

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ  
РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное автономное образовательное учреждение  
в высшего образования

«Северо-Кавказский федеральный университет»

Невинномысский технологический институт (филиал)

Методические указания к выполнению

Лабораторных работ

«основы компьютерного моделирования»

ПОСТРОЕНИЕ ВЕРОЯТНОСТНЫХ СТАТИЧЕСКИХ МОДЕЛЕЙ  
СИСТЕМ

для студентов направления 15.03.04 –

Автоматизация технологических процессов и производств

Настоящие методические указания предназначены для студентов направления 15.03.04 «Автоматизация технологических процессов и производств». Они разработаны в соответствии с государственным образовательным стандартом и основной образовательной программой специальности.

В методических указаниях рассмотрены правила идентификации детерминированных динамических объектов, определены требования к содержанию лабораторной работы, даны варианты заданий и приведен список рекомендуемой литературных источников

Составитель      канд. техн. наук доцент Д.В. Болдырев

Отв. Редактор    канд. техн. наук, доцент А.А. Евдокимов

## СОДЕРЖАНИЕ

	<b>ВВЕДЕНИЕ .....</b>	<b>4</b>
<b>1</b>	<b>ТЕОРЕТИЧЕСКАЯ ЧАСТЬ .....</b>	<b>4</b>
	1.1 Понятие о регрессии .....	4
	1.2 Множественная линейная регрессия .....	7
	1.3 Пошаговая регрессия .....	9
<b>2</b>	<b>ПОРЯДОК ВЫПОЛНЕНИЯ РАБОТЫ .....</b>	<b>11</b>
	<b>КОНТРОЛЬНЫЕ ВОПРОСЫ .....</b>	<b>12</b>
	<b>ЛИТЕРАТУРА .....</b>	<b>12</b>
	<b>ПРИЛОЖЕНИЯ .....</b>	<b>14</b>

## ВВЕДЕНИЕ

Цель работы – усвоение основных правил построения регрессионных моделей систем и оценки их качества с использованием методов математической статистики.

Содержание работы – определение параметров уравнения множественной линейной регрессии, представление полученных результатов в форме отчета и доказательство их правильности.

## 1 ТЕОРЕТИЧЕСКАЯ ЧАСТЬ

### 1.1 Понятие о регрессии

Если все параметры объекта определяются точно, его модель можно считать детерминированной и представить в виде функциональной зависимости. На практике вследствие неточности измерения или влияния физической природы объекта эти параметры находятся с некоторой ошибкой  $\epsilon$ , что делает величины на его входе и выходе  $x$  и  $h$  случайными (см. рисунок 1.1).

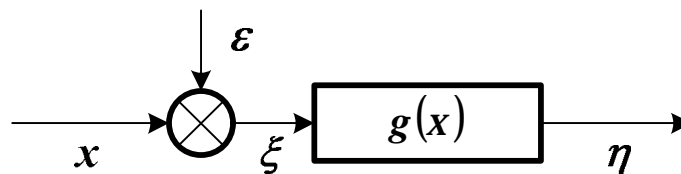


Рисунок 1.1 – Структурная схема модели объекта со случайной ошибкой входного сигнала

Информация о структуре объекта также может отсутствовать. В этих условиях зависимость  $h = g(x)$  будет вероятностной (*стохастической*). Параметр  $x$  считается независимой переменной (**фактором** или **предиктором**), а параметр  $h$  – зависимой переменной (или **откликом**).

Среди всех функций  $g(x)$  необходимо найти ту, которая наилучшим образом воспроизводит значения  $h$ . Общепринятым способом решения этой задачи является *метод наименьших квадратов*.

Зависимость  $g(x)$  называется наилучшим приближением  $h$  по методу наименьших квадратов, если математическое ожидание  $M\{h - g(x)\}^2$  принимает наименьшее возможное значение. Зависимость  $g(x)$  называется **среднеквадратической регрессией** величины  $h$  на величину  $x$ .

В случае нормального распределения  $x$  и  $h$  эта регрессия является линейной функцией

$$\frac{g(x) - Mh}{S_h} = r_{xh} \times \frac{x - Mx}{S_x}, \quad (1.1)$$

где  $Mx$ ,  $Mh$ ,  $S_x$ ,  $S_h$  – математические ожидания и среднеквадратические отклонения  $x$  и  $h$ ,  $r_{xh}$  – коэффициент их корреляции.

Уравнение (1.1) определяет *прямую регрессии*

$$\hat{h} = g(x) = Mh + r_{xh} \times \frac{S_h}{S_x} \times (x - Mx) = b_0 + b_1 \times x, \quad (1.2)$$

где  $b_0$  – смещение линии регрессии по оси ординат (т. н. *пересечение*),  $b_1$  – коэффициент регрессии  $h$  на  $x$ . Их значения находятся по формулам

$$b_1 = r_{xh} \times \frac{S_h}{S_x}, \quad b_0 = Mh - b_1 \times Mx. \quad (1.3)$$

Разность  $h - \hat{h}$  называется *остатком*  $h$  относительно  $x$ . Математическое ожидание квадрата остатка

$$M\{h - \hat{h}\}^2 = S_h^2 \times (1 - r_{xh}^2) \quad (1.4)$$

называется *остаточной дисперсией*  $h$  относительно  $x$ . Она показывает, насколько можно уменьшить дисперсию  $h$ , вычитая из переменной значение регрессии.

Для оценки точности регрессии можно рассмотреть следующее тождество ( $n$  – число наблюдений зависимой переменной).

$$(h_i - Mh) = (\hat{h}_i - Mh) + (h_i - \hat{h}_i), \quad i = 1, \dots, n. \quad (1.5)$$

Геометрическая интерпретация (1.5) показана на рисунке 1.2.

