

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ  
ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО  
ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ

«Московский государственный технический университет имени Н.Э. Баумана»  
(МГТУ им. Н.Э. Баумана)

ФАКУЛЬТЕТ ФН

КАФЕДРА

«ВЫЧИСЛИТЕЛЬНАЯ МАТЕМАТИКА И МАТЕМАТИЧЕСКАЯ ФИЗИКА»

Направление: Математика и компьютерные науки

Дисциплина: Численные методы

Домашняя работа №1-2

«Метод наименьших квадратов и модели регрессии»

Группа: ФН11-52Б

Вариант №8

Студент: Зеликова В.И.

Преподаватель: Кутыркин В.А.

Оценка:

Москва 2022

## **ЗАДАНИЕ 2.1**

Дана модель линейной регрессии:

$$Y = x_*^0 + z_1 x_*^1 + z_2 x_*^2 + z_3 x_*^3 + z_4 x_*^4 + z_5 x_*^5 + z_6 x_*^6 + \varepsilon. \quad (1)$$

Для оценки неизвестных вектора тренда  $\mathbf{x}_* = [x_*^0, x_*^1, \dots, x_*^k] \in \mathbf{E}^{k+1}$  и параметра  $\sigma$  случайной составляющей  $\varepsilon \sim N(0, \sigma)$  модели линейной регрессии (20) проводился эксперимент, в котором получены  $m=20$  значений  $y^1, \dots, y^m \in \mathbf{R}$  (см. *Таблицы 2 и 3*) регрессора модели (20) для  $m$  различных наборов  $\mathbf{z}^1 = [z_1^1, \dots, z_6^1], \dots, \mathbf{z}^m = [z_1^m, \dots, z_6^m] \in \mathbf{R}^6$  (см. *Таблицу 4*) шести факторов модели (1).

Требуется получить оценки вектора тренда  $\mathbf{x}_* = [x_*^0, x_*^1, \dots, x_*^k] \in \mathbf{E}^{k+1}$  и параметра  $\sigma$  случайной составляющей  $\varepsilon \sim N(0, \sigma)$  модели линейной регрессии (1). Если возможно, редуцировать модель регрессии (1) до приведённой модели. Результаты расчётов проиллюстрировать графически, сопроводив их необходимыми комментариями. ►

### **Решение:**

Исходные данные:  $N=8$ ;  $n=52$ ;  $\alpha=0,1$ . Значения регрессора модели для 8 варианта задания:

$y + \alpha$	$y$
14,65	14,55
15,06	14,96
13,37	13,27
15,9	15,8
10,41	10,31
14,25	14,15
15,14	15,04
11,54	11,44
10,05	9,95
12,38	12,28
13,8	13,7
14,62	14,52
13,11	13,01
12,6	12,5
12,8	12,7
9,82	9,72
13,3	13,2
15,14	15,04
15,01	14,91
12,44	12,34

Заданные наборы факторов модели.

$z_1$	$z_2$	$z_3$	$z_4$	$z_5$	$z_6$
1,158574	1,194067	1,745872	1,566271	1,825556	1,942503
1,238868	1,913419	1,182653	1,044649	1,304209	1,924039
1,564043	1,561357	1,070589	1,778954	1,226447	1,824122
1,737266	1,798975	1,952239	1,752281	1,247871	1,54796
1,364544	1,03122	1,380596	1,688101	1,987396	1,058504
1,535295	1,742973	1,580401	1,063356	1,999237	1,425459
1,780725	1,306711	1,972594	1,68627	1,582629	1,767235
1,135044	1,139164	1,686178	1,220069	1,034577	1,019745
1,246498	1,114597	1,079653	1,333415	1,054445	1,156743
1,416456	1,349223	1,68038	1,003235	1,471908	1,095523
1,611866	1,972991	1,443953	1,014008	1,91699	1,182531
1,520585	1,427992	1,464156	1,011505	1,108341	1,981536
1,229896	1,304392	1,852107	1,705496	1,725639	1,21482
1,726829	1,866756	1,074984	1,09888	1,983154	1,256935
1,77279	1,363353	1,227454	1,076754	1,656758	1,675253
1,418256	1,072481	1,123447	1,438917	1,059481	1,080325
1,119724	1,947356	1,372631	1,635578	1,94058	1,112827
1,728446	1,802332	1,365001	1,184759	1,119633	1,880032
1,161107	1,359294	1,956206	1,143406	1,49144	1,688437
1,963561	1,271859	1,250008	1,19367	1,466262	1,624409

Средствами Microsoft Excel строим для заданных  $y$  и  $z_1, z_2, z_3, z_4, z_5, z_6$  факторов модели линейную регрессию. Получили следующие данные:

	<i>t-статистика</i>	<i>Нижние 95%</i>	<i>Верхние 95%</i>
Y-пересечение	-12,94173785	-0,127603904	-0,091096219
$z_1$	0,480942136	-0,005462293	0,008590798
$z_2$	1092,251018	2,996339201	3,008215639
$z_3$	1126,83178	2,996780466	3,008293429
$z_4$	-0,820912798	-0,008444615	0,00379407
$z_5$	1,021176019	-0,002787248	0,007784244
$z_6$	1193,405432	2,994358289	3,005219057

Подпрограмма «Регрессия», расположенная в разделе «Анализ данных» программы Excel, позволяет получить доверительные интервалы уровня значимости 5% для каждой компоненты вектора параметров тренда модели линейной регрессии. Может возникнуть ситуация, когда оценки некоторых компонент этого вектора на заданном уровне значимости не отличаются от нуля. Тогда ту из этих компонент, оценка которой наименее отличается от

нуля (согласно  $t$ -статистике), можно считать нулевой, если доверительный интервал для оценки этой компоненты содержит значение нуля.

Таковой можно считать компоненту  $z_1$ . Соответственно можно рассмотреть редуцированную модель линейной регрессии, зависящую от факторов  $z_2, z_3, z_4, z_5, z_6$ .

Аналогично исключим  $z_4$ :

	<i>t-статистика</i>	<i>Нижние 95%</i>	<i>Верхние 95%</i>
Y-пересечение	-14,78146326	-0,123050465	-0,091866121
$z_2$	1130,521008	2,996727525	3,008119721
$z_3$	1177,360671	2,99684232	3,007780885
$z_4$	-0,872325828	-0,008296124	0,003498873
$z_5$	1,086999497	-0,002510068	0,007668842
$z_6$	1279,538964	2,99510035	3,005158093

Далее исключим  $z_5$ :

	<i>t-статистика</i>	<i>Нижние 95%</i>	<i>Верхние 95%</i>
Y-пересечение	-17,56145463	-0,123970001	-0,097134365
$z_2$	1165,684534	2,99741811	3,008399724
$z_3$	1209,219187	2,99659286	3,007175499
$z_5$	1,026985388	-0,002591072	0,007409685
$z_6$	1290,213129	2,995219492	3,005132154

Согласно полученным оценкам редуцировать регрессию дальше нельзя.

Таким образом, оценки параметров вектора тренда (столбец «Коэффициенты»):

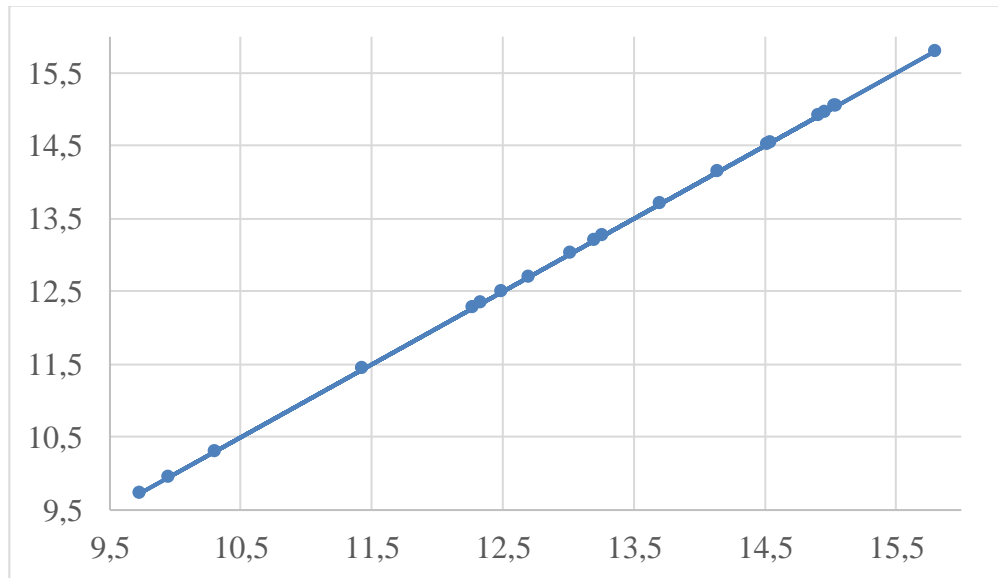
	<i>Коэффициенты</i>	<i>t-статистика</i>	<i>Нижние 95%</i>	<i>Верхние 95%</i>
Y-пересечение	-0,107947602	-18,7034735	-0,120182692	-0,095712511
$z_2$	3,003816776	1239,30174	2,998678554	3,008954997
$z_3$	3,002383087	1231,161395	2,99721336	3,007552813
$z_6$	2,99946835	1348,231592	2,994752107	3,004184594

Оценка параметра  $\sigma$  : 0,003281466

<i>Наблюдение</i>	<i>Предсказанное <math>y</math></i>	<i>Остатки</i>
1	14,54706372	0,002936282
2	14,96148394	-0,001483939
3	13,26779726	0,00220274
4	15,80027006	-0,000270065
5	10,30967566	0,00032434
6	14,14821232	0,001787677
7	15,04042113	-0,000421134
8	11,43513749	0,004862507
9	9,95124349	-0,00124349
10	12,27600214	0,003997863
11	13,70082024	-0,000820235
12	14,52099045	-0,000990451
13	13,01475584	-0,004755843
14	12,49709592	0,002904082
15	12,69747049	0,002529507
16	9,727007735	-0,007007735
17	13,20060648	-0,000606484
18	15,04327989	-0,003279893
19	14,91281567	-0,00281567
20	12,33785006	0,00214994

<i>Предсказанное <math>y</math></i>	<i><math>y</math></i>
14,54706372	14,5470637
14,96148394	14,9614839
13,26779726	13,2677973
15,80027006	15,8002701
10,30967566	10,3096757
14,14821232	14,1482123
15,04042113	15,0404211
11,43513749	11,4351375
9,95124349	9,95124349
12,27600214	12,2760021
13,70082024	13,7008202
14,52099045	14,5209905
13,01475584	13,0147558
12,49709592	12,4970959
12,69747049	12,6974705
9,727007735	9,72700773
13,20060648	13,2006065
15,04327989	15,0432799
14,91281567	14,9128157
12,33785006	12,3378501

Ниже представлен