

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РФ
Федеральное государственное автономное образовательное учреждение
высшего образования
«СЕВЕРО-КАВКАЗСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»

Методические указания

по выполнению лабораторных работ
по дисциплине

«Обработка экспериментальных данных»

для направления подготовки 15.03.04 Автоматизация технологических
процессов и производств
направленность (профиль) Информационно-управляющие системы

Невинномысск 2023

Содержание

Лабораторная работа №1 Статистическая обработка данных экспериментальных исследований в Mathcad.

Лабораторная работа №2 Многомерные вычисления в Matcard.

Лабораторная работа №3 Нейросетевая аппроксимация функции одной переменной в Matlab

Лабораторная работа №4 Планирование экспериментальных исследований. Методы статической оптимизации объектов исследования при наличии ошибок измерений на базе метода крутого восхождения (Бокса-Уилсона).

Лабораторная работа №5 Оценка ошибок измерений. Статистическая проверка гипотез. Исключение грубых ошибок (промахов).

Лабораторная работа №6 Изучение методов планирования эксперимента для получения модели регрессии второго порядка. Статистическая обработка моделей регрессии при планировании второго порядка

Лабораторная работа №7 Расчёт аппроксимации экспериментальных данных с использованием Microsoft Excel.

Лабораторная работа №8 Расчёт аппроксимации экспериментальных данных с использованием MathCAD.

Лабораторная работа №9 Расчёт аппроксимации экспериментальных данных с использованием MatLAB.

Лабораторная работа № 1 «Статистическая обработка в Mathcad»

Цель работы: научиться использовать математические и статистические функции в системе MathCAD.

Задания на выполнение

1. Линейная и сплайновая интерполяции.

Исходные данные для выполнения задания 1 в таблице 1.

Таблица 1.

варианты	1	2	3	4	5	6	7	8	9
Исходные данные	$\begin{bmatrix} 1 & 0 \\ 2 & 3 \\ 4 & 6 \\ 8 & 5 \end{bmatrix}$	$\begin{bmatrix} 2 & 4 \\ 5 & 8 \\ 9 & 3 \\ 1 & 4 \end{bmatrix}$	$\begin{bmatrix} 1 & 5 \\ 6 & 3 \\ 2 & 4 \\ 7 & 1 \end{bmatrix}$	$\begin{bmatrix} 2 & 0 \\ 3 & 5 \\ 4 & 9 \\ 8 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 3 & 5 \\ 5 & 7 \\ 7 & 9 \end{bmatrix}$	$\begin{bmatrix} 2 & 3 \\ 4 & 6 \\ 6 & 8 \\ 9 & 0 \end{bmatrix}$	$\begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \\ 8 & 9 \end{bmatrix}$	$\begin{bmatrix} 3 & 5 \\ 4 & 7 \\ 5 & 8 \\ 6 & 9 \end{bmatrix}$	$\begin{bmatrix} 3 & 5 \\ 4 & 7 \\ 5 & 8 \\ 6 & 9 \end{bmatrix}$

Рассмотрим линейную интерполяцию

Линейная интерполяция

$$A := \begin{pmatrix} 1 & 0 \\ 2 & 3 \\ 3 & 5 \\ 4 & 1 \\ 5 & 2 \\ 6 & 4 \end{pmatrix} \quad \text{Векторы исходных данных X и Y}$$

$$A := \text{csort}(A, 0) \quad X := A^{(0)} \quad Y := A^{(1)}$$

$$f(x) := \text{linterp}(X, Y, z) \quad f(2) = 3 \quad f(7) = 6$$

Сплайновая интерполяция

$$S := \text{scpline}(X, Y) \quad fs(x) := \text{interp}(S, X, Y, x)$$

$$fs(2) = 3 \quad fs(7) = -2.067$$

Как мы видим значения точек при линейной и сплайновой интерполяции различны.

Построим графики линейной и сплайновой интерполяций. Для этого задается количество точек и шаг.

```

i := 0..length(X)-1    scale := 100    j := 0..scale
x_j = min(X) + j * (max(X)-min(X))/scale

```

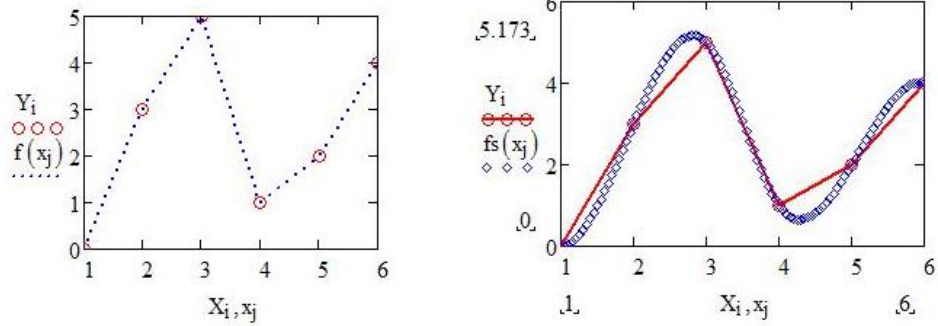


Рис. 1. Линейная и сплайновая интерполяции

2. Линейная регрессия.

Исходные данные для выполнения задания 2 в таблице 2.

№	Заданные вектора	№	Заданные вектора
1	VX=[3, 2, 4, 5] VY=[7, 8, 9, 5]	4	VX=[7, 18, 3, 11] VY=[1, 5, 3, 9]
2	VX=[12, 14, 7, 11] VY=[6, 8, 10, 15]	5	VX=[24, 9, 12, 27] VY=[9, 3, 17, 11]
3	VX=[3, 9, 12, 14] VY=[7, 9, 11, 13]	6	VX=[4, 15, 2, 19] VY=[11, 17, 1, 13]

$$\begin{matrix}
 VX := \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{pmatrix} &
 VY := \begin{pmatrix} 4 \\ 11 \\ 15 \\ 21 \\ 28 \end{pmatrix} &
 \text{Исходные данные} \\
 ORIGIN := 1
 \end{matrix}$$

Вычислим коэффициенты a и b линейной регрессии

$$a := \text{intercept}(VX, VY) \quad b := \text{slope}(VX, VY)$$

$$i := 1..6 \quad f(x) := a + b \cdot x$$

$$1 = -1.6 \quad b = 5.8 \quad \text{corr}(VX, VY) = 0,996$$

$$f(3) = 15,8 \quad \text{linterp}(VX, VY, 3) = 15$$

По этим данным строим график линейной регрессии

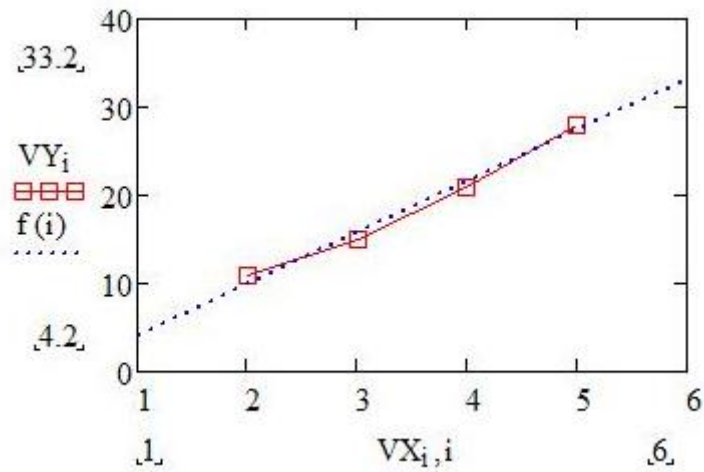


Рис. 2. Линейная регрессия

3. Линейная регрессия общего вида.

Исходные данные для выполнения задания 3 в таблице 2.

$$VX := \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{pmatrix} \quad VY := \begin{pmatrix} 4 \\ 11 \\ 15 \\ 21 \\ 28 \end{pmatrix} \quad F(x) := \begin{pmatrix} \frac{1}{x^2} \\ x \\ 2 \cdot x \\ \exp(x) \end{pmatrix}$$

Функции $F(x)$ подбираются так, чтобы коэффициент детерминации стремился к единице.

$$K := \text{linfit}(VX, VY, F) \quad g(t) := F(t) \cdot K$$

$$R^2 = \frac{\sum (g(VX) - \text{mean}(VY))^2}{\sum (VY - \text{mean}(VY))^2} = 0.999 \quad \text{коэффициент детерминации}$$

$$i := 0..4 \quad r := 1, 1.25..5 \quad K := \begin{pmatrix} -5.242 \\ -0.918 \\ 1.613 \\ 3.226 \\ 0.079 \end{pmatrix}$$

$g(t)$ – это функция регрессии

K – коэффициент функции регрессии

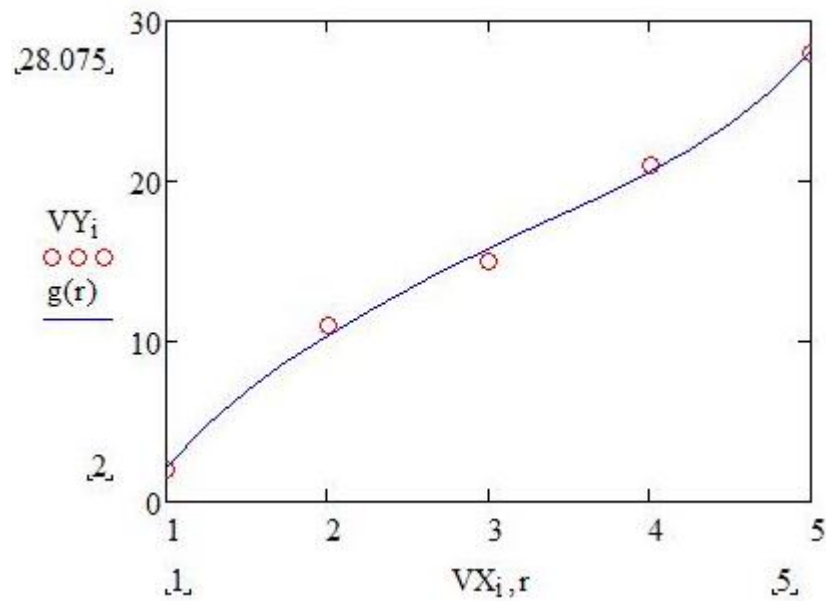


Рис. 3. Линейная регрессия общего вида

4. Полиномиальная регрессия

Исходные данные для выполнения задания 4 в таблице 2.

$$VX := \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{pmatrix} \quad VY := \begin{pmatrix} 4 \\ 11 \\ 15 \\ 21 \\ 28 \end{pmatrix} \quad k := 3 \quad \text{Исходные данные}$$

k – степень полинома, от него зависит насколько точно будет построена функция.

$$z := \text{regress}(VX, VY, K)$$

$$F(x) := \text{interp}(z, VX, VY, x)$$

$$\text{Coeffs} := \text{submatrix}(z, 3, \text{length}(z)-1, 0, 0)$$

$$\text{Coeffs}^T = (-6.2 \quad 12.81 \quad -2.857 \quad 0.333)$$

$F(x)$ - выдаёт интерполированное значение в x от коэффициентов вектора напротив, и оригинальных данных в VX и VY .

Coeffs - выдаёт субматрицу массива состоящую из элементов в строках ir через jr и столбцах ic через jc из z .

Coeffs^T - преобразует коэффициент в горизонтальную запись

$$R^2 = \frac{\sum (F(VX) - \text{mean}(VY))^2}{\sum (VY - \text{mean}(VY))^2} = 0.998 \quad \text{коэффициент детерминации}$$

$$i := 0.6 \quad j := 0.4 \quad tx_j := \min(VX) + j \cdot \frac{\max(VX) - \min(VX)}{50}$$

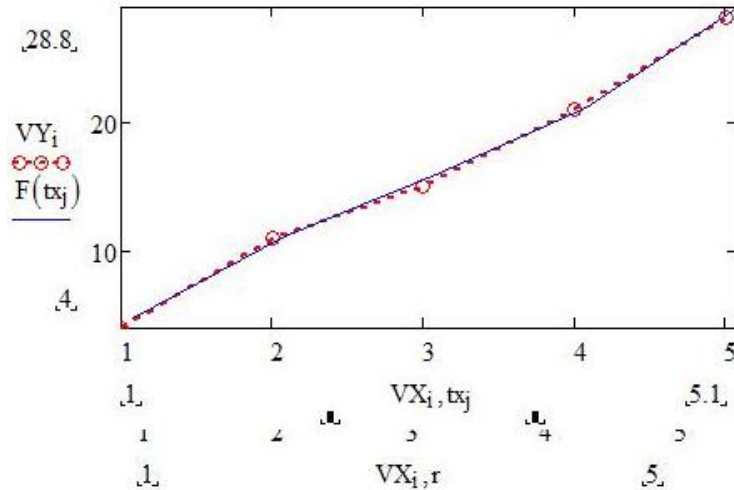


Рис. 4. Полиномиальная регрессия

5. Нелинейная регрессия общего вида.

Исходные данные для выполнения задания 5 в таблице 2

$$F(x, a, b) := a \cdot \exp(-b \cdot x) + a \cdot b$$

$$ORIGIN := 1$$

$$VX := \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{pmatrix} \quad VY := \begin{pmatrix} 4 \\ 11 \\ 15 \\ 21 \\ 28 \end{pmatrix} \quad VS := \begin{pmatrix} 2 \\ 2 \end{pmatrix}$$

$$F1(x, k) := \begin{pmatrix} k_1 \exp(-k_2 x + k_1 k_2) \\ \exp(-k_2 x) + k_2 \\ -k_1 x \cdot \exp(-k_2 x) + k_1 \end{pmatrix}$$

$$P := \text{genfit}(VX, VY, VS, F1) \quad G(x) := F1(x, P)_1 \quad x := 0, 0.1..4$$

$$P = \begin{pmatrix} 0.762 \\ 20.722 \end{pmatrix}$$

Вектор P возвращает значения a=k1 и b=k2 для наилучшего среднеквадратического приближения F(x, a, b)

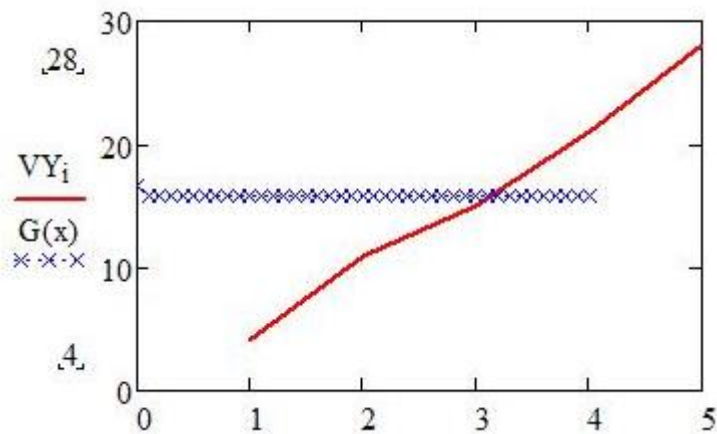


Рис. 5. Нелинейная регрессия общего вида

В заключение работы, следует сделать *вывод* о целесообразности применения той или иной регрессии, чем они друг от друга отличаются.

Контрольные вопросы

1. Что такое ранжированная переменная и для решения, каких задач она используется?
2. Какие категории функций имеются в системе MathCAD?
3. Дать понятие функции пользователя?
4. Назовите виды операторов системы MathCAD и поясните их назначение.
5. Как вывести результаты вычислений в виде таблиц?
6. Организация вложенных циклов.
7. Правила задания многомерных функций.

Лабораторная работа № 2 Многомерные вычисления в Mathcard

Цель работы: знакомство с правилами организации вложенных циклов, многомерных вычислений в программе Mathcard. Построение трехмерных графиков в Mathcard.

Задание к работе

1. Построить график поверхности в системе Matcard.

Исходные данные в Таблице 1.

Таблица 1.			
№	Функция	№	Функция
1	$z(x,y)=3\cos(x*y)$	2	$z(x,y)=x^2*y$
3	$z(x,y)=\sin(x*y)$	4	$z(x,y)=5\cos(x*y)$
5	$z(x,y)=\text{tg}(x*y)$	6	$z(x,y)=2x+y^2$

Построение поверхностей по матрице аппликат их точек. Поскольку элементы матрицы M – переменные с целочисленными индексами, то перед созданием матрицы требуется задать индексы в виде ранжированных переменных с целочисленными значениями, а затем уже из них формировать сетку значений x и y – координат для аппликат $z(x,y)$. Значения x и y при этом обычно должны быть вещественными числами, нередко как положительными, так и отрицательными. После выполнения указанных выше определений вводится шаблон графика (команда Surface Plot). (Рисунок 8.1)

Построение Графика Поверхности

$z(x,y) := \cos(x*y)$ Функция двух переменных x, y
 $i := 0..20$ $j := 0..20$ Целочисленные индексы

$$M_{i,j} := z \left[\frac{(i-10)}{5}, \frac{(j-10)}{5} \right]$$

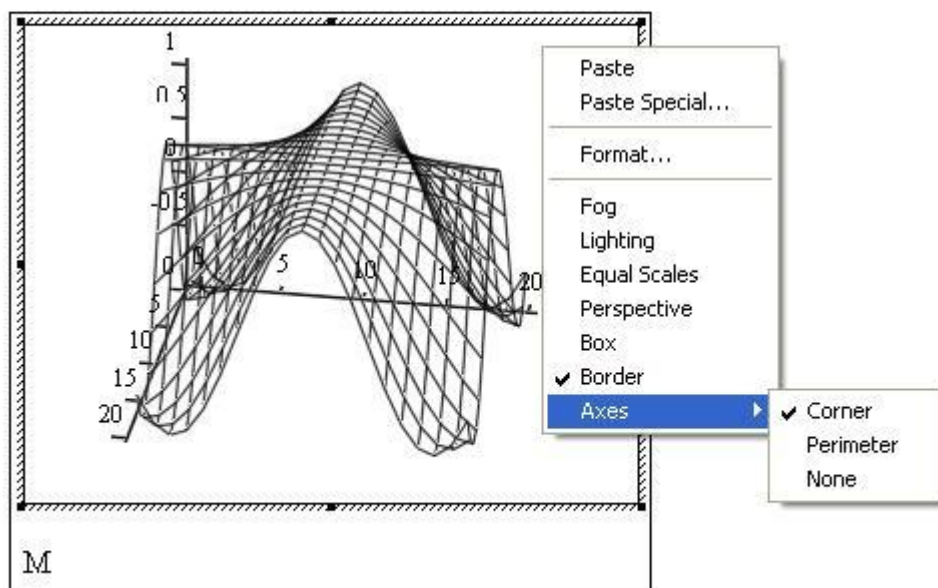


Рис. 1. Построение поверхности без удаления невидимых линий

2. Построить график поверхности в системе Matcard отформатировав его, применив алгоритм функциональной окраски и удаление невидимых линий.