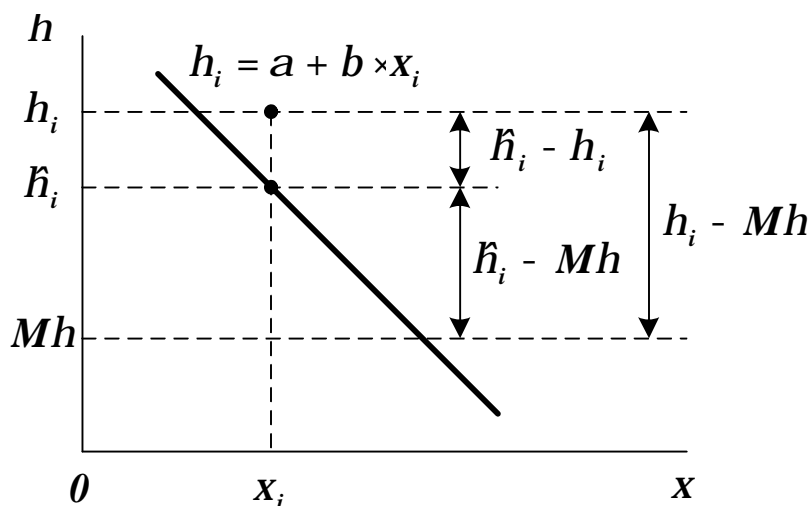


Рисунок 1.2 – Геометрическая интерпретация тождества (1.5)

Возводя обе части (1.5) в квадрат и используя формулу (1.2), можно получить основное уравнение дисперсионного анализа

$$\sum_{i=1}^n (h_i - Mh)^2 = \sum_{i=1}^n (\hat{h}_i - Mh)^2 + \sum_{i=1}^n (h_i - \hat{h}_i)^2, \quad (1.6)$$

$$SS_M = SS_R + SS_D,$$

где  $SS_M$  – общая сумма квадратов (*Square Sum*) отклонений от среднего (*Mean*),  $SS_R$  – сумма квадратов отклонений, обусловленная регрессией (*Regression*),  $SS_D$  – сумма квадратов остатка (*Deviation*). Регрессия считается удовлетворительной, если близок к единице коэффициент детерминации

$$R^2 = \frac{SS_R}{SS_M} = 1 - \frac{SS_D}{SS_M}. \quad (1.7)$$

Можно составить таблицу дисперсионного анализа, включающую значения элементов уравнения (1.6).

**Примечание.** Любая сумма квадратов связана с числом ее степеней свободы, которое показывает, как много независимых элементов выборки из  $n$  независимых чисел требуется для образования данной суммы квадратов.

Таблица 1.1 – Таблица дисперсионного анализа

<i>Источник вариации</i>	<i>Число степеней свободы</i>	<i>SS</i>
Регрессия	1	$\dot{a} (\bar{h}_i - Mh)^2$
Остаток	$n - 2$	$\dot{a} (h_i - \bar{h}_i)^2$
Общий	$n - 1$	$\dot{a} (h_i - Mh)^2$

Регрессия считается статистически значимой, если отношение

$$F = \frac{\dot{a} (\bar{h}_i - Mh)^2}{\dot{a} (h_i - \bar{h}_i)^2} \times (n - 2) \quad (1.8)$$

превышает критерий Фишера  $F_a(1, n - 2)$  для заданного уровня доверительной вероятности  $a$ .

Оценки коэффициентов  $b_0$  и  $b_1$  считаются статистически значимыми, если значения выражений

$$t_0 = \frac{|b_0|}{s} \times \sqrt{\frac{n \times \dot{a} (x_i - Mx)^2}{\dot{a} x_i^2}}, \quad t_1 = \frac{|b_1|}{s} \times \sqrt{\dot{a} (x_i - Mx)^2} \quad (1.9)$$

превышают критерий Стьюдента  $t_{1-a/2}(n - 2)$  для заданного уровня доверительной вероятности  $a$ .

## 1.2 Множественная линейная регрессия

Пусть  $x_1, \dots, x_m$  – случайные величины, имеющие математические ожидания  $Mx_i$  и дисперсии  $Dx_i$  ( $i = 1, \dots, m$ ). Если  $x_i$  – отклик, а все остальные – факторы, можно определить зависимость

$$\hat{h} = g(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_m), \quad (1.10)$$

где  $\hat{h} = x_i$ . Форма  $g(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_m)$  должна подбираться так, чтобы выполнялось условие

$$M\{h - g(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_m)\}^2 \text{ @ } \min. \quad (1.11)$$

Если зависимость (1.10) представляется линейной функцией

$$\hat{h} = Mx_i + \sum_{j=1, j \neq i}^m b_{ij} \times (x_j - Mx_j), \quad (1.12)$$

то ее называют **множественной линейной регрессией** величины  $x_i$  на остальные величины. Коэффициенты регрессии  $b_{ij}$  находятся по формуле

$$b_{ij} = - \frac{D_{ij}^{-1}}{D_{ii}^{-1}}, \quad (1.13)$$

где  $D$  – матрица центральных корреляционных моментов. Каждый ее элемент равен

$$D_{ij} = M\{(x_i - Mx_i) \times (x_j - Mx_j)\} = \begin{cases} \hat{D}x_i, & i = j, \\ \hat{cov}\{x_i, x_j\}, & i \neq j. \end{cases} \quad (1.14)$$

*Если матрица  $D$  положительно определена, то коэффициенты регрессии всегда определяются однозначно.*

Таблица дисперсионного анализа для множественной линейной регрессии приведена ниже.

Таблица 1.2 – Таблица дисперсионного анализа для множественной линейной регрессии

<i>Источник вариации</i>	<i>Число степеней свободы</i>	<i>SS</i>
Регрессия	$m - 1$	$\hat{a} (\hat{h}_i - Mh)^2$
Остаток	$n - m$	$\hat{a} (h_i - \hat{h}_i)^2$
Общий	$n - 1$	$\hat{a} (h_i - Mh)^2$



Величина  $R^2$  в данном случае является мерой статистической взаимосвязи между  $x_i$  и всеми остальными величинами и называется *сводным коэффициентом корреляции*.

