом токарном цехе производит цилиндрические болты определенного типа. Из одной партии болтов взята выборка объемом 20 и произведены измерения длины каждого болта, по которым рассчитаны статистические характеристики $\overline{x} = 18$ мм, $s^2 = 0,000784$ мм^2, $s = 0,028$ мм. Контролер по опыту знает точность (а тем самым и среднее квадратическое отклонение как меру рассеивания) станка при производстве болтов данного типа, которая в общем 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 $s = 0,028$ мм указывает на несоответствие точности станка предъявленным требованиям?

Задача 24. 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 25. 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 миллиметрах квадратных):

$$0,067; \quad 0,136; \quad 0,168; \quad 0,068; \quad 0,066; \quad 0,102; \quad 0,137.$$

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

Задача 26. Исследовалась работа промышленного агрегата по процессу извлечения гелия из природного Оренбургского газа. Целью работы установки является более полное (близкое к 100%) извлечение гелия. Испытывались два технологических режима для того, чтобы выбрать лучший по признаку наибольшего процента извлечения гелия. Можно ли при уровне значимости $\alpha = 0,05$ считать, что обе технологии дают одинаковые результаты?

Технология	$\overline{x_i}$	s_i^2	n_i
1-я	$99,2$	$0,14$	110
2-я	$99,34$	$0,123$	60

Задача 27. Определение концентрации SiO_2 в мартеновском шлаке проводилось весовым (5 опытов) и фотокалориметрическим (6 опытов) методами. Для первого метода получены следующие значения среднего и дисперсии: $20,5$ и $0,0529$; для второго метода: $21,3$ и $0,16$. Свидетельствует ли при уровне значимости $\alpha = 0,1$ различие между средними значениями о систематическом расхождении между результатами применения первого и второго методов?

Задача 28. Выпускаемый промышленностью поливинилбутираль может содержать до 2–3% сорбированной на его поверхности остаточной влаги. Количественное определение влаги проводится по ГОСТ 11736-68 методом электрометрического титрования. При большой чувствительности и достаточной точности этот метод продолжителен по времени, т. к. включает стадию растворения полимера. Исследовалась возможность определения влаги непосредственно в порошкообразных образцах поливинилбутираля методом газовой хроматографии. Даны результаты определения влаги (%) двумя методами для семи образцов. Можно ли использовать более простой метод газовой хроматографии вместо метода электрометрического титрования? (Принять уровень значимости $\alpha = 0,05$.)

Хроматографическое определение	0,51	1,17	0,28	0,99	0,96	0,85	0,66
Электрометрическое титрование	0,49	1,19	0,27	1	1	0,88	0,65

Вспомогательные величины: $\sum x_{1i} = 5,42$; $\sum x_{1i}^2 = 4,7672$; $\sum x_{2i} = 5,48$; $\sum x_{2i}^2 = 4,926$; $\sum \Delta x_i = -0,06$; $\sum (\Delta x_i)^2 = 0,0036$.

Задача 29. Одной из важных характеристик качества колумбийской кормовой патоки является число градусов плотности Брикса. Это показатель количества твердого вещества в патоке и основной фактор, рассматриваемый при ее производстве. Поставщиками патоки являются два различных района страны. Приведены 8 выборочных показателей для каждого из районов. Одинаково ли среднее число градусов Брикса для этих двух районов? (Принять уровень значимости $\alpha = 0,05$.)

1-й район	81,6	81,3	82	79,6	78,4	81,8	80,2	80,7
2-й район	81,8	84,7	82	85,6	79,9	83,2	84,1	85

Вспомогательные величины: $\sum x_{1i} = 645,6$; $\sum x_{1i}^2 = 52110,74$; $\sum x_{2i} = 666,3$; $\sum x_{2i}^2 = 55520,75$; $\sum \Delta x_i = 20,7$; $\sum (\Delta x_i)^2 = 85,51$.

Ответы и указания к решению задач 20–29.

20. $H_0 : \mu = 12$ ($\sigma^2 = 0,25$ известна); $\overline{H} : \mu < 12$. Используем критерий (8): $u_{\text{расч}} = 3,6$; $u_{\text{табл}} = u_{0,1} = 1,645$. Итак, следует признать, что станок изготавливает детали уменьшенного размера.

21. $H_0 : \mu = 2,11$ (σ^2 не известна); $\overline{H} : \mu \neq 2,11$. Учитывая, что дан сгруппированный статистический ряд, вычисляем $n = \sum n_i = 30$; $\overline{x} = \frac{1}{n}\sum x_i n_i = 2,1$; $s^2 = \frac{1}{29}(132,31 - 30 \cdot 2,1^2) = 0,000345$ (по формуле (2)). Используем критерий (9): $t_{\text{расч}} = 2,95$; $t_{\text{табл}} = t_{0,05;\,29} \approx t_{0,05;\,30} = 2,04$. Следовательно, гипотеза о соответствии продукции номиналу должна быть отвергнута.

22. $H_0 : \mu = 24$ (σ^2 не известна); $\overline{H} : \mu \neq 24$. Определяем $n = 8$; $\overline{x} = 24,5625$; $s^2 = 1,35125$. Используем критерий (9): $t_{\text{расч}} = 1,37$; $t_{\text{табл}} = t_{0,05;\,7} \approx \approx \frac{1}{8-5}(2 \cdot t_{0,05;\,8} + 1 \cdot t_{0,05;\,5}) = 2,4$. Следовательно, нет оснований утверждать, что полученные результаты несовместимы с предполагаемым значением средней скорости полимеризации.

23. $H_0 : \sigma^2 = \sigma_0^2$; $\overline{H} : \sigma^2 > \sigma_0^2$. Используем критерий (10): $\chi^2_{\text{расч}} = 37,24$; $\chi^2_{0,05;\,19} \approx \chi^2_{0,05;\,20} = 31,41$. Следовательно, точность станка не соответствует предъявленным требованиям.

24. $H_0 : \sigma_1^2 = \sigma_2^2$; $\overline{H} : \sigma_1^2 < \sigma_2^2$. Используем критерий (11): $F_{\text{расч}} = 4,375$; $F_{0,05;\,5;\,3} \approx \frac{1}{4}(19,3 + 19,2 + 6,39 + 6,16) = 12,745$. Итак, нет оснований считать дисперсии неоднородными. Следовательно, усовершенствование неэффективно.

25. $H_0 : \sigma_i^2 = \text{const}$; $\overline{H} : \sigma_i^2 \neq \text{const}$. Используем критерий Кохрена (12): $G_{\text{расч}} = 0,226$; $G_{0,05;\,16;\,7} \approx \frac{1}{2}(0,3135 + 0,2462) = 0,27985$. Итак, нет оснований считать дисперсии неоднородными; разладка машины не наблюдается.

26. Проверяется гипотеза $H_0 : \mu_1 = \mu_2$ при альтернативе $\overline{H} : \mu_1 \neq \mu_2$. Для выбора формы критерия Стьюдента необходимо проверить вспомогательную гипотезу $H_0' : \sigma_1^2 = \sigma_2^2$; $\overline{H'} : \sigma_1^2 \neq \sigma_2^2$. По критерию Фишера (11): $F_{\text{расч}} = 1,14$; $F_{0,025;\,109;\,59} \approx F_{0,025;\,109;\,60} > F_{0,025;\,\infty;\,60} = 1,48$. Итак, нет оснований считать дисперсии неоднородными. Следовательно, для проверки гипотезы H_0 критерий Стьюдента применяется в форме (14): $t_{\text{расч}} = 2,38$; $t_{\text{табл}} = t_{0,05;\,168} \approx \approx \frac{1}{2}(1,98 + 1,97) = 1,975$. Гипотеза о том, что обе технологии дают одинаковые результаты, отвергается.

27. $H_0 : \mu_1 = \mu_2$; $\overline{H} : \mu_1 \neq \mu_2$. Вспомогательная гипотеза $H_0' : \sigma_1^2 = \sigma_2^2$ при альтернативе $\overline{H'} : \sigma_1^2 \neq \sigma_2^2$ проверяется по критерию Фишера (11): $F_{\text{расч}} = 3,02$; $F_{0,05;\,5;\,4} \approx \frac{1}{2}(6,39 + 6,16) = 6,275$. Таким образом, нет оснований считать дисперсии неоднородными. Итак, для проверки гипотезы H_0 используем (14): $t_{\text{расч}} = 3,94$; $t_{\text{табл}} = t_{0,1;\,9} \approx \frac{1}{2}(1,86 + 1,81) = 1,835$. Следует признать, что между результатами применения двух методов имеются систематические расхождения.

28. Поскольку мы имеем парные наблюдения (для каждого из 7 образцов проводились измерения влаги двумя методами, при этом количество влаги может варьироваться от образца к образцу), то проверяемая гипотеза имеет вид $H_0 : \mu_{\Delta x} = 0$; $\overline{H} : \mu_{\Delta x} \neq 0$. Вычисляем $\overline{\Delta x} = \frac{1}{7}\sum \Delta x_i = -0,0086$; $s^2 = \frac{1}{6}(0,0036 - 7 \cdot (-0,0086)^2) = 0,0005$ (по формуле (2)). Используем критерий (9): $t_{\text{расч}} = 1,02$; $t_{\text{табл}} = t_{0,05;\,6} \approx \frac{1}{8-5}(1 \cdot t_{0,05;\,8} + 2 \cdot t_{0,05;\,5}) = 2,48$. Методы дают одинаковые результаты, можно использовать более простой метод.

29. Даны две независимые выборки, необходимо проверить гипотезу $H_0 : \mu_1 = \mu_2$ при альтернативе $\overline{H} : \mu_1 \neq \mu_2$. Для выбора формы критерия Стьюдента должна быть проверена вспомогательная гипотеза $H_0' : \sigma_1^2 = \sigma_2^2$ при альтернативе $\overline{H'} : \sigma_1^2 \neq \sigma_2^2$. Вычисляем $\overline{x_1} = \frac{1}{8}\sum x_{1i} = 80,7$; $\overline{x_2} = \frac{1}{8}\sum x_{2i} = 83,3$; $s_1^2 = \frac{1}{7}(52110,74 - 8 \cdot (80,7)^2) = 1,55$; $s_2^2 = \frac{1}{7}(55520,75 - 8 \cdot (83,3)^2) = 3,76$ (по формуле (2)). По критерию Фишера (11): $F_{\text{расч}} = 2,43$; $F_{0,025;\,7;\,7} \approx \approx \frac{1}{4}(5,82 + 4,65 + 5,60 + 4,43) = 5,13$. Итак, нет оснований считать дисперсии неоднородными. Поэтому для проверки гипотезы H_0 используем (14): $t_{\text{расч}} = 3,19$; $t_{\text{табл}} = t_{0,05;\,14} \approx \frac{1}{15-12}(2 \cdot t_{0,05;\,15} + 1 \cdot t_{0,05;\,12}) = 2,15$. Следует признать, что среднее число градусов Брикса для двух районов различно.

Раздел 2. ЭЛЕМЕНТЫ РЕГРЕССИОННОГО АНАЛИЗА

Основные задачи корреляционного и регрессионного анализа

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 одно, строго определенное значение другой; 3) связаны статистической зависимостью, при которой каждому значению одной величины соответствует множество возможных значений другой. **Статистической** (или **вероятностной**, или **стохастической**) называется зависимость, при которой изменение значения одной величины влечет изменение распределения другой. Например, статистической будет зависимость урожайности некоторой 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. Такая зависимость может быть представлена в виде: $M(\eta|\xi = x) = \varphi(x)$; $M(\xi|\eta = y) = \psi(y)$. Эти уравнения называют модельными уравнениями регрессии соответственно η на ξ (или η по ξ) и ξ на η; функции $\varphi(x)$, $\psi(y)$ — функциями регрессии (модельными), а их графики — линиями регрессии (модельными).

На практике по экспериментальным данным можно найти только оценку (приближенное выражение) функции регрессии. Говорят, 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.

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.

Выборочный коэффициент корреляции и его свойства

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

$$r_{x,y} = \frac{\sum\limits_{i=1}^{N} x_i y_i - N \cdot \overline{x} \cdot \overline{y}}{\sqrt{\left(\sum\limits_{i=1}^{N} x_i^2 - N \cdot (\overline{x})^2\right)\left(\sum\limits_{i=1}^{N} y_i^2 - N \cdot (\overline{y})^2\right)}}, \tag{16}$$

где $\overline{x} = \frac{1}{N}\sum\limits_{i=1}^{N} x_i$; $\overline{y} = \frac{1}{N}\sum\limits_{i=1}^{N} y_i$; N — объем выборки (количество наблюдений $(x_i; y_i)$).

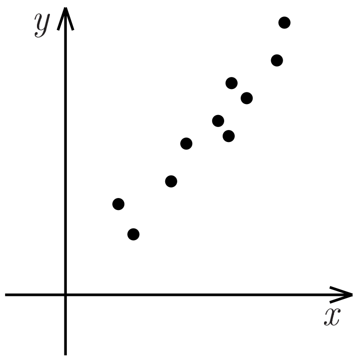
Выборочный коэффициент корреляции обладает следующими свойствами.

1. $-1 \leqslant r_{x,y} \leqslant 1$.

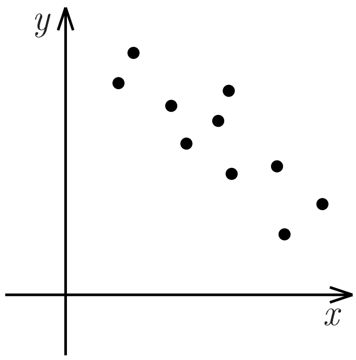
2. Если с ростом значений одной величины значения второй также в основном возрастают, то $r_{x,y} > 0$; если с ростом значений одной величины значения второй в основном убывают, то $r_{x,y} < 0$.

3. Если $|r_{x,y}| = 1$ (или близок к 1), то наблюдаемые величины x и y связаны линейной зависимостью, т. е. $y = b_0 + b_1 x$ (экспериментальные точки на корреляционном поле расположены вдоль некоторой прямой).

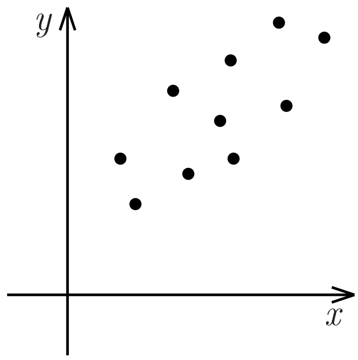
4. Если наблюдаемые величины независимы, то $r_{x,y} \approx 0$.



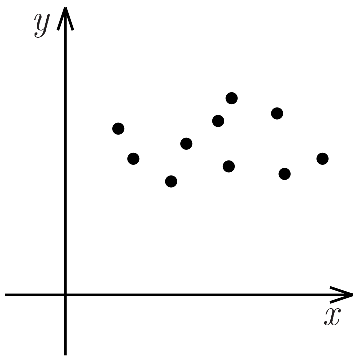
$r_{x,y} \approx 0,95$



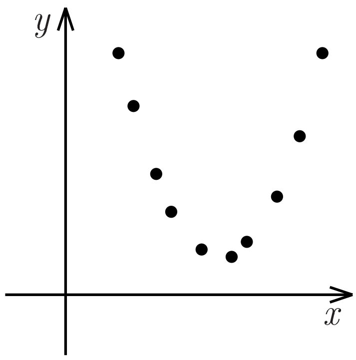
$r_{x,y} \approx -0,85$



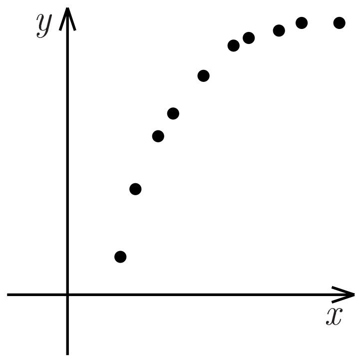
$r_{x,y} \approx 0,7$



$r_{x,y} \approx 0$



$r_{x,y} \approx 0$



$r_{x,y} \approx 0,9$

Для корреляционного поля, изображенного на первом рисунке, выборочный коэффициент корреляции $r_{x,y} \approx 0,95$, т. е. близок к 1, экспериментальные точки расположены вдоль некоторой прямой. Как можно видеть по второму и третьему рисункам, 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. На четвертом корреляционном поле точки разбросаны, зависимость между увеличением значений одной переменной и увеличением или уменьшением значений второй не прослеживается, коэффициент корреляции близок к нулю. Пятый рисунок показывает, что равенство нулю коэффициента корреляции не означает независимости между переменными.

Подчеркнем, что коэффициент корреляции является мерой именно *линейной* зависимости. На пятом и шестом рисунках точки образуют не прямые, а кривые линии, т. е. зависимость между переменными хотя и явно присутствует, но отличается от линейной, и 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 (проверка того, значимо ли отличается выборочный коэффициент корреляции от нуля) производится в случае выборки из нормального распределения по критерию Стьюдента.

Нулевая гипотеза $H_0: \ \varrho = 0.$

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

Гипотезу H_0 при заданном уровне значимости α принимают, если

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

где N — количество наблюдений, а значение $t_{\text{табл}}$ определяется по таблице приложения 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.

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 y и x, т. е. оценкой для модельного уравнения регрессии y на x является эмпирическое линейное уравнение регрессии

$$\widehat{y} = b_0 + b_1 x.$$

Далее по выборочным данным нужно оценить коэффициенты зависимости, т. е. получить оценки для b_0 и b_1.

Для этого используют **метод наименьших квадратов** (МНК). Идея МНК заключается в следующем: значения коэффициентов b_0 и b_1 выбирают так, чтобы сумма квадратов отклонений наблюдаемых значений y_i от предсказываемых по уравнению регрессии

$\widehat{y_i} = b_0 + b_1 x_i$ была наименьшей:

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

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

$$\begin{cases} N b_0 + b_1 \sum\limits_{i=1}^{N} x_i = \sum\limits_{i=1}^{N} y_i, \\ b_0 \sum\limits_{i=1}^{N} x_i + b_1 \sum\limits_{i=1}^{N} x_i^2 = \sum\limits_{i=1}^{N} x_i y_i. \end{cases} \tag{17}$$

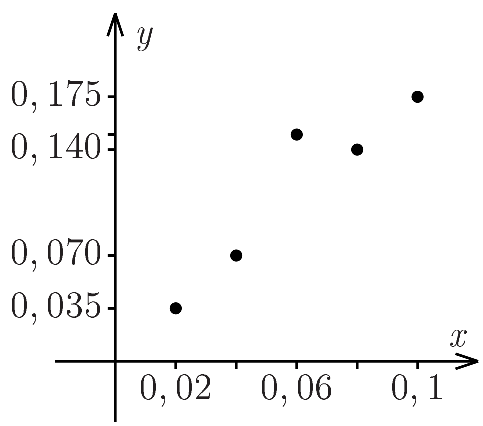
Примеры решения задач

Задача 30. При контроле качества пищевых продуктов для определения концентрации тех или иных веществ находят эмпирическое линейное уравнение зависимости оптической плотности градуировочного раствора от концентрации. Имеются данные для определения концентрации фосфора в мясных изделиях.

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 прямую на корреляционном поле.

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



2) По виду корреляционного поля можно предположить, что выборочный коэффициент корреляции положителен и значимо отличается от 0. Рассчитаем его по формуле (16).

Число опытов $N = 5$; вычислим необходимые суммы:

$$\sum_{i=1}^{5} x_i = 0,02 + 0,04 + 0,06 + 0,08 + 0,10 = 0,30;$$

$$\sum_{i=1}^{5} y_i = 0,035 + 0,070 + 0,150 + 0,140 + 0,175 = 0,570;$$

$$\sum_{i=1}^{5} x_i y_i = 0,02 \cdot 0,035 + 0,04 \cdot 0,070 + 0,06 \cdot 0,150 + 0,08 \cdot 0,140 +$$

$$+0,10 \cdot 0,175 = 0,0007 + 0,0028 + 0,0090 + 0,0112 + 0,0175 = 0,0412;$$

$$\sum_{i=1}^{5} x_i^2 = 0,02^2 + 0,04^2 + 0,06^2 + 0,08^2 + 0,10^2 =$$

$$= 0,0004 + 0,0016 + 0,0036 + 0,0064 + 0,0100 = 0,0220;$$

$$\sum_{i=1}^{5} y_i^2 = 0,035^2 + 0,070^2 + 0,150^2 + 0,140^2 + 0,175^2 =$$

$$= 0,001225 + 0,004900 + 0,022500 + 0,019600 + 0,030625 = 0,078850.$$

Для расчета этих сумм удобно составить вспомогательную таблицу.

№ п/п	x_i — концентрация	y_i — оптическая плотность	$x_i y_i$	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

Подставляя полученные значения в (16), получим:

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

Поскольку

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

то при уровне значимости $\alpha = 0,05$ 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. $y = b_0 + b_1x$.

3) Для того чтобы найти коэффициенты b_0 и b_1 эмпирического линейного уравнения регрессии $\widehat{y} = b_0 + b_1x$ по методу наименьших квадратов, составим систему нормальных уравнений по формуле (17) и решим ее:

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

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

Упрощая второе уравнение, получим:

$$0,06 \cdot (0,57 - 0,3b_1) + 0,022b_1 = 0,0412;$$

$$0,0342 - 0,018b_1 + 0,022b_1 = 0,0412;$$

$$0,004b_1 = 0,007;$$

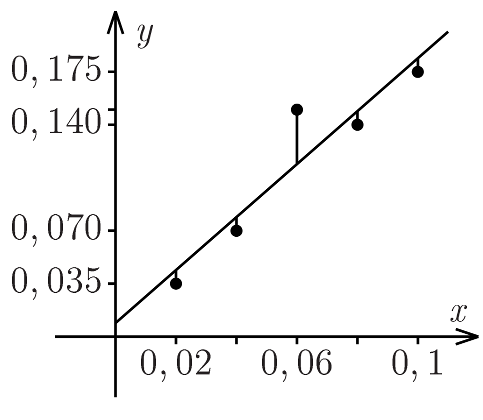
$$b_1 = 1,75.$$

Подставляя полученное значение в первое уравнение, определим

$$b_0 = \frac{0,57 - 0,3 \cdot 1,75}{5} = 0,009.$$

Итак, эмпирическое линейное уравнение регрессии имеет вид $\widehat{y} = 0,009 + 1,75x$.

Чтобы построить прямую на корреляционном поле, подставим в это уравнение какие-нибудь два значения переменной x. Например, при $x = 0$ получим $\widehat{y} = 0,009$; при $x = 0,1$ получим $\widehat{y} = 0,184$. Согласно методу наименьших квадратов, построенная прямая приближает экспериментальные данные наилучшим образом в том смысле, что будет наименьшей сумма квадратов длин вертикальных отрезков, показанных на рисунке.



Задачи для самостоятельного решения

Задача 31. Для разработки методики прогнозирования морозостойкости кирпича керамического на ОАО «Радашковичский кирпичный завод» проводился сбор статистической информации. Для изделий определялись: резервная пористость x_1, %, водопоглощение под вакуумом x_2, %, водопоглощение при атмосферном давлении x_3, %, и морозостойкость y, циклы.

x_1	$19,5$	$18,5$	$37,9$	$25,8$	$31,7$	$33,1$	34	$27,2$
x_2	$18,4$	$17,9$	$18,9$	19	$18,7$	$18,2$	$18,9$	$19,2$
x_3	$15,4$	$15,4$	$13,7$	$15,1$	$15,4$	$14,2$	$14,1$	$15,1$
y	30	30	76	41	60	54	65	40

По имеющимся данным для переменных x_1 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; построить прямую на корреляционном поле.

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

Задача 32. Решить задачу **31** для переменных x_2 и y.

Вспомогательные величины: $\sum x_{2i} = 149,2$; $\sum x_{2i}^2 = 2783,96$; $\sum x_{2i}y_i = 7405,7$.

Задача 33. Решить задачу **31** для переменных x_3 и y.

Вспомогательные величины: $\sum x_{3i} = 118,4$; $\sum x_{3i}^2 = 1755,64$; $\sum x_{3i}y_i = 5795,6$.

Ответы к задачам 31–33.

31. 2) $r_{x,y} = 0,964$; $t_{\text{расч}} = 8,93 > t_{\text{табл}} = t_{0,05;6} \approx \frac{1}{8-5}(1 \cdot t_{0,05;\,8} + 2 \cdot t_{0,05;\,5}) = 2,48$. Следовательно, при уровне значимости $\alpha = 0,05$ можно принять гипотезу о линейной зависимости величины y от x_1; 3) $\widehat{y} = -17,01 + 2,34x_1$.

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) $r_{x,y} \approx -0,8$, коэффициент корреляции значимо отличается от нуля, принимается гипотеза о линейной зависимости между переменными; 3) $\widehat{y} = 340,15 - 19,64x_3$.

Криволинейная регрессия

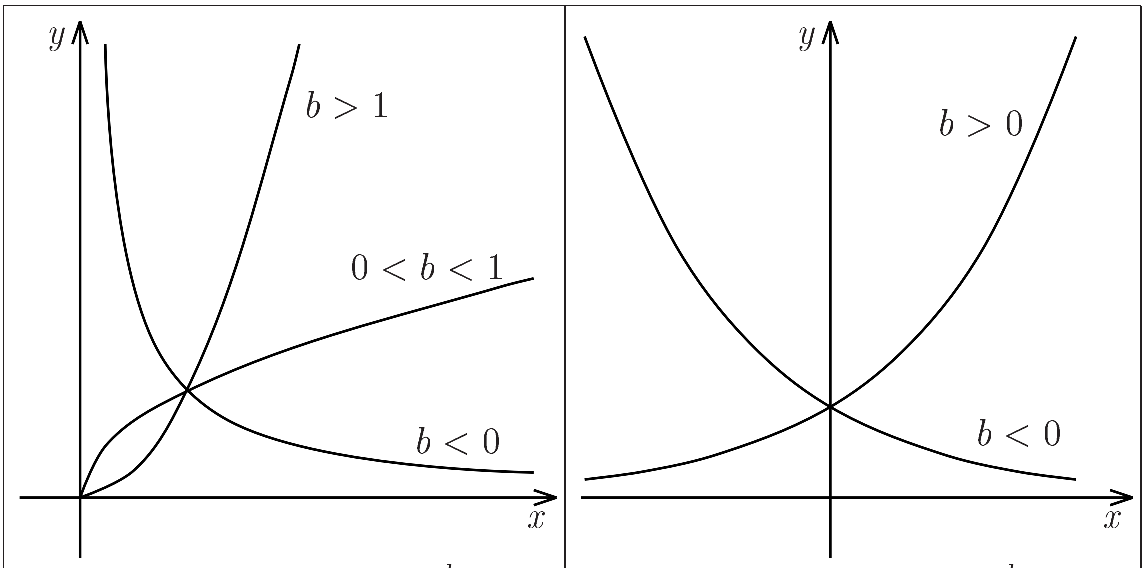
Зависимость между двумя наблюдаемыми величинами далеко не всегда можно выразить линейной функцией. Иногда видно, что точки корреляционного поля образуют некоторую кривую. При выборе вида эмпирической функции регрессии необходимо учитывать теоретические сведения и опыт предыдущих аналогичных исследований.