На рисунке 2 показано, как отформатировать график, применение алгоритма функциональной окраски поверхности и удаление невидимых линий.

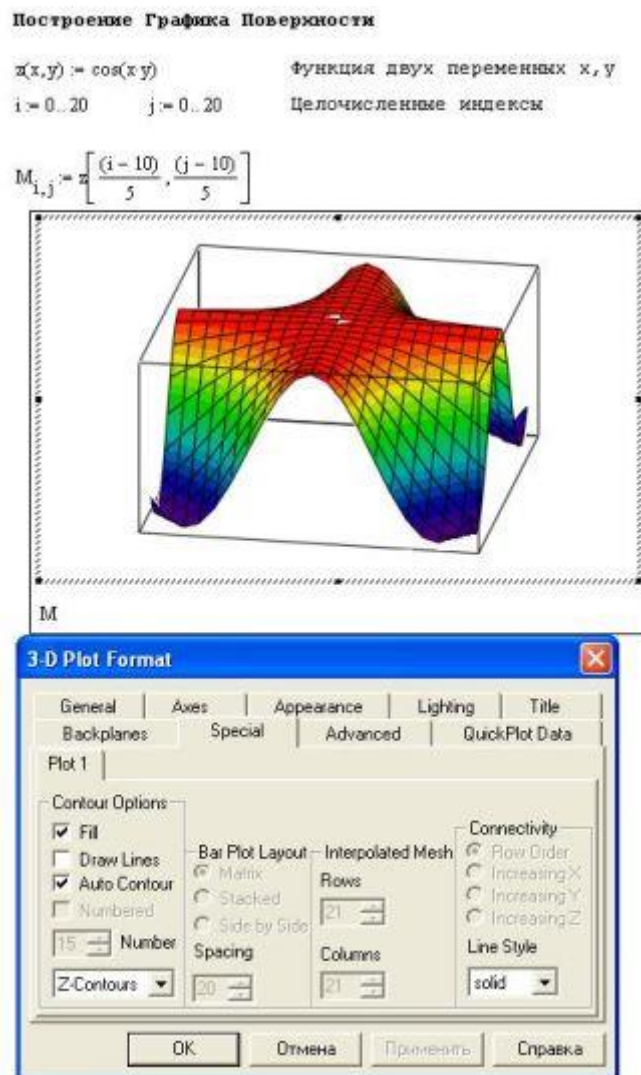


Рис. 2.

3. Построить трехмерный график в системе Matcard

В программе Mathcad есть возможность построения трехмерных графиков – без задания матриц аппликат поверхностей. Единственным недостатком такого упрощенного метода построения поверхностей является неопределенность в масштабировании, поэтому графики требуют форматирования.

Построение графика поверхности без задания матрицы

$z(x,y) = \cos(x \cdot y)$ функция двух переменных x, y

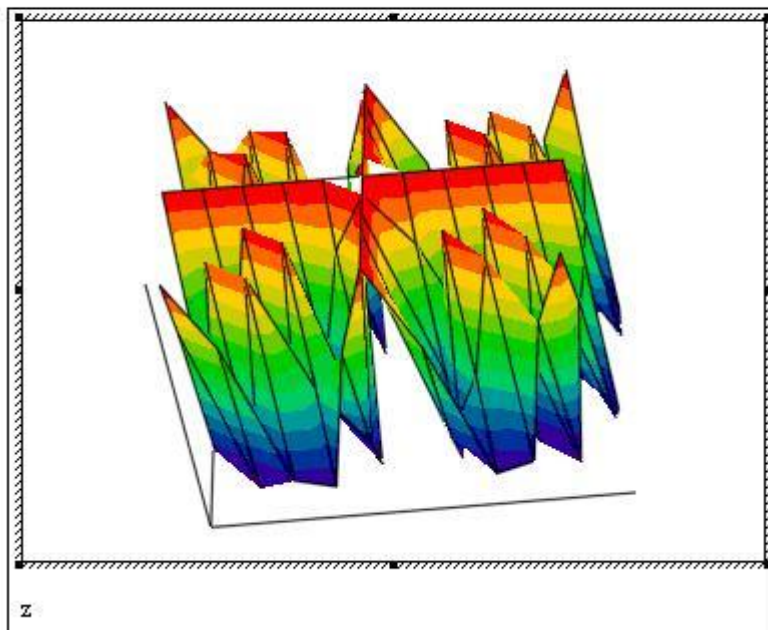


Рис. 3. Построение графика поверхности без задания матрицы

4. Построение графика поверхности без задания матрицы

Варианты задания в Таблице 2, как построить данный график показано на рисунке 4.

Таблица 2.	
№	Функция
1	$H(u,v)=3(u^2 \cdot v)$
2	$H(u,v)=3\sin(u \cdot v)$
3	$H(u,v)=7\cos(u \cdot v)$
4	$H(u,v)=\cos(u \cdot v)$
5	$H(u,v)=\text{tg}(u \cdot v)$
6	$H(u,v)=\sin(u \cdot v)$

Применение графической функции CreateMesh

1. Зададим функцию двух переменных $H(u, v) := \sin(u \cdot v)$
2. Зададим пределы изменения переменных $u0 := -2 \quad u1 := 2$
 $v0 := -2 \quad v1 := 2$
3. Используя функцию CreateMesh создадим матрицу поверхности
 $C := \text{CreateMesh}(H, u0, u1, v0, v1)$
4. Задав построение графика типа Surface Plot, получим

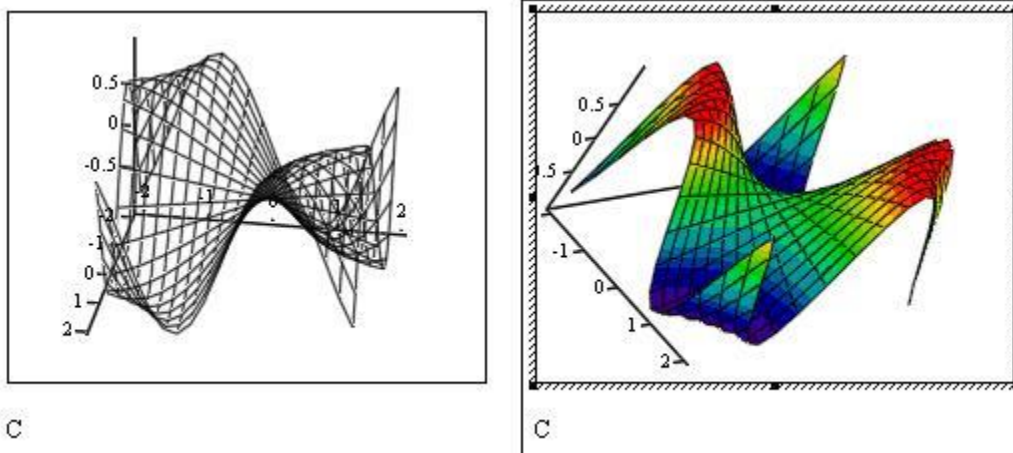


Рис. 4. Пример построения графика поверхности с применением функции CreateMesh

Еще один пример применения функции CreateMesh – построение объемной фигуры, которая получается вращением кривой, заданной функцией $f(x)$, вокруг оси X или Y.

Исходные данные в таблице 3.

Таблица 3.			
№	Функции	№	Функции
1	$f(x)=\cos(x^2)$ $G(u,v)=f(u)*\sin(v)$ $H(u,v)=f(u)*\cos(v)$	4	$f(x)=3\cos(x^2)$ $G(u,v)=f(u)*4\sin(v)$ $H(u,v)=f(u)*6\cos(v)$
2	$f(x)=3(x^2)$ $G(u,v)=f(u)*v$ $H(u,v)=f(u)*3\cos(v)$	5	$f(x)=3x*\cos(x^2)$ $G(u,v)=f(u)*(v^2)$ $H(u,v)=f(u)*3(v)$
3	$f(x)=\text{tg}(x^2)$ $G(u,v)=f(u)*(v^2)$ $H(u,v)=f(u)*3(v)$		

На рисунке 5 показан пример решения данной задачи.

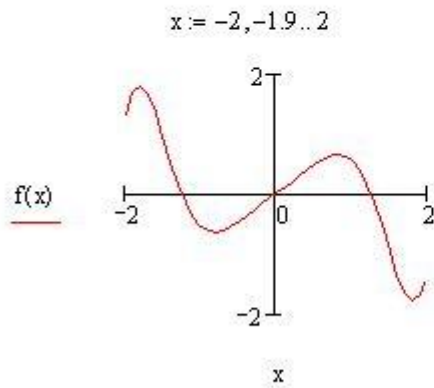
Построение фигуры, полученной вращением кривой вокруг оси X

$f(x) := x \cos(x^2)$ $a := -2$ $b := 2$ $mesh := 30$

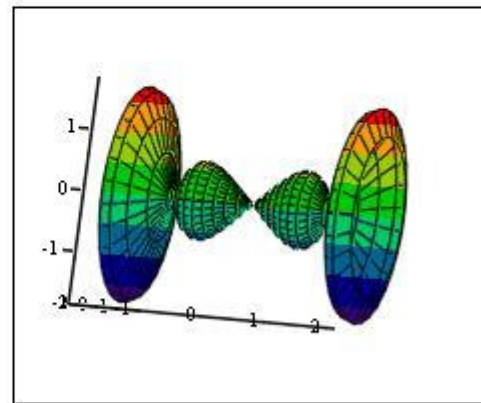
$F(u, v) := u$ $G(u, v) := f(u) \cdot \cos(v)$ $H(u, v) := f(u) \cdot \sin(v)$

$S := \text{CreateMesh}(F, G, H, a, b, 0, 2\pi, mesh)$

График функции $f(x)$



Поверхность вращения графика
 $f(x)$



S

Рис. 5. Фигура полученная вращением кривой.

Контрольные вопросы

1. Организация вложенных циклов.
2. Правила задания многомерных функций.
3. Вывод многомерных результатов в форме таблицы.
4. Трехмерная графика с функциональной окраской раскраской.
5. Назначение команды Surface Plot.
6. Для чего используется функция CreateMesh?
7. Как построить фигуру с помощью вращения во круг оси?

Лабораторная работа № 3

Нейросетевая аппроксимация функции одной переменной в Matlab

1. Цель работы

Научиться работать с сетью прямой передачи сигнала, функция newff. Разобраться с алгоритмом обратного распространения ошибки.

2. Краткие теоретические сведения

В лабораторной работе рассматривается нейронная сеть с прямой передачей сигнала (с прямой связью), то есть сеть, в которой сигналы передаются только в направлении от входного слоя к выходному, и элементы одного слоя связаны со всеми элементами следующего слоя. Важнейшим для реализации нейронных сетей является определение алгоритма обучения сети.

В настоящее время одним из самых эффективных и обоснованных методов облужения нейронных сетей является *алгоритм обратного распространения ошибки*, который применим к *однонаправленным многослойным сетям*. В многослойных нейронных сетях имеется множество скрытых нейронов, входы и выходы которых не являются входами и выходами нейронной сети, а соединяют нейроны внутри сети, то есть *скрытые нейроны*. Занумеруем выходы нейронной сети индексом $j = 1, 2, \dots, n$, а обучающие примеры индексом $M = 1, 2, \dots, M_0$. Тогда в качестве целевой функции можно выбрать функцию ошибки как сумму квадратов расстояний между реальными выходными состояниями y_{jM} нейронной сети, выдаваемых сетью на входных данных примеров, и правильными значениями функции d_{jM} , соответствующими этим примерам. Пусть $\mathbf{x} = \{x_i\}$ – столбец входных значений, где $i=1, 2, \dots, n$. Тогда $\mathbf{y} = \{y_j\}$ – выходные значения, где $j=1, 2, \dots, m$. В общем случае $n \neq m$. Рассмотрим разность $y_{jM} - d_{jM}$, где d_{ji} – точное (правильное) значение из примера. Эта разность должна быть минимальна. Введем расстояния согласно евклидовой метрике, определив норму

$$\|\mathbf{y} - \mathbf{d}\| = \sqrt{(\mathbf{y} - \mathbf{d}, \mathbf{y} - \mathbf{d})^2} . \quad (1)$$

Пусть целевая функция имеет вид

$$E = \frac{1}{2} \sum_{j,M} (y_{j,M} - d_{j,M})^2 . \quad (2)$$

Задача обучения нейронной сети состоит в том, чтобы найти такие коэффициенты $w_{\beta k}$, при которых достигается минимум $E(\mathbf{w})$ ($E \geq 0$).

На рис. 1 показана архитектура нейронной сети с прямой передачей сигнала.

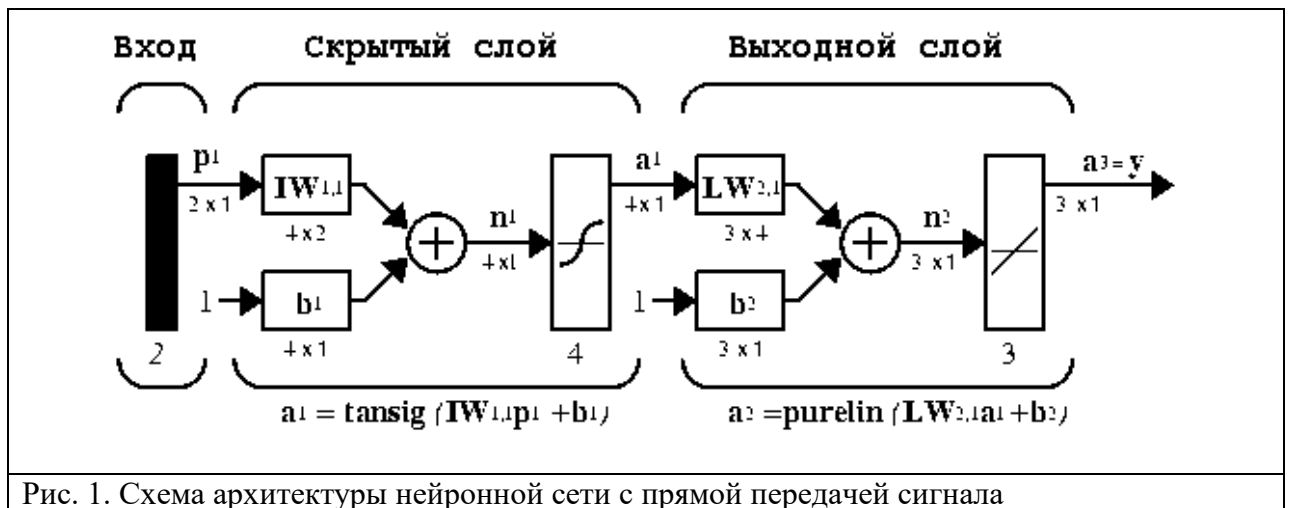


Рис. 1. Схема архитектуры нейронной сети с прямой передачей сигнала

Здесь приняты обозначения: p^1 - вектор входа, $IW^{i,j}$, $LW^{i,j}$ - матрицы весов входа и выхода, b^i - смещение, a^i - выход слоя, y - выход сети, *tansig* (гиперболическая тангенциальная), *purelin* (линейная) - соответствующие функции активации.

Весы и смещения определяются с помощью алгоритма обратного распространения ошибок.

Обучение сети обратного распространения требует выполнения следующих операций:

1. Выбрать очередную обучающую пару из обучающего множества; подать входной вектор на вход сети.
2. Вычислить выход сети.
3. Вычислить разность между выходом сети и требуемым выходом (целевым вектором обучающей пары).
4. Скорректировать веса сети так, чтобы минимизировать ошибку.
5. Повторять шаги с 1 по 4 для каждого вектора обучающего множества до тех пор, пока ошибка на всем множестве не достигнет приемлемого уровня.

3. Пример решения типовой задачи

Выполнение лабораторной работы состоит из следующих этапов: прежде всего, необходимо оцифровать график функции $y=f(x)$, то есть получить ряд соответствующих значений по горизонтальной и вертикальной осям.

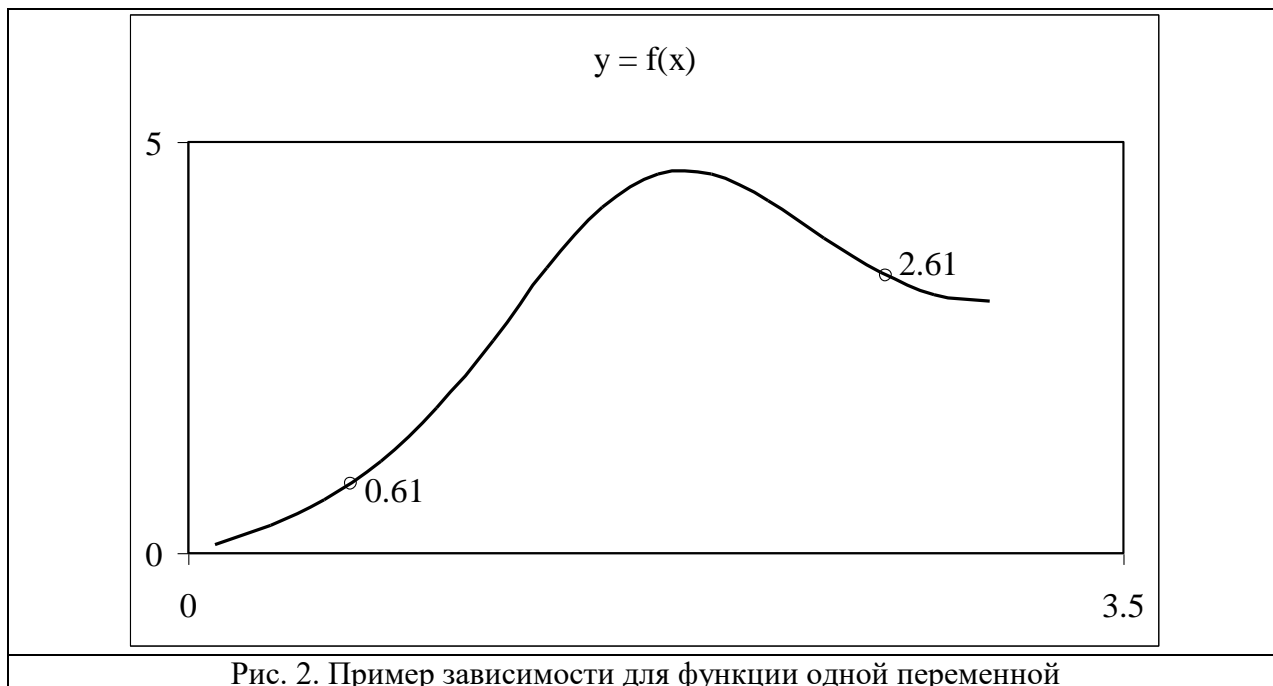


Рис. 2. Пример зависимости для функции одной переменной

В примере, показанном на рис. 2 были получены два массива, каждый из которых состоит из 15 значений.

По горизонтальной оси – [0.10 0.31 0.51 0.72 0.93 1.14 1.34 1.55 1.76 1.96 2.17 2.38 2.59 2.79 3.00].