Регрессия считается статистически значимой, если отношение

$$F = \frac{\dot{a} (\hat{h}_i - Mh)^2}{\dot{a} (h_i - \hat{h}_i)^2} \times \frac{n - m}{m - 1} \quad (1.15)$$

превышает критерий Фишера  $F_a(m - 1, n - m)$  для заданного уровня доверительной вероятности  $a$ .

Оценки коэффициентов  $b_{ij}$  считаются статистически значимыми, если значения выражений

$$t_j = \sqrt{\frac{r_{ij}^2}{1 - r_{ij}^2} \times (n - m)} \quad (1.16)$$

превышают критерий Стьюдента  $t_{1-a/2}(n - m)$  для заданного уровня доверительной вероятности  $a$ .

Параметры  $r_{ij}$ , входящие в (1.16), называемые *частными коэффициентами корреляции*, определяют меру статистической взаимосвязи  $x_i$  и  $x_j$  после устранения влияния всех остальных факторов. Они вычисляются по формуле

$$r_{ij} = - \frac{D_{ij}^{-1}}{\sqrt{D_{ii}^{-1} \times D_{jj}^{-1}}}. \quad (1.17)$$

### 1.3 Пошаговая регрессия

Для одного и того же объекта можно построить несколько регрессионных моделей, отличающихся набором независимых переменных. Для выбора оптимальной модели обычно используют следующие критерии:

- минимальная сумма квадратов остатка;
- минимальное число коэффициентов, совместимое с разумной ошибкой;

- достаточные физические основания для включения фактора в модель.

Одним из методов получения оптимальной модели является пошаговая регрессия. Она заключается в последовательном включении в уравнение независимых переменных и исследовании его на значимость.

На первом шаге выбирается переменная  $x_j$ , наиболее сильно коррелированная с величиной  $h$ , составляется уравнение  $\hat{h} = g(x_j)$ , проверяется общая значимость регрессии по критерию (1.15) и значимость коэффициента  $b_{ij}$  по критерию (1.16). Если результат проверки отрицательный, наилучшей считают модель  $\hat{h} = Mh$ .

На втором шаге вычисляется остаток  $h - Mh - b_{ij} \times (x_j - Mx_j)$ , и определяется наиболее сильно коррелированная с ним переменная  $x_k$ . Составляется уравнение  $\hat{h} = g(x_j, x_k)$ , проверяется общая значимость регрессии и значимость коэффициентов  $b_{ij}$  и  $b_{ik}$ . Незначимые переменные из уравнения регрессии исключаются.

На третьем шаге вычисляется остаток регрессии  $h - Mh - b_{ij} \times (x_j - Mx_j) - b_{ik} \times (x_k - Mx_k)$ , определяется наиболее сильно коррелированная с ним переменная  $x_l$ , и процедура формирования уравнения регрессии повторяется.

Процесс прекращается, если дальнейшее расширение или сокращение модели делают регрессию незначимой.

## 2 ПОРЯДОК ВЫПОЛНЕНИЯ РАБОТЫ

До выполнения работы необходимо самостоятельно изучить основы регрессионного анализа и правила построения множественных регрессионных моделей.

При выполнении работы необходимо по исходным данным получить полностью значимую при уровне доверительной вероятности  $\alpha = 0,95$  полиномиальную модель вида

$$y = \sum_{j=0}^m b_j \times x^j. \quad (1.18)$$

Для приведения зависимости (1.18) к виду (1.12) необходимо сформировать матрицу переменных  $x_i$  следующего вида

$$X = \begin{pmatrix} \hat{y}_1 & x_1 & x_1^2 & \dots & x_1^m \\ \hat{y}_2 & \dots & \dots & \dots & \dots \\ \hat{y}_n & x_n & x_n^2 & \dots & x_n^m \end{pmatrix} \quad (1.19)$$

где  $x_1$  – отклик,  $x_j = x^{j-1}$  ( $j = 2, \dots, m+1$ ) – факторы. С помощью пошаговой процедуры в регрессию

$$\hat{h} = g(x_2, \dots, x_m) \quad (1.20)$$

должны быть включены только значимые факторы.

Содержание отчета о лабораторной работе:

- общая постановка задачи;
- расчет основных статистических характеристик независимых переменных;
- решение задачи пошаговой регрессии.

Результаты расчетов должны приводиться как в числовой, так и в графической форме (например, в виде графика остатков). Для каждой модели должны определяться среднеквадратическая ошибка и коэффициент детерминации.

Варианты заданий приведены в приложении **A**. Пример решения задачи средствами **MathCAD** показан в приложении **B**.

## КОНТРОЛЬНЫЕ ВОПРОСЫ

1. Что считается среднеквадратической регрессией одной переменной относительно другой переменной?
2. Как строится уравнение линейной регрессии? Как определяются коэффициенты ее уравнения?
3. Как строится уравнение множественной линейной регрессии? Как определяются коэффициенты ее уравнения?
4. Как строится таблица дисперсионного анализа для уравнения регрессии?
5. Как определяется значимость регрессии в целом? Как определяется значимость переменных, входящих в уравнение регрессии?
6. Как выполняется процедура пошаговой регрессии?

## ЛИТЕРАТУРА

1. Вентцель, Е.С., Овчаров, Л.А. Теория случайных процессов и ее инженерные приложения [Текст] : Учеб. пособие для втузов. / Вентцель Е.С., Овчаров Л.А. – 2-е изд., стер. – М. : Высш. шк., 2000. – 383 с. : ил. – (Высшая математика для втузов). – ISBN 5-06-003831-9.
2. Случайные процессы [Текст] : Учеб. для вузов. / Волков И.К., Зуев С.М., Цветкова Г.М. ; Под ред. В.С. Зарубина, А.П. Крищенко. – М. : Изд-во МГТУ им. Н.Э. Баумана, 1999. – 448 с. – (Математика в техническом университете. Вып. XVIII). – ISBN 5-7038-1267-4 (Вып. XVIII) ; ISBN 5-7038-1270-4.
3. Вентцель, Е.С., Овчаров, Л.А. Теория вероятностей и ее инженерные приложения [Текст] : Учеб. пособие для втузов. / Вентцель Е.С., Овчаров Л.А. – 2-е изд., стер. – М. : Высш. шк., 2000. – 480 с. : ил. – (Высшая математика для втузов). – ISBN 5-06-003830-0.
4. Теория вероятностей [Текст] : Учеб. для вузов. / А.В. Печин-

кин, О.И. Тескин, Г.М. Цветкова и др. ; Под ред. В.С. Зарубина, А.П. Крищенко. – М. : Изд-во МГТУ им. Н.Э. Баумана, 1999. – 456 с. : ил. – (Математика в техническом университете. Вып. XVI). – ISBN 5-7038-1307-7 (Вып. XVI) ; ISBN 5-7038-1270-4.

5. Дрейпер, Н., Смит, Г. Прикладной регрессионный анализ [Текст] : монография : в 2 т. / Дрейпер Н., Смит Г. – 2-е изд., перераб. и доп. – М. : Финансы и статистика, 1986-1987. – (Математико-статистические методы за рубежом).

Т. 1 : – М. : Финансы и статистика, 1986. – 366 с. : ил.

Т. 2 : – М. : Финансы и статистика, 1987. – 351 с. : ил.

6. Химмельблау, Д. Анализ процессов статистическими методами [Текст] / Д. Химмельблау. – М. : Мир, 1973. – [957 с.] : ил.

## ПРИЛОЖЕНИЕ А

### Варианты заданий для выполнения лабораторной работы

№ варианта																			
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Начальное значение независимой переменной $x$																			
1,0	1,1	1,2	1,3	1,4	1,5	1,6	1,7	1,8	1,9	2,0	1,9	1,8	1,7	1,6	1,5	1,4	1,3	1,2	1,1
Шаг изменения независимой переменной $x$																			
0,1	0,2	0,3	0,2	0,3	0,1	0,3	0,2	0,1	0,2	0,1	0,3	0,2	0,1	0,2	0,3	0,2	0,3	0,1	0,2
Значения зависимой переменной $y$																			
9,0	6,0	3,5	4,0	297	7,0	40,0	2,0	12,8	8,8	2,8	29,0	2,3	92,0	2,0	32,0	2,0	5,0	26,8	0,1
5,4	7,0	3,0	12,0	24,5	11,5	16,5	3,2	6,0	4,0	3,6	11,0	2,6	52,0	2,5	18,5	2,5	4,4	23,6	1,4
4,2	9,0	2,8	20,0	10,5	15,0	11,5	3,8	4,8	2,3	5,0	8,0	2,9	36,0	2,5	12,5	2,5	3,4	20,2	3,4
3,6	11,0	2,6	32,0	7,0	18,5	9,0	4,4	4,2	1,4	7,0	7,0	3,3	26,0	3,0	9,0	2,5	2,6	16,2	6,2
3,2	14,0	2,5	44,0	5,5	21,5	7,5	4,6	4,0	0,8	10,6	6,0	3,7	18,0	4,0	6,0	3,0	1,6	11,4	10,7
3,0	17,0	2,4	58,0	4,5	24,5	7,0	4,8	3,8	0,5	14,8	6,0	4,0	12,0	6,0	4,5	3,0	1,2	7,4	15,1
3,0	20,0	2,4	74,0	4,0	27,5	6,0	4,8	3,6	0,3	20,2	5,2	4,3	8,0	12,5	3,0	3,5	0,6	3,6	19,6
2,8	25,0	2,3	90,0	3,5	30,0	6,0	5,0	3,6	0,2	24,4	5,1	4,5	6,0	31,5	2,5	4,0	0,4	1,6	22,5
2,6	30,0	2,3	108	3,5	32,5	5,5	5,0	3,4	0,1	30,6	4,9	4,7	4,0	48,9	1,5	5,0	0,2	0,2	25,9
2,6	37,0	2,3	126	3,5	35,0	5,0	5,0	3,4	0,1	30,6	4,7	4,7	4,0	68,3	1,5	5,0	0,2	0,2	25,9
Степень аппроксимирующего полинома																			
5	6	6	5	5	6	6	5	5	6	6	5	5	6	6	5	5	6	6	5

**ПРИЛОЖЕНИЕ Б****Решение пошаговой регрессионной задачи  
(документ MathCAD)****ПОШАГОВАЯ РЕГРЕССИЯ.**

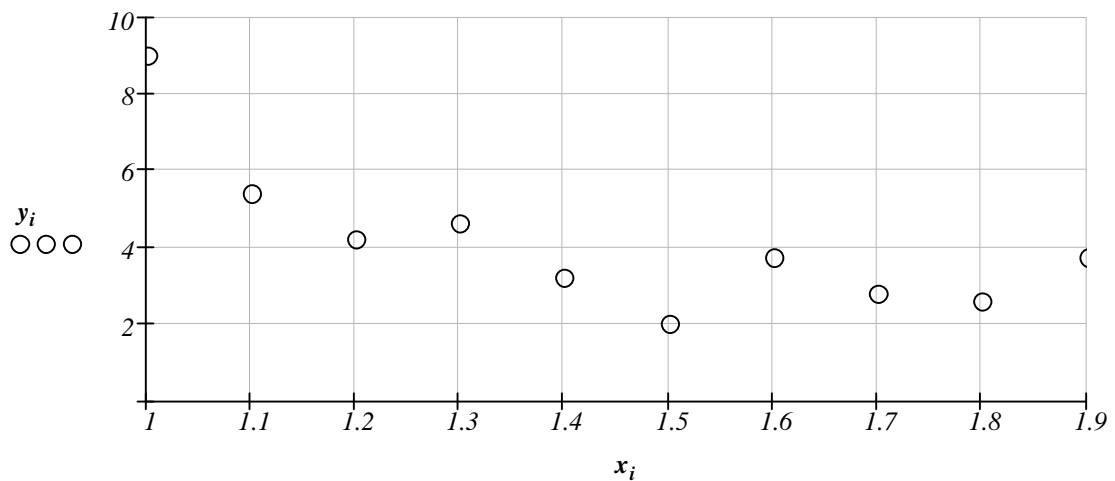
Исходные данные:

$$x = \begin{pmatrix} 1.0 \\ 1.1 \\ 1.2 \\ 1.3 \\ 1.4 \\ 1.5 \\ 1.6 \\ 1.7 \\ 1.8 \\ 1.9 \end{pmatrix} \quad y = \begin{pmatrix} 9.0 \\ 5.4 \\ 4.2 \\ 4.6 \\ 3.2 \\ 2.0 \\ 3.7 \\ 2.8 \\ 2.6 \\ 3.7 \end{pmatrix}$$

$$n = \text{rows}(x)$$

$$i = 0.. \text{last}(x)$$

График экспериментальных данных:



