|  |  |  |
| --- | --- | --- |
| Министерство науки и высшего образования  Российской Федерации | | |
| Федеральное государственное бюджетное  образовательное учреждение высшего образования | | |
| «Новосибирский государственный технический университет» | | |
|  | | |
| Кафедра теоретической и прикладной информатики | | |
|  | | |
| Лабораторная работа № 7 | | |
| по дисциплине «Статистические методы анализа данных» | | |
|  | | |
|  | | |
|  | | |
|  | Факультет: | ПМИ |
| Группа: | ПМИ-02 |
| Вариант: | 6 |
| Студент: | Сидоров Даниил, |
|  | Дюков Богдан |
| Преподаватель: | Попов Александр Александрович. |
|  |  |
|
|  |  |
| Новосибирск | | |
| 2023 | | |

**Постановка задачи**

1. Изучить функциональные возможности программного комплекса ОДА, пользуясь данными методическими рекомендациями и встроенной контекстной помощью.
2. Подготовить исходные данные из лабораторных работ № 2, № 6.
3. Решить задачу выбора «наилучшей» модели регрессии для задач из лабораторных работ № 2 и 6.

**Ход работы**

**Исходные данные из лабораторной работы №2**

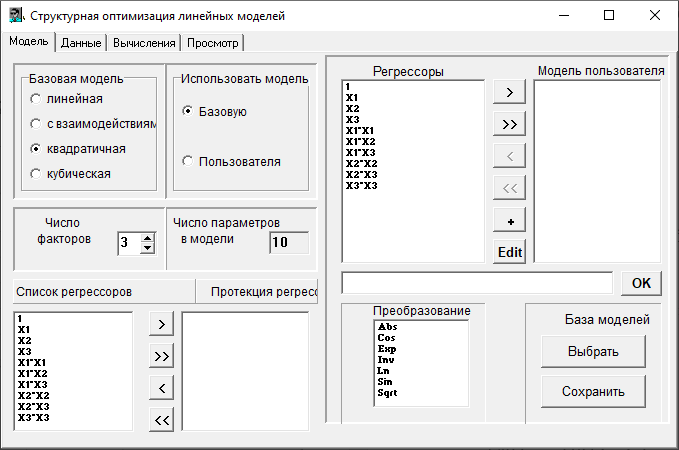
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **№** | **x1** | **x2** | **x3** | **y** |
| 1 | -1 | -1 | -1 | -7,421921045 |
| 2 | -1 | -1 | 1 | -2,139324492 |
| 3 | -1 | -0,33 | -1 | -2,090986381 |
| 4 | -1 | -0,33 | 1 | -1,350575114 |
| 5 | -1 | 0,33 | -1 | -1,69529944 |
| 6 | -1 | 0,33 | 1 | -0,586504844 |
| 7 | -1 | 1 | -1 | 0,650083657 |
| 8 | -1 | 1 | 1 | 5,35306572 |
| 9 | 0 | -1 | -1 | -8,867318951 |
| 10 | 0 | -1 | 1 | -0,362126435 |
| 11 | 0 | -0,33 | -1 | -3,687216239 |
| 12 | 0 | -0,33 | 1 | 6,217390024 |
| 13 | 0 | 0,33 | -1 | -1,994746844 |
| 14 | 0 | 0,33 | 1 | 7,613558733 |
| 15 | 0 | 1 | -1 | 3,420614445 |
| 16 | 0 | 1 | 1 | 9,791004984 |
| 17 | 1 | -1 | -1 | -8,710567529 |
| 18 | 1 | -1 | 1 | 4,429908371 |
| 19 | 1 | -0,33 | -1 | -4,660703112 |
| 20 | 1 | -0,33 | 1 | 4,743967873 |
| 21 | 1 | 0,33 | -1 | -0,130615528 |
| 22 | 1 | 0,33 | 1 | 12,95403137 |
| 23 | 1 | 1 | -1 | 3,408641288 |
| 24 | 1 | 1 | 1 | 16,0689259 |

Истинная модель:

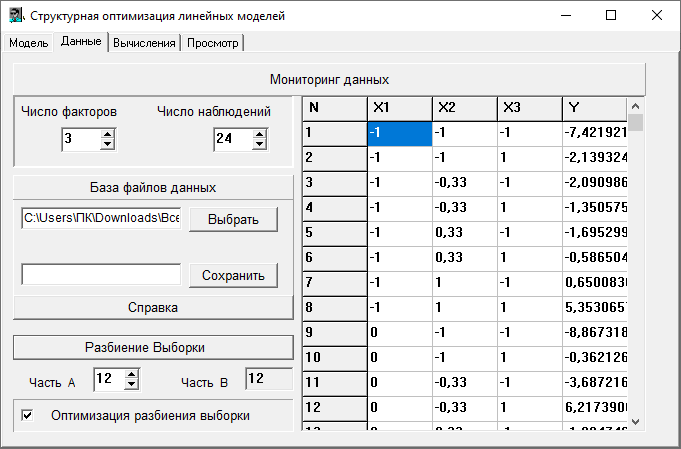
;

**Поиск «наилучшей» модели регрессии для задачи из лабораторной работы №2**

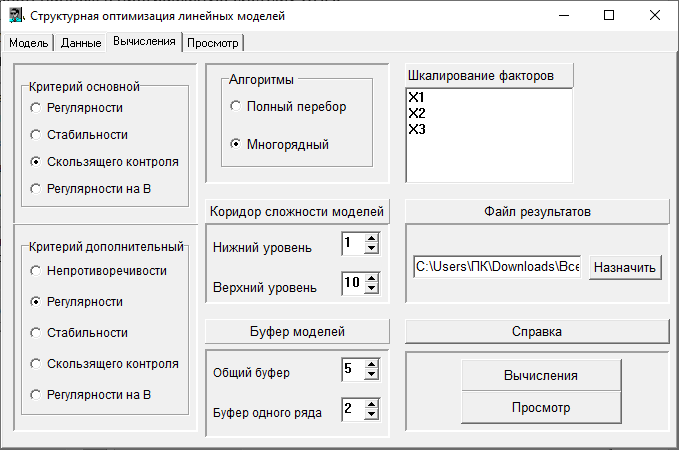
1. Ввод данных в программную систему ОДА:



Наша истинная модель дополняется до полной.



Выборка разбивалась на две равные части.



Основным критерием был выбран критерий скользящего контроля, а дополнительным - критерий регулярности. Число лучших моделей, отбираемых по основному критерию, равно 5. Среди них будет отбираться наиболее подходящая. Для буфера одного ряда, в случае многорядного алгоритма, рекомендуется выбирать значение, которое больше 1, что и было сделано.

1. Содержимое файла результата:

Вектор оценок параметров полной модели

-6.38E-0004 2.34E+0000 5.06E+0000 3.94E+0000-3.40E-0001

1.35E+0000 2.28E+0000 2.70E-0002-1.47E-0001 1.50E+0000

Сумма квадратов отклонений для полной модели= 3.53E+0001

Список регрессоров:

1): 1

2): X1

3): X2

4): X3

5): X1\*X1

6): X1\*X2

7): X1\*X3

8): X2\*X2

9): X2\*X3

10): X3\*X3

Список регрессоров лучших моделей

---------------------------------------------------------------------

m 1 -я мод. 2 -я мод. 3 -я мод. 4 -я мод. 5 -я мод.

---------------------------------------------------------------------

1): 0 1 1 1 1

---------------------------------------------------------------------

2): 2 2 2 2 2

---------------------------------------------------------------------

3): 3 3 3 3 3

---------------------------------------------------------------------

4): 4 4 4 4 4

---------------------------------------------------------------------

5): 0 0 0 0 0

---------------------------------------------------------------------

6): 6 6 6 6 6

---------------------------------------------------------------------

7): 7 7 7 7 7

---------------------------------------------------------------------

8): 0 0 0 0 0

---------------------------------------------------------------------

9): 0 0 9 0 9

---------------------------------------------------------------------

10): 10 0 0 10 10

Значения 1внешнего критерия:

2.5205 2.5205 2.7170 2.2638 2.4241

Лучшая по 1 критерию 4 модель.

Значения 2 внешнего критерия:

154.7933 154.7933 154.6034 154.7933 154.6034

Лучшая по 2 критерию 3 модель.

