

ФЕДЕРАЛЬНОЕ АГЕНТСТВО ПО ОБРАЗОВАНИЮ
ВОЛГОГРАДСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ
ЛАБОРАТОРИЯ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ ХТФ
КАФЕДРА ХИМИИ И ТЕХНОЛОГИИ ПЕРЕРАБОТКИ ЭЛАСТОМЕРОВ

А.Н. Гайдадин, С.А. Ефремова, Н.Н.Печурина

МОДЕЛИРОВАНИЕ
ТЕХНОЛОГИЧЕСКИХ ПРОЦЕССОВ С ПОМОЩЬЮ
МЕТОДА НАИМЕНЬШИХ КВАДРАТОВ

Методические указания



Волгоград
2008

УДК 678.04

Рецензент

профессор кафедры «Промышленная экология и безопасность жизнедеятельности»
А.Б. Голованчиков

Издается по решению редакционно-издательского совета
Волгоградского государственного технического университета

Моделирование технологических процессов с помощью метода наименьших квадратов: метод. указания / сост. А.Н.Гайдадин, С.А.Ефремова, Н.Н.Печурина; ВолгГТУ. – Волгоград, 2008. – 16 с.

В методических указаниях описана процедура проведения подбора типа математической модели технологических процессов и определения ее коэффициентов с помощью метода наименьших квадратов и метода относительных наименьших квадратов. Для студентов по направлениям 240100 «Химическая технология и биотехнология», 260100 «Технология продуктов питания», специальности 240502 «Технология переработки пластических масс и эластомеров», а также для студентов, обучающихся по магистерским программам 240115 «Технология переработки эластомеров» и 240110 «Химическая технология высокомолекулярных соединений».

© Волгоградский государственный
технический университет, 2008

ВВЕДЕНИЕ

Методы математического моделирования позволяют проводить прогноз свойств и оптимизацию показателей технических объектов с меньшими затратами, чем проведение эксперимента. К сожалению, в химической и пищевой промышленности количество математических моделей крайне мало. В результате при обработке экспериментальных данных возникает необходимость получения математических моделей с помощью доступных способов. Одним из них является метод наименьших квадратов. При использовании этого метода задачей исследователя является получение и анализ математической модели. В результате разработанная эмпирическая математическая модель может быть использована для оценки механизмов процессов, протекающих внутри объекта

1. Цель и задачи лабораторной работы

Целью лабораторной работы является ознакомление с методикой подбора типа математической модели технологического процесса с помощью метода наименьших квадратов. Приобретение студентами навыков выбора типа уравнения, приведение его к линейному типу, определение значения коэффициентов, обоснование значений статистических критериев и определения адекватности исследуемой модели.

В ходе лабораторной работы студент должен овладеть теоретическими подходами к использованию метода наименьших квадратов (МНК) или метода относительных наименьших квадратов (МОНК), способами обоснования выбора метода МНК или МОНК в конкретной технологической ситуации, использование методик для линейных и нелинейных моделей. Студент должен обоснованно выбирать значения статистических критериев и определять адекватность полученной модели.

2. Теоретические основы

В практике научных исследований наряду с физическими моделями все большее распространение получают математические модели, в качестве которых часто используются рассчитанные по результатам эксперимента уравнения регрессии. Статистические математические модели получают, описывая зависимости выходных параметров Y (свойств, откликов) объекта от изменения входных параметров X (факторов) с помощью различных функций:

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

Одно из основных требований к математической модели объекта - это точность описания (предсказывания) поведения реального объекта при изменении условий.