Порядок аппроксимирующего полинома:

$$m = 4$$

$$j = 0..m$$

$$k = 0..m$$

Факторы уравнения регрессии:

$$X^{(j)} = x^j$$

$$X^{(0)} = y$$

Матрица корреляционных моментов:

$$D_{j,k} = \text{cvar}(X^{(j)}, X^{(k)})$$

$$\text{rank}(D) = 5.0000$$

Матрица полного ранга, следовательно коэффициенты регрессии будут определяться однозначно.

Частные коэффициенты корреляции:

$$r_{j,k} = -\frac{(D^{-1})_{j,k}}{\sqrt{(D^{-1})_{j,j} \cdot (D^{-1})_{k,k}}}$$

$$r^{(0)T} = (-1.0000 \quad -0.5409 \quad 0.5153 \quad -0.4959 \quad 0.4813)$$

В регрессию включается фактор 1 (x).

**ШАГ 1.**

$$m = 1$$

$$j = 0..m$$

$$k = 0..m$$

Факторы уравнения регрессии:

$$x^{(0)} = X^{(0)}$$

$$x^{(1)} = X^{(1)}$$

Средние значения переменных:

$$M_x = 0$$

$$M_{x_j} = \text{mean}(x^{(j)})$$

$$M_x^T = (4.1200 \quad 1.4500)$$



**Матрица корреляционных моментов:**

$$D = 0$$

$$D_{j,k} = \text{cvar}(x^{\langle j \rangle}, x^{\langle k \rangle})$$

$$\text{rank}(D) = 2.0000$$

**Матрица полного ранга, следовательно коэффициенты регрессии будут определяться однозначно.**

**Частные коэффициенты корреляции:**

$$r = 0$$

$$r_{j,k} = -\frac{(D^{-1})_{j,k}}{\sqrt{(D^{-1})_{j,j} \cdot (D^{-1})_{k,k}}}$$

$$r^{\langle 0 \rangle T} = (-1.0000 \quad -0.7231)$$

**Коэффициенты регрессии:**

$$b = 0$$

$$b_j = -\frac{(D^{-1})_{0,j}}{(D^{-1})_{0,0}}$$

$$b_0 = M_{x_0} - \sum_{j=1}^m b_j \cdot M_{x_j}$$

$$b^T = (10.9921 \quad -4.7394)$$

**Расчетные значения зависимой переменной:**

$$z = b_0 + \sum_{j=1}^m b_j \cdot x^{\langle j \rangle}$$

**Сумма квадратов, относящихся к регрессии, и сумма квадратов остатка:**

$$SS_R = (|y - M_{x_0}|)^2$$

$$SS_D = (|y - z|)^2$$

**Сводный коэффициент корреляции:**

$$\sqrt{1 - \frac{SS_D}{SS_R}} = 0.7231$$

**Качество регрессии низкое.**

**Дисперсия и стандартная ошибка оценки:**

$$S_2 = \frac{SS_D}{n - m}$$

$$\sqrt{S_2} = 1.3705$$

Значения F-критерия (рассчитанное и пороговое):

$$F = \frac{SS_R}{SS_D} \cdot \frac{n - m - 1}{m}$$

$$F = 16.7695$$

$$qF(0.95, m, n - m - 1) = 5.3177$$

Регрессия статистически значима.

Значения t-критерия для коэффициентов (рассчитанные и пороговые):

$$j = 1..m$$

$$tb_j = \sqrt{\frac{(r_{0,j})^2 \cdot (n - m - 1)}{1 - (r_{0,j})^2}}$$

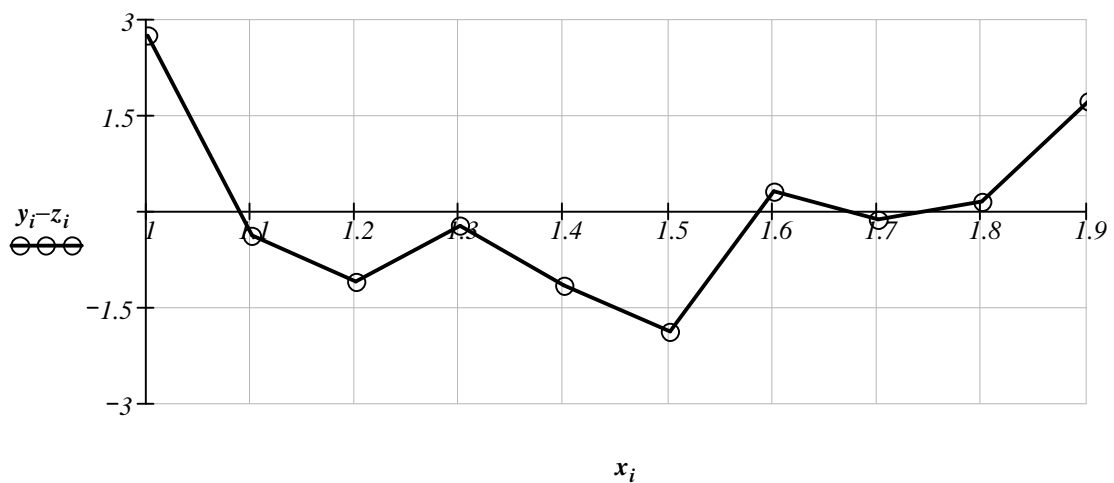
$$tb_0 = \frac{|b_0| \cdot \sqrt{n}}{\sqrt{S_2}}$$

$$tb^T = ( 25.3627 \quad 2.9613 )$$

$$qt(0.975, n - m - 1) = 2.3060$$

Все коэффициенты уравнения значимы.

График остатков:



Определение следующего значимого фактора:

$$x = 0$$

$$x \langle 0 \rangle = X \langle 0 \rangle - z$$

$$x \langle 1 \rangle = X \langle 2 \rangle$$

$$x \langle 2 \rangle = X \langle 3 \rangle$$

$$x \langle 3 \rangle = X \langle 4 \rangle$$

$$j = 0..3$$

$$k = 0..3$$

**Матрица корреляционных моментов:**

$$D = 0$$

$$D_{j,k} = \text{cvar}(x^{\langle j \rangle}, x^{\langle k \rangle})$$

$$\text{rank}(D) = 4.0000$$

**Матрица полного ранга, следовательно коэффициенты регрессии будут определяться однозначно.**

**Частные коэффициенты корреляции:**

$$r = 0$$

$$r_{j,k} = -\frac{(D^{-1})_{j,k}}{\sqrt{(D^{-1})_{j,j} \cdot (D^{-1})_{k,k}}}$$

$$r^{\langle 0 \rangle T} = (-1.0000 \quad -0.6057 \quad 0.5600 \quad -0.5094)$$

**В регрессию включается фактор 2 ( $x^2$ ).**

**ШАГ 2.**

$$m = 2$$

$$j = 0..m$$

$$k = 0..m$$

**Факторы уравнения регрессии:**

$$x = 0$$

$$x^{\langle 0 \rangle} = X^{\langle 0 \rangle}$$

$$x^{\langle 1 \rangle} = X^{\langle 1 \rangle}$$

$$x^{\langle 2 \rangle} = X^{\langle 2 \rangle}$$

**Средние значения переменных:**

$$M_x = 0$$

$$M_{x_j} = \text{mean}(x^{\langle j \rangle})$$

$$M_x^T = (4.1200 \quad 1.4500 \quad 2.1850)$$

**Матрица корреляционных моментов:**

$$D = 0$$

$$D_{j,k} = \text{cvar}(x^{\langle j \rangle}, x^{\langle k \rangle})$$

$$\text{rank}(D) = 3.0000$$

**Матрица полного ранга, следовательно коэффициенты регрессии будут определяться однозначно.**

■ Частные коэффициенты корреляции:

$$r = 0$$

$$r_{j,k} = -\frac{(D^{-1})_{j,k}}{\sqrt{(D^{-1})_{j,j} \cdot (D^{-1})_{k,k}}}$$

$$r^{(0)T} = (-1.0000 \quad -0.8599 \quad 0.8362)$$

Коэффициенты регрессии:

$$b = 0$$

$$b_j = -\frac{(D^{-1})_{0,j}}{(D^{-1})_{0,0}}$$

$$b_0 = M_{x_0} - \sum_{j=1}^m b_j \cdot M_{x_j}$$

$$b^T = (41.2156 \quad -48.1295 \quad 14.9621)$$

Расчетные значения зависимой переменной:

$$z = b_0 + \sum_{j=1}^m b_j \cdot x^{(j)}$$

Сумма квадратов, относящихся к регрессии, и сумма квадратов остатка:

$$SS_R = (|y - M_{x_0}|)^2$$

$$SS_D = (|y - z|)^2$$

Сводный коэффициент корреляции:

$$\sqrt{1 - \frac{SS_D}{SS_R}} = 0.9255$$

Качество регрессии высокое.

Дисперсия и стандартная ошибка оценки:

$$S_2 = \frac{SS_D}{n - m}$$

$$\sqrt{S_2} = 0.7973$$

Значения F-критерия (рассчитанное и пороговое):

$$F = \frac{SS_R}{SS_D} \cdot \frac{n - m - 1}{m}$$

$$F = 24.3911$$

$$qF(0.95, m, n - m - 1) = 4.7374$$

Регрессия статистически значима.

■ Значения t-критерия для коэффициентов (рассчитанные и пороговое):

$$j = 1..m$$

$$tb_j = \sqrt{\frac{(r_{0,j})^2 \cdot (n - m - 1)}{1 - (r_{0,j})^2}}$$

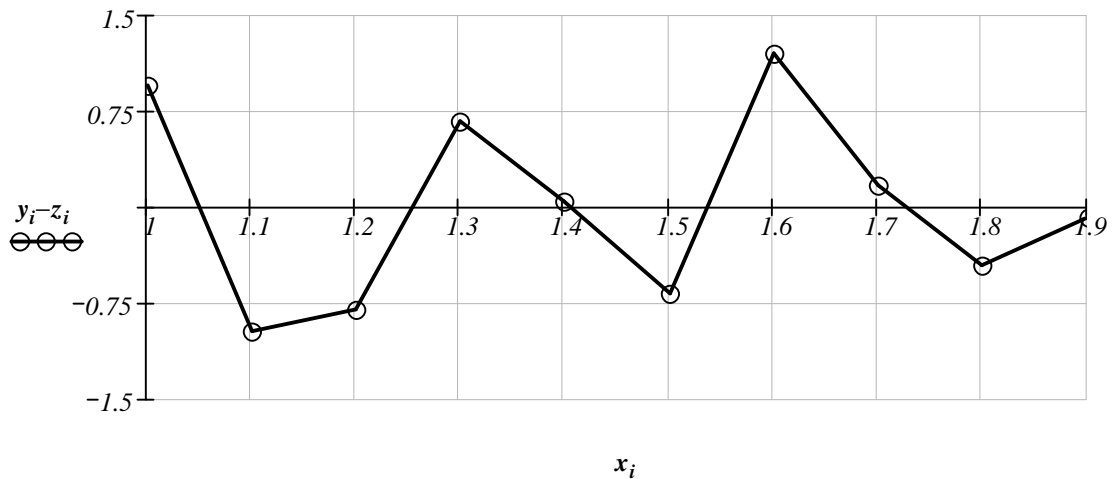
$$tb_0 = \frac{|b_0| \cdot \sqrt{n}}{\sqrt{S_2}}$$

$$tb^T = ( 163.4804 \quad 4.4575 \quad 4.0338 )$$

$$qt(0.975, n - m - 1) = 2.3646$$

Все коэффициенты уравнения значимы.

