ОБРАБОТКА И ПЛАНИРОВАНИЕ ЭКСПЕРИМЕНТА

Методические указания к практическим занятиям для студентов магистратуры направления 18.04.01

САНКТ-ПЕТЕРБУРГ
2021

Министерство науки и высшего образования Российской Федерации

Федеральное государственное бюджетное образовательное учреждение
высшего образования
Санкт-Петербургский горный университет

Кафедра химических технологий и переработки энергоносителей

ОБРАБОТКА И ПЛАНИРОВАНИЕ ЭКСПЕРИМЕНТА

Методические указания к практическим занятиям для студентов магистратуры направления 18.04.01

САНКТ-ПЕТЕРБУРГ
2021

УДК 662.62+ 665.6/.7 (073)

ОБРАБОТКА И ПЛАНИРОВАНИЕ ЭКСПЕРИМЕНТА: Методические указания к практическим занятиям / Санкт-Петербургский горный университет. Сост.: *Е.В. Саламатова.* СПб, 2021. 44 с.

Изложены основные положения дисциплины «Обработка и планирование эксперимента», приведены примеры решения задач по данной дисциплине. Даны краткие теоретические сведения, необходимые для решения задач, основные расчетные формулы, справочные таблицы.

Предназначены для студентов магистратуры направления 18.04.01 «Химическая технология».

Научный редактор проф. *Н.К. Кондрашева*

Рецензенты д.т.н. *В.В. Васильев* (СПбГЭУ)

ВВЕДЕНИЕ

Дисциплина «Обработка и планирование эксперимента» относится к циклу дисциплин, формирующих профессиональный уровень студента, специализирующегося в области химических технологий. Целью дисциплины является получение будущими выпускниками знаний в области планирования и обработки инженерного и химического эксперимента.

В научно-исследовательской деятельности химика-технолога (бакалавра, магистра, аспиранта), кроме теоретических исследований, значительную долю составляют экспериментальные работы по определению параметров и характеристик различных химических систем, которые, зависят от множества внутренних и внешних факторов. Следовательно, исследователю – необходимо владеть методами теории математического планирования эксперимента для успешного решения научных, производственных и технологических проблем и задач.

Освоение содержания данного учебного пособия позволяет достигнуть основной учебной цели: сформировать у студентов компетентность в области теории и практики проведения экспериментального научного исследования. Студенты должны научиться правильно проводить математическую обработку экспериментальных данных, исключать грубые ошибки, определять необходимое количество измерений и проводить корреляционно-регрессионный анализ экспериментальных данных.

МАТЕМАТИЧЕСКАЯ ОБРАБОТКА РЕЗУЛЬТАТОВ ЭКСПЕРИМЕНТА

Сбор экспериментальных данных

Чтобы определить закон распределения случайной величины, нужно провести серию измерений или подсчётов для интересующей нас случайной величины (признака).

В результате получаем статистический ряд – это совокупность числовых данных или выборка объёмом n: затем производят упорядочивание членов выборки – эта операция называется ранжирование.

Ранжирование - это расположение всех имеющихся вариант по возрастанию. Получаем ранжированный статистический ряд.

Составление вариационного ряда

Вариационный ряд (статистическое распределение) - набор пар значение – частота, с которой это значение встретилось в выборке.

Если случайная величина изменяется дискретно, то составляем дискретный вариационный ряд. Графическое представление дискретного вариационного ряда - это полигон частот.

Если признак изменяется непрерывно, то составляется интервальный вариационный ряд: набор пар вид интервал – частота. Для построения интервального вариационного ряда выборку разбивают на интервалы. Есть несколько рекомендаций по вычислению числа интервалов:

$k=\log_2 n+1$ (формула Стерджесса), $k=\sqrt{n}$ и др

Длина интервала ΔX рассчитывается по формуле:

$$\Delta x = \frac{x_{max} - x_{min}}{k}$$

Графическая характеристика непрерывного вариационного ряда - Гистограмма.

Закономерности распределения генеральной совокупности оцениваются по выборочной совокупности.

При увеличении объёма выборки (n→∞), относительные частоты стремятся к вероятностям соответствующих значений с.в., то есть *к закону распределения*.

$$\frac{m_i}{n} \to P(\mathrm{x}_i)$$

Статистические характеристики совокупности

Характеристики генеральной совокупности:

- Математическое ожидание M[X]
- дисперсия D[X]
- среднее квадратическое отклонение σ[X]

Характеристики выборки (статистики)

- $\overline{x}$ - среднее арифметическое
- S_n^2 - дисперсия
- S_n - стандартное отклонение (среднее квадратическое)

$$M[X] = \sum_{i=1}^{k} x_1 \cdot P(x_1) = \frac{\sum_{i=1}^{n} x_1}{n} \qquad \sigma[X] = \sqrt{D[X]}$$

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

$$D[X] = \sum_{i=1}^{k}\left(x_i - M[X]\right)^2 \cdot P(x_i) = \frac{\sum_{i=1}^{n}\left(x_i - M[X]\right)^2}{n}$$

Ошибка среднего арифметического

Извлечём из генеральной совокупности N выборок, тогда их средние арифметические сами будут являться значениями случайной величины

Все эти значения имеют отклонения (рассеивание) от истинного значения M[X].

Это отклонение называется ошибка среднего арифметического, она в n раз меньше отклонения каждого x_i от для данной выборки объёмом n

$$\boxed{S_{\bar{x}} = \frac{S_n}{\sqrt{n}} = \sqrt{\frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}{n\cdot\left(n-1\right)}}}$$

$S_{\bar{x}}$ показывает насколько выборочное среднее арифметическое близко к матожиданию M[X] генеральной совокупности.

Чем больше объём выборки n, тем ближе среднее арифметическое к M[X] генеральной совокупности (т.е., ошибка меньше, чем больше n). Этот вывод получил название Закон больших чисел.

Доверительный интервал и доверительная вероятность

И Истинные значения M[X] и D[X] можно найти по генеральной совокупности, что практически невозможно. По выборке из этой совокупности мы находим лишь их точечные оценки и S_n^2, но насколько их значения близки истинным M[X] и D[X]. Например, как велика разность

$$\Delta x = \left|\bar{x} - M\left[X\right]\right|$$

Поэтому наряду с точечными оценками, применяют интервальные оценки параметров генеральной совокупности по выборке. То есть мы хотим найти интервал ΔX, такой что:

$$\bar{x} - \Delta x \le M\left[X\right] \le \bar{x} + \Delta x \quad \text{или} \quad M\left[X\right] - \Delta x \le \bar{x} \le M\left[X\right] \pm \Delta x$$

Если известна функция распределения, то этот интервал можно найти из соотношения:

$$\int_{M[X]-\Delta x}^{M[X]+\Delta x} f(x)dx = F\left(M\left[X\right] + \Delta x\right) - F\left(M\left[X\right] - \Delta x\right) = P\left(M\left[X\right] - \Delta x \le \bar{x} \le M\left[X\right] + \Delta x\right)$$

зная границы интервала, можно найти вероятность случайной величины $\bar{X}\left\{\bar{x}_1 .\ \bar{x}_2,\ \ldots \bar{x}_N\right\}$ принимать значения из данного интервала.

Но нам требуется решить обратную задачу: определить границы интервала, следовательно, для этого надо заранее задать вероятность, с которой мы этот интервал будем определять. Эту вероятность называют доверительной вероятностью $P_Д$, а определённый с её помощью интервал - доверительным интервалом $\Delta X_д$.

Доверительным интервалом какого-либо параметра, называют такой интервал, о котором можно сказать, что с вероятностью $P_Д$ он содержит в себе этот параметр.

Доверительную вероятность обычно берут равной $P_Д$=0,95, но в особо ответственных случаях принимают $P_Д$=0,99 или даже $P_Д$=0,999. С доверительной вероятностью связан уровень значимости $\alpha=1-P_Д$.

Уровень значимости α --это вероятность того, что значение исследуемого параметра не попадёт в доверительный интервал. Но для малых выборок (n<30) распределение может значительно отличаться от нормального.

В 1908 г английский математик и химик Уильям Госсет под псевдонимом Стьюдент предложил распределение случайной величины для малых выборок.