Объект исследований рассматривается в качестве системы «черный ящик» (рис. 1) [1]:

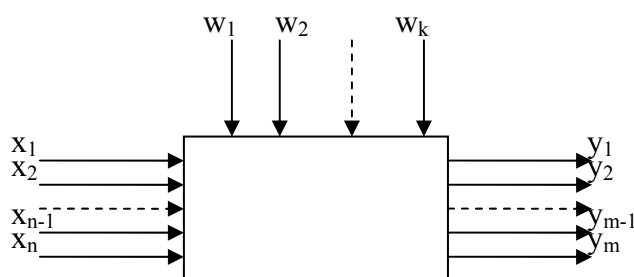


Рисунок 1 – Система «Черный ящик»

Суть данного принципа состоит в изучении зависимости отклика системы $Y(y_1, y_2, \dots, y_m)$ на изменение входных измеряемых и регулируемых параметров $X(x_1, x_2, \dots, x_n)$ при действии случайных факторов $W(w_1, w_2, \dots, w_k)$ («шум» объекта). Комплекс параметров X называют основным, он определяет условия эксперимента. Выходными параметрами Y могут являться любые технологические или технические показатели исследуемого процесса. В качестве случайных параметров W рассматриваются параметры, которые по различным причинам трудно ли нет необходимости учитывать. Случайным будет считаться любой фактор, не вошедший в основной комплекс

входных параметров.

Уравнение приближенной регрессии существенно зависит от выбираемого метода приближения. В качестве такого метода часто выбирают метод наименьших квадратов. По методу наименьших квадратов можно обрабатывать любые экспериментальные данные, однако оптимальность этой процедуры доказывается только для статистически нормального распределения.

Пусть задан некоторый класс функций $f(x)$, накладывающий на выборку одинаковое число связей l . Число связей l равно числу неопределенных коэффициентов, входящих в аналитическое выражение этой функции. Чаще всего используют многочлены различной степени. Наилучшее уравнение приближенной регрессии дает та функция из рассматриваемого класса, для которой сумма квадратов имеет наименьшее значение, другими словами[2]:

$$\Phi = \sum_{i=1}^n [y_i - f(x_i)]^2 \rightarrow \min. \quad (1)$$

Задача определения параметров уравнения регрессии сводится практически к определению минимума функции многих переменных.

Процедура метода наименьших квадратов позволяет с высокой точностью определить значения коэффициентов линейного уравнения вида:

$$y = a + b \cdot x, \quad (2)$$

где значения a и b представляют собой искомые коэффициенты уравнения.

Соотношение (1) с учетом (2) можно записать иначе:

$$\Phi = \sum_{i=1}^n [y_i - a_0 - b \cdot x_i]^2 \rightarrow \min \quad (3)$$

Минимум функции достигается при одновременном равенстве нулю частных производных по всем неизвестным, т.е. [3]:

$$\frac{\partial \Phi}{\partial a} = -2 \sum_{i=1}^n [y_i - a - b \cdot x_i] = 0 \quad (4)$$

$$\frac{\partial \Phi}{\partial b} = -2 \sum_{i=1}^n [y_i - a - b \cdot x_i]^2 x_i = 0$$

Решение системы уравнения (4) дает единственное решение:

$$a = \frac{\sum_{i=1}^n y_i \sum_{i=1}^n x_i^2 - \sum_{i=1}^n (x_i y_i) \sum_{i=1}^n x_i}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2},$$

$$b = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2}. \quad (5)$$

На рисунке 2 показаны эмпирические данные (таблица 1), полученные при исследовании влияния количества вносимого обезжиренного молока на степень выделения белка, и уравнение линейной регрессии, коэффициенты которой рассчитаны с помощью МНК по формулам (5).

Таблица 1 - Экспериментальные данные, полученные при исследовании влияния количества вносимого обезжиренного молока (X) на степень выделения белка (Y) [4].

Входящий фактор, X	Значения параллельных опытов выходящего параметра					Среднее значение Y _{ср}
	Y ₁	Y ₂	Y ₃	Y ₄	Y ₅	
10	17,07188	15,26061	15,71343	18,27939	17,52469	16,77
20	20,67188	18,86061	19,31343	21,87939	21,12469	20,37
30	22,87188	21,06061	21,51343	24,07939	23,32469	22,57
40	26,37188	24,56061	25,01343	27,57939	26,82469	26,07
50	29,47188	27,66061	28,11343	30,67939	29,92469	29,17
60	35,06188	33,25061	33,70343	36,26939	35,51469	34,76
70	37,66188	35,85061	36,30343	38,86939	38,11469	37,36
80	40,36188	38,55061	39,00343	41,56939	40,81469	40,06
90	44,86188	43,05061	43,50343	46,06939	45,31469	44,56

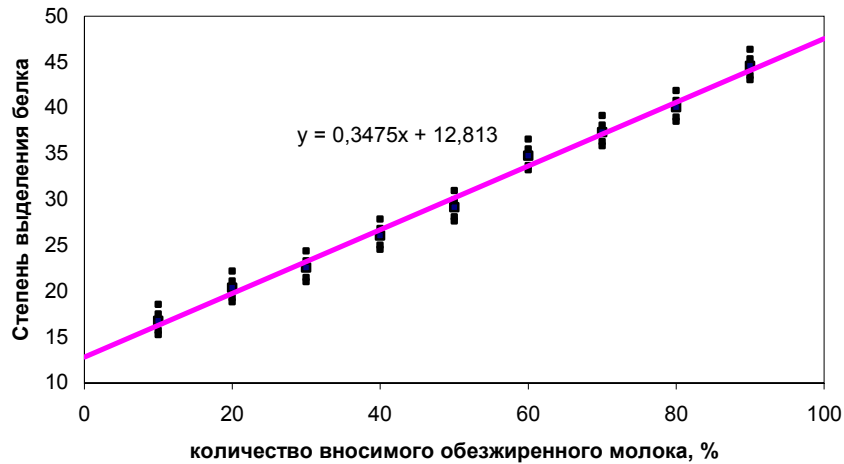


Рисунок 2 – Эмпирические данные и расчетная линия регрессии.

Расчет коэффициентов a и b с помощью метода относительных наименьших квадратов (МОНК) проводится по формулам:

$$a = \frac{\sum_{i=1}^n \left(\frac{1}{y_i}\right) \sum_{i=1}^n \left(\frac{x_i}{y_i}\right)^2 - \sum_{i=1}^n \left(\frac{x_i}{y_i}\right) \sum_{i=1}^n \left(\frac{x_i}{y_i^2}\right)}{\sum_{i=1}^n \left(\frac{1}{y_i}\right)^2 \sum_{i=1}^n \left(\frac{x_i}{y_i}\right)^2 - \left[\sum_{i=1}^n \left(\frac{x_i}{y_i^2}\right)\right]^2}, \quad (6)$$

$$b = \frac{\sum_{i=1}^n \left(\frac{1}{y_i}\right)^2 \sum_{i=1}^n \left(\frac{x_i}{y_i}\right) - \sum_{i=1}^n \left(\frac{x_i}{y_i}\right) \sum_{i=1}^n \left(\frac{x_i}{y_i^2}\right)}{\sum_{i=1}^n \left(\frac{1}{y_i}\right)^2 \sum_{i=1}^n \left(\frac{x_i}{y_i}\right)^2 - \left[\sum_{i=1}^n \left(\frac{x_i}{y_i^2}\right)\right]^2} \quad (7)$$

После того как уравнение регрессии найдено, необходимо провести статистический анализ результатов. Этот анализ заключается в проверке значимости всех коэффициентов регрессии в сравнении с ошибкой воспроизводимости и адекватности уравнения. Такое исследование называется регрессионным анализом.

Коэффициенты уравнения регрессии связаны друг с другом.

Для определения адекватности выбранной модели используем критерий Фишера [5]:

$$F_p = S_{ad}^2 / S_{воспр}^2, \quad (8)$$

где $S_{ад}$ – дисперсия адекватности; $S_{воспр}$ – дисперсия воспроизводимости.

В случае, когда оценка параллельных опытов проводится только в одной экспериментальной точке, дисперсия адекватности и дисперсия воспроизводимости рассчитываются по формулам:

$$S^2_{ad} = \frac{\sum_{i=1}^n (y_{cp} - y_{pac})^2}{n - l},$$

$$S^2_{воспр} = \frac{\sum_{j=1}^m (y_j - y_{cp})^2}{(m - 1)}$$
(9)

где n – объем выборки, l – число коэффициентов в уравнении регрессии (в данном случае $l=2$), y_{pac} – значение функции (2) в данной экспериментальной точке x , y_j – значение выходного параметра при j - параллельном при данном x , y_{cp} – среднее значение y , полученное по формуле:

$$y_{cp} = \frac{\sum_{j=1}^m y_j}{m},$$

где m – число параллельных опытов.

В случае, когда расчет дисперсии воспроизводимости проводится с учетом параллельных опытов в каждой экспериментальной точке формулы (9) записываются в следующем виде:

$$S^2_{ad} = \frac{m \sum_{i=1}^n (y_{cp} - y_{pac})^2}{n - l},$$

$$S^2_{воспр} = \frac{\sum_{i=1}^n \sum_{j=1}^m (y_{ij} - y_{cp})^2}{n(m - 1)}$$
(10)

где y_{pac} – значение функции (2) при x_i , y_{ij} – значение выходного параметра при j - параллельном опыте при x_i , y_{cp} – среднее значение y при x_i .

Если расчетное значение критерия Фишера окажется меньше табличного значения $F_{1-p}(f1, f2)$ для уровня значимости p и чисел степеней свободы $f1=f_{ad}=n-l$ и $f2=f_{воспр}=n(m-1)$, то уравнение адекватно эксперименту.

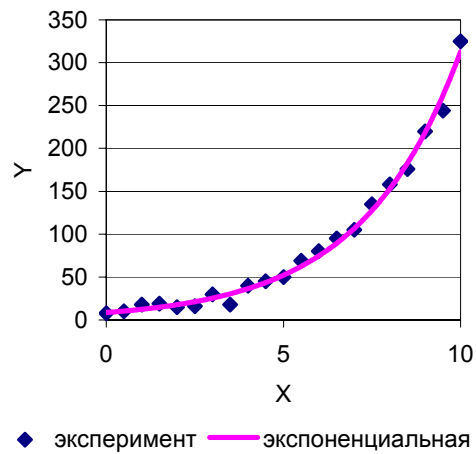
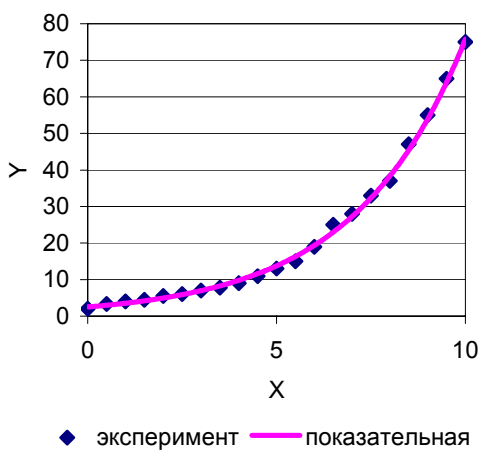
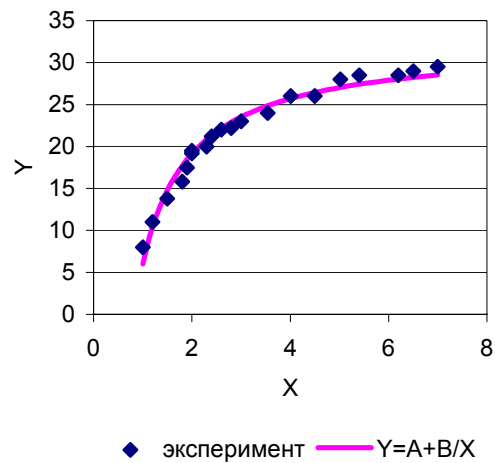
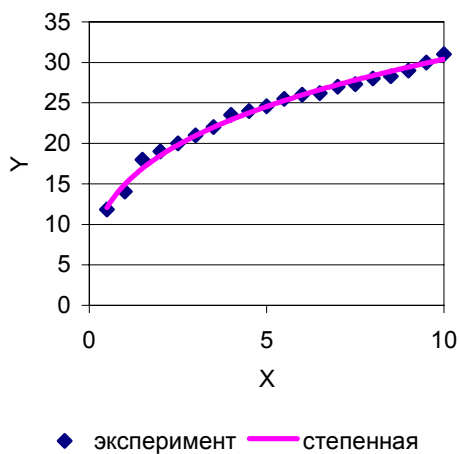
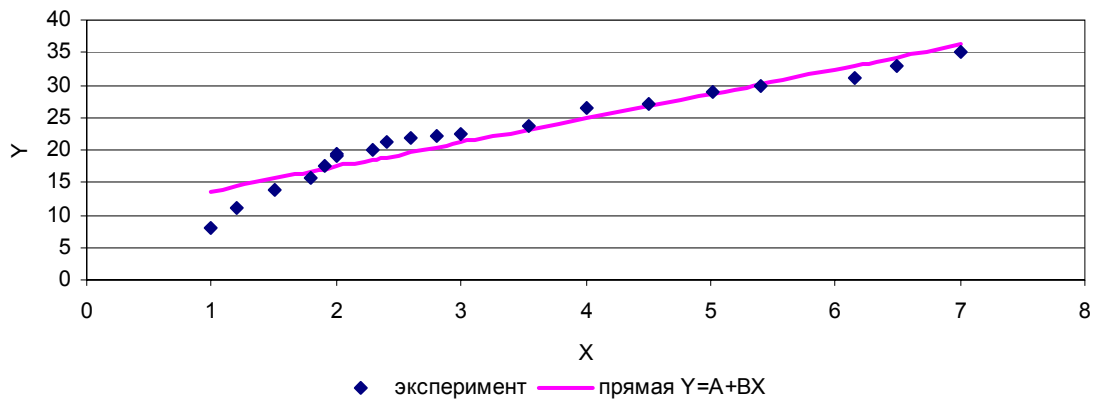