График остатков:



Определение следующего значимого фактора:

$$x = 0$$

$$x \langle 0 \rangle = X \langle 0 \rangle - z$$

$$x \langle 1 \rangle = X \langle 3 \rangle$$

$$x \langle 2 \rangle = X \langle 4 \rangle$$

$$j = 0..2$$

$$k = 0..2$$

Матрица корреляционных моментов:

$$D = 0$$

$$D_{j,k} = \text{cvar}(x \langle j \rangle, x \langle k \rangle)$$

$$\text{rank}(D) = 3.0000$$

Матрица полного ранга, следовательно коэффициенты регрессии будут определяться однозначно.

**Частные коэффициенты корреляции:**

$$r = 0$$

$$r_{j,k} = -\frac{(D^{-1})_{j,k}}{\sqrt{(D^{-1})_{j,j} \cdot (D^{-1})_{k,k}}}$$

$$r^{(0)T} = ( -1.0000 \quad 0.0917 \quad -0.0923 )$$

**В регрессию включается фактор 4 ( $x^4$ ).**

**ШАГ 3.**

$$m = 3$$

$$j = 0..m$$

$$k = 0..m$$

**Факторы уравнения регрессии:**

$$x = 0$$

$$x^{(0)} = X^{(0)}$$

$$x^{(1)} = X^{(1)}$$

$$x^{(2)} = X^{(2)}$$

$$x^{(3)} = X^{(4)}$$

**Средние значения переменных:**

$$M_x = 0$$

$$M_{x_j} = \text{mean}(x^{(j)})$$

$$M_x^T = ( 4.1200 \quad 1.4500 \quad 2.1850 \quad 5.4733 )$$

**Матрица корреляционных моментов:**

$$D = 0$$

$$D_{j,k} = \text{cvar}(x^{(j)}, x^{(k)})$$

$$\text{rank}(D) = 4.0000$$

**Матрица полного ранга, следовательно коэффициенты регрессии будут определяться однозначно.**

**Частные коэффициенты корреляции:**

$$r = 0$$

$$r_{j,k} = -\frac{(D^{-1})_{j,k}}{\sqrt{(D^{-1})_{j,j} \cdot (D^{-1})_{k,k}}}$$

$$r^{(0)T} = ( -1.0000 \quad -0.6107 \quad 0.5437 \quad -0.4234 )$$

■ **Коэффициенты регрессии:**

$$b = 0$$

$$b_j = -\frac{(D^{-1})_{0,j}}{(D^{-1})_{0,0}}$$

$$b_0 = M_{x_0} - \sum_{j=1}^m b_j \cdot M_{x_j}$$

$$\mathbf{b}^T = ( 78.0784 \quad -119.5542 \quad 52.9033 \quad -2.9595 )$$

**Расчетные значения зависимой переменной:**

$$z = b_0 + \sum_{j=1}^m b_j \cdot x_{(j)}$$

**Сумма квадратов, относящихся к регрессии, и сумма квадратов остатка:**

$$SS_R = \left( |y - M_{x_0}| \right)^2$$

$$SS_D = \left( |y - z| \right)^2$$

**Сводный коэффициент корреляции:**

$$\sqrt{1 - \frac{SS_D}{SS_R}} = 0.9393$$

**Качество регрессии высокое.**

**Дисперсия и стандартная ошибка оценки:**

$$S_2 = \frac{SS_D}{n - m}$$

$$\sqrt{S_2} = 0.7721$$

**Значения F-критерия (рассчитанное и пороговое):**

$$F = \frac{SS_R}{SS_D} \cdot \frac{n - m - 1}{m}$$

$$F = 16.9821$$

$$qF(0.95, m, n - m - 1) = 4.7571$$

**Регрессия статистически значима.**

**Значения t-критерия для коэффициентов (рассчитанные и пороговое):**

$$j = 1 .. m$$

$$t_{b_j} = \sqrt{\frac{(r_{0,j})^2 \cdot (n - m - 1)}{1 - (r_{0,j})^2}}$$

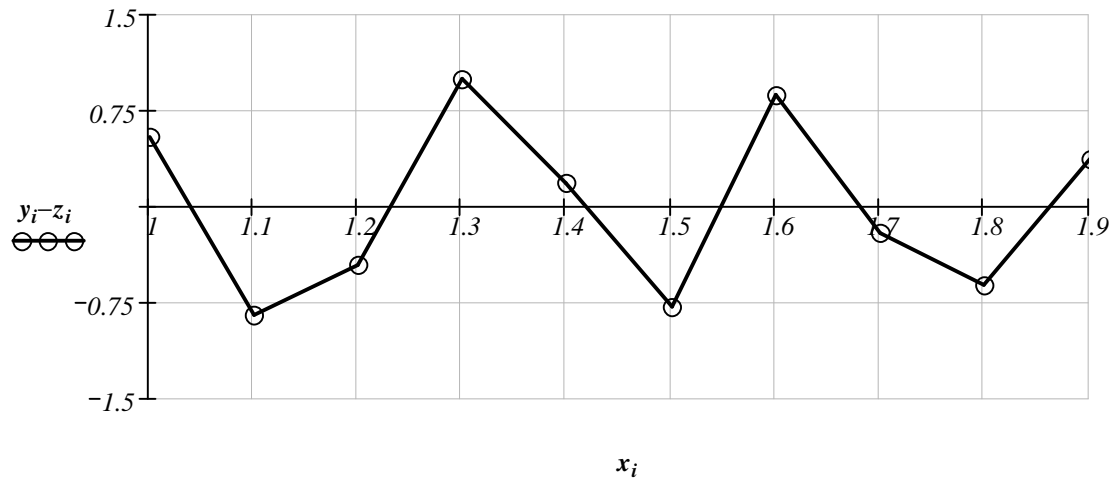
$$t_{b_0} = \frac{|b_0| \cdot \sqrt{n}}{\sqrt{S_2}}$$

$$t_b^T = ( 319.7701 \quad 1.8893 \quad 1.5867 \quad 1.1448 )$$

$$qt(0.975, n - m - 1) = 2.4469$$

Три коэффициента уравнения незначимы.