Нормированная случайная величина вычисляется по формуле:

$$\boxed{t=\frac{\bar{x}-M[X]}{S_{\bar{x}}}}$$

Плотность вероятности случайной величины:

$$S(t_{St},n)=B_n\cdot\left(1+\frac{t_{St}^{\ 2}}{n-1}\right)^{\frac{-n}{2}}$$

где B_n - параметр , зависит от n.

По мере увеличения объёма выборок n, распределение Стьюдента довольно быстро приближается к нормальному распределению Гаусса и при n>30 практически не отличается от него.

Практическим следствием этого открытия явилась возможность определять границы доверительного интервала для M[X] с заданной доверительной вероятностью $P_Д$:

$$\boxed{\Delta X_D = t_{St}(P_D, n) \cdot S_{\bar{x}} = t_{St}(P_D, n) \cdot \sqrt{\frac{\sum_{i=1}^{n} (x - \bar{x})^2}{n \cdot (n-1)}}}$$

Таким образом, определив доверительный интервал, можно записать:

$$\boxed{M[X] = \bar{x} \pm \Delta X_D}$$

Погрешностью измерения называется отклонение результата измерения от истинного значения измеряемой величины. Различают виды погрешностей результатов измерений:

1. **Грубая погрешность (или промахи)** – погрешность, существенно превышающая погрешность большинства результатов наблюдений при заданных условиях. Причиной такого вида погрешности может являться сбой измерительной аппаратуры. Грубые погрешности обнаруживают статистическими методами и обычно исключают из рассмотрения.

2. **Систематическая погрешность** – погрешность измерения, которая либо остается постоянной, либо закономерно изменяется в большую или меньшую сторону. Причиной такой погрешности может быть неправильная калибровка прибора или, например, резкое изменение внешних воздействий на исследуемую систему, когда крайне необходимо их постоянное влияние. Обнаружение и устранение систематической погрешности является одной из самых сложных задач при измерениях. Такие погрешности можно выявить с помощью образцовых приборов и эталонов. Оценка систематической погрешности может быть проведена также по результатам измерения искомой величины независимыми методами с применением различной аппаратуры. В результате обнаруженная и оцененная систематическая погрешность должна устраняться путем введения поправок. Однако даже введением поправок не удастся полностью избавиться от этой погрешности и какая-то ее остаточная часть, оставшаяся не устраненной, будет

представлять собой систематическую составляющую погрешности результата измерения.

3. **Случайная погрешность** – неустранимая погрешность измерения, представляющая собой отклонения от истинного значения измерения, меняющиеся при повторных измерениях. Данная погрешность обусловлена влиянием многих факторов, воздействие которых невозможно заранее предсказать. Такими причинами могут быть перепады давления, температуры, напряжения в сети, а также ошибки, связанные с действиями экспериментатора. Случайные ошибки нельзя исключить из результатов измерения. Однако они могут быть уменьшены с помощью статистической обработки результатов эксперимента.

4. **Инструментальная погрешность** – погрешность средств измерений, обусловленная конструктивными и/или технологическими недостатками средств измерений, а также следствиями их износа, неисправности. Является видом систематической погрешности.

5. **Методическая погрешность** – вид систематической погрешности, связанный не с прибором измерения, а с методом его использования. Полная погрешность измерения является суммой указанных выше составляющих.

По форме представления погрешности различают:

1. **Абсолютная погрешность** – разность между измеренным и истинным значением величины. Поскольку истинное значение величины неизвестно, то на практике вместо него используют действительное значение измеряемой величины, находимое экспериментально.

2. **Относительная погрешность** – отношение абсолютной погрешности измерения к истинному (действительному) значению измеряемой величины

Алгоритм обработки результатов прямых измерений

1.Провести серию измерений, не менее трех

$$\left\{x_1, x_2, \ldots x_N\right\}, \quad N \ge 3.$$

2.Найти среднее арифметическое

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \ldots + x_N}{N}.$$

3.Вычислить доверительный интервал (случайную ошибку).

для заданной доверительной вероятности, например, $P_D = 0{,}95.$

$$\boxed{\Delta\bar{x}_{\text{сл}} = t_{St} \cdot \sqrt{\frac{\sum_{i=1}^{N}\left(x_i - \bar{x}\right)^2}{N \cdot (N-1)}}}$$

4.Найти систематическую ошибку.

а). если указан класс точности прибора:

$$Кл.т. = \frac{\Delta x_{сист}}{x_{шкалы}} \cdot 100\% \quad \Rightarrow \quad \Delta x_{сист} = \frac{Кл.т. \cdot x_{шкалы}}{100\%}$$

где X шкалы – это предел шкалы (максимальное значение на шкале)

б). если класс точности не указан (например, линейка или термометр)

$$\Delta x_{сист} = \frac{цена\_деления}{2}$$

5. Вычислить общую ошибку:

$$\Delta\bar{x}_{общ} = \sqrt{\Delta\bar{x}^2_{сл.} + \Delta\bar{x}^2_{сист.}}$$

6. Записать окончательный результат:

$$\boxed{x = \bar{x} \pm \Delta\bar{x}_{\text{общ}}, \text{ для } P_D = 0{,}95}$$

7. Кроме абсолютной ошибки желательно также найти коэффициент вариации (или относительную ошибку, выраженную в процентах):

$$\omega\% = \frac{\Delta\bar{x}}{\bar{x}} \cdot 100\%.$$

ИСКЛЮЧЕНИЕ ГРУБЫХ ОШИБОК. НЕОБХОДИМОЕ ЧИСЛО ИЗМЕРЕНИЙ

Основные определения и теория процесса

Методы В некоторых случаях оказывается, что результат того или иного отдельного измерения резко отличается от результатов других измерений, выполненных в тех же условиях, т.е. такой результат содержит грубую погрешность (промах). Грубыми погрешностями измерений называют погрешности, существенно превышающие ожидаемые при данных условиях измерений. Причинами этих погрешностей являются ошибки оператора, неисправность измерительной аппаратуры, резкое изменение условий измерений, возникновение сильной кратковременной помехи, нарушение электрического контакта, толчок и т.д.

Результаты измерений, содержащие грубые погрешности, должны быть исключены из рассмотрения. Однако обнаружить грубую погрешность не всегда бывает легко и особенно опасны они при однократных измерениях. Чем больше ряд повторных измерений, тем легче обнаружить грубые погрешности.

Погрешности измерений, как и результаты измерений, следует рассматривать как случайные величины, подчиняющиеся статистическим законам распределения вероятностей. Грубые погрешности – это большие случайные погрешности. Их появление теоретически маловероятно (например, несколько погрешностей на 1 млн. измерений), но все же они возможны, и не исключена возможность, что уже один из первых результатов измерений будет содержать такую погрешность.

Между результатами, содержащими грубые погрешности, и результатами, заслуживающими доверия, зачастую на практике нельзя провести четкую границу. В этом случае вопрос о том, содержит ли данный результат измерений грубую погрешность, решается одним из методов проверки статистических гипотез. При выявлении грубых погрешностей проверяемая гипотеза предполагает, что результат наблюдения x_i не содержит грубой погрешности и является одним из значений случайной величины X с определенным законом распределения вероятностей (обычно – это нормальный закон). Сомнительным может быть в первую очередь наибольший (x_{max}) или наименьший (x_{min}) из результатов наблюдений. Методика проверки гипотезы сводится к следующему:

-на основании полученного ряда результатов наблюдений находят статистические оценки параметров нормального закона распределения вероятностей - математического ожидания и СКО:

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad \text{и} \quad \hat{v} = S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{X})^2}$$

-находят отношения

$$v_1 = \frac{x_{max} - \overline{X}}{S}$$

$$v_2 = \frac{\overline{X} - x_{min}}{S}$$

-по справочной таблице по заданной доверительной вероятности «*P*» и числу результатов наблюдений «*n*» находят наибольшее значение ν_p , которое случайная величина ν может принимать еще по чисто случайным причинам;

-если вычисленное по опытным данным значение ν_1 (или ν_2) окажется меньше ν_p, то гипотеза принимается; в противном случае результат x_{max} (или x_{min}) следует считать содержащим грубую погрешность и не учитывать его при дальнейшей обработке результатов измерений. В этом случае результат x_{max} (или x_{min}) исключают из ряда наблюдений и вновь полученный ряд проверяют на наличие грубых погрешностей по описанной методике.

Предельные значения ν_p для проверки грубой погрешности

n	ν_p при P, равной				n	ν_p при P, равной			
	0,9	0,95	0,98	0,99		0,9	0,95	0,98	0,99
3	1,41	1,41	1,41	1,41	13	2,26	2,43	2,56	2,71
4	1,64	1,69	1,71	1,72	14	2,30	2,46	2,60	2,76
5	1,73	1,78	1,92	1,96	15	2,33	2,49	2,64	2,81
6	1,89	2,00	2,07	2,13	20	2,49	2,64	2,80	2,90
7	1,97	2,09	2,18	2,65	25	2,62	2,78	2,96	3,08
8	2,04	2,17	2,27	2,37	30	2,72	2,88	3,07	3,20
9	2,10	2,24	2,35	2,46	35	2,79	2,96	3,16	3,29
10	2,15	2,29	2,41	2,54	40	2,85	3,00	3,22	3,36
11	2,19	2,38	2,47	2,61	45	2,90	3,08	3,28	3,42
12	2,23	2,39	2,52	2,66	50	2,99	3,16	3,37	3,52

Необходимое число измерений

Увеличение количества измерений (числа проб, образцов и т.п.), даже при неизменной их точности (σx = const), может увеличить доверительную вероятность Р или сузить доверительный интервал $\pm\delta$ для определения действительного значения измеряемой величины (математического ожидания).

Необходимое количество измерений (образцов, проб и т.п.) n для достижения требуемой точности δ при заданной доверительной вероятности Р можно определить заранее в том случае, когда известно действительное значение среднеквадратичного отклонения σx, а экспериментальные данные (измерения) подчиняются нормальному закону распределения.

Действительно, при этих допущениях число измерений можно определить из выражения:

$$n \geq (\frac{z_{1-\frac{\alpha}{2}} \cdot \sigma_x}{\delta})^2 = z^2_{1-\frac{\alpha}{2}} \cdot (\frac{\sigma_x}{\delta})^2 = z^2_{1-\frac{\alpha}{2}} \cdot (\varepsilon)^2,$$

где

$$\varepsilon = \frac{\sigma_x}{\delta}$$

Таким образом, число измерений n определяется требуемой доверительной вероятностью Р = 1 – α и относительным (по отношению к среднеквадратичному отклонению) значением половины ширины доверительного интервала δ, т.е. требуемой точностью определения измеряемой величины. Так, при Р = 0,95, z0,975 = 1,96 и при δ = σх число измерений равно 4. При увеличении необходимой точности измерений в 2 раза, т.е. сужении доверительного интервала до величины δ=(1/2)σх, необходимое число измерений составит 16. Нетрудно заметить, что необходимое число измерений с увеличением точности возрастает в квадратичной зависимости.

Как правило, действительное значение среднеквадратичной ошибки (σх) неизвестно, а имеется только ее оценка (Sх). В этом случае следует воспользоваться соотношением, т.е. критерием Стьюдента, и необходимое число измерений определять из соотношения:

$$n \geq \frac{t^2_{\text{а.т}} \cdot S_x^2}{\delta^2} = t^2_{a.\text{т}} \cdot (\frac{S_x}{\delta})^2 = t^2_{a.\text{т}} \cdot \varepsilon^2$$

где

$$\varepsilon = \frac{S_x}{\delta}$$

При расчетах по этому уравнению следует иметь в виду, что значение критерия Стьюдента зависит не только от α, но и от числа степеней свободы m, последние же определяются числом измерений. В связи с этим уравнение следует решать методом последовательных приближений. В качестве начального приближения можно задать, в частности, число измерений. Так, если решить последнее уравнение методом последовательных приближений, то можно показать, что при Р = 0,95 (α =0,05) для определения доверительного интервала с точностью δ=Sx требуется 7 измерений, а с точностью δ=0,5Sx – 19. С повышением необходимой точности различие в числе измерений уменьшается и, как показывают расчеты, при величине $\delta \le 0{,}2$Sx они практически совпадают.

ТОЧЕЧНАЯ ОЦЕНКА ПАРАМЕТРОВ РЕГРЕССИОННОЙ ЗАВИСИМОСТИ

Регрессионный анализ – статистический метод исследования зависимости между зависимой переменной y и одной или несколькими независимыми переменными $x_1, x_2, ..., x_n$.

При полном факторном эксперименте влияние факторов на результат эксперимента раскрывается с помощью уравнения регрессии.

Регрессия - величина, выражающая зависимость среднего значения случайной величины y от значений случайной величины x.

Уравнение регрессии выражает среднюю величину одного признака как функцию другого.

Построение уравнение регрессии осуществляется в два этапа:

– определение вида аналитической зависимости y=f(x);

– оценка параметров выбранной модели.

Для определения вида аналитической зависимости применяются три основных метода:

– графический (на основе анализа поля корреляций);

– аналитический, т. е. исходя из теории изучаемой взаимосвязи;

– экспериментальный, т.е. путем сравнения величины остаточной дисперсии или средней ошибки аппроксимации, рассчитанных для различных моделей регрессии (метод перебора).

Наиболее часто при проведении исследований встречаются линейные, параболические, гиперболические, степенные, экспоненциальные зависимости

Для оценки параметров выбранной модели используют метод наименьших квадратов (МНК). Он позволяет получить такие оценки параметров, при которых сумма квадратов отклонений фактических значений результативного признака y_i от теоретических значений y_0 при тех же значениях фактора x минимальна, т. е.:

$$\sum_{i=1}^{n}(y_1 - y_0)^2 \to \min$$

В случае линейной регрессии ($y_0 = a \pm bx$) параметры a и b находятся из следующей системы нормальных уравнений метода наименьших квадратов:

$$\begin{cases} na + b\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i \\ a\sum_{i-1}^{n} x_i + b\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} y_i x_i \end{cases},$$

где n – количество наблюдений.

Коэффициенты a и b при факторной переменной x имеют следующую интерпретацию:

– коэффициент a показывает значение средней величины y при x=0;

– коэффициент b показывает, насколько изменится в среднем величина y при изменении фактора x на 1 единицу измерения.

Точность аппроксимации уравнения регрессии оценивают с помощью средней ошибки аппроксимации – среднее относительной отклонение расчетных значений (y_0) от фактических (y_i). Ее определяют по следующему выражению:

$$\varepsilon = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - y_0}{y_i}\right| \cdot 100\%$$

В случае линейной регрессии ($y_0 = a \pm bx$) параметры a и b находятся из следующей системы нормальных уравнений метода наименьших квадратов:

$$\begin{cases} na + b\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i \\ a\sum_{i-1}^{n} x_i + b\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} y_i x_i \end{cases},$$

Коэффициенты a и b при факторной переменной x имеют следующую интерпретацию:

– коэффициент a показывает значение средней величины y при x=0;

– коэффициент b показывает, насколько изменится в среднем величина y при изменении фактора x на 1 единицу измерения.

Точность аппроксимации уравнения регрессии оценивают с помощью средней ошибки аппроксимации - среднее относительной отклонение расчетных значений (y_0) от фактических (y_i). Ее определяют по следующему выражению:

$$\varepsilon = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - y_0}{y_i}\right| \cdot 100\%$$

Построенное уравнение регрессии считается удовлетворительным, если средняя ошибка аппроксимации не превышает 10 %.

Для оценки значимости уравнения регрессии в целом используется F-критерий Фишера.

F-критерий Фишера заключается в проверке гипотезы H_o о статистической незначимости уравнения регрессии. Для этого выполняется сравнение фактического – $F_{факт}$ и критического (табличного) $F_{табл}$ значений F-критерия Фишера.

$F_{факт}$ определяется из соотношения значений факторной и остаточной дисперсий, рассчитанных на одну степень свободы

$$F_{факт} = \frac{r^2}{1 - r^2} \cdot \frac{n - m - 1}{m},$$

где n – число единиц совокупности (количество наблюдений); m – число параметров при переменных. Для линейной регрессии m = 1.

$F_{факт}$ – максимально возможное значение критерия под влиянием случайных факторов при степенях свободы

$$k_1 = m,$$

$$k_2 = n - m - 1,$$

и уровне значимости α.

Уровень значимости α. – вероятность отвергнуть правильную гипотезу при условии, что она верна. Обычно величина α. принимается равной 0,05 или 0,01.

Если $F_{табл} < F_{факт}$, то H_o-гипотеза о случайной природе оцениваемых характеристик отклоняется и признается их статистическая значимость и надежность. Если $F_{факт} > F_{табл}$, то гипотеза H_o не отклоняется и признается статистическая незначимость, ненадежность уравнения регрессии.

ОДНОФАКТОРНЫЙ ЭКСПЕРИМЕНТ. ПОЛНЫЙ ФАКТОРНЫЙ ЭКСПЕРИМЕНТ

Описание исследуемого объекта нельзя получить в виде точной формулы, справедливой во всем диапазоне существования аргументов. Оно может быть лишь приближенным и на небольшом участке в окрестностях выбранной базовой точки. Аппроксимация искомой математической зависимости представляет собой некоторый полином – отрезок ряда Тейлора.

Рассмотрим снова матрицу планирования 2^2. Предположим, что для движения к оптимуму нужна линейная модель типа $Y = b_0 + b_1 \cdot X_1 + b2 \cdot X_2$.

Наша задача – найти неизвестные коэффициенты, причем эксперимент, содержащий конечное число опытов, позволяет получить только выборочные оценки для коэффициентов уравнения. Их точность зависит от свойств выборки и нуждается в стат. проверке.

Оценки коэффициентов вычисляются по простой формуле:

$$b_j = \frac{\sum_i X_{ji} \cdot Y_i}{N}, \qquad j = 0, 1, 2, \dots k.$$

$$b_1 = \frac{(-1)Y_1 + (+1)Y_2 + (-1)Y_3 + (+1)Y_4}{4},$$

$$b_2 = \frac{(-1)Y_1 + (-1)Y_2 + (+1)Y_3 + (+1)Y_4}{4}.$$

Коэффициенты при переменных указывают на силу влияния факторов. Чем больше коэффициент по абсолютной величине, тем больше влияние на эксперимент оказывает данный фактор.

Знак плюс говорит о том, что параметр оптимизации увеличивается с увеличением фактора, минус – наоборот.

Планируя эксперимент на первом этапе, мы стремимся получить линейную модель. Однако у нас нет гарантии, что в выбранных интервалах варьирования модель линейна. Один из часто встречающихся видов нелинейности связан с тем, что эффект одного фактора зависит от уровня, на котором находится другой фактор, то есть просматривается эффект взаимодействия двух факторов.

АЛГОРИТМ ЭКСПЕРИМЕНТАЛЬНОГО ПОИСКА ЭКСТРЕМУМА

Основные определения и теория процесса

Метод градиентного спуска

Рассматривается задача безусловной минимизации целевой функции $f(x): R^n \to R$.

$$f(x) \to \min_{x \in R^n}.$$

Идея методов приближенного решения поставленной задачи состоит в построении последовательности точек, удовлетворяющих условию.

$$f\left(x^{(0)}\right) \geq f\left(x^{(1)}\right) \geq \ldots \geq f\left(x^{(k)}\right) \geq \ldots$$

Процедура построения последовательности по методу градиентного спуска

$$x^{(k+1)} = x^{(k)} + \alpha^{(k)} y^{(k)}, k = 0, 1, 2, \ldots,$$

$$y^{(k)} = -\nabla f\left(x^{(k)}\right),$$

где величины шага спуска α(k) > 0 выбираются достаточно малыми для того, чтобы выполнялось условие последовательного убывания значений целевой функции.

Приведем анализ метода градиентного спуска:

1. Итерационная процедура спуска с постоянным шагом α = const проста, но при реализации требует знания величины шага α, которая на практике определяется подбором. При этом выбор большого шага α может привести к нарушению условия убывания целевой функции

$$f\left(x^{k+1}\right) = f(x^{(k)} + \alpha y^{(k)}) < f(x^{(k)}) \qquad (6.1)$$

С другой стороны, достаточно малый шаг обеспечит выполнение неравенства (5.4), но для достижения требуемой точности потребуется большое количество итераций. Поэтому сначала фиксируют некоторое значение шага спуска α>0 и производят вычисления по формуле. При этом если неравенство не выполняется, то для текущей итерации величину шага спуска а уменьшают (дробят) до тех пор, пока условие убывания функции не будет выполнено, и продолжают вычисления.

2. Условие окончания процедуры спуска. Выполняется требование на точность решения задачи минимизации, например

$$\left\|x^{(k+1)} - x^{(k)}\right\| \leq \varepsilon_1 \text{ или}$$

$$\left|f\left(x^{(k+1)}\right) - f\left(x^{(k)}\right)\right| \leq \varepsilon_2$$

ПРИМЕРЫ РЕШЕНИЯ ЗАДАЧ

Задача №1

Выполнить статистическую обработку результатов измерений pH KOH, приведенных в таблице

X1	X2	X3	X4	X5	X6	X7	X8	X9
9,791	9,795	9,789	9,784	9,796	10,025	9,793	9,793	9,765
X10	X11	X12						
9,794	9,797	9,761						

Решение

1 Значения результатов наблюдения упорядочивают по возрастающим значениям в вариационный ряд x_1, x_2,...,x_n.

Вариационный ряд результатов наблюдений при измерении сопротивления **R** число наблюдений n = 10:

9,992; 9,995; 9,997; 9,999; 10,000; 10,001; 10,003; 10,005; 10,007; 10,121.

2. Среднее арифметическое значение результатов наблюдений

$$\overline{R} = \frac{1}{n} \cdot \sum_{i=1}^{n} R_i = 10{,}012$$

3. Вычисляется оценка среднего квадратичного отклонения результатов наблюдений

$$\sigma_x = S = \sqrt{\sum_{i=1}^{9} (R_i - \overline{R})^2 \cdot \frac{1}{n-1}} = \sqrt{\sum_{i=1}^{10} (R_i - 10{,}012)^2 \cdot \frac{1}{10-1}}$$
$$= 0{,}04$$

4. Если значения x_i резко отличаются от других членов вариационного ряда (промах, грубая погрешность), то их отбрасывают и в обработке результатов наблюдений не учитывают. Для проверки вида погрешности (грубая или значительная случайная) используется статистический критерий обнаружения грубых погрешностей (ГОСТ 11.002-73).

Суть статистического способа оценки результатов наблюдений заключается в том, что грубыми признают те погрешности, вероятность появления которых не превышает некоторого, заранее выбранного критерия.

Воспользуемся отбраковкой некоторых результатов измерений по критерию превышения отклонения среднего удвоенного значения среднего квадратичного отклонения результатов наблюдений

$$\Delta R_i = \left(R_i - \overline{R}\right) > 2S.$$

В случае обнаружения грубых погрешностей результаты наблюдений, их содержащие, исключаются и математическая обработка повторяется. Для данного ряда проверим значение R_{10} = 10,121.

ΔR_i= 10,121 -10,012=0,109, ΔR_i= 0,109>2·0,04.

Отбрасываем R_{10} , принимаем n = 9 и повторяем п.2 и 3:

$$\overline{\mathrm{R}} = \frac{1}{9} \cdot \sum_{i=1}^{9} R_i = 10{,}000$$

$$S = \sqrt{\sum_{i=1}^{9} (R_i - 10{,}000)^2) \cdot \frac{1}{9-1}} = 4{,}78 \cdot 10^3$$

5. Определяется доверительный интервал (границы) случайной погрешности результатов наблюдений как

E = t·S,

где t - коэффициент (квантиль нормального распределения) Стьюдента, который в зависимости от вероятности P и числа результатов наблюдений берется из справочных материалов.

6. При нормальном законе распределения результатов наблюдений (при числе наблюдений n≤15 принадлежность их нормальному закону не проверяют) математическое ожидание случайной величины M(x) с заданной вероятностью должно находиться в границах (доверительном интервале)

$$\overline{x} - t\sigma_{\overline{x}} < M(x) < \overline{x} + t \cdot \sigma_{\overline{x}},$$

Где $\sigma_{\overline{x}}$ - среднее квадратичное отклонение действительного значения (среднего арифметического) результатов наблюдений,

$$\sigma_{\overline{x}} = \frac{S}{\sqrt{n}} \quad \text{или} \quad \sigma_{\overline{x}} = \frac{0{,}00478}{\sqrt{9}} = 0{,}0016$$

Коэффициент Стьюдента для n-1=8 и P=0,95; t=2,31.

Следовательно, доверительный интервал 10-2,31·0,0016 <R< 10+2,31·0,0016; или 9,996 < R <10,004.

Таким образом, при P=0,95 доверительный интервал R=(10±0,004).

Задача №2

При определении концентрации KI в растворе были получены следующие результаты (в мг/л):110, 112, 115, 113, 114. Найти среднее значение, стандартное отклонение и доверительный интервал для Рд=0.95.

Решение

$$\overline{x} = \frac{110 + 112 + 115 + 113 + 114}{5} = 112{,}8$$

$$S_n = \sqrt{\frac{\sum_{j=1}^{n}(x_j - \overline{x})^2}{n-1}}$$

$$= \sqrt{\frac{(110 + 112{,}8)^2 + (112 - 112{,}8)^2 + (115 - 112{,}8)^2 + (113 - 112{,}8)^2 + (114 - 112{,}8)^2}{5-1}}$$

$$= \sqrt{\frac{2{,}8^2 + 0{,}8^2 + 2{,}2^2 + 0{,}2^2 + 1{,}2^2}{4}}$$

$$= \sqrt{\frac{7{,}84 + 0{,}64 + 4{,}84 + 0{,}04 + 1{,}44}{4}} \sqrt{\frac{14{,}8}{4}} 1{,}92$$

$$S_{\bar{x}} = \frac{S_n}{\sqrt{n}} = \frac{1,92}{\sqrt{5}} = \frac{1,92}{2,24} = 0,86$$

$$S_{\bar{x}} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n(n-1)}} =$$

$$\sqrt{\frac{(110 + 112,8)^2 + (112 - 112,8)^2 + (115 - 112,8)^2 + (113 - 112,8)^2 + (114 - 112,8)^2}{5(5-1)}}$$

$$= \sqrt{\frac{14,8}{20}} = \sqrt{0,74} = 0,86$$

$\Delta x_D = t_{st} \cdot S_{\bar{x}} = 2,8 \cdot 0,86 = 2,4$ Для Рд=0,95

$M\left[X\right] = 112,8 \pm 2,4(мг/л)$

Задача №3

В параллельных опытах по измерению вязкости раствора полимера с использованием капиллярного вискозиметра время вытекания жидкости составило величины, представленные в таблице, определить грубую ошибку.

X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}
258,5	255,4	256,6	256,6	256,7	257,0	256,5	256,7	255,3	256,0
X_{11}	X_{12}	X_{13}	X_{14}	X_{15}					
256,3	256,5	256,3	256,9	266,0					

Решение

Вычисляем среднее значение времени:

$$\bar{\tau} = 257,7 \text{ сек.}$$

Определяем среднеквадратичное отклонение для числа степеней свободы:

$$S(\tau) = \sqrt{\frac{\sum_{i=1}^{n}(\tau_i - \bar{\tau})^2}{n-1}} 2,6 \text{ сек.}$$

Вычисляем максимальный критерий Стьюдента:

$$t_{\text{макс}} = \frac{|\tau_{\text{макс}} - \overline{\tau}|}{S(\tau)} = \frac{|266{,}0 - 257{,}1|}{2{,}6} 3{,}4$$

С сравниваем $t_{\text{макс}}$ с табличным значением для заданного уровня значимости, например α= 0,05. В итоге имеем 3,4>2,49, и результат τ =266,0 сек является промахом.

Задача №4

При измерении абсолютного давления в ректификационной колонне разделения бензола и толуола получены результаты, представленные в таблице. Систематические ошибки в результатах измерений исключены: все результаты одинаковой точности. Необходимо провести выявление грубых ошибок в результатах измерений (P = 0.95); при их выявлении исключить результаты, содержащие грубые ошибки из серии.

X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9
1,256	1,243	1,264	1,223	1,237	1,247	1,226	1,213	1,254
X_{10}	X_{11}	X_{12}	X_{13}					
1,224	1,322	1,227	1,254					

Решение

1) Вычисляем среднее значение по формуле (2.1):

$$\overline{X}_{13} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{13}\sum_{i=1}^{13} x_i = 1{,}247.$$

2) Вычисляем среднюю дисперсию по формуле (2.2):

$$D_{x.13} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{X})^2 = \frac{1}{12}\sum_{i=1}^{13}(x_i - 1{,}247)^2 = 7{,}38 \cdot 10^{-4}.$$

3) В серии результатов выбираем $X_{max} = X_{11} = 1{,}322$ и $X_{min} = X_8 = 1{,}213$; определяем по формулам:

$$v_{1.13} = \frac{X_{max} - \overline{X}}{\sqrt{D_x}} = \frac{1{,}322 - 1{,}247}{\sqrt{7{,}38 \cdot 10^{-4}}} = 2{,}76$$

$$v_{2.13} = \frac{\overline{X} - X_{min}}{\sqrt{D_x}} = \frac{1{,}247 - 1{,}213}{\sqrt{7{,}38 \cdot 10^{-4}}} = 1{,}25$$

4) В справочной литературе находим значение соответствующее $n = 13$ и $P = 0{,}95$ - $v_p = 2{,}43$.Поскольку $v_{1.13} = 2{,}76 > v_p = 2{,}43$, $X_{max} = X_{11} = 1{,}322$ содержит грубую ошибку. Исключаем $X_{11} = 1{,}322$ из ряда результатов; получаем «новый» ряд, содержащий 12 результатов измерений, и проверяем его на наличие грубых ошибок.

5) вычисляем среднее значение результатов измерений в «новом» ряду по формуле:

$$\overline{X}_{12} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{12}\sum_{i=1}^{12} x_i = 1{,}227$$

6) вычисляем дисперсию результатов измерений в «новом» ряду по формуле:

$$D_{x.12} = \frac{1}{n-1}\sum_{i-1}^{n}(x_i - \overline{X})^2 = \frac{1}{11}\sum_{i=1}^{12}(x_i - 1{,}227)^2 = 2{,}47 \cdot 10^{-4}$$

7) В серии результатов выбираем $X_{max} = X_3 = 1{,}264$ $X_{min} = X_8 = 1{,}213$; определяем по формулам:

$$v_{1.12} = \frac{X_{max} - \overline{X}}{\sqrt{D_X}} = \frac{1{,}264 - 1{,}227}{\sqrt{2{,}47 \cdot 10^{-4}}} = 1{,}5$$

$$v_{2.12} = \frac{\overline{X} - X_{min}}{\sqrt{D_X}} = \frac{1{,}227 - 1{,}213}{\sqrt{2{,}47 \cdot 10^{-4}}} = 1{,}72$$

8) В справочной литературе находим значение соответствующее $n = 12$ и $P = 0{,}95$ - $v_p = 2{,}39$.

Результаты измерений, входящие в «новый ряд» с вероятностью $P = 0{,}95$ не содержат грубых ошибок, поскольку $v_{1.12} = 1{,}5 < v_p = 2{,}39$, $v_{2.12} = 1{,}72 < v_p = 2{,}39$.

Задача №5

На условной установке оценивали влияние температуры на выход продукта, результаты измерений представлены в таблице. Провести точечный регрессионный анализ.

температура	117	128	134	136	141	155	159	163	168	172
выход	365	355	342	339	328	315	308	301	287	277

Решение

1) Определяем средние значения x и y.

$$\overline{x} = \frac{117+128+134+136+141+155+159+163+168+172+}{10} = \frac{1473}{10} = 147{,}3$$

$$\overline{y} = \frac{365+355+342+339+328+315+308+301+287+277}{10} = \frac{3217}{10} = 321{,}7$$

2) Пошагово находим коэффициент корреляции. Для удобства расчет составляющих формулы выполним с использованием таблицы:

№ п/п	$x_i - \overline{x}$	$y_i - \overline{y}$	$(x_i - \overline{x})(y_i - \overline{y})$	$(x_i - \overline{x})^2$	$(y_i - \overline{y})^2$
1	-30,3	43,3	-1311,99	918,09	1874,89
2	-19,3	33,3	-642,69	372,49	1108,89
3	-13,3	20,3	-269,99	176,89	412,09
4	-11,3	17,3	-195,49	127,69	299,29
5	-6,3	6,3	-39,69	39,69	39,69
6	7,7	-6,7	-51,59	59,29	44,89
7	11,7	-13,7	-160,29	136,89	187,69
8	15,7	-20,7	-324,99	246,49	428,49
9	20,7	-34,7	-718,29	428,49	1204,09
10	24,7	-44,7	-1104,09	610,09	1998,09
Сумма	-	-	-4819,1	3116,1	7598,1

Рассчитываем коэффициент корреляции

$$r = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2 \sum_{i=1}^{n}(y_i - \overline{y})^2}} = \frac{-4819{,}1}{\sqrt{3116{,}1 \cdot 7598{,}1}} = -0{,}99$$

Полученное значение коэффициента корреляции показывает, что сила связи параметров x и y сильная, а характер связи – обратный (таблица).

4) Определяем значения коэффициентов для линейной аналитической зависимости. Для удобства выполнения расчетов подготовим данные в таблице:

№ п/п	Значения x	Значения y	x^2	$y \cdot x$
1	117	365	13689	42705
2	128-	355	16384	45440
3	134	342	17956	45828
4	136	339	18496	46104
5	141	328	19881	46248
6	155	315	24025	48825
7	159	308	25281	48972
8	163	301	26569	49063
9	168	287	28224	48216
10	172	277	29584	47644
Сумма	1473	3217	220089	469045

5) Подставляем полученные исходные данные в систему уравнений.

$$\begin{cases} na + b\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i \\ a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} y_i x_i \end{cases}$$

$$\begin{cases} 10a + 1473b = 3217 \\ 1473a + 220089b = 469045 \end{cases}$$

6) Выражаем из первого уравнения системы коэффициента

$$10a = 3217 - 1473b$$

$$a = \frac{3217 - 1473b}{10} = 321{,}7 - 147{,}3b$$

7) Полученное выражение подставляем во второе уравнение системы и находим значение коэффициента b.

$$1473\,(321{,}7 - 147{,}3b) + 220089b = 469045$$

$$473864{,}1 - 216972{,}9b + 220089b = 469045$$

$$-216972{,}9b + 220089b = 469045 - 473864{,}1$$

$$3116{,}1b = -4819{,}1$$

$$b = \frac{-4819{,}1}{3116{,}1} \approx -1{,}55$$

8) Определяем величину коэффициента а.

$$10a = 3217 - 1473 \cdot (-1{,}55)$$

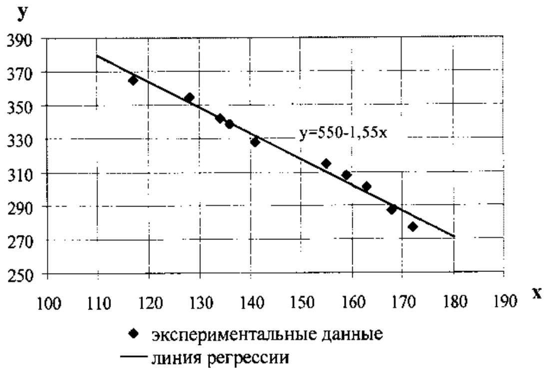
$$10a = 3217 + 2283{,}15$$

$$a = \frac{3217 + 2283{,}15}{10} = \frac{5500{,}15}{10} \approx 550$$

8) Подставляем полученные значения коэффициентов в уравнение регрессии.

$$y_0 = 550 - 1{,}55x.$$

9) Строим корреляционное поле экспериментальных данных и теоретическую линию, полученную по уравнению регрессии:



10) Определяем среднюю ошибку аппроксимации. Расчет произведем пошагово с использованием таблицы:

№ п/п	x	y	$y_0 = a - bx$	$y - y_0$	$\left\lvert \frac{y_i - y_0}{y_i} \right\rvert$
1	117	365	368,7	-3,65	0,0100
2	128	355	351,6	3,4	0,0096
3	134	342	342,3	-0,3	0,0009
4	136	339	339,2	-0,2	0,0006
5	141	328	331,5	-3,45	0,0105
6	155	315	309,8	5,25	0,0167
7	159	308	303,6	4,45	0,0144
8	163	301	297,4	3,65	0,0121
9	168	287	289,6	-2,6	0,0091
10	172	277	283,4	-6,4	0,0231
Сумма	-	-	-	-	0,1070

Рассчитываем среднюю ошибку аппроксимации

$$\varepsilon = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - y_0}{y_i}\right| \cdot 100\% = \frac{0{,}1070}{10} \cdot 100\% = 1{,}07\%$$

Точность аппроксимации уравнения регрессии удовлетворительна, так как ее средняя ошибка не превышает 10 %.

11) Определяем фактическое значение F-критерия Фишера

$$F_{\text{факт}} = \frac{r^2}{1-r^2} \cdot \frac{n-m-1}{m} = \frac{(-0{,}99)^2}{1-(-0{,}99)^2} \cdot \frac{10-1-1}{1} 410{,}39$$

12) Определяем число степеней свободы

$$k_1 = m = 1$$

$$k_2 = n - m - 1 = 10 - 1 - 1 = 8$$

По справочным данным определяем максимально возможное значение критерия

$$F_{\text{табл}} = 238{,}89$$

Уравнение регрессии достоверно описывает экспериментальные данные и является адекватным в заданных пределах варьирования исследуемых факторов, так как фактическое значение F-критерия Фишера значительно превышает табличное

$$F_{\text{факт}} = 410{,}39 \geq F_{\text{табл}} = 238{,}89$$

Задача №6

Рассмотрим химический процесс, в котором выход продукта реакции y (%) зависит от температуры реакционной смеси x_1 (°C) и концентрации реагента x_2 (%). Требуется с помощью полного факторного эксперимента найти математическое описание этого процесса в окрестности точки факторного пространства с координатами x_{01} = 50 °C и x_{02} = 25 %.

Решение

Математическое описание рассматриваемого процесса будем искать в виде уравнения регрессии $Y = b_0 + b_1X_1 + b_2X_2$, где кодированные переменные связаны с температурой и концентрацией следующими соотношениями:

$$X_1 = \frac{x_1 - x_{01}}{\Delta x_1}$$

$$X_{12} = \frac{x_2 - x_{02}}{\Delta x_2}$$

Основные характеристика экспериментов

Характеристика эксперимента	X_1, ^{0}C	X_2, %
Основной уровень	50	25
Интервал варьирования	5	1
Верхний уровень	55	26
Нижний уровень	45	24

Полный двухфакторный эксперимент

Номер опыта	X_1	X_1	x_1, ^{0}C	x_1, %	Y, %
1	-1	-1	45	24	35,5
2	+1	-1	55	24	38,7
3	-1	+1	45	26	32,6
4	+1	+1	55	26	36,2

На основании результатов полного факторного эксперимента рассчитаем коэффициенты регрессии:

$$b_0 = \frac{1}{4}(35{,}5 + 38{,}7 + 32{,}6 + 36{,}2) = 35{,}6$$

$$b_0 = \frac{1}{4}(-35{,}5 + 38{,}7 - 32{,}6 + 36{,}2) = 1{,}95$$

$$b_0 = \frac{1}{4}(-35{,}5 - 38{,}7 + 32{,}6 + 36{,}2) = -1{,}35$$

Будем считать, что оценка дисперсии среднего значения S^2_y равна 0,42. Примем также, что с этой величиной связаны три степени свободы:

$$(f = N(k-1) = 3(2-1) = 3)$$

Ошибку в определении коэффициентов регрессии вычислим по формуле:

$$S_b = \sqrt{\frac{S_y^2}{N}} = \sqrt{\frac{0{,}42}{4}} = 0{,}32$$

Пользуясь справочными материалами, найдем, что для доверительной вероятности Р = 0,95 и трех степеней свободы значение критерия Стьюдента t = 3,18. Тогда S_b t = 0,32 ⦾3,18 = 1,03.

Для оценки значимости коэффициентов регрессии рассмотрим следующие соотношения:

$$|b_0| = 35{,}6 > S_6 \cdot t, |b_1| = |1{,}95 > S_6 \cdot t|, |b_2| = 1{,}35 > S_6 \cdot t$$

Видно, что все коэффициенты регрессии значимы. Следовательно, искомое уравнение имеет вид:

$$y = 35{,}6 + 1{,}95X_1 - 1{,}35X_2$$

Для проверки адекватности уравнения регрессии найдем расчетные значения функции отклика:

$$Y_1^P = 35{,}6 + 1{,}95(-1) - 1{,}35(-1) = 35{,}0$$
$$Y_2^P = 35{,}6 + 1{,}95(+1) - 1{,}35(-1) = 38{,}9$$
$$Y_3^P = 35{,}6 + 1{,}95(-1) - 1{,}35(+1) = 32{,}3$$
$$Y_1^P = 35{,}6 + 1{,}95(+1) - 1{,}35(+1) = 36{,}2$$

По формуле вычислим оценку дисперсии адекватности:

$$S_{\text{ад}}^2 = \frac{1}{N-(k+1)} \sum_{j=1}^{N} (y_j^{\text{э}} - y_j^p)^2$$
$$= \frac{1}{4-3}[(35{,}5 - 35{,}0)^2 + (32{,}6 - 32{,}3)^2 + (36{,}2 - 36{,}2)^2]$$
$$= 0{,}38$$

С ней связано число степеней свободы f = N – (k + 1) = 4 – 3 = 1. Расчетное значение критерия Фишера находим по формуле:

$$F = \frac{S_{\text{ад}}^2}{S^2} = \frac{0,38}{0,42} = 0,905$$

Оно не превосходит значения, приведенного в справочных данных, (Fтаб = 10,13), следовательно, нельзя сказать, что уравнение регрессии неадекватно.

Пример №7

Минимизировать целевую функцию методом градиентного спуска, завершив расчет при

$$\left|\frac{\partial f(x^{(k)})}{\partial x_i}\right| \leq 0,05\,, i = 1,2$$

Решение

Выберем начальное приближение x(0) = (0,0) и величину шага спуска α =1, построим последовательность (с дроблением шага спуска α), записывая результаты вычислений в таблицу.

k	$x_1^{(k)}$	$x_2^{(k)}$	$f(x^{(k)})$	$\frac{\partial f(x^{(k)})}{\partial x_1}$	$\frac{\partial f(x^{(k)})}{\partial x_2}$	α	Примечание
0	0	0	1	1	1	1	
	-1	-1	3.145	-	-		Условие (6.1) нарушено. Уменьшим α в 2 раза
	0	0	1	1	1	0.5	
	-0.5	-0.5	1.118	-	-		Условие (6.1) нарушено. Уменьшим α в 2 раза
	0	0	1	1	1	0.25	
1	-0.25	-0.25	0.794	0.106	-0.393	0.25	Условие (6.1) выполнено.
2	-0.277	-0.152	0.774	0.098	0.045	0.25	Условие (6.1) выполнено.
3	-0.301	-0.163	0.772	0.026	0.023	-	Точность достигнута

Следовательно, получим

$$x^{\circ} \approx x^{(3)} = (-0.301, -0.163)$$

$$f^{\circ} \approx f(-0.301, -0.163) = 0.772$$

Пример №8

Пусть в результате полного факторного эксперимента получено адекватное уравнение регрессии

$$y_1 = 35{,}6 + 1{,}95X_1 - 1{,}35X_2$$

Здесь, y_1 – выход продукта реакции, X_1 – температура, X_2 – концентрация реагента. Введем также в рассмотрение функцию отклика y_2, характеризующую скорость химической реакции (кмоль· м$^{-3}$· ч$^{-1}$). Требуется выполнение условия $y_2 \geq 2{,}5$.

<u>Решение</u>

Допустим, что ограничения на влияющие факторы имеют вид:

$$30° \leq x_1 \leq 120°$$

$$10\% \leq x_2 \leq 70\%$$

Будем оптимизировать выход продукта реакции методом крутого восхождения. качестве базового фактора возьмем температуру и примем шаг движения на крутом восхождении 4°, а Δx=5 (полнофакторный эксперимент), тогда

$$\gamma = \frac{\Delta x_1^*}{b_1 \Delta x_1} = \frac{4}{1{,}95 \cdot 5} = 0{,}41$$

Шаг по концентрации на крутом восхождении можно рассчитать по уравнению:

$$\Delta X_2^* = \gamma b_2 \Delta X_2 = 0{,}41(-1{,}35) \cdot 1 = -0{,}55\%$$

Результаты приведены в таблице:

Результаты опытов по методу крутого восхождения

Характеристика опыта	X_1	X_2	$y_1^э$	$y_2^э$
Центр плана	50	25	35,1	2,9
Интервал варьирования	5	1	-	-
Шаг движения	4	-0,5	-	-
Крутое восхождение				
Номер опыта				
1	54	24,5	36,9	3,2
2	58	24,0	37,2	3,7
3	62	23,5	38,5	2,8
4	66	23,0	40,7	2,3
5	70	22,5	38,1	1,9
6	74	22,0	37,2	1,6

Как видно из таблицы, в опыте № 4 достигнут максимальный выход продукта реакции, однако скорость процесса в этом случае меньше допустимого значения. По-видимому, оптимальным режимом процесса следует считать условия опыта № 3. Ограничения на x_1 и x_2 в ходе оптимизации не нарушены.

ЗАДАЧИ ДЛЯ САМОСТОЯТЕЛЬНОГО РЕШЕНИЯ

1) Десять измерений концентрации соляной кислоты в растворе дали результаты (в мМ) представленные в таблице. Провести обработку серии этих измерений для Рд=0,95

X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}
2,83	2,82	2,81	2,85	2,87	2,86	2,83	2,85	2,83	2,84

2) Приведены измерения оптической плотности раствора, содержащего при различной концентрации. Оптическая плотность измерялась 3 раза при каждой длине волны. Получены следующие результаты:

Концентрация	D1	D2	D3
50	0,22	0,18	0,20
100	0,43	0,39	0,42
150	0,64	0,60	0,61
200	0,85	0,81	0,83
250	0,104	0,103	0,102
300	0,123	0,125	0,121
350	0,143	0,146	0,142

Провести обработку результатов измерений при Рд=0,9. Построить график зависимости оптической плотности от концентрации с указанием доверительных интервалов.

3) Выполнить статистическую обработку результатов измерений температуры в реакторе сернокислого алкилирования, приведенных в таблице:

X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
9, 91	9, 95	9, 89	9,94	9, 96	9,93	9, 94	9,99
X_9	X_{10}	X_{11}	X_{12}				
9,95	9,79	9, 97	9,92				

4) При навешивании образцов сырья для выделения парафинов из тяжелых нефтяных остатков получены образцы с весом, представленным в таблице, какой из образцов нельзя использовать для проведения химического анализа.

X_1	X_2	X_3	X_4	X_5	X_6
1,332	1,345	1,341	1,325	1,296	1,354
X_7	X_8	X_9	X_{10}		
1,349	1,321	1,311	1,304		

5) При измерении pH бидистиллированной воды на открытом воздухе были получены результаты, представленные в таблице. Систематические ошибки в результатах измерений исключены: все результаты одинаковой точности. Необходимо провести выявление грубых ошибок в результатах измерений (P = 0.95); при их выявлении исключить результаты, содержащие грубые ошибки из серии.

X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}
5,28	5,56	5,43	7,0	5,74	5,65	5,44	5,81	5,58	5,49
X_{11}	X_{12}	X_{13}	X_{14}	X_{15}	X_{16}	X_{17}	X_{18}	X_{19}	X_{20}
5,59	4,99	5,67	5,45	6,98	5,05	5,33	5,49	5,76	6,0

6) При измерении температуры в реакторе риформинга получены результаты, представленные в таблице. Систематические ошибки в результатах измерений исключены: все результаты одинаковой точности. Необходимо провести выявление грубых ошибок в результатах измерений (P =0.95); при их выявлении исключить результаты, содержащие грубые ошибки из серии.

X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}
482	502	494	504	486	519	492	499	461	489
X_{11}	X_{12}	X_{13}	X_{14}	X_{15}	X_{16}	X_{17}	X_{18}	X_{19}	X_{20}
559	493	491	481	510	505	490	490	502	491
X_{21}	X_{22}	X_{23}	X_{24}	X_{25}	X_{26}				
476	466	506	514	522	559				

7) Было выполнено 4 измерения концентрации серной кислоты в растворе, полученные результаты приведены в таблице. Достаточно ли количество измерений для определения концентрации серной кислоты? Вычислите необходимое количество измерений.

X_1	X_2	X_3	X_4
25 мМ	75мМ	50мМ	20мМ

8) Было выполнено 10 измерений оптической плотности элюата после хроматографического разделения тяжелых нефтяных остатков, полученные результаты приведены в таблице. Достаточно ли количество измерений для определения концентрации серной кислоты? Вычислите необходимое количество измерений.

X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}
1,21	1,23	1,20	1,24	1,21	1,22	1,23	1,25	1,20	1,21

9) В таблице представлены данные pH исследуемого раствора KOH для точности 0,9, какое количество измерений необходимо провести для получения точности 0,99

X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}
9,34	9,43	9,36	9,38	9,42	9,41	9,35	9,38	9,40	9,32

10) В таблице представлены значения температуры в колонне пиролиза для точности 0,9, какое количество измерений необходимо провести для получения точности 0,99

X_1	X_2	X_3	X_4	X_5
950	960	955	958	956

11) В эксперименте оценивали влияние повышения температуры на скорость протекания реакции. Полученные результаты представлены в таблице. Провести точечный регрессионный анализ.

t	200	250	300	350	400	450	500	550	600
Выход, кг/ч	10	100	1000	2200	4400	8000	9000	9500	10000

12) В эксперименте оценивали влияние добавления дисперсной платины в реакционную среду на скорость протекания реакции. Полученные результаты представлены в таблице. Провести точечный регрессионный анализ.

Pt	10	20	30	40	50	60	70	80	90
Выход, кг/ч	1130	2600	4360	5200	6700	8000	9120	9500	10000

13) В эксперименте оценивали влияние щавелевоуксусной кислоты на флюоресценцию радахлорина. Полученные результаты представлены в таблице. Провести точечный регрессионный анализ.

ЩУК, мкМ	100	200	300	400	500	600	700	800	900
D	3,5	3,2	3,1	2,9	2,5	2,3	2,0	1,7	1,3

14) В эксперименте оценивали влияние изменения давления на скорость протекания реакции изомеризации. Полученные результаты представлены в таблице. Провести точечный регрессионный анализ.

P, атм	1,0	1,3	1,6	1,9	2,1	2,4	2,7	3,0	3,3
Выход, кг/ч	1010	980	1000	993	959	1005	967	979	990

15) Изучается напряжение при удлинении 300% (функция отклика y) типовой протекторной резины на основе 70% СКД и 30% СКИ-3 в зависимости от содержания (в вес. ч.) трех компонентов: серы (z1), технического углерода (z2) и пластификатора (z3). При исследовании влиянии серы, технического углерода и пластификатора были выбраны пределы измерения дозировок: для серы - 1,1 -2,5 вес.ч., для технического углерода - 45 – 65 вес.ч., для пластификатора ПН6 - 2-16 вес.ч. Получить уравнение регрессиии при планировании эксперимента для труех факторов.

16) На основании результатов ПФЭ, приведенных в таблице, найти линейную зависимость мощности смесителя N от массы смеси $m_{см}$ и прочности смеси $\sigma_{см}$:

Номер опыта	Исходные значения		Кодовые значения		y: N, Вт			$\overline{N}$ Вт
	$m_{см}$, кг	$\sigma_{см}$, кПа	X_1	X_2				
1	3	60	-	-	205	210	215	210,000
2	5	100	+	+	375	385	390	383,333
3	5	60	+	-	230	235	245	236,667
4	3	100	-	+	310	315	325	316,667

Факторы в эксперименте принимали следующие значения

$$x_1: m_{см}=(4\pm1) \text{ кг};$$

$$x_2: \sigma_{см}=(80\pm20) \text{ кПа}$$

17) Минимизировать целевую функцию, иные условия из примера 1:

$$f(x) = f(x_1, x_2) = x_1^2 + 2x_2^2 + e^{x_1+x_2}, x^{(0)} = (1,1), \alpha = 0,1;$$

18) Методом градиентного спуска найти оптимум функции:

$$f(x) = f(x_1, x_2) = e^{x_1^2} + (x_1 + x_2 + x_3)^2, x^{(0)} = (1, 1, 1), \alpha = 0,1;$$

19) Методом крутого восхождения найти оптимальное значение для полнофакторного эксперимента, описываемого уравнением регрессии $y_1 = 40 + 2X_1 - 1.78X_2$, условие максимального y_1, при $X_1 < 70$.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Ренкин Н.М.* Методы обработки результатов химического эксперимента: учеб. пособие. Самара: Самар. гос. техн. ун-т, 2012. – 107 с.

2. *Пономарев В.Б., Лошкарев А.Б.* Математическое моделирование технологических процессов. Екатеринбург: УГТУ, 2006. – 129 с.

3. *Чарыкин А.К.* Математическая обработка результатов химического анализа. Л.: Химия, 1984. – 168 с.

4. *Пономарев В.Б., Лошкарев А.Б.* Математическая обработка результатов инженерного эксперимента. Екатеринбург: УрФУ, 2016. – 100 с.

5. *Ахназарова С.Л., Кафаров В.В.* Методы оптимизации эксперимента в химической технологии. М.: Высшая школа, 1985. - 327 с.

6. *Фаддеев М.А.* Элементарная обработка результатов эксперимента: Учеб. пособие. Н. Новгород: Изд-во Нижегородского госуниверситета, 2002. – 108 с.

7. *Спиридонов В.П.* Математическая обработка физико-химических данных. М.: Изд-во Московского университета, 1970. – 222 с.

8. *Кобзарь А.И.* Прикладная математическая статистика. Для инженеров и научных работников. - М: ФИЗМАТЛИТ, 2006. – 816 с.

9. *Адлер Ю.П., Маркова Е.В., Грановский Ю.В.* Планирование эксперимента при поиске оптимальных условий. М.: Наука, 1976. - 278 с.

10. ГОСТ 8.207-76. Прямые измерения с многократными наблюдениями. Методы обработки результатов измерений.

11. *Бондарь А.Г., Статюха Г.А., Потяженко И.А.* Планирование эксперимента при оптимизации процессов химической технологии. Киев: Высшая школа, 1980. 264с.

12. *Саутин С.Н.* Планирование эксперимента в химии и химической технологии. Л.: Химия, 1975.

13. *Рузинов Л.П.* Статистические методы оптимизации химических процессов. М.: Химия, 1972. 200 с.
14. *Хартман К.* и др. Планирование эксперимента в исследовании технологических процессов. М.: Мир, 1977. 520 с.
15. *Гмурман В.Е.* Теория вероятностей и математическая статистика. М.: Высшая школа, 1999. 325 с.

СОДЕРЖАНИЕ

ОБРАБОТКА И ПЛАНИРОВАНИЕ ЭКСПЕРИМЕНТА

Методические указания к практическим занятиям для студентов магистратуры направления 18.04.01

Сост. *Е.В. Саламатова*

Печатается с оригинал-макета, подготовленного кафедрой химических технологий и переработки энергоносителей

Ответственный за выпуск *Е.В. Саламатова*

Лицензия ИД № 06517 от 09.01.2002

Подписано к печати 29.04.2021. Формат 60×84/16.
Усл. печ. л. 2,6. Усл.кр.-отт. 2,6. Уч.-изд.л. 2,3. Тираж 75 экз. Заказ 377.

Санкт-Петербургский горный университет
РИЦ Санкт-Петербургского горного университета
Адрес университета и РИЦ: 199106 Санкт-Петербург, 21-я линия, 2