Рисунок 2 – Примеры функциональных зависимостей

К сожалению, линейная модель не всегда наилучшим образом подходит к рассматриваемому распределению. Наиболее часто встречающиеся на практике эмпирические зависимости (рис.2):

1) линейная функция: $y = a + bx$ (например, при равномерном движении тел зависимость между расстоянием S и временем t задается формулой

$S=v_0t+S_0$, где v_0 – первоначальная скорость движения тела, S_0 – первоначальное положение тела);

2) степенная функция: $y = ax^b$ (например, уравнение скорости реакции при постоянной концентрации [B]: $v = k_{эфф}[A]^{V_A}$, где V_A - стехиометрический коэффициент реагирующего вещества А, $k_{эфф}$ - эффективная константа скорости, $k_{эфф} = k[B]^{V_B}$).

3) показательная функция: $y = ab^x$ (например, используется при описании процессов ядерной физики: общая формула для процесса распада радиоактивного вещества: $m = m_0(1/2)^{-t/t_0}$, где m_0 - первоначальная масса вещества, t_0 - период полураспада);

4) экспоненциальная $y = a \cdot e^{b \cdot x}$ (например, уравнение Аррениуса – температурная зависимость константы скорости $k = A \cdot \exp(-E_a / RT)$, где A – предэкспоненциальный множитель, E_a – энергия активации, R – газовая постоянная)

Приведем примеры приведения уравнения нелинейного вида к уравнению прямой.

- Уравнение вида $y=a+b/x^2$ нелинейно. Приведение осуществляется за счет замены аргумента $(1/x^2)$ на значение $x1$. В случае для реализации методов МНК получено линейное уравнение $y=a+b \cdot x1$, корни которого находят по выше приведенным формулам.

Для расчета удобно воспользоваться таблицей:

Таблица 2 – Расчет коэффициентов регрессии $y=a+b/x^2$

$N\varnothing$ n/n	$x_{эксн}$	$y_{эксн}$	$x1=1/x_{ксп}^2$	$x1^2$	$x1 \cdot y_{эксн}$
1	x_1	y_1	$x1_1=1/x_{1эксн}^2$	$x1_1^2$	$x1_1 \cdot y_1$
...
i	x_i	y_i	$x1_i=1/x_{iэксн}^2$	$x1_i^2$	$x1_i \cdot y_i$
...
n	x_n	y_n	$x1_n=1/x_{nэксн}^2$	$x1_n^2$	$x1_n \cdot y_n$
		$\Sigma y_{эксн}$	$\Sigma x1$	$\Sigma x1^2$	$\Sigma (x1 \cdot y)$

Коэффициенты регрессии рассчитаем по преобразованным формулам (5):

$$a = \frac{\sum_{i=1}^n y_i \sum_{i=1}^n x_i^2 - \sum_{i=1}^n (x_i y_i) \sum_{i=1}^n x_i}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} \quad b = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2}$$

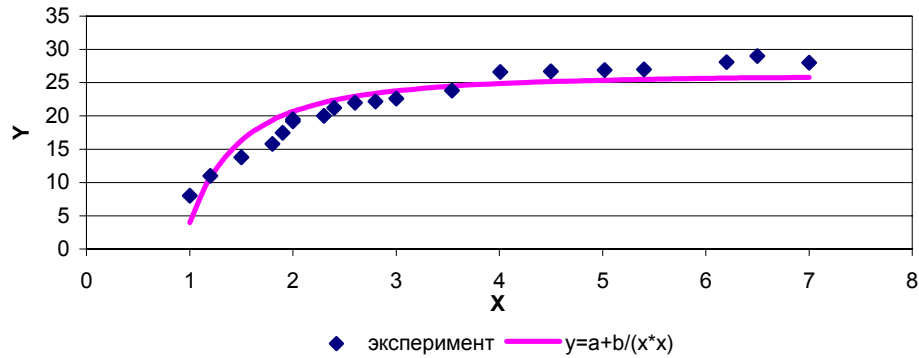


Рисунок 3 – График зависимости $y=a+b/x^2$, полученной с помощью МНК.

- Уравнение вида $y=a \cdot \exp(-b \cdot x)$ не линейно. Приведение осуществляется в две стадии. На первой стадии проводится логарифмирование левой и правой части уравнения. В результате получаем $\ln(y)=\ln(a)-b \cdot x$. При замене $y_l=\ln(y)$, $a_l=\ln(a)$, $b_l=-b$ получаем линейное уравнение $y_l=a_l+b_l \cdot x$.

Для расчета удобно воспользоваться таблице:

Таблица 3 – Расчет коэффициентов регрессии $y=a \cdot \exp(-b \cdot x)$

$\frac{№}{n/n}$	$x_{эксн}$	$y_{эксн}$	$y_l=\ln(y)$	$x_{эксн}^2$	$x \cdot y_l$
1	x_1	y_1	$y_{l1}=\ln(y_1)$	x_1^2	$x_1 \cdot y_{l1}$
...
i	x_i	y_i	$y_{li}=\ln(y_i)$	x_i^2	$x_i \cdot y_{li}$
...
n	x_n	y_n	$y_{ln}=\ln(y_n)$	x_n^2	$x_n \cdot y_{ln}$
	$\sum x_{эксн}$		$\sum y_l$	$\sum x_{эксн}^2$	$\sum (x \cdot y_l)$

Коэффициенты регрессии рассчитаем по преобразованным формулам (5):

$$a1 = \frac{\sum_{i=1}^n y1_i \sum_{i=1}^n x_i^2 - \sum_{i=1}^n (x_i y1_i) \sum_{i=1}^n x_i}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} \quad b1 = \frac{n \sum_{i=1}^n x_i y1_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2}$$

Для использования первоначального уравнения приведенные коэффициенты преобразуются:

$$a = \exp(a1); \quad b = -b1;$$

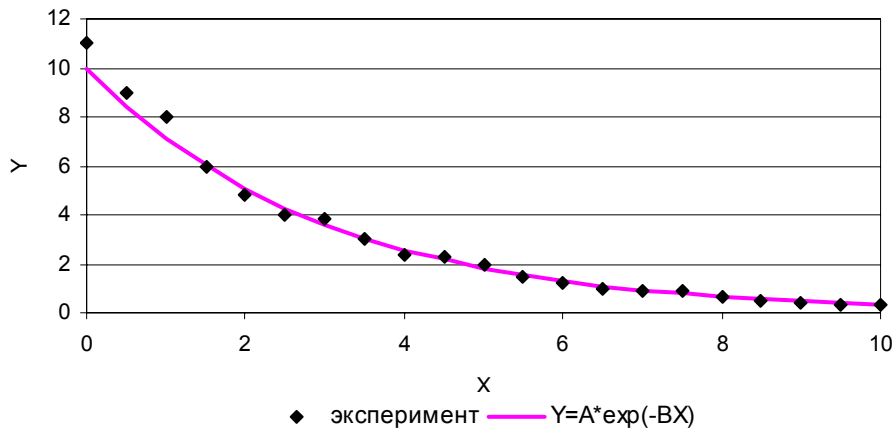


Рисунок 3 - График зависимости $y = a \cdot \exp(b \cdot x)$, полученной с помощью МНК.

С помощью метода наименьших квадратов можно обрабатывать любые экспериментальные данные, однако, надо отметить, что оптимальность данного метода доказывается только для нормального распределения [6].

К достоинствам методов МНК и МОНК можно отнести следующее: позволяют получать коэффициенты приближенной регрессии, позволяют работать с нелинейными моделями, минимизируют ошибку. Наряду с этим надо отметить и недостатки, а именно: не все модели могут быть приведены к линейному виду, методы дают большую погрешность, когда функция отклика существенно изменяется в выбранном интервале.

3. Порядок выполнения работы

1. Подготовить экспериментальные данные и получить допуск на проведение лабораторной работы у преподавателя.

2. Произвести оценку полученных данных, определить однородность и нормальность распределения выборки Y_n . Обосновать использование методов МНК и МОНК.

3. По экспериментальным данным построить график исследуемой зависимости. По типу графика выбрать уравнения, предположительно описывающие экспериментальную кривую. При необходимости привести уравнение к линейному виду.

4. По полученным данным с использованием МНК либо МОНК заполнить таблицу 4.

Таблица 4 - Натуральные и приведенные значения показателей.

№ п/п	Значения аргумента	Значения функции	Приведенные значения аргумента	Приведенные значения функции
1	2	3	4	5

5. Провести расчет значений сумм аргументов и функций для МНК либо МОНК и заполнить таблицу 5.

Таблица 5 - Значения сумм функций и аргументов.

№ п/п	Для метода наименьших квадратов				Для метода относительных наименьших квадратов				
	x	x^2	y	$x \cdot y$	$1/y$	(x/y^2)	$(x/y)^2$	(x/y)	$(1/y)^2$
1	2	3	4	5	6	7	8	9	10
l	x_l	x_l^2	y_l	y_l^2	$1/y_l$	x_l/y_l^2	$(x_l/y_l)^2$	x_l/y_l	$(1/y_l)^2$
...
i	x_i	x_i^2	y_i	y_i^2	$1/y_i$	x_i/y_i^2	$(x_i/y_i)^2$	x_i/y_i	$(1/y_i)^2$
...
n	x_n	x_n^2	y_n	y_n^2	$1/y_n$	x_n/y_n^2	$(x_n/y_n)^2$	x_n/y_n	$(1/y_n)^2$
	Σx_i	Σx_i^2	Σy_i	Σy_i^2	$\Sigma(1/y_i)$	$\Sigma(x_i/y_i^2)$	$\Sigma(x_i/y_i)^2$	$\Sigma(x_i/y_i)$	$\Sigma(1/y_i)^2$

6. Рассчитать значения коэффициентов линейного уравнения по методикам МНК либо МОНК.

7. В случае использования приведенных значений, преобразовать полученные коэффициенты в коэффициенты первоначального уравнения.

8. Определить расчетные значения исследуемой функции и заполнить таблицу 5. Сделать выводы о соответствии выбранной модели изучаемому процессу.

Таблица 5 - Сравнение экспериментальных и расчетных значений.

№ п/п	Экспериментальные значения функции	Расчетные значения функции	Относительное отклонение, %
1	2	3	4

9. Определить значения критерия Фишера для исследуемого процесса. Оценить адекватность математической модели.

4. Контрольные вопросы.

1. Суть принципа системы «Черный ящик». Определение объекта исследования.
2. Сущность метода наименьших квадратов.
3. Терминология, алгоритм и особенности применения МНК и МОНК в химической технологии и биотехнологии.
4. Сущность и назначение приведения уравнения к линейному виду. Понятие и особенности использования приведенных коэффициентов.
5. Назначение статистических критериев, обоснование величины степени надежности, оценка адекватности математической модели.
6. Обоснование отличий в использовании МНК и МОНК при моделировании технических объектов.
7. Укажите достоинства методов МНК и МОНК.
8. Укажите недостатки методов МНК и МОНК.

5.Список рекомендуемой литературы

1. Ахназарова С.Л., Кафаров В.В. Методы оптимизации эксперимента химической технологии: Учеб. пособие для хим. – технол. спец. вузов. – 2-е изд., перераб. и доп. – М.: Высш. шк., 1985.-327 с..
2. Зельдович Я. Б., Мышкис А. Д. Элементы прикладной математики. - М.: Наука, 1972г. –592с.
3. Адлер Ю.П., Маркова Е.В., Грановский Ю.В. Планирование эксперимента при поиске оптимальных условий. – М.Наука, 1976 – 296 с.
4. Львовский Е.Н. Статистические методы построения эмпирических формул: Учеб. пособие для вузов – 2-е изд., перераб. и доп. – М.: Высш. школа, 1988. – 239 стр.
5. Храпцов А.Г., Евдокимов И.А., Костина В.В., Федорова Л.А. Математическое моделирование процесса совместной коагуляции белков подсырной сыворотки и обезжиренного молока. Сборник научных трудов Северо-Кавказского государственного технического университета. Серия «продовольствие». Выпуск 5. Ставрополь: ГОУВПО «СевКавГТУ», 2002. С.20-22.
6. Математическая статистика: Учеб. для вузов/В.Б.Гориянов, И.В. Павлов, Г.М. Цветкова и др.; Под ред. В.С. Зарубина, А.П. Крищенко. – М.: Изд-во МГТУ им. Н.Э.Баумана, 2001. – 424 с.

Алексей Николаевич **Гайдадин**
Светлана Анатольевна **Ефремова**
Наталья Николаевна **Печурин**

МОДЕЛИРОВАНИЕ ТЕХНОЛОГИЧЕСКИХ ПРОЦЕССОВ С ПОМОЩЬЮ МЕТОДА
НАИМЕНЬШИХ КВАДРАТОВ
Методические указания к лабораторной работе

Редактор *Л. Н. Рыжих*

Темплан выпуска электронных изданий 2008 г., поз. №53 .

На магнитоносителе. Уч.-изд. л. 1,0.

Подписано на «Выпуск в свет» 16.04.2008 г. Заказ № .

Волгоградский государственный технический университет.
400131, г. Волгоград, пр. им. В. И. Ленина, 28.

РИО РПК «Политехник»
Волгоградского государственного технического университета.
400131, г. Волгоград, ул. Советская, 35.