Определение коэффициентов однопараметрических моделей технологических процессов статистическими методами

(с помощью Matlab)

ЗАДАНИЕ.

Цель работы

1) изучить методику обработки экспериментальных данных и получить параметры модели технологического процесса методом наименьших квадратов (МНК);
2) проверить адекватность полученной модели по критерию Фишера.

В данной работе по экспериментальным данным, полученным на промышленных технологических установках ректификации, дегидрирования углеводов, сушки и т. д., необходимо выработать вид однопараметрической линии регрессии, определить параметры и оценить адекватность предложенной регрессионной модели реальному процессу.

РЕШЕНИЕ.

Алгоритм решения

Построение графика по исходным данным. Выбор уравнения регрессии

На основании исходных данных табл. 1 построим эмпирическую линию регрессии y по x (рис. 1).

Таблица 1

Вариант задания к выполнению работы

№	1	2	3	4	5	6	7	8	9	10
x	1	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9
y	12.95	14.95	16.78	18.46	20.02	21.47	22.84	24.15	25.38	26.43
№	11	12	13	14	15	16	17	18	19	20
x	2	2.1	2.2	2.3	2.4	2.5	2.6	2.7	2.8	2.9
y	27.59	28.62	29.6	30.46	31.38	32.27	33.08	33.08	33.91	34.58

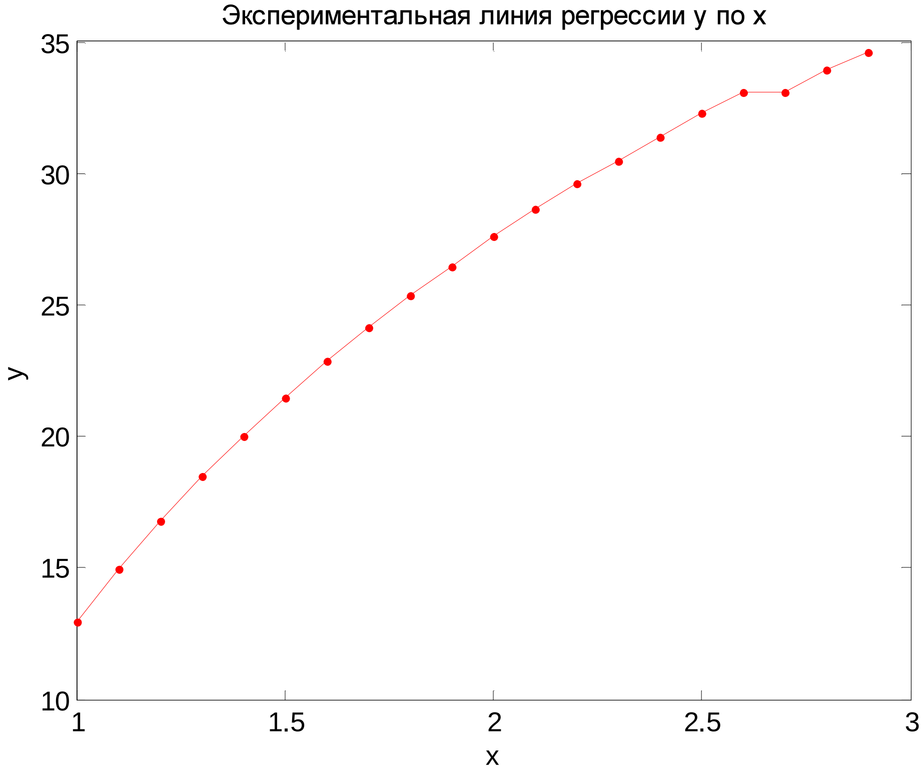
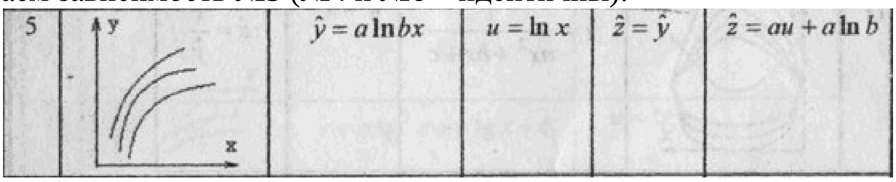


Рисунок 1. Эмпирическая линия регрессии y по x

По виду построенной линии подберем уравнение регрессии, пользуясь табл. методических указаний (выбираем из №4, №5, №6, №14, №15).
Выбираем зависимость №5 (№4 и №6 – идентичны):



5		$\hat{y} = a\ln bx$	$u = \ln x$	$\hat{z} = \hat{y}$	$\hat{z} = au + a\ln b$

Проверим полученные предположения, построив эмпирическую линию регрессии y по $\ln x$ (рис. 2).
На рис. 2 отображена зависимость, похожая на линейную (не на квадратичную), что подтверждает правильность нашего выбора.

Таким образом, выбранный вид зависимости:

$$\hat{z} = a \cdot u + a\ln b = A \cdot u + B, \tag{1}$$

где $A = a$, $B = a\ln b$. Следовательно, $b = e^{\frac{B}{a}}$.

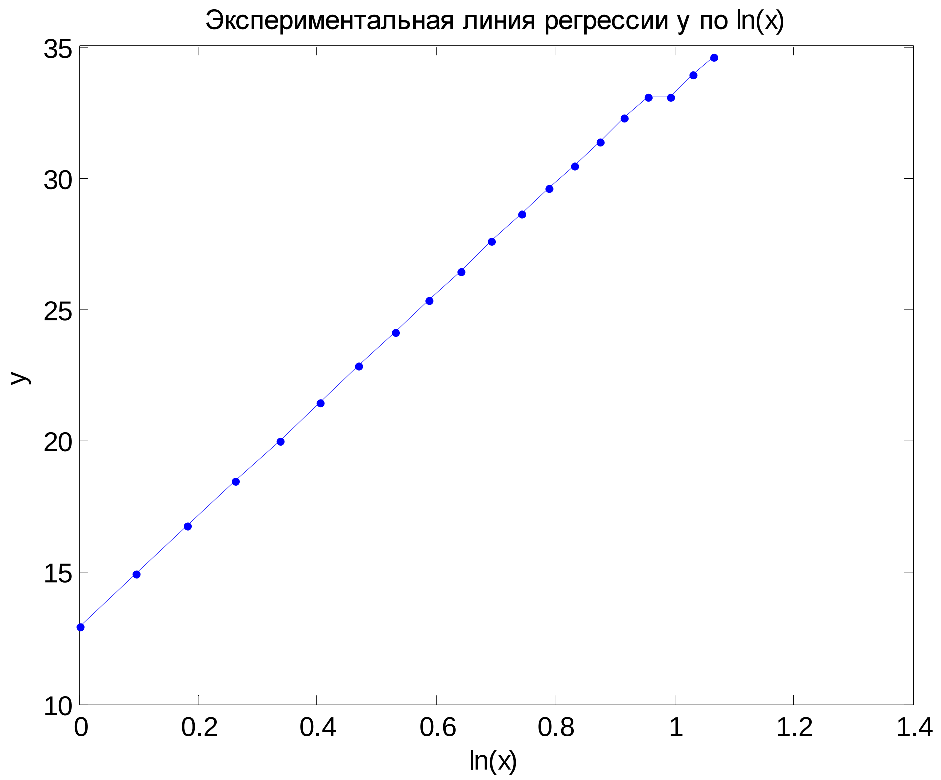


Рисунок 2. Эмпирическая линия регрессии y по $\ln x$

Расчет коэффициентов регрессионной модели

Для определения параметров A, B линейной регрессии (1)

$$\hat{z} = A \cdot u + B, \qquad (2)$$

используя набор экспериментальных данных по фактору u и отклику z, проводится расчет коэффициентов из условия:

$$\Phi = \sum_{i=1}^{N} (z_i - \hat{z}_i)^2 \xrightarrow[A,\,B]{} \min, \qquad (3)$$

где $N = 20$ – объем выборки.

Исходя из условия существования экстремума функции нескольких переменных, необходимым условием минимума $\Phi(A, B)$ является выполнение неравенств:

$$\frac{\partial \Phi}{\partial A} = 0; \quad \frac{\partial \Phi}{\partial B} = 0. \qquad (4)$$

Таким образом, для случая (2) требуется решить систему уравнений:

$$\begin{cases} A\sum_{i=1}^{N} u_i^2 + B\sum_{i=1}^{N} u_i = \sum_{i=1}^{N} z_i u_i\,, \\ A\sum_{i=1}^{N} u_i + B\cdot N = \sum_{i=1}^{N} z_i. \end{cases} \tag{5}$$

Или в общем виде:

$$\begin{cases} Ak_{11} + Bk_{12} = p_1, \\ Ak_{21} + Bk_{22} = p2, \end{cases} \tag{6}$$

где $k_{11},\ldots k_{22}$ – коэффициенты системы (5):

$$k_{11} = \sum_{i=1}^{N} u_i^2, \quad k_{12} = \sum_{i=1}^{N} u_i = k_{21}, \\ k_{22} = N. \tag{7}$$

p_1 и p_2 – правые части системы уравнений (5):

$$p_1 = \sum_{i=1}^{N} z_i u_i\,, \quad p_2 = \sum_{i=1}^{N} z_i. \tag{8}$$

Для решения системы уравнений (6) используется правило Крамера:

$$\Delta = \begin{vmatrix} k_{11} & k_{12} \\ k_{21} & k_{22} \end{vmatrix} = k_{11}\cdot k_{22} - k_{21}\cdot k_{12}\,,$$

$$\Delta A = \begin{vmatrix} p_1 & k_{12} \\ p_2 & k_{22} \end{vmatrix} = p_1\cdot k_{22} - p_2\cdot k_{12}\,, \tag{9}$$

$$\Delta B = \begin{vmatrix} k_{11} & p_1 \\ k_{21} & p_2 \end{vmatrix} = k_{11}\cdot p_2 - k_{21}\cdot p_1\,,$$

$$A = \frac{\Delta A}{\Delta}, \quad B = \frac{\Delta B}{\Delta}. \tag{10}$$

Для расчета параметров выбранного уравнения регрессии (№5 табл. методических указаний) воспользуемся формулами:

$$a = A, \quad b = e^{\frac{B}{A}}. \tag{11}$$

Значения отклика по модели $\hat{y}_i$ найдем как

$$\hat{y}_i = a\ln bx_i\,. \tag{12}$$

Проверка адекватности модели при отсутствии параллельных опытов

Адекватность полученного уравнения регрессии при отсутствии параллельных опытов проверяется по критерию Фишера сравнением остаточной дисперсии $S^2_{ост}$ и дисперсии относительного среднего S^2_y:

$$F = \frac{S^2_y}{S^2_{ост}}, \tag{13}$$

где

$$S^2_y = \frac{\sum_{i=1}^{N}(y_i - \bar{y})^2}{N-1}, \quad S^2_{ост} = \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{N-l}, \tag{14}$$

$$\bar{y} = \frac{\sum_{i=1}^{N} y_i}{N} \text{ – среднее значение выхода объекта;} \tag{15}$$

$\hat{y}_i$ – значения отклика, рассчитанные по модели; $N = 20$ – объем выборки; l – число связей, наложенных на выборку, равное числу определенных коэффициентов в уравнении (для линейной регрессии (2) $l = 2$).

Дисперсия $S^2_{ост}$ характеризует отклонение рассчитанных по модели значений от экспериментальных значений, а S^2_y – отклонение экспериментальных значений от средних.

Если F больше некоторого критического значения

$$F = \frac{S^2_y}{S^2_{ост}} > F_p(f_1, f_2, p), \tag{16}$$

то модель адекватна объекту.

$F_p(f_1, f_2, p)$ – табличное (критическое) значение критерия Фишера, которое зависит от чисел f_1, f_2 степеней свободы для дисперсий S^2_y и $S^2_{ост}$ соответственно, а также от уровня значимости p.

Число степеней свободы f любой дисперсии определяется разностью между общим количеством опытов и числом характеристик, рассчитанных по этим опытам и используемых при расчете дисперсий:

$$f_1 = N-1, \quad f_2 = N-l. \tag{17}$$

Для имеющихся исходных данных табл. 1 и выбранного уравнения регрессии (12) $f_1 = N-1 = 20-1 = 19, \quad f_2 = N-l = 20-2 = 18$; табличное значение $F_p(f_1, f_2, p) = F_p(19, 18, 5\%) = 2{,}2033$.

Оценка тесноты связи

В случае нелинейной связи между x и y (принимая во внимание вид модели (12)) оценка тесноты связи характеризуется величиной корреляционного отношения θ

$$\theta = \sqrt{1-\xi}, \tag{18}$$

где, учитывая (14),

$$\xi = \frac{(N-l)\cdot S_{ост}^2}{(N-1)\cdot S_y^2}. \tag{19}$$

Величина корреляционного отношения θ изменяется от 0 до 1. Чем больше θ, тем сильнее связь. При θ =1 наблюдается функциональная зависимость между x и y.