По вертикальной оси – [0.1010 0.3365 0.6551 1.1159 1.7632 2.5847 3.4686 4.2115 4.6152 4.6095 4.2887 3.8349 3.4160 3.1388 3.0603].

Ниже приводится программа создания, обучения нейронной сети и вывода результатов.

```
x=[0.10 0.31 0.51 0.72 0.93 1.14 ...
    1.34 1.55 1.76 1.96 2.17 2.38 ...
    2.59 2.79 3.00];
y=[0.1010 0.3365 0.6551 1.1159 1.7632 2.5847 ...
    3.4686 4.2115 4.6152 4.6095 4.2887 3.8349 ...
    3.4160 3.1388 3.0603];
net=newff([0 3],[5,1],{'tansig','purelin'},'trainbfg');
net.trainParam.epochs=300;
net.trainParam.show=50;
net.trainParam.goal=1.37e-2;
[net,tr]=train(net,x,y);
an=sim(net,x);
plot(x,y,'+r',x,an,'-g'); hold on;
xx=[0.61 2.61];
v=sim(net,xx)
plot(xx,v,'ob','MarkerSize',5,'LineWidth',2)
```

В результате выполнения программы получают следующие результаты, отражённые на рис. 3 и 4:

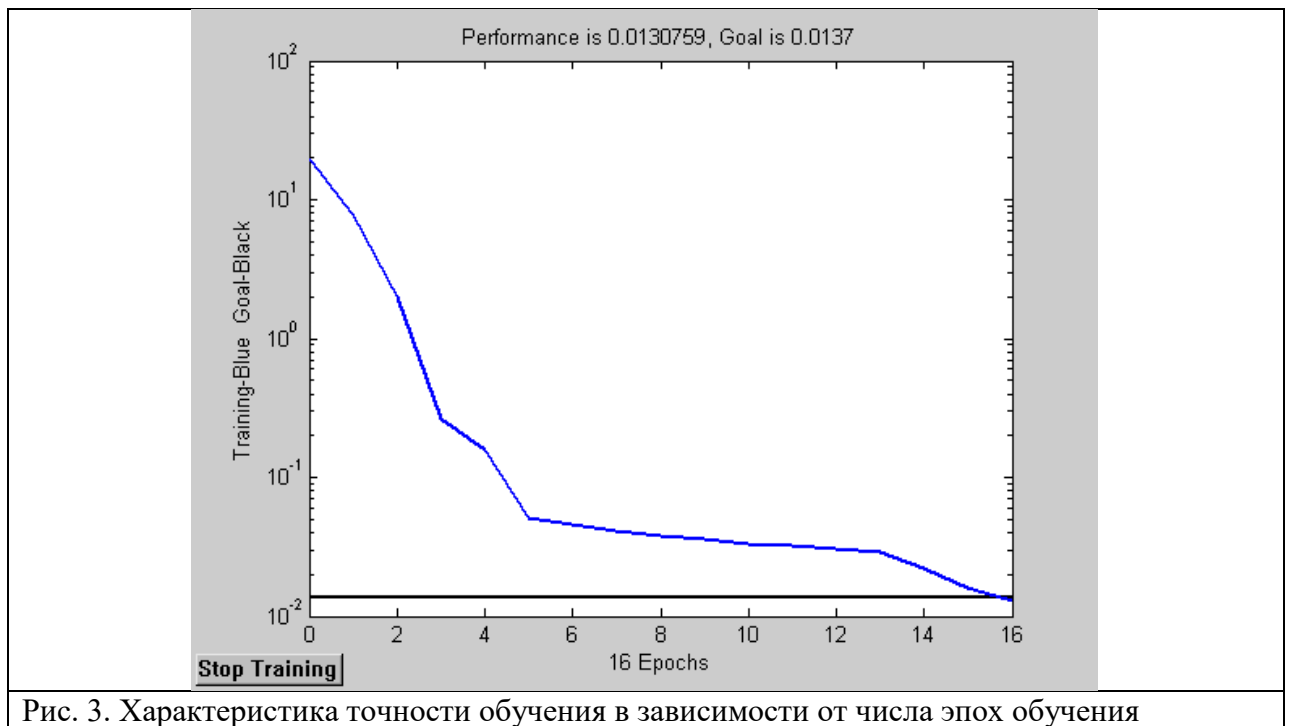


Рис. 3. Характеристика точности обучения в зависимости от числа эпох обучения

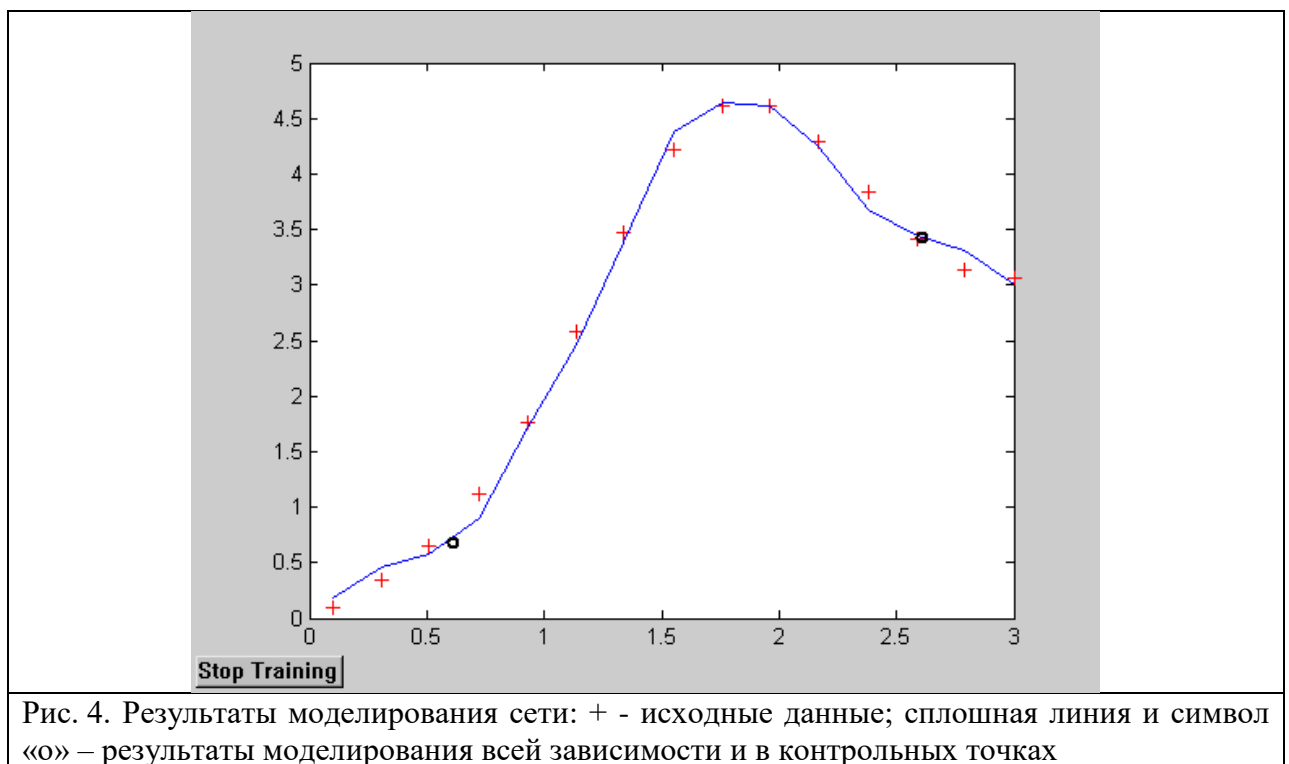


Рис. 4. Результаты моделирования сети: + - исходные данные; сплошная линия и символ «o» – результаты моделирования всей зависимости и в контрольных точках

В массиве v содержатся приближённые значения для двух контрольных точек, указанных на графике (рис. 2) $xx=[0.61 \ 2.61]$. При данных параметрах сети получены значения: $v = [1.05 \ 3.35]$. Сравнив эти приближённые значения с точными значениями $[0.85 \ 3.37]$, можно сделать вывод о корректности построения нейронной сети.

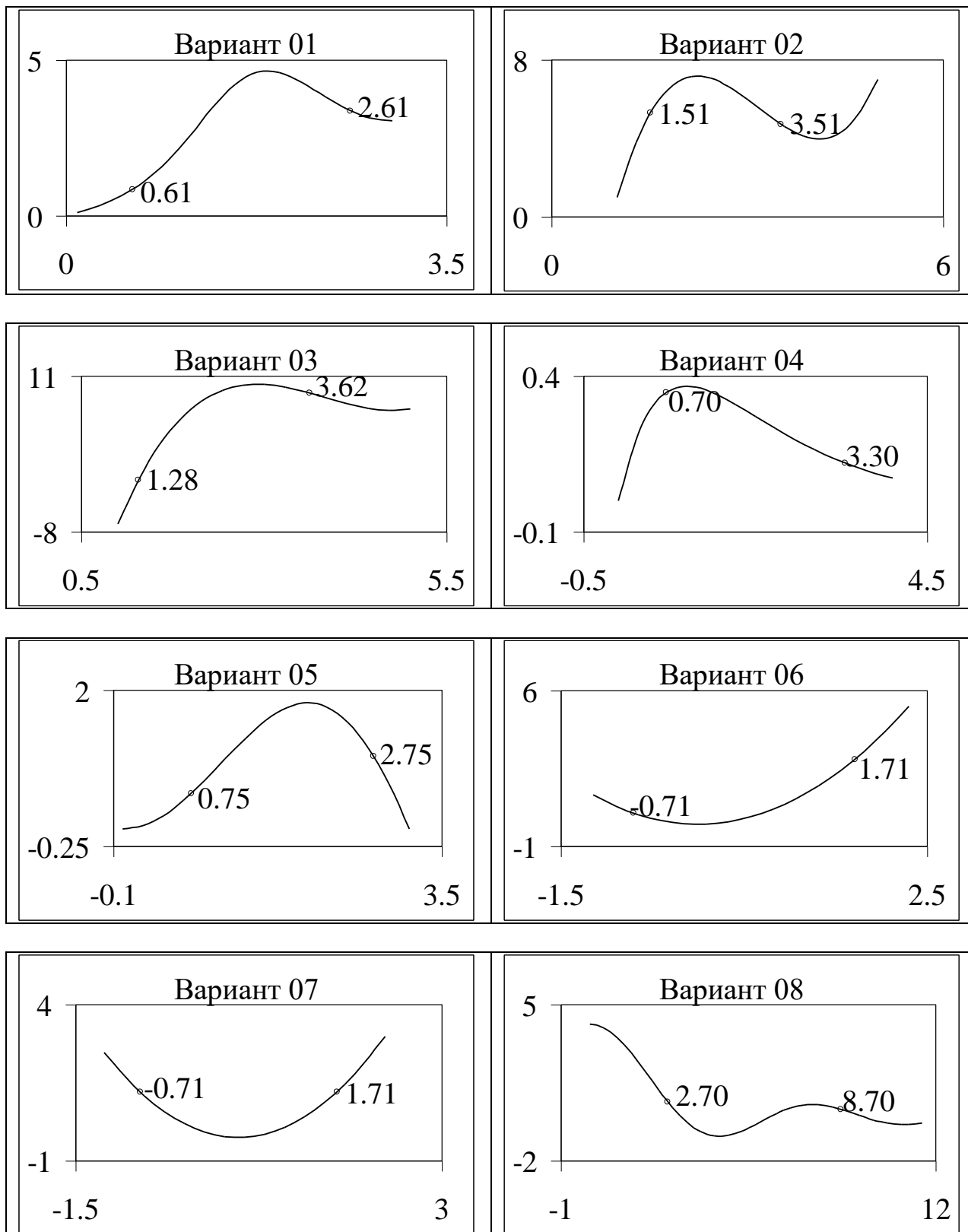
4. Отчёт о выполнении работы

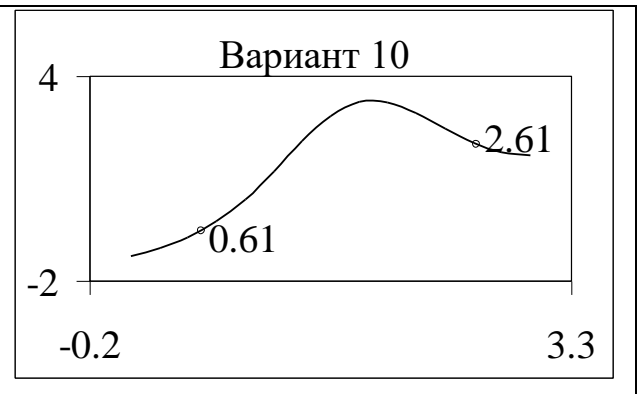
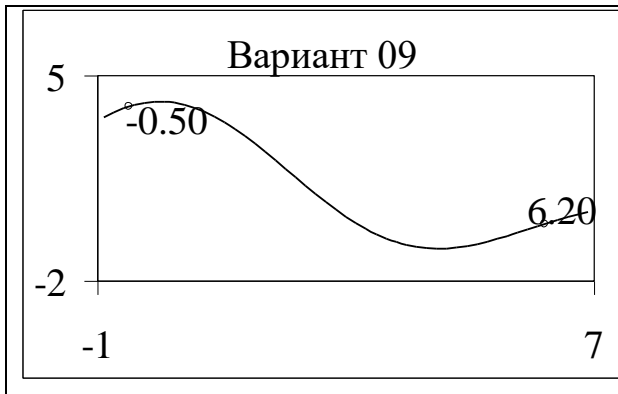
Отчёт о выполнении лабораторной работы должен быть выполнен на листах формата A4 и содержать следующие результаты:

1. Исходные данные (рис. 2);

2. Текст программы с подробными комментариями;
3. Характеристику точности обучения (рис. 3);
4. Результаты моделирования (рис. 4);
5. Сопоставление результатов в контрольных точках;
6. Краткие выводы о результатах работы.

Контрольные задания





Контрольные вопросы:

1. Дайте определение нейронной сети?
2. Чем характеризуется НС?
3. Из чего состоит простейшая сеть?
4. Перечислите недостатки однослойной сети?
5. Что называется процессом обучения НС?
6. Чем характеризуется результат самообучения?
7. Когда обучение НС считается законченным?
8. В чем заключается алгоритм обратного распространения ошибки?

Лабораторная работа № 4.

Планирование экспериментальных исследований. Методы статической оптимизации объектов исследования при наличии ошибок измерений на базе метода крутого восхождения (Бокса-Уилсона).

Целью работы является ознакомление с основами методов статической оптимизации объектов исследования при наличии ошибок измерений на базе метода крутого восхождения (Бокса-Уилсона).

ТЕОРЕТИЧЕСКОЕ ОПИСАНИЕ РАБОТЫ

При поиске оптимального сочетания факторов, действующих в объекте исследования, то есть наилучших значений в конкретных условиях, задача решается следующим образом.

Задаются критерием оптимизации y , зависящим от вектора управляемых факторов $X (x_1, x_2, \dots, x_k)$, $k = 1, 2, \dots, i$ - число факторов.

Задача оптимизации сводится к отысканию таких значений факторов, при которых целевая функция достигает экстремума (максимума или минимума).

Зависимость

$$y(\vec{X}) = f(x_1, x_2, \dots, x_k) \quad (2.1)$$

образует некоторую поверхность в $(k + 1)$ -мерном пространстве x_1, x_2, \dots, x_k, y . Эту поверхность называют поверхностью отклика, а отдельные её точки или значения y в точках \vec{X} факторного пространства – просто откликом.

В большинстве практических случаев аналитическая зависимость (2.1) неизвестна и единственное, чем располагает исследователь, – это возможностью наблюдать значение отклика при любой комбинации варьируемых факторов (x_1, x_2, \dots, x_k) . Поскольку такое наблюдение связано с проведением опытов и измерений, то фактически наблюдается сумма истинного значения критерия оптимизации $y_{ист}$ и случайной ошибки опыта ε , то есть $y_{набл} = y_{ист} + \varepsilon$.

При оптимизации объекта исследования методом крутого восхождения (МКВ) рабочее движение совершается в направлении наиболее быстрого возрастания критерия оптимизации (при поиске максимума) или быстрого убывания (при поиске минимума), то есть в направлении градиента. МКВ представляет собой метод экспериментальной оптимизации, сочетающей полный или дробный факторный эксперимент с движением по градиенту функции отклика.

Из аналитической геометрии известно, что для многомерной функции градиент задается выражением:

$$\text{grad } \hat{y} = \frac{\partial y}{\partial x_1} \vec{i}_1 + \frac{\partial y}{\partial x_2} \vec{i}_2 + \dots + \frac{\partial y}{\partial x_k} \vec{i}_n$$

(2.2)

где \vec{e}_i направляющий вектор координатной оси x_i .

Координатами вектора

$$\text{grad } \hat{y} = \frac{\partial y}{\partial x_1} + \frac{\partial y}{\partial x_2} + \dots + \frac{\partial y}{\partial x_k},$$

(2.3)

служат координаты модели регрессии при линейных членах разложения

функции y в ряд Тэйлора по степеням x_i ($i = 1, 2, \dots, k$). Тогда соответствующие компоненты вектора градиента могут быть получены как коэффициенты

b_1, b_2, \dots, b_k линейной аппроксимации поверхности отклика вблизи исходной точки X :

$$1) \quad y(\vec{X}) \cong b_0 + b_1 x_1 + b_2 x_2 + \dots + b_k x_k.$$

2) (2.4)

Для получения оценок b_1, b_2, \dots, b_k необходимо выполнить полный

факторный эксперимент (ПФЭ) или дробный факторный эксперимент (ДФЭ) с центром в базовой точке k_0 (рисунок 2.1), на котором представлена графическая иллюстрация МКВ на двухфакторном эксперименте. Цифрами 20, 30, 40 и т.д. обозначены линии равного уровня поверхности отклика y в некоторых относительных единицах.

В МКВ для поиска экстремальной области S процесс осуществляется в 2...3 цикла и в первом цикле выполняются:

1. Выбирается базовая точка k_0 (рисунок 2.1). Координаты базовой точки выбираются произвольно на основании предварительного изучения объекта исследования, но в области планирования факторов.

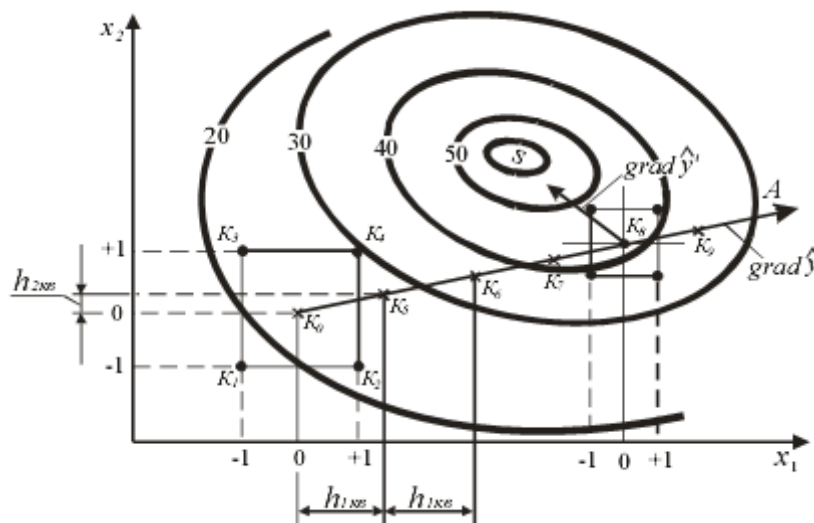


Рисунок 2.1 – Иллюстрация метода кругого восхождения

2. Вокруг базовой точки выполняется ПФЭ или ДФЭ. При этом факторы нормируются по формуле:

$$x_i = \frac{\tilde{x}_i - \tilde{x}_{i,0}}{\Delta \tilde{x}_i},$$

i) (2.5)

ii)

где x_i - нормированное (безразмерное) значение i -го фактора;

\tilde{x}_i - натуральное (с единицами измерения) значение i -го фактора на верхнем и нижнем уровнях; ⁱ

$x_{i,0}$ - натуральное значение i -го фактора на нулевом уровне;

Δx_i - интервал варьирования i -го фактора. В МКВ рекомендуется ориентировочно выбирать интервал варьирования 0,05...0,3 от области планирования.

При выполнении ПФЭ или ДФЭ назначается число m параллельных опытов в каждой точке плана, которые рандомизируются.

3. Вычисляются оценки коэффициентов регрессии нормированной

модели $\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + \dots + b_k x_k$ по формулам:

$$b_0 = \frac{\sum_{u=1}^N \bar{y}_u}{N},$$

(2.6)

где y_u - среднее значение отклика в u -ой строке матрицы плана ($u = 1, 2, \dots, N$);

N - число строк матрицы плана (МП).

$$b_i = \frac{\sum_{u=1}^N x_{iu} \cdot \bar{y}_u}{N},$$

(2.7)

где x_{iu} - значение i -го фактора в u -ой строке матрицы плана (факторы

нормированы).

4. Важной особенностью МКВ является проведение статистического анализа промежуточных результатов на пути к оптимуму, который включает:

- а) проверку однородности наблюдений в строках МП;
- б) проверку значимости оценок коэффициентов b_i линейной модели объекта;
- в) проверку адекватности полученной модели объекту.

5. Проверка однородности наблюдений в строке МП состоит в проверке

гипотезы о равенстве дисперсий $\sigma^2\{y_1\} = \sigma^2\{y_2\} = \dots = \sigma^2\{y_n\}$ при

экспериментах в соответствующих строках МП. Оценки дисперсий в u -ой строке МП находятся по формуле:

$$s_u^2 = \frac{1}{m-1} \sum_{j=1}^m (y_{uj} - \bar{y}_u)^2 \quad (2.8)$$

Проверка однородности выполняется по G -критерию Кохрена (см. лабораторную работу №1), для чего вычисляется

$$G_{\max} = \frac{s_{u\max}^2}{\sum_{u=1}^N s_u^2}, \quad (2.9)$$

где

s_u^2 - дисперсия в u -ой строке МП;

$s_{u\max}^2$ - максимальная из всех дисперсий.

Если вычисленное значение критерия G_{\max} окажется меньше $G_{кр}$,

найденного по таблице 3 приложения для $f_1 \ll m \ll 1$ и $f_2 \ll N$ и выбранного уровня значимости q , то гипотеза об однородности дисперсий принимается.

Если проверка на однородность дала отрицательный результат, то остается признать, что управляемые и неконтролируемые факторы создают на выходе большой уровень шума. При этом следует попытаться увеличить число параллельных опытов. Существуют и другие способы преобразования данных.

- б. После определения оценок коэффициентов регрессии необходимо проверить гипотезу о значимости коэффициентов b_i . Проверка гипотезы проводится с помощью критерия Стьюдента, который при проверке нуль-гипотезы формируется как

$$t = \frac{|b_i|}{s\{b_i\}}, \quad (2.10)$$

где $s\{b_i\}$ - среднеквадратичное отклонение ошибки определения коэффициента b_i :

$$s\{b_i\} = \frac{s\{y\}}{\sqrt{N \cdot m}} \quad (2.11)$$

где $s\{y\}$ - среднеквадратичное отклонение ошибки опытов и при равномерном дублировании опытов в строках МП определяется по формуле:

$$s^2\{y\} = \frac{\sum_{m=1}^N \sum_1^m (y_m - \bar{y}_m)^2}{N \cdot (m-1)}, \quad (2.12)$$

где y_{iu} - число дублирующих опытов в u -ой строке МП.

Если расчетная величина t превышает значение $t_{кр}$, определенное по таблице 1 приложения для числа степеней свободы $f = N(m-1)$ при заданном уровне значимости q , то нуль-гипотеза отвергается и коэффициент b_i считается значимым. В противном случае нуль-гипотеза принимается и коэффициент b_i считают статистически незначимым (то есть $b_i=0$).

Статистическая незначимость коэффициента b_i обусловлена следующими причинами:

1. Уровень базового региона МП близок к точке частного экстремума по фактору x_i .
2. Интервал варьирования $\Delta \tilde{x}$ выбран малым.
3. Данный фактор не имеет функциональной связи с откликом y .
4. Величина ошибки эксперимента велика вследствие наличия неуправляемых и неконтролируемых факторов.

Если какой-либо коэффициент b_i оказался незначимым, он может быть отброшен без пересчета остальных, так как МП ортогональна и оценки коэффициентов регрессии определяются независимо друг от друга.

Модель регрессии объекта исследования составляется только для значимых коэффициентов.

7. Чтобы проверить гипотезу об адекватности представления результатов эксперимента найденной моделью регрессии, необходимо оценить отклонение предсказанной моделью \hat{y}_u от результатов эксперимента y_u в тех же точках (строках МП).

Дисперсия адекватности $s_{ад}^2$ при равномерном дублировании опытов в строках МП определяется по формуле:

$$s_{ад}^2 = \frac{m}{N-d} \sum_{u=1}^N (\bar{y}_u - \hat{y}_u)^2, \quad (2.13)$$

где m - число параллельных опытов в строках МП;

d - число оценок коэффициентов регрессии модели, включая и b_0 ;

\hat{y}_u - расчетное значение критерия оптимизации в u -ой строке МП.

Проверка гипотезы об адекватности проводится с помощью F - критерия Фишера, который позволяет проверить нуль-гипотезу о равенстве

двух дисперсий s_{ad}^2 и $s^2\{y\}$.

F -критерий вычисляется как отношение

$$F = \frac{s_{ad}^2}{s^2\{y\}}. \quad (2.14)$$

Если вычисленное значение F меньше $F_{кр}$, найденного по таблице 2

приложения для числа степеней свободы числителя $f_{числ} = f_{ад} = N - d$ и $f_{знам} = f_y = N - 1$ при заданном уровне значимости q , то нуль-гипотеза принимается и модель считается адекватной и наоборот.

Если гипотеза адекватности отвергается, необходимо переходить к более сложной форме модели регрессии или проводить эксперименты, при возможности, с меньшим интервалом варьирования. Однако с уменьшением интервала варьирования появляются другие трудности:

1. Растет отношение помех (шума) к полезному сигналу, что приводит к необходимости увеличивать число дублирующих опытов для выделения полезного сигнала на фоне шума.
2. Уменьшаются значения коэффициентов b_i . Коэффициенты могут стать статистически незначимыми.

Если линейная модель регрессии после статистической обработки оказалась не адекватной, то её можно дополнить эффектами парных взаимодействий и она будет иметь вид (неполная полиномиальная модель):

$$\hat{y}(\bar{X}) \cong b_0 + b_1x_1 + \dots + b_kx_k + b_{12}x_1x_2 + \dots + b_{ij}x_ix_j,$$

где b_{ij} - оценки коэффициентов регрессии, характеризующие парные взаимодействия.

Они вычисляются по формуле:

$$b_{ij} = \frac{\sum_{i \neq j} x_{iu} \cdot x_{ju} \cdot \bar{y}_u}{N}, \quad (2.16)$$

где x_{iu} и x_{ju} - значения i -го и j -го факторов в u -ой строке МП.

Статистическая обработка не полной полиномиальной модели (2.15) (оценка однородности наблюдений, значимости оценок коэффициентов регрессии, адекватности модели) осуществляется аналогично и по тем же аналитическим формулам, что и линейной модели (2.4).

8. В МКВ после получения модели регрессии (2.4) начинается шаговое движение в область экстремума по градиенту функции отклика.

Движение начинается из базовой точки K_0 (рисунок 2.1).

При этом вычисляется произведение $b_i \Delta \tilde{x}_i$, где $\Delta \tilde{x}_i$ - интервал варьирования i -го фактора. Фактор, для которого это произведение максимально, принимается за базовый ($b_0 \Delta \tilde{x}_0$).

На примере однофакторного эксперимента (рисунок 2.2) проследим шаговое движение по градиенту.

Известно, что первая производная в геометрической интерпретации представляет собой тангенс угла наклона касательной (градиента) к кривой и можно записать:

$$b_i \Delta \tilde{x}_i \approx \text{tg} \varphi.$$

Значение функции в точках, лежащих на градиенте, можно

определить, допустим, $y \approx OC \approx \text{tg} \varphi \cdot x_i \approx \text{tg} \varphi \cdot b_i \Delta \tilde{x}_i$.

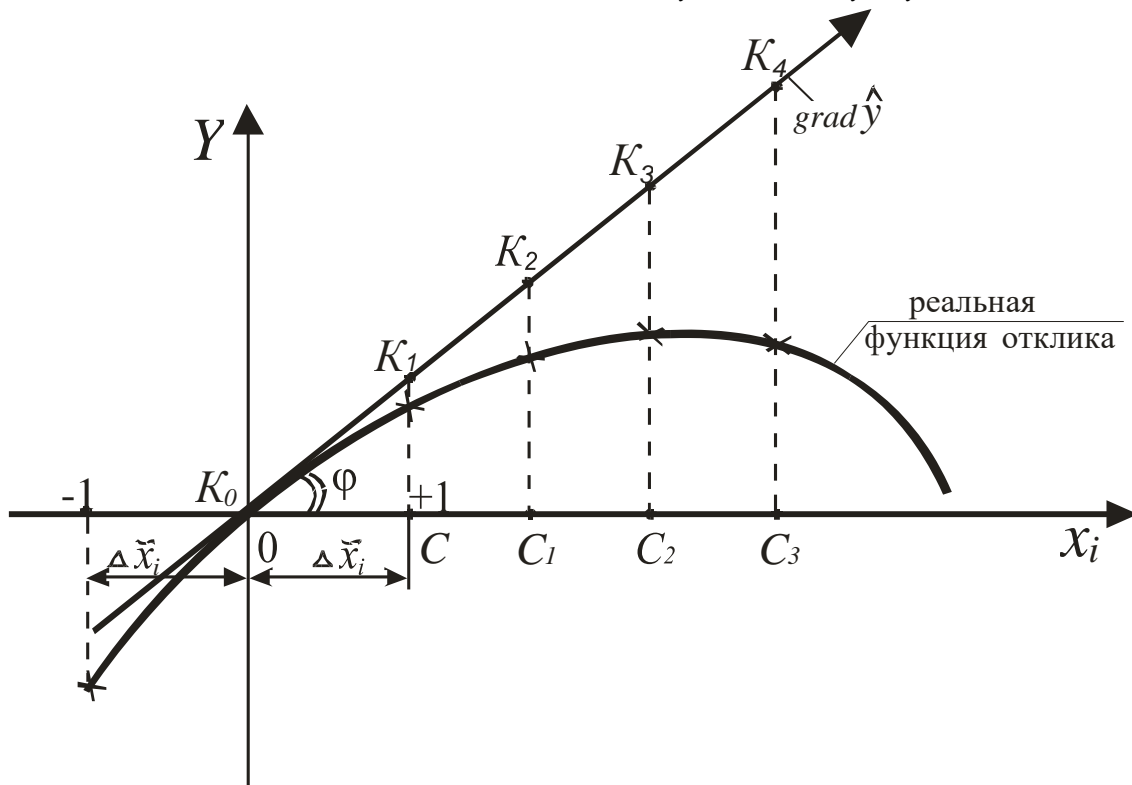


Рисунок 2.2 – Шаговое движение по градиенту

Из рисунка видно, что движение по градиенту (при поиске максимума функции отклика) представляет собой шаговую процедуру, которая продолжается до тех пор, пока последующие результаты не будут получаться ниже предыдущих.

В точке K_4 движение по градиенту прекращается.

В случае многомерной функции отклика при числе факторов $k = 1, 2, \dots, i$ назначается шаг движения по градиенту для базового фактора, а так как все факторы

связаны в математическую модель, то при шаговом движении по одному из них, должны изменяться и другие, пропорционально

произведению коэффициента b_i , умноженного на интервал варьирования

фактора $\Delta \tilde{x}_i$, то есть $b_i \Delta \tilde{x}_i$.

В реальных условиях на изучаемом объекте исследования всегда имеем дело с факторами, выраженными в реальном масштабе, а в модели регрессии факторы нормированы (безразмерные). При оценке составляющих градиента следует учитывать интервал варьирования факторами.

Для получения шагов других факторов h_i необходимо для каждого

фактора произведение $b_i \Delta \tilde{x}_i$ уменьшить во столько раз, во сколько уменьшится

$b_0 \Delta \tilde{x}_0$ для базового фактора при делении его на шаг, то есть

$$h_i = \frac{b_i \Delta \tilde{x}_i}{b_0 \Delta \tilde{x}_0} \cdot h_0, \quad (2.17)$$

где h_0 - шаг базового фактора, натуральное значение.

Знаки h_i должны соответствовать знакам b_i .

Выбор шага базового фактора не формализован. Если шаг принять очень малым, то в дальнейшем необходимо будет поставить много опытов для достижения области экстремума. При слишком же большом шаге имеется опасность «перешагнуть» через область экстремума. Для удобства выполнения опытов шаги факторов округляют. Для базового фактора шаг

варьирования принимают равным $h_i \approx \Delta x_0$ или вводят более мелкий, чтобы

первая рабочая точка не выходила за границы области планирования. Точка

K_5 на рисунке 2.1.

На расчет градиента свободный член модели регрессии b_0 не влияет.

Незначимые факторы стабилизируются на любом уровне в интервале $[0, 1]$.

Если при шаговом движении по градиенту каким-либо фактором невозможно варьировать, то он стабилизируется на лучшем уровне и движение осуществляется по оставшимся.

Шаговое движение по градиенту осуществляется от базовой точки K_0 . Для этого вычисляются координаты K -х рабочих точек ($K = 1, 2, \dots$) на

направлении градиента K_0A (рис.2.1) в реальном масштабе

$$\tilde{x}_{i,k} = \tilde{x}_{i,0} + K \cdot h_i, \quad (2.18)$$

в них последовательно выполняются проверочные (реальные) и мысленные опыты.

Мысленные опыты заключаются в получении предсказанных

(расчетных) значений отклика \hat{y}_{np} по модели регрессии (2.4). Для этого в уравнение (2.4) следует подставлять нормированные по уравнению (2.5) значения координат факторов.

Таки образом уравнение (2.4) принимает вид

$$\hat{y}_{np} = b_0 + b_1 \cdot k \frac{h_1}{\Delta \tilde{x}_1} + b_2 \cdot k \frac{h_2}{\Delta \tilde{x}_2} + \dots, \quad (2.19)$$

Где $k \in \{1, 2, \dots\}$ - число шагов.

Мысленные опыты позволяют сократить объём реальных опытов, то есть увеличивать скорость продвижения к экстремуму.

Реальные (проверочные) опыты в начале движения из базовой точки вдоль градиента KOA ставятся не в каждой рабочей точке, а через 2...4 точки. Затем, при заметном уменьшении приращений отклика в последующих реальных опытах проверочные опыты ставятся чаще.

Рабочее движение продолжается пока не будет достигнут частный экстремум (точка K_8 на рисунке 2.1). Признаком достижения частного экстремума (максимума) является уменьшение отклика в последующих проверочных опытах.

Точка частного экстремума на направлении KOA принимается за новую базовую точку, организуется второй цикл крутого восхождения (спуска). Порядок работы на втором цикле тот же, что и на первом. Различие

b) ~

состоит в том, что интервалы варьирования

Δx_i

при постановке ПФЭ или

ДФЭ и размер рабочих шагов

h_i назначаются меньшими, в связи с

увеличением кривизны с приближением к экстремуму. В случае необходимости выполняется третий цикл.

В МКВ поисковое движение прекращают после достижения области экстремума S , признаком этого является статистическая незначимость всех оценок коэффициентов регрессии b_i на последнем цикле МКВ.

Задание и порядок выполнения работы

1. Взять у преподавателя индивидуальный вариант результатов эксперимента для двухфакторной МП, уровни и интервал варьирования факторами, цель эксперимента (максимум или минимум).

2. Рассчитать оценки коэффициентов регрессии модели (2.4) для двух факторов. Оценить однородность наблюдений по критерию Кохрена, значимость оценок коэффициентов регрессии по критерию Стьюдента и адекватность модели по критерию Фишера. Построить линейную модель регрессии.

3. Выполнить процедуру крутого восхождения (спуска) по направлению градиента поверхности отклика с помощью мысленных опытов.

Мысленные опыты продолжаются до тех пор, пока выполняется

неравенство:

$$\hat{y}_{np} \leq (1...2) \cdot y_{max},$$

где

y_{max} - максимально возможный выход, определяемый из физических

соображений (задан в вариантах).

При поиске минимума:

$$\hat{y}_{np} \geq (1...2) \cdot y_{min},$$

где y_{min} - минимально возможный выход, определяемый из физических соображений (задан в вариантах).

Лабораторная работа №5. Оценка ошибок измерений. Статистическая проверка гипотез. Исключение грубых ошибок (промахов).

Целью работы является оценка ошибок измерений, возникающих при выполнении опытов в научных исследованиях, изучение методов расчета статистических характеристик случайных величин, а также методики применения статистических критериев для проверки гипотез и отыскания доверительных границ.

Теоретическое описание работы

В экспериментальных исследованиях необходимо знать и уметь рассчитывать ошибки измерений.

Численное значение физической величины получается в результате ее измерения, то есть сравнения ее с другой величиной того же рода, принятой за единицу.

Ошибкой измерения называется разность $x_i - x_0$ между результатом измерения x_i и истинным значением x_0 измеряемой величины.

Одной из важнейших задач математической обработки результатов опытов как раз и является оценка истинного значения измеряемой величины. Обычно неизвестно значение ошибок и истинного значения величины. Поэтому, ставится задача вычисления x_0 с минимальной ошибкой.

В зависимости от способа измерения ошибки подразделяют на прямые и косвенные.

Прямые – это те, которые получают непосредственно из опытных данных. Косвенные – получают по результатам прямых измерений путем вычисления по аналитическим выражениям. Например, определение мощности электрического тока по данным измерений тока и напряжения сети.

В данной работе рассматриваются только прямые измерения. Ошибки измерений классифицируются на 3 вида:

1. Грубые ошибки или промахи – это ошибки, являющиеся результатом низкой квалификации лица, производящего измерения, его небрежности или неожиданных сильных внешних воздействий на измерения. Промахи исключаются из обработки результатов после аналитической проверки.

2. Систематические ошибки вызываются факторами, действующими одинаковым образом при многократном повторении одних и тех же измерений. Например, после проведения серии измерений обнаруживается неправильная регулировка прибора на нулевую отметку.

В случае систематических ошибок их обнаруживают и вводят коррекцию в результаты наблюдений. Однако полностью эти ошибки всегда исключить не удается.

3. Случайные ошибки измерения вызываются большим количеством таких факторов, эффекты действия которых столь незначительны, что их нельзя учесть в отдельности. Случайную ошибку можно рассматривать как суммарный эффект действия таких факторов. Случайные ошибки являются неустраняемыми, но с помощью методов теории вероятностей их можно учесть и внести соответствующие поправки к истинному значению.

Если измерения выполняются неоднократно, то их результаты обозначим x_1, x_2, \dots, x_n ($n = 1, 2, \dots, i$), тогда разности $\Delta x_i = x_i - x_0$ называются абсолютной ошибкой.

Качество результатов измерений удобно характеризовать не абсолютной, а относительной ошибкой, которая равняется отношению абсолютной ошибки к истинному значению измеряемой величины в процентах, то есть:

$$\frac{\Delta x_i}{x_0} \cdot 100, \%$$

При измерениях физических величин, когда основную роль играют случайные ошибки, все оценки точности измерений можно сделать только с некоторой вероятностью. При этом часть ошибок будет положительной, часть – отрицательной, что очень неудобно.

Поэтому, ошибки измерений рассматривают как случайную величину в статистической совокупности.

Статистическая совокупность – это совокупность значений случайной переменной, которые получены в результате измерений и на практике исследователь всегда располагает лишь ограниченным числом значений случайной величины, представляющим собой некоторую выборку из генеральной совокупности.

По выборке могут быть рассчитаны выборочные статистические характеристики (выборочное среднее, выборочная дисперсия и т.п.), которые являются оценками соответствующих генеральных статистических характеристик (параметров).

Математическая статистика предъявляет к оценкам требования состоятельности, несмещенности и эффективности.

Оценка называется состоятельной, если с увеличением объема выборки она стремится (по вероятности) к оцениваемому параметру.

Оценка называется несмещенной, если ее математическое ожидание при любом объеме выборки равно оцениваемому параметру.

Оценка называется эффективной, если ее дисперсия минимально возможная из всех возможных оценок.

Для оценки истинного значения физической величины принимается ее среднее арифметическое значение, определяемое по формуле:

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{1}{n} \sum_{i=1}^n x_i, \quad (1.1)$$

где n - число повторностей опыта (число измерений).

В математической статистике доказано, что в большинстве случаев случайные ошибки подчиняются нормальному закону распределения.

Для оценки случайной ошибки измерений существует несколько способов. Наиболее распространена оценка с помощью среднего квадратичного отклонения (стандарта), определяемая по формуле:

$$s_i = \sqrt{\frac{(x_1 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n-1}} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}. \quad (1.2)$$

Квадрат этой величины называется дисперсией измерений. Дисперсию, а затем и среднюю квадратическую ошибку при работе на клавишных вычислительных машинах удобнее рассчитывать по алгебраически эквивалентной формуле, которую можно получить разложением суммы квадратов $\sum_{i=1}^n (x_i - \bar{x})^2$

$$\begin{aligned} \sum_{i=1}^n (x_i - \bar{x})^2 &= \sum_{i=1}^n (x_i^2 - 2x_i\bar{x} + \bar{x}^2) = x_1^2 - 2x_1\bar{x} + \bar{x}^2 + x_2^2 - 2x_2\bar{x} + \bar{x}^2 + \dots + x_n^2 - 2x_n\bar{x} + \bar{x}^2 = \\ &= (x_1^2 + x_2^2 + \dots + x_n^2) - 2\bar{x}(x_1 + x_2 + \dots + x_n) + (\bar{x}^2 + \bar{x}^2 + \dots + \bar{x}^2) = \sum_{i=1}^n x_i^2 - \frac{2\sum_{i=1}^n x_i}{n} \sum_{i=1}^n x_i + n \left(\frac{\sum_{i=1}^n x_i}{n} \right)^2 = \\ &= \sum_{i=1}^n x_i^2 - \frac{\left(\sum_{i=1}^n x_i \right)^2}{n}. \end{aligned}$$

Тогда

$$s_i^2 = \frac{\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2 / n}{n-1}. \quad (1.3)$$

В приведенных формулах необходимо учитывать, что при вычислении оценки дисперсии величина x принимается с тем же количеством значащих цифр, что и величины x_i .

Пусть α означает вероятность того, что результат измерений

отличается от истинного значения на величину, не большую, чем Δx . Это принято записывать в виде:

$$P(x - \Delta x \leq x_0 \leq x + \Delta x) = \alpha.$$

Вероятность называется доверительной вероятностью или

коэффициентом надежности. Интервал $x - \Delta x$ до $x + \Delta x$ называется доверительным интервалом.

Чем большей надежности мы требуем, тем большим получается доверительный интервал и наоборот: чем больше доверительный интервал мы задаем, тем вероятнее, что результаты измерений не выйдут за его пределы.

Таким образом, приходим к очень важному заключению: для характеристики случайной ошибки необходимо задать два числа, а именно: величину самой ошибки (или доверительного интервала) и величину доверительной вероятности. Задание

величины ошибки без доверительной вероятности не имеет смысла, так как имеется неопределенность в обозначении границ изменения случайной величины (ошибки).

Статистическая проверка гипотез

Под статистическими гипотезами понимаются некоторые предположения относительно значений генеральных статистических характеристик и генеральных распределений вероятности.

Проверка гипотезы заключается в сопоставлении некоторых статистических показателей, критериев проверки, вычисляемых по данным выборки, со значениями этих показателей, определенными теоретически в предположении, что проверяемая гипотеза верна.

Для критериев проверки выбираются надлежащие уровни значимости $q = 10\%$, 5% , 2% и т.д.), отвечающие событиям, которые при проводимых исследованиях считаются (с некоторым риском) практически невозможными. Следующим этапом является определение критической области применяемого критерия, вероятность попадания в которую в случае, если гипотеза верна, в точности равна уровню значимости.

Если q - уровень значимости, то вероятность попадания в область

допустимых значений равна $1 - \frac{q}{100}$.

Статистические критерии являются правилами, которые позволяют делать выводы о свойствах параметров генеральной совокупности, с

принятым уровнем значимости, на основании информации, полученной по данным выборок.

Рассмотрим те из них, которые наиболее часто применяются в теории планирования эксперимента (ТПЭ).

1. t - критерий Стьюдента применяется, когда необходимо сделать

вывод, равно ли математическое ожидание

$M\{x\}$ генеральной совокупности

некоторому предполагаемому значению C , или когда нужно построить

доверительный интервал для

$M\{x\}$.

Доказано, что случайная величина

$$t = \sqrt{N} \frac{\bar{x} - M\{x\}}{s\{x\}} = \frac{\bar{x} - M\{x\}}{s\{\bar{x}\}} \quad (1.4)$$

распределена по закону Стьюдента, если распределение x нормально.

Здесь N - объем выборки; $s^2\{x\}$ - среднее квадратическое отклонение, случайной величины x .

$s^2\{\bar{x}\}$ - среднее квадратическое отклонение среднего

$$s^2\{\bar{x}\} = \frac{s^2\{x\}}{N}.$$

На рисунке 1.1 показаны кривые дифференциального закона распределения $\varphi(t)$ для различных степеней свободы $f = N - 1$, по которым вычисляется несмещенная оценка дисперсии $s^2\{x\}$. При $N \rightarrow \infty$ кривая $\varphi(t)$ приближается к кривой нормированного нормального распределения.

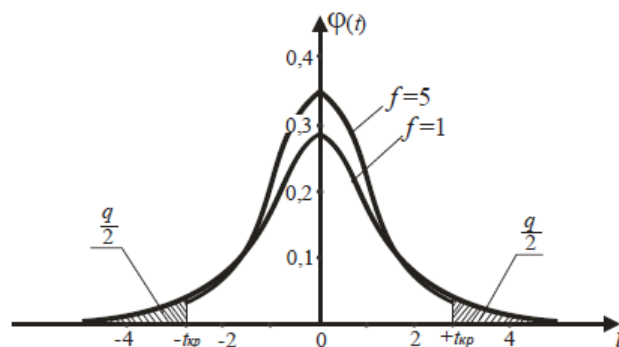


Рисунок 1.1 - Дифференциальные кривые t-распределения Стьюдента

Из рисунка 1.1 видно, что t - распределение симметрично относительно $t = 0$, поэтому в таблицах, где приводятся критические значения $t_{кр} = t_{q,f}$ для принятого уровня значимости q и имеющегося числа степеней свободы f , задаются лишь положительные $t_{кр}$ (приложение 1).

Если при расчете по формуле (1.4) окажется, что $t < t_{кр}$, то делается вывод, что гипотеза $M = \mu$ не противоречит результатам наблюдений при принятом уровне значимости q .

Расчет доверительного интервала выполняется по формуле:

$$\Delta x = \frac{1}{\sqrt{N}} \cdot t_{кр} \cdot s\{x\}. \quad (1.5)$$

2. F - критерий Фишера называется дисперсионным отношением, так как он формируется как отношение двух сравниваемых несмещенных оценок дисперсий:

$$F = \frac{s_1^2\{x\}}{s_2^2\{x\}} > 1, \quad (1.6)$$

причем в числитель ставится большая из двух дисперсий. Расчетное F сравнивается с критическим $F_{кр} = F_{q, f_1, f_2}$, где F_{q, f_1, f_2} находится из приложения 2 для степеней свободы:

$$f_1 = f_{\text{числителя}} = N_1 - 1,$$

$$f_2 = f_{\text{знаменателя}} = N_2 - 1,$$

где N_1 – число элементов выборки, по которым вычислена s_1^2 , а N_2 – число элементов выборки, по которым получена s_2^2 , причем f_1 – находится в горизонтальном заголовке таблицы, а f_2 – слева в вертикальном заголовке, для принятого уровня значимости q .

На рисунке 1.2 показаны кривые F – распределения Фишера.

Заштрихована область критических значений F .

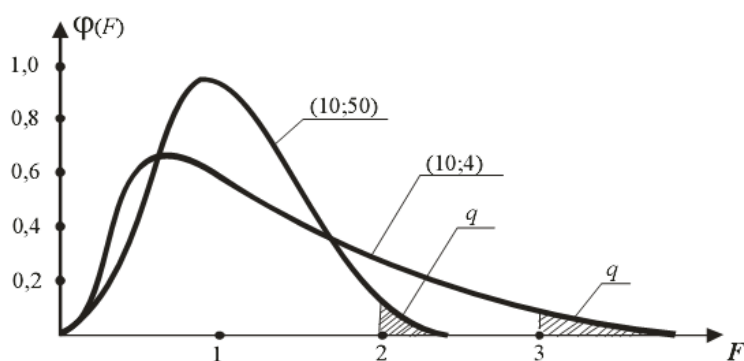


Рисунок 1.2 - Дифференциальные кривые F -распределения Фишера

При проверке нуль-гипотезы ($\sigma_1^2 = \sigma_2^2$) о равенстве двух генеральных дисперсий, вычисляется F , вычисляются s_1^2 и s_2^2 по данным выборок и сравнивается с табличным значением $F_q(f_1, f_2)$. Если найденное из наблюдений значение F -критерия оказывается меньше табличного, $F_q(f_1, f_2)$, то нуль-гипотеза о равенстве двух дисперсий не отвергается.

Когда с нуль-гипотезой конкурирует альтернативная гипотеза

($\sigma_1^2 \neq \sigma_2^2$), то принято отбрасывать нуль-гипотезу, если F расчетное превосходит верхнее критическое значение. При этом вся критическая область критерия будет отвечать уровню $\frac{q}{100}$.

Если вычисленное по уравнению (1.6)

$$F \leq F_{кр},$$

то принимается нуль-гипотеза о равенстве генеральных дисперсий

$\sigma_1^2\{x\} = \sigma_2^2\{x\}$, при принятом уровне значимости.

3. G – критерий Кохрена применяется для оценки однородности

несмещенных оценок дисперсий, вычисленных по одинаковому числу N наблюдений. Он формируется как отношение максимальной из сравниваемых дисперсий к сумме всех k – дисперсий:

$$G_{max} = \frac{s_j^2\{x\}_{max}}{\sum_{j=1}^k s_j^2\{x\}}$$

Если

$$G_{max} \leq G_{кр} = G_{q, f_1, f_2},$$

то выборочные дисперсии однородны.

Распределение G табулировано (таблица 3 приложения). Числа

степеней свободы, по которым выбирается из таблиц

$G_{кр}$, равны:

$$f_1 = f_{числителя} = N_1 - 1,$$

$$f_2 = f_{знаменателя} = k,$$

где f_1 – число степеней свободы числителя;

f_2 – число степеней свободы знаменателя и определяется числом

всех сравниваемых дисперсий.

Порядок применения G – критерия

1. По данным выборок вычисляются дисперсии

$$s_k^2\{x\};$$

2. Из всех дисперсий выбирается s_{max} ;

3. Вычисляется сумма $\sum_1^k s_i^2\{x\}$;

4. Рассчитывается эмпирическое $G_{max} \{ G_{кр} = G_{q, f_1, f_2} \}$.

Если это условие выполняется, то дисперсии однородны.

Проверка гипотезы о равенстве математического ожидания заданному значению C

Часто встречающаяся задача статистической проверки гипотез заключается в том, что требуется определить равенство математического ожидания m_x какому-либо постоянному числу C , то есть $m_x \approx C$. Критерием

проверки принимается величина

$$t = \sqrt{N} \left(\frac{\bar{x} - C}{s\{x\}} \right), \quad (1.7)$$

называемая t – критерием Стьюдента с $f \approx N - 1$ степенями свободы, где N – объем выборки.

Если вычисленное значение t – критерия не превышает $t_{кр} \approx t_{q, f}$, найденного по таблице 1 приложения по заданному уровню значимости q и числу степеней свободы f , то исходная гипотеза $m_x \approx C$ принимается.

a. Определение необходимого числа повторностей опытов (объема выборки, числа параллельных опытов)

При обычных технических измерениях бывает достаточным принять доверительный интервал $\Delta x \approx 0,05 \dots 0,1 \Delta x$ и доверительную вероятность

$P \approx 0,95$, а в ответственных случаях $P \approx 0,99$.

Если в распоряжении исследователя имеется значение среднего квадратичного отклонения $s\{x\}$, то необходимое число измерений

определяется из Δx по формуле

$$N = \frac{t_{кр}^2 S^2\{x\}}{\Delta x^2}, \quad (1.8)$$

где $t_{кр}$ – критическое значение критерия Стьюдента, принятое по таблице 1 приложения с числом степеней свободы, с которым вычислена $s^2\{x\}$ и принятом уровне значимости q .

b.

c. Исключение грубых ошибок (промахов)

Если в исследуемом ряду имеются сильно отклоняющиеся результаты

наблюдений, то необходимо произвести проверку принадлежности подозреваемого измерения к исследуемому ряду.

Для исключения грубых ошибок используют t – критерий Стьюдента:

$$\frac{x_{\min}^{\max} - \bar{x}}{s\{x\}} = t,$$

(1.9)

где x_{\min}^{\max} – результаты, которые ставятся под сомнение;

\bar{x} – среднее значение, вычисленное без подозреваемого наблюдения;

$s\{x\}$ – среднеквадратичное отклонение, вычисленное без x_{\min}^{\max} .

Вначале вычисления проводят с наблюдением x_{\min}^{\max} , а затем с x_{\max}^{\min} .

Если расчетное значение $t \leq t_{q,f}$ принятое по таблице 1 приложения, то результат исключается.

Если $t > t_{q,f}$ то результат не исключается.

Задание и порядок выполнения работы

1. Взять у преподавателя индивидуальный вариант – две выборки опытов по 10 случайных чисел.

2. Результаты опытов и расчетов записать карандашом в таблицу отчета.

Фамилию студента и номер группы записать ручкой.

а) Найти для каждой выборки оценки математического ожидания и

дисперсии (\bar{x} и s^2).

б) По одной из выборок найти доверительные границы математического ожидания (с доверительной вероятностью $P = 0,95$).

в) Проверить гипотезу о равенстве нулю математического ожидания по одной из выборок (с уровнем значимости $q = 5\%$).

г) Проверить гипотезу о равенстве дисперсий в двух полученных выборках (с уровнем значимости $q = 5\%$).

д) Рассчитать необходимое число повторностей измерений для одной из выборок (с уровнем значимости $q = 5\%$ и доверительным интервалом в пределах 5...10 % от среднего значения).

ж) Для сильно отклоняющихся наблюдений по данным одной из выборок произвести проверку принадлежности подозреваемых наблюдений к исследуемому ряду (с уровнем значимости $q = 5\%$).

Лабораторная работа №6.
Изучение методов планирования эксперимента для получения модели регрессии второго порядка. Статистическая обработка моделей регрессии при планировании второго порядка

Целью работы является изучение методов планирования эксперимента для получения модели регрессии второго порядка.

ТЕОРЕТИЧЕСКОЕ ОПИСАНИЕ

Если линейная модель регрессии (2.4) в лабораторной работе № 2 оказалась неадекватной результатам эксперимента, то зависимость критерия оптимизации от входных факторов необходимо искать в виде полинома более высокого порядка. В технических исследованиях с достаточной точностью

эти зависимости аппроксимируются полиномом второго порядка:

$$\hat{y} = b_0 + \sum_{i=1}^k b_i x_i + \sum_{\substack{i \neq j \\ i < j}} b_{ij} x_i x_j + \sum_{i=1}^k b_{ii} x_i^2, \quad (3.1)$$

где i – число факторов ($i = 1, 2, \dots, k$);

b_0, b_i, b_{ij}, b_{ii} – оценки коэффициентов регрессии модели второго порядка.

Для получения модели второго порядка (3.1) пользуются результатами эксперимента, проведенного по плану второго порядка (при активном эксперименте).

Планы второго порядка отличаются от линейных тем, что факторы варьируют на нескольких уровнях, минимум на трех.

При построении планов второго порядка используют центральные композиционные (то есть строящиеся последовательно) планы (ЦКП) включающие точки полного или дробного факторного эксперимента типа 2^k ,

«звездные точки», расположенные по факторным осям симметрично относительно центра плана. Общее число точек ЦКП второго порядка определяется формулой: $N = 2^k + 2k + k_0$, где k_0 – число нулевых (центральных точек).

При построении планов используют различные критерии оптимальности.

Наиболее широкое распространение получили следующие планы: ортогональные, ротатабельные и D – оптимальные.

При ортогональном планировании оценки коэффициентов модели регрессии (3.1) получаются независимыми друг от друга с минимальными дисперсиями. Факторы с незначимыми коэффициентами можно сразу отбрасывать без пересчета оставшихся значимых оценок коэффициентов, как это необходимо при неортогональных планах.

Однако при ортогональном планировании второго порядка точность предсказания критерия оптимизации при повороте факторных осей неодинакова. Наилучшим с этой точки зрения служит такое математическое описание, которое дает одинаковую точность предсказания критерия

□
оптимизации U во всех направлениях на одинаковом расстоянии от центра

плана. Такое математическое описание получается при ротатабельном композиционном планировании.

Точность оценивания коэффициентов модели регрессии характеризуется эллипсоидом рассеяния их оценок. Планирование, при котором требуется, чтобы объем эллипсоида расстояния оценок коэффициентов был минимальным, называется D – оптимальным.

В настоящее время в теории планирования эксперимента разработаны методы построения всех типов планов, формулы для расчета оценок коэффициентов регрессии и статистическая обработка результатов эксперимента.

Свойством ротатабельности обладает и ряд некомпозиционных планов. На рисунке 3.1 показан ротатабельный план второго порядка для двух факторов в виде правильного шестиугольника (план на шестиугольнике).

Особенности данного плана заключаются в том, что один фактор (x_2) варьирует на трех уровнях, а другой (x_1) – на пяти уровнях. В некоторых конкретных ситуациях это бывает удобно.

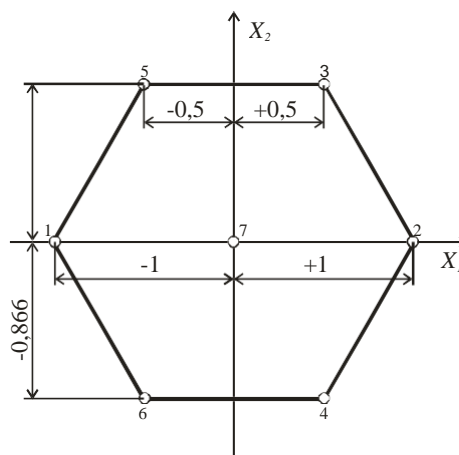


Рисунок 3.1 – Наименьший ротатабельный план на шестиугольнике для двух факторов

Факторы нормированы. Матрица некомпозиционного ротатабельного плана второго порядка для двух факторов представлена в таблице 3.1.

Таблица 3.1 – Матрица наименьшего ротатабельного плана второго порядка для двух факторов

Номера точек	План			Расчетная матрица			Критерий оптимизации
	x_0	x_1	x_2	$x_1 x_2$	x_1^2	x_2^2	\bar{y}_n
1	+ 1	- 1	0	0	1	0	\bar{y}_1
2	+ 1	+ 1	0	0	1	0	\bar{y}_2
3	+ 1	0,5	0,866	0,433	0,25	0,75	\bar{y}_3
4	+ 1	0,5	-0,866	-0,433	0,25	0,75	\bar{y}_4
5	+ 1	-0,5	0,866	-0,433	0,25	0,75	\bar{y}_5
6	+ 1	-0,5	-0,866	0,433	0,25	0,75	\bar{y}_6
7	+ 1	0	0	0	0	0	\bar{y}_0

Оценки коэффициентов регрессии модели (3.1) рассчитываются по следующим формулам.

Свободный член

$$b_0 = \bar{y}_0, \quad (3.2)$$

где \bar{y}_0 – среднее по результатам опытов в центре плана.

Линейные оценки коэффициентов регрессии факторов:

$$b_j = T_3 \sum_{n=1}^N x_{jn} \bar{y}_n, \quad (3.3)$$

где T_3 – расчетный коэффициент, при $k = 2$ $T_3 = 0,33333$.

Оценки коэффициентов при парных взаимодействиях:

$$b_{ij} = T_6 \sum_{n=1}^N x_{in} x_{jn} \bar{y}_n, \quad (3.4)$$

где T_6 – расчетный коэффициент. При $k = 2$ $T_6 = 1,33333$.

Оценки коэффициентов квадратичных эффектов факторов:

$$b_{ii} = T_4 \sum_{n=1}^N x_{in}^2 \bar{y}_n + T_5 \sum_{i=1}^k \sum_{n=1}^N x_{in}^2 \bar{y}_n - T_2 \sum_{n=1}^N x_0 \bar{y}_n, \quad (3.5)$$

где T_2, T_4, T_5 – расчетные коэффициенты. При $k = 2$ $T_2 = 1,0$; $T_4 = 0,66667$;

$$T_5 = 0,83333.$$

Статистическая обработка моделей регрессии при планировании второго порядка

Статистическая обработка результатов планирования второго порядка производится по аналогии с линейными планами (см. лабораторная работа № 2), но имеет некоторые особенности.

Например, вычисление оценок дисперсий производится отдельно для различных групп коэффициентов (b_0, b_i, b_{ij} и b_{ii}).

Вначале осуществляется проверка однородности наблюдений в строках МП с помощью критерия Кохрэна, для чего вычисляется:

$$G_{max} = \frac{S_{u_{max}}^2}{\sum_{u=1}^N S_u^2}, \quad (3.6)$$

где S_u^2 – дисперсия в u -ой строке МП:

$$s_u^2 = \frac{1}{m-1} \sum_{j=1}^m (y_{uj} - \bar{y}_u)^2. \quad (3.7)$$

Здесь m – число параллельных опытов в строках МП.

Выражения (3.6) и (3.7) справедливы, когда число параллельных опытов одинаково в строках МП.

Если вычисленное по выражению (3.6) значение G_{max} окажется меньше критического $G_{кр}$, найденного по таблице 3 приложения для $f_1 = m - 1$

и $f_2 = N$ и выбранного уровня значимости, то гипотеза об однородности наблюдений принимается.

Проверка гипотезы о значимости оценок коэффициентов регрессии производят с помощью t – критерия Стьюдента:

$$t_i = \frac{|b_i|}{s\{b_i\}}, \quad (3.8)$$

где $s\{b_i\}$ – среднеквадратическое отклонение коэффициента b_i .

При планировании второго порядка $s\{b_i\}$ определяется отдельно для разных групп оценок (b_0, b_i, b_{ij} и b_{ii}) коэффициентов и оценки $s\{b_i\}$ в уравнении

(3.8) будут одинаковы только внутри данной группы оценок коэффициентов.

Дисперсии для оценок коэффициентов регрессии определяются по

формулам (для наименьшего ротатабельного плана второго порядка для двух факторов):

$$s\{b_0\} = T_7 s\{y\} \quad (3.9)$$

где T_7 – расчетный коэффициент, при $k = 2$ $T_7 = 1,0$;

$s\{y\}$ – среднеквадратичное отклонение ошибки опытов и при равномерном дублировании опытов в строках МП определяется по формуле:

$$s\{y\} = \sqrt{\frac{\sum_{i=1}^N \sum_{j=1}^m (y_{ij} - \bar{y}_i)^2}{N(m-1)}},$$

$$s\{b_i\} = T_8 s\{y\}, \quad (3.10)$$

$$s\{b_i\} = T_8 s\{y\}, \quad (3.11)$$

где T_8 – расчетный коэффициент и при $k = 2$ $T_8 = 0,57735$.

$$s\{b_{ij}\} = T_{10} s\{y\}, \quad (3.12)$$

где T_9 – расчетный коэффициент и при $k = 2$ $T_9 = 1,1547$.

$$s\{b_{ii}\} = T_{10} s\{y\}, \quad (3.13)$$

где T_{10} – расчетный коэффициент и при $k = 2$ $T_{10} = 1,22474$.

Если расчетная величина по выражению (3.8) t – превышает значение, определенное по таблице 1 приложения для числа степеней свободы $f = N(m-1)$ при заданном уровне значимости q , то коэффициент считается значимым.

Для проверки гипотезы об адекватности представления результатов эксперимента моделью второго порядка необходимо оценить

отклонение предсказанной моделью \hat{y}_u в тех же строках МП.

Дисперсия адекватности $s_{\text{ад}}^2$ при равномерном дублировании опытов, определяется по формуле:

$$s_{\text{ад}}^2 = \frac{m}{N-d} \sum_{i=1}^N (\bar{y}_i - \hat{y}_i)^2, \quad (3.14)$$

где d – число оценок коэффициентов регрессии модели, включая и b_0 ;

\hat{y}_u – расчетное значение критерия оптимизации в u -ой строке МП.

Проверка адекватности выполняется по F – критерию Фишера, который вычисляется как отношение:

$$F = \frac{s_{\text{ад}}^2}{s^2\{y\}},$$

(3.15)

Если вычисленное по выражению (3.15) значение F меньше $F_{кр}$, найденного по таблице 2 приложения для числа степеней свободы числителя $f_{числ} = N - d$ и $f_{знам} = N(m-1)$ при заданном уровне значимости q , то модель считается адекватной и наоборот.

ЗАДАНИЕ И ПОРЯДОК ВЫПОЛНЕНИЯ РАБОТЫ

1. Взять у преподавателя индивидуальный вариант результатов эксперимента по наименьшему ротатабельному плану для построения модели регрессии второго порядка, уровни и интервал варьирования факторами для двух факторов.

2. Рассчитать оценки коэффициентов регрессии модели (3.1) для двух факторов. Оценить однородность наблюдений по критерию Кохрэна, значимость оценок коэффициентов регрессии по критерию Стьюдента и адекватность модели по критерию Фишера. Построить квадратичную модель регрессии.

Лабораторная работа №7. Расчёт аппроксимации экспериментальных данных с использованием Microsoft Excel.

Теоретические сведения

Аппроксимация – это процесс подбора эмпирической функции $\varphi(x)$ для установления из опыта функциональной зависимости $y = \varphi(x)$. Пусть величина y является функцией аргумента x . Это означает, что любому значению x из области определения поставлено в соответствии значение y . Вместе с тем на практике часто неизвестна явная связь между y и x , т.е. невозможно записать эту связь в виде $y = f(x)$. Наиболее распространенным и практически важным случаем, когда вид связи между параметрами x и y неизвестен, является задание этой связи в виде некоторой таблицы $\{x_i, y_i\}$. Это означает, что дискретному множеству значений аргумента $\{x_i\}$ поставлено в соответствие множество значений функции $\{y_i\}$ ($i=0, 1, n$). Эти значения – либо результаты расчетов, либо экспериментальные данные. На практике могут понадобиться значения величины y и в других точках, отличных от узлов x_i . Однако получить эти значения можно лишь путем очень сложных расчетов или проведением дорогостоящих экспериментов. Таким образом, с точки зрения экономии времени и средств появляется необходимость использования имеющихся табличных данных для приближенного вычисления искомого параметра y при любом значении (из некоторой области) определяющего параметра x , поскольку точная связь $y = f(x)$ неизвестна. Этой цели и служит задача об аппроксимации функций: данную функцию $f(x)$ требуется приближенно заменить (аппроксимировать) некоторой функцией $g(x)$ так, чтобы отклонение (в некотором смысле) $g(x)$ от $f(x)$ в заданной области было минимальным. Функция $g(x)$ при этом называется аппроксимирующей.

Обычно определение параметров при известном виде зависимости осуществляют по методу наименьших квадратов. Метод наименьших квадратов позволяет по экспериментальным данным подобрать такую аналитическую функцию, которая проходит настолько близко к экспериментальным точкам, насколько это возможно. Именно поэтому он является полезным при обработке экспериментальных данных.

Метод наименьших квадратов

Пусть задана некоторая функция $f(x)$ и существует некоторая функция $\varphi(x)$. При этом функция $\varphi(x)$ считается наилучшим приближением к $f(x)$, если для нее сумма квадратов невязок $\varepsilon_i = f(x_i) - \varphi(x_i)$ или отклонений “теоретических” значений $\varphi(x_i)$, найденных по эмпирической формуле, от соответствующих опытных значений y_i , имеет наименьшее значение по сравнению с другими функциями, из числа которых выбирается искомое приближение.

$$\sigma = \sum_{i=0}^n [f(x_i) - \varphi(x_i)]^2 \rightarrow \min \quad (1)$$

Используя методы дифференциального исчисления, метод наименьших квадратов формулирует аналитические условия достижения суммой квадратов отклонений σ (1) своего наименьшего значения. Так, если функция $\varphi(x)$ вполне определяется своими параметрами a, b, c, \dots , то наилучшие значения этих параметров находятся из решения системы уравнений.

Например, в простейшем случае, когда функция $\varphi(x)$ представлена линейным уравнением $y = ax + b$, система имеет вид:

$$\begin{cases} a \cdot \sum_{i=1}^n x_i^2 + b \cdot \sum_{i=1}^n x_i = \sum_{i=1}^n x_i \cdot y_i \\ a \cdot \sum_{i=1}^n x_i + b \cdot n = \sum_{i=1}^n y_i \end{cases} \quad (2)$$

Выяснить вид функции можно либо из теоретических соображений, либо анализируя расположение точек (x_i, y_i) на координатной плоскости.

Например, пусть точки расположены так, как показано на рис.1.



Рис.1.

Учитывая то, что практические данные получены с некоторой погрешностью, обусловленной неточностью измерений, необходимостью округления результатов и т.д., естественно предположить, что здесь имеет место линейная зависимость $y=ax+b$. Чтобы функция приняла конкретный вид, необходимо каким-то образом вычислить a и b . Для этого можно решить систему (2). Расположение экспериментальных точек в виде кривой на рис.2 наводит на мысль, что зависимость обратно пропорциональна и функцию $\varphi(x)$ нужно подбирать в виде $y=a+b/x$. Здесь также необходимо вычислить параметры a и b .

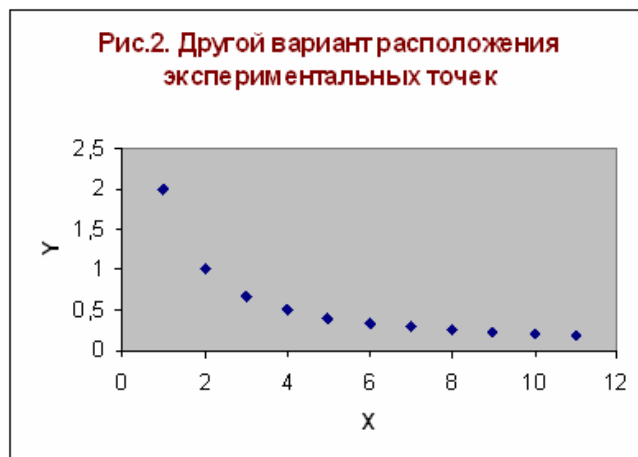


Рис.2.

Таким образом, расположение экспериментальных точек может иметь самый различный вид, и каждому соответствует конкретный тип функции. Построение эмпирической функции сводится к вычислению входящих в нее параметров, так чтобы из всех функций такого вида выбрать ту, которая лучше других описывает зависимость между изучаемыми величинами. То есть сумма квадратов разности между табличными значениями функции в некоторых точках и значениями, вычислениями по полученной формуле, должна быть минимальна.

Аппроксимация экспериментальных данных в программе Microsoft Excel

В MS Excel аппроксимация экспериментальных данных осуществляется путем построения их графика (x – отвлеченные величины) или точечного графика (x – имеет конкретные значения) с последующим подбором подходящей аппроксимирующей функции (линии тренда).

Возможны следующие варианты функций:

- Линейная – $y=ax+b$. Обычно применяется в простейших случаях, когда экспериментальные данные возрастают или убывают с постоянной скоростью.
- Полиномиальная – $y=a_0+a_1x+a_2x^2+\dots+a_nx^n$, где до шестого порядка включительно ($n \leq 6$), a_i – константы. Используется для описания экспериментальных данных, попеременно возрастающих и убывающих. Степень полинома определяется количеством экстремумов (максимумов или минимумов) кривой. Полином второй степени можно описать только один максимум или минимум, полином третьей степени может иметь один или два экстремума, четвертой степени – не более трех экстремумов и т.д.
- Логарифмическая – $y=a \cdot \ln x + b$, где a и b – константы, \ln – функция натурального логарифма. Функция применяется для описания экспериментальных данных, которые вначале быстро растут или убывают, а затем постепенно стабилизируются.
- Степенная – $y=b \cdot x^a$, где a и b – константы. Аппроксимация степенной функцией используется для экспериментальных данных с постоянно увеличивающейся (или убывающей) скоростью роста. Данные не должны иметь нулевых или отрицательных значений.
- Экспоненциальная – $y=b \cdot e^{ax}$, a и b – константы, e – основание натурального логарифма. Применяется для описания экспериментальных данных, которые быстро растут или убывают, а затем постепенно стабилизируются. Часто ее использование вытекает из теоретических соображений.

Степень близости аппроксимации экспериментальных данных выбранной функцией оценивается коэффициентом детерминации (R^2). Таким образом, если есть несколько подходящих вариантов типов аппроксимирующих функций, можно выбрать функцию с большим коэффициентом детерминации (стремящимся к 1).

Расчёт поставленной задачи посредством разных программных средств. Расчёт аппроксимации экспериментальных данных с использованием Microsoft Excel

Исходные данные:

Вариант 9. Функция $y=f(x)$ задана таблицей 1

Таблица 1

x_i	y_i	x_i	y_i	x_i	y_i	x_i	y_i	x_i	y_i
1.05	3.45	3.65	40.43	5.08	95.06	7.13	87.95	9.54	25.97
1.65	6.76	4.05	53.87	5.43	100.98	7.34	72.08	9.85	18.64
2.08	9.08	4.15	59.96	5.89	121.76	8.01	60.87	10.06	11.43
2.76	17.98	4.39	70.08	6.43	112.83	8.54	55.08	10.42	8.87
2.99	27.78	4.76	85.96	6.91	99.05	9.01	44.41	10.89	5.51

Для расчёта исходные данные подставляются в таблицу 2. В ячейках A1:A25 и B1:B25 (табл. 1) занесены значения x и y соответственно.

Таблица 2 Вычисление средних значений и их сумм

	A	B	C	D	E	F	G	H	I
1	1,05	3,45	1,10	3,62	1,16	1,22	3,80	1,24	1,30
2	1,65	6,76	2,72	11,15	4,49	7,41	18,40	1,91	3,15
3	2,08	9,08	4,33	18,89	9,00	18,72	39,28	2,21	4,59
4	2,76	17,98	7,62	49,62	21,02	58,03	136,96	2,89	7,97
5	2,99	27,78	8,94	83,06	26,73	79,93	248,36	3,32	9,94
6	3,65	40,43	13,32	147,57	48,63	177,49	538,63	3,70	13,50
7	4,05	53,87	16,40	218,17	66,43	269,04	883,60	3,99	16,15
8	4,15	59,96	17,22	248,83	71,47	296,61	1032,66	4,09	16,99
9	4,39	70,08	19,27	307,65	84,60	371,41	1350,59	4,25	18,66
10	4,76	85,96	22,66	409,17	107,85	513,37	1947,65	4,45	21,20
11	5,08	95,06	25,81	482,90	131,10	665,97	2453,16	4,55	23,14
12	5,43	100,98	29,48	548,32	160,10	869,36	2977,39	4,61	25,06
13	5,89	121,76	34,69	717,17	204,34	1203,54	4224,11	4,80	28,28
14	6,43	112,83	41,34	725,50	265,85	1709,40	4664,95	4,73	30,39
15	6,91	99,05	47,75	684,44	329,94	2279,88	4729,45	4,60	31,76
16	7,13	87,95	50,84	627,08	362,47	2584,39	4471,11	4,48	31,92
17	7,34	72,08	53,88	529,07	395,45	2902,58	3883,35	4,28	31,40
18	8,01	60,87	64,16	487,57	513,92	4116,52	3905,43	4,11	32,91
19	8,54	55,08	72,93	470,38	622,84	5319,02	4017,07	4,01	34,24
20	9,01	44,41	81,18	400,13	731,43	6590,21	3605,21	3,79	34,18
21	9,54	25,97	91,01	247,75	868,25	8283,11	2363,57	3,26	31,07
22	9,85	18,64	97,02	183,60	955,67	9413,37	1808,50	2,93	28,81
23	10,06	11,43	101,20	114,99	1018,11	10242,17	1156,76	2,44	24,51
24	10,42	8,87	108,58	92,43	1131,37	11788,83	963,07	2,18	22,74
25	10,89	5,51	118,59	60,00	1291,47	14064,09	653,44	1,71	18,58
26	152,06	1295,84	1132,05	7869,08	9423,68	83825,66	52076,49	88,52	542,44

В ячейку C1 вводится формула =A1^2. Затем она копируется в ячейки C2:C25.

В ячейку D1 вводится формула =A1*B1. Затем она копируется в ячейки D2:D25.

В ячейку E1 вводится формула =A1^3. Затем она копируется в ячейки E2:E25.

В ячейку F1 вводится формула =A1^4. Ячейки F2:F25 заполняются копированием.

В ячейку G1 вводится формула = A1^2*B1. Ячейки G2:G25 заполняются копированием.

В ячейку H1 вводится формула = LN(B1). Ячейки H2:H25 заполняются копированием.

В ячейку I1 вводится формула =A1*LN(B1). Ячейки I2:I25 заполняются копированием.

Ячейки A26:I26 являются суммами всех ячеек соответствующих столбцов.

Линейная аппроксимация функции:

Решив систему (1), воспользовавшись значениями, расположенными в ячейках A26, B26, C26, D26, находим a_1 и a_2 (табл. 3).

$$\begin{cases} a_1 n + a_2 \sum_{i=1}^n x_i = \sum_{i=1}^n y_i \\ a_1 \sum_{i=1}^n x_i + a_2 \sum_{i=1}^n x_i^2 = \sum_{i=1}^n x_i y_i \end{cases} \quad (1)$$

Таблица 3 Нахождение коэффициентов a_1 и a_2

	A	B	C	D	E
28	25,00	152,06	1295,84		
29	152,06	1132,05	7869,08		
30					
31	Обратная матрица				
32	0,22	-0,03		a1=	52,21
33	-0,03	0,00		a2=	-0,06

Линейная аппроксимация имеет вид:

$$y = 52,21 - 0,06 * x.$$

В ячейках A32:B33 записана формула {=МОБР(A28:B29)}.

В ячейках E22:E33 записана формула

$$\{=\text{МУМНОЖ}(\text{A32:В33},\text{C28:C29})\}.$$

Квадратичная аппроксимация функции:

Для аппроксимации функции $y=f(x)$ квадратичной функцией $y = a_1 + a_2x + a_3x^2$ решим систему (2) используя значения, расположенные в ячейках A26, B26, C26, D26, E26, F26, G26. Таким образом определим коэффициенты a_1, a_2, a_3 (табл. 4).

$$\begin{cases} a_1 n + a_2 \sum_{i=1}^n x_i + a_3 \sum_{i=1}^n x_i^2 = \sum_{i=1}^n y_i \\ a_1 \sum_{i=1}^n x_i + a_2 \sum_{i=1}^n x_i^2 + a_3 \sum_{i=1}^n x_i^3 = \sum_{i=1}^n x_i y_i \\ a_1 \sum_{i=1}^n x_i^2 + a_2 \sum_{i=1}^n x_i^3 + a_3 \sum_{i=1}^n x_i^4 = \sum_{i=1}^n x_i^2 y_i \end{cases} \quad (2)$$

Таблица 4 Нахождение коэффициентов a_1, a_2, a_3

	A	B	C	D	E	F
36	25,00	152,06	1132,05	1295,84		
37	152,06	1132,05	9423,68	7869,08		
38	1132,05	9423,68	83825,66	52076,49		
39						
40	Обратная матрица					
41	0,80	-0,27	0,02		a1=	-76,14
42	-0,27	0,11	-0,01		a2=	53,72
43	0,02	-0,01	0,00		a3=	-4,39

Квадратичная аппроксимация имеет вид: $y = -74,14 + 53,72 \cdot x - 4,39 \cdot x^2$.

В ячейках A41:C43 записана формула $\{=\text{МОБР}(\text{A36:C38})\}$.

В ячейках F41:F43 записана формула $\{=\text{МУМНОЖ}(\text{A41:A43},\text{D36:D38})\}$.

Экспоненциальная аппроксимация функции:

Аппроксимируем функцию $y=f(x)$ экспоненциальной функцией $y = a_1 \cdot e^{a_2 x}$.
 Определим коэффициенты a_1 и a_2 прологарифмировав значения y и используя значения, полученные в ячейках A26, C26, H26, I26 получим систему:

$$\begin{cases} 25 \cdot c + \text{A26} \cdot a_2 = \text{H26} \\ \text{A26} \cdot c + \text{C26} \cdot a_2 = \text{I26} \end{cases} \quad (3)$$

где $c = \ln(a_1)$.

Решив систему (3), нашли c и a_2 , a_1 нашли потенцированием (табл. 5).

Таблица 5 Нахождение коэффициентов c, a_1, a_2

	A	B	C	D	E	
46	25,00	152,06	88,52			
47	152,06	1132,05	542,44			
48						
49	Обратная матрица				c=	3,42
50	0,22	-0,03		a2=	0,02	
51	-0,03	0,00		a1=	30,64	

В ячейках A50:A51 записана формула $\{=\text{МОБР}(\text{A46:B47})\}$.

В ячейках E49:E50 записана формула {=МУМНОЖ(A50:B51,C46:C47)}.

В ячейке E51 записана формула = EXP(E49).

Среднее арифметическое всех x и y вычисляется по формулам:

$$\bar{x} = \frac{1}{25} \sum_{i=1}^{25} x_i; \quad \bar{y} = \frac{1}{25} \sum_{i=1}^{25} y_i.$$

(табл. 6)

Таблица 6 Нахождение среднего арифметического всех x и y

	A	B
54	X _{ср} =	6,08
55	Y _{ср} =	51,83

В ячейке B54 записана формула =A26/25.

В ячейке B55 записана формула =B26/25.

Расчёт коэффициента корреляции и коэффициента детерминации:

Данные занесём в таблицу 7, которая является продолжением таблицы 2.

Таблица 7

	A	B	J	K	L	M	N	O	I
1	1,05	3,45	243,49	25,33	2340,97	2371,00	785,34	774,15	1,30
2	1,65	6,76	199,78	19,65	2031,63	2056,27	38,59	619,06	3,15
3	2,08	9,08	171,12	16,02	1827,87	1848,97	56,67	521,07	4,59
4	2,76	17,98	112,48	11,04	1146,07	1159,94	428,96	205,98	7,97
5	2,99	27,78	74,38	9,56	578,58	587,76	304,91	22,06	9,94
6	3,65	40,43	27,74	5,92	130,04	133,47	442,27	56,74	13,50
7	4,05	53,87	-4,14	4,13	4,15	3,65	242,06	429,13	16,15
8	4,15	59,96	-15,70	3,73	66,04	64,12	126,36	715,08	16,99
9	4,39	70,08	-30,88	2,86	332,93	329,15	25,17	1347,28	18,66
10	4,76	85,96	-45,13	1,75	1164,61	1159,07	34,20	2739,90	21,20
11	5,08	95,06	-43,33	1,00	1868,52	1863,20	134,10	3749,60	23,14
12	5,43	100,98	-32,06	0,43	2415,37	2411,43	220,39	4478,65	25,06
13	5,89	121,76	-13,45	0,04	4889,70	4888,05	1140,58	7638,12	28,28
14	6,43	112,83	21,20	0,12	3720,56	3723,17	626,89	6100,10	30,39
15	6,91	99,05	39,08	0,68	2229,39	2234,20	184,44	4095,63	31,76
16	7,13	87,95	37,84	1,10	1304,39	1309,05	17,82	2782,21	31,92
17	7,34	72,08	25,46	1,58	409,92	413,05	91,98	1349,27	31,40
18	8,01	60,87	17,42	3,72	81,66	83,81	135,63	627,92	32,91
19	8,54	55,08	7,98	6,04	10,54	11,54	54,79	357,09	34,24
20	9,01	44,41	-21,73	8,57	55,11	52,47	50,57	62,32	34,18
21	9,54	25,97	-89,43	11,95	668,93	657,98	118,03	119,34	31,07
22	9,85	18,64	-125,06	14,19	1101,82	1086,49	71,58	341,43	28,81
23	10,06	11,43	-160,71	15,82	1632,45	1612,75	73,91	667,70	24,51
24	10,42	8,87	-186,36	18,81	1845,87	1823,03	3,49	821,51	22,74
25	10,89	5,51	-222,71	23,11	2145,88	2118,58	296,81	1047,63	18,58
26	152,06	1295,84	-12,73	207,16	34002,98	34002,20	5705,54	41668,97	542,44

Ячейки A1:A26 и B1:B26 заполняются как в таблице 2.

В ячейку J1 вводим формулу =(A1-\$B\$54)*(B1-\$B\$55).

Ячейки J2:J25 заполняются копированием.

В ячейку K1 вводим формулу =(A1-\$B\$54)^2.

Ячейки K2:K25 заполняются копированием.

В ячейку L1 вводится формула =(B1-\$B\$55)^2. Ячейки L2:L25 заполняются копированием.

В ячейку M1 вводится формула =(\$E\$32+\$E\$33*A1-B1)^2. Ячейки M2:M25 заполняются копированием.

В ячейку N1 вводим формулу =(\$F\$41+\$F\$42*A1+\$F\$43*A1^2-B1)^2. Ячейки N2:N25 заполняются копированием.

В ячейку O1 вводим формулу =(\$E\$51*EXP(\$E\$50*A1)-B1)^2. Ячейки O2:O25 заполняются копированием.

Ячейки J26:O26 являются суммами всех ячеек соответствующих столбцов.

По формулам (4) и (5) рассчитаем коэффициенты корреляции и детерминации (табл. 8). Коэффициент корреляции рассчитывается только для линейной аппроксимации.

$$\rho = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (4)$$

где $\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$, $\bar{y} = \frac{\sum_{i=1}^n y_i}{n}$, \bar{x} и \bar{y} — среднее арифметическое значение соответственно по x и y .

$$r^2 = 1 - \frac{S_{\text{ост}}}{S_{\text{полн}}} \quad (5)$$

где $S_{\text{ост}}$ – остаточная сумма квадратов, $S_{\text{полн}}$ – сумма остаточной суммы квадратов и регрессионной суммы квадратов.

Таблица 8 Нахождение коэффициента корреляции и коэффициента детерминации

	А	В
57	Коэффициент корреляции	-0,004798
58	Коэффициент детерминированности	
59	(линейная аппроксимация)	0,000023
60	Коэффициент детерминированности	
61	(квадратичная аппроксимация)	0,832205
62	Коэффициент детерминированности	
63	(экспоненциальная аппроксимация)	-0,225451

В ячейку В57 вводим формулу =J26/(K26*L26)^(1/2).

В ячейку В59 вводим формулу =1-M26/L26.

В ячейку В61 вводим формулу =1-N26/L26.

В ячейку В63 вводим формулу =1-O26/L26.

Построение графика в Excel и использование функции ЛИНЕЙН:

Построим график функции и линию тренда, которой является прямая $y=a_1+a_2*x$.

Для построения линии тренда необходимо:

1. Активизируем диаграмму дважды щёлкнув по ней и нажимаем на одну из точек.
2. Пункт меню Вставка-«Линия тренда».
3. На вкладке «Тип» выбираем линейный тип для графика линейной аппроксимации (рис.3.), полиномиальный тип степень 2 для графика квадратичной аппроксимации (рис.4.) и экспоненциальный тип для графика экспоненциальной аппроксимации (рис.5.).
4. На вкладке «Параметры» потребуем показать уравнение тренда и значение R^2 .

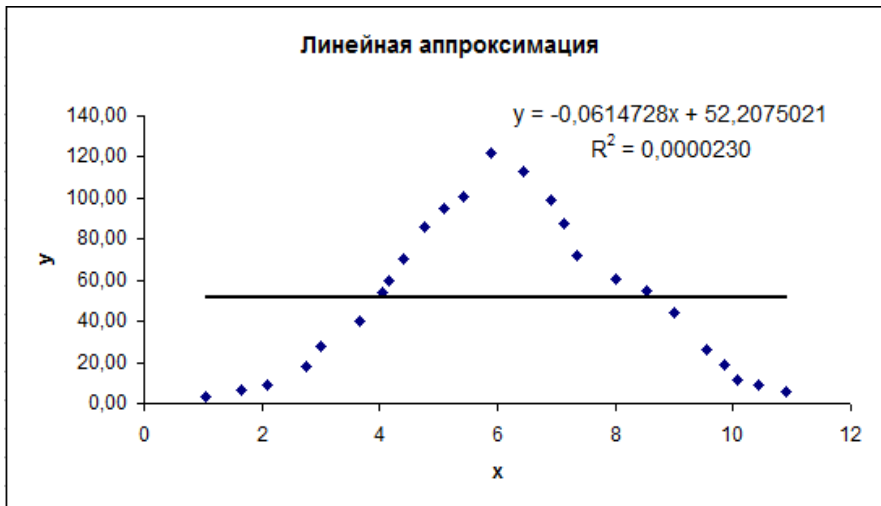


Рис.3. Исходная функция и линия тренда для линейной аппроксимации.

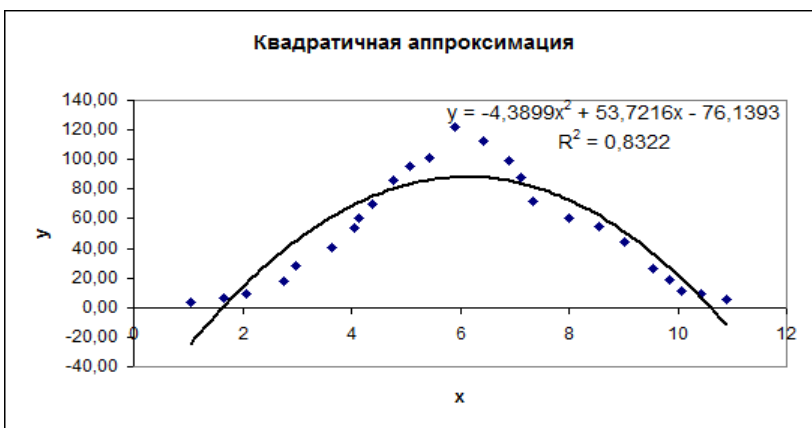


Рис.4. Исходная функция и линия тренда для квадратичной аппроксимации.

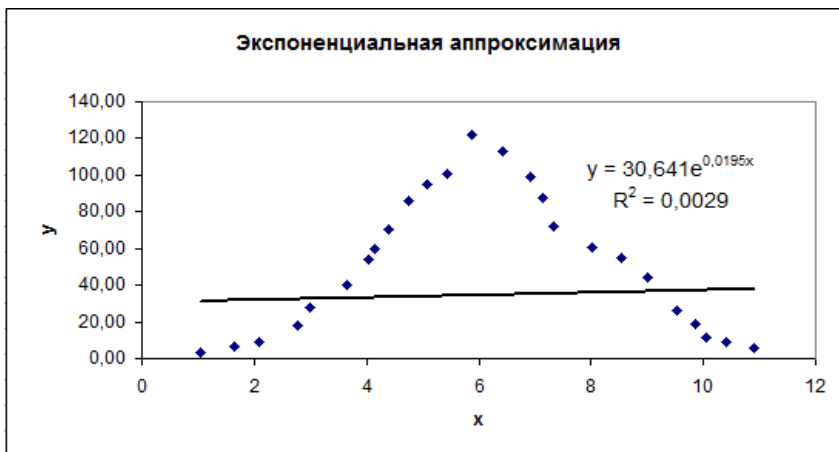


Рис.5. Исходная функция и линия тренда для экспоненциальной аппроксимации.

Получение числовых характеристик зависимости:

Выделяем область A65:B69 и вызываем Мастер функций. Выбираем функцию ЛИНЕЙН. В качестве «изв_знач_y» указываем B1:B25, а «изв_знач_x» A1:A25. В поле «стат» выбираем истина. Нажимаем Закончить. Устанавливаем курсор в строку формул и нажимаем Ctrl+Shift+Enter (табл. 9).

Таблица 9 Числовые характеристики зависимости

	A	B
65	-0,061472792	52,20750211
66	2,671363575	17,97613885
67	0,00002302	38,44940059
68	0,000529541	23
69	0,78285084	34002,19733

В ячейке A65 расположено значения коэффициента a_2 , в ячейке B65 расположено значение коэффициента a_1 .

В A66 расположено значение стандартной ошибки a_1 , в B66 расположено значение стандартной ошибки коэффициента a_2 .

В ячейке A67 расположено значение коэффициента детерминированности.

В ячейке A68 расположено значение F-критерия, в ячейке B68 расположено число степеней свободы.

В ячейке B69 расположено значение остаточной суммы квадратов, а в ячейке A69 значение регрессионной суммы квадратов.

В ходе расчётов выяснили, что квадратичная аппроксимация наилучшим способом описывает экспериментальные данные. Сравнивая результаты, полученные при помощи функции ЛИНЕЙН, видим, что они полностью совпадают с вычислениями, проведёнными выше. Полученное при построении линии тренда значения коэффициента детерминированности для экспоненциальной зависимости $R^2 = 0,0029$ не совпадает с истинным значением $R^2 = -0,2254$ поскольку при вычислении коэффициента детерминированности с помощью функции ЛИНЕЙН используются не истинные значения y , а преобразованные значения $\ln y$ с дальнейшей линеаризацией.

Лабораторная работа №8.

Расчёт аппроксимации экспериментальных данных с использованием MathCAD.