Как правило, до начала исследования должен быть определен вид эмпирической функции регрессии с точностью до нескольких параметров, значения которых оцениваются по результатам эксперимента. В том случае, если функция регрессии линейна по параметрам или может быть сведена к таковой с помощью замены переменных, для определения оценок параметров используют МНК.

Например, степенная зависимость вида $y = ax^b$ может быть сведена к линейной с помощью логарифмирования:

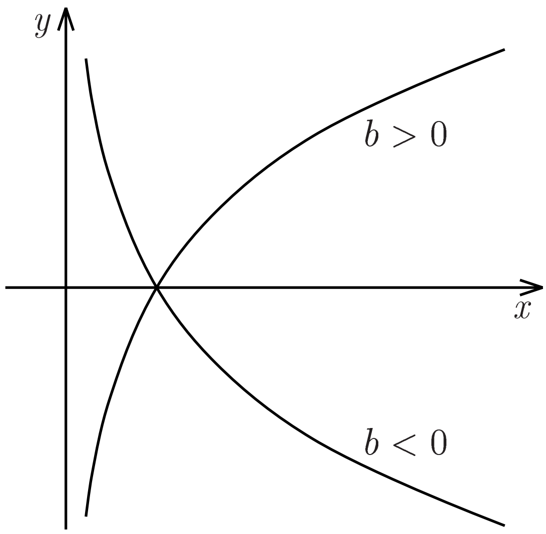
$$\ln y = \ln a + \ln x^b \;\Rightarrow\; \ln y = \ln a + b \ln x.$$

Если ввести новые переменные $Y = \ln y,\ X = \ln x$, исходная зависимость сведется к линейной $Y = b_0 + b_1 X$, коэффициенты которой могут быть найдены по МНК. Тогда коэффициенты искомой зависимости определятся из соотношений $a = \mathrm{e}^{b_0},\ b = b_1$.

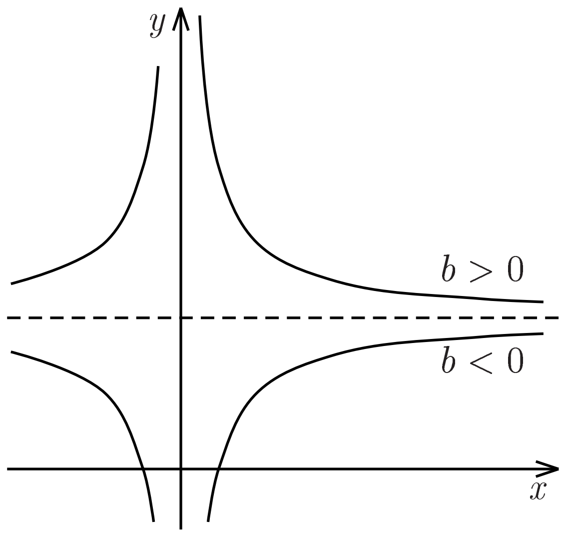


Зависимость $y = ax^b$ сводится к $Y = b_0 + b_1 X$ заменой $Y = \ln y,\ X = \ln x$ $(a = \mathrm{e}^{b_0},\ b = b_1)$

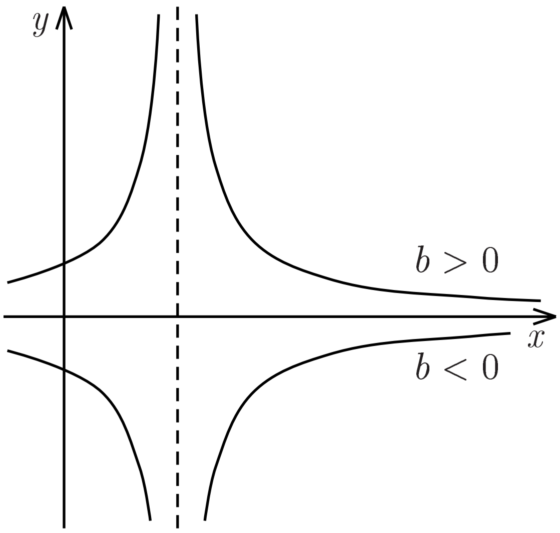
Зависимость $y = a\,\mathrm{e}^{bx}$ сводится к $Y = b_0 + b_1 X$ заменой $Y = \ln y,\ X = x$ $(a = \mathrm{e}^{b_0},\ b = b_1)$



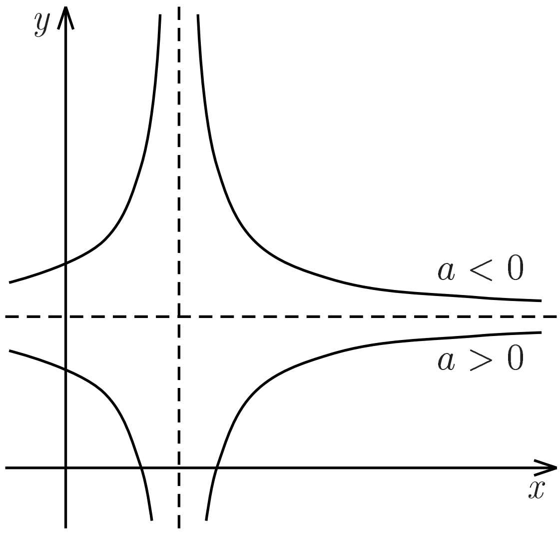
Зависимость $y = a + b\ln x$
сводится к $Y = b_0 + b_1 X$
заменой $Y = y,\ X = \ln x$
$(a = b_0,\ b = b_1)$



Зависимость $y = a + \frac{b}{x}$
сводится к $Y = b_0 + b_1 X$
заменой $Y = y,\ X = \frac{1}{x}$
$(a = b_0,\ b = b_1)$



Зависимость $y = \frac{1}{a+bx}$
сводится к $Y = b_0 + b_1 X$
заменой $Y = \frac{1}{y},\ X = x$
$(a = b_0,\ b = b_1)$



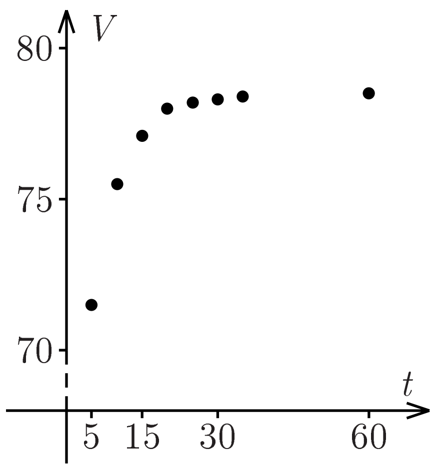
Зависимость $y = \frac{x}{a+bx}$
сводится к $Y = b_0 + b_1 X$
заменой $Y = \frac{x}{y},\ X = x$
$(a = b_0,\ b = b_1)$

Для проверки того, удачно ли выбран вид зависимости, следует построить новое корреляционное поле на плоскости OXY. Если вид зависимости y от x подобран правильно, то точки $(X_i; Y_i)$ будут располагаться вдоль прямой.

Задача 34. Установить зависимость количества выделившейся сыворотки (V, %) от времени (t, мин) при сычужном способе сквашивания молока.

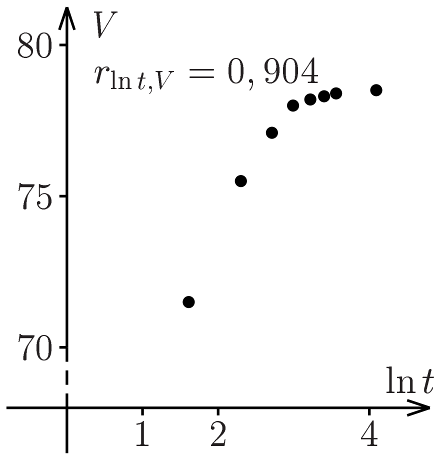
t	5	10	15	20	25	30	35	60
V	$71,5$	$75,5$	$77,1$	78	$78,2$	$78,3$	$78,4$	$78,5$

Решение. Построим корреляционное поле, отмечая по горизонтальной оси время, а по вертикальной — количество выделившейся сыворотки.



Предположим, что имеет место логарифмическая зависимость $V = a + b\ln t$. Сделаем замену переменных и построим новое корреляционное поле, отмечая по осям координат на плоскости OXY значения $X = \ln t$ и $Y = V$.

$X = \ln t$	$Y = V$
$1,61$	$71,5$
$2,3$	$75,5$
$2,71$	$77,1$
$2,99$	78
$3,22$	$78,2$
$3,4$	$78,3$
$3,56$	$78,4$
$4,09$	$78,5$

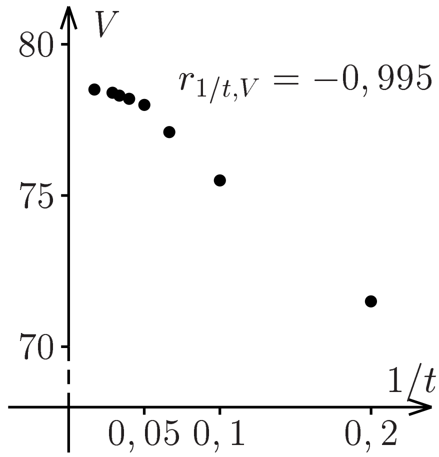


Хотя коэффициент корреляции $r_{\ln t,V} = 0,904$ довольно высок, точки корреляционного поля группируются вдоль какой-то кривой, а не прямой линии. Поэтому следует попытаться подобрать другую зависимость.

Поскольку с увеличением времени количество выделившейся сыворотки сперва резко возрастает, а затем практически стабилизируется, можно предположить, что искомая зависимость имеет вид $V = a + \frac{b}{t}$ или $V = \frac{t}{a+bt}$. Чтобы выбрать подходящую зависимость, построим новые корреляционные поля и рассчитаем коэффициенты корреляции, сделав соответствующие замены переменных.

Зависимость $V = a + \frac{b}{t}$ линеаризуется при помощи замены $Y = V,\ X = \frac{1}{t}$.

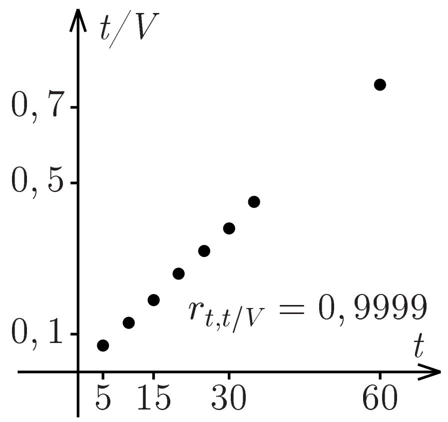
$X = \frac{1}{t}$	$Y = V$
$0,2$	$71,5$
$0,1$	$75,5$
$0,067$	$77,1$
$0,05$	78
$0,04$	$78,2$
$0,033$	$78,3$
$0,029$	$78,4$
$0,017$	$78,5$



Вычислив коэффициент корреляции между новыми переменными, получим $r_{1/t,V} = -0,995$, т. е. он близок по модулю к единице. По виду корреляционного поля, а также на основании значения коэффициента корреляции заключаем, что эта зависимость лучше аппроксимирует экспериментальные данные, чем логарифмическая.

Рассмотрим еще зависимость $V = \frac{t}{a+bt}$.

$X = t$	$Y = \frac{t}{V}$
5	$0,07$
10	$0,13$
15	$0,19$
20	$0,26$
25	$0,32$
30	$0,38$
35	$0,45$
60	$0,76$



Коэффициент корреляции между переменными $X = t$ и $Y = \frac{t}{V}$ очень высок: $r_{t,t/V} = 0,9999$, точки корреляционного поля практически идеально выстраиваются вдоль прямой. Следовательно, нужно определить коэффициенты зависимости $V = \frac{t}{a+bt}$.

Для этого определим по методу наименьших квадратов коэффициенты линейного уравнения регрессии $\widehat{Y} = b_0 + b_1 X$, где $X = t$, $Y = \frac{t}{V}$, $b_0 = a,\ b_1 = b$. Составим систему нормальных уравнений по формуле (17). Необходимые вспомогательные вычисления проведены в таблице.

$$\begin{cases} 8b_0 + 200b_1 = 2,56, \\ 200b_0 + 7100b_1 = 90,45. \end{cases}$$

№ п/п	X_i	Y_i	X_iY_i	X_i^2
1	5	0,07	0,35	25
2	10	0,13	1,3	100
3	15	0,19	2,85	225
4	20	0,26	5,2	400
5	25	0,32	8	625
6	30	0,38	11,4	900
7	35	0,45	15,75	1225
8	60	0,76	45,6	3600
$\sum$	200	2,56	90,45	7100

Выражая из первого уравнения b_0 и подставляя во второе, получим:

$$b_0 = \frac{2,56 - 200b_1}{8} = 0,32 - 25b_1;$$

$$200 \cdot (0,32 - 25b_1) + 7100b_1 = 90,45,$$

откуда $b_1 \approx 0,013$; $b_0 \approx 0,0051$.

Итак, линейное уравнение регрессии $Y = \frac{t}{V}$ на $X = t$ имеет вид $\widehat{Y} = 0,005 + 0,013X$, а искомая зависимость запишется как $\widehat{V} = \dfrac{t}{0,005 + 0,013t}$.

Задачи для самостоятельного решения

Задача 35. Установить зависимость частоты y сокращений сердца червя от температуры T, °C.

T	0	5	10	15	20	25	30
y	42	54	74	100	136	182	254

Ответ к задаче 35. **35.** $y = 40,8\,\mathrm{e}^{0,06T}$.

Множественная регрессия

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

$$\widehat{y} = b_0 + \sum_{j=1}^{l} b_j X_j + \sum_{j<t} b_{jt} X_j X_t + \sum_{j=1}^{l} b_{jj} X_j^2 + \ldots \tag{18}$$

Коэффициент b_0 называется **оценкой свободного члена уравнения регрессии**, коэффициенты b_j — **оценками линейных эффектов**, b_{jt} — **оценками эффектов взаимодействия**, b_{jj} — **оценками квадратичных эффектов**.

На практике обычно ограничиваются рассмотрением задачи определения коэффициентов **линейного**

$$\widehat{y} = b_0 + b_1X_1 + b_2X_2 + \ldots + b_lX_l \tag{19}$$

или **квадратичного**

$$\widehat{y} = b_0 + \sum_{j=1}^{l} b_jX_j + \sum_{j<t} b_{jt}X_jX_t + \sum_{j=1}^{l} b_{jj}X_j^2 \tag{20}$$

уравнения регрессии. В некоторых случаях получают уравнение вида

$$\widehat{y} = b_0 + \sum_{j=1}^{l} b_jX_j + \sum_{j<t} b_{jt}X_jX_t, \tag{21}$$

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 $X_1, X_2, \ldots, X_l$ требуется оценить коэффициенты эмпирического уравнения регрессии (18).

Поскольку уравнение (18) линейно относительно параметров b_0, b_j, b_{jt}, b_{jj}, ..., для оценки коэффициентов можно использовать МНК. Идея обобщения метода наименьших квадратов на случай регрессионной модели вида (18) 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:

$$X_{l+1} = X_1X_2,\ X_{l+2} = X_1X_3,\ \ldots \tag{22}$$

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

Тогда уравнение (18) будет записываться как однородное линейное уравнение

$$\widehat{y} = b_0X_0 + b_1X_1 + b_2X_2 + \ldots + b_kX_k. \tag{23}$$

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

$$
\begin{cases}
b_0 \sum\limits_{i=1}^{N} X_{0i}^2 + b_1 \sum\limits_{i=1}^{N} X_{0i}X_{1i} + b_2 \sum\limits_{i=1}^{N} X_{0i}X_{2i} + \ldots + b_k \sum\limits_{i=1}^{N} X_{0i}X_{ki} = \\
\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad = \sum\limits_{i=1}^{N} X_{0i}y_i, \\
b_0 \sum\limits_{i=1}^{N} X_{0i}X_{1i} + b_1 \sum\limits_{i=1}^{N} X_{1i}^2 + b_2 \sum\limits_{i=1}^{N} X_{1i}X_{2i} + \ldots + b_k \sum\limits_{i=1}^{N} X_{1i}X_{ki} = \\
\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad = \sum\limits_{i=1}^{N} X_{1i}y_i, \\
\ldots \\
b_0 \sum\limits_{i=1}^{N} X_{0i}X_{ki} + b_1 \sum\limits_{i=1}^{N} X_{1i}X_{ki} + b_2 \sum\limits_{i=1}^{N} X_{2i}X_{ki} + \ldots + b_k \sum\limits_{i=1}^{N} X_{ki}^2 = \\
\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad = \sum\limits_{i=1}^{N} X_{ki}y_i.
\end{cases}
\tag{24}
$$

Здесь X_{ji} 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 $b_0, b_1, b_2, \ldots, b_k$, следовательно, система имеет $k+1$ уравнение.

Запомнить систему нормальных уравнений можно следующим образом. Запишем уравнение (23) 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, получим второе уравнение системы. Если умножать 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 $b_0, b_1, b_2, \ldots, b_k$ входили в уравнение регрессии л и н е й н о и чтобы число оцениваемых коэффициентов было не больше числа р а з л и ч н ы х опытов. При вычислении коэффициентов нелинейного уравнения регрессии вводят замену переменных: обозначают через $X_0, X_1, X_2, \ldots, X_k$ те факторы или те функции от исходных факторов, при которых стоят неизвестные искомые коэффициенты $b_0, b_1, b_2, \ldots, b_k$. При этом необходимо, чтобы новые переменные были линейно независимы (т. е. их значения не могут быть пропорциональны, не могут линейно выражаться друг через друга).

Задача 36. По имеющимся результатам эксперимента получить (если это возможно):

№ оп.	X_{1i}	X_{2i}	y_i
1	1	0	6
2	0	2	1
3	1	−1	8
4	0	1	2

1) линейное уравнение регрессии;
2) уравнение вида $\widehat{y} = b_0 + b_{12}X_1X_2 + b_{22}X_2^2$;
3) квадратичное уравнение регрессии;
4) уравнение $\widehat{y} = b_0 + b_1X_1 + b_2X_2 + b_{11}X_1^2$.

Решение. 1) Линейное уравнение регрессии в случае двух факторов имеет вид $\widehat{y} = b_0 + b_1X_1 + b_2X_2$.

Для нахождения его коэффициентов составим систему нормальных уравнений по формуле (24). Запишем систему в общем виде, учитывая, что $X_{0i} = 1$ для всех опытов и число опытов $N = 4$:

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

Для вычисления нужных сумм составим таблицу.

№ оп.	X_{1i}	X_{2i}	y_i	X_{1i}^2	$X_{1i}X_{2i}$	$X_{1i}y_i$	X_{2i}^2	$X_{2i}y_i$
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
$\sum$	2	2	17	2	−1	14	6	−4

Запишем систему нормальных уравнений:

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

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

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

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

$$14b_0 + 11 \cdot \frac{45 - 8b_0}{6} = 80;$$
$$6 \cdot 14b_0 + 11 \cdot (45 - 8b_0) = 6 \cdot 80;$$
$$-4b_0 = -15;$$
$$b_0 = 3,75.$$

Следовательно,

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

Таким образом, искомая линейная зависимость имеет вид

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

2) Поскольку указанное уравнение имеет три неизвестных коэффициента, система нормальных уравнений будет содержать три уравнения. Запишем систему в общем виде с помощью формулы (24):

$$\begin{cases} 4b_0 + b_{12} \sum\limits_{i=1}^{N} X_{1i}X_{2i} + b_{22} \sum\limits_{i=1}^{N} X_{2i}^2 = \sum\limits_{i=1}^{N} y_i, \\ b_0 \sum\limits_{i=1}^{N} X_{1i}X_{2i} + b_{12} \sum\limits_{i=1}^{N} X_{1i}^2X_{2i}^2 + b_{22} \sum\limits_{i=1}^{N} X_{1i}X_{2i}^3 = \sum\limits_{i=1}^{N} X_{1i}X_{2i}y_i, \\ b_0 \sum\limits_{i=1}^{N} X_{2i}^2 + b_{12} \sum\limits_{i=1}^{N} X_{1i}X_{2i}^3 + b_{22} \sum\limits_{i=1}^{N} X_{2i}^4 = \sum\limits_{i=1}^{N} X_{2i}^2y_i. \end{cases}$$

Для вычисления нужных сумм составим таблицу.

№ оп.	X_{1i}	X_{2i}	y_i	$X_{1i}X_{2i}$	X_{2i}^2	$X_{1i}^2X_{2i}^2$	$X_{1i}X_{2i}^3$	$X_{1i}X_{2i}y_i$	X_{2i}^4	$X_{2i}^2y_i$
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
$\sum$	2	2	17	-1	6	1	-1	-8	18	14

Запишем систему нормальных уравнений:

$$\begin{cases} 4b_0 - b_{12} + 6b_{22} = 17, \\ -b_0 + b_{12} - b_{22} = -8, \\ 6b_0 - b_{12} + 18b_{22} = 14. \end{cases}$$

Решив систему, получим $b_0 \approx 4,73$, $b_{12} \approx -4,31$, $b_{22} \approx -1,04$.

Итак, искомая зависимость имеет вид

$$\widehat{y} = 4,73 - 4,31X_1X_2 - 1,04X_2^2.$$

3) Квадратичное уравнение регрессии в случае двух факторов имеет вид

$$\widehat{y} = b_0 + b_1X_1 + b_2X_2 + b_{12}X_1X_2 + b_{11}X_1^2 + b_{22}X_2^2,$$

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 $\widehat{y} = b_0 + b_1X_1 + b_2X_2 + b_{11}X_1^2$ содержит 4 коэффициента, однако по результатам данного эксперимента рассчитать значения этих коэффициентов невозможно, поскольку переменные $X_0 \equiv 1, X_1, X_2, X_3 = X_1^2$ линейно зависимы (значения X_1 и $X_3 = X_1^2$ во всех опытах совпадают).

Задачи для самостоятельного решения

Задача 37. По имеющимся результатам эксперимента получить (если это возможно):

№ оп.	X_{1i}	X_{2i}	y_i
1	-1	0	2
2	1	0	3
3	0	-1	4
4	0	1	5

1) линейное уравнение регрессии;
2) уравнение с парным взаимодействием;
3) уравнение $\widehat{y} = b_0+b_1X_1+b_2X_2+b_{11}X_1^2+b_{22}X_2^2$;
4) уравнение $\widehat{y} = b_0 + b_1X_1 + b_2X_2 + b_{11}X_1^2$.

Задача 38. Получить линейное уравнение регрессии по данным результатам эксперимента.

X_{1i}	-1	1	1	-1	-1	-1	1	1
X_{2i}	-1	-1	-1	1	1	1	1	1
y_i	1	2	3	4	6	8	10	12

Задача 39. Записать в общем виде через значения X_{1i}, X_{2i}, y_i систему нормальных уравнений для определения по результатам 12 опытов коэффициентов зависимости $\widehat{y} = b_0 + b_2X_2 + b_{12}X_1X_2$.

Задача 40. Записать в общем виде через значения X_{1i}, X_{2i}, X_{3i}, y_i систему нормальных уравнений для определения по результатам 15 опытов коэффициентов эмпирического уравнения $\widehat{y} = b_0 + b_2 \ln X_2 + b_{11}X_1^2 + b_{13}X_1X_3$.

Ответы к задачам 37–40. **37.** 1) $\widehat{y} = 3,5 + 0,5X_1 + 0,5X_2$. 2) Уравнение $\widehat{y} = b_0 + b_1X_1 + b_2X_2 + b_{12}X_1X_2$ получить невозможно, поскольку переменная $X_3 = X_1X_2$ принимает во всех опытах одно и то же значение 0. 3) Невозможно, т. к. количество опытов меньше количества неизвестных коэффициентов. 4) $\widehat{y} = 4,5 + 0,5X_1 + 0,5X_2 - 2X_1^2$. **38.** $\widehat{y} = 4,875 + 1,875X_1 + 3,5X_2$.

39. $$\begin{cases} 12b_0 + b_2\sum\limits_{i=1}^{N} X_{2i} + b_{12}\sum\limits_{i=1}^{N} X_{1i}X_{2i} = \sum\limits_{i=1}^{N} y_i, \\ b_0\sum\limits_{i=1}^{N} X_{2i} + b_2\sum\limits_{i=1}^{N} X_{2i}^2 + b_{12}\sum\limits_{i=1}^{N} X_{1i}X_{2i}^2 = \sum\limits_{i=1}^{N} X_{2i}y_i, \\ b_0\sum\limits_{i=1}^{N} X_{1i}X_{2i} + b_1\sum\limits_{i=1}^{N} X_{1i}X_{2i}^2 + b_{12}\sum\limits_{i=1}^{N} X_{1i}^2X_{2i}^2 = \sum\limits_{i=1}^{N} X_{1i}X_{2i}y_i. \end{cases}$$

40. $$\begin{cases} 15b_0 + b_2\sum\limits_{i=1}^{N} \ln X_{2i} + b_{11}\sum\limits_{i=1}^{N} X_{1i}^2 + b_{13}\sum\limits_{i=1}^{N} X_{1i}X_{3i} = \sum\limits_{i=1}^{N} y_i, \\ b_0\sum\limits_{i=1}^{N} \ln X_{2i} + b_2\sum\limits_{i=1}^{N} \ln^2 X_{2i} + b_{11}\sum\limits_{i=1}^{N} X_{1i}^2 \ln X_{2i} + b_{13}\sum\limits_{i=1}^{N} X_{1i}X_{3i}\ln X_{2i} = \\ \qquad = \sum\limits_{i=1}^{N} y_i \ln X_{2i}, \\ b_0\sum\limits_{i=1}^{N} X_{1i}^2 + b_2\sum\limits_{i=1}^{N} X_{1i}^2 \ln X_{2i} + b_{11}\sum\limits_{i=1}^{N} X_{1i}^4 + b_{13}\sum\limits_{i=1}^{N} X_{1i}^3X_{3i} = \sum\limits_{i=1}^{N} X_{1i}^2 y_i, \\ b_0\sum\limits_{i=1}^{N} X_{1i}X_{3i} + b_2\sum\limits_{i=1}^{N} X_{1i}X_{3i}\ln X_{2i} + b_{11}\sum\limits_{i=1}^{N} X_{1i}^3X_{3i} + b_{13}\sum\limits_{i=1}^{N} X_{1i}^2X_{3i}^2 = \\ \qquad = \sum\limits_{i=1}^{N} X_{1i}X_{3i}y_i. \end{cases}$$

Метод наименьших квадратов в матричной форме записи

При определении коэффициентов эмпирического уравнения регрессии
$$\widehat{y} = b_0X_0 + b_1X_1 + \ldots + b_kX_k$$
исходные данные могут быть представлены в виде двух матриц:
$$X_{N\times(k+1)} = \begin{pmatrix} X_{01} & X_{11} & \ldots & X_{k1} \\ X_{02} & X_{12} & \ldots & X_{k2} \\ \ldots & \ldots & \ldots & \ldots \\ X_{0N} & X_{1N} & \ldots & X_{kN} \end{pmatrix} \text{ и } Y_{N\times 1} = \begin{pmatrix} y_1 \\ y_2 \\ \ldots \\ y_N \end{pmatrix}.$$

Матрица X называется **матрицей базисных функций**, количество строк этой матрицы равно количеству опытов, а количество столбцов — количеству определяемых коэффициентов. В строках этой матрицы записаны наблюдаемые в соответствующих опытах значения переменных $X_0, X_1, \ldots, X_k$, входящих в уравнение регрессии. Матрица Y называется **столбцом наблюдений**.