Остаточные суммы квадратов для лучших моделей:

36.1643 36.1643 35.8766 36.1643 35.8766

Критерий Маллоуса для лучших моделей:

2.3604 2.3604 4.2462 4.3604 6.2462

Оценки параметров лучших моделей:

1 0.0000 1.2897 1.2897 -0.0006 -0.0006

2 2.3366 2.3366 2.3366 2.3366 2.3366

3 5.0628 5.0628 5.0628 5.0628 5.0628

4 3.9381 3.9381 3.9381 3.9381 3.9381

5 0.0000 0.0000 0.0000 0.0000 0.0000

6 1.3544 1.3544 1.3544 1.3544 1.3544

7 2.2785 2.2785 2.2785 2.2785 2.2785

8 0.0000 0.0000 0.0000 0.0000 0.0000

9 0.0000 0.0000 -0.1470 0.0000 -0.1470

10 1.2897 0.0000 0.0000 1.2904 1.2904

Сравнение прогноза и отклика для модели 4

N Y Y^ Y-Y^ (Y-Y^)/Y

1 -7.42 -6.41 1.01 -13.57

2 -2.14 -3.10 -0.96 44.70

3 0.65 1.00 0.35 54.13

4 5.35 4.32 -1.03 -19.28

5 -8.71 -9.01 -0.30 3.41

6 4.43 3.43 -1.00 -22.67

7 3.41 3.83 0.42 12.27

8 16.07 16.26 0.19 1.19

9 -2.09 -3.93 -1.84 87.96

10 -1.35 -0.61 0.74 -54.76

11 -1.70 -1.48 0.21 -12.54

12 -0.59 1.84 2.42 -413.13

13 -8.87 -7.71 1.16 -13.04

14 -0.36 0.16 0.53 -145.55

15 -3.69 -4.32 -0.63 17.14

16 6.22 3.56 -2.66 -42.79

17 -1.99 -0.98 1.02 -50.99

18 7.61 6.90 -0.72 -9.39

19 3.42 2.41 -1.01 -29.41

20 9.79 10.29 0.50 5.10

21 -4.66 -4.71 -0.05 1.01

22 4.74 7.73 2.98 62.84

23 -0.13 -0.47 -0.34 261.78

24 12.95 11.96 -0.99 -7.67

Сравнение прогноза и отклика для модели 3

N Y Y^ Y-Y^ (Y-Y^)/Y

1 -7.42 -6.56 0.86 -11.59

2 -2.14 -2.95 -0.81 37.83

3 0.65 1.15 0.50 76.74

4 5.35 4.17 -1.18 -22.02

5 -8.71 -9.15 -0.44 5.10

6 4.43 3.57 -0.86 -19.35

7 3.41 3.97 0.57 16.59

8 16.07 16.11 0.04 0.27

9 -2.09 -3.98 -1.89 90.28

10 -1.35 -0.56 0.79 -58.35

11 -1.70 -1.43 0.26 -15.40

12 -0.59 1.79 2.37 -404.86

13 -8.87 -7.86 1.01 -11.38

14 -0.36 0.31 0.67 -186.15

15 -3.69 -4.37 -0.68 18.45

16 6.22 3.61 -2.61 -42.01

17 -1.99 -0.93 1.07 -53.42

18 7.61 6.85 -0.76 -10.03

19 3.42 2.56 -0.86 -25.12

20 9.79 10.14 0.35 3.60

21 -4.66 -4.76 -0.10 2.05

22 4.74 7.77 3.03 63.86

23 -0.13 -0.42 -0.29 224.63

24 12.95 11.91 -1.04 -8.04

1. Сравнение оценок

Программа выбрала 2 наилучшие модели (модель 4 и модель 3), среди которых была выбрана модель 3, так она даёт меньшую ошибку прогноза по сравнению с моделью 4. Сравним следующие оценки:

|  |  |  |  |
| --- | --- | --- | --- |
| Истинное значение | из л.р. №2 | с помощью ОДА для полной модели | с помощью ОДА для модели 3 |
| 1 | 1,50143 | -0.0006 | 1.2897 |
| 2 | 2,33657 | 2.34 | 2.3366 |
| 5 | 5,06282 | 5.06 | 5.0628 |
| 4 | 3,93806 | 3.94 | 3.9381 |
| 1.5 | 1,35442 | -0.34 | 0.0000 |
| 2.5 | 2,27846 | 1.35 | 1.3544 |
| 0.02 | -0,34001 | 2.28 | 2.2785 |
| 0.01 | 0,027 | 0.027 | 0.0000 |
| 0.0000 | 0.0000 | -0.147 | -0.1470 |
| 0.0000 | 0.0000 | 1.50 | 0.0000 |

Если мы рассмотрим среднее абсолютное отклонение, то оценки из л.р. №2 лучше всех остальных оценок, которые получены с помощью ОДА.

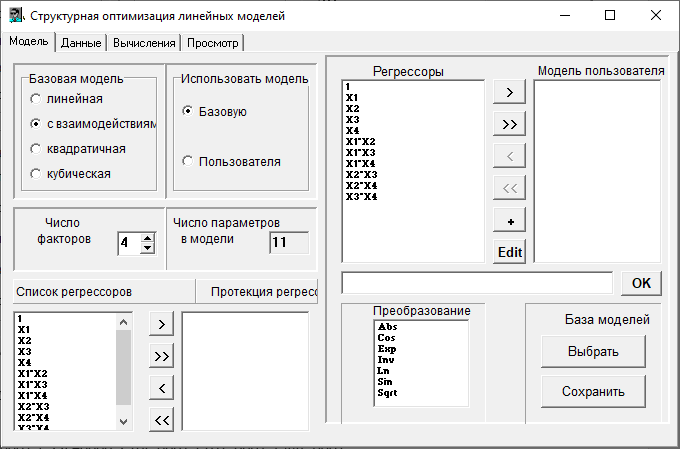
**Исходные данные из лабораторной работы №6**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **№** | **X1** | **X2** | **X3** | **X4** | **Y** |
| 1 | 7 | 26 | 6 | 60 | 78.5 |
| 2 | 1 | 29 | 15 | 52 | 74.3 |
| 3 | 11 | 56 | 8 | 20 | 104.3 |
| 4 | 11 | 31 | 8 | 47 | 87.6 |
| 5 | 7 | 52 | 6 | 33 | 95.9 |
| 6 | 11 | 55 | 9 | 22 | 109.2 |
| 7 | 3 | 71 | 17 | 6 | 102.7 |
| 8 | 1 | 31 | 22 | 44 | 72.5 |
| 9 | 2 | 54 | 18 | 22 | 93.1 |
| 10 | 21 | 47 | 4 | 26 | 115.9 |
| 11 | 21 | 40 | 23 | 34 | 83.8 |
| 12 | 11 | 66 | 9 | 12 | 113.3 |
| 13 | 10 | 68 | 8 | 12 | 109.4 |

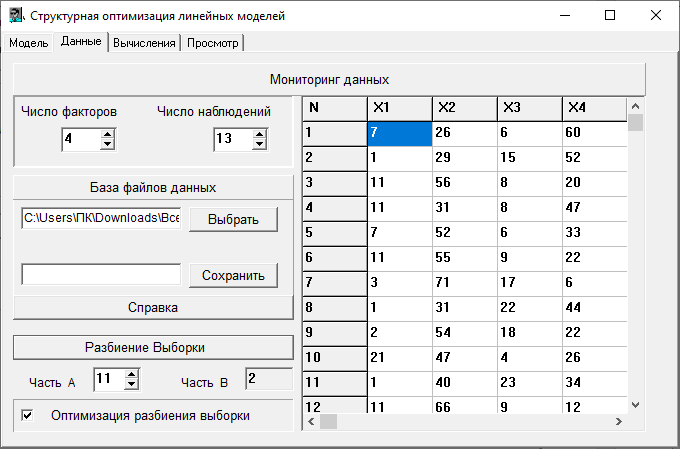
В качестве полной модели выбиралась следующая:

**Поиск «наилучшей» модели регрессии для задачи из лабораторной работы №6**

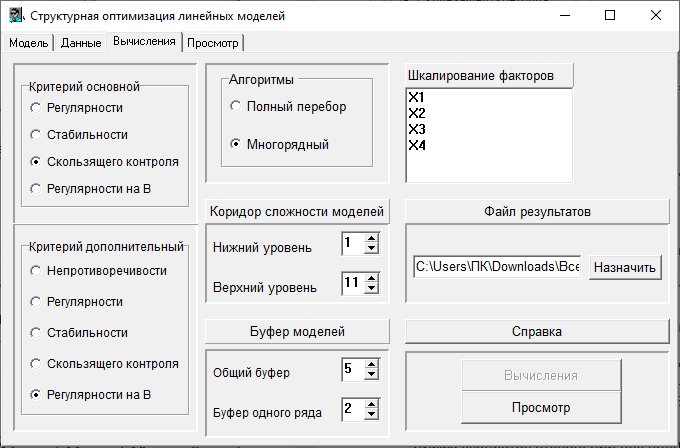
1. Ввод данных в программную систему ОДА:



В программе выбиралась 4-х факторная модель с взаимодействиями.



Выборка разбивалась на две части: А – 11 наблюдений, В – 2 наблюдения.



1. Содержимое файла результата:

Вектор оценок параметров полной модели

-4.66E+0001 4.21E+0000 9.18E-0001-3.01E+0000 9.84E-0001

-6.53E-0004 1.04E-0001-5.21E-0002 8.50E-0002 1.51E-0002

3.22E-0002

Сумма квадратов отклонений для полной модели= 1.73E+0000

Список регрессоров:

1): 1

2): X1

3): X2

4): X3

5): X4

6): X1\*X2

7): X1\*X3

8): X1\*X4

9): X2\*X3

10): X2\*X4

11): X3\*X4

Список регрессоров лучших моделей

---------------------------------------------------------------------

m 1 -я мод. 2 -я мод. 3 -я мод. 4 -я мод. 5 -я мод.

---------------------------------------------------------------------

1): 1 1 1 1 1

---------------------------------------------------------------------

2): 0 0 0 0 2

---------------------------------------------------------------------

3): 3 3 3 3 3

---------------------------------------------------------------------

4): 0 0 4 4 0

---------------------------------------------------------------------

5): 5 5 5 5 5

---------------------------------------------------------------------

6): 6 6 6 6 0

---------------------------------------------------------------------

7): 0 0 7 0 0

---------------------------------------------------------------------

8): 0 0 0 0 0

---------------------------------------------------------------------

9): 0 9 9 9 0

---------------------------------------------------------------------

10): 10 10 10 10 0

---------------------------------------------------------------------

11): 11 11 11 11 0

Значения 1внешнего критерия:

6.5899 6.0536 1.8169 6.1528 6.5655

Лучшая по 1 критерию 3 модель.

Значения 2 внешнего критерия:

3.6698 1.5601 4.3660 0.7394 8.3828

Лучшая по 2 критерию 4 модель.

Остаточные суммы квадратов для лучших моделей:

35.1406 17.0138 2.0795 10.8597 47.9727

Критерий Маллоуса для лучших моделей:

39.6029 20.6585 7.4027 15.5478 50.4298

Оценки параметров лучших моделей:

1 124.2731 38.8279 14.0746 50.8519 71.6483

2 0.0000 0.0000 0.0000 0.0000 1.4519

3 -0.3129 0.2153 0.0917 -0.0794 0.4161

4 0.0000 0.0000 -5.4608 -3.8730 0.0000

5 -0.7236 0.2121 0.2753 -0.0183 -0.2365

6 0.0207 0.0473 0.0591 0.0521 0.0000

7 0.0000 0.0000 0.0707 0.0000 0.0000

8 0.0000 0.0000 0.0000 0.0000 0.0000

9 0.0000 0.0287 0.1192 0.0860 0.0000

10 0.0046 0.0095 0.0158 0.0121 0.0000

11 -0.0155 -0.0154 0.0595 0.0353 0.0000

Сравнение прогноза и отклика для модели 3

N Y Y^ Y-Y^ (Y-Y^)/Y

1 83.80 83.26 -0.54 -0.64

2 95.90 95.79 -0.11 -0.12

3 72.50 72.82 0.32 0.44

4 74.30 74.09 -0.21 -0.28

5 115.90 115.89 -0.01 -0.01

6 78.50 78.67 0.17 0.21

7 87.60 87.55 -0.05 -0.05

8 93.10 93.97 0.87 0.94

9 109.20 108.71 -0.49 -0.44

10 104.30 104.30 -0.00 -0.00

11 102.70 102.29 -0.41 -0.40

12 113.30 113.94 0.64 0.56

13 109.40 109.23 -0.17 -0.16

Сравнение прогноза и отклика для модели 4

N Y Y^ Y-Y^ (Y-Y^)/Y

1 83.80 83.16 -0.64 -0.77

2 95.90 96.38 0.48 0.51

3 72.50 73.24 0.74 1.02

4 74.30 74.14 -0.16 -0.22

5 115.90 117.16 1.26 1.09

6 78.50 78.89 0.39 0.50

7 87.60 86.50 -1.10 -1.26

8 93.10 93.95 0.85 0.92

9 109.20 106.90 -2.30 -2.11

10 104.30 104.83 0.53 0.51

11 102.70 102.86 0.16 0.16

12 113.30 112.80 -0.50 -0.45

13 109.40 109.68 0.28 0.26

1. Вывод

Программа выбрала 2 наилучшие модели (модель 3 и модель 4), среди которых была выбрана модель 3, так она даёт меньшую ошибку прогноза по сравнению с моделью 4. Таким образом, оптимальная модель: