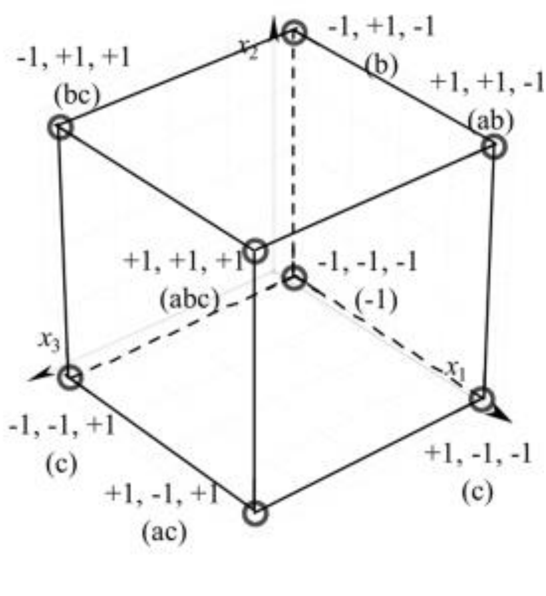
И. В. Ремизова

ПЛАНИРОВАНИЕ ЭКСПЕРИМЕНТА
при разработке систем управления

Практикум

Часть 1



Санкт-Петербург
2020

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ
УНИВЕРСИТЕТ ПРОМЫШЛЕННЫХ ТЕХНОЛОГИЙ И ДИЗАЙНА»

ВЫСШАЯ ШКОЛА ТЕХНОЛОГИИ И ЭНЕРГЕТИКИ

И. В. Ремизова

ПЛАНИРОВАНИЕ ЭКСПЕРИМЕНТА
при разработке систем управления

Практикум

Часть 1

Санкт-Петербург
2020

УДК 62-5 (075)
ББК 32.973я7
Р 380

Ремизова И.В. Планирование эксперимента при разработке систем управления: практикум /ВШТЭ СПбГУПТД. – СПб., 2020. Часть 1. – 52 с.

Практикум содержит необходимые теоретические сведения и задания для проведения практических работ, необходимые для успешного освоения дисциплины «Планирование эксперимента при разработке АСУ».

Предназначен для магистрантов, обучающихся по направлению 15.04.04 «Автоматизация технологических процессов и производств» всех форм обучения. Может быть использован при самостоятельном изучении курсов и подготовке к занятиям обучающимися других направлений.

Рецензент:

доцент кафедры ИИТСУ ВШТЭ СПбГУПТД, канд. техн. наук Е.П. Дятлова

Подготовлен и рекомендован к печати кафедрой информационно-измерительных технологий и систем управления ВШТЭ СПбГУПТД (протокол № 9 от 31.08.2020).

Утвержден к изданию методической комиссией института энергетики и автоматизации ВШТЭ СПбГУПТД (протокол № 2 от 29.10.2020).

Рекомендован к изданию Редакционно-издательским советом ВШТЭ СПбГУПТД в качестве практикума.

© Высшая школа технологии
и энергетики СПбГУПТД, 2020
© Ремизова И.В., 2020

ВВЕДЕНИЕ

Основы курса планирование эксперимента могут быть использованы для решения разнообразных прикладных задач. К таким задачам относятся: расчет оптимального технологического режима с использованием математической модели процесса, построение интерполяционных формул, уточнение параметров теоретических моделей и получение экспериментальных моделей систем автоматического управления и т.д. Основным результатом проведения экспериментов является получение математических моделей объекта, описывающих взаимосвязь между входными и выходными параметрами объекта.

В результате выполнения лабораторных и практических заданий обучающийся получит знания о принципах планирования эксперимента, лежащих в основе организации фундаментальных и прикладных научных исследований. Для достижения этой цели решаются задачи по рассмотрению экспериментальных исследований как базы для выполнения наукоемких разработок, рассматриваются вопросы классификации экспериментов, изучаются возможные способы моделирования систем.

Правильная организация эксперимента при проведении научно-исследовательских работ позволит получить математические модели изучаемых технологических процессов и на их основе осуществить оптимизацию соответствующих конструктивных и режимных параметров.

Основная задача пособия - научить умению использовать теоретические положения и современные методы планирования и обработки эксперимента при проведении научных исследований.

По результатам выполнения работ оформляется отчет, содержащий титульный лист, выполненную работу, включая формулы, расчеты, графики, выводы по работе и ответы на контрольные вопросы по указанию преподавателя. Расчеты выполняются в пакете *Microsoft Excel*. Номер варианта определяется преподавателем.

1. ПОЛУЧЕНИЕ ЛИНЕЙНОЙ РЕГРЕССИОННОЙ МОДЕЛИ

Теория планирования экспериментов позволяет решать задачи на основании полученной в результате эксперимента информации об изучаемом технологическом процессе. На основании априорной (доопытной) информации о рабочем процессе составляется план проведения экспериментов. После каждого проведенного этапа экспериментальных исследований план корректируется, тем самым появляется возможность повышения управляемости рассматриваемого технологического процесса с появлением новых, полученных в результате эксперимента сведений о нем.

При недостаточности сведений об исследуемом объекте управления, он рассматривается как «черный ящик», представленный на рис.1.1.

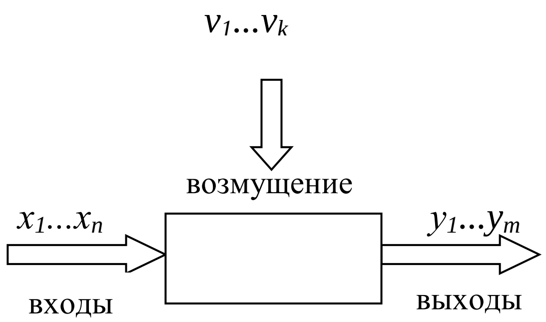


Рис. 1.1. Объект управления как «черный ящик»:
$x_1 \ldots x_n$ – управляемые независимые параметры (факторы); $y_1 \ldots y_m$ – функции выхода (отклика) представляют собой реакцию системы на воздействие факторов $x_1 \ldots x_n$; $v_1 \ldots v_k$ – случайные возмущающие воздействия

«Черный ящик» представляет собой некоторый элемент сложной системы, о которой пользователю известны лишь входные и выходные параметры, а также функциональное назначение. С точки зрения решения задач автоматизации, внутреннее устройство «черного ящика» неизвестно и не имеет никакого значения.

Переменные $x_1 \ldots x_n$ принято называть *факторами*. *Фактор* – это некоторая переменная величина, принимающая в каждый момент времени определенное значение из своей области определения и отражающая внешнее воздействие на объект или его отклик на это воздействие.

Важнейшей задачей методов обработки полученной в ходе эксперимента информации является задача построения математической модели изучаемого явления, процесса, объекта [1].

С общефилософской точки зрения, *эксперимент* (от лат. experimentium – проба, опыт) – это чувственно-предметная деятельность в науке, в более узком смысле слова – опыт, воспроизведение объекта познания, проверка гипотез и т. д.

В технической литературе эксперимент определяется следующим образом: *эксперимент* – это система операций, воздействий или наблюдений, направленных на получение информации об объекте исследования [2].

Хорошо спланированный эксперимент обеспечивает оптимальную обработку результатов и, следовательно, возможность четких статистических выводов.

Цель планирования эксперимента – нахождение таких условий и правил проведения опытов, при которых удается получить надежную и достоверную информацию об объекте с наименьшей затратой труда, а также представить эту информацию в компактной и удобной форме с количественной оценкой точности [2].

Зависимость между случайными величинами называется *регрессией.* Она понимается как зависимость между математическими ожиданиями этих величин.

Построение регрессионных моделей – это многоступенчатый, итерационный процесс. Первая построенная модель в процессе статистического анализа может оказаться не адекватной данным. Диагностика регрессионных моделей позволяет обнаружить несоответствие модели данным и наметить пути для дальнейшего улучшения построенной модели [3].

Существует два вида уравнения регрессии:

1. Простая «парная» регрессия представляет собой линейную модель, т.е. зависимость переменной y как функции одной независимой переменной х. Модели соответствует кривая регрессии в виде *простой линии*.

В неявном виде парная регрессия – это модель вида:

$$y = f(x).$$

В явном виде[1]:

$$y= b_0+ b_1 \cdot x \quad , \tag{1.1}$$

где x – независимая переменная; y – зависимая переменная; b_0 – свободный член уравнения регрессии; b_1 – коэффициент уравнения регрессии.

2. Множественная (зависящая от множества факторов) регрессия представляет собой модель, показывающую зависимость переменной y как функции нескольких независимых переменных – факторов $x_1, x_2, \dots x_n$.

В неявном виде множественная регрессия – это модель вида:

$$y = f(x_1, x_2, x_3 \dots x_n).$$

В явном виде[1]:

$$y = b_0+b_1x_1+b_2x_2+\dots+b_nx_n \, ,$$

где $x_1, x_2, x_3 \dots x_n$ – независимые переменные; y – зависимая переменная; b_0 – свободный член уравнения регрессии; $b_{1 \dots} b_n$ – коэффициенты уравнения регрессии.

Большинство вероятностно-статистических моделей можно свести к парной регрессии, поэтому данная регрессия получила широкое распространение.

Парная регрессия при построении регрессионных моделей может быть представлена различными видами моделей, например:

- линейная регрессия: $y= b_0+ b_1 \cdot x+\varepsilon$;
- равносторонняя гипербола: $y = b_0 + \frac{b_1}{x} + \varepsilon$;
- степенная регрессия: $y = b_0 \cdot x^{b_1} \cdot \varepsilon$;
- показательная регрессия: $y = b_0+{b_1}^{x} \cdot \varepsilon$;

[1] Для линейной регрессии.

– экспоненциальная регрессия: $y = e^{(b_0+b_1*x)} + \varepsilon$,

где b_1 – коэффициент; b_0 – свободный член уравнения регрессии; ε – ошибка.

Наибольшее распространение получила парная линейная регрессия

$$y = b_0 + b_1 \cdot x + \varepsilon_i ,$$

где b_0 – сдвиг (длина отрезка, отсекаемого на координатной оси прямой y); b_1 – наклон прямой y; ε_i – случайная ошибка переменной y в i-м наблюдении.

Пусть дан ряд входных и выходных данных:

$$x_1 \quad x_2 \quad x_3 \quad \ldots \quad x_N;$$

$$y_1 \quad y_2 \quad y_3 \quad \ldots \quad y_N,$$

где N – число экспериментов.

Каждый эксперимент характеризуется набором из двух переменных x_i и y_i. Расчёт парной линейной регрессии требует проведения ряда подготовительных вычислений, необходимо получить:

1. Средние значения $\overline{x}$ и $\overline{y}$, которые находятся следующим образом:

$$\overline{x} = \frac{\sum_1^N x}{N}; \tag{1.2}$$

$$\overline{y} = \frac{\sum_1^N y}{N},$$

где N – количество случайных величин, число экспериментов.

2. Дисперсии σ_x^2, σ_y^2 случайной величины $\overline{x}$, $\overline{y}$ - мера разброса данной случайной величины, т.е. её отклонения от математического ожидания (от среднего). В силу линейности математического ожидания справедливы формулы:

$$\sigma_x^2 = \overline{x^2} - \overline{x}^2; \tag{1.3}$$

$$\sigma_y^2 = \overline{y^2} - \overline{y}^2,$$

Для получения уравнения парной регрессии необходимо рассчитать свободный член b_0 и коэффициент b_1:

$$b_1 = \frac{\overline{xy} - \overline{x} * \overline{y}}{\sigma_x^2}; \tag{1.4}$$

$$\overline{\mathrm{xy}}=\frac{\sum_1^N x_i * y_i}{N}, i = 1 \dots N;$$

$$b_0 = \overline{y} - b_1\overline{x} .$$

Корреляционная связь показывает меру зависимости между двумя и более случайными величинами. Степень тесноты (силы) парной линейной корреляционной зависимости определяет линейный *коэффициент корреляции* r_{xy}:

$$r_{xy} = \frac{\overline{\mathrm{xy}} - \overline{\mathrm{x}} * \overline{\mathrm{y}}}{\sqrt{\sigma_{\mathrm{x}}^2 * \sigma_{\mathrm{y}}^2}}. \quad (1.5)$$

Качественная характеристика показателя тесноты (силы) корреляционной зависимости может быть дана на основе коэффициента корреляции по шкале Чеддока, представленной в Приложении 1.

Оценить качество подбора регрессионной модели позволяет также квадрат коэффициента корреляции, называемый *коэффициентом детерминации* $R = (r_{xy})^2$.

Данный показатель позволяет определить, насколько уравнение регрессии соответствует реальным данным.

На практике, чем ближе коэффициент детерминации к 1, тем выше качество модели. Если коэффициент детерминации ближе к 0, то это означает низкую значимость модели, когда входная переменная плохо «объясняет» поведение выходной, т.е. линейная зависимость между ними отсутствует, и такая модель будет иметь низкую эффективность.

Оценка стандартного отклонения, или стандартная ошибка уравнения регрессии, или среднеквадратичное отклонение – S_y используется для оценки отклонения (разброса) значений от их средней величины и рассчитывается как корень квадратный из дисперсии[2]

$$S_y = \sqrt{\frac{\sum_{i=1}^N (y_i - \widehat{y}_i)^2}{N-2}}, i=1 \dots N, \quad (1.6)$$

[2] Под дисперсией подразумевается несмещенная оценка дисперсии.

где N – число экспериментов, i – порядковый номер эксперимента, $\widehat{y_i}$ рассчитывается, как $\widehat{y_i} = b_0 + b_1 \cdot x_i$.

Среднеквадратичные ошибки, или стандартные отклонения для коэффициента b_1 и свободного члена b_0 уравнения регрессии находятся следующим образом:

$$S_{b0} = S_y \cdot \sqrt{\frac{\sum_{i=1}^{N} x_i^2}{N * \sum_{i=1}^{N}(x_i - \overline{x})^2}}; \qquad (1.7)$$

$$S_{b1} = \frac{S_y}{\sqrt{\sum_{i=1}^{N}(x_i - \overline{x})^2}},$$

где S_{b0} – среднеквадратичная ошибка оценивания свободного члена b_0, S_{b1} – среднеквадратичная ошибка оценивания коэффициента b_1.

На основе построенной модели может осуществляться прогнозирование зависимой переменной y, что предполагает использование полученной взаимосвязи переменных на период прогноза.

В случае, когда для каждого момента времени определяется только одно значение прогнозируемого показателя, полученный таким образом прогноз называют *точечным*.

Значения фактора x подставляют в модель и получают *точечные оценки, или точечный прогноз* изучаемого показателя y:

$$y = b_0 + b_1 \cdot x \pm \varepsilon.$$

В дополнении к точечному прогнозу определяются границы возможного изменения прогнозируемого показателя – *границы доверительного интервала*, тем самым задается диапазон возможных значений прогнозируемого показателя, т.е. получают *интервальный* прогноз.

Для вычисления *границ доверительных интервалов* и *оценки значимости коэффициентов* уравнения регрессии используют t-критерий Стьюдента.

Критерий Стьюдента – это метод статистической проверки гипотез, основанный на распределении Стьюдента. Для применения данного критерия необходимо, чтобы исходные данные подчинялись нормальному закону распределения.

Значение критерия Стьюдента для каждого коэффициента b_i рассчитывается следующим образом:

$$t_{bi} = \frac{|bi|}{S_{bi}}, \tag{1.8}$$

где t_{bi} – критерий Стьюдента для i-го коэффициента b_i.

С помощью таблицы, представленной в Приложении 2, t-критерия Стьюдента требуется найти критическое значение ($t_{табл}$) критерия для заданного уровня значимости p и числа степеней свободы φ.

Для того чтобы определить, входят ли в модель свободный член b_0 и коэффициент b_1, необходимо сравнить критическое и рассчитанное значения критерия Стьюдента.

Если рассчитанное (t_{bi}) значение t-критерия Стьюдента *равно или больше критического* ($t_{табл}$), найденного по таблице, делается вывод о статистической значимости различий между сравниваемыми величинами. Например,

$$t_{bi} \geq t_{табл}.$$

Если значение рассчитанного (t_{bi}) t-критерия Стьюдента *меньше* критического ($t_{табл}$), то различия сравниваемых величин статистически не значимы. Например,

$$t_{bi} \leq t_{табл}.$$

После определения t-критерия Стьюдента возможен расчет границ доверительных интервалов по каждому из показателей, для чего определяется *предельная ошибка показателя* – Δ. Предельная ошибка показателя показывает, как максимально сильно значение показателя может отклониться от фактического. Она используется для расчета точности прогнозирования, что позволяет оценить, как точно и корректно сформирован прогноз.

Для нахождения предельной ошибки показателя находят произведение среднейквадратичной ошибки на t-критерий Стьюдента для заданной доверительной вероятности. Размеры предельной ошибки зависят от критерия $t_{табл}$, исходя из заданной вероятности:

$$\Delta_{bi} = t_{\text{табл}} \cdot S_{bi} \, , \tag{1.9}$$

где Δ_{bi} – предельная ошибка i-го коэффициента b_i.

Доверительные интервалы являются интервальными оценками и рассчитываются для заданной вероятности или заданного уровня значимости[3]. Они позволяют сделать утверждение, что истинное значение неизвестного статистического параметра генеральной совокупности находится в полученном диапазоне значений с вероятностью, которая задана выбранным уровнем статистической значимости.

Границы доверительных интервалов, или *доверительные границы прогноза* – это границы средних или относительных величин, выход за пределы которых вследствие случайных колебаний имеет незначительную вероятность.

Границы доверительных интервалов, или *доверительные границы прогноза* для коэффициентов регрессии рассчитываются как

$$(\widehat{b_i} - \Delta_{bi}) < b_i < (\widehat{b_i} + \Delta_{bi}) \, . \tag{1.10}$$

Смысл вычисления доверительного интервала заключается в построении по данным выборки такого интервала, чтобы можно было утверждать с заданной вероятностью, что значение оцениваемого параметра находится в этом интервале. Другими словами, доверительный интервал с определенной вероятностью содержит неизвестное значение оцениваемой величины. Чем шире интервал, тем выше неточность.

---

[3] Уровню значимости p соответствует разность между 1 и доверительной вероятностью β. $p=1-\beta$.

Если в границы доверительного интервала попадает ноль, т.е. нижняя граница отрицательна, а верхняя положительна, то оцениваемый параметр принимается нулевым, так как он не может одновременно принимать и положительное, и отрицательное значения.

Практическая работа № 1

Цель работы: ознакомление с основными типами регрессионных моделей, получение линейной регрессионной модели.

Задание:

На основании полученных экспериментальных данных сделать предварительный вывод - является ли регрессия линейной, построив поле корреляции. Получить парное линейное уравнение регрессии.

Вычислить коэффициенты корреляции и детерминации, стандартную ошибку регрессии и среднеквадратичные ошибки коэффициентов уравнения. С помощью критерия Стьюдента оценить значимость коэффициентов уравнения регрессии и величину доверительных интервалов для уровня значимости $p = 0{,}05$.

Построить на одних осях корреляционное поле и линию тренда. Сделать выводы о проделанной работе.

Работа выполняется в *Microsoft Excel*.

Варианты:

I												
N	1	2	3	4	5	6	7	8	9	10	11	12
X	451	163	513	112	123	561	312	123	456	156	126	131
У	145	123	235	145	120	160	100	240	135	123	230	125

II												
N	1	2	3	4	5	6	7	8	9	10	11	12
X	52	110	170	141	150	160	200	230	240	260	270	300
У	100	90	130	31	60	39	58	70	80	150	120	130

III												
N	1	2	3	4	5	6	7	8	9	10	11	12
X	120	150	140	156	216	156	153	324	102	233	200	123
У	354	325	205	385	524	562	655	244	241	145	254	145

IV												
N	1	2	3	4	5	6	7	8	9	10	11	12
X	26	48	65	23	150	123	264	156	152	154	520	415
У	221	153	155	102	156	264	435	156	203	325	456	163

V												
N	1	2	3	4	5	6	7	8	9	10	11	12
X	15	26	65	320	652	156	896	123	16	263	459	213
У	125	163	162	263	563	23	463	126	133	152	213	159

VI												
N	1	2	3	4	5	6	7	8	9	10	11	12
X	651	165	849	156	156	132	456	251	489	265	56	256
У	156	859	156	165	49	566	145	156	566	456	56	156

VII												
N	1	2	3	4	5	6	7	8	9	10	11	12
X	156	165	415	894	465	163	496	131	566	561	562	354
У	526	414	561	561	456	321	648	654	321	456	464	456

VIII												
N	1	2	3	4	5	6	7	8	9	10	11	12
X	463	135	563	456	131	452	496	131	566	561	562	354
У	526	414	561	561	456	321	156	123	542	563	451	456

IX												
N	1	2	3	4	5	6	7	8	9	10	11	12
X	123	465	789	744	456	452	453	131	566	879	562	789
У	465	123	123	416	435	746	156	325	748	563	464	456

X												
N	1	2	3	4	5	6	7	8	9	10	11	12
X	165	156	498	3,15	496	132	981	135	158	864	456	894
У	465	135	496	159	791	168	489	163	89	213	489	564

Пример выполнения работы:

-												
N	1	2	3	4	5	6	7	8	9	10	11	12
X	98	88	151	29	60	37	41	69	79	151	110	131
У	126	108	170	139	150	155	201	225	241	255	270	300

Поле корреляции – это поле точек, координаты которых (x; y) определяются значениями входных и выходных данных.

Чтобы построить поле корреляции, в *Excel* в закладке «*Вставка*» следует выбрать: «*Рекомендуемые диаграммы*» - «*Точечная*». Диаграмма строится по данным x и y (рис. 1.2):

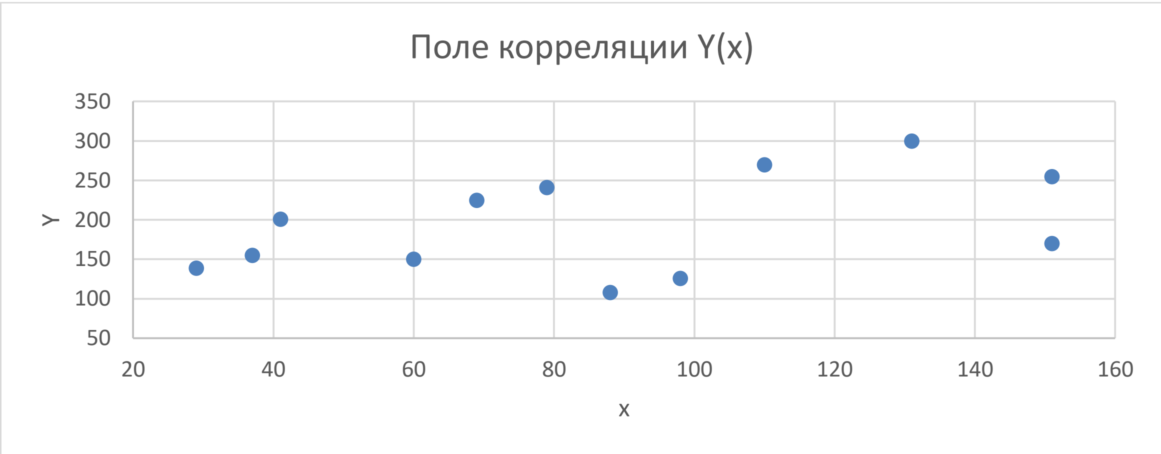


Рис. 1.2. Поле корреляции y от x

По характеру расположения точек на поле корреляции можно сделать вывод о наличии или отсутствии связи, о характере связи, т. е. о виде математической модели.

Линия тренда. Для оценки уравнения регрессии для поля корреляции рис.1.2 строится линия тренда.

Линия тренда строится на осях имеющегося поля корреляции. Для построения нужно: выполнить один щелчок левой кнопкой мыши в любом месте диаграммы и затем нажать иконку с символом *плюс* (+) рядом с диаграммой, чтобы открыть меню «*Элементы диаграммы*» (*Chart elements*).

Другой вариант: нажать кнопку «*Добавить элемент диаграммы*» (*Add Chart Elements*), которая находится в разделе «*Макеты диаграмм*» (*Chart Layouts*) на вкладке «*Конструктор*» (*Design*).

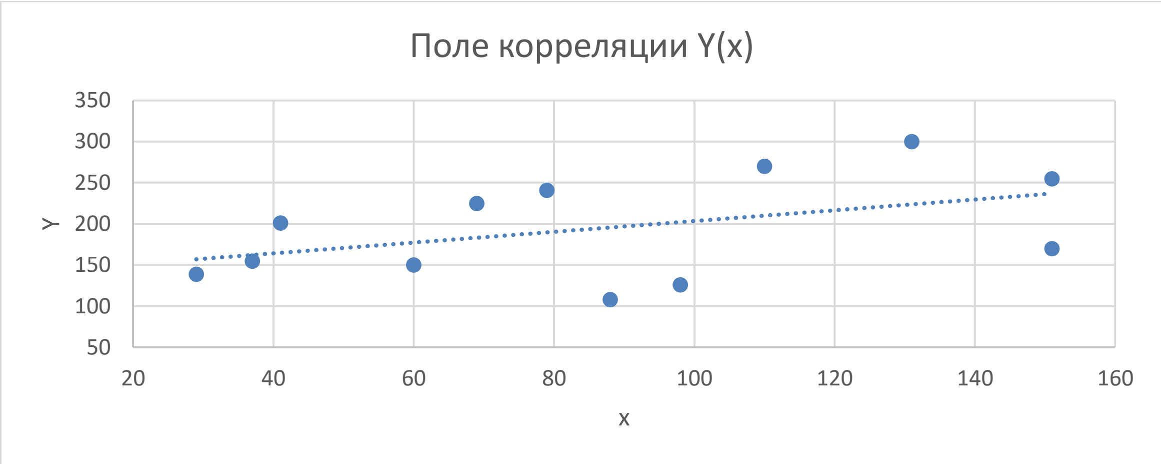


Рис. 1.3. Линия тренда для полученного уравнения регрессии

Таблица 1.1

Промежуточные результаты для получения математической модели

№	*x*	$y_{эксп}$	*(x*y)*	x^2	y^2	$(x-\bar{x})$	$(x-\bar{x})^2$
1	98	126	12348	9604	15876	11	121
2	88	108	9504	7744	11664	1	1
3	151	170	25670	22801	28900	64	4096
4	29	139	4031	841	19321	-58	3364
5	60	150	9000	3600	22500	-27	729
6	37	155	5735	1369	24025	-50	2500
7	41	201	8241	1681	40401	-46	2116
8	69	225	15525	4761	50625	-18	324
9	79	241	19039	6241	58081	-8	64
10	151	255	38505	22801	65025	64	4096
11	110	270	29700	12100	72900	23	529
12	131	300	39300	17161	90000	44	1936
Сумма:				110704			19876
Среднее:	$\bar{x}$	$\bar{y}$	$\overline{xy}$	$\overline{x^2}$	$\overline{y^2}$		
	87	195	18049,83	9225,3	41609,83		

Для расчета параметров уравнения линейной парной регрессии выполняется ряд промежуточных расчетов (табл. 1.1): произведение: *(x*y)*; квадраты чисел: x^2, y^2; разности: $(\mathrm{x}-\overline{\mathrm{x}}), (\mathrm{y}-\overline{\mathrm{y}})$, где $\overline{\mathrm{x}}, \overline{\mathrm{y}}$ находятся по формулам (1.2); квадраты разности: $(\mathrm{x}-\overline{\mathrm{x}})^2$, $(\mathrm{y}-\overline{\mathrm{y}})^2$; суммы: $\sum \mathrm{x}$, $\sum \mathrm{y}$, $\sum(\mathrm{x}*\mathrm{y})$, $\sum \mathrm{x}^2$, $\sum \mathrm{y}^2$, $\sum(\mathrm{x}-\overline{\mathrm{x}})$, $\sum(\mathrm{x}-\overline{\mathrm{x}})^2$; средние значения: $\overline{\mathrm{xy}}, \overline{\mathrm{x}^2}, \overline{\mathrm{y}^2}$.

Произведение управляемой независимой переменной *x* на функцию выхода *y*:

$(x*y)_1 = 98*126 = 12348$

...

$(x*y)_{12} = 131*300 = 39300.$

Квадрат независимой переменной *x* и функции выхода *y*:

$x^2_1 = 98^2 = 9604$ $y^2_1 = 126^2 = 15876$

... ...

$x^2_{12} = 131^2 = 17161$ $y^2_{12} = 300^2 = 90000.$

Разность управляемой независимой переменной *x* и ее среднего значения $\overline{\mathrm{x}}$:

$(\mathrm{x}_1-\overline{\mathrm{x}}) = 98-87 = 11$

...

$(\mathrm{x}_{12}-\overline{\mathrm{x}}) = 131-87 = 44.$

Средние значения выражений (в *Excel* это функция *СРЗНАЧ()*):

$$\overline{\mathrm{x}} = \frac{(98+88+\ldots+110+131)}{12} = 87$$

...

$$\overline{\mathrm{y}^2} = \frac{(15876+\ldots+90000)}{12} = 41609{,}83.$$

Для расчета коэффициентов уравнения регрессии находятся дисперсии случайных величин $\overline{\mathrm{x}}, \overline{\mathrm{y}}$, формула (1.3):

$$\sigma_{\mathrm{x}}^2 = \overline{\mathrm{x}^2} - \overline{\mathrm{x}}^2 = 9225{,}3 - (87)^2 = 1656{,}3\ ,$$

$$\sigma_{\mathrm{y}}^2 = \overline{\mathrm{y}^2} - \overline{\mathrm{y}}^2 = 41609{,}83 - (195)^2 = 3584{,}83.$$

Параметры a и b линейного уравнения регрессии $\hat{y} = b_0 + b_1 \cdot x$ рассчитываются по формулам (1.4):

$$b_1 = \frac{\overline{xy} - \bar{x} * \bar{y}}{\sigma_x^2} = \frac{18049{,}83 - 87*195}{1656{,}3} = 0{,}65;$$

$$b_0 = \bar{y} - b_1 \cdot \bar{x} = 195 - 0{,}65 * 87 = 138{,}02.$$

Полученные коэффициенты подставляются в уравнение регрессии согласно (1.1):

$$\hat{y} = 138{,}02 + 0{,}65 \cdot x \ .$$

Значение линейного коэффициента парной корреляции для определения связи рассчитывается по формуле (1.5):

$$r_{xy} = \frac{\overline{xy} - \overline{x} * \overline{y}}{\sqrt{\sigma_x^2 * \sigma_y^2}} = \frac{18049{,}83 - 87 * 195}{\sqrt{1656{,}3 * 3584{,}83}} = 0{,}445 \ .$$

Интервал, в котором находится рассчитанный коэффициент, определяется по шкале Чеддока (см. Приложение 1).

Коэффициент находится в интервале от 0,3 до 0,5, связь умеренная, прямая.

Знная коэффициент корреляции, определяют коэффициент детерминации:

$R = (r_{xy})^2 = (0{,}445)^2 = 0{,}198$.

Соответствие уравнения регрессии реальным данным составляет 19,8 %, что говорит о низкой значимости полученной математической модели.

Для оценки полученного уравнения регрессии сформирована таблица промежуточных результатов, табл.1.2.

Табличные расчеты:

Уравнение регрессии для каждого *x:*

$\hat{y}_1 = 138{,}02 + 0{,}65x = 138{,}02 + 0{,}65*98 = 202{,}2$

...

$\hat{y}_{12} = 138{,}02 + 0{,}65*131 = 223{,}82$.

Вычитание $(y - \hat{y})$:

$(y - \hat{y})_1 = 129 - 202{,}2 = -76{,}2$

...

$(y - \hat{y})_{12} = 300 - 223{,}82 = 76{,}18$.

Возведение в квадрат $(y - \hat{y})^2$:

$(y - \hat{y})^2_1 = (-76{,}2)^2 = 5807{,}14$

...

$(y - \hat{y})^2_{12} = (76{,}18)^2 = 5803{,}66$.

Таблица 1.2

Промежуточные результаты для анализа полученного уравнения регрессии

№	x	$y_{эксп}$	$\hat{y}_{рас}$	$(y - \hat{y})$	$(y - \hat{y})^2$
1	98	129	202,2	-76,2	5807,14
2	88	108	195,65	-87,66	7683,39
3	150	170	236,92	-66,92	4477,95
4	29	139	157,01	-18,01	324,44
5	60	150	177,32	-27,32	746,17
6	37	155	162,25	-7,25	52,59
7	41	202	164,87	36,13	1305,25
8	69	225	183,21	41,79	1746,35
9	79	241	189,76	51,24	2625,51
10	151	255	236,92	18,08	326,98
11	110	270	210,06	59,94	3592,31
12	131	300	223,82	76,18	5803,66
Сумма					34491,72

Далее рассчитывается *стандартная ошибка уравнения регрессии*, формулы (1.6):

$$S_y = \sqrt{\frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{N-2}} = \sqrt{\frac{34491{,}72}{12-2}} = 58{,}73 \ .$$

Для коэффициента b_1 и свободного члена b_0 полученного уравнения регрессии по формулам (1.7) рассчитываются среднеквадратичные ошибки коэффициентов

$$S_{b0} = S_y \cdot \sqrt{\frac{\sum_{i=1}^{N} x_i^2}{N * \sum_{i=1}^{N}(x_i - \overline{x})^2}} = 58{,}73 \cdot \sqrt{\frac{110704}{12 \cdot 19876}} = 40{,}01 \ ,$$

$$S_{b1} = \frac{S_y}{\sqrt{\sum_{i=1}^{N}(x_i - \overline{x})^2}} = \frac{58.73}{\sqrt{19876}} = 0{,}417 \ .$$

Для проверки правильности выполнения своей работы в Microsoft Excel можно воспользоваться функцией *ЛИНЕЙН()*. Для того чтобы применить функцию, необходимо на Листе *Excel* выделить диапазон из 2 столбцов и 5 строк (10 ячеек). После чего в строке формул ввести функцию в следующем формате: ***=ЛИНЕЙН(****диапазон значений у****;****диапазон значений х****;;ИСТИНА)*** После чего следует одновременно нажать «*Ctrl + Shift + Enter*».

Функция рассчитает следующие показатели:

b1 - коэффициент	0,65	138,02	b0 - свободный член
Sb1 – среднеквадратичная ошибка оценивания коэффициента b1	0,417	40,01	Sb0 – среднеквадратичная ошибка оценивания свободного члена b0
R – коэффициент детерминации	0,198	58,73	Sy – среднеквадратичное отклонение y
F-статистика (F-критерий Фишера)	2,47	10	Число степеней свободы
Регрессионная сумма квадратов	8526,28	34491,72	Остаточная сумма квадратов

Первая строчка – коэффициенты уравнения регрессии. Вторая строчка – средние квадратичные ошибки коэффициентов. Если одна из ошибок по модулю больше, чем сам коэффициент, то коэффициент считается нулевым. Третья строка – коэффициент детерминации, характеризует качество связи между факторами. Полученное значение 0,198 говорит о слабой связи факторов. *F*-статистика проверяет гипотезу о адекватности регрессионной модели. Если *F*-статистика больше, чем *F*-критическое, значит, регрессионная модель адекватна. В последней строке приведены регрессионная и остаточная суммы сумма квадратов. Регрессионная сумма, т.е. сумма, подтверждаемая регрессией, должна быть намного больше остаточной, вызванной случайными факторами. В рассматриваемом примере это условие не выполняется, что говорит о слабой регрессионной связи.

Значимость коэффициентов полученного уравнения регрессии оценивается с использованием *t*-критерия Стьюдента.

t – критерий Стьюдента для обоих параметров рассчитывается по формулам (1.8):

$$t_{b0} = \frac{b_0}{s_{b0}} = \frac{138,02}{40,01} = 3,45 \ ,$$

$$t_{b1} = \frac{b_1}{s_{b1}} = \frac{0,65}{0,417} = 1,57 \ .$$

Полученное значение критерия Стьюдента сравнивается с критическим, взятым для уровня значимости $p = 0{,}05$ и числа степеней свободы φ = 11 (φ $=N\text{-}1$) (см. Приложение 2):

$$t_{табл} = 2{,}201 .$$

Так как рассчитанное значение критерия $t_{b0} = 3{,}45$ больше критического, можно сделать вывод, что наблюдаемые различия статистически значимы, коэффициент b_0 надёжен. Значение критерия $t_{b1} = 1{,}57$ меньше табличного, значит различия сравниваемых величин статистически незначимы, коэффициент b_1 ненадёжен при данном уровне значимости.

Для расчета доверительного интервала определяется предельная ошибка показателей Δ, формулы (1.9):

$$\Delta_{b0} = t_{\text{табл}} \cdot S_{b0} = 2{,}201 \cdot 40{,}01 = 88{,}07 ,$$

$$\Delta_{b1} = t_{\text{табл}} \cdot S_{b1} = 2{,}201 \cdot 0{,}417 = 0{,}917 .$$

Доверительные интервалы для коэффициентов регрессии рассчитываются по формулам (1.10):

$$(b_0 - \Delta_{b0}) < b_0 < (b_0 + \Delta_{b0})$$

$$49{,}95 < b_0 < 226{,}08 ;$$

$$(b_1 - \Delta_{b1}) < b_1 < (b_1 + \Delta_{b1})$$

$$-0{,}26 < b_1 < 1{,}57 .$$

Для доверительных интервалов справедливы общие правила: чем уже доверительный интервал, тем точнее оценка параметра; если доверительный интервал включает нулевое значение, то оцениваемый параметр статистически незначим (равен нулю).

С вероятностью 95 % при неограниченно большом числе наблюдений коэффициент регрессии b_0 попадет в диапазон от 49,95 до 226,08, показатель b_1 в диапазон (-0,26; 1,57). Значение коэффициента b_1 в рамках доверительного интервала меняет свое значение с отрицательного на положительное, что говорит о незначимости данного коэффициента.

Таким образом, можно утверждать, с учетом низкой значимости полученной математической модели, большой ошибки уравнения регрессии и ненадёжности коэффициента b_1: полученное уравнение регрессии $\hat{y} = 138{,}02 + 0{,}65 \cdot x$ не может использоваться для моделирования и прогнозирования данных.

2. ПОЛУЧЕНИЕ МАТЕМАТИЧЕСКОЙ МОДЕЛИ ПРИ ПОЛНОМ ФАКТОРНОМ ЭКСПЕРИМЕНТЕ

План эксперимента – это совокупность данных, определяющих число, условия и порядок реализации опытов [4].

Каждому определенному состоянию объекта соответствует конкретное сочетание входных факторов.

При анализе априорной информации и постановке задачи исследования для каждого входного фактора определяется область, в пределах которой он будет изменяться во время эксперимента. Сочетание таких областей по всем входным факторам будет факторным пространством эксперимента.

Построение плана эксперимента предполагает выбор уровней входных факторов. Первоначально выбирается нулевой уровень каждого входного фактора, в качестве которого может быть взята любая точка факторного пространства эксперимента. Далее относительно нулевого рассчитываются нижний и верхний уровни входного фактора. Для этого вводится интервал варьирования каждого входного фактора. *Верхний уровень* входного фактора определяется при сложении нулевого уровня данного фактора и интервала варьирования, а *нижний уровень* – при вычитании интервала варьирования из нулевого уровня фактора.

При планировании *полного факторного эксперимента* (ПФЭ) реализуются все возможные комбинации факторов на всех выбранных уровнях исследования.

Обычно при первичном планировании эксперимента количество уровней по всем входным факторам выбирается одинаковым. Тогда количество опытов при полном факторном эксперименте N определяется по формуле

$$N=n^k ,$$

где n – количество уровней каждого входного фактора; k – число входных факторов в эксперименте.

Если на основе анализа априорной информации становится понятно, что исследуемая зависимость является линейной, то достаточно реализовать эксперимент, в котором каждому входному фактору соответствует два уровня, т.е. $N=2^k$. Если исследуемая зависимость предположительно нелинейна, то реализуется трехуровневый факторный эксперимент [7].

Полный факторный эксперимент страдает избыточностью опытов. Можно сократить число опытов, отбросив несущественную информацию, т.е. исключить из эксперимента незначимые взаимодействия уровней факторов. Дробный факторный эксперимент проводится, когда пропущены некоторые сочетания уровней факторов.

Для повышения точности расчета коэффициентов модели, а также для упрощения обработки данных факторы, выраженные в натуральных единицах, переводят в относительные единицы x, применяя процедуру нормировки. В этом случае соотношение между факторами x в относительных единицах и этими же факторами в натуральных единицах z выводится следующим образом.

Для каждого из входных факторов находятся максимальное z_i^{max} и минимальное z_i^{min} значения технологического параметра.

Определяются координаты центра плана Z_i^0:

$$Z_i^0 = \frac{z_i^{min}+z_i^{max}}{2}, i = 1,2,3, \dots, k. \qquad (2.1)$$

Рассчитывается интервал варьирования Δz_i:

$$\Delta z_i = \frac{Z_i^{max}-Z_i^{min}}{2}, i = 1,2,3, \dots, k. \qquad (2.2)$$

Переход к новой безразмерной системе координат x_1, …, x_k осуществляется с помощью преобразования:

$$x_i = \frac{Z_i-Z_i^0}{\Delta Z_i}, i = 1,2,3, \dots, k. \qquad (2.3)$$

В безразмерной системе координат верхний уровень принимает значение (+1), нижний уровень соответственно (-1), координаты центра равны нулю и совпадают с началом координат.

Условия полного факторного эксперимента записывают в виде таблицы – *матрицы планирования* эксперимента.

Геометрическая интерпретация полного факторного эксперимента типа 2^2 представлена на рис.2.1.; матрица планирования в табл.2.1.

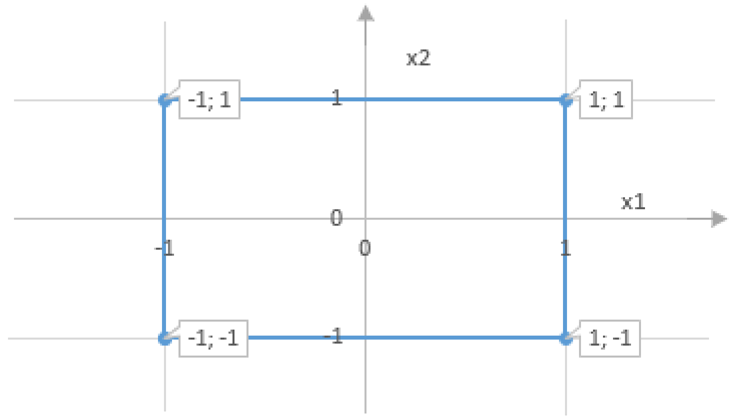


Рис. 2.1. Геометрическая интерпретация полного факторного эксперимента типа 2^2

Таблица 2.1

Матрица планирования эксперимента

Номер опыта	Входные факторы		Функция отклика
	x_1	x_2	$Y_{эксп}$
1	–1	–1	Y_1
2	–1	+1	Y_2
3	+1	–1	Y_3
4	+1	+1	Y_4

Пусть требуется провести полный факторный двухуровневый эксперимент для трёх факторов.

План полного факторного эксперимента типа 2^3 представлен в натуральных и относительных единицах в табл.2.2.

Для нахождения коэффициентов уравнения необходимо составить расширенную матрицу планирования ПФЭ.

Расширенная матрица планирования ПФЭ, дополнительно к входным и выходным факторам, включает фиктивную переменную x_0=1 (соответствующую постоянной составляющей уравнения регрессии b_0) и взаимодействия факторов между собой x_1x_2, x_1x_3 , x_2x_3 , $x_1x_2x_3$ (табл.2.3.). Взаимодействия факторов получаются перемножением нормированных значений соответствующих факторов [9].

Математическая модель объекта (уравнение регрессии) запишется в следующем виде:

y=$b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_{12}x_1x_2 + b_{13}x_1x_3 + b_{23}x_2x_3 + b_{123}x_1x_2x_3$,(2.4)

где b_{12}, b_{13}, b_{23} – коэффициенты парного взаимодействия, b_{123} – коэффициент тройного взаимодействия.

Таблица 2.2

План полного факторного эксперимента типа *2*^{3} в натуральных и относительных единицах

№ п/п	Входные факторы						Выходной фактор
	в натуральных единицах			в относительных единицах			
	z_1	z_2	z_3	x_1	x_2	x_3	$y_э$
1	z_1^{min}	z_2^{min}	z_3^{min}	–1	–1	–1	y_1
2	z_1^{max}	z_2^{min}	z_3^{min}	+1	–1	–1	y_2
3	z_1^{min}	z_2^{max}	z_3^{min}	–1	+1	–1	y_3
4	z_1^{max}	z_2^{max}	z_3^{min}	+1	+1	–1	y_4
5	z_1^{min}	z_2^{min}	z_3^{max}	–1	–1	+1	y_5
6	z_1^{max}	z_2^{min}	z_3^{max}	+1	–1	+1	y_6
7	z_1^{min}	z_2^{max}	z_3^{max}	–1	+1	+1	y_7
8	z_1^{max}	z_2^{max}	z_3^{max}	+1	+1	+1	y_8

Коэффициенты при независимых переменных указывают на силу влияния факторов. Чем больше численная величина коэффициента, тем большее влияние оказывает фактор. Если коэффициент имеет знак плюс, то с увеличением значения фактора параметр оптимизации увеличивается, а если минус, то уменьшается [7]. Полный факторный эксперимент также позволяет количественно оценивать эффекты взаимодействия.

Таблица 2.3

Расширенная матрица планирования полного факторного эксперимента 2^3

№ п/п	x_0	x_1	x_2	x_3	x_1x_2	x_1x_3	x_2x_3	$x_1x_2x_3$	$у_э$
1	+1	–1	–1	–1	+1	+1	+1	–1	$у_1$
2	+1	+1	–1	–1	–1	–1	+1	+1	$у_2$
3	+1	–1	+1	–1	–1	+1	–1	+1	$у_3$
4	+1	+1	+1	–1	+1	–1	–1	–1	$у_4$
5	+1	–1	–1	+1	+1	–1	–1	+1	$у_5$
6	+1	+1	–1	+1	–1	+1	–1	–1	$у_6$
7	+1	–1	+1	+1	–1	–1	+1	–1	$у_7$
8	+1	+1	+1	+1	+1	+1	+1	+1	$у_8$

Матрица полного факторного эксперимента[4] обладает следующими свойствами:

– симметричность относительно центра эксперимента – сумма элементов каждого столбца равна нулю;

– свойство нормировки – сумма квадратов элементов каждого столбца равна числу опытов:

$$\sum_{i=1}^{N} x_{ji}^2 = N,$$

где $i = 1 \dots N$- число опытов; $j = 1 \dots k$ –входные факторы.

– ортогональность – сумма почленных произведений любых двух столбцов матрицы равна нулю:

$$\sum_{i=1}^{N} x_{ji}x_{vi} = 0, j \neq v,$$

где j, v – номера входных факторов.

– ротатабельность, точки в матрице планирования подбираются так, что точность предсказания значений параметра оптимизации одинакова на равных расстояниях от центра эксперимента и не зависит от направления [7].

Коэффициенты уравнения регрессии вычисляются по формуле

[4] Рассматриваются свойства матрицы планирования без учета фиктивного столбца.

$$b_j = \frac{1}{N}\sum_{i=1}^{N} x_{ji} y_i, \qquad (2.5)$$

обоснование данной формулы приведено в практической работе №3.

После расчета коэффициентов уравнения регрессии требуется определить, входит ли в математическую модель каждый из коэффициентов b_i, т.е. необходимо сравнить критическое и рассчитанное значения критерия Стьюдента для каждого из коэффициентов (см. практическую работу 1).