Аппроксимация экспериментальных данных в программе MathCAD

В MathCAD существует несколько функций, позволяющих выполнить регрессию с использованием зависимостей, наиболее часто встречающихся на практике. Таких функций в MathCAD всего шесть. Вот некоторые из них:

- $\text{expfit}(vx,vy,vg)$ - регрессия экспоненциальной функцией $y = a \cdot e^{b \cdot x} + c$.
- $\text{sinfit}(vx,vy,vg)$ - регрессия синусоидальной функцией $y = a \cdot \sin(x+b) + c$.
- $\text{pwrfit}(vx,vy,vg)$ - регрессия степенной функцией $y = a \cdot x^b + c$.

Перечисленные функции используют трехпараметрическую аппроксимирующую функцию, нелинейную по параметрам. При вычислении оптимальных значений трех параметров регрессионной функции по методу наименьших квадратов возникает необходимость в решении сложной системы из трех нелинейных уравнений. Такая система часто может иметь несколько решений. Поэтому в функциях MathCAD, которые проводят регрессию трехпараметрическими зависимостями, введен дополнительный аргумент vg . Данный аргумент - это трехкомпонентный вектор, содержащий приблизительные значения параметров a , b и c , входящих в аппроксимирующую функцию. Неправильный выбор элементов вектора vg может привести к неудовлетворительному результату регрессии. В MathCAD существуют средства для проведения регрессии самого общего вида. Это означает, что можно использовать любые функции в качестве аппроксимирующих и находить оптимальные значения любых их параметров, как линейных, так и нелинейных. В том случае, если регрессионная функция является линейной по всем параметрам, т.е. представляет линейную комбинацию жестко заданных функций, провести регрессию можно с помощью встроенной функции $\text{linfit}(vx,vy,F)$. Аргумент F - это векторная функция, из элементов которой должна быть построена линейная комбинация, наилучшим образом аппроксимирующая заданную последовательность точек. Результатом работы функции linfit является вектор линейных коэффициентов. Каждый элемент этого вектора - коэффициент при функции, стоящей на соответствующем месте в векторе F . Таким образом, для того чтобы получить регрессионную функцию, достаточно скалярно перемножить эти два вектора.

Расчёт аппроксимации экспериментальных данных с использованием MathCAD

Ввод исходных данных:

$x :=$	1.05	$y :=$	3.45
	1.65		6.76
	2.08		9.08
	2.76		17.98
	2.99		27.78
	3.65		40.43
	4.05		53.87
	4.15		59.96
	4.39		70.08
	4.76		85.96
	5.08		95.06
	5.43		100.98
	5.89		121.76
	6.43		112.83
	6.91		99.05
	7.13		87.95
	7.34		72.08
	8.01		60.87
	8.54		55.08
	9.01		44.41
	9.54		25.97
	9.85		18.64
	10.06		11.43
	10.42		8.87
	10.89		5.51
	$n := \text{last}(x)$		$i := 0 .. n$

Линейная аппроксимация функции:
Нахождение коэффициентов A и B.

$A := \text{intercept}(x, y)$ $B := \text{slope}(x, y)$
 $f(i) := A + B \cdot c$

$A = 52.208$ $B = -0.061$

$c := 0, 0.1 .. 12$

Построение графика линейной аппроксимации в MathCAD:

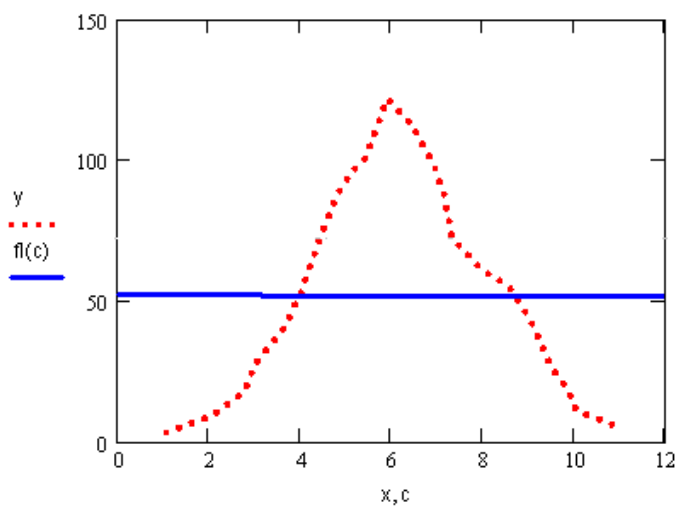


Рис.6. Исходная функция и линия тренда для линейной аппроксимации.

Квадратичная аппроксимация данных:
Нахождение коэффициентов $d2$, $f2(c)$, $k2$.

```

p := 2
d2 := regress(x, y, p)
f2(c) := interp(d2, x, y, c)
k2 := submatrix(d2, 3, length(d2) - 1, 0, 0)
k2T = (-76.139  53.722  -4.39)

```

Построение графика квадратичной аппроксимации в MathCAD:

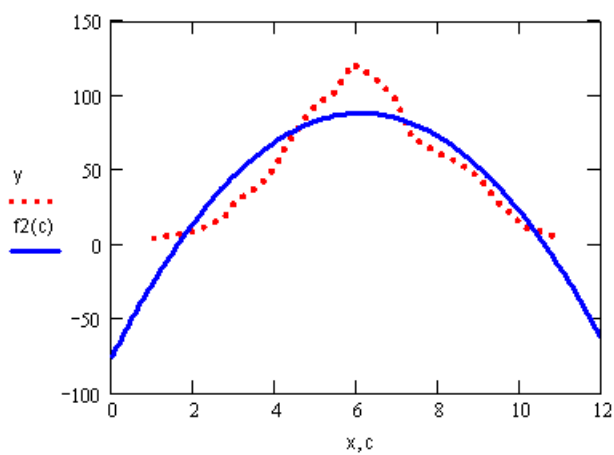


Рис.7. Исходная функция и линия тренда для квадратичной аппроксимации.

Экспоненциальная аппроксимация данных:
Нахождение коэффициентов $a1$, $a2$, z .


```

m := ln(y)
z := intercept(x, m)
z = 3.422
a2 := slope(x, m)
a2 = 0.019
a1 := exp(z)
a1 = 30.641
f3(c) := a1 · exp(a2 · c)

```

Построение графика экспоненциальной аппроксимации в MathCAD:

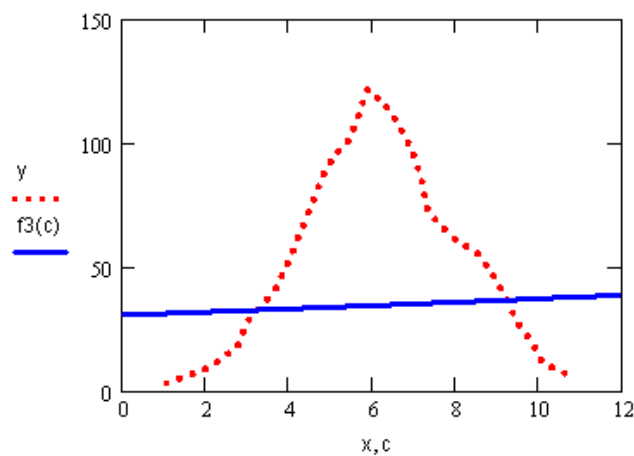


Рис.8. Исходная функция и линия тренда для экспоненциальной аппроксимации.

Практическое занятие № 3. Расчёт аппроксимации экспериментальных данных с использованием MatLAB

Ввод исходных данных:

```

>> x=[1.05 1.65 2.08 2.76 2.99 3.65 4.05 4.15 4.39 4.76 5.08 5.43 5.89 6.43 6.91 7.13
7.34 8.01 8.54 9.01 9.54 9.85 10.06 10.42 10.89];
>> y=[3.45 6.76 9.08 17.98 27.78 40.43 53.87 59.96 70.08 85.96 95.06 100.98 121.76
112.83 99.05 87.95 72.08 60.87 55.08 44.41 25.97 18.64 11.43 8.87 5.51];

```

Линейная аппроксимация функции:
Нахождение коэффициентов линейной функции.

```

>> p=polyfit(x,y,1)
p =
-0.0615 52.2075
>> t=polyval(p,x);
>> plot(x,y,'ob',x,t,'-g')
>>

```

Построение графика линейной аппроксимации в MatLAB:

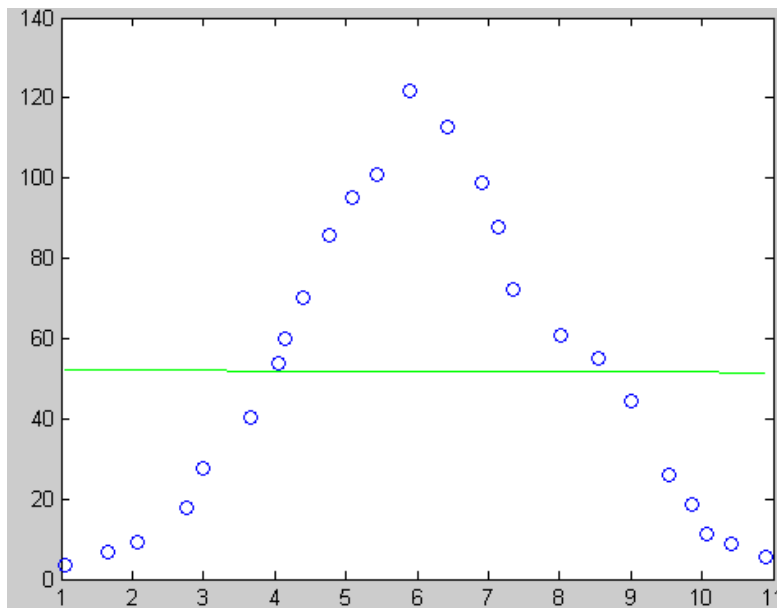


Рис.9. Исходные данные и линия тренда для линейной аппроксимации.

Квадратичная аппроксимация данных:
Нахождение коэффициентов квадратичной функции.

```
>> p=polyfit(x,y,2)
p =
-4.3899 53.7216 -76.1393
>> t=polyval(p,x);
>> plot(x,y,'ob',x,t,'-g')
>>
```

Построение графика квадратичной аппроксимации в MatLAB:

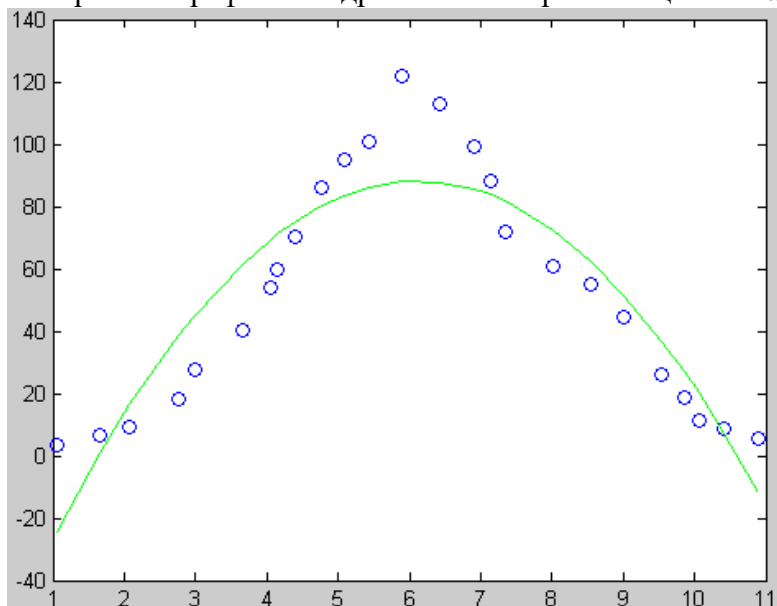


Рис.10. Исходные данные и линия тренда для квадратичной аппроксимации.

Экспоненциальная аппроксимация данных:
Нахождение коэффициентов экспоненциальной функции.

```
>> b=log(y);
>> p=polyfit(x,b,1)
```

```

p =
0.0195 3.4223
>> c=3.4223
c =
3.4223
>> a2=0.0195
a2 =
0.0195
>> a1=exp(c)
a1 =
30.6398
>> t=a1*exp(a2*x);
>> plot(x,y,'ob',x,t,'-g')
>>

```

Построение графика экспоненциальной аппроксимации в MatLAB:

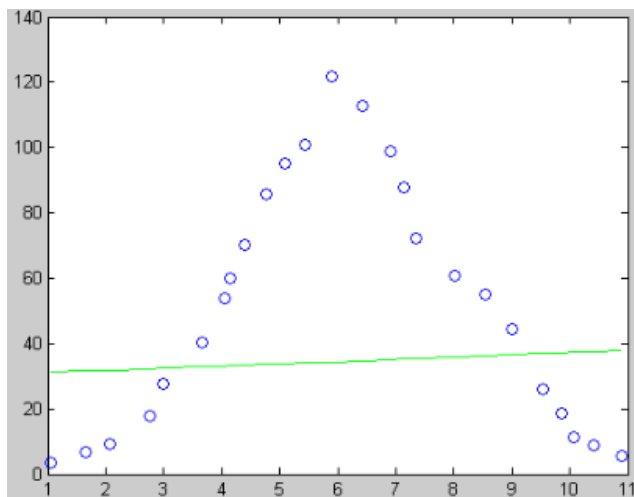


Рис.11. Исходные данные и линия тренда для экспоненциальной аппроксимации

Варианты заданий

Вариант									
0		1		2		3		4	
x_i	y_i	x_i	y_i	x_i	y_i	x_i	y_i	x_i	y_i
1,05	2,10	2,03	-1,88	1,05	-1,88	2,03	-1,88	1,05	3,12
1,57	3,46	2,31	-0,66	1,57	-0,66	2,31	-0,66	1,57	4,34
2,09	5,37	2,59	0,71	2,09	0,71	2,59	0,71	2,09	5,71
2,61	7,81	2,87	2,24	2,61	2,24	2,87	2,24	2,61	7,24
3,13	10,80	3,15	3,92	3,13	3,92	3,15	3,92	3,13	8,92
3,65	14,32	3,43	5,76	3,65	5,76	3,43	5,76	3,65	10,76
4,17	18,39	3,71	7,76	4,17	7,76	3,71	7,76	4,17	12,76
4,69	23,00	3,99	9,92	4,69	9,92	3,99	9,92	4,69	14,92
5,21	28,14	4,27	12,23	5,21	12,23	4,27	12,23	5,21	17,23
5,73	33,83	4,55	14,70	5,73	14,70	4,55	14,70	5,73	19,7
6,25	40,06	4,83	17,33	6,25	17,33	4,83	17,33	6,25	22,33
6,77	46,83	5,11	20,11	6,77	12,23	5,11	12,23	6,77	17,23
7,29	54,14	5,39	23,05	7,29	9,92	5,39	9,92	7,29	14,92
7,81	62,00	5,67	26,15	7,81	7,76	5,67	7,76	7,81	12,76
8,33	70,39	5,95	29,40	8,33	5,76	5,95	5,76	8,33	10,76

8,85	79,32	6,23	32,81	8,85	3,92	6,23	3,92	8,85	8,92
9,37	88,80	6,51	36,38	9,37	2,24	6,51	2,24	9,37	7,24
9,89	98,81	6,79	40,10	9,89	0,71	6,79	0,71	9,89	5,71
10,41	109,3 7	7,07	43,98	10,41	-0,66	7,07	-0,66	10,41	4,34
10,93	120,4 6	7,35	48,02	10,93	-1,88	7,35	-1,88	10,93	3,12

Вариант									
5		6		7		8		9	
x _i	y _i	x _i	y _i	x _i	y _i	x _i	y _i	x _i	y _i
2,03	2,10	2,03	6,33	2,03	2,12	1,05	2,12	2,03	3,12
2,31	3,46	2,31	10,03	2,31	3,34	1,57	3,34	2,31	4,34
2,59	5,37	2,59	14,79	2,59	4,71	2,09	4,71	2,59	5,71
2,87	7,81	2,87	20,78	2,87	6,24	2,61	6,24	2,87	7,24
3,15	10,80	3,15	28,10	3,15	7,92	3,13	7,92	3,15	8,92
3,43	14,32	3,43	36,91	3,43	9,76	3,65	9,76	3,43	10,76
3,71	18,39	3,71	47,34	3,71	11,7 6	4,17	11,76	3,71	12,76
3,99	23,00	3,99	59,53	3,99	13,9 2	4,69	13,92	3,99	14,92
4,27	28,14	4,27	73,57	4,27	16,2 3	5,21	16,23	4,27	17,23
4,55	33,83	4,55	89,64	4,55	18,7	5,73	18,7	4,55	19,7
4,83	40,06	4,83	107,8 5	4,83	23,3 3	6,25	23,33	4,83	22,33
5,11	46,83	5,11	88,05	5,11	18,2 3	6,77	18,23	5,11	18,23
5,39	54,14	5,39	80,42	5,39	15,9 2	7,29	15,92	5,39	15,92
5,67	62,00	5,67	72,35	5,67	13,7 6	7,81	13,76	5,67	13,76
5,95	70,39	5,95	64,02	5,95	11,7 6	8,33	11,76	5,95	11,76
6,23	79,32	6,23	55,57	6,23	9,92	8,85	9,92	6,23	9,92
6,51	88,80	6,51	47,13	6,51	8,24	9,37	8,24	6,51	8,24
6,79	98,81	6,79	38,77	6,79	6,71	9,89	6,71	6,79	6,71
7,07	109,37	7,07	30,68	7,07	5,34	10,41	5,34	7,07	5,34
7,35	120,46	7,35	22,93	7,35	4,12	10,93	4,12	7,35	4,12

Лабораторная работа №9.
Расчёт аппроксимации экспериментальных данных с использованием
MatLAB.

Аппроксимация экспериментальных данных в программе MatLAB

MatLAB (сокращение от англ. «Matrix Laboratory») — пакет прикладных программ для решения задач технических вычислений и одноимённый язык программирования, используемый в этом пакете. Язык MATLAB является высокоуровневым интерпретируемым языком программирования, включающим основанные на матрицах структуры данных, широкий спектр функций, интегрированную среду разработки, объектно-ориентированные возможности и интерфейсы к программам, написанным на других языках программирования.

Идея метода наименьших квадратов заключается в том, что функцию: $Y=f(x, a_0, a_1, \dots, a_k)$ необходимо подобрать таким образом, чтобы сумма квадратов отклонений измеренных значений y_i от расчетных y была наименьшей.

Программы, написанные на MATLAB, бывают двух типов — функции и скрипты. Функции имеют входные и выходные аргументы, а также собственное рабочее пространство для хранения промежуточных результатов вычислений и переменных. Скрипты же используют общее рабочее пространство. Как скрипты, так и функции не компилируются в машинный код и сохраняются в виде текстовых файлов. Существует также возможность сохранять так называемые pre-parsed программы — функции и скрипты, обработанные в вид, удобный для машинного исполнения. В общем случае такие программы выполняются быстрее обычных, особенно если функция содержит команды построения графиков.