Обозначим столбец неизвестных коэффициентов через

$$B_{(k+1)\times 1} = \begin{pmatrix} b_0 \\ b_1 \\ \dots \\ b_k \end{pmatrix}.$$

Тогда система нормальных уравнений (24) может быть записана в матричном виде как

$$X^T X B = X^T Y, \tag{25}$$

ее решение определяется по формуле

$$B = (X^T X)^{-1} X^T Y.$$

Для существования обратной матрицы $(X^T X)^{-1}$ требуется, чтобы переменные $X_0, X_1, \dots, X_k$ были линейно независимыми, т. е. в матрице базисных функций ни один из столбцов не должен быть линейной комбинацией остальных. Из этого требования, в частности, следует, что $k + 1 \leqslant N$, т. е. число определяемых коэффициентов должно быть не больше числа различных опытов.

На практике при отборе факторов для получения регрессионной зависимости по экспериментальным данным для проверки линейной независимости факторов вычисляют выборочные коэффициенты корреляции для каждой пары факторов. Если коэффициент корреляции между какими-то двумя факторами оказывается близким к ± 1, то в уравнение регрессии можно включить только один из факторов.

Понятие о планировании регрессионных экспериментов

Под **экспериментом** понимают совокупность опытов, проводимых в определенном порядке с целью получения информации об объекте исследования. **Опыт** — это воспроизведение исследуемого явления при определенных заданных условиях.

План эксперимента — это совокупность данных, определяющих число, условия и порядок проведения опытов.

Планирование эксперимента — это совокупность приемов, позволяющих исследователю выбрать план эксперимента так, чтобы получить максимум информации при минимуме затрат.

В основе планирования эксперимента лежит возможность проводить опыты при заданных исследователем условиях. Такой эксперимент называется **активным** в отличие от **пассивного** эксперимента, при котором исследователь не может управлять условиями, а имеет возможность только наблюдать за ходом процесса. К пассивному эксперименту относится, например, сбор данных о каком-либо процессе в режиме нормальной эксплуатации промышленного объекта. Из-за невозможности целенаправленно задавать условия проведения опытов эффективность пассивного эксперимента значительно ниже, чем активного.

Если целью эксперимента является получение зависимости отклика y от факторов $X_1, X_2, \ldots, X_k$ в виде линейного уравнения регрессии, то вся информация о плане эксперимента содержится в матрице базисных функций X (матрица составлена из значений факторов в каждом из опытов; столбцы соответствуют факторам, а строки — условиям проведения опытов). Эту матрицу иногда называют также **матрицей плана**. В случае построения зависимости более сложной, чем линейная, целесообразно различать понятия матрицы плана, содержащей значения изучаемых факторов в каждом из опытов, и матрицы базисных функций, включающей значения тех функций от факторов, которые присутствуют в уравнении регрессии.

Матрица $X^T X$, построенная по матрице базисных функций, называется **информационной матрицей плана** и играет фундаментальную роль в планировании регрессионных экспериментов (т. е. экспериментов, целью которых является получение уравнения регрессии). Матрица $X^T X$ не зависит от значений отклика y, но от ее свойств существенно зависит точность оценок коэффициентов b_j уравнения регрессии. Поэтому, выбирая специальным образом план эксперимента, можно получить хорошие статистические оценки коэффициентов уравнения регрессии независимо от наблюдаемых значений отклика y. Поскольку информационная матрица строится по матрице базисных функций, планирование регрессионного эксперимента проводится с учетом вида уравнения регрессии.

Проверка воспроизводимости эксперимента и расчет дисперсии воспроизводимости

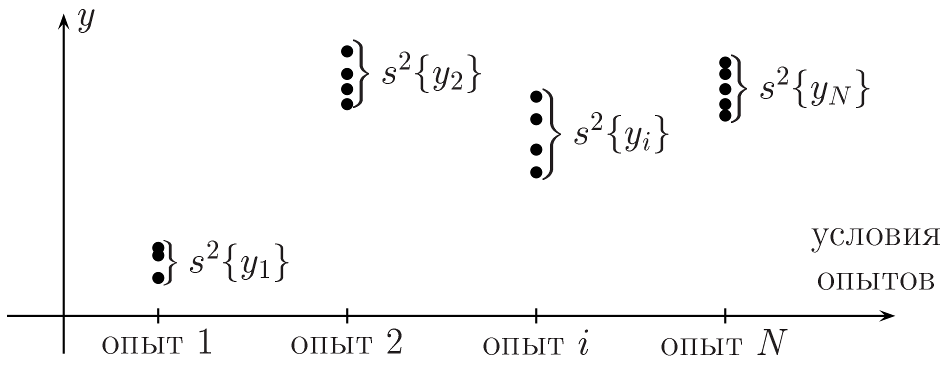
Для статистического анализа уравнения регрессии необходимо иметь количественную оценку ошибок эксперимента в целом, т. е. общую для всех опытов дисперсию параметра y, которую будем обозначать $s^2\{y\}$. Эта дисперсия называется **дисперсией воспроизводимости**.

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.

Пусть производится **дублирование опытов**, т. е. каждый из N различных опытов проводится соответственно $m_1, m_2, \ldots, m_N$ раз. Обозначим y_{iu} — наблюдаемое значение отклика при u-м повторении i-го опыта. Для каждого опыта нужно рассчитать

$$\overline{y_i} = \frac{1}{m_i}\sum_{u=1}^{m_i} y_{iu}, \quad s^2\{y_i\} = \frac{1}{m_i - 1}\sum_{u=1}^{m_i}\left(y_{iu} - \overline{y_i}\right)^2, \quad f_i = m_i - 1.$$

Здесь $\overline{y_i}$ — среднее значение отклика, наблюдаемое в i-м опыте (при i-м комплексе условий), дисперсия $s^2\{y_i\}$ характеризует степень разброса значений отклика в i-м опыте, ее число 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, т. е.

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

В случае **равномерного дублирования опытов**, когда все опыты повторяются одинаковое количество раз, эта проверка проводится по критерию Кохрена (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.

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:

$$s^2\{y\} = \frac{f_1 s^2\{y_1\} + f_2 s^2\{y_2\} + \ldots + f_N s^2\{y_N\}}{f_1 + f_2 + \ldots + f_N} = \frac{\sum\limits_{i=1}^{N} f_i s^2\{y_i\}}{\sum\limits_{i=1}^{N} f_i}, \quad (26)$$

ее число степеней свободы равно $f_{\text{воспр}} = f_1 + f_2 + \ldots + f_N$.

Как правило, планируется равномерное дублирование опытов, т. е. $m_1 = m_2 = \ldots = m_N = m$. В этом случае формула (26) для расчета дисперсии воспроизводимости упрощается:

$$s^2\{y\} = \frac{1}{N}\sum_{i=1}^{N} s^2\{y_i\}, \qquad f_{\text{воспр}} = N(m-1). \quad (27)$$

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.

Примеры решения задач

Задача 41. Оценить воспроизводимость эксперимента, результаты которого приведены в задаче **38.**

Решение. Рассмотрим, при каких условиях (при каких значениях факторов X_1 и 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.

№ оп.	X_{1i}	X_{2i}	y_{i1}	y_{i2}	y_{i3}	$\overline{y_i}$	$s^2\{y_i\}$	f_i
1	-1	-1	1	—	—	1	—	—
2	1	-1	2	3	—	2, 5	0, 5	1
3	-1	1	4	6	8	6	4	2
4	1	1	10	12	—	11	2	1

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:

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

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

$$\overline{y_3} = \frac{1}{3} \cdot (4 + 6 + 8) = 6;$$

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

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

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

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 (26):

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

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

Задачи для самостоятельного решения

В задачах **42**–**45** требуется:
1) проверить гипотезу о воспроизводимости эксперимента при уровне значимости $\alpha = 0,05$; 2) сделать вывод о целесообразности расчета дисперсии воспроизводимости;
3) рассчитать дисперсию воспроизводимости и найти ее число степеней свободы.

Задача 42.

№ оп.	y_{i1}	y_{i2}	y_{i3}	y_{i4}	y_{i5}
1	11	8	8	5	—
2	3	5	7	1	4
3	4	8	12	—	—

Задача 43.

№ оп.	y_{i1}	y_{i2}	y_{i3}
1	15	17	22
2	10	12	14
3	12	16	14
4	18	22	14
5	10	6	14

Задача 44.

№ оп.	$s^2\{y_i\}$	f_i
1	120	8
2	55	8
3	85	8
4	90	8
5	65	8

Задача 45.

№ оп.	$s^2\{y_i\}$	f_i
1	230	20
2	140	12
3	180	10
4	212	8
5	80	15

Ответы к задачам 42–45. **42.** 1) эксперимент воспроизводим, $F_{\text{расч}} = 3,2 < F_{0,025;2;4} = 10,7$; 3) $s^2\{y\} = 7,78$, $f_{\text{воспр}} = 9$. **43.** 1) воспроизводимость эксперимента может быть проверена по критерию Кохрена: $G_{\text{расч}} = 0,3 < G_{0,05;2;5} = 0,6838$, или по критерию Фишера: $F_{\text{расч}} = 4 < F_{0,025;2;2} = 39$; эксперимент воспроизводим; 3) $s^2\{y\} = 10,6$, $f_{\text{воспр}} = 10$. **44.** 1) воспроизводимость эксперимента может быть проверена по критерию Кохрена: $G_{\text{расч}} = 0,29 < G_{0,05;8;5} = 0,4387$, или по критерию Фишера: $F_{\text{расч}} = 2,18 < F_{0,025;8;8} = 4,43$; эксперимент воспроизводим; 3) $s^2\{y\} = 83$, $f_{\text{воспр}} = 40$. **45.** 1) $F_{0,025;20;15} = 2,76$; эксперимент не воспроизводим, расчет дисперсии воспроизводимости нецелесообразен; 3) $s^2\{y\} = 168,87$, $f_{\text{воспр}} = 65$.

Проверка адекватности эмпирического уравнения регрессии

После определения по результатам эксперимента коэффициентов эмпирического уравнения регрессии нужно проверить адекватность 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 $m_1, m_2, \ldots, m_N$ раз, получено уравнение регрессии

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

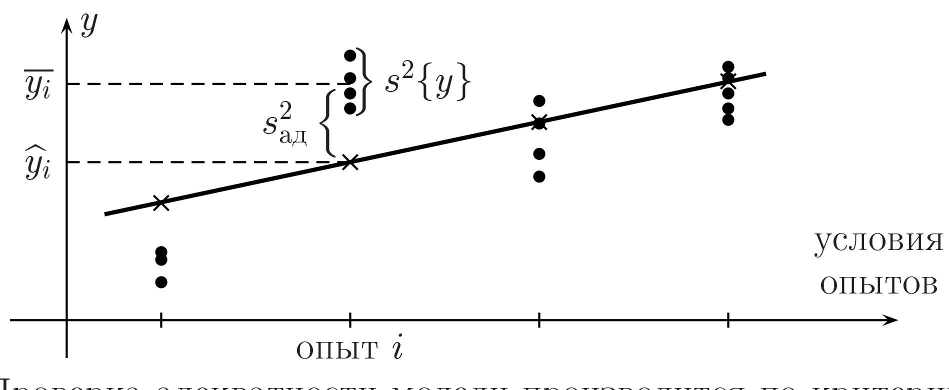
Обозначим $\widehat{y_i} = b_0 X_{0i} + b_1 X_{1i} + \ldots + b_k X_{ki}$ — значение отклика, предсказываемое по полученному уравнению регрессии для i-го опыта; $\overline{y_i}$ — среднее наблюдаемых значений отклика в i-м опыте.

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

$$s^2_{\text{ад}} = \frac{1}{N-d} \sum_{i=1}^{N} m_i \left(\widehat{y_i} - \overline{y_i}\right)^2, \tag{28}$$

число степеней свободы дисперсии адекватности равно $f_{\text{ад}} = N - d$, где N — число различных опытов; d — число коэффициентов проверяемого уравнения регрессии, которые определялись по результатам эксперимента; m_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 значительные расхождения между $\widehat{y_i}$ и результатами эксперимента, то дисперсия адекватности, как правило, значительно больше, чем дисперсия воспроизводимости. Гипотеза об адекватности модели при заданном уровне значимости α принимается, если

$$F_{\text{расч}} = \frac{s^2_{\text{ад}}}{s^2\{y\}} < F_{\text{табл}} = F_{\alpha; f_{\text{ад}}; f_{\text{воспр}}}. \tag{29}$$

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 $\boxed{N > d}$.

Примеры решения задач

Задача 46. В задаче **38** проверить адекватность линейного уравнения регрессии при уровне значимости $\alpha = 0,05$.

Решение. В условии задачи **38** 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 **38** получено линейное уравнение регрессии:

$$\widehat{y} = 4,875 + 1,875X_1 + 3,5X_2.$$

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

$$\widehat{y}_1 = 4,875 + 1,875 \cdot (-1) + 3,5 \cdot (-1) = -0,5;$$

$$\widehat{y}_2 = 4,875 + 1,875 \cdot 1 + 3,5 \cdot (-1) = 3,25;$$

$$\widehat{y}_3 = 4,875 + 1,875 \cdot (-1) + 3,5 \cdot 1 = 6,5;$$

$$\widehat{y}_4 = 4,875 + 1,875 \cdot 1 + 3,5 \cdot 1 = 10,25.$$

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

№ оп.	X_{1i}	X_{2i}	y_{i1}	y_{i2}	y_{i3}	$\overline{y_i}$	$\widehat{y}_i$	$(\widehat{y}_i - \overline{y_i})^2$	m_i
1	-1	-1	1	—	—	1	$-0,5$	$2,25$	1
2	1	-1	2	3	—	$2,5$	$3,25$	$0,5625$	2
3	-1	1	4	6	8	6	$6,5$	$0,25$	3
4	1	1	10	12	—	11	$10,25$	$0,5625$	2

Вычислим дисперсию адекватности по формуле (29):

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

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

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 47. В пункте 2 задачи **36** 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 **36** 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 (27). Следовательно, дисперсия $s^2\{y\} = 2$ имеет число степеней свободы $f_{\text{воспр}} = 4 \cdot (3 - 1) = 8$.

В пункте 2 задачи **36** получено следующее уравнение регрессии:

$$\widehat{y} = 4,73 - 4,31X_1X_2 - 1,04X_2^2.$$

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

$$\widehat{y}_1 = 4,73 - 4,31 \cdot 1 \cdot 0 - 1,04 \cdot 0^2 = 4,73;$$

$$\widehat{y}_2 = 4,73 - 4,31 \cdot 0 \cdot 2 - 1,04 \cdot 2^2 = 0,57;$$

$$\widehat{y}_3 = 4,73 - 4,31 \cdot 1 \cdot (-1) - 1,04 \cdot (-1)^2 = 8;$$

$$\widehat{y}_4 = 4,73 - 4,31 \cdot 0 \cdot 1 - 1,04 \cdot 1^2 = 3,69.$$

Данные эксперимента и результаты расчетов занесем в таблицу.

№ оп.	X_{1i}	X_{2i}	$\overline{y_i}$	$\widehat{y}_i$	$(\widehat{y}_i - \overline{y_i})^2$
1	1	0	6	4,73	1,6129
2	0	2	1	0,57	0,1849
3	1	−1	8	8	0
4	0	1	2	3,69	2,8591

Вычислим дисперсию адекватности по формуле (29), учитывая, что число повторений всех опытов равно $m_i = 3$:

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

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

Проверим гипотезу об адекватности модели по критерию Фишера. Поскольку

$$F_{\text{расч}} = \frac{13,9617}{2} \approx 6,98 > F_{\text{табл}} = F_{0,05;1;8} = 5,32,$$

то полученное уравнение регрессии следует признать неадекватным.

Задачи для самостоятельного решения

Задача 48. В пункте 1 задачи **36** 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 49. В пункте 1 задачи **37** 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 50. Проверьте при уровне значимости $\alpha = 0,05$ адекватность уравнения $\widehat{y} = 0,45 - 0,09X + 0,8X^2$, полученного по данным в таблице 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 48–50. **48.** $s^2_{\text{ад}} = 0,25$, $f_{\text{ад}} = 4 - 3 = 1$; $s^2\{y\} = 0,1$, $f_{\text{воспр}} = 5 - 1 = 4$; $F_{\text{расч}} = 2,5 < F_{0,05;1;4} = 7,71$. Модель адекватна. **49.** $s^2_{\text{ад}} = 16$, $f_{\text{ад}} = 4 - 3 = 1$; $s^2\{y\} = 1$, $f_{\text{воспр}} = 4 \cdot (4 - 1) = 12$; $F_{\text{расч}} = 16 > F_{0,05;1;12} = 4,75$. Модель неадекватна. **50.** $s^2_{\text{ад}} = 1,44$, $f_{\text{ад}} = 5 - 3 = 2$; $s^2\{y\} = 0,148$, $f_{\text{воспр}} = 2 + 1 + 2 = 5$; $F_{\text{расч}} = 9,73 > F_{0,05;2;5} \approx \frac{1}{2}(6,94 + 5,14) = 6,04$. Модель неадекватна.

Раздел 3. МЕТОДЫ ПЛАНИРОВАНИЯ МНОГОФАКТОРНЫХ РЕГРЕССИОННЫХ ЭКСПЕРИМЕНТОВ

Понятие о критериях оптимальности планов регрессионных экспериментов

Регрессионным называется эксперимент, целью которого является получение регрессионной модели. Среди планов регрессионных экспериментов выделяют планы 1-го и 2-го порядков.

Планы 1-го порядка предназначены для получения линейного уравнения регрессии (19). К планам 1-го порядка относятся планы полного факторного эксперимента (ПФЭ) типа 2^k и дробного факторного эксперимента (ДФЭ) типа 2^{k-p}, симплекс-планы.

Планы 2-го порядка предназначены для получения квадратичного уравнения регрессии (20). К таким планам относятся ПФЭ типа 3^k, центрально-композиционные ортогональный и ротатабельный планы (ЦКОП и ЦКРП), B-план и другие.

В основе планирования эксперимента лежит возможность проводить опыты при заданных исследователем условиях, т. е. осуществлять активный эксперимент. Планирование эксперимента — это совокупность приемов, позволяющих исследователю выбрать план эксперимента, т. е. число, порядок и условия проведения опытов, так, чтобы получить максимум информации при минимуме затрат.

При выборе плана эксперимента придерживаются **принципа оптимальности планирования**: план эксперимента должен обладать некоторыми оптимальными свойствами, быть в каком-то смысле наилучшим из возможных планов эксперимента. Критерии оптимальности планов эксперимента могут выбираться по-разному, не существует одного, наилучшего во всех смыслах плана. Известные критерии оптимальности планов фактически формализуют, переводят на математический язык те или иные интуитивные представления специалистов-экспериментаторов о качественном, хорошем эксперименте.

Выбор критерия оптимальности плана эксперимента зависит прежде всего от решаемой задачи. Критерии оптимальности планов регрессионных экспериментов принято подразделять на две группы: 1) критерии, связанные с точностью оценивания коэффициентов регрессии; 2) критерии, связанные с предсказательными свойствами эмпирической модели (насколько точно можно предсказать значения отклика по уравнению регрессии).

В настоящее время известно несколько десятков различных критериев оптимальности планов регрессионного анализа. Назовем некоторые наиболее употребительные из них.

План регрессионного эксперимента называется **ортогональным**, если матрица $(X^T X)^{-1}$ диагональна (напомним, что X — матрица базисных функций). Название объясняется тем, что в этом случае столбцы матрицы X ортогональны, т. е. для любых двух столбцов сумма попарных произведений их элементов равна нулю.

Ортогональные планы часто стремятся использовать, поскольку при таком плане эксперимента оценки коэффициентов регрессии получаются независимыми, что существенно облегчает их анализ и интерпретацию. Кроме того, значительно упрощается расчет коэффициентов регрессии (не требуется составлять и решать систему нормальных уравнений).

Среди названных выше планов регрессионных экспериментов ортогональными являются планы ПФЭ 2^k, ДФЭ 2^{k-p}, ЦКОП.

Одним из наиболее важных и часто используемых является критерий **D-оптимальности**, согласно которому минимизируется обобщенная дисперсия оценок коэффициентов регрессии (план обладает свойством D-оптимальности, если он имеет минимальный определитель матрицы $(X^T X)^{-1}$). D-оптимальными являются планы ПФЭ 2^k и ДФЭ 2^{k-p}; B-план 2-го порядка близок к D-оптимальному.

План эксперимента называется **ротатабельным**, если точность предсказания значений отклика по уравнению регрессии одинакова во всех равноудаленных от центра плана точках. Планы ПФЭ 2^k и ДФЭ 2^{k-p} обладают свойством ротатабельности в случае линейных регрессионных моделей, из планов 2-го порядка ротатабельным является ЦКРП.

Отметим также одно свойство планов, которое является весьма желательным с точки зрения минимизации затрат на проведение эксперимента. Это степень насыщенности плана. План называется **насыщенным**, если число опытов равно числу определяемых коэффициентов уравнения регрессии. Желательно, чтобы любой реальный план был близок к насыщенному. Это особенно важно на этапе предварительного исследования, когда требуется получить приближенное представление об объекте при минимальных затратах на проведение эксперимента.

Построение линейной модели с помощью полного факторного эксперимента типа 2^k

Полным факторным экспериментом (ПФЭ) называется эксперимент, в котором реализуются все возможные комбинации факторов на всех выбранных для исследования уровнях.

Во многих случаях целью регрессионного эксперимента является получение линейной модели. В частности, в начале исследования, когда об изучаемой зависимости почти ничего неизвестно, стремятся получить наиболее простое — линейное уравнение регрессии.

Для нахождения эмпирического линейного уравнения регрессии

$$\widehat{y} = b_0 + b_1 X_1 + \ldots + b_k X_k$$

каждый из факторов должен варьироваться по крайней мере на двух уровнях. План полного факторного эксперимента, в котором число уровней каждого фактора равно двум, обозначается как **ПФЭ типа** 2^k, где k — число факторов. Число различных опытов в таком эксперименте равно $N = 2^k$.

Рассмотрим ПФЭ 2^2, т. е. эксперимент, в котором реализуются все возможные комбинации двух факторов на двух уровнях — верхнем и нижнем. Число опытов $N = 2^2 = 4$. Перечислим эти опыты: в первом опыте оба фактора фиксируются на нижнем уровне; во втором — первый на верхнем, а второй на нижнем уровне; в третьем опыте первый фактор будет на нижнем, а второй на верхнем уровне; в четвертом опыте оба фактора устанавливаются на верхнем уровне.

№ оп.	X_1	X_2
1	$X_1^{\min}$	$X_2^{\min}$
2	$X_1^{\max}$	$X_2^{\min}$
3	$X_1^{\min}$	$X_2^{\max}$
4	$X_1^{\max}$	$X_2^{\max}$

Для упрощения планирования и расчетов перейдем от факторов X_j в натуральном масштабе к **кодированным (нормализованным, стандартизированным)** переменным

$$x_j = \frac{X_j - X_j^0}{\Delta X_j}, \tag{30}$$

где формулами

$$X_j^0 = \frac{X_j^{\max} + X_j^{\min}}{2}, \qquad \Delta X_j = \frac{X_j^{\max} - X_j^{\min}}{2} \tag{31}$$

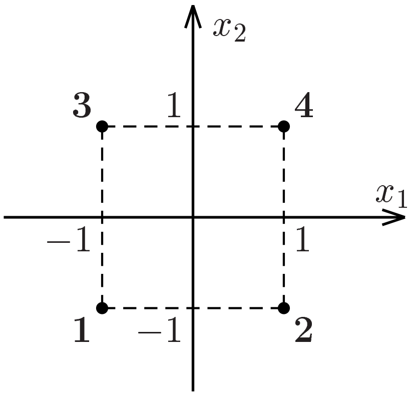
определяются соответственно **основной (базовый, нулевой) уровень** и **шаг варьирования (интервал варьирования)** фактора X_j.

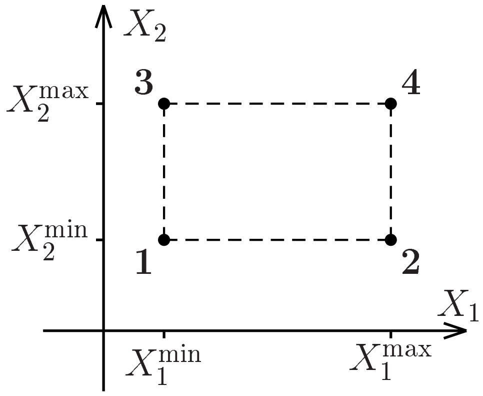
Точка $(X_1^0; X_2^0; \ldots; X_k^0)$ называется **центром плана**. Формулы (30) называются **формулами перехода от натуральных значений факторов к кодированным**.

При замене (30) получаем:

$$X_j = X_j^{\max} \to x_j = +1;$$
$$X_j = X_j^{\min} \to x_j = -1.$$

Полезно иметь в виду геометрическую интерпретацию плана ПФЭ. На факторной плоскости Ox_1x_2 (в кодированных переменных) условия опытов ПФЭ 2^2 образуют квадрат с центром в начале координат. На факторной плоскости OX_1X_2 (в натуральных переменных) условия этих опытов образуют прямоугольник.





При планировании ПФЭ типа 2^k в качестве центра плана обычно выбирают такую комбинацию уровней факторов, которая соответствует начальным условиям эксперимента, определенным из априорной информации. Интервал варьирования для каждого фактора выбирают так, чтобы: 1) значения верхнего и нижнего уровней не выходили за пределы области определения фактора, т. е. могли быть реализованы в эксперименте; 2) интервал варьирования фактора не был меньше ошибки измерения (фиксации) фактора.

При составлении матрицы плана ПФЭ типа 2^k в кодированных переменных обычно придерживаются следующих правил:

1) уровни варьирования 1-го фактора чередуются от опыта к опыту;

2) частота смены уровней каждого последующего фактора вдвое ниже, чем предыдущего.

№ оп.	x_0	x_1	x_2	x_3	…
1	1	−1	−1	−1	
2	1	+1	−1	−1	
3	1	−1	+1	−1	
4	1	+1	+1	−1	
5	1	−1	−1	+1	
6	1	+1	−1	+1	
7	1	−1	+1	+1	
8	1	+1	+1	+1	
…	…	…	…	…	…

Замечание. Матрицы планов в таблицах указывают только условия проведения опытов. Последовательность постановки опытов не должна соответствовать нумерации опытов, поскольку возможно накопление систематических ошибок. Например, за счет неучтенных факторов, которые оказывают влияние на отклик и действуют по-разному в начале и конце эксперимента. На практике применяется **рандомизация**, т. е. проведение запланированных опытов в случайной последовательности. Для установления этой последовательности пользуются таблицей случайных чисел.

Свойства матрицы плана ПФЭ 2^k в кодированных переменных.

1. Симметричность относительно центра эксперимента: сумма элементов каждого столбца, соответствующего фактору, равна 0, т. е. $\sum\limits_{i=1}^{N} x_{ji} = 0,\ 1 \leqslant j \leqslant k$.

2. Свойство нормировки: сумма квадратов элементов каждого столбца (включая столбец, соответствующий фиктивному фактору x_0) равна числу опытов, т. е. $\sum\limits_{i=1}^{N} x_{ji}^2 = N,\ 0 \leqslant j \leqslant k$.

3. Свойство ортогональности: сумма попарных произведений элементов различных столбцов равна 0, т. е. $\sum\limits_{i=1}^{N} x_{ji}x_{ti} = 0,\ j \neq t$.

Свойства симметричности и ортогональности приводят к диагональной информационной матрице $X^T X$, а следовательно, к ортогональному плану, что позволяет получить независимые оценки коэффициентов регрессии, а также существенно облегчает расчет коэффициентов уравнения регрессии.

Рассмотрим для простоты случай $k = 2$. По результатам $N = 2^2 = 4$ опытов требуется получить линейное уравнение регрессии в кодированных переменных, т. е. $\widehat{y} = b_0 + b_1x_1 + b_2x_2$. Система нормальных уравнений метода наименьших квадратов имеет вид

$$\begin{cases} b_0 \sum\limits_{i=1}^{N} x_{0i}^2 + b_1 \sum\limits_{i=1}^{N} x_{0i}x_{1i} + b_2 \sum\limits_{i=1}^{N} x_{0i}x_{2i} = \sum\limits_{i=1}^{N} x_{0i}y_i, \\ b_0 \sum\limits_{i=1}^{N} x_{0i}x_{1i} + b_1 \sum\limits_{i=1}^{N} x_{1i}^2 + b_2 \sum\limits_{i=1}^{N} x_{1i}x_{2i} = \sum\limits_{i=1}^{N} x_{1i}y_i, \\ b_0 \sum\limits_{i=1}^{N} x_{0i}x_{2i} + b_1 \sum\limits_{i=1}^{N} x_{1i}x_{2i} + b_2 \sum\limits_{i=1}^{N} x_{2i}^2 = \sum\limits_{i=1}^{N} x_{2i}y_i. \end{cases}$$

В силу свойств симметричности, ортогональности и нормировки матрицы базисных функций

$$X = \begin{pmatrix} 1 & -1 & -1 \\ 1 & +1 & -1 \\ 1 & -1 & +1 \\ 1 & +1 & +1 \end{pmatrix}$$

получим:

$$\begin{cases} 4b_0 = \sum\limits_{i=1}^{N} x_{0i}y_i, \\ 4b_1 = \sum\limits_{i=1}^{N} x_{1i}y_i, \\ 4b_2 = \sum\limits_{i=1}^{N} x_{2i}y_i; \end{cases} \Rightarrow \begin{cases} b_0 = \frac{1}{4}\sum\limits_{i=1}^{N} x_{0i}y_i, \\ b_1 = \frac{1}{4}\sum\limits_{i=1}^{N} x_{1i}y_i, \\ b_2 = \frac{1}{4}\sum\limits_{i=1}^{N} x_{2i}y_i. \end{cases}$$

Аналогично в общем случае ПФЭ типа 2^k коэффициенты линейного уравнения регрессии в кодированных переменных рассчитываются независимо друг от друга по следующей простой формуле:

$$b_j = \frac{1}{N}\sum_{i=1}^{N} x_{ji}y_i, \tag{32}$$

где $N = 2^k$ — число опытов ПФЭ; x_{ji} — кодированное значение j-го фактора в i-м опыте; y_i — значение отклика в i-м опыте.

Замечание. В случае ПФЭ 2^k с равномерным дублированием опытов, когда каждый опыт повторяется m раз, расчет коэффициентов линейного уравнения регрессии в кодированных переменных осуществляется по формуле

$$b_j = \frac{1}{N}\sum_{i=1}^{N} x_{ji}\overline{y_i}, \tag{33}$$

где $\overline{y_i}$ — среднее значение отклика в i-м опыте. При неравномерном дублировании опытов эти формулы неприменимы и для определения коэффициентов уравнения регрессии необходимо составлять систему нормальных уравнений (24).

Отметим свойства плана ПФЭ 2^k. План ПФЭ типа 2^k является ортогональным и D-оптимальным при данном количестве факторов k, числе опытов $N = 2^k$ и заданном виде уравнения регрессии, а также обладает важным свойством ротатабельности, т. е. способностью линейной модели, полученной в результате эксперимента, предсказывать значения параметра оптимизации с одинаковой точностью на равных расстояниях от центра эксперимента независимо от выбранного направления. Кроме того, план ПФЭ 2^k удовлетворяет и другим критериям оптимальности.

Примеры решения задач

Задача 51. Предварительными исследованиями установлено, что важнейшими факторами гидролиза древесной массы, влияющими на предел прочности при изгибе y, МПа, являются: температура X_1, °C; время X_2, мин; кислотность древесной массы X_3, pH. Предполагается провести ПФЭ 2^3 для установления зависимости y от указанных факторов в пределах: $20 \le X_1 \le 60$; $10 \le X_2 \le 60$; $4,5 \le X_3 \le 5,2$. При каких условиях должны быть проведены опыты? Определить центр плана и интервалы варьирования факторов.

Решение. Имеем случай $k = 3$ факторов. ПФЭ типа 2^3 содержит $2^3 = 8$ опытов. Определим центр плана и интервалы варьирования факторов по формулам (31). По условию задачи верхний уровень фактора X_1 равен $X_1^{\max} = 60$°C, нижний $X_1^{\min} = 20$°C. Следовательно, основной (базовый) уровень и интервал варьирования

$$X_1^0 = \frac{X_1^{\max} + X_1^{\min}}{2} = 40\text{°C}; \qquad \Delta X_1 = \frac{X_1^{\max} - X_1^{\min}}{2} = 20\text{°C}.$$

Аналогично для факторов X_2 и X_3 имеем:

$$X_2^{\max} = 60 \text{ мин}; X_2^{\min} = 10 \text{ мин}; X_2^0 = 35 \text{ мин}; \Delta X_2 = 25 \text{ мин};$$

$$X_3^{\max} = 5,2 \text{ pH}; X_3^{\min} = 4,5 \text{ pH}; X_3^0 = 4,85 \text{ pH}; \Delta X_3 = 0,35 \text{ pH}.$$

Факторы	X_1	X_2	X_3
X_j^0	40	35	4,85
ΔX_j	20	25	0,35
$X_j^{\max}$	60	60	5,2
$X_j^{\min}$	20	10	4,5

Уровни и интервалы варьирования факторов занесем в таблицу. Переход от натуральных переменных X_1, X_2, X_3 к кодированным x_1, x_2, x_3, которые принимают только значения $+1$ и -1, задается формулами (30), которые в нашем случае примут вид:

$$x_1 = \frac{X_1 - 40}{20}; \quad x_2 = \frac{X_2 - 35}{25}; \quad x_3 = \frac{X_3 - 4,85}{0,35}.$$

Запишем в таблицу матрицу плана ПФЭ типа 2^3 в кодированных переменных (знак «$-$» означает, что в данном опыте соответствующий фактор фиксируется на нижнем уровне, знак «$+$» означает верхний уровень фактора) и укажем соотвествующие натуральные значения факторов.

ПФЭ	Кодированные значения			Натуральные значения		
№ оп.	x_1	x_2	x_3	X_1	X_2	X_3
1	−	−	−	20	10	4,5
2	+	−	−	60	10	4,5
3	−	+	−	20	60	4,5
4	+	+	−	60	60	4,5
5	−	−	+	20	10	5,2
6	+	−	+	60	10	5,2
7	−	+	+	20	60	5,2
8	+	+	+	60	60	5,2

Таким образом, опыты нужно провести при следующих условиях: первый опыт — температура 20°C, время 10 мин, кислотность 4,5 pH; во втором опыте надо установить температуру 60°C, время 10 мин, кислотность 4,5 pH; в третьем опыте температура 20°C, время 60 мин, кислотность 4,5 pH; в четвертом опыте температура 60°C, время 60 мин, кислотность 4,5 pH. Далее такие же опыты повторяются при кислотности 5,2 pH.

Задача 52. Получить линейное уравнение регрессии по данным результатам эксперимента.

x_1	−1	+1	−1	+1	−1	+1	−1	+1
x_2	−1	−1	+1	+1	−1	−1	+1	+1
y	4	6	8	10	2	4	10	6

Решение. Имеем результаты эксперимента с двумя факторами, каждый из которых принимает только два значения: +1 и −1. Проведено $N = 4$ различных опыта, каждый повторен $m = 2$ раза: при одинаковых условиях проведены соответственно первый и пятый, второй и шестой, третий и седьмой, четвертый и восьмой опыты. Таким образом, убеждаемся, что даны результаты ПФЭ типа 2^2 в кодированных переменных с равномерным дублированием опытов. В этом случае коэффициенты линейного уравнения регрессии $\widehat{y} = b_0 + b_1x_1 + b_2x_2$ рассчитываются по формуле (33).

Для удобства расчетов составим таблицу, в которой строки будут соответствовать различным опытам, и вычислим средние значения $\overline{y_i}$ отклика в каждом из опытов.

№ оп.	x_{0i}	x_{1i}	x_{2i}	y_{i1}	y_{i2}	$\overline{y_i}$
1	+	−	−	4	2	3
2	+	+	−	6	4	5
3	+	−	+	8	10	9
4	+	+	+	10	6	8
b_j	6,25	0,25	2,25			

Подставляя в (33) средние значения отклика и кодированные значения (знаки) факторов из таблицы, получим:

$$b_0 = \frac{1}{4}(3 + 5 + 9 + 8) = 6,25;$$

$$b_1 = \frac{1}{4}(-3 + 5 - 9 + 8) = 0,25;$$

$$b_2 = \frac{1}{4}(-3 - 5 + 9 + 8) = 2,25.$$

Следовательно, линейное уравнение регрессии имеет вид

$$\widehat{y} = 6,25 + 0,25x_1 + 2,25x_2.$$

Задача 53. Получить линейное уравнение регрессии по данным результатам эксперимента.

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. Имеем результаты эксперимента с тремя факторами, каждый из которых принимает только два значения. Проведено $N = 8$ различных опытов. Следовательно, даны результаты ПФЭ типа 2^3 в натуральных переменных без дублирования опытов.

Определим центр плана и интервалы варьирования факторов по формулам (31):

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

$$X_2^{\max} = 80;\ X_2^{\min} = 20;\ X_2^0 = \frac{80 + 20}{2} = 50;\ \Delta X_2 = \frac{80 - 20}{2} = 30;$$

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

Переход от натуральных переменных X_1, X_2, X_3 к кодированным x_1, x_2, x_3 задается формулами:

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

№ оп.	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
b_j	4,75	−0,25	1,25	1,25	

Запишем результаты эксперимента в таблицу, перейдя к кодированным переменным и сохранив порядок следования опытов (фактор X_1 в первом опыте фиксируется на верхнем уровне, во втором — на нижнем и т. д.). Коэффициенты линейного уравнения регрессии в кодированных переменных $\widehat{y} = b_0 + b_1x_1 + b_2x_2 + b_3x_3$ определим по формуле (32):

$$b_0 = \frac{1}{8}(2+3+4+5+4+5+8+7) = 4,75;$$

$$b_1 = \frac{1}{8}(2-3+4-5+4-5+8-7) = -0,25;$$

$$b_2 = \frac{1}{8}(-2-3+4+5-4-5+8+7) = 1,25;$$

$$b_3 = \frac{1}{8}(-2-3-4-5+4+5+8+7) = 1,25.$$

Следовательно, линейное уравнение регрессии в кодированных переменных примет вид

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

Подставив формулы, связывающие натуральные и кодированные значения факторов, получим линейное уравнение регрессии в натуральных переменных:

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

Задача 54. Получить линейное уравнение регрессии по данным результатам эксперимента.

x_1	-1	$+1$	-1	$+1$	-1	$+1$
x_2	-1	-1	$+1$	$+1$	$+1$	-1
y	2	3	4	5	5	2

Решение. Имеем результаты эксперимента с двумя факторами, каждый из которых принимает только два значения: $+1$ и -1. Проведено 6 опытов, из них 4 различных: при одинаковых условиях проведены соответственно второй и шестой, третий и пятый опыты; первый и четвертый опыты не повторялись. Таким образом, убеждаемся, что даны результаты ПФЭ типа 2^2 в кодированных переменных с неравномерным дублированием опытов. В этом случае формулы (32), (33) для расчета коэффициентов регрессии неприменимы, для определения коэффициентов необходимо составить систему нормальных уравнений. Для линейного уравнения регрессии с двумя факторами $\widehat{y} = b_0 + b_1x_1 + b_2x_2$ эта система имеет вид

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

Для вычисления нужных сумм составим таблицу, в которой будут учтены все 6 проведенных опытов.

№ оп.	x_{1i}	x_{2i}	y_i	x_{1i}^2	$x_{1i}x_{2i}$	$x_{1i}y_i$	x_{2i}^2	$x_{2i}y_i$
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
$\sum$	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

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

Задачи для самостоятельного решения

Задача 55. Исследуется химический процесс, в котором выход продукта реакции y, %, зависит от температуры реакционной смеси X_1, °C, и концентрации реагента X_2, %. Предполагается провести полный факторный эксперимент для определения линейного уравнения регрессии в окрестности точки $X_1^0 = 50°C$, $X_2^0 = 25$ %. Интервалы варьирования факторов $\Delta X_1 = 5°C$, $\Delta X_2^0 = 1$ %. При каких условиях должны быть проведены опыты?

В задачах **56–59** требуется:

1) определить, являются ли приведенные данные результатами ПФЭ в кодированных или натуральных переменных, имеет ли место дублирование опытов, и если да, то является ли дублирование равномерным;

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

Задача 56.

x_1	-1	$+1$	-1	$+1$
x_2	-1	$+1$	$+1$	-1
y	2	3	4	5

Задача 57.

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 58.

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

Задача 59.

x_1	20	20	80	60
x_2	40	40	50	50
y	2	3	4	5

Ответы к задачам 55–59.

55.

Факторы	X_1	X_2
X_j^0	50	25
ΔX_j	5	1
$X_j^{\max}$	55	26
$X_j^{\min}$	45	24

ПФЭ	Кодированные значения		Натуральные значения	
№ оп.	x_1	x_2	X_1	X_2
1	−	−	45°C	24 %
2	+	−	55°C	24 %
3	−	+	45°C	26 %
4	+	+	55°C	26 %

56. 1) ПФЭ 2^2 в кодированных переменных без дублирования опытов; 2) $\widehat{y} = 3,5 + 0,5x_1$. **57.** 1) ПФЭ 2^2 в натуральных переменных с равномерным дублированием опытов, каждый опыт повторяется $m = 2$ раза; 2) $\widehat{y} = 6,5 + 2x_1 + 0,5x_2$; $\widehat{y} = 6,5 + 2 \cdot \dfrac{X_1 - 30}{10} + 0,5 \cdot \dfrac{X_2 - 20}{10}$. **58.** 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) $\widehat{y} = 4,5 + 1,5x_2$; $\widehat{y} = 4,5 + 1,5 \cdot \dfrac{X_2 - 10}{10}$. **59.** 1) эксперимент не является ПФЭ типа 2^k; для определения коэффициентов линейного уравнения регрессии необходимо составить систему нормальных уравнений, причем для упрощения расчетов имеет смысл перейти к кодированным переменным; 2) при замене $x_1 = \dfrac{X_1 - 40}{20}$, $x_2 = \dfrac{X_2 - 45}{5}$ получим $\widehat{y} = 3,75 - x_1 + 2,25x_2$, т. е. $\widehat{y} = 3,75 - \dfrac{X_1 - 40}{20} + 2,25 \cdot \dfrac{X_2 - 45}{5}$; выбрав замену $x_1 = \dfrac{X_1 - 50}{30}$, $x_2 = \dfrac{X_2 - 45}{5}$ получим $\widehat{y} = 3,25 - 1,5x_1 + 2,25x_2$, т. е. $\widehat{y} = 3,25 - 1,5 \cdot \dfrac{X_1 - 50}{30} + 2,25 \cdot \dfrac{X_2 - 45}{5}$.

Получение моделей со взаимодействиями по результатам полного факторного эксперимента типа 2^k

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 $x_j^3 = x_j^2 \cdot x_j = x_j$ 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 модель, 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:

$$\widehat{y} = b_0 + \sum_{j=1}^{k} b_j x_j + \sum_{j<t} b_{jt} x_j x_t + \sum_{j<t<s} b_{jts} x_j x_t x_s + \ldots$$

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

$$\widehat{y} = b_0 + b_1 x_1 + b_2 x_2 + b_{12} x_1 x_2;$$

наиболее полная модель, которая может быть получена по результатам ПФЭ 2^3, включает парные и тройное взаимодействия:

$$\widehat{y} = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + b_{12} x_1 x_2 + b_{13} x_1 x_3 + b_{23} x_2 x_3 + b_{123} x_1 x_2 x_3;$$

ПФЭ 2^4 позволяет определить коэффициенты уравнения

$$\widehat{y} = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + b_4 x_4 +$$
$$+ b_{12} x_1 x_2 + b_{13} x_1 x_3 + b_{14} x_1 x_4 + b_{23} x_2 x_3 + b_{24} x_2 x_4 + b_{34} x_3 x_4 +$$
$$+ b_{123} x_1 x_2 x_3 + b_{124} x_1 x_2 x_4 + b_{134} x_1 x_3 x_4 + b_{234} x_2 x_3 x_4 + b_{1234} x_1 x_2 x_3 x_4.$$

Отметим, что число коэффициентов в этих моделях равно количеству опытов соответствующего ПФЭ 2^k, т. е. имеет место насыщенное планирование.

Укажем также важный случай моделей с парными взаимодействиями (21). Во многих случаях стремятся получить именно такие модели, считая эффекты тройного взаимодействия и взаимодействий более высоких порядков менее значимыми. Для случаев трех и четырех факторов уравнения с парными взаимодействиями имеют вид

$$\widehat{y} = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_{12}x_1x_2 + b_{13}x_1x_3 + b_{23}x_2x_3;$$

$$\widehat{y} = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4 +$$
$$+ b_{12}x_1x_2 + b_{13}x_1x_3 + b_{14}x_1x_4 + b_{23}x_2x_3 + b_{24}x_2x_4 + b_{34}x_3x_4.$$

Для оценки b_{jt} и других эффектов взаимодействия нужно ввести в матрицу базисных функций X соответствующие столбцы x_jx_t и другие. Записав матрицу базисных функций, увидим, что в ней нет совпадающих столбцов. В данном случае, поскольку все элементы матрицы либо $+1$, либо -1, это означает, что столбцы линейно независимы.

Более того, матрица сохраняет все свойства матрицы плана ПФЭ: симметричность, нормировку, ортогональность. Поэтому в случае ПФЭ типа 2^k все коэффициенты уравнения со взаимодействиями в кодированных переменных рассчитываются независимо друг от друга по таким же, как (32), простым формулам:

$$b_j = \frac{1}{N}\sum_{i=1}^{N} x_{ji}y_i;\ b_{jt} = \frac{1}{N}\sum_{i=1}^{N} x_{ji}x_{ti}y_i;\ b_{jts} = \frac{1}{N}\sum_{i=1}^{N} x_{ji}x_{ti}x_{si}y_i. \quad (34)$$

Замечание. В случае ПФЭ 2^k с равномерным дублированием опытов для расчета коэффициентов в формулы (34) подставляют средние значения отклика $\overline{y_i}$. При неравномерном дублировании опытов эти формулы неприменимы, для определения коэффициентов уравнения регрессии необходимо составлять систему нормальных уравнений (24).

План ПФЭ типа 2^k в случае модели со взаимодействиями обладает свойствами ортогональности и D-оптимальности (при данном количестве факторов k, числе опытов $N = 2^k$ и заданном виде уравнения регрессии), но, в отличие от случая линейной модели, не обладает свойством ротатабельности.

Примеры решения задач

Задача 60. По приведенным в задаче **53** результатам эксперимента получить уравнение регрессии с парными взаимодействиями.

Решение. Имеем результаты ПФЭ типа 2^3 в натуральных переменных без дублирования опытов. При решении задачи **53** были определены формулы, задающие переход от натуральных значений факторов к кодированным, и составлена матрица плана в кодированных переменных. Для расчета коэффициентов уравнения с парными взаимодействиями в кодированных переменных

$$\widehat{y} = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_{12}x_1x_2 + b_{13}x_1x_3 + b_{23}x_2x_3$$

введем в матрицу столбцы, отвечающие парным взаимодействиям x_1x_2, x_1x_3 и x_2x_3. Для этого перемножим соответствующие значения исходных столбцов x_1, x_2 и x_3.

№ оп.	x_0	x_1	x_2	x_3	x_1x_2	x_1x_3	x_2x_3	y
1	+	+	−	−	−	−	+	2
2	+	−	−	−	+	+	+	3
3	+	+	+	−	+	−	−	4
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 в кодированных переменных вычисляются независимо друг от друга по формулам (34). Свободный член уравнения b_0 и линейные эффекты b_1, b_2, b_3 были рассчитаны при решении задачи **53**, их значения не изменятся. Определим эффекты парных взаимодействий:

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

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

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

Итак, уравнение регрессии в кодированных переменных запишется в виде

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

$$+ 0,25x_1x_2 + 0,25x_1x_3 + 0,25x_2x_3.$$

Подставив формулы, связывающие натуральные и кодированные значения факторов, получим уравнение регрессии в натуральных переменных:

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

$$+0,25\cdot\frac{X_1-50}{10}\cdot\frac{X_2-50}{30} + 0,25\cdot\frac{X_1-50}{10}\cdot\frac{X_3-5}{5} + 0,25\cdot\frac{X_2-50}{30}\cdot\frac{X_3-5}{5}.$$

Задача 61. По данным задачи **53** получить уравнение регрессии со взаимодействиями.

Задача 62. По данным задачи **56** получить уравнение регрессии с парным взаимодействием.

Задача 63. По данным задачи **57** получить уравнение регрессии со взаимодействиями.

Задача 64. Изучается влияние на выход продукта y, %, трех факторов: температуры X_1, изменяющейся в диапазоне 100–200°C, давления X_2, $2\text{–}6 \cdot 10^5$ Па, и времени пребывания X_3, изменяющегося в окрестности 15 мин с интервалом 5 мин. Предполагается получить уравнение регрессии с парными взаимодействиями. Требуется предложить план эксперимента, составить матрицу плана в кодированных переменных, указать условия проведения опытов.

Ответы к задачам 61–64. **61.** Уравнение в кодированных переменных: $\widehat{y} = 4,75 - 0,25x_1 + 1,25x_2 + 1,25x_3 + 0,25x_1x_2 + 0,25x_1x_3 + 0,25x_2x_3 + 0,25x_1x_2x_3$; уравнение в натуральных переменных: $\widehat{y} = 4,75 - 0,25 \cdot \dfrac{X_1 - 50}{10} + 1,25 \cdot \dfrac{X_2 - 50}{30} + 1,25 \cdot \dfrac{X_3 - 5}{5} + 0,25 \cdot \dfrac{X_1 - 50}{10} \cdot \dfrac{X_2 - 50}{30} + 0,25 \cdot \dfrac{X_1 - 50}{10} \cdot \dfrac{X_3 - 5}{5} + 0,25 \cdot \dfrac{X_2 - 50}{30} \times \dfrac{X_3 - 5}{5} + 0,25 \cdot \dfrac{X_1 - 50}{10} \cdot \dfrac{X_2 - 50}{30} \cdot \dfrac{X_3 - 5}{5}$. **62.** $\widehat{y} = 3,5 + 0,5x_1 - x_1x_2$. **63.** Эффект парного взаимодействия незначим: $b_{12} = 0$; уравнение в кодированных переменных: $\widehat{y} = 6,5 + 2x_1 + 0,5x_2$; уравнение в натуральных переменных: $\widehat{y} = 6,5 + 2 \cdot \dfrac{X_1 - 30}{10} + 0,5 \cdot \dfrac{X_2 - 20}{10}$. **64.** ПФЭ типа 2^3;

Факторы	X_1	X_2	X_3
X_j^0	150	4	15
ΔX_j	50	2	5
$X_j^{\max}$	200	6	20
$X_j^{\min}$	100	2	10

ПФЭ	Матрица базисных функций в кодированных переменных							Натуральные значения факторов		
№ оп.	x_0	x_1	x_2	x_3	x_1x_2	x_1x_3	x_2x_3	X_1	X_2	X_3
1	+	−	−	−	+	+	+	100	2	10
2	+	+	−	−	−	−	+	200	2	10
3	+	−	+	−	−	+	−	100	6	10
4	+	+	+	−	+	−	−	200	6	10
5	+	−	−	+	+	−	−	100	2	20
6	+	+	−	+	−	+	−	200	2	20
7	+	−	+	+	−	−	+	100	6	20
8	+	+	+	+	+	+	+	200	6	20

Таким образом, опыты нужно проводить при следующих условиях: первый опыт провести при температуре 100°C и давлении $2\cdot10^5$ Па в течение 10 мин; во втором опыте установить температуру 200°C, давление $2\cdot10^5$ Па, время 10 мин; в третьем опыте температура 100°C, давление $6\cdot10^5$ Па, время 10 мин; в четвертом опыте температура 200°C, давление $6\cdot10^5$ Па, время 10 мин. Далее такие же опыты повторяются для времени 20 мин.

Дробный факторный эксперимент типа 2^{k-p}

С ростом числа факторов количество опытов в ПФЭ 2^k резко возрастает:

k	2	3	4	5	6	7	8	9	...
$N=2^k$	4	8	16	32	64	128	256	512	...