После следует проверить адекватность полученной модели. Для проверки гипотезы адекватности необходимо сравнить две дисперсии:

– дисперсию адекватности, характеризующую адекватность модели, представляющую сумму квадратов разностей между рассчитанными по модели $y_{\text{расч}}$ и экспериментальными $y_{\text{эксп}}$ значениями выходной величины:

$$\mathrm{S}_{\text{адекв}} = \sum_{i=1}^{N} v\left(y_{\text{расч}}^{i} - y_{\text{эксп}}^{i}\right)^2, \qquad (2.6)$$

где v – число параллельных опытов. Число степеней свободы дисперсии адекватности определяется выражением

$$\varphi_1 = N - (k+1),$$

где N – число экспериментов; k – количество коэффициентов модели, не считая нулевого (свободного члена);

– дисперсию воспроизводимости, представляющую сумму квадратов, характеризующую ошибку эксперимента:

$$S_e = \sum_{\boldsymbol{i=1}}^{N} \sum_{\mathrm{j}=1}^{v}\left(y_{\text{э}}^{ij} - y_{\text{э}}^{i}\right)^2, \qquad (2.7)$$

где $y_{\text{э}}^{i}$ –среднее значение наблюдений по параллельным опытам в i-м эксперименте. Число степеней свободы дисперсии воспроизводимости

$$\varphi_2 = N \cdot v - N = N(v-1).$$

Задавшись уровнем значимости проверки гипотезы об адекватности, рассчитывают значение показателя $F = \frac{S_{D:\varphi_1}}{S_e:\varphi_2}$, подчиняющегося распределению Фишера для заданных чисел степеней свободы φ_1 и φ_2.

Расчетное значение сравнивается с критическим, полученным из таблицы распределения Фишера (Приложение 3).

Если рассчитанное значение F меньше критического $F_{расч} < F_{крит}$, т.е. находится в зоне доверительного интервала плотности распределения величины F, то гипотеза об адекватности принимается. В противном случае гипотеза об адекватности модели отклоняется.

Практическая работа № 2

Цель работы: изучение метода полного факторного эксперимента при построении математической модели процесса.

Задание:

Согласно заданным диапазонам изменения входных факторов и результатам эксперимента, полученным в виде серии параллельных опытов, требуется получить математическую модель процесса в виде уравнения регрессии. При построении математической модели учесть взаимодействия факторов между собой.

При выполнении работы составить расширенную матрицу планирования ПФЭ с фиктивной переменной.

Получить коэффициенты уравнения регрессии.

Проверить значимость коэффициентов уравнения регрессии и адекватность полученной математической модели для уровня значимости 0,05.

Сформировать уравнение регрессии для выходного параметра в натуральных единицах.

Работа выполняется в *Microsoft Excel*.

Варианты:

Вариант	Нечетный			Четный		
Уровни факторов	входные факторы			входные факторы		
	X_1	X_2	X_3	X_1	X_2	X_3
Нижний	4	10	0,8	4	20	0,2
Верхний	12	39	0,9	14	92	0,6

Варианты: выходной фактор – результаты наблюдений									
I	II	III	IV	V	VI	VII	VIII	IX	X
Y_1 Y_2 Y_3	Y_2 Y_3 Y_4	Y_3 Y_4 Y_5	Y_4 Y_5 Y_6	Y_5 Y_6 Y_7	Y_6 Y_7 Y_8	Y_7 Y_8 Y_9	Y_8 Y_9 Y_{10}	Y_1 Y_3 Y_5	Y_2 Y_4 Y_6

Результаты параллельных опытов для матрицы планирования соотв.табл.2.2									
Y_1	Y_2	Y_3	Y_4	Y_5	Y_6	Y_7	Y_8	Y_9	Y_{10}
0,12	0,12	0,11	0,12	0,12	0,12	0,11	0,12	0,12	1,11
0,06	0,06	0,06	0,08	0,06	0,07	0,06	0,08	0,06	0,06
0,20	0,18	0,20	0,20	0,20	0,18	0,21	0,20	0,20	0,18
0,18	0,16	0,18	0,16	0,18	0,11	0,18	0,12	0,15	0,16
0,14	0,12	0,14	0,16	0,13	0,12	0,14	0,16	0,14	0,12
0,11	0,12	0,10	0,10	0,11	0,12	0,11	0,10	0,11	0,12
0,20	0,23	0,16	0,21	0,16	0,18	0,22	0,21	0,17	0,23
0,20	0,22	0,20	0,18	0,20	0,21	0,20	0,18	0,25	0,22

Пример выполнения работы

Исходные данные:

Уровни факторов	Входные факторы		
	X_1	X_2	X_3
Нижний	5	13	48
Верхний	18	25	64

Дан диапазон изменения входных факторов и результаты наблюдений параллельных опытов.

Y1	Y2	Y3	Y4
0,12	0,11	0,10	0,11
0,06	0,07	0,08	0,05
0,20	0,19	0,20	0,25
0,18	0,21	0,17	0,15
0,12	0,17	0,16	0,2
0,12	0,09	0,10	0,18
0,23	0,20	0,21	0,28
0,12	0,15	0,13	0,2

Составляется матрица планирования эксперимента согласно табл. 2.2.

№ п/п	Входные факторы						Выходной фактор				
	в натуральных единицах			в относительных единицах							
	z_1	z_2	z_3	x_1	x_2	x_3	y_1	y_2	y_3	y_4	$\bar{y}$
1	5	13	48	–1	–1	–1	0,12	0,11	0,10	0,11	0,11
2	18	13	48	+1	–1	–1	0,06	0,07	0,08	0,05	0,065
3	5	25	48	–1	+1	–1	0,20	0,19	0,20	0,25	0,21
4	18	25	48	+1	+1	–1	0,18	0,21	0,17	0,15	0,1775
5	5	13	64	–1	–1	+1	0,12	0,17	0,16	0,2	0,1625
6	18	13	64	+1	–1	+1	0,12	0,09	0,10	0,18	0,1225
7	5	25	64	–1	+1	+1	0,23	0,20	0,21	0,28	0,23
8	18	25	64	+1	+1	+1	0,12	0,15	0,13	0,2	0,15

Рассчитывается среднее значение параллельных опытов

$$\bar{y}^i = \frac{1}{v}(y^{i1} + y^{i2} + \dots + y^{iv})$$

$$\bar{y}^1 = \frac{1}{4}(0{,}12 + 0{,}11 + 0{,}10 + 0{,}11) = 0{,}11\ ,$$

…

$$\bar{y}^8 = \frac{1}{4}(0{,}12 + 0{,}15 + 0{,}13 + 0{,}2) = 0{,}15\ .$$

Уравнение регрессии с учетом взаимодействия факторов согласно (2.4)

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_{12}x_1x_2 + b_{13}x_1x_3 + b_{23}x_2x_3 + b_{123}x_1x_2x_3\ .$$

Расширенная матрица планирования полного факторного эксперимента 2^3

№ п/п	x_0	x_1	x_2	x_3	x_1x_2	x_1x_3	x_2x_3	$x_1x_2x_3$	$\bar{y}$
1	+1	–1	–1	–1	+1	+1	+1	–1	0,11
2	+1	+1	–1	–1	–1	–1	+1	+1	0,065
3	+1	–1	+1	–1	–1	+1	–1	+1	0,21
4	+1	+1	+1	–1	+1	–1	–1	–1	0,1775
5	+1	–1	–1	+1	+1	–1	–1	+1	0,1625
6	+1	+1	–1	+1	–1	+1	–1	–1	0,1225
7	+1	–1	+1	+1	–1	–1	+1	–1	0,23
8	+1	+1	+1	+1	+1	+1	+1	+1	0,15

В соответствии с формулой (2.5) определяются коэффициенты уравнения регрессии:

$$b_0 = \frac{1}{8}\sum_{i=1}^{8} x_{0i} y_i = \frac{1}{8}\big((+1)0{,}11 + (+1)0{,}065 + \cdots + (+1)0{,}15\big) \cong 0.153$$

…

$$b_{12} = \frac{\sum_{i=0}^{8}(x_1\, x_2) y_i}{8} = \frac{\big((+1)0{,}11 + (-1)0{,}065 + \cdots + (+1)0{,}15\big)}{8} \cong -0.0034$$

…

Результаты расчета коэффициентов:

$b_0 = 0{,}153$ b_1 =-0,025 b_2 =0,038 b_3 =0,013

b_{12} =-0,0034 b_{13} =-0,0053 b_{23} =-0,015 b_{123} =-0,0066

После подстановки коэффициентов уравнение регрессии будет иметь следующий вид:

$$y_{\text{р}} = 0{,}153 - 0{,}025x_1 + 0{,}038x_2 + 0{,}013x_3 - 0{,}0034x_1x_2 - 0{,}0053x_1x_3 - 0{,}015x_2x_3 - 0{,}0066x_1x_2x_3\,.$$

Далее следует проверить значимость коэффициентов уравнения регрессии. Каждому коэффициенту уравнения регрессии соответствует переменная Стьюдента t_{bi}. Согласно (1.8) $t_{bi} = \frac{|bi|}{S_{bi}}$, где b_i – коэффициент уравнения регрессии, среднеквадратичная ошибка коэффициента $S_{bi} = \frac{S_y}{\sqrt{N}}$; N=8 – количество экспериментов; S_y – стандартная ошибка уравнения регрессии $S_y = \sqrt{\frac{S_e}{v\varphi}}$; v=4 – число параллельных опытов; S_e – ошибка эксперимента, см. (2.7); $\varphi = N(v-1) = 8(4-1) = 24$

$S_e = \sum_{i=1}^{8}\sum_{j=1}^{4}\left(y_э{}^{ij} - y_э{}^{i}\right)^2 = (0{,}12-0{,}11)^2 + (0{,}11-0{,}11)^2 + (0{,}10-0{,}11)^2 +$
$+ (0{,}11-0{,}11)^2 + (0{,}06-0{,}065)^2 + \cdots = 0{,}020525$

$S_y = \sqrt{\frac{0{,}020525}{4\cdot 24}}$=0,0146 $\quad S_{\text{bi}} = \frac{0{,}0146}{\sqrt{8}} = 0{,}005$

$t_{b0} = \frac{|0{,}153|}{0{,}005} = 29{,}68$ $\quad t_{b1} = 4{,}78$ $\quad t_{b2} = 7{,}44$ $\quad t_{b3} = 2{,}47$

$t_{b12} = 0{,}66$ $\quad t_{b13} = 1{,}03$ $\quad t_{b23} = 2{,}84$ $\quad t_{b123} = 1{,}27$.

Критическое значение Стьюдента для уровня значимости 0,05 и $\varphi = 24$, согласно Приложению 2, $t_{\text{табл}} = 2{,}064$.

Сравнив критическое и рассчитанные значения критерия Стьюдента для каждого из коэффициентов, определяют незначимые коэффициенты и исключают их из уравнения регрессии. В данном примере незначимые коэффициенты $- b_{12}, b_{13,} b_{123}$.

Уравнение регрессии будет иметь следующий вид:

$$y_р = 0{,}153 - 0{,}025x_1 + 0{,}038x_2 + 0{,}013x_3 - 0{,}015x_2x_3 \ .$$

Далее рассчитываются показатели проверки адекватности модели.

Подставив исходные данные в полученное уравнение регрессии находятся расчетные значения выходного показателя
$y_р$= (0,112; 0,063; 0,218; 0,169; 0,167; 0,118; 0,215; 0,165).

Дисперсия адекватности S_D (см. 2.6) при $\varphi_1 \ = N - (k+1) = 8 - (5+1) = 2$

$$S_D = \sum_{i=1}^{8} 4 \cdot ((0{,}11 - 0{,}112)^2 + (0{,}065 - 0{,}063)^2 + \cdots) = 0{,}0027$$

Дисперсия воспроизводимости $S_e = 0{,}020525$ для $\varphi_2 = 8 \cdot (4-1) = 24$.

Критическое значение распределения Фишера для числа степеней свободы $\varphi_1 \ = 2 \quad \varphi_2 \ = 24$ и уровня значимости *p=0,05*, согласно Приложению 3, составляет $F_{крит}$ =3,4.

Расчетное значениē критерия Фишера:

$$F_{расч} = \frac{S_D:\varphi_1}{S_e:\varphi_2} = \frac{0{,}0027:24}{0{,}021:2} = 0{,}011 < F_{кр} = 3{,}4.$$

Таким образом, результаты подтверждают гипотезу об адекватности модели.