Схема алгоритма решения

Представленный алгоритм (рис.3) предназначен для вычисления параметров модели, величины корреляционного отношения и установления адекватности логарифмической модели при отсутствии параллельных опытов в эксперименте.


1
НАЧАЛО
2
$N = 20, l = 2,$
$F_p(f_1, f_2, p), x_i, y_i,$
$i = \overline{1, N}$
Ввод исходных данных
3
$u_i = \ln x_i$
$z_i = y_i$
Переход к линейной модели
4
$p_1 = p_2 =$
$= k_{11} = k_{12}$
Обнуление ячеек для накопления сумм
5
$i = \overline{1, N}$
$k_{11} = k_{11} + u_i^2$
$k_{12} = k_{12} + u_i$
6
$p_1 = p_1 + z_i u_i$
$p_2 = p_2 + z_i$
Вычисление сумм по пересчитан-ным данным u_i, z_i для расчета коэффициен тов k_{11}, k_{12}, p_1, p_2 в системе (6)
7
$k_{21} = k_{12}$
$k_{22} = N$
Расчет коэффициен тов по формулам (7)
8

Рисунок 3. Схема алгоритма решения задачи

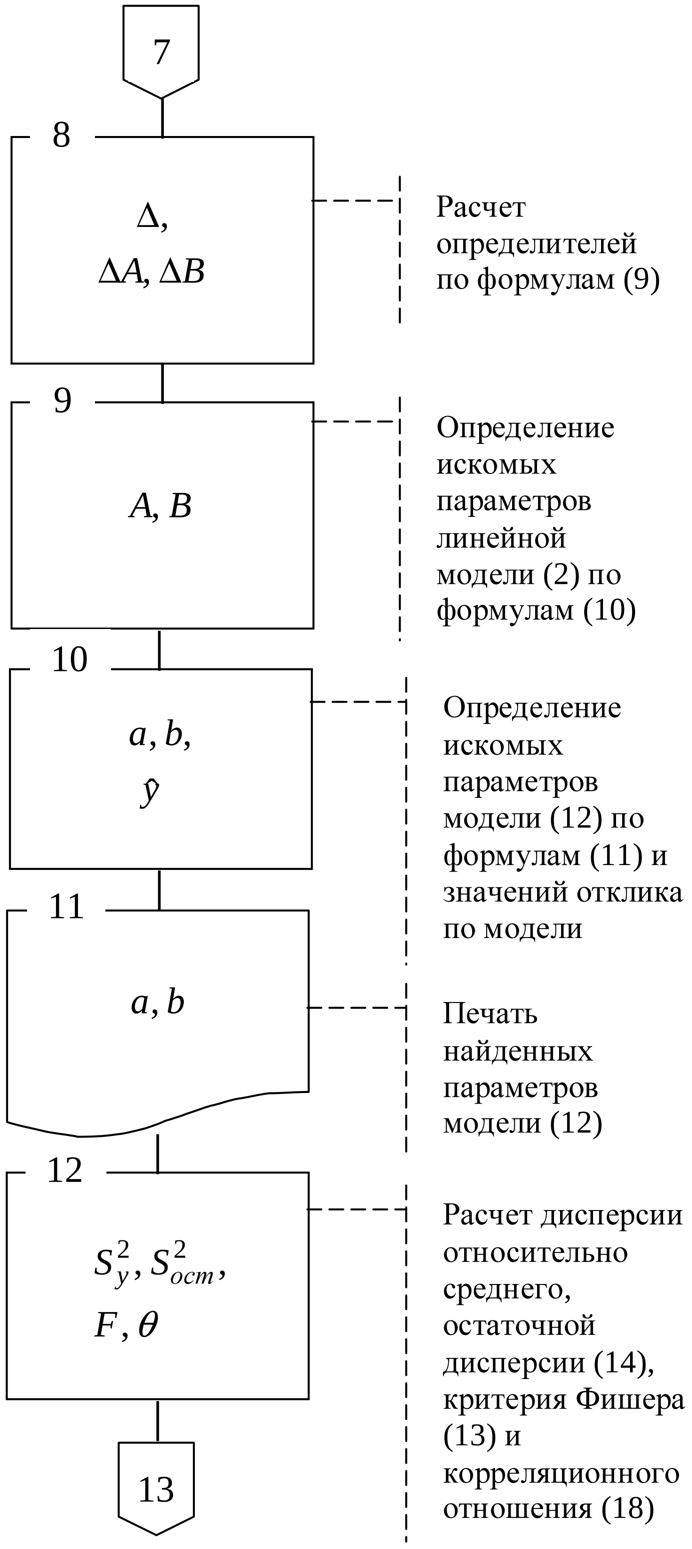


Рисунок 3. Продолжение

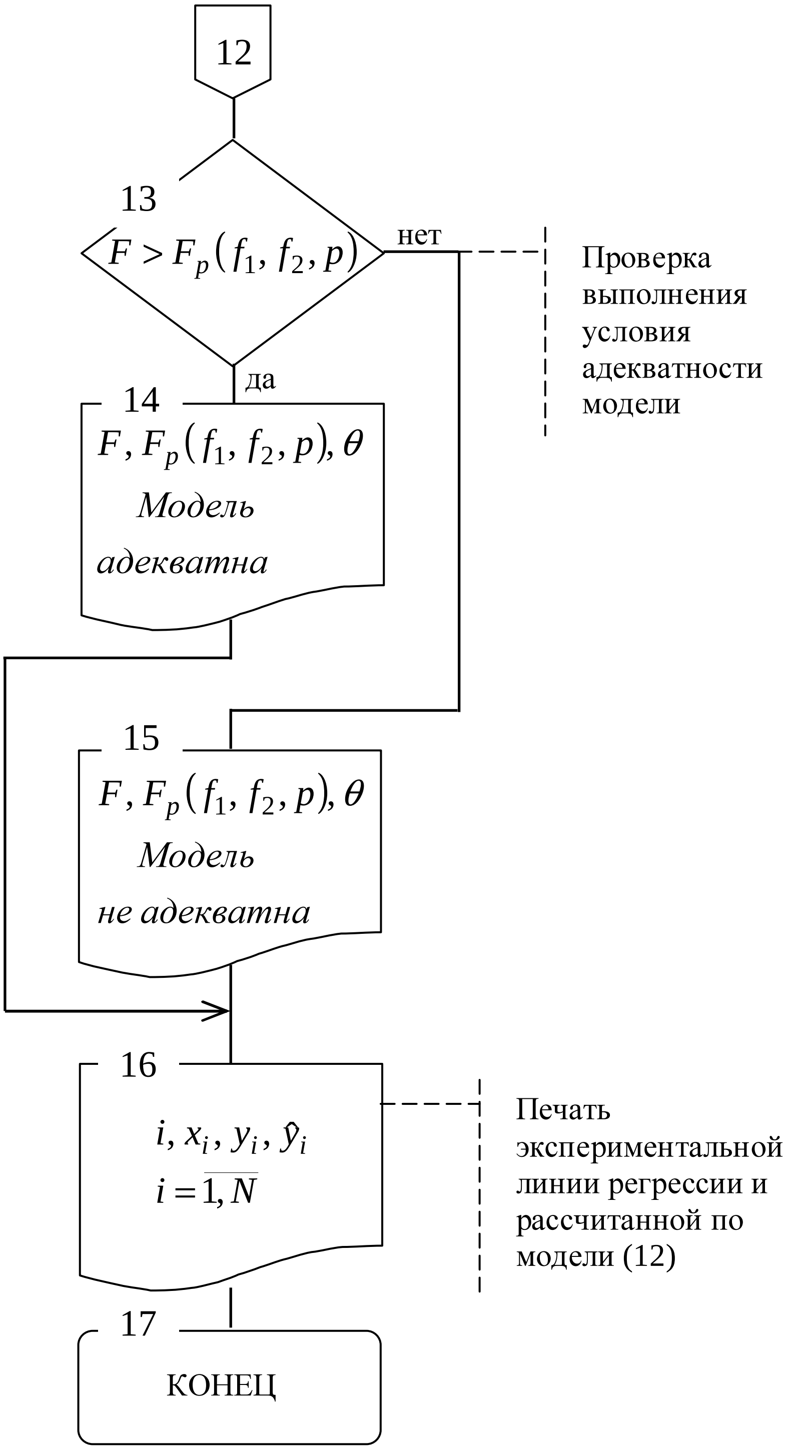


Рисунок 3. Окончание

На основании схемы алгоритма решения задачи (рис. 3) составим программу расчета параметров модели, критерия Фишера, а также коэффициента корреляционного отношения, используя математический пакет MATLAB.