График остатков:



Введение фактора снизило качество уравнения регрессии.

В регрессию включается фактор 3 ( $x^3$ ).

**ШАГ 4.**

$$m = 3$$

$$j = 0..m$$

$$k = 0..m$$

Факторы уравнения регрессии:

$$x = 0$$

$$x \langle 0 \rangle = X \langle 0 \rangle$$

$$x \langle 1 \rangle = X \langle 1 \rangle$$

$$x \langle 2 \rangle = X \langle 2 \rangle$$

$$x \langle 3 \rangle = X \langle 3 \rangle$$

Средние значения переменных:

$$M_x = 0$$

$$M_{x_j} = \text{mean}(x \langle j \rangle)$$

$$M_x^T = ( 4.1200 \quad 1.4500 \quad 2.1850 \quad 3.4075 )$$

Матрица корреляционных моментов:

$$D = 0$$

$$D_{j,k} = \text{cvar}(x \langle j \rangle, x \langle k \rangle)$$

$$\text{rank}(D) = 4.0000$$



■ Матрица полного ранга, следовательно коэффициенты регрессии будут определяться однозначно.

Частные коэффициенты корреляции:

$$r = 0$$

$$r_{j,k} = -\frac{(D^{-1})_{j,k}}{\sqrt{(D^{-1})_{j,j} \cdot (D^{-1})_{k,k}}}$$

$$r^{(0)T} = ( -1.0000 \quad -0.5741 \quad 0.5048 \quad -0.4409 )$$

Коэффициенты регрессии:

$$b = 0$$

$$b_j = -\frac{(D^{-1})_{0,j}}{(D^{-1})_{0,0}}$$

$$b_0 = M_{x_0} - \sum_{j=1}^m b_j \cdot M_{x_j}$$

$$b^T = ( 91.9565 \quad -158.3526 \quad 92.7855 \quad -17.8904 )$$

Расчетные значения зависимой переменной:

$$z = b_0 + \sum_{j=1}^m b_j \cdot x^{(j)}$$

Сумма квадратов, относящихся к регрессии, и сумма квадратов остатка:

$$SS_R = (|y - M_{x_0}|)^2$$

$$SS_D = (|y - z|)^2$$

Сводный коэффициент корреляции:

$$\sqrt{1 - \frac{SS_D}{SS_R}} = 0.9404$$

Качество регрессии высокое.

Дисперсия и стандартная ошибка оценки:

$$S_2 = \frac{SS_D}{n - m}$$

$$\sqrt{S_2} = 0.7650$$

Значения F-критерия (рассчитанное и пороговое):

$$F = \frac{SS_R}{SS_D} \cdot \frac{n - m - 1}{m}$$

$$F = 17.3016$$

$$qF(0.95, m, n - m - 1) = 4.7571$$

Регрессия статистически значима.

Значения t-критерия для коэффициентов (рассчитанные и пороговое):

$$j = 1..m$$

$$tb_j = \sqrt{\frac{(r_{0,j})^2 \cdot (n - m - 1)}{1 - (r_{0,j})^2}}$$

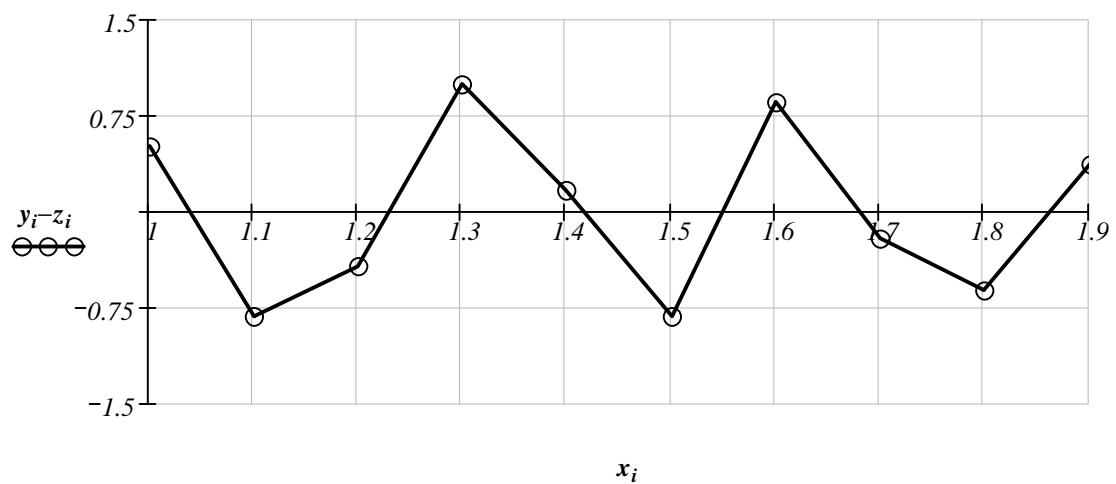
$$tb_0 = \frac{|b_0| \cdot \sqrt{n}}{\sqrt{S_2}}$$

$$tb^T = ( 380.1344 \quad 1.7176 \quad 1.4325 \quad 1.2034 )$$

$$qt(0.975, n - m - 1) = 2.4469$$

Три коэффициента уравнения незначимы.

График остатков:



Введение фактора не повысило качества уравнения регрессии.

Принимается модель, полученная на шаге 2.

$$f(x) = 41.2156 - 48.1295 \cdot x + 14.9621 \cdot x^2$$

## ■ График аппроксимирующей характеристики.