Приведем математическую модель от относительных единиц к натуральным. Для этого потребуется рассчитать центр плана (2.1) и интервал варьирования (2.2):

$$Z_1^0 = \frac{5+18}{2} = 11{,}5; \ \ Z_2^0 = \frac{13+25}{2} = 19; \ \ Z_3^0 = \frac{48+64}{2} = 56;$$

$$\Delta Z_1 = \frac{18-5}{2} = 6{,}5; \ \ \Delta Z_2 = \frac{25-13}{2} = 6; \ \ \Delta Z_3 = \frac{64-48}{2} = 8;$$

$$x_1 = \frac{Z_1 - 11{,}5}{6{,}5}; \ x_2 = \frac{Z_2 - 18}{6}; \ x_3 = \frac{Z_3 - 56}{8}.$$

С учетом проведенных преобразований уравнение регрессии в натуральных единицах будет иметь вид

$$y_{\text{р}} = 0{,}153 - 0{,}025\frac{Z_1 - 11{,}5}{6{,}5} + 0{,}038\frac{Z_2 - 18}{6} + 0{,}013\frac{Z_3 - 56}{8} - 0{,}015\frac{Z_2 - 18}{6} \cdot \frac{Z_3 - 56}{8},$$

$$\text{y}_{\text{р}} = 3{,}8 \cdot 10^{-4} \cdot \text{Z}_1 + 0{,}023 \cdot \text{Z}_2 + 7{,}2 \cdot 10^{-3} \cdot \text{Z}_3 - 3{,}1 \cdot 10^{-4} \cdot \text{Z}_2\text{Z}_3 - 0{,}327.$$

Выводы: в результате обработки данных эксперимента получена математическая модель процесса:

$$y_{\text{р}} = 0{,}153 - 0{,}025x_1 + 0{,}038x_2 + 0{,}013x_3 - 0{,}015x_2x_3\ ,$$

$$\text{y}_{\text{р}} = 3{,}8 \cdot 10^{-4} \cdot \text{Z}_1 + 0{,}023 \cdot \text{Z}_2 + 7{,}2 \cdot 10^{-3} \cdot \text{Z}_3 - 3{,}1 \cdot 10^{-4} \cdot \text{Z}_2\text{Z}_3 - 0{,}327.$$

Полученная модель адекватна для уровня значимости $p=0{,}05$.

3. ПОСТРОЕНИЕ МАТЕМАТИЧЕСКОЙ МОДЕЛИ С ПОМОЩЬЮ МЕТОДА НАИМЕНЬШИХ КВАДРАТОВ

В теории планирования эксперимента различают *пассивный* и *активный* эксперименты.

Если разработчик модели имеет возможность контролировать и управлять изменениями всех значимых параметров – факторов, влияющих на модель, то такой эксперимент можно назвать *активным*. Активный эксперимент позволяет максимально быстро решить поставленные перед исследователем задачи. В то же время он более сложен, требует больших материальных затрат и влияет на нормальный ход технологического процесса.

Если исследователь может только наблюдать за ходом технологического процесса при изменениях технологических параметров и регистрировать полученные изменения, то такой эксперимент называется *пассивным*.

Задача планирования эксперимента сводится к оптимальной организации сбора данных и выбору количества и частоты измерений, а также выбору метода обработки результатов измерений.

В случае допустимого выбора между пассивным и активным экспериментом предпочтение отдают последнему.

Модель — это отображение знаний или представлений об объекте, выраженное в форме, зависящей: от целей моделирования; от средств; от уровня знаний об объекте. Нельзя построить модель объекта, о котором ничего не известно. Модель строится на основе данных опыта или наблюдения и отражает только те свойства объекта, которые подлежат изучению.

Модель – это объект любой природы, который при исследовании способен замещать реально существующий объект с целью получения новой информации о последнем [4].

Моделирование — это процесс создания и исследования некоторой системы, имеющей определенное сходство с объектом.

Модель может быть представлена в физической и математической формах. Физическая модель строится в уменьшенном геометрическом размере с соблюдением всех критериев подобия. Математическая модель – совокупность математических объектов (уравнений, систем уравнений и неравенств, алгебраических выражений и т.д.), описывающих языком математических символов исследуемый объект и его отношения с окружающим миром.

Чтобы математическую модель можно было использовать для исследования реального объекта, она должна удовлетворять следующим требованиям [5]:

– быть практически полезной;
– быть адекватной реальному объекту;
– быть адекватной решаемым задачам.

Так как всякий эксперимент связан с появлением случайных ошибок, то при построении математических моделей на основе экспериментальных данных используют методы математической статистики. В этом случае наиболее часто используют метод наименьших квадратов.

Пусть требуется на основе экспериментальных данных построить математическую модель некоторого процесса. Найти функциональную взаимосвязь между некоторыми независимыми входными переменными, представленными вектором $\bar{x} = (x_{1,} x_2, x_3, \dots x_n)$, и выходной зависимой переменной *y,* т.е. получить зависимость *y=y(x).*

В общем случае математическая модель исследуемого объекта записывается в виде полинома n-степени, т.е. представляет часть ряда Тейлора, в который раскладывается неизвестная функция:

$$y(x_1, \dots, x_n) = b_0 + \sum_{i=1} b_i * x_i + \sum_{\substack{i,j=1 \\ i \neq j}}^{n} b_{ij} * x_j * x_i + \sum_{\substack{i,j,u=1 \\ i \neq j \neq u}}^{n} b_{iju} * x_j * x_i * x_u + \dots,$$

где b_0 – свободный член; b_i – линейные эффекты; b_{ij} – члены парного взаимодействия; b_{ii} – квадратичные эффекты; b_{iju} – члены тройного взаимодействия.

Если структура модели выражена линейной зависимостью, то модель процесса может описываться как

$$y=b_0 + b_1x_1 + b_2x_2 + \cdots + b_nx_n, \qquad (3.1)$$

где b_0, b_1, $b_2 \dots b_n$ – неизвестные параметры процесса, оценки которых надо найти путем обработки экспериментальных данных.

Если характер связи описывается квадратичной функцией, будет получена модель процесса в виде [6, 8]:

$$\begin{aligned} y=&b_0+b_1x_1 + b_2x_2 + \cdots +b_nx_n + \\ &+b_{n+1}x_1x_1 + b_{n+2}x_1x_2 + \cdots +b_{2n}x_1x_n + \\ &+b_{2n+1}x_2x_2 + b_{2n+2}x_2x_3 + \cdots + b_{3n-1}x_2x_n + \cdots + b_kx_nx_n. \end{aligned} \qquad (3.2)$$

Пусть *y=y(x,b)* – модель, линейная относительно коэффициентов $b = (b_0, b_1 \dots b_k)'$,[5] т.е.

$$y(b,x) = b_0f_0(x) + b_1f_1(x) + \cdots + b_kf_k(x), \qquad (3.3)$$

где $f_0(x), f_1(x) \dots f_k(x)$ – известные функции – компоненты вектора $f(x)$.

В векторной форме модель (3.3) запишется как

$$y = b'f(x) = f'(x)b. \qquad (3.4)$$

Для модели (3.1) вектор $f(x)$ будет выглядеть

$$f(x) = (1, x_1, x_2, \dots, x_n)'. \qquad (3.5)$$

Для модели вида (3.2):

$$f(x) = (1, x_1, x_2, \dots, x_n, x_1x_1, x_1x_2, \dots, x_1x_n, x_2x_2, x_2x_3, \dots x_nx_n).$$

Для *реальных* (действительных) значений вектора коэффициентов $\bar{b}$ требуется найти *расчетные* оценки $\hat{b}$, используя для этого результаты

---

[5] Штрихом здесь и далее обозначается операция транспонирования. Транспонирование вектора $\bar{a} = \begin{bmatrix} a_1 \\ a_2 \\ \dots \\ a_n \end{bmatrix}$ представляет собой замену столбца на строку $\bar{a}' = [a_1, a_2, \dots, a_n]$.

эксперимента $y_э$. При этом расчетное значение $y_р$ будет получено по формуле (3.4):

$$y_р = \hat{b}' f(x) = f'(x)\hat{b} \qquad (3.6)$$

Пусть пространство эксперимента ограничено N точками $x^1, x^2, \dots, x^i, \dots, x^N$ с координатами

$$x^1 = (x_1^1, x_{2,}^1 \dots, x_n^1),$$
$$x^2 = (x_1^2, x_{2,}^2 \dots, x_n^2),$$
$$\dots$$
$$x^N = (x_1^N, x_2^N, \dots, x_n^N).$$

Множество всех точек эксперимента $x^i, i = \overline{1, N}$ называется *планом эксперимента*, тогда *матрица плана* эксперимента будет иметь вид:

$$X = (x_j^i) = \begin{pmatrix} x_1^1 & x_2^1 & \dots & x_n^1 \\ x_1^2 & x_2^2 & \dots & x_n^2 \\ \dots & \dots & \dots & \dots \\ x_1^N & x_2^N & \dots & x_n^N \end{pmatrix}.$$

Результаты эксперимента, полученные в точках x^i, $i = \overline{1, N}$, обозначаются с помощью вектора наблюдений

$$Y_э = ({y_э}^1, {y_э}^2, \dots {y_э}^N). \qquad (3.7)$$

Если в каждой точке x^i поставлено v параллельных опытов, то для данной точки будет получено v результатов ${y_э}^{i1}$, ${y_э}^{i2}, \dots, {y_э}^{iv}$.

В этом случае в качестве ${y_э}^i$ используется среднее значение параллельных опытов в точке x^i:

$${y_э}^i = \frac{1}{v}({y_э}^{i1} + {y_э}^{i2} + \dots + {y_э}^{iv}).$$

Задача состоит в том, чтобы по результатам эксперимента (3.7) найти наилучшие оценки $\hat{b}$ и $y_р$. Для этого требуется сопоставить результаты эксперимента $Y_э$ (3.7) и значения вектора $Y_р$ = (${y_р}^1, {y_р}^2, \dots {y_р}^N$), полученные расчетным путем, формула (3.6) для вектора $\hat{b}$ и координат соответствующих точек эксперимента x^i:

$$y_{\text{р}}{}^{i} = \hat{b}' f(x^i) = f'(x^i)\hat{b}.$$

Соответственно, для вектора $Y_{\text{р}}$ модель будет выглядеть как

$$Y_{\text{р}} = F * \hat{b}, \tag{3.8}$$

где F представляет собой следующую таблицу векторов $f_j(x^i)$:

$$F = \left(f_j(x^i)\right) = \begin{pmatrix} f_0(x^1) & f_1(x^1) & \dots & f_k(x^1) \\ f_0(x^2) & f_1(x^2) & \dots & f_k(x^2) \\ \dots & \dots & \dots & \dots \\ f_0(x^N) & f_1(x^N) & \dots & f_k(x^N) \end{pmatrix}.$$

Для линейной модели (3.1) F будет иметь следующий вид:

$$F = \begin{pmatrix} 1 & x_1^1 & \dots & x_n^1 \\ 1 & x_1^2 & \dots & x_n^2 \\ \dots & \dots & \dots & \dots \\ 1 & x_1^N & \dots & x_n^N \end{pmatrix}. \tag{3.9}$$

Результаты эксперимента $y_{\text{э}}{}^{i}$ в точках $x^i, i = \overline{1, N}$ зависят от случайной ошибки e^i: $e^i = y_{\text{э}}{}^{i} - \bar{y}^i$. Множество всех значений ошибок в N экспериментальных точках задается вектором $e = Y_{\text{э}} - \bar{Y}$, т.е. разницей между множеством значений, полученных в результате эксперимента, и множеством действительных значений.

Для применения метода наименьших квадратов на результаты эксперимента накладываются следующие ограничения:

1. Результаты эксперимента не содержат систематических ошибок, т.е. математическое ожидание величины $y_{\text{э}}{}^{i}$ равно действительному значению $\bar{y}^i$:

$$\mathrm{M}(Y_{\text{э}}) = \bar{Y} \text{ и } \mathrm{M}(e) = 0.$$

2. Результаты наблюдений в точке x^i не зависят от результатов в точке x^j:

$$\mathrm{M}\left(\left(y_{\text{э}}{}^{i} - \bar{y}^i\right) \cdot \left(y_{\text{э}}{}^{j} - \bar{y}^j\right)\right) = 0 \text{ для } i \neq j$$

или

$$\mathrm{M}\left(e^i \cdot e^j\right) = 0 \text{ для } i \neq j.$$

3. Дисперсия результатов во всех точках эксперимента x^i одинакова:

$$D\left(\tilde{y}^i\right) = \sigma^2 \text{ для } \forall i$$

или $$D(\tilde{e}^i) = \sigma^2 \text{ для } \forall i.$$

Теоретическую основу метода наименьших квадратов составляет утверждение: сумма квадратов отклонений расчетных и экспериментальных оценок минимальна:

$$S = \sum_{i=1}^{N} \left(y_э{}^i - y_р{}^i\right)^2 = |Y_э - \bar{Y}|^2 = (Y_э - \bar{Y})' \cdot (Y_э - \bar{Y}),$$

$$\text{т.е. } S(\hat{b}) = \min_b S(b).$$

Оценка $\hat{b}$ имеет единственный минимум при

$$\hat{b} = (F'F)^{-1}F'Y_э, \tag{3.10}$$

где $|F'F| \neq 0$.

Матрица $\mathrm{C} = (F'F)^{-1}$ называется дисперсионной матрицей. Тогда $\hat{b} = \mathrm{C}F'\tilde{Y}$.

Практическая работа № 3

Цель работы: изучение метода наименьших квадратов при построении модели процесса по экспериментальным данным, проверка адекватности полученной модели.

Задание:

По результатам проведенного эксперимента требуется получить линейную модель процесса, описывающую зависимость некоторой величины y от переменных (факторов) x_1 , x_2, x_3 , x_4 и их парных взаимодействий x_1x_3 , x_2x_3, остальные парные взаимодействия принимаются незначимыми.

Требуется записать искомый вид модели, сформировать дробный факторный план эксперимента, найти оценки параметров модели.

Далее найти доверительные интервалы для параметров модели с доверительной вероятностью 0,95 и проверить адекватность полученной модели для уровня значимости 0,01.

При построения математической модели следует ввести в модель фактор x_4, равный взаимодействию x_1 и x_2: $x_4 = x_1x_2$.

Работа выполняется в *Microsoft Excel*.

Варианты:

I				
x_1	x_2	x_3	y^{i1}	y^{i2}
+	-	-	12	21
-	+	-	43	43
+	+	-	24	34
-	-	+	31	37
+	-	+	62	46
+	+	+	63	44
-	+	+	21	22
-	-	-	6	13

II				
x_1	x_2	x_3	y^{i1}	y^{i2}
+	+	-	11	22
-	-	+	44	42
+	-	+	23	31
+	+	+	34	43
-	+	+	65	43
+	-	-	66	44
-	+	-	27	22
-	-	-	5	10

III				
x_1	x_2	x_3	y^{i1}	y^{i2}
+	-	+	11	22
+	+	+	44	42
-	+	+	23	31
+	-	-	34	43
-	+	-	65	43
+	+	-	66	44
-	-	+	27	22
-	-	-	9	10

IV				
x_1	x_2	x_3	y^{i1}	y^{i2}
-	+	+	12	22
+	-	-	23	12
-	+	-	34	21
+	+	-	41	43
-	-	+	52	43
+	-	+	63	54
+	+	+	21	32
-	-	-	8	7

V				
x_1	x_2	x_3	y^{i1}	y^{i2}
-	+	-	12	22
+	+	-	23	42
-	-	+	34	31
+	-	+	41	43
+	+	+	52	43
-	+	+	63	44
+	-	-	21	22
-	-	-	3	11

VI				
x_1	x_2	x_3	y^{i1}	y^{i2}
-	-	+	11	21
+	-	+	44	43
+	+	+	23	34
-	+	+	34	37
+	-	-	65	46
-	+	-	66	44
+	+	-	27	22
-	-	-	8	14

VII				
x_1	x_2	x_3	y^{i1}	y^{i2}
+	+	+	11	22
-	+	+	44	42
+	-	-	23	31
-	+	-	34	43
+	+	-	65	43
-	-	+	66	44
+	-	+	27	22
-	-	-	1	11

VIII				
x_1	x_2	x_3	y^{i1}	y^{i2}
+	-	-	11	22
-	+	-	44	42
+	-	+	23	31
+	+	+	34	43
+	+	-	65	43
-	-	+	66	44
-	+	+	27	22
-	-	-	5	11

IX				
x_1	x_2	x_3	y^{i1}	y^{i2}
-	-	+	11	22
+	-	+	44	42
+	+	+	23	31
+	-	-	34	43
-	+	-	65	43
+	+	-	66	44
-	+	+	27	22
-	-	-	5	11

X				
x_1	x_2	x_3	y^{i1}	y^{i2}
+	-	+	11	22
+	+	+	44	42
-	+	+	27	31
+	-	-	36	43
-	+	-	45	43
+	+	-	67	44
-	-	+	25	22
-	-	-	5	11

Пример выполнения работы:

Модель выходной величины y в виде линейной функции согласно (3.1) запишется в виде

$$y=b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4 + b_{13}x_1x_3 + b_{23}x_2x_3 \quad ,$$

где x – входные факторы эксперимента; b – коэффициенты модели, являются неизвестными параметрами, оценки которых надо найти путем обработки экспериментальных данных.

К результатам эксперимента следует добавить недостающие столбцы: свободного члена x_0, фактора x_4 и значимых парных взаимодействий x_1x_3 , x_2x_3.

-								
x_0	x_1	x_2	x_3	x_4	x_1x_3	x_2x_3	y^{i1}	y^{i2}
	+	-	-				11	22
	-	+	-				44	42
	+	+	-				23	31
	-	-	+				34	43
	+	-	+				65	43
	+	+	+				66	44
	-	+	+				27	22
	-	-	-				5	11

Знаки «+» и «–» заменяются на 1 и -1, соответственно. Подобным образом заполняются столбцы x_4 ,x_1x_3, x_2x_3, учитывая, что фактор $x_4 = x_1x_2$. Столбец свободного члена принимается равным единице x_0 =1, см. (3.5):

-								
x_0	x_1	x_2	x_3	x_4	x_1x_3	x_2x_3	y^{i1}	y^{i2}
1	1	-1	-1	-1	-1	1	11	22
1	-1	1	-1	-1	1	-1	44	42
1	1	1	-1	1	-1	-1	23	31
1	-1	-1	1	1	-1	-1	34	43
1	1	-1	1	-1	1	-1	65	43
1	1	1	1	1	1	1	66	44
1	-1	1	1	-1	-1	1	27	22
1	-1	-1	-1	1	1	1	5	11

Для нахождения оценок параметров модели требуется формула (3.10):

$$\hat{b} = (F'F)^{-1}F'Y_э,$$

где матрица F, согласно (2.9), и вектор $Y_э$ будут иметь следующий вид:

$$F = \begin{pmatrix} 1 & 1 & -1 & -1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 \\ 1 & 1 & 1 & -1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & -1 & -1 & -1 & 1 & 1 & 1 \end{pmatrix} Y_э = \begin{pmatrix} 16{,}5 \\ 43 \\ 27 \\ 38{,}5 \\ 54 \\ 55 \\ 24{,}5 \\ 8 \end{pmatrix},$$

вектор $Y_э$ представляет собой среднее арифметическое исходных наблюдаемых значений $y_э^{i1}$ и $y_э^{i2}$.

Транспонированную матрицу можно получить, скопировав в *Microsoft Excel* исходную матрицу и вставив ее с помощью: *правая кнопка мыши* - *специальная вставка* - *транспонировать*.

-									
x_0	x_1	x_2	x_3	x_4	x_1x_3	x_2x_3	y^{i1}	y^{i2}	$\bar{Y}$
1	1	-1	-1	-1	-1	1	11	22	16,5
1	-1	1	-1	-1	1	-1	44	42	43
1	1	1	-1	1	-1	-1	23	31	27
1	-1	-1	1	1	-1	-1	34	43	38,5
1	1	-1	1	-1	1	-1	65	43	54
1	1	1	1	1	1	1	66	44	55
1	-1	1	1	-1	-1	1	27	22	24,5
1	-1	-1	-1	1	1	1	5	11	8

Умножение матриц в *Microsoft Excel* осуществляется с помощью оператора *МУМНОЖ()*, при этом необходимо выделить поле высотой, равной количеству строк первой матрицы, и шириной, равной количеству столбцов второй матрицы, после завершения ввода данных нужно нажать сочетание *Ctrl+Shift+Enter*.

$$F' \cdot F = \begin{pmatrix} 8 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 8 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 8 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 8 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 8 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 8 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 8 \end{pmatrix} = 8 \cdot E_7,$$

где E_7 – единичная матрица[6].

Дисперсионная матрица $C = (F'F)^{-1} = \frac{1}{8}E_7$, тогда $C \cdot F' = \frac{1}{8}F'$.

Для вычисления обратной матрицы в *MS Excel* существует специальная функция *МОБР()*.

$$\hat{b} = \begin{pmatrix} \hat{b}_0 \\ \hat{b}_1 \\ \hat{b}_2 \\ \hat{b}_3 \\ \hat{b}_4 \\ \hat{b}_{13} \\ \hat{b}_{23} \end{pmatrix} = CF'\tilde{Y} =$$

$$= \frac{1}{8}\begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 & 1 & -1 & -1 \\ -1 & 1 & 1 & -1 & -1 & 1 & 1 & -1 \\ -1 & -1 & -1 & 1 & 1 & 1 & 1 & -1 \\ -1 & -1 & 1 & 1 & -1 & 1 & -1 & 1 \\ -1 & 1 & -1 & -1 & 1 & 1 & -1 & 1 \\ 1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} 16,5 \\ 43 \\ 27 \\ 38,5 \\ 54 \\ 55 \\ 24,5 \\ 8 \end{pmatrix} = \begin{pmatrix} 33,31 \\ 4,81 \\ 4,06 \\ 9,69 \\ -1,19 \\ 6,69 \\ -7,31 \end{pmatrix}.$$

Полученная математическая модель будет иметь следующий вид:

$$y = 33,31 + 4,81x_1 + 4,06x_2 + 9,69x_3 - 1,19x_4 + 6,69x_1x_3 - 7,31x_2x_3 \quad .$$

---

[6] Единичная матрица – это диагональная матрица, каждый элемент главной диагонали которой равен единице. Пример, $E_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.

Оценки $\hat{b}$, рассчитанные выше, отличаются от истинных значений коэффициентов, причем ошибка i-го коэффициента тем больше, чем больше дисперсия ошибок наблюдений S^2. Требуется определить точность коэффициентов выбранной модели, или другими словами, определить доверительные интервалы для параметров модели.

Прежде всего следует найти расчётные значения $\hat{Y}$, согласно (3.8)

$$\hat{Y} = F \cdot \hat{b} = (11{,}56,\ 38{,}06,\ 31{,}94,\ 33{,}56,\ 58{,}94,\ 50{,}06,\ 29{,}44,\ 12{,}94)'.$$

В каждой из N точек x_i эксперимента поставлено $v = 2$ параллельных опыта. Для расчета оценок использовалось среднее значение $\bar{y}_э{}^i$.

Сумма квадратов, характеризующая ошибку эксперимента, рассчитывается как

$$s_e = \sum_{\boldsymbol{i=1}}^{N} \sum_{\mathrm{j}=1}^{v} \left(y_э{}^{ij} - y_э{}^{i}\right)^2. \qquad (3.11)$$

При этом s_e имеет число степеней свободы φ = $N \cdot v - N = N(v-1)$.

Для расчётного примера φ = $8 \cdot (2-1) = 8$,

$S_e = \sum_{i=1}^{8} \sum_{\mathrm{j}=1}^{2} \left(y_э{}^{ij} - y_э{}^{i}\right)^2 = (11 - 16{,}5)^2 + (22 - 16{,}5)^2 + (44 - 43)^2 +$
$+ (42 - 43)^2 \ldots$ = 649,5.

Дисперсия ошибок наблюдений S^2 получается с использованием суммы квадратов s_e, как

$$S^2 = \frac{S_e}{v\varphi} = \frac{649{,}5}{2 \cdot 8} = 40{,}59.$$

Стандартная ошибка $\hat{y}$, или среднеквадратичное отклонение наблюдаемых значений $S = \sqrt{S^2} = \sqrt{40{,}59} = 6{,}37$.

Среднеквадратичное отклонение коэффициента b_i: $S_{bi} = \sqrt{C_{ii}}S$, где $\sqrt{C_{ii}}$ – значение диагонального коэффициента дисперсионной матрицы С.

$$S_{bi} = \sqrt{\frac{1}{8}} \cdot 6{,}37 = 2{,}25.$$

При построении доверительного интервала коэффициента b_i для заданной доверительной вероятности P может быть использовано выражение (1.10)

$$\widehat{b_i} - t_{\text{табл}} \cdot S_{bi} < b_i < \widehat{b_i} + t_{\text{табл}} \cdot S_{bi},$$

где $t_{\text{табл}}$ – критическое значение Стьюдента, зависит от числа степеней свободы φ для заданного уровня значимости *p=1-P*. (см. таблицу распределения Стьюдента (Приложение 2)). В рассматриваемом случае p =0,05, φ = 8, по табл. Стьюдента $t_{\text{табл}} = 2{,}306$.

$$\widehat{b_i} - 2{,}306 \cdot 2{,}25 < b_i < \widehat{b_i} + 2{,}306 \cdot 2{,}25,$$

$$\widehat{b_i} - 5{,}19 < b_i < \widehat{b_i} + 5{,}19.$$

Например, для $b_0 = 33{,}31$ доверительный интервал составит

$$33{,}31 - 2{,}306 \cdot 2.25 < b_0 < 33{,}31 + 2{,}306 \cdot 2.25$$

$$28{,}12 < b_0 < 38{,}5$$

Подобным образом рассчитываются доверительные интервалы для остальных коэффициентов.

После следует проверить адекватность полученной модели. Для проверки гипотезы адекватности сравниваются дисперсия адекватности (2.6) и дисперсия воспроизводимости (2.7) (см. практическую работу № 2).

Требуется рассчитать дисперсию адекватности

$$S_D = \sum_{i=1}^{N} v\left(\widetilde{y}^{\,i} - \hat{y}^i\right)^2,$$

где k – количество коэффициентов модели, не считая нулевого (свободного члена), в случае рассматриваемого примера k=6;

при $\varphi_1 = N - (k+1) = 8 - (6+1) = 1$

$S_D = \sum_{i=1}^{8} 2 \cdot ((16{,}5 - 11{,}56)^2 + (43 - 38{,}06)^2 + \cdots) = 390{,}0625 \cong 390.$

Дисперсия воспроизводимости (формула 3.11) $S_e = 649{,}5$ для $\varphi_2 = 8 \cdot (2-1) = 8$.

Критическое значение распределения Фишера для числа степеней свободы $\varphi_1 = 1$, $\varphi_2 = 8$ и уровня значимости *p=0,01*, согласно Приложению 3 составляет $F_{\text{крит}}$ =11,26.

Расчетное значениē критерия Фишера:

$$F_{\text{расч}} = \frac{S_D : \varphi_1}{S_e : \varphi_2} = \frac{390:1}{649.5:8} = 4{,}8 < F_{\text{кр}} = 11{,}26.$$

Таким образом, результаты подтверждают гипотезу об адекватности модели.

Выводы: в результате обработки данных эксперимента получена математическая модель процесса

$$y = 33{,}31 + 4{,}81x_1 + 4{,}06x_2 + 9{,}69x_3 - 1{,}19x_4 + 6{,}69x_1x_3 - 7{,}31x_2x_3.$$

Получены доверительные интервалы для параметров модели с доверительной вероятностью 0,95:

$$\widehat{b_i} - 5{,}19 < b_i < \widehat{b_i} + 5{,}19,$$

$$28{,}12 < b_0 < 38{,}5.$$

Полученная модель адекватна для уровня значимости $p=0{,}01$.

ПРИЛОЖЕНИЯ

Приложение 1

Шкала Чеддока

Сила корреляционной зависимости	Значение коэффициента корреляции при наличии	
	прямой связи	обратной связи
Слабая	0,1–0,3	(-0,3)–(-0,1)
Умеренная	0,3–0,5	(-0,5)–(-0,3)
Заметная	0,5–0,7	(-0,7)–(-0,5)
Высокая	0,7–0,9	(-0,7)–(-0,9)
Весьма высокая	0,9–1	(-1)–(-0,9)

Приложение 2

Таблица распределения t-критерия Стьюдента

Число степеней свободы	Уровень					
	0,10	0,05	0,02	0,01	0,002	0,001
1	6,314	12,71	31,82	63,66	318,3	636,6
2	2,920	4,303	6,965	9,925	22,33	31,60
3	2,353	3,182	4,541	5,841	10,21	12,92
4	2,132	2,776	3,747	4,604	7,173	8,610
5	2,015	2,571	3,365	4,032	5,893	6,869
6	1,943	2,447	3,143	3,707	5,208	5,959
7	1,895	2,365	2,998	3,499	4,785	5,408
8	1,860	2,306	2,896	3,355	4,501	5,041
9	1,833	2,262	2,821	3,250	4,297	4,781
10	1,812	2,228	2,764	3,169	4,144	4,587
11	1,796	2,201	2,718	3,106	4,025	4,437
12	1,782	2,179	2,681	3,055	3,930	4,318
13	1,771	2,160	2,650	3,012	3,852	4,221
14	1,761	2,145	2,624	2,977	3,787	4,140
15	1,753	2,131	2,602	2,947	3,733	4,073
16	1,746	2,120	2,583	2,921	3,686	4,015
17	1,740	2,110	2,567	2,898	3,646	3,965
18	1,734	2,101	2,552	2,878	3,610	3,922
19	1,729	2,093	2,539	2,861	3,579	3,883
20	1,725	2,086	2,528	2,845	3,552	3,850
21	1,721	2,080	2,518	2,831	3,527	3,819
22	1,717	2,074	2,508	2,819	3,505	3,792
23	1,714	2,069	2,500	2,807	3,485	3,767
24	1,711	2,064	2,492	2.797	3,467	3,745
25	1,708	2,060	2,485	2,787	3,450	3,725
26	1,706	2,056	2,479	2,779	3,435	3,707
27	1,703	2,052	2,473	2,771	3,421	3,690
28	1,701	2,048	2,467	2,763	3,408	3,674
29	1,699	2,045	2,462	2,756	3,396	3,659
30	1,697	2,042	2,457	2,750	3,385	3,646
00	1,645	1,960	2,326	2,576	3,090	3,291

Приложение 3

Распределение Фишера (*F*-распределение)[7]

φ_2	φ_1							
	1	2	3	4	5	6	7	8
1	161,4 4052	199,5 4999	215,7 5403	224,6 5625	230,2 5764	234 5859	238,9 5981	243,9 6106
2	18,51 98,49	19 99,01	19,16 0,17	19,25 99,25	19,3 99,3	19,33 99,33	19,37 99,36	19,41 99,42
3	10,13 34,12	9,55 30,81	9,28 29,46	9,12 28,71	9,01 28,24	8,94 27,91	8,84 27,49	8,74 27,05
4	7,71 21,2	6,94 18	6,59 16,69	6,39 15,98	6,26 15,52	6,16 15,21	6,04 14,8	5,91 14,37
5	6,61 16,26	5,79 13,27	5,41 12,06	5,19 11,39	5,05 10,97	4,95 10,67	4,82 10,27	4,68 9,89
6	5,99 13,74	5,14 10,92	4,76 9,78	4,53 9,15	4,39 8,75	4,28 8,47	4,15 8,1	4 7,72
7	5,59 12,25	4,74 9,55	4,35 8,45	4,12 7,85	3,97 7,46	3,87 7,19	3,73 6,84	3,57 6,47
8	5,32 11,26	4,46 8,65	4,07 7,59	3,84 7,1	3,69 6,63	3,58 6,37	3,44 6,03	3,28 5,67
9	5,12 10,56	4,26 8,02	3,86 6,99	3,63 6,42	3,48 6,06	3,37 5,8	3,23 5,47	3,07 5,11
10	4,96 10,04	4,1 7,56	3,71 6,55	3,48 5,99	3,33 5,64	3,22 5,39	3,07 5,06	2,91 4,71
11	4,84 9,7	3,98 7,2	359 6,2	3,36 5,7	3,2 5,3	3,09 5,1	3,01 4,4	2,9 4
12	4,75 9,3	3,89 6,9	3,49 6	3,26 5,4	3,11 5,1	3 4,8	2,91 4,2	2,85 3,8
13	4,67 9,1	3,81 6,7	3,41 5,7	3,18 5,2	3,03 4,9	2,92 4,6	2,83 4	2,77 3,6
14	4,6 8,9	3,74 6,5	3,34 5,6	3,11 5	2,96 4,7	2,85 4,5	2,76 3,8	2,7 3,4
15	4,54 8,7	3,68 6,4	3,29 5,4	3,06 4,9	2,9 4,6	2,79 4,3	2,71 3,7	2,64 3,3
16	4,49 8,5	3,63 6,2	3,24 5,3	3,01 4,8	2,85 4,4	2,74 4,2	2,66 3,6	2,59 3,2
17	4,45 8,4	3,59 6,1	3,2 5,2	2,96 4,7	2,81 4,3	2,7 4,1	2,61 3,5	2,55 3,1
18	4,41 8,3	3,55 6	3,16 5,1	2,93 4,6	2,77 4,3	2,66 4	2,58 3,4	2,51 3

[7] Уровень значимости: верхняя строка – p=0,05, нижняя – p= 0,01

Окончание Приложения 3

φ_2	φ_1							
	1	2	3	4	5	6	7	8
19	4,38 8,2	3,52 5,9	3,13 5	2,9 4,5	2,74 4,2	2,63 3,9	2,54 3,3	2,48 2,9
20	4,35 8,1	3,49 5,9	3,1 4,9	2,87 4,4	2,71 4,1	2,6 3,9	2,51 3,2	2,45 2,9
22	4,3 7,9	3,44 5,7	3,05 4,8	2,82 4,3	2,66 4	2,55 3,8	2,46 3,1	2,4 2,8
24	4,26 7,8	3,4 5,6	3,01 4,7	2,78 4,2	2,62 3,9	2,51 3,7	2,42 3	2,36 2,7
26	4,23 7,7	3,37 5,5	2,98 4,6	2,74 4,1	2,59 3,8	2,47 3,6	2,39 3	2,32 2,6
28	4,2 7,6	3,34 5,5	2,95 4,6	2,71 4,1	2,56 3,8	2,45 3,5	2,36 2,9	2,29 2,5
30	4,17 7,6	3,32 5,4	2,92 4,5	2,69 4	2,53 3,7	2,42 3,5	2,33 2,8	2,27 2,5

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Математическое планирование эксперимента в научных исследованиях [Электрон. ресурс] - URL: https://studbooks.net/ 2243889/ matematika_himiya_fizika/matematicheskoe_planirovanie_eksperimenta_nauchnyh_issledovaniyah, 09.11.2020.

2. Основы планирования эксперимента [Электрон. ресурс] - URL: https://studme.org/232707/matematika_himiya_fizik/osnovy_planirovaniya_eksperimenta, 09.11.2020.

3. Диагностика регрессионных моделей [Электрон. ресурс] - URL: https://docplayer.ru/29962706-Diagnostika-regressionnyh-modeley.html, 09.11.2020.

4. Штерензон, В.А. Моделирование технологических процессов: конспект лекций / В.А. Штерензон. – Екатеринбург: Изд-во Рос. гос. проф.-пед. ун-та, 2010. - 66 с.

5. Зобнин, Б.Б. Моделирование систем: конспект лекций / Б.Б. Зобнин. – Екатеринбург: Изд-во УГГГА, 2001. 129 с.

6. Бойко, Н.Г. Теория и методы инженерного эксперимента: курс лекций / Н.Г. Бойко [и др.]. - Донецк: ЛО-ГОС, 2009.- 155 с.

7. Теория планирования эксперимента [Электрон. ресурс] - URL: http://appmath.narod.ru/page6.html, 09.11.2020.

8. Построение математической модели с помощью метода наименьших квадратов. [Электрон. ресурс]. Режим доступа: http://pgsksaa07.narod.ru/examples_method_square.htm, 21.03.2019.

9. Спирин, Н.А. Методы планирования и обработки результатов инженерного эксперимента: учебное пособие / Н. А. Спирин [и др.]. – Екатеринбург: УГТУ – УПИ. 2003. – 260 с.

СОДЕРЖАНИЕ

Редактор и корректор Н.П.Новикова
Техн.редактор Л.Я.Титова Темплан 2020 г. поз.112

Подп. к печати 16.11.2020. Формат 60х84/16. Бумага тип № 1.
Печать офсетная. Усл. печ. л. 3,25; 3,25 уч. - изд.л. Тираж 50 экз. Изд. № 112.
Цена «С». Заказ

Ризограф Высшей школы технологии и энергетики СПбГУПТД.
198095, СПб., ул. Ивана Черных, 4.