По результатам такого эксперимента можно получить уравнение регрессии со взаимодействиями, включив в него парные x_jx_t и тройные $x_jx_tx_s$ взаимодействия, а также взаимодействия более высоких порядков. Однако ПФЭ 2^k не позволяет оценить квадратичные эффекты x_j^2 и получить квадратичное уравнение регрессии. Во многих случаях учет взаимодействий высоких порядков при невозможности учета квадратичных эффектов не является необходимым.

К тому же обычно на начальном этапе исследования требуется получить какую-то, хотя бы приблизительную информацию о процессе при минимальном числе опытов. Для этой цели достаточно линейной модели, однако при построении линейного уравнения регрессии при большом количестве факторов ПФЭ является неэкономным.

Для сокращения числа опытов был предложен дробный факторный эксперимент (ДФЭ).

Дробным факторным называется эксперимент, реализующий часть (дробную реплику) полного факторного эксперимента. При этом дробные реплики выбираются так, чтобы сохранялись основные свойства матрицы плана ПФЭ. Для этого в качестве основы плана ДФЭ берут план ПФЭ с меньшим числом факторов.

Рассмотрим идею построения ДФЭ на примере $k=3$ факторов. Если мы оцениваем коэффициенты линейной модели с тремя факторами $\widehat{y}=b_0+b_1x_1+b_2x_2+b_3x_3$, то нам надо оценить 4 коэффициента и для этого в принципе достаточно 4 опытов. Используем матрицу плана ПФЭ 2^2 и положим $x_3=x_1x_2$, т. е. будем изменять фактор x_3 по такой программе:

№ оп.	x_0	x_1	x_2	$x_3 = x_1x_2$
1	+	−	−	+
2	+	+	−	−
3	+	−	+	−
4	+	+	+	+

Матрица плана такого эксперимента сохраняет все свойства матрицы плана ПФЭ 2^k (симметричность, нормировка, ортогональность), и поэтому коэффициенты уравнения регрессии в кодированных переменных вычисляются по тем же простым формулам (32).

Полученная матрица плана представляет собой половину матрицы плана ПФЭ 2^3, поэтому соответствующий план называется **полурепликой** от ПФЭ 2^k и обозначается как ДФЭ 2^{3-1} (3 — число факторов, 1 — число линейных эффектов, приравненных к эффектам взаимодействия).

В общем случае рассматривают ДФЭ типа 2^{k-p}, где k — число факторов, а p — степень дробности реплики.

В рассмотренном примере фактор x_3 варьируется одинаково с парным взаимодействием x_1x_2, поэтому по результатам такого эксперимента влияние фактора x_3 нельзя отделить от влияния взаимодействия x_1x_2; говорят, что коэффициент b_3 дает **совместную**, или **смешанную**, оценку для истинных коэффициентов регрессии β_3 и β_{12}. Это символически записывается так: $b_3 \to \beta_3 + \beta_{12}$.

Если построить расширенную матрицу плана, включив парные и тройное взаимодействия (все эффекты, влияние которых может быть оценено по результатам ПФЭ 2^3), то по совпадающим столбцам можно определить, с чем смешиваются оценки остальных коэффициентов:

№ оп.	x_0	x_1	x_2	x_3	x_1x_2	x_1x_3	x_2x_3	$x_1x_2x_3$
1	+	−	−	+	+	−	−	+
2	+	+	−	−	−	−	+	+
3	+	−	+	−	−	+	−	+
4	+	+	+	+	+	+	+	+

$$\text{Поскольку} \begin{cases} x_0 = x_1x_2x_3, \\ x_1 = x_2x_3, \\ x_2 = x_1x_3, \\ x_3 = x_1x_2, \end{cases} \text{то} \begin{cases} b_0 \to \beta_0 + \beta_{123}, \\ b_1 \to \beta_1 + \beta_{23}, \\ b_2 \to \beta_2 + \beta_{13}, \\ b_3 \to \beta_3 + \beta_{12}. \end{cases}$$

Это **полная система смешивания оценок** ДФЭ 2^{3-1} с **генерирующим соотношением** $x_3 = x_1x_2$.

Чтобы получить систему смешивания оценок, не выписывая расширенную матрицу плана, удобно использовать следующий прием.

Как было видно на примере ДФЭ 2^{3-1}, при построении ДФЭ все исследуемые факторы условно делятся на основные и дополнительные (в рассмотренном примере основными являлись факторы x_1 и x_2, а дополнительным — фактор x_3). При составлении матрицы плана для основных факторов записывается матрица ПФЭ, а дополнительные факторы приравниваются к некоторым взаимодействиям (произведениям) основных факторов.

Соотношение, привязывающее дополнительный фактор к определенному взаимодействию основных факторов и служащее для построения ДФЭ, называется **генерирующим соотношением** (ГС). Оно показывает, с каким из эффектов взаимодействия связан данный фактор.

В рассмотренном примере ГС $x_3 = x_1x_2$. Умножив обе части данного ГС на x_3, получим $x_3^2 = x_1x_2x_3$. Поскольку кодированные переменные принимают только значения $+1$ и -1, то $x_3^2 = (\pm 1)^2 = 1$, а следовательно,

$$1 = x_1x_2x_3.$$

Это соотношение называется **определяющим контрастом** (ОК) данной реплики (данного ДФЭ).

Определяющий контраст — это соотношение, в левой части которого стоит 1, а в правой — некоторое произведение факторов, т. е. соотношение, задающее элементы 1-го столбца матрицы плана, отвечающего фиктивной переменной.

Умножая ОК последовательно на независимые переменные, получим всю систему смешивания оценок:

$$\begin{cases} x_0 = x_1x_2x_3, \\ x_1 = x_1^2x_2x_3 = x_2x_3, \\ x_2 = x_1x_2^2x_3 = x_1x_3, \\ x_3 = x_1x_2x_3^2 = x_1x_2, \end{cases} \quad \text{т. е.} \quad \begin{cases} b_0 \to \beta_0 + \beta_{123}, \\ b_1 \to \beta_1 + \beta_{23}, \\ b_2 \to \beta_2 + \beta_{13}, \\ b_3 \to \beta_3 + \beta_{12}. \end{cases}$$

Рассмотрим реплики более высокой степени дробности.

При построении ДФЭ стремятся, чтобы число опытов было как можно меньше, однако этих опытов должно быть достаточно для определения коэффициентов заданного уравнения регрессии. Наиболее простое уравнение регрессии — линейное, в случае k факторов оно содержит $k+1$ коэффициент. Следовательно, степень дробности ДФЭ выбирают так, чтобы $k + 1 \leqslant 2^{k-p}$. В связи с этим при числе факторов $k = 3$ и $k = 4$ возможно использование только полуреплик (ДФЭ типа 2^{k-1}).

Рассмотрим $k = 5$ и ДФЭ 2^{5-2} — четверть-реплику от ПФЭ 2^5. В основе этого эксперимента лежит ПФЭ 2^3. Будем считать основными факторами x_1, x_2, x_3. Программа изменения дополнительных факторов x_4 и x_5 должна быть задана с помощью ГС, представляющих собой некоторые произведения основных факторов.

ГС могут быть выбраны по-разному. Как правило, в качестве ГС выбирают взаимодействия возможно более высоких порядков, которые обычно считаются менее значимыми.

Выберем ГС для факторов x_4 и x_5 и запишем соответствующие ОК:

$$\begin{array}{ll} \text{ГС:} & x_4 = x_1x_2x_3, \quad x_5 = x_1x_2; \\ \text{ОК:} & 1 = x_1x_2x_3x_4, \quad 1 = x_1x_2x_5. \end{array}$$

Перемножив исходные ОК между собой, получим новый ОК: $1 = x_3x_4x_5$.

Запишем **обобщенный определяющий контраст** (ООК), включающий все возможные ОК:

$$1 = x_1x_2x_3x_4 = x_1x_2x_5 = x_3x_4x_5.$$

Обобщенный определяющий контраст получается при перемножении всех возможных ОК данной реплики во всех возможных сочетаниях (по одному, по два, по три и т. д.). ООК содержит 2^p (где p — степень дробности реплики) соотношений.

Зная ООК, легко получить систему смешивания оценок данного ДФЭ.

Замечание. Полная система смешивания оценок ДФЭ 2^{k-p} будет содержать 2^{k-p} строк по 2^p соотношений (k — число факторов, p — степень дробности реплики). Чтобы получить полную систему смешивания оценок, нужно выписать все эффекты, влияние которых можно оценить по результатам ПФЭ 2^k, и определить, какие из них смешиваются между собой в данном ДФЭ. Поскольку ДФЭ 2^{k-p} построен на основе ПФЭ для $k-p$ факторов, то по нему можно определить 2^{k-p} различных эффектов, следовательно, в системе смешивания оценок будет 2^{k-p} различных соотношений.

В нашем случае $k = 5$, $p = 2$, $2^{k-p} = 8$. Система смешивания оценок будет содержать 8 различных соотношений. Одно из них — ООК, еще 5 соотношений получим, умножив ООК последовательно на переменные x_1, x_2, x_3, x_4, x_5.

Далее будем умножать ООК на парные взаимодействия. Умножение на x_1x_2 не даст нового соотношения, т. к. в силу заданных ГС x_1x_2 смешивается с x_5. Поэтому умножим ООК на x_1x_3 и x_1x_4. В результате получим 8 соотношений, в которые войдут все 32 различных эффекта, которые можно было бы определить по результатам ПФЭ 2^5.

Итак, система смешивания оценок ДФЭ 2^{5-2} с ГС $x_4 = x_1x_2x_3$, $x_5 = x_1x_2$ имеет вид:

$$
\begin{aligned}
&1 = x_1x_2x_3x_4 = x_1x_2x_5 = x_3x_4x_5;\\
&x_1 = x_2x_3x_4 = x_2x_5 = x_1x_3x_4x_5;\\
&x_2 = x_1x_3x_4 = x_1x_5 = x_2x_3x_4x_5;\\
&x_3 = x_1x_2x_4 = x_1x_2x_3x_5 = x_4x_5;\\
&x_4 = x_1x_2x_3 = x_1x_2x_4x_5 = x_3x_5;\\
&x_5 = x_1x_2x_3x_4x_5 = x_1x_2 = x_3x_4;\\
&x_1x_3 = x_2x_4 = x_2x_3x_5 = x_1x_4x_5;\\
&x_1x_4 = x_2x_3 = x_2x_4x_5 = x_1x_3x_4x_5.
\end{aligned}
$$

В заключение отметим следующее.

1. ДФЭ применяется с целью сокращения числа опытов при решении следующих задач: 1) построение линейного уравнения регрессии в локальной области изменения факторов; 2) описание процессов, в которых заведомо не могут иметь место хотя бы некоторые из парных взаимодействий.

2. План ДФЭ типа 2^{k-p}, как и ПФЭ типа 2^k, обладает свойствами ортогональности, D-оптимальности, а в случае линейной модели — и свойством ротатабельности.

3. Коэффициенты уравнения регрессии в кодированных переменных вычисляются по результатам ДФЭ типа 2^{k-p} с помощью тех же простых формул (32)–(34), что и в случае ПФЭ типа 2^k.

4. Эффективность применения дробной реплики зависит от удачного выбора системы смешивания оценок. При построении ДФЭ следует использовать всю априорную информацию о процессе и выбирать дробную реплику так, чтобы линейные эффекты и существенные (значимые) эффекты взаимодействия оценивались раздельно (не смешивались между собой).

Выбор дробной реплики осложняется тем, что нельзя заранее сказать, возможно ли построить план ДФЭ при заданной степени дробности реплики и указанном наборе существенных переменных. Задача выбора дробной реплики решается путем рассмотрения всех возможностей.

Задача 65. Составить план наиболее экономного ДФЭ при данном списке существенных переменных, записать матрицу плана:

а) $x_0,\ x_1,\ x_2,\ x_3,\ x_4,\ x_1x_2,\ x_2x_3,\ x_2x_4$;

б) $x_0,\ x_1,\ x_2,\ x_3,\ x_4,\ x_1x_2,\ x_2x_3,\ x_3x_4$.

Решение. а) По условию имеем 4 фактора, 8 существенных переменных. Требуется составить план эксперимента, позволяющий оценить коэффициенты уравнения

$$\widehat{y} = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4 + b_{12}x_1x_2 + b_{23}x_2x_3 + b_{24}x_2x_4.$$

Для этого необходимо как минимум 8 опытов. Число опытов ДФЭ должно удовлетворять условию $2^{4-p} \geqslant 8$. Следовательно, можно попытаться использовать ДФЭ типа 2^{4-1}, содержащий 8 опытов. Для построения этого эксперимента нужно выбрать три основных фактора (пусть это будут x_1, x_2 и x_3) и генерирующее соотношение, задающее дополнительный фактор x_4 через некоторое взаимодействие основных факторов.

Выпишем возможные взаимодействия основных факторов: $x_1x_2x_3,\ x_1x_2,\ x_1x_3,\ x_2x_3$. Следовательно, имеем 8 различных вариантов выбора ГС:

$$\begin{array}{llll} 1)\ x_4 = x_1x_2x_3; & 2)\ x_4 = x_1x_2; & 3)\ x_4 = x_1x_3; & 4)\ x_4 = x_2x_3; \\ 5)\ x_4 = -x_1x_2x_3; & 6)\ x_4 = -x_1x_2; & 7)\ x_4 = -x_1x_3; & 8)\ x_4 = -x_2x_3. \end{array}$$

Далее для каждого способа задания ДФЭ нужно проверить, не смешиваются ли между собой существенные переменные. Для этого запишем ООК и выпишем систему смешивания оценок, умножая ООК последовательно на существенные переменные.

Рассмотрим первый вариант ДФЭ. По выбранному генерирующему соотношению запишем обобщенный определяющий контраст:

$$\text{ГС: } x_4 = x_1x_2x_3;$$

$$\text{ООК: } 1 = x_1x_2x_3x_4;$$

система смешивания оценок будет иметь вид:

$$\begin{aligned} &\underline{1} = x_1x_2x_3x_4; \\ &\underline{x_1} = x_2x_3x_4; \\ &\underline{x_2} = x_1x_3x_4; \\ &\underline{x_3} = x_1x_2x_4; \\ &\underline{x_4} = x_1x_2x_3; \end{aligned}$$

$$\underline{x_1x_2} = x_3x_4;$$
$$\underline{x_2x_3} = x_1x_4;$$
$$\underline{x_2x_4} = x_1x_3.$$

Подчеркнув существенные переменные в системе смешивания оценок, видим, что все существенные переменные входят в разные соотношения, т. е. не смешиваются между собой. Следовательно, этот способ задания ДФЭ удовлетворяет условию задачи.

№ оп.	x_1	x_2	x_3	$x_4 = x_1x_2x_3$
1	−	−	−	−
2	+	−	−	+
3	−	+	−	+
4	+	+	−	−
5	−	−	+	+
6	+	−	+	−
7	−	+	+	−
8	+	+	+	+

Для того чтобы составить матрицу плана в кодированных переменных, запишем матрицу плана ПФЭ 2^3 для основных факторов x_1, x_2 и x_3, а столбец для фактора x_4 получим, перемножив соответствующие значения столбцов для x_1, x_2, x_3 по формуле $x_4 = x_1x_2x_3$.

Аналогично можно показать, что варианты

3) ДФЭ 2^{4-1} с ГС $x_4 = x_1x_3$;

5) ДФЭ 2^{4-1} с ГС $x_4 = -x_1x_2x_3$;

7) ДФЭ 2^{4-1} с ГС $x_4 = -x_1x_3$

также будут допустимыми при данном списке существенных переменных.

Остальные варианты не подходят, поскольку используют существенные переменные в генерирующем соотношении, тем самым смешивая их с x_4.

б) По условию имеем 4 фактора, 8 существенных переменных. Требуется составить план эксперимента, позволяющий оценить коэффициенты уравнения

$$\widehat{y} = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4 + b_{12}x_1x_2 + b_{23}x_2x_3 + b_{34}x_3x_4.$$

Для этого необходимо как минимум 8 опытов. Попытаемся использовать ДФЭ типа 2^{4-1}, содержащий 8 опытов. Считая основными факторы x_1, x_2 и x_3, выпишем все возможные варианты выбора ГС, задающего x_4 через произведение основных факторов:

1) $x_4 = x_1x_2x_3$; 2) $x_4 = x_1x_2$; 3) $x_4 = x_1x_3$; 4) $x_4 = x_2x_3$;
5) $x_4 = -x_1x_2x_3$; 6) $x_4 = -x_1x_2$; 7) $x_4 = -x_1x_3$; 8) $x_4 = -x_2x_3$.

Заметим, что по существу должны быть рассмотрены только варианты 1–4, т. к. варианты 5–8 отличаются от них только знаком, на группирование смешивающихся переменных это не влияет.

Варианты 2 и 4 не подходят, поскольку смешивают существенные переменные x_1x_2 или x_2x_3 с существенной переменной x_4 уже в генерирующем соотношении.

Для оставшихся вариантов 1 и 3 выпишем системы смешивания оценок.

$$
\begin{array}{ll}
(1)\ \text{ГС: } x_4 = x_1x_2x_3; & (3)\ \text{ГС: } x_4 = x_1x_3; \\
\underline{1} = x_1x_2x_3x_4; & \underline{1} = x_1x_3x_4; \\
\underline{x_1} = x_2x_3x_4; & \underline{x_1} = \underline{x_3x_4}; \\
\underline{x_2} = x_1x_3x_4; & \underline{x_2} = x_1x_2x_3x_4; \\
\underline{x_3} = x_1x_2x_4; & \underline{x_3} = x_1x_4; \\
\underline{x_4} = x_1x_2x_3; & \underline{x_4} = x_1x_3; \\
\underline{x_1x_2} = \underline{x_3x_4}; & \underline{x_1x_2} = x_2x_3x_4; \\
\underline{x_2x_3} = x_1x_4; & \underline{x_2x_3} = x_1x_2x_4; \\
x_2x_4 = x_1x_3. & x_2x_4 = x_1x_2x_4.
\end{array}
$$

Подчеркнув существенные переменные в системе смешивания оценок, видим, что в случае 1 смешиваются парные взаимодействия x_1x_2 и x_3x_4, а в случае 3 — фактор x_1 с существенным взаимодействием x_3x_4. Таким образом, эти варианты ДФЭ также не удовлетворяют условию.

Вывод: для получения указанного уравнения регрессии 8 опытов ДФЭ недостаточно, необходимо использовать ПФЭ 2^4.

Матрица плана этого эксперимента записана в таблице. Здесь фактор x_1 изменяется от опыта к опыту, второй фактор — в два раза реже, третий фактор — еще в два раза реже, фактор x_4 в первой половине эксперимента фиксируется на нижнем, а во второй — на верхнем уровне.

№ оп.	x_1	x_2	x_3	x_4
1	−	−	−	−
2	+	−	−	−
3	−	+	−	−
4	+	+	−	−
5	−	−	+	−
6	+	−	+	−
7	−	+	+	−
8	+	+	+	−
9	−	−	−	+
10	+	−	−	+
11	−	+	−	+
12	+	+	−	+
13	−	−	+	+
14	+	−	+	+
15	−	+	+	+
16	+	+	+	+

Задача 66. Для расчета аппаратов конденсационной системы при проектировании хлораторов необходимо знать коэффициент удельной теплопроводности возгонов, образующихся при хлорировании титанового шлака в расплаве. В качестве независимых переменных, влияющих на теплопроводность, выбраны: X_1 — насыпной вес, г/см^3; X_2 — содержание хлора в возгонах, весовые проценты; X_3 — отношение концентраций SiO_2 и TiO_2; X_4 — температура, °C. Предполагается получить уравнение регрессии, изменяя

факторы в следующей области: $0,72 \leqslant X_1 \leqslant 1,02$; $35 \leqslant X_2 \leqslant 45$; $0,75 \leqslant X_3 \leqslant 1,25$; $200 \leqslant X_4 \leqslant 300$.

Исходя из общетеоретических соображений, можно априори предполагать, что на теплопроводность возгонов могут оказывать влияние только два парных взаимодействия: взаимодействие факторов состава x_2x_3 и взаимодействие x_3x_4 (механизм теплопроводности определяется отношением концентраций SiO_2/TiO_2 и может изменяться с изменением температуры). Остальными взаимодействиями можно пренебречь.

Требуется составить план наиболее экономного по количеству опытов ДФЭ; записать матрицу плана в кодированных переменных; указать условия проведения первых трех опытов.

Решение. По условию имеем 4 фактора и следующий список существенных переменных: x_0, x_1, x_2, x_3, x_4, x_2x_3, x_3x_4. Поскольку этот список содержит 7 переменных, для определения коэффициентов соответствующего уравнения регрессии

$$\widehat{y} = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4 + b_{23}x_2x_3 + b_{34}x_3x_4$$

необходимо как минимум 7 различных опытов. Это означает, что в основу плана ДФЭ должен быть положен план ПФЭ 2^3, содержащий 8 опытов. Следовательно, мы можем попытаться провести ДФЭ типа 2^{4-1}. Для этого нужно выбрать три основных фактора (пусть это будут факторы x_1, x_2, x_3), а четвертый фактор приравнять в эксперименте к некоторому взаимодействию основных факторов.

Существуют следующие взаимодействия трех основных факторов: x_1x_2, x_1x_3, x_2x_3, $x_1x_2x_3$. Взаимодействие x_2x_3 входит в список существенных переменных, поэтому его нельзя использовать в качестве генерирующего соотношения. Возможны следующие генерирующие соотношения (ГС) (с точностью до выбора знаков): $x_4 = x_1x_2x_3$; $x_4 = x_1x_2$; $x_4 = x_1x_3$.

Рассмотрим первый вариант — ДФЭ 2^{4-1} с ГС $x_4 = x_1x_2x_3$. В данном случае, т. е. в случае полуреплики, ООК совпадает с ОК. Он получится, если домножить обе части ГС на его левую часть x_4: $1 = x_1x_2x_3x_4$. Систему смешивания оценок можем получить, умножая ООК последовательно на все существенные переменные:

$$\begin{aligned}
\underline{1} &= x_1x_2x_3x_4;\\
\underline{x_1} &= x_2x_3x_4;\\
\underline{x_2} &= x_1x_3x_4;\\
\underline{x_3} &= x_1x_2x_4;
\end{aligned}$$

$$\underline{x_4} = x_1 x_2 x_3;$$
$$\underline{x_2 x_3} = x_1 x_4;$$
$$\underline{x_3 x_4} = x_1 x_2;$$
$$x_1 x_3 = x_2 x_4.$$

В системе смешивания оценок должно быть столько же соотношений, сколько различных опытов в ДФЭ. Поэтому при нахождении последнего, восьмого, соотношения мы умножили ООК на взаимодействие x_1x_3, которое не встречается в первых семи соотношениях системы смешивания.

Поскольку существенные переменные (подчеркнуты в системе смешивания оценок) не смешиваются между собой (не встречаются в одном соотношении системы смешивания), этот вариант ДФЭ можно использовать для решения нашей задачи.

Можно показать, что ДФЭ 2^{4-1} с ГС $x_4 = x_1x_2$ также удовлетворяет условию задачи. Соответственно, возможно также использовать ДФЭ с ГС $x_4 = -x_1x_2x_3$ или $x_4 = -x_1x_2$.

Покажем, что ДФЭ 2^{4-1} с ГС $x_4 = x_1x_3$ не удовлетворяет условию задачи. В этом случае ООК: $1 = x_1x_3x_4$. Домножив ООК на x_1, получим $\underline{x_1} = \underline{x_3x_4}$, т. е. смешиваются существенные переменные x_1 и x_3x_4.

Итак, решением задачи является ДФЭ с ГС $x_4 = \pm x_1x_2x_3$ или $x_4 = \pm x_1x_2$.

Составим матрицу плана и определим условия проведения опытов ДФЭ с ГС $x_4 = x_1x_2x_3$. Предварительно найдем центр плана и интервалы варьирования факторов по формулам (31) и запишем в соответствии с (30) формулы, задающие связь между натуральными и кодированными значениями факторов:

$$x_1 = \frac{X_1 - 0,87}{0,15}; \quad x_2 = \frac{X_2 - 40}{5}; \quad x_3 = \frac{X_3 - 1}{0,25}; \quad x_4 = \frac{X_4 - 250}{50}.$$

Уровни и интервалы варьирования факторов занесем в таблицу.

Факторы	X_1	X_2	X_3	X_4
X_j^0	$0,87$	40	1	250
ΔX_j	$0,15$	5	$0,25$	50
$X_j^{\max}$	$1,02$	45	$1,25$	300
$X_j^{\min}$	$0,72$	35	$0,75$	200

При составлении матрицы плана в кодированных переменных используем матрицу плана ПФЭ 2^3 для основных факторов x_1, x_2 и x_3, а столбец для фактора x_4 получим, перемножив соответствующие значения столбцов для основных факторов x_1, x_2, x_3 согласно заданному ГС $x_4 = x_1x_2x_3$. Введем дополнительно в матрицу плана столбцы x_2x_3 и x_3x_4, отвечающие существенным взаимодействиям, чтобы убедиться еще раз, что все столбцы матрицы базисных функций различны, а следовательно, существенные переменные не смешиваются между собой.

ДФЭ	Матрица базисных функций в кодированных переменных							Натуральные значения факторов			
№ оп.	x_0	x_1	x_2	x_3	$x_4 = x_1x_2x_3$	x_2x_3	x_3x_4	X_1	X_2	X_3	X_4
1	+	−	−	−	−	+	+	0,72	35	0,75	200
2	+	+	−	−	+	+	−	1,02	35	0,75	300
3	+	−	+	−	+	−	−	0,72	45	0,75	300
4	+	+	+	−	−	−	+	1,02	45	0,75	200
5	+	−	−	+	+	−	+	0,72	35	1,25	300
6	+	+	−	+	−	−	−	1,02	35	1,25	200
7	+	−	+	+	−	+	−	0,72	45	1,25	200
8	+	+	+	+	+	+	+	1,02	45	1,25	300

Укажем условия проведения первых трех опытов. В первом опыте следует задать насыпной вес $0,72$ г/см^3, содержание хлора в возгонах 35%, отношение концентраций SiO_2 и TiO_2 $0,75$, температуру 200°C; во втором опыте — насыпной вес $1,02$ г/см^3, содержание хлора в возгонах 35%, отношение концентраций SiO_2 и TiO_2 $0,75$, температуру 300°C; третий опыт проводится при насыпном весе $0,72$ г/см^3, содержании хлора в возгонах 45%, отношении концентраций SiO_2 и TiO_2 $0,75$, температуре 300°C.

Задача 67. При исследовании возможности получения азотно-фосфорно-калийного удобрения путем частичной замены поташа аммиаком предполагается получить линейное уравнение зависимости степени усвояемости образующихся фосфорных соединений от следующих независимых факторов: X_1 — температура аммонизации, °C; X_2 — продолжительность аммонизации, мин; X_3 — норма аммиака в процентах от стехиометрической нормы; X_4 — температура при взаимодействии компонентов аммонизированной вытяжки с раствором поташа, °C; X_5 — продолжительность взаимодействия с раствором поташа, мин; X_6 — норма поташа в процентах от стехиометрической нормы. Базовые уровни и интервалы варьирования факторов даны в таблице.

Факторы	X_1	X_2	X_3	X_4	X_5	X_6
X_j^0	50	20	125	50	45	110
ΔX_j	20	10	25	20	15	10

Требуется составить план наиболее экономного по количеству опытов ДФЭ; записать матрицу плана в кодированных переменных; указать условия проведения опытов.

Решение. По условию имеем 6 факторов. Линейное уравнение регрессии имеет вид

$$\widehat{y} = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4 + b_5x_5 + b_6x_6,$$

т. е. содержит 7 коэффициентов. Для их определения необходимо как минимум 7 различных опытов, поэтому в основу плана ДФЭ должен быть положен план ПФЭ 2^3, содержащий 8 опытов. Следовательно, мы будем планировать ДФЭ типа 2^{6-3}.

Выберем три основных фактора x_1, x_2, x_3 и ГС, задающие программу изменения остальных факторов. ГС могут быть заданы с помощью следующих произведений трех основных факторов: x_1x_2, x_1x_3, x_2x_3, $x_1x_2x_3$.

Выберем ДФЭ 2^{6-3} с ГС $x_4 = x_1x_2x_3$; $x_5 = x_1x_2$; $x_6 = x_1x_3$. Запишем в таблицу матрицу плана и условия проведения опытов для данного ДФЭ. Предварительно определим верхний и нижний уровни каждого фактора по формулам

$$X_j^{\max} = X_j^0 + \Delta X_j; \qquad X_j^{\min} = X_j^0 - \Delta X_j.$$

Согласно (30), связь между натуральными и кодированными значениями факторов задается формулами

$$x_1 = \frac{X_1 - 50}{20}; \quad x_2 = \frac{X_2 - 20}{10}; \quad x_3 = \frac{X_3 - 125}{25};$$

$$x_4 = \frac{X_4 - 50}{20}; \quad x_5 = \frac{X_5 - 45}{15}; \quad x_6 = \frac{X_6 - 110}{10}.$$

Уровни и интервалы варьирования факторов занесем в таблицу.

Факторы	X_1	X_2	X_3	X_4	X_5	X_6
X_j^0	50	20	125	50	45	110
ΔX_j	20	10	25	20	15	10
$X_j^{\max}$	70	30	150	70	60	120
$X_j^{\min}$	30	10	100	30	30	100

При составлении матрицы плана в кодированных переменных используем матрицу плана ПФЭ 2^3 для основных факторов x_1, x_2 и x_3, а столбцы для факторов x_4, x_5, x_6 получим, перемножив соответствующие значения столбцов для основных факторов x_1, x_2, x_3 в соответствии с заданными ГС $x_4 = x_1x_2x_3$; $x_5 = x_1x_2$; $x_6 = x_1x_3$. Поскольку все столбцы матрицы базисных функций различны, убеждаемся, что ДФЭ 2^{6-3} с ГС $x_4 = x_1x_2x_3$; $x_5 = x_1x_2$; $x_6 = x_1x_3$ удовлетворяет условию задачи.

ДФЭ	Матрица базисных функций в кодированных переменных							Натуральные значения факторов					
№ оп.	x_0	x_1	x_2	x_3	x_4	x_5	x_6	X_1	X_2	X_3	X_4	X_5	X_6
1	+	−	−	−	−	+	+	30	10	100	30	60	120
2	+	+	−	−	+	−	−	70	10	100	70	30	100
3	+	−	+	−	+	−	+	30	30	100	70	30	120
4	+	+	+	−	−	+	−	70	30	100	30	60	100
5	+	−	−	+	+	+	−	30	10	150	70	60	100
6	+	+	−	+	−	−	+	70	10	150	30	30	120
7	+	−	+	+	−	−	−	30	30	150	30	30	100
8	+	+	+	+	+	+	+	70	30	150	70	60	120

Условия проведения опытов указаны в таблице. В частности, в первом опыте нужно установить: температуру аммонизации 30°С, продолжительность аммонизации 10 мин, норму аммиака 100% от стехиометрической нормы, температуру при взаимодействии компонентов аммонизированной вытяжки с раствором поташа 30°С, продолжительность взаимодействия с раствором поташа 60 мин, норму поташа 120% от стехиометрической нормы.

Задачи для самостоятельного решения

Задача 68. Составить матрицу плана ДФЭ 2^{k-p} с ГС $x_4 = x_1x_3$; $x_5 = -x_2x_3$; $x_6 = -x_1x_2$. Указать тип ДФЭ; записать ООК.

Задача 69. Составить матрицу плана ДФЭ 2^{k-p} с ГС $x_5 = x_1x_2x_3x_4$; $x_6 = -x_2x_3x_4$; $x_7 = -x_1x_2x_3$. Указать тип ДФЭ; записать ООК.

Задача 70. Предложить наиболее экономный ДФЭ при данном списке существенных переменных:

а) x_0, x_1, x_2, x_3, x_4, x_5, x_1x_2;

б) x_0, x_1, x_2, x_3, x_4, x_5, x_1x_2, x_2x_3;

в) x_0, x_1, x_2, x_3, x_4, x_5, x_1x_2, x_4x_5.

Ответы к задачам 68–70. **68.** ДФЭ типа 2^{6-3} с ООК: $1 = x_1x_3x_4 = -x_2x_3x_5 = -x_1x_2x_6 = -x_1x_2x_4x_5 = -x_2x_3x_4x_6 = x_1x_3x_5x_6 = x_4x_5x_6$.

№ оп.	x_0	x_1	x_2	x_3	$x_4 = x_1x_3$	$x_5 = -x_2x_3$	$x_6 = -x_1x_2$
1	+	−	−	−	+	−	−
2	+	+	−	−	−	−	+
3	+	−	+	−	+	+	+
4	+	+	+	−	−	+	−
5	+	−	−	+	−	+	−
6	+	+	−	+	+	+	+
7	+	−	+	+	−	−	+
8	+	+	+	+	+	−	−

69. ДФЭ типа 2^{7-3} с ООК: $1 = x_1x_2x_3x_4x_5 = -x_2x_3x_4x_6 = -x_1x_2x_3x_7 = -x_1x_5x_6 = -x_4x_5x_7 = x_1x_4x_6x_7 = x_2x_3x_5x_6x_7$.

№ оп.	x_0	x_1	x_2	x_3	x_4	$x_5 = x_1x_2x_3x_4$	$x_6 = -x_2x_3x_4$	$x_7 = -x_1x_2x_3$
1	+	−	−	−	−	+	+	+
2	+	+	−	−	−	−	+	−
3	+	−	+	−	−	−	−	−
4	+	+	+	−	−	+	−	+
5	+	−	−	+	−	−	−	−
6	+	+	−	+	−	+	−	+
7	+	−	+	+	−	+	+	+
8	+	+	+	+	−	−	+	−
9	+	−	−	−	+	−	−	+
10	+	+	−	−	+	+	−	−
11	+	−	+	−	+	+	+	−
12	+	+	+	−	+	−	+	+
13	+	−	−	+	+	+	+	−
14	+	+	−	+	+	−	+	+
15	+	−	+	+	+	−	−	+
16	+	+	+	+	+	+	−	−

70. а) ДФЭ 2^{5-2}; б) ДФЭ 2^{5-2} с ГС $x_4 = \pm x_1x_2x_3$, $x_5 = \pm x_1x_3$ или $x_4 = \pm x_1x_3$, $x_5 = \pm x_1x_2x_3$; в) ДФЭ 2^{5-1}.

Планирование 2-го порядка

Если линейное уравнение регрессии оказывается неадекватным, то используются планы 2-го порядка, которые предназначены для получения регрессионной модели в виде полинома 2-го порядка (20). Для построения линейного уравнения регрессии независимые факторы должны варьироваться не менее чем на двух уровнях, а для получения квадратичной модели — как минимум на трех уровнях. (В случае одного фактора для построения прямой необходимо две точки, для построения параболы — три точки.)

Трехуровневый план, в котором реализуются все возможные комбинации k факторов на 3 уровнях, представляет собой **полный факторный эксперимент типа** 3^k.

№ оп.	x_0	x_1	x_2
1	1	−1	−1
2	1	0	−1
3	1	+1	−1
4	1	−1	0
5	1	0	0
6	1	+1	0
7	1	−1	+1
8	1	0	+1
9	1	+1	+1

Например, в таблице приведена матрица плана ПФЭ 3^2 в кодированных переменных. Этот эксперимент содержит 9 опытов. Уравнение, для получения которого он предназначен, имеет $d = 6$ членов и записывается как

$$\widehat{y} = b_0 + b_1x_1 + b_2x_2 + b_{12}x_1x_2 + b_{11}x_1^2 + b_{22}x_2^2. \qquad (35)$$

В случае $k = 3$ факторов квадратичное уравнение регрессии

$$\widehat{y} = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_{12}x_1x_2 + b_{13}x_1x_3 + b_{23}x_2x_3 + \\ + b_{11}x_1^2 + b_{22}x_2^2 + b_{33}x_3^2 \qquad (36)$$

имеет $d = 10$ коэффициентов, а число опытов ПФЭ 3^3 равно $N = 27$.

В общем случае ПФЭ 3^k содержит $N = 3^k$ опытов. С ростом числа факторов количество опытов в ПФЭ 3^k резко возрастает и значительно превосходит число d коэффициентов квадратичной модели, поэтому ПФЭ 3^k на практике применяется очень редко. Для уменьшения числа опытов в плане 2-го порядка используется идея композиционного планирования.

Композиционные планы 2-го порядка получаются путем добавления дополнительных экспериментальных точек (т. е. опытов) к плану 1-го порядка (т. е. к плану, предназначенному для получения линейной регрессионной модели).

Такая достройка плана очень удобна, если вид модели заранее неизвестен. Тогда на начальном этапе строят линейную модель, а если она оказывается неадекватной, проводят дополнительные опыты и получают модель в виде квадратичного уравнения регрессии.

Если дополнительные экспериментальные точки расположены симметрично вокруг центра плана 1-го порядка (т. е. основного уровня по каждому из факторов), то такой план 2-го порядка называется **центрально-композиционным**.

Центрально-композиционный план (ЦКП) 2-го порядка состоит из трех частей. В качестве **ядра** плана 2-го порядка используется ПФЭ 2^k или ДФЭ 2^{k-p}, позволяющий получить независимые раздельные (несмешанные) оценки всех линейных эффектов и эффектов взаимодействия. (Как правило, если число факторов $k \leqslant 4$, используют ПФЭ 2^k; при $5 \leqslant k \leqslant 7$ — ДФЭ 2^{k-1}; если $k > 7$ — ДФЭ 2^{k-2}.)

К ядру плана добавляют n_0 опытов в центре плана (в точке $(0; 0; 0; ...; 0)$ факторного пространства в кодированных переменных) и $2k$ **«звездных точек»**, расположенных на координатных осях факторного пространства в кодированных переменных. Координаты звездных точек: $(\pm\alpha; 0; 0; ...; 0)$; $(0; \pm\alpha; 0; ...; 0)$; ...; $(0; 0; 0; ...; \pm\alpha)$, где α — расстояние от центра плана до звездной точки, называемое **звездным плечом**.

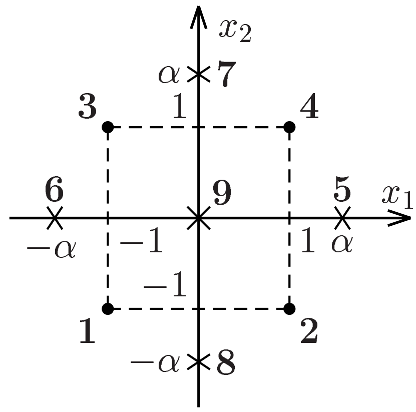
Таким образом, общее число опытов в ЦКП равно

$$N = N_1 + n_0 + 2k,$$

где $N_1 = 2^{k-p}$ — количество опытов в ядре плана.

Число n_0 опытов в центре плана и величину звездного плеча α выбирают исходя из заданного критерия оптимальности планирования, т. е. так, чтобы план обладал заданным свойством оптимальности или был близок к оптимальному.

На рисунке дана геометрическая интерпретация ЦКП в случае двух факторов. Точки 1, 2, 3, 4 образуют ПФЭ 2^2 и составляют ядро плана 2-го порядка. Точки 5, 6, 7, 8 — звездные точки. Опыту в центре плана соответствует точка 9.



Запишем матрицу базисных функций и систему нормальных уравнений для определения по ЦКП коэффициентов квадратичного уравнения регрессии

$$\widehat{y} = b_0 + \sum_{j=1}^{k} b_j x_j + \sum_{j<t} b_{jt} x_j x_t + \sum_{j=1}^{k} b_{jj} x_j^2. \tag{37}$$

Для наглядности рассмотрим случай $k = 2$. Квадратичная модель для случая двух факторов имеет вид (35). Поскольку в уравнении 6 коэффициентов, матрица базисных функций будет содержать 6 столбцов. Если положить $n_0 = 1$, то ЦКП в случае двух факторов содержит $N = 9$ опытов.

№ оп.	x_0	x_1	x_2	x_1x_2	x_1^2	x_2^2	Примечания
1	1	-1	-1	1	1	1	ядро
2	1	1	-1	-1	1	1	плана —
3	1	-1	1	-1	1	1	ПФЭ 2^2
4	1	1	1	1	1	1	
5	1	α	0	0	α^2	0	
6	1	$-\alpha$	0	0	α^2	0	звездные
7	1	0	α	0	0	α^2	точки
8	1	0	$-\alpha$	0	0	α^2	
9	1	0	0	0	0	0	центр плана

Можно показать, что система нормальных уравнений для определения по результатам ЦКП коэффициентов квадратичного уравнения регрессии в общем случае имеет вид

$$
\begin{cases}
Nb_0 + (N_1 + 2\alpha^2)\sum\limits_{j=1}^{k} b_{jj} = \sum\limits_{i=1}^{N} y_i,\\
(N_1 + 2\alpha^2) b_j = \sum\limits_{i=1}^{N} x_{ji} y_i, \quad 1 \leqslant j \leqslant k,\\
N_1 b_{jt} = \sum\limits_{i=1}^{N} x_{ji} x_{ti} y_i, \quad 1 \leqslant j < t \leqslant k,\\
(N_1 + 2\alpha^2) b_{jj} + N_1 \sum\limits_{\substack{t=1\\ t \neq j}}^{k} b_{tt} = \sum\limits_{i=1}^{N} x_{ji}^2 y_i, \quad 1 \leqslant j \leqslant k.
\end{cases}
\tag{38}
$$

Выбирая определенным образом величину звездного плеча α и число n_0 опытов в центре плана, получают планы, удовлетворяющие тем или иным критериям оптимальности. Укажем способы выбора параметров планирования для получения планов, обладающих свойствами ротатабельности, D-оптимальности (точнее, планов, близких к D-оптимальным) и ортогональности.

Если предполагается использовать уравнение регрессии для прогнозирования отклика y при тех или иных значениях факторов, то стремятся, чтобы точность предсказания была одинакова во всех равноудаленных от центра плана точках, т. е. чтобы план удовлетворял критерию ротатабельности.

При построении **центрально-композиционного ротатабельного плана (ЦКРП)**, ядром которого является ДФЭ 2^{k-p}, величина звездного плеча α определяется по формуле

$$\alpha = 2^{(k-p)/4},$$

а число n_0 опытов в центре плана выбирается из условия униформности плана, т. е. чтобы точность предсказания значений отклика была примерно одинакова в окрестности центра плана.

В таблице приведены параметры униформ-ротатабельного ЦКП для числа факторов $2 \leqslant k \leqslant 7$.

k	2	3	4	5 $(p=0)$	5 $(p=1)$	6 $(p=0)$	6 $(p=1)$	7 $(p=1)$
α	$1,414$	$1,682$	2	$2,378$	2	$2,828$	$2,378$	$2,828$
n_0	5	6	7	10	6	15	9	14
N	13	20	31	52	32	91	53	92

Поскольку в центре плана ставится несколько опытов ($n_0 > 1$), это позволяет не проводить дублирование опытов. Для статистического анализа дисперсия воспроизводимости рассчитывается по повторным опытам в центре плана.

Одними из наиболее простых ЦКП 2-го порядка являются B-**планы**, содержащие ядро ПФЭ 2^k или ДФЭ 2^{k-p} и $2k$ звездных точек с $\alpha = 1$. Как правило, для B-планов $n_0 = 0$, но иногда добавляют опыты в центре плана ($n_0 > 0$).

B-планы часто используются, поскольку отличаются простотой построения и удовлетворяют критерию D-оптимальности (оказываются близкими к D-оптимальным планам).

Для построения **центрально-композиционного ортогонального плана (ЦКОП)** используют специальное преобразование базисных функций: вместо квадратичных членов x_j^2 вводят новые переменные x_j' по формуле

$$x_j' = x_j^2 - \frac{1}{N}\sum_{i=1}^{N} x_{ji}^2 = x_j^2 - \frac{N_1 + 2\alpha^2}{N}. \qquad (39)$$

Таким образом, вместо квадратичного уравнения (37) следует определить коэффициенты модели

$$\widehat{y} = b_0' + \sum_{j=1}^{k} b_j x_j + \sum_{j<t} b_{jt} x_j x_t + \sum_{j=1}^{k} b_{jj}' x_j'. \qquad (40)$$

По полученному уравнению (40) легко можно определить коэффициенты уравнения (37), поскольку

$$b_{jj} = b_{jj}'; \qquad b_0 = b_0' - \frac{N_1 + 2\alpha^2}{N}\sum_{j=1}^{k} b_{jj}'.$$

Для получения ортогонального плана величину звездного плеча α определяют по формуле

$$\alpha = \sqrt{\frac{1}{2}\cdot\left(\sqrt{N\cdot N_1} - N_1\right)}.$$

В таблице приведены параметры ЦКОП для числа факторов $2 \leqslant k \leqslant 7$, если в центре плана ставится один опыт ($n_0 = 1$).

k	2	3	4	5 ($p=0$)	5 ($p=1$)	6 ($p=0$)	6 ($p=1$)	7 ($p=1$)
α	1	1,215	1,414	1,596	1,547	1,761	1,724	1,885
N	9	15	25	43	27	77	45	79

Можно показать, что в случае ЦКОП коэффициенты уравнения (40) рассчитываются независимо друг от друга по следующим формулам:

$$b'_0 = \frac{1}{N}\sum_{i=1}^{N} y_i; \qquad b_j = \frac{1}{N_1 + 2\alpha^2}\sum_{i=1}^{N} x_{ji} y_i;$$
$$b_{jt} = \frac{1}{N_1}\sum_{i=1}^{N} x_{ji} x_{ti} y_i; \qquad b'_{jj} = \frac{1}{2\alpha^4}\sum_{i=1}^{N} x'_{ji} y_i, \tag{41}$$

где y_i, x_{ji}, x_{ti}, x'_{ji} — соответственно значения отклика, факторов x_j и x_t, переменных x'_{ji} в i-м опыте. Переменные x'_{ji} определяются формулой (39).

Примеры решения задач

Задача 71. При дозревании суперфосфата, получаемого из апатита, исследовалась зависимость коэффициента разложения y, %, от времени хранения (X_1, сут), при различном содержании H_2SO_4 (X_2, мас. ч.). Получить квадратичное уравнение регрессии по данным в таблице результатам эксперимента, проведенного по ЦКОП при $X_1^0 = 12,5$, $\Delta X_1 = 7,5$, $X_2^0 = 66$, $\Delta X_2 = 6$. Указать условия проведения опытов.

№ оп.	x_1	x_2	y
1	−1	−1	76
2	+1	−1	85
3	−1	+1	91
4	+1	+1	96
5	+1	0	91
6	−1	0	85
7	0	+1	93
8	0	−1	81
9	0	0	88

Решение. Даны условия проведения опытов в кодированных переменных. Связь между кодированными и натуральными переменными определяется, согласно (30), формулами

$$x_1 = \frac{X_1 - 12,5}{7,5}; \quad x_2 = \frac{X_2 - 66}{6}.$$

Кодированному значению $x_j = 0$ соответствует натуральное значение $X_j = X_j^0$; значению $x_j = +1$ соответствует $X_j = X_j^{\max}$; значению $x_j = -1$ — значение $X_j = X_j^{\min}$. Следовательно,

$$X_1^{\max} = 12,5 + 7,5 = 20; \qquad X_1^{\min} = 12,5 - 7,5 = 5;$$

$$X_2^{\max} = 66 + 6 = 72; \qquad X_2^{\min} = 66 - 6 = 60.$$

Факторы	X_1	X_2
X_j^0	12,5	66
ΔX_j	7,5	6
$X_j^{\max}$	20	72
$X_j^{\min}$	5	60

Уровни и интервалы варьирования факторов занесем в таблицу.

Требуется получить квадратичное уравнение вида (35). Для этого, поскольку даны результаты ЦКОП, рассчитаем по формулам (41) коэффициенты вспомогательного уравнения

$$\widehat{y} = b_0' + b_1x_1 + b_2x_2 + b_{12}x_1x_2 + b_{11}'x_1' + b_{22}'x_2'.$$

Определим параметры данного ЦКОП. Имеем $k = 2$ фактора, число опытов в центре плана $n_0 = 1$. Ядром плана является ПФЭ типа 2^2, содержащий $N_1 = 2^2 = 4$ опыта. Общее число опытов ЦКОП $N = N_1 + 2k + n_0 = 4 + 2 \cdot 2 + 1 = 9$. Величина звездного плеча $\alpha = 1$. Вспомогательные переменные x_j' вводятся по формулам (39), которые в нашем случае примут вид

$$x_j' = x_j^2 - \frac{2}{3}.$$

Запишем в таблицу условия проведения опытов, а также матрицу базисных функций ЦКОП в кодированных переменных.

№ оп.	x_0	x_1	x_2	x_1x_2	x_1'	x_2'	X_1	X_2	y
1	$+1$	-1	-1	$+1$	$1/3$	$1/3$	5	60	76
2	$+1$	$+1$	-1	-1	$1/3$	$1/3$	20	60	85
3	$+1$	-1	$+1$	-1	$1/3$	$1/3$	5	72	91
4	$+1$	$+1$	$+1$	$+1$	$1/3$	$1/3$	20	72	96
5	$+1$	$+1$	0	0	$1/3$	$-2/3$	20	66	91
6	$+1$	-1	0	0	$1/3$	$-2/3$	5	66	85
7	$+1$	0	$+1$	0	$-2/3$	$1/3$	12,5	72	93
8	$+1$	0	-1	0	$-2/3$	$1/3$	12,5	60	81
9	$+1$	0	0	0	$-2/3$	$-2/3$	12,5	66	88
b_j	87,33	3,33	6,33	-1	0	-1			

Подставляя в формулы (41) значения отклика и кодированные значения факторов из таблицы, рассчитаем коэффициенты уравнения регрессии:

$$b_0' = \frac{1}{9}(76 + 85 + 91 + 96 + 91 + 85 + 93 + 81 + 88) = 87,33;$$

$$b_1 = \frac{1}{6}(-76 + 85 - 91 + 96 + 91 - 85) = 3,33;$$

$$b_2 = \frac{1}{6}(-76 - 85 + 91 + 96 + 93 - 81) = 6,33;$$

$$b_{12} = \frac{1}{4}(76 - 85 - 91 + 96) = -1;$$

$$b_{11}' = \frac{1}{2}\left(\frac{1}{3} \cdot 76 + \frac{1}{3} \cdot 85 + \frac{1}{3} \cdot 91 + \frac{1}{3} \cdot 96 + \frac{1}{3} \cdot 91 + \frac{1}{3} \cdot 85 - \right.$$

$$-\frac{2}{3}\cdot 93-\frac{2}{3}\cdot 81-\frac{2}{3}\cdot 88\Big)=0;$$

$$b'_{22}=\frac{1}{2}\Big(\frac{1}{3}\cdot 76+\frac{1}{3}\cdot 85+\frac{1}{3}\cdot 91+\frac{1}{3}\cdot 96-\frac{2}{3}\cdot 91-\frac{2}{3}\cdot 85+$$

$$+\frac{1}{3}\cdot 93+\frac{1}{3}\cdot 81-\frac{2}{3}\cdot 88\Big)=-1.$$

Следовательно, уравнение регрессии имеет вид

$$\widehat{y}=87{,}33+3{,}33x_1+6{,}33x_2-x_1x_2-x'_1.$$

Подставив в это уравнение выражение для x'_1, получим квадратичное уравнение регрессии в кодированных переменных:

$$\widehat{y}=87{,}33+3{,}33x_1+6{,}33x_2-x_1x_2-\left(x_1^2-\frac{2}{3}\right).$$

Наконец, подставив формулы, связывающие кодированные и натуральные значения факторов, получим квадратичное уравнение в натуральных переменных:

$$\widehat{y}=87{,}33+3{,}33\cdot\frac{X_1-12{,}5}{7{,}5}+6{,}33\cdot\frac{X_2-66}{6}-$$

$$-\frac{X_1-12{,}5}{7{,}5}\cdot\frac{X_2-66}{6}-\left(\left(\frac{X_1-12{,}5}{7{,}5}\right)^2-\frac{2}{3}\right).$$

Задача 72. Качество беления ткани зависит от стабильности концентрации y раствора в пропиточной ванне. Предложить наиболее экономный по количеству опытов центрально-композиционный план эксперимента, позволяющий получить квадратичное уравнение зависимости y от количества питающего раствора X_1, л/мин, подаваемого в ванну в единицу времени, массы ткани в единице ее длины X_2, г/м, и концентрации подаваемого раствора X_3, г/л; указать условия проведения опытов. Базовые уровни и интервалы варьирования факторов приведены в таблице.

Факторы	X_1	X_2	X_3
X_j^0	1,12	150	200
ΔX_j	0,22	30	20

Решение. Имеем случай $k=3$ факторов. Квадратичное уравнение регрессии имеет вид (36) и содержит 10 коэффициентов. Следовательно, в эксперименте должно быть не менее 10 опытов.

Центрально-композиционный план 2-го порядка состоит из трех частей: ядро плана — ПФЭ 2^3, содержащий $N_1 = 2^3 = 8$ опытов, $2k = 6$ звездных точек и n_0 опытов в центре плана. Всего

$$N = N_1 + 2k + n_0 = 14 + n_0$$

опытов. Наименьшее число опытов достигается при отсутствии их в центре плана, т. е. при $n_0 = 0$.

Определим параметры центрально-композиционных планов, удовлетворяющих различным критериям оптимальности.

Для униформ-ротатабельного плана в случае $k = 3$ факторов следует взять число опытов в центре плана $n_0 = 6$ и величину звездного плеча $\alpha = 1,682$. Число опытов такого эксперимента будет $N = 20$. Таким образом, униформ-ротатабельный план не удовлетворяет условию минимизации количества опытов.

В случае B-плана $n_0 = 0$, $\alpha = 1$. Число опытов $N = 14$.

ЦКОП с $n_0 = 0$ также имеет $N = 14$ опытов, величина звездного плеча α для этого плана определяется по формуле

$$\alpha = \sqrt{\frac{1}{2} \cdot \left(\sqrt{14 \cdot 8} - 8\right)} = 1,136.$$

Таким образом, для решения задачи можно взять B-план или ЦКОП с $n_0 = 0$ и $\alpha = 1,136$. Выберем B-план как более простой.

Формулы связи между натуральными и кодированными значениями факторов (30) в нашем случае примут вид

$$x_1 = \frac{X_1 - 1,12}{0,22}; \quad x_2 = \frac{X_2 - 150}{30}; \quad x_3 = \frac{X_3 - 200}{20}.$$

Уровни и интервалы варьирования факторов занесем в таблицу. В случае B-плана каждый фактор изменяется только на трех уровнях: верхнем, нижнем и базовом, которым соответствуют кодированные значения $+1$, -1 и 0.

Факторы	X_1	X_2	X_3
X_j^0	1,12	150	200
ΔX_j	0,22	30	20
$X_j^{\max}$	1,34	170	220
$X_j^{\min}$	0,9	120	180

Запишем условия проведения опытов в кодированных и натуральных переменных в таблицу.

Итак, в первом опыте следует подавать $X_1 = 0,9$ л питающего раствора в минуту, взять массу ткани в единице ее длины $X_2 = 120$ г/м, концентрацию подаваемого раствора $X_3 = 180$ г/л; второй опыт провести при тех же значениях $X_2 = 120$ г/м и

B-план	Кодированные значения			Натуральные значения		
№ оп.	x_1	x_2	x_3	X_1	X_2	X_3
1	-1	-1	-1	$0,9$	120	180
2	$+1$	-1	-1	$1,34$	120	180
3	-1	$+1$	-1	$0,9$	170	180
4	$+1$	$+1$	-1	$1,34$	170	180
5	-1	-1	$+1$	$0,9$	120	220
6	$+1$	-1	$+1$	$1,34$	120	220
7	-1	$+1$	$+1$	$0,9$	170	220
8	$+1$	$+1$	$+1$	$1,34$	170	220
9	$+1$	0	0	$1,34$	150	200
10	-1	0	0	$0,9$	150	200
11	0	$+1$	0	$1,12$	170	200
12	0	-1	0	$1,12$	120	200
13	0	0	$+1$	$1,12$	150	220
14	0	0	-1	$1,12$	150	180

$X_3 = 180$ г/л, но увеличить скорость подачи питающего раствора до $X_1 = 1,34$ л/мин; третий опыт провести при $X_1 = 0,9$ л/мин, $X_2 = 170$ г/м, $X_3 = 180$ г/л и т. д.; в последнем опыте подавать $X_1 = 1,12$ л питающего раствора в минуту, взять массу ткани в единице ее длины $X_2 = 150$ г/м, концентрацию подаваемого раствора $X_3 = 180$ г/л.

Задача 73. В условиях задачи **72** предложить план эксперимента, если предполагается по квадратичному уравнению регрессии прогнозировать значения отклика в области планирования.

Решение. Учитывая, что уравнение регрессии предполагается использовать для прогнозирования значений отклика, используем униформ-ротатабельный план, позволяющий получить уравнение регрессии, точность предсказания значений отклика по которому примерно одинакова в окрестности центра плана.

При решении задачи **72** были определены параметры этого плана для случая $k = 3$ факторов: число опытов в центре плана $n_0 = 6$, величина звездного плеча $\alpha = 1,682$. Число опытов такого эксперимента равно $N = 20$.

Матрица плана в кодированных переменных приведена в таблице. При решении задачи **72** были записаны формулы, связывающие кодированные и натуральные значения переменных. Выражая из этих формул X_1, X_2, X_3, получим формулы для расчета натуральных значений факторов:

$$X_1 = 1,12 + 0,22x_1,\ X_2 = 150 + 30x_2,\ X_3 = 200 + 20x_3.$$

ЦКРП	Кодированные значения			Натуральные значения		
№ оп.	x_1	x_2	x_3	X_1	X_2	X_3
1	-1	-1	-1	$0,9$	120	180
2	$+1$	-1	-1	$1,34$	120	180
3	-1	$+1$	-1	$0,9$	170	180
4	$+1$	$+1$	-1	$1,34$	170	180
5	-1	-1	$+1$	$0,9$	120	220
6	$+1$	-1	$+1$	$1,34$	120	220
7	-1	$+1$	$+1$	$0,9$	170	220
8	$+1$	$+1$	$+1$	$1,34$	170	220
9	$+1,682$	0	0	$1,49$	150	200
10	$-1,682$	0	0	$0,75$	150	200
11	0	$+1,682$	0	$1,12$	200	200
12	0	$-1,682$	0	$1,12$	100	200
13	0	0	$+1,682$	$1,12$	150	234
14	0	0	$-1,682$	$1,12$	150	166
15–20	0	0	0	$1,12$	150	200

Условия проведения опытов в натуральных переменных приведены в таблице. Например, в 9-м опыте, который соответствует звездной точке, следует подавать $X_1 = 1,12 + 0,22 \cdot 1,682 \approx 1,49$ л питающего раствора в минуту, взять массу ткани в единице ее длины $X_2 = 150$ г/м, концентрацию подаваемого раствора $X_3 = 180$ г/л. Заметим, что проводится 6 опытов при одних и тех же условиях $X_1 = 1,12$ л/мин, $X_2 = 150$ г/м, $X_3 = 200$ г/л (в центре плана).

Задачи для самостоятельного решения

Задача 74. По данным задачи **55** предложить наиболее экономный по количеству опытов центрально-композиционный план эксперимента, позволяющий получить квадратичное уравнение зависимости; указать условия проведения опытов.

Задача 75. По данным задачи **55** предложить план эксперимента, если предполагается по квадратичному уравнению регрессии прогнозировать значения отклика в области планирования; указать условия проведения опытов.

Задача 76. Определить параметры наиболее экономного по количеству опытов центрально-композиционного плана эксперимента, позволяющего получить квадратичное уравнение зависимости в случае 4 факторов. Сколько опытов содержит такой эксперимент?

Задача 77. Определить параметры ЦКРП 2-го порядка в случае 4 факторов.

Задача 78. Определить параметры наиболее экономного по количеству опытов ЦКРП 2-го порядка в случае 5 факторов. Сколько опытов содержит такой эксперимент?

Ответы к задачам 74–78.

74. B-план с $n_0 = 0$ или ЦКОП с $n_0 = 0$, $\alpha = 0,91$.

Факторы	X_1	X_2
X_j^0	50	25
ΔX_j	5	1
$X_j^{\max}$	55	26
$X_j^{\min}$	45	24

B-план	Кодированные значения		Натуральные значения	
№ оп.	x_1	x_2	X_1	X_2
1	-1	-1	45°C	24 %
2	$+1$	-1	55°C	24 %
3	-1	$+1$	45°C	26 %
4	$+1$	$+1$	55°C	26 %
5	$+1$	0	55°C	25 %
6	-1	0	45°C	25 %
7	0	$+1$	50°C	26 %
8	0	-1	50°C	24 %

75. ЦКРП с $n_0 = 5$, $\alpha = 1,414$.

Факторы	X_1	X_2
X_j^0	50	25
ΔX_j	5	1
$X_j^{\max}$	55	26
$X_j^{\min}$	45	24

ЦКРП	Кодированные значения		Натуральные значения	
№ оп.	x_1	x_2	X_1	X_2
1	-1	-1	45°C	24 %
2	$+1$	-1	55°C	24 %
3	-1	$+1$	45°C	26 %
4	$+1$	$+1$	55°C	26 %
5	$+1,414$	0	57,1°C	25 %
6	$-1,414$	0	42,9°C	25 %
7	0	$+1,414$	50°C	26,4 %
8	0	$-1,414$	50°C	23,6 %
9–13	0	-1	50°C	24 %

76. 24 опыта: B-план с $n_0 = 0$ или ЦКОП с $n_0 = 0$, $\alpha = 1,341$. **77.** $n_0 = 7$, $\alpha = 2$, $N = 31$. **78.** $n_0 = 6$, $\alpha = 2$, $N = 32$.

Раздел 4. ЭКСПЕРИМЕНТАЛЬНЫЕ МЕТОДЫ ПОИСКА ОПТИМАЛЬНЫХ УСЛОВИЙ

Постановка задачи экспериментальной оптимизации

Задача оптимизации — одна из наиболее распространенных и важных задач прикладных исследований. Задача оптимизации ставится следующим образом: при исследовании зависимости параметра y от нескольких факторов $X_1, X_2, \ldots, X_k$ требуется определить те значения факторов, при которых параметр y принимает максимальное (или минимальное) значение.

Например, для качества черепицы и фасадных керамических изделий существенны такие показатели, как морозостойкость, водопоглощение, прочность, зависящие в свою очередь от состава формовочной смеси, количества различных добавок, температуры обжига и других технологических параметров. Таким образом, может быть поставлена задача определения оптимальных технологических условий изготовления черепицы, при которых получаются изделия с наибольшей морозостойкостью или с наименьшим водопоглощением.

Как правило, в практических задачах вид теоретической функции отклика неизвестен и нет возможности получить достаточно хорошее приближение данной функции в интересующей исследователя области изменения факторов (для этого требуется много опытов), можно только получить некоторое приближение функции отклика в локальной области варьирования факторов. Задачу оптимизации приходится решать экспериментальным путем.

Большинство методов экспериментальной оптимизации основано на принципе последовательного проведения экспериментов. Это означает, что условия проведения очередной серии опытов определяются каждый раз по результатам предыдущей серии.

Метод крутого восхождения

Метод крутого восхождения (МКВ), предложенный американскими учеными Дж. Боксом и К. Уилсоном в 1951 г., основан на свойстве градиента функции нескольких переменных: градиент направлен в сторону наискорейшего возрастания функции.

В математическом анализе градиент функции нескольких переменных определяется как вектор, координаты которого равны частным производным этой функции. Согласно МКВ, для оценки градиента функции отклика в некоторой точке факторного пространства по результатам эксперимента с центром в этой точке (как правило, ПФЭ или ДФЭ) получают линейное уравнение регрессии и определяют градиент линейного приближения.

Для увеличения параметра оптимизации y надо двигаться в направлении градиента, для уменьшения y — в противоположном направлении. Движение в направлении градиента (в случае поиска минимума — в направлении антиградиента) осуществляется до достижения частного экстремума в данном направлении, после чего нужно снова определить направление градиента и повторить указанную процедуру.

Описанная процедура должна быть многократной, поскольку, сделав шаг к экстремуму, мы попадаем в точку, где направление градиента, вообще говоря, другое. Для его оценки надо вновь поставить эксперимент.

Опишем основные этапы МКВ.

1. Выбор начальной точки $X^0 = (X_1^0; X_2^0; \ldots; X_k^0)$, т. е. базовых уровней всех факторов, и интервалов варьирования факторов ΔX_j.

Если имеется априорная информация, какие-то проведенные ранее опыты, то в качестве начальной точки выбирают точку, отвечающую наилучшему из предварительных опытов. Если априорной информации нет, в качестве начальной точки берут центр исследуемой области.

Интервалы варьирования факторов выбирают не слишком малыми, чтобы изменение отклика было больше ошибки в определении отклика, и не слишком большими, чтобы получить более точную оценку направления градиента.

По мере приближения к экстремуму, когда область поиска уменьшается, интервалы варьирования факторов следует уменьшать.

2. Определяют верхний и нижний уровни для каждого фактора и проводят ПФЭ или ДФЭ для определения компонент градиента; по результатам эксперимента получают линейное уравнение регрессии в кодированных переменных:

$$\widehat{y} = b_0 + b_1x_1 + b_2x_2 + \ldots + b_kx_k.$$

Преимущество от использования ПФЭ или ДФЭ заключается в том, что коэффициенты уравнения регрессии рассчитываются независимо друг от друга и с наибольшей точностью.

3. Вычисляют интервалы изменения натуральных переменных $b_j \Delta X_j$; выбирают **базовый фактор** X_{j^*}, для которого

$$|b_{j^*} \Delta X_{j^*}| > |b_j \Delta X_j|;$$

выбирают шаг варьирования δ_{j^*} для базового фактора, причем **знак** δ_{j^*} должен совпадать со знаком $b_{j^*} \Delta X_{j^*}$ при поиске максимума и быть противоположным знаку $b_{j^*} \Delta X_{j^*}$ при поиске минимума.

Для остальных факторов интервалы варьирования определяются из пропорции

$$\frac{\delta_j}{b_j \Delta X_j} = \frac{\delta_{j^*}}{b_{j^*} \Delta X_{j^*}} \Rightarrow \delta_j = b_j \Delta X_j \cdot \frac{\delta_{j^*}}{b_{j^*} \Delta X_{j^*}}.$$

4. Намечаются условия серии опытов в направлении градиента (или антиградиента в задаче минимизации):

$$X_j^{(k)} = X_j^0 + k\delta_j, \quad k = 1, 2, 3, \ldots$$

5. Движение в направлении градиента осуществляют до достижения частного экстремума в данном направлении, т. е. пока значения отклика не начнут ухудшаться.

6. Точку частного экстремума на первоначальном направлении выбирают в качестве начальной точки и организуют второй цикл крутого восхождения.

Признаком достижения области экстремума является незначимость коэффициентов уравнения регрессии или если движение в направлении градиента не приводит к улучшению результата.

Симплексный метод поиска экстремума

Этот метод планирования многофакторного эксперимента возник в 1962 г. Он предусматривает экономный многошаговый процесс движения к экстремуму поверхности отклика с одновременным, если это необходимо, получением линейного уравнения регрессии в локальной области изменения факторов.

Условия проведения опытов в ПФЭ геометрически изображаются в виде многомерного куба в факторном пространстве. Название симплекс-метода (от лат. simplex — простой) объясняется тем, что экспериментальные точки образуют простейшую фигуру в факторном пространстве — симплекс.

k-мерным симплексом называется многогранник в k-мерном факторном пространстве, имеющий $k+1$ вершину. Например, в случае двух факторов это будет треугольник, в случае трех факторов — треугольная пирамида. **Регулярным** называется симплекс, у которого все ребра равны, т. е. все вершины находятся друг от друга на одинаковом расстоянии. Двумерный правильный симплекс — равносторонний треугольник, трехмерный — тетраэдр.

При планировании экспериментов обычно используют регулярные симплексы в кодированных переменных.

Симплексный метод поиска экстремума заключается в следующем. Проводится эксперимент, состоящий из $k+1$ опытов, условия которых соответствуют вершинам k-мерного регулярного симплекса в факторном пространстве. Из этих опытов выбирается опыт с наихудшим значением отклика y. Для достижения экстремума надо двигаться в противоположном направлении. Для этого строится новый симплекс: наихудшая точка зеркально отражается относительно противоположной грани.

Координаты новой точки в кодированных переменных вычисляются по формуле

$$x_j^{(\text{нов})} = \frac{2}{k} \sum_{\substack{t=1 \\ t \neq u}}^{k+1} x_j^{(t)} - x_j^{(u)},$$

где j — номер фактора; t — номер опыта; u — номер наихудшего опыта.

Для постановки опытов нужно переходить к натуральным переменным по формулам (30). Можно показать, что координаты новой вершины симплекса в натуральных переменных могут рассчитываться так же, как кодированные:

$$X_j^{(\text{нов})} = \frac{2}{k} \sum_{\substack{t=1 \\ t \neq u}}^{k+1} X_j^{(t)} - X_j^{(u)}.$$

Если одно и то же наихудшее значение отклика наблюдается в нескольких вершинах симплекса, выбирают любую.

Если на очередном шаге в новой вершине отклик оказывается наихудшим, выбирают вершину не с наихудшим, а со вторым из самых плохих значением.

Критерием остановки является переход к вращению симплекса вокруг одной из вершин. Считается, что симплекс начал вращаться, когда в более чем $N_0 = 1,65k + 0,05k^2$ последовательных положениях симплекса сохраняется одна общая вершина. При этом следует повторить наилучший опыт, чтобы выяснить, не вызвано ли завышенное значение отклика случайными помехами.

Условия опыта, давшего самый хороший результат, можно считать оптимальными на данном этапе исследования.

Для уточнения оптимума можно попытаться продолжить процесс поиска экстремума, уменьшив размеры симплекса. Для этого симплекс стягивают к его лучшей вершине, рассчитывая координаты остальных вершин нового симплекса по формуле

$$X_j^{(\text{нов})} = (1 - \alpha)X_j^* + \alpha X_j,$$

где $X_j^{(\text{нов})}$ — j-я координата вершины нового (уменьшенного) симплекса; X_j — j-я координата соответствующей вершины старого симплекса; X_j^* — j-я координата наилучшей вершины. При таком преобразовании ребра исходного симплекса уменьшаются в $1/\alpha$ раз.

Симплексный метод поиска экстремума прост в реализации и не требует сложных расчетов. При большом числе факторов и наличии помех симплекс-метод оказывается эффективнее МКВ. Кроме того, в симплекс-методе параметр оптимизации может измеряться приближенно, поскольку для движения к экстремуму достаточно выбрать опыт с наихудшим результатом. По этой же причине симплекс-метод дает возможность работать с качественными параметрами оптимизации или учитывать несколько параметров оптимизации.

ВОПРОСЫ ПО КУРСУ «ПЛАНИРОВАНИЕ И ОРГАНИЗАЦИЯ ЭКСПЕРИМЕНТА»

Тема 1. Графическое описание статистических рядов.

1. Задачи математической статистики. Генеральная и выборочная совокупности. Повторная и бесповторная выборки. Репрезентативная выборка.

2. Вариационный ряд. Полигон. Гистограмма частот и относительных частот.

3. Эмпирическая функция распределения и ее свойства.

4. Как оценить по выборке функцию распределения и плотность распределения?

Тема 2. Статистическое оценивание параметров.

5. Понятия статистики и точечной оценки параметра. Состоятельные, смещенные и несмещенные оценки. Примеры несмещенных оценок для математического ожидания и дисперсии.

6. Выборочные среднее и дисперсия. Несмещенные оценки для математического ожидания и дисперсии.

7. Интервальные оценки. Доверительный интервал и доверительная вероятность, уровень значимости.

8. Построение доверительного интервала для математического ожидания нормального распределения с известной и неизвестной дисперсией.

9. Построение доверительного интервала для дисперсии нормально распределенной случайной величины.

10. Планирование эксперимента при построении интервальных оценок.

Тема 3. Проверка статистических гипотез.

11. Основные понятия теории проверки гипотез. Простая и сложная гипотезы. Нулевая и альтернативная гипотезы. Статистический критерий. Ошибки первого и второго родов. Уровень значимости и мощность критерия.

12. Проверка гипотезы о виде закона распределения генеральной совокупности. Критерии согласия.

13. Критерии значимости. Проверка гипотез о математических ожиданиях одной и двух независимых нормальных выборок.

14. Критерии значимости. Проверка гипотез о дисперсиях одной и двух независимых нормальных выборок.

15. Критерии значимости. Проверка гипотез о математических ожиданиях двух зависимых и независимых нормальных выборок.

16. Использование распределения Стьюдента при построении доверительных интервалов и проверке статистических гипотез.

17. Использование нормального распределения при построении доверительных интервалов и проверке статистических гипотез.

18. Использование χ^2-распределения при построении доверительных интервалов и проверке статистических гипотез.

19. Использование F-распределения Фишера в задачах проверки статистических гипотез.

20. Проверка однородности дисперсий нескольких независимых нормальных выборок.

Тема 4. Элементы корреляционного анализа. Выборочный коэффициент корреляции.

21. Виды зависимостей между случайными величинами. Основные задачи корреляционного и регрессионного анализа.

22. Выборочный коэффициент корреляции и его свойства.

23. Проверка значимости коэффициента корреляции двух случайных величин, распределенных нормально.

Тема 5. Получение уравнения регрессии методом наименьших квадратов.

24. Отыскание параметров эмпирического уравнения линейной регрессии методом наименьших квадратов.

25. Система нормальных уравнений метода наименьших квадратов.

Тема 6. Метод наименьших квадратов в матричной форме записи.

26. Система нормальных уравнений метода наименьших квадратов в матричной форме записи. Матрица базисных функций и информационная матрица плана.

27. План эксперимента. Понятия активного и пассивного эксперимента.

Тема 7. Оценка воспроизводимости эксперимента.

28. Равномерное и неравномерное дублирование опытов. Способы проверки воспроизводимости эксперимента.

29. Расчет дисперсии воспроизводимости, число степеней свободы дисперсии воспроизводимости.

Тема 8. Статистический анализ эмпирического уравнения регрессии.

30. Проверка значимости коэффициентов регрессии.

31. Проверка адекватности эмпирического уравнения регрессии экспериментальным данным.

Тема 9. Полный факторный эксперимент.

32. Полный факторный эксперимент (ПФЭ) типа 2^k. Свойства ПФЭ типа 2^k.

33. Матрица плана (матрица базисных функций) ПФЭ типа 2^k и ее свойства.

34. Формулы перехода от натуральных переменных к кодированным в ПФЭ типа 2^k.

35. Вычисление оценок коэффициентов уравнения регрессии в кодированных переменных по результатам ПФЭ 2^k, получение уравнения регрессии в натуральных переменных.

Тема 10. Дробный факторный эксперимент.

36. Дробный факторный эксперимент (ДФЭ) типа 2^{k-p}, генерирующие соотношения, определяющий контраст, обобщенный определяющий контраст, система смешивания оценок.

37. Построение и свойства матрицы плана ДФЭ.

Тема 11. Планы 2-го порядка.

38. ПФЭ типа 3^k и типа 2^k, назначение и принцип построения.

39. Идея композиционного планирования. Структура центрально-композиционных планов 2-го порядка.

40. Композиционные планы 2-го порядка: ЦКОП, ЦКРП, B-планы.

41. Принцип построения центрально-композиционных ортогональных планов 2-го порядка.

Тема 12. Свойства планов регрессионных экспериментов.

42. Планы 1-го порядка и планы 2-го порядка. Примеры.

43. Насыщенный план. Примеры.

44. Ортогональные планы. Примеры.

45. Ротатабельные планы. Примеры.

46. Рандомизация.

Тема 13. Экспериментальные методы поиска оптимальных условий.

47. Постановка задачи экспериментальной оптимизации.

48. Метод крутого восхождения.

49. Симплексный метод поиска экстремума.

ЗАДАНИЯ К КОНТРОЛЬНОЙ РАБОТЕ

Контрольная работа состоит из 5 заданий. Номера задач в каждом из заданий выбираются по первым пяти буквам фамилии студента в соответствии со следующей ниже таблицей. Если фамилия студента имеет менее пяти букв, то после фамилии записывается имя и учитываются первые пять букв полученной записи.

1	2	3	4	5	6	7	8	9	10
А Й У	Б К Ф	В Л Х	Г М Ц	Д Н Ч	Е О Ш, Щ	Ё П Ъ, Ы, Ь	Ж Р Э	З С Ю	И Т Я

Например, студентка Блинова Анна (Б–2, Л–3, И–10, Н–5, О–6) выполняет задачи 1.2, 2.3, 3.10, 4.5, 5.6, а студент Кот Иван (К–2, О–6, Т–10, И–10, В–3) — задачи 1.2, 2.6, 3.10, 4.10, 5.3.

Задание 1

В задачах **1.1**–**1.5** по данному интервальному статистическому ряду требуется: 1) найти несмещенные оценки для математического ожидания и дисперсии; 2) построить гистограмму относительных частот; 3) записать эмпирическую функцию распределения и построить ее график.

1.1.

$[x_{i-1}; x_i)$	$[-1; 1)$	$[1; 3)$	$[3; 5)$	$[5; 7)$
n_i	10	45	30	15

1.2.

$[x_{i-1}; x_i)$	$[-1; 1)$	$[1; 3)$	$[3; 5)$	$[5; 7)$
n_i	2	9	6	3

1.3.

$[x_{i-1}; x_i)$	$[-2; 2)$	$[2; 6)$	$[6; 10)$	$[10; 14)$
n_i	20	90	60	30

1.4.

$[x_{i-1}; x_i)$	$[-3; 1)$	$[1; 5)$	$[5; 9)$	$[9; 13)$
n_i	10	45	30	15

1.5.

$[x_{i-1}; x_i)$	$[-10; -6)$	$[-6; -2)$	$[-2; 2)$	$[2; 6)$	$[6; 10)$
n_i	10	15	45	25	5

В задачах **1.6**–**1.10** по данному статистическому ряду требуется: 1) найти несмещенные оценки для математического ожидания и дисперсии; 2) построить полигон частот; 3) записать эмпирическую функцию распределения и построить ее график.

1.6.

x_i	0	1	2	3
n_i	30	45	20	5

1.7.

x_i	1	2	3	4	5	6
n_i	7	5	10	10	15	13

1.8.

x_i	0	1	2	3
n_i	30	50	10	10

1.9.

x_i	0	1	2	3
n_i	35	40	15	10

1.10.

x_i	1	2	3	4	5	6
n_i	13	15	10	10	5	7

Задание 2

2.1. Приведены данные о дополнительных часах сна после употребления снотворных А и В у четырех пациентов. Требуется при уровне значимости $\alpha = 0,05$ проверить, имеется ли существенное различие между действием снотворных средств А и В.

Как изменилась бы процедура проверки гипотезы в случае, если бы в эксперименте участвовали две группы пациентов: одна группа принимала бы снотворное А, а другая — снотворное В?

Пациент	1	2	3	4
Снотворное А	$1,9$	0	$0,3$	$1,6$
Снотворное В	$1,3$	$0,2$	$-0,2$	$1,7$

2.2. Согласно стандарту, сопротивление некоторой проволоки должно быть равно 40 условным единицам. Приведены результаты, полученные при исследовании 5 образцов:

$$36; \quad 37; \quad 41; \quad 39; \quad 37.$$

Можно ли при уровне значимости $\alpha = 0,05$ принять гипотезу о том, что сопротивление проволоки соответствует стандарту или образцы имеют меньшее сопротивление?

2.3. В лаборатории, изучающей воздействие окружающей среды на человека, были исследованы 5 мужчин и 5 женщин для того, чтобы установить комнатную температуру, при которой они чувствуют себя наиболее комфортабельно. Имеются ли при уровне значимости $\alpha = 0,05$ достаточные основания для отклонения гипотезы о том, что в среднем температура наибольшего комфорта для мужчин и женщин одинакова?

Мужчины	23	21	25	24	22
Женщины	24	25	26	26	24

2.4. Шариковые подшипники проходят проверку на овальность в специальном измерительном устройстве, которое автоматически фиксирует отклонение от заданных условий. Возникло подозрение, что время, необходимое для проведения проверки, у разных контролеров различно. Были выбраны 2 контролера, и время (в секундах), необходимое каждому из них на проведение проверки, регистрировалось. Каждый контролер проверил по 6 изделий (всего было проверено 12 изделий). Можно ли при уровне значимости $\alpha = 0,05$ утверждать, что наблюдаются различия в скорости работы контролеров?

Как изменилась бы процедура проверки гипотезы, если бы контролеры проверили одни и те же 6 изделий?

A	13	12	11	10	13	13
B	14	12	16	15	16	17

2.5. Имеются данные (в микронах) об измерениях неровностей поверхностей одних и тех же образцов на двух двойных микроскопах с заводскими номерами № 61 и № 263. Можно ли при уровне значимости $\alpha = 0,05$ считать, что между показаниями приборов нет систематических расхождений?

Микроскоп № 61	$0,9$	$1,9$	$3,0$	$3,5$	$3,8$
Микроскоп № 263	$1,4$	$2,1$	$3,1$	$3,6$	$3,4$

2.6. С помощью двух измерителей длины — механического и оптического — измерили длину 5 осей. Отклонения от размера $12,9$ мм занесены в таблицу, цена деления измерителей $1/100$ мм. Определить при уровне значимости $0,05$, имеются ли существенные различия между результатами, полученными с помощью двух приборов.

№ измерения	1	2	3	4	5
Механический	2	3	4	8	13
Оптический	3	5	1	6	10

2.7. 69 приборов типа A и 12 приборов типа B для измерения осадков были рассредоточены случайным образом на некотором участке. За определенный период времени на этом участке прошло 6 ливней. В таблице приведено среднее количество осадков, замеренное двумя типами приборов. Проверить при уровне значимости $\alpha = 0,05$ гипотезу, что оба типа приборов дают одинаковые результаты.

A	20	32	11	53	5	20
B	26	38	7	50	10	16

2.8. Для проверки работы двух станков проведены измерения размера выпускаемых ими однотипных изделий (в микронах). Можно ли на основании этих величин при уровне значимости $\alpha = 0,05$ сделать вывод о том, что станки выпускают изделия одинакового размера?

1-й станок	210	212	218	208	—
2-й станок	208	212	210	210	215

2.9. Для сравнения удельного веса кирпичей из двух зон обжига (А и В) отобрали и взвесили 16 кирпичей из зоны А и 10 кирпичей из зоны В, отмечая всякий раз отклонения от номинального значения 1800 кГ/м^3. Следует ли считать различие между средними значениями выборок А и В существенным при уровне значимости $\alpha = 0,05$?

Завод	$\overline{x_i}$	s_i^2	n_i
А	$2,43$	$16,4$	16
В	$4,9$	$22,5$	10

2.10. Для определения прочности на разрыв целлофановых мешков разработан специальный критерий. Исследовались 10 мешков типа А и 12 мешков типа В. Каждый из мешков наполняли и бросали до тех пор, пока он не разрывался. Обозначим число падений мешка до момента разрыва через x. Можно ли на уровне значимости $\alpha = 0,05$ говорить, что мешки одного типа прочнее, чем другого?

Завод	$\overline{x_i}$	s_i^2	n_i
А	$75,5$	$83,17$	10
В	$89,3$	$128,2$	12

Задание 3

В задачах **3.1**–**3.8** требуется по приведенным данным проверить воспроизводимость эксперимента на уровне значимости $\alpha = 0,05$, рассчитать дисперсию воспроизводимости и указать ее число степеней свободы.

3.1. В таблице приведены результаты эксперимента по определению зависимости содержания Fe (y, %) в кристаллах медного купороса $CuSO_4 \cdot 5H_2O$ от содержания $FeSO_4$ (x, г/л) в маточном растворе. Каждый опыт повторялся два раза.

№ оп.	x_i	y_{i1}	y_{i2}
1	50	$0,66$	$0,84$
2	60	$0,96$	$0,84$
3	70	$0,96$	$1,2$
4	85	$1,33$	$1,47$
5	100	$1,74$	$1,86$
6	105	$2,32$	$2,48$

3.2. Проведен эксперимент по определению зависимости срока службы вставных токарных резцов от скорости резания при обработке деталей из твердой никелевой стали. Для трех резцов, использованных на низкой скорости резания, срок службы составил 44, 48 и 52 минуты; четыре резца, использованные на средней скорости, работали 34, 33, 31 и 30 минут соответственно; резцы, использованные на высокой скорости, работали 21, 18, 15 и 18 минут.

3.3. При производстве синтетического волокна для уменьшения последующей усадки продукция, движущаяся не-

Температура	y_{i1}	y_{i2}	y_{i3}	y_{i4}
120°C	$3,1$	$3,5$	$3,4$	$3,2$
140°C	$4,3$	$4,1$	$4,5$	$4,7$

прерывным потоком, подвергается термической обработке. Исследовалась зависимость величины усадки (в процентах) от температуры обработки.

3.4. Качество беления ткани зависит от стабильности концентрации y раствора в пропиточной ванне. В таблице приведены результаты эксперимента по исследованию зависимости y от количества питающего раствора X_1, л/мин, подаваемого в ванну в единицу времени, и массы ткани в единице ее длины X_2, г/м.

№ оп.	X_{1i}	X_{2i}	y_{i1}	y_{i2}	y_{i3}	y_{i4}	y_{i5}
1	0,9	108	26,4	26,6	26,7	26,3	—
2	1,34	108	32,2	32,8	32,8	32,4	32,3
3	0,9	186	25	25,2	25,4	—	—

3.5. При разработке рецептуры нового напитка «Пикантный» на основе молочной сыворотки с использованием томатной пасты исследовалась зависимость органолептических показателей качества (y, баллы) от концентрации томатной пасты (X_1, %), соли (X_2, %) и кислотности сыворотки (X_3, °).

№ оп.	X_{1i}	X_{2i}	X_{3i}	y_{i1}	y_{i2}	y_{i3}
1	15	0,25	20	52	53	—
2	20	0,75	20	78	79	77
3	15	0,75	65	37	36	35
4	20	0,25	65	60	64	62

3.6. Исследовалась зависимость производительности пиления древесины ели цепными моторными пилами y, см^2/с, от диаметра реза X_1, см, рабочей длины пильного аппарата X_2, см, и скорости резания X_3, м/с.

№ оп.	X_{1i}	X_{2i}	X_{3i}	y_{i1}	y_{i2}	y_{i3}
1	30	75	11,5	32	33	34
2	50	35	11,5	38	39	37
3	30	35	15,5	57	56	55
4	50	75	15,5	56	54	52

3.7. В таблице приведены результаты эксперимента по определению зависимости степени окисления (y, %) хромита Cr_2O_3 в хромат CrO_3 от продолжительности прокаливания (x, ч) шихты при 830°. Каждый опыт повторялся два раза.

№ оп.	x_i	y_{i1}	y_{i2}
1	0,3	8,2	12,2
2	1,2	28	32
3	2	44,2	47,2
4	3	49,3	51,3
5	4	53	58
6	5	56,4	57,4

3.8. При получении фосфора возгонкой из фосфатов кальция исследовалась зависимость степени восстановления фосфата от температуры. Получены следующие результаты: при температуре 1100°C —

$8,5\%$, $11,6\%$ и $10,2\%$; при температуре 1150°C — $29,5\%$, $30,6\%$ и $28,4\%$; при температуре 1200°C — $50,4\%$, 50% и 52%.

3.9. Для сравнения точности станков, производящих одинаковую продукцию, было отобрано 6 проб, по одной пробе с каждого станка. С 1-го станка было взято 12 единиц продукции, со 2-го — 16 единиц и т. д. Числовые характеристики, полученные по каждой из выборок, приведены в таблице. Проверить при уровне значимости $0,05$ гипотезу об однородности полученных дисперсий. Для характеристики средней точности станков вычислить общую дисперсию и указать ее число степеней свободы.

№ станка	1	2	3	4	5	6
$\overline{x}_i$	$45,3$	$46,1$	$45,8$	$46,2$	$45,5$	$45,8$
s_i^2	$1,2$	$0,8$	$0,4$	$0,45$	$1,2$	$1,6$
n_i	12	16	12	10	15	5

3.10. Станок в автоматическом токарном цехе производит цилиндрические болты определенного типа. Для проверки равномерности работы станка за 6 смен работы было отобрано 6 проб, по одной пробе в смену, численностью 16 болтов каждая. По данным каждой из этих проб подсчитаны оценки s_i^2 дисперсий (в миллиметрах квадратных):

$$0,000067;\ 0,000162;\ 0,000144;\ 0,000102;\ 0,000088;\ 0,000067.$$

Проверить при $\alpha = 0,05$ гипотезу об отсутствии разладки станка по рассеиванию размеров болтов за 6 смен работы (об однородности полученных дисперсий). Для характеристики точности станка вычислить общую дисперсию и указать ее число степеней свободы.

Задание 4

№ оп.	X_{1i}	X_{2i}	y_{i1}	y_{i2}	y_{i3}
1	-1	0	1	2	3
2	1	0	2	3	4
3	0	-1	5	4	3
4	0	1	3	5	7

4.1. По данным в таблице результатам эксперимента, каждый из четырех опытов которого был повторен три раза, получить линейное уравнение регрессии и проверить его адекватность при уровне значимости $\alpha = 0,05$.

4.2. Исследуется зависимость продуктивности резания древесины цепными моторными пилами y, см^2/с, от диаметра резца X_1, см, и рабочей длины пилящего аппарата X_2, см. В качестве базового уровня приняты значения $X_1 = 40$ см, $X_2 = 55$ см; интервалы варьирования факторов $\Delta X_1 = 10$ см, $\Delta X_2 = 20$ см. По данным в таблице результатам ПФЭ типа 2^2 в кодированных переменных, каждый опыт которого повторен два раза, получить линейное уравнение регрессии в кодированных переменных и проверить его адекватность при уровне значимости $\alpha = 0,05$; записать уравнение регрессии в натуральных переменных.

№ оп.	x_{1i}	x_{2i}	y_{i1}	y_{i2}
1	-1	-1	22	24
2	1	-1	44	40
3	-1	1	32	36
4	1	1	44	38

4.3. По данным в таблице результатам ПФЭ типа 2^2, каждый опыт которого был повторен два раза, получить линейное уравнение регрессии в кодированных переменных и проверить его адекватность при уровне значимости $\alpha = 0,05$; записать уравнение регрессии в натуральных переменных.

№ оп.	X_{1i}	X_{2i}	y_{i1}	y_{i2}
1	20	10	$4,5$	$5,5$
2	20	30	$5,2$	$4,8$
3	40	10	8	9
4	40	30	10	9

4.4. По данным в таблице результатам ПФЭ типа 2^3 получить в кодированных переменных уравнение регрессии с парными взаимодействиями и проверить его адекватность при уровне значимости $0,05$, считая, что каждое значение $\overline{y_i}$ есть среднее из 3 параллельных опытов, рандомизированных во времени, и дисперсия воспроизводимости равна $s^2\{y\} = 1,8$; записать уравнение регрессии в натуральных переменных.

№ оп.	X_{1i}	X_{2i}	X_{3i}	$\overline{y_i}$
1	40	20	0	2
2	60	20	0	3
3	40	80	0	4
4	60	80	0	5
5	40	20	10	6
6	60	20	10	1
7	40	80	10	8
8	60	80	10	7

4.5. По данным в таблице результатам эксперимента получить уравнение регрессии с парным взаимодействием и проверить его адекватность при уровне значимости $0,05$, считая, что дисперсия воспроизводимости получена по отдельной серии из 3 опытов и равна $s^2\{y\} = 1,8$.

№ оп.	X_{1i}	X_{2i}	y_i
1	-1	-1	9
2	0	-1	$7,5$
3	2	0	5
4	0	1	$9,5$
5	1	1	13

4.6. По данным в таблице результатам эксперимента получить уравнение регрессии $\widehat{y} = b_0 + b_1 X_1 + b_2 X_2 + b_{11} X_1^2$ и проверить его адек-

№ оп.	X_{1i}	X_{2i}	y_i
1	−1	−1	10
2	0	−1	11
3	0	0	10
4	0	1	15
5	1	1	8

ватность при уровне значимости $0,05$, считая, что дисперсия воспроизводимости получена по отдельной серии из 3 опытов и равна $s^2\{y\} = 1,5$.

4.7. По данным в таблице результатам эксперимента получить линейное уравнение регрессии и проверить его адекватность при уровне значимости $\alpha = 0,05$, считая, что каждое значение $\overline{y_i}$ есть среднее из 3 параллельных наблюдений и дисперсия воспроизводимости равна $s^2\{y\} = 5$.

№ оп.	X_{1i}	X_{2i}	X_{3i}	$\overline{y_i}$
1	−1	−1	0	5
2	0	−1	1	10
3	2	0	0	7
4	0	1	−1	−2
5	1	1	0	−1

4.8. По данным в таблице результатам эксперимента получить уравнение регрессии

$$\widehat{y} = b_0 + b_{12}X_1X_2 + b_{33}X_3^2$$

и проверить его адекватность при уровне значимости $\alpha = 0,05$, считая, что каждое значение $\overline{y_i}$ есть среднее из двух параллельных опытов и дисперсия воспроизводимости равна $s^2\{y\} = 0,8$.

№ оп.	X_{1i}	X_{2i}	X_{3i}	$\overline{y_i}$
1	−1	−1	0	3
2	0	−1	1	2
3	2	0	0	6
4	0	1	−1	0
5	1	1	0	1

4.9. По данным в таблице результатам эксперимента получить уравнение регрессии

$$\widehat{y} = b_0 + b_{12}X_1X_2 + b_{33}X_3^2$$

№ оп.	X_{1i}	X_{2i}	X_{3i}	y_i
1	−1	−1	0	2
2	1	−1	−1	3
3	−1	0	−1	6
4	1	0	1	4
5	−1	1	1	5
6	1	1	0	8

и проверить его адекватность при уровне значимости $0,05$, если для оценки дисперсии воспроизводимости проведена отдельная серия из 5 параллельных опытов при $X_1 = X_2 = X_3 = 0$, в результате которых получены следующие значения отклика: $y_{01} = 3$, $y_{02} = 5,2$, $y_{03} = 3,2$, $y_{04} = 3,8$, $y_{05} = 4,8$.

4.10. По данным в таблице результатам эксперимента получить линейное уравнение регрессии и проверить его адекватность при уровне значимости $\alpha = 0,05$, если для оцен-

№ оп.	X_{1i}	X_{2i}	y_i
1	−1	−1	3
2	1	−1	6
3	−1	0	6
4	1	0	4
5	−1	1	6
6	1	1	11

ки дисперсии воспроизводимости проведена отдельная серия из 5 параллельных опытов при $X_1 = X_2 = 0$, в результате которых получены следующие значения отклика: $y_{01} = 6,2$, $y_{02} = 4,2$, $y_{03} = 4$, $y_{04} = 5,8$, $y_{05} = 4,8$.

Задание 5

В задачах **5.1–5.10** требуется предложить наиболее экономный по количеству опытов план эксперимента (ПФЭ типа 2^k, ДФЭ типа 2^{k-p}, композиционные планы 2-го порядка), позволяющий получить указанное уравнение регрессии; составить матрицу плана в кодированных переменных; указать условия проведения опытов.

5.1. Исследуется зависимость прочности бетона y, МПа, от расхода цемента на 1 м^3 (X_1, кг/м^3), количества добавки суперпластификатора (X_2, %) и количества добавки ускорителя твердения (X_3, %). Предполагается получить уравнение с парными взаимодействиями. Базовые уровни и интервалы варьирования факторов даны в таблице.

Факторы	X_1	X_2	X_3
X_j^0	200	$0,15$	$1,5$
ΔX_j	50	$0,05$	$0,5$

5.2. Изучается влияние на производительность пиления древесины ели y, см^2/с, трех факторов: диаметра реза X_1, 30–53 см; рабочей длины пильного аппарата X_2, 48–66 см; скорости резания X_3, 1,5–15,5 см/с. Предполагается получить уравнение с парными взаимодействиями.

5.3. Исследуется зависимость выхода некоторого продукта y от соотношения NaOH : исходный продукт 1 (фактор X_1), соотношения исходный продукт 1 : исходный продукт 2 (фактор X_2), температуры реакции (фактор X_3, °C) и времени реакции (фактор X_4, ч). Предполагается получить линейное уравнение регрессии в области изменения факторов $1 \leqslant X_1 \leqslant 1,5$; $1 \leqslant X_2 \leqslant 1,5$; $20 \leqslant X_3 \leqslant 30$; $10 \leqslant X_4 \leqslant 20$.

5.4. При исследовании процесса хлорирования титансодержащего концентрата изучается зависимость выхода в расплав хлористого железа от следующих независимых факторов: X_1 — концентрация руды в расплаве, %; X_2 — температура, °C; X_3 — концентрация KCl в расплаве, весовые проценты; X_4 — концентрация углерода в расплаве, весовые проценты. Предполагается получить уравнение регрессии в виде

$$\widehat{y} = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4 + b_{13}x_1x_3 + b_{23}x_2x_3 + b_{34}x_3x_4,$$

изменяя факторы в следующей области: $6,5 \leqslant X_1 \leqslant 8,5$; $700 \leqslant X_2 \leqslant 750$; $55 \leqslant X_3 \leqslant 75$; $3 \leqslant X_4 \leqslant 5$.

Факторы	X_1	X_2	X_3	X_4	X_5
X_j^0	4	4	4	8	80
ΔX_j	2	2	2	2	20

5.5. На стадии разработки лабораторного регламента исследуется стадия твердофазной экстракции процесса извлечения биологически активного препарата феллавина из листьев бархата амурского. (Лекарственный препарат феллавин рекомендуется применять в медицинской практике в качестве противовирусного и антигепатоксического средства.) Предполагается получить линейное уравнение зависимости выхода феллавина (в процентах от его содержания в сырье) от следующих факторов: X_1, X_2, X_3 — продолжительность первой, второй и третьей экстракций, ч; X_4 — соотношение растворитель : сырье; X_5 — температура экстракции , °C. Базовые уровни и интервалы варьирования каждого фактора даны в таблице.

Факторы	X_1	X_2	X_3	X_4	X_5
X_j^0	2	$0,65$	$0,1$	$0,25$	$1,2$
ΔX_j	$0,2$	$0,15$	$0,025$	$0,05$	$0,2$

5.6. Изучается реакция, протекающая по схеме $A + B + C \to D$ в водно-спиртовом растворе. На качество и количество продукта D влияют следующие факторы: X_1 — время реакции, ч; X_2 — содержание спирта в водно-спиртовом растворе, мол. доли; X_3 — концентрация вещества C, мол. доли; X_4 — концентрация вещества B, мол. доли; X_5 — молярное соотношение веществ B и A. Предполагается получить линейное уравнение регрессии, проведя ДФЭ. Базовые уровни и интервалы варьирования факторов даны в таблице.

5.7. Исследуется прочность сплава на основе железа в зависимости от состава сплава. В качестве независимых переменных рассматриваются 7 факторов — процентное содержание хрома (X_1), никеля (X_2), молибдена (X_3), ванадия (X_4), ниобия (X_5), марганца (X_6), углерода (X_7). Предполагается получить линейное уравнение регрессии, проведя ДФЭ. Базовые уровни и интервалы варьирования факторов даны в таблице.

Факторы	X_1	X_2	X_3	X_4	X_5	X_6	X_7
X_j^0	4	2	$0,1$	$0,02$	$0,1$	$0,4$	$0,4$
ΔX_j	1	1	$0,1$	$0,02$	$0,1$	$0,1$	$0,1$

5.8. Исследуется влияние основных технологических факторов процесса прессования фанеры на ее прочность. Предполагается получить линейное уравнение зависимости параметра y (предела прочности на скалывание по клеевому слою, МПа) от следующих факторов: X_1 — вязкость смолы, 50–200°Э; X_2 — давление прессования, 1,6–2,2 МПа; X_3 — температура прессования, 130–150°C; X_4 — расход смолы, 110–150 г/м^2; X_5 — время прессования, 11,5–14,5 минут; X_6 — коэффициент качества шпона, 0,95–0,99.

5.9. Качество беления ткани зависит от стабильности концентрации y раствора в пропиточной ванне. Исследуется зависимость y от количества питающего раствора X_1, л/мин, подаваемого в ванну в единицу времени и массы ткани в единице ее длины X_2, г/м. Предполагается получить квадратичное уравнение регрессии. Базовые уровни и интервалы варьирования факторов даны в таблице.

Факторы	X_1	X_2
X_j^0	$1,12$	150
ΔX_j	$0,22$	30

5.10. При полиэфиризации жирных кислот гликолем представляет интерес влияние концентрации катализатора X_1, $4–16 \cdot 10^{-4}$ граммолекул/100 г, и температуры X_2, $175-225$°C, на процент конверсии. Предполагается получить квадратичное уравнение регрессии.

ПРИЛОЖЕНИЕ 1

Таблица значений $\Phi(x) = \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_0^x \mathrm{e}^{-\frac{t^2}{2}}\, dt$

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
0	0	0,4	0,1554	0,8	0,2881	1,2	0,3849	1,6	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
0,1	0,0398	0,5	0,1915	0,9	0,3159	1,3	0,4032	1,7	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
0,2	0,0793	0,6	0,2257	1	0,3413	1,4	0,4192	1,8	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
0,3	0,1179	0,7	0,2580	1,1	0,3643	1,5	0,4332	1,9	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

Окончание приложения 1

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
2	0,4772	2,3	0,4893	2,6	0,4953	2,9	0,4981	3,2	0,499313
2,01	0,4778	2,31	0,4896	2,61	0,4955	2,91	0,4982	3,21	0,499336
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
2,1	0,4821	2,4	0,4918	2,7	0,4965	3	0,4987	3,3	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
2,2	0,4861	2,5	0,4938	2,8	0,4974	3,1	0,4990	3,4	0,499663
2,21	0,4864	2,51	0,4940	2,81	0,4975	3,11	0,4991	3,5	0,499767
2,22	0,4868	2,52	0,4941	2,82	0,4976	3,12	0,4991	3,6	0,499841
2,23	0,4871	2,53	0,4943	2,83	0,4977	3,13	0,4991	3,7	0,499892
2,24	0,4875	2,54	0,4945	2,84	0,4977	3,14	0,4992	3,8	0,499928
2,25	0,4878	2,55	0,4946	2,85	0,4978	3,15	0,4992	3,9	0,499952
2,26	0,4881	2,56	0,4948	2,86	0,4979	3,16	0,4992	4	0,499968
2,27	0,4884	2,57	0,4949	2,87	0,4979	3,17	0,4992	4,5	0,499997
2,28	0,4887	2,58	0,4951	2,88	0,4980	3,18	0,4993	5	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: $\mathbf{P}\left(\chi^2_\nu > \chi^2_{\alpha;\nu}\right) = \alpha$)

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: $\mathbf{P}\left(|t_\nu| > t_{\alpha;\nu}\right) = \alpha$)

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: $\mathbf{P}\left(F_{\nu_1,\nu_2} > F_{\alpha;\nu_1,\nu_2}\right) = \alpha$)

$\alpha = 0,05$												
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 Кохрена

(значения $G_{\alpha;\nu,N}$ в зависимости от числа степеней свободы ν и N при заданной вероятности $\alpha = 0,05$: $\mathbf{P}\left(G_{\nu,N} > G_{0,05;\nu,N}\right) = 0,05$)

N	ν								
	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

ЛИТЕРАТУРА

Основная

1. Ахназарова, С. Л. Методы оптимизации эксперимента в химической технологии / С. Л. Ахназарова, В. В. Кафаров. — М.: Высшая школа, 1985.

2. Красовский, Г. И. Планирование эксперимента / Г. И. Красовский, Г. Ф. Филаретов. — Минск: БГУ, 1982.

3. Пижурин, А. А. Исследование процессов деревообработки / А. А. Пижурин, М. С. Розенблит. — М.: Лесная промышленность, 1984.

Дополнительная

4. Адлер, Ю. П. Планирование эксперимента при поиске оптимальных условий / Ю. П. Адлер, Е. В. Маркова, Ю. В. Грановский. — М.: Наука, 1976.

5. Блинова, Е. И. Планирование и организация эксперимента: практикум по выполнению расчетных заданий с использованием ЭВМ для студентов II курса специальности 1-54 01 03 / Е. И. Блинова. — Минск: БГТУ, 2004.

6. Герасимович, А. И. Математическая статистика / А. И. Герасимович. — Минск: Вышэйшая школа, 1983.

7. Грачев, Ю. П. Математические методы планирования экспериментов / Ю. П. Грачев, Ю. М. Плаксин. — М.: ДеЛипринт-Москва, 2005.

8. Джонсон, Н. Статистика и планирование эксперимента в технике и науке. Методы обработки данных / Н. Джонсон, Ф. Лион. — М.: Мир, 1980.

9. Жарский, И. М. Планирование и организация эксперимента: учеб. пособие / И. М. Жарский, Б. А. Каледин, И. Ф. Кузьмицкий. — Минск: БГТУ, 2003.

10. Кремер, Н. Ш. Теория вероятностей и математическая статистика: учеб. для вузов / Н. Ш. Кремер. — М.: ЮНИТИ-ДАНА, 2001.

11. Марченко, В. М. Методы оптимизации и статистической обработки результатов измерений: учеб. пособие для студентов физ.-хим. и инж.-техн. специальностей / В. М. Марченко, Т. Б. Копейкина. — Минск: БГТУ, 2007.

12. Налимов, В. В. Статистические методы планирования экстремальных экспериментов / В. В. Налимов, Н. А. Чернова. — М.: Наука, 1965.

ОГЛАВЛЕНИЕ

Учебное издание

Блинова Елена Ивановна

ПЛАНИРОВАНИЕ И ОРГАНИЗАЦИЯ ЭКСПЕРИМЕНТА

Учебно-методическое пособие

Редактор *Р. М. Рябая*
Компьютерный набор и верстка *Е. И. Блинова*

Подписано в печать 16.07.2010. Формат $60 \times 84 \frac{1}{16}$.
Бумага офсетная. Гарнитура Адонис. Печать офсетная.
Усл. печ. л. 7,6. Уч.-изд. л. 7,4.
Тираж 80 экз. Заказ .

Отпечатано в Центре издательско-полиграфических
и информационных технологий учреждения образования
«Белорусский государственный технологический университет».
220006. Минск, Свердлова, 13а.
ЛИ № 02330/0549423 от 08.04.2009.
ЛП № 02330/0150477 от 16.01.2009.