Листинг программы

```
clear, clc
%1

%2 Ввод исходных данных
N=20
L=2
Fp=2.2033
x=1:0.1:2.9
y=[12.95 14.95 16.78 18.46 20.02 21.47 22.84 24.15 25.38 26.43
27.59 28.62 29.6 30.46 31.38 32.27 33.08 33.08 33.91 34.58]
% Печать экспериментальной линии регрессии y по x
figure(1)
plot(x,y,'r',x,y,'r .')
title('Экспериментальная линия регрессии y по x')
xlabel('x')
ylabel('y')

%3 Переход к линейной модели
u=log(x);
z=y;
% Печать экспериментальной линии регрессии y по ln(x)
figure(2)
plot(u,z,u,z,'b .')
title('Экспериментальная линия регрессии y по ln(x)')
xlabel('ln(x)')
ylabel('y')

%4 Обнуление ячеек для накопления сумм
p1=0;
p2=0;
k11=0;
k12=0;

%5-6 Вычисление сумм по пересчитанным данным
for i=1:N
    k11=k11+u(i)^2;
    k12=k12+u(i);
    p1=p1+z(i)*u(i);
    p2=p2+z(i);
end

%7 Расчет коэффициентов по формулам (7)
k21=k12;
k22=N;

%8 Расчет определителей по формулам (9)
D=k11*k22-k21*k12;
```

```
DA=p1*k22-p2*k12;
DB=k11*p2-k21*p1;


%9 Определение искомых параметров
% линейной модели (2) по формулам (10)
A=DA/D;
B=DB/D;

%10 Определение искомых параметров модели (12)
% по формулам (11) и значений отклика по модели
a=A;
b=exp(B/A);
i=1:N;
ym(i)=a*log(b*x(i));

%11 Печать найденных параметров модели (12)
'Параметры модели y(i)=a*log(b*x(i))',a,b

%12 Расчет дисперсии относительно среднего,
% остаточной дисперсии (14), критерия Фишера (13)
% и корреляционного отношения (18)
% расчет среднего
S=0;
for i=1:N
    S=S+y(i);
end
y_=S/N;

% расчет дисперсии относительно среднего,
% остаточной дисперсии
Sy=0;
So=0;
for i=1:N
    Sy=Sy+(y(i)-y_)^2;
    So=So+(y(i)-ym(i))^2;
end
S2y=Sy/(N-1);
S2o=So/(N-L);

% расчет критерия Фишера
F=S2y/S2o;

% расчет корреляционного отношения
ksi=(N-L)*S2o/((N-1)*S2y);
tet=sqrt(1-ksi);

%13-14-15 Проверка выполнения условия адекватности модели
if F>Fp
```

```
   'расчетное значение по критерию Фишера',F,
   'критическое значение по критерию Фишера',Fp,
   'величина корреляционного отношения', tet,
   'F>Fp: модель адекватна'
else 'расчетное значение по критерию Фишера',F,
   'критическое значение по критерию Фишера',Fp,
   'величина корреляционного отношения', tet,
   'F<=Fp: модель не адекватна'
end

%16 Печать экспериментальной линии регрессии
% и рассчитанной по модели (12)
figure(3)
plot(x,y,'r',x,y,'r .',x,ym,'b--',x,ym,'b.')
legend('экспериментальная','линия регрессии','модельная','линия
регрессии',4)
title('Линия регрессии y по x')
xlabel('x')
ylabel('y')
```

Результат работы программы

```
N =

    20


L =

     2


Fp =

    2.2033


x =

  Columns 1 through 8

    1.0000    1.1000    1.2000    1.3000    1.4000    1.5000
1.6000    1.7000

  Columns 9 through 16
```

```
    1.8000    1.9000    2.0000    2.1000    2.2000    2.3000
2.4000    2.5000

  Columns 17 through 20

    2.6000    2.7000    2.8000    2.9000


y =

  Columns 1 through 8

   12.9500   14.9500   16.7800   18.4600   20.0200   21.4700
22.8400   24.1500

  Columns 9 through 16

   25.3800   26.4300   27.5900   28.6200   29.6000   30.4600
31.3800   32.2700

  Columns 17 through 20

   33.0800   33.0800   33.9100   34.5800


ans =

Параметры модели y(i)=a*log(b*x(i))


a =

   20.6068


b =

    1.8902


ans =

расчетное значение по критерию Фишера


F =

  714.1568
```

```
ans =

критическое значение по критерию Фишера


Fp =

    2.2033


ans =

величина корреляционного отношения


tet =

    0.9993


ans =

F>Fp: модель адекватна
```

В результате расчетов получаем, что выбранная модель регрессии (12) с найденными параметрами адекватна и имеет вид

$$\hat{y}_i = 20.6068 \cdot \ln(1.8902 \cdot x_i). \qquad (20)$$

Корреляционное отношение $\theta = 0.9993 \approx 1.00$, что говорит об очень сильной связи между x и y. Таким образом, найденная модель (20) пригодна к использованию. Графики экспериментальной и расчетной кривых, представленные на рис.4, подтверждают качество модели (20).

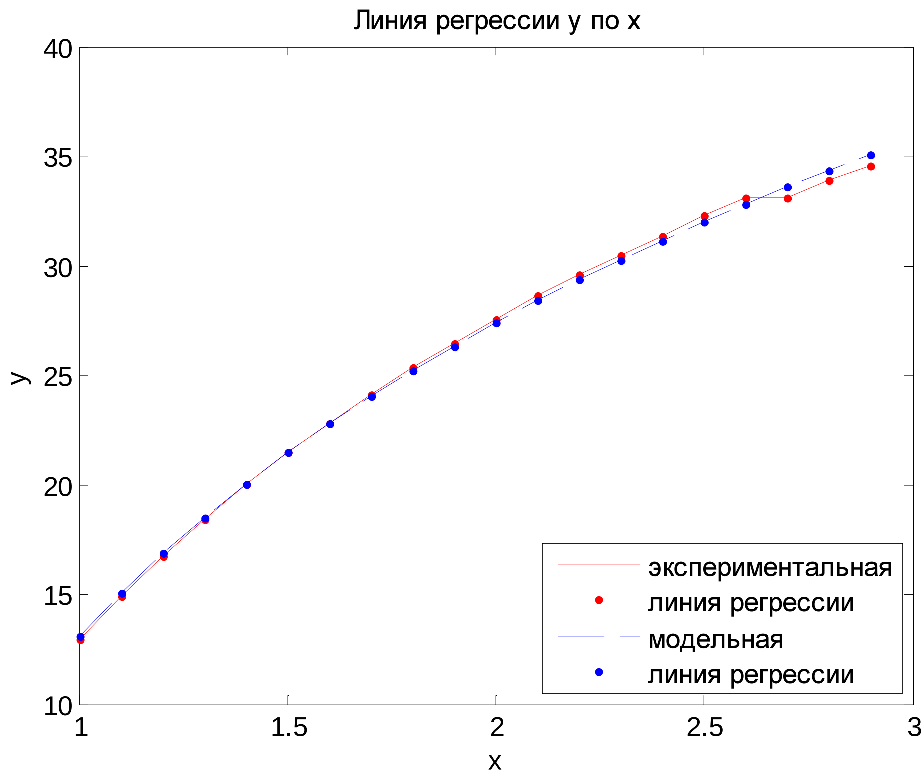


Рисунок 4. Графики экспериментальной и расчетной кривых

Выводы

В данной работе изучены методика обработки экспериментальных данных и получения параметров модели технологического процесса методом наименьших квадратов (МНК), а также методика проверки адекватности полученной модели по критерию Фишера и оценки тесноты связи между переменными.

На основании предложенных экспериментальных данных, полученных на промышленных технологических установках ректификации, дегидрирования углеводов, сушки и т. д., выработан вид однопараметрической линии регрессии вида $\hat{y}_i = a \ln bx_i$.

Получен алгоритм расчета параметров регрессии, построена схема алгоритма и составлена программа расчета параметров модели, критерия Фишера и коэффициента корреляционного отношения. В ходе работы программы определены требуемые параметры регрессии так, что $\hat{y}_i = 20.6068 \cdot \ln(1.8902 \cdot x_i)$.

Выполнена оценка адекватности предложенной регрессионной модели реальному процессу (модель адекватна), произведен расчет величины корреляционного отношения, что позволяет судить о сильной связи между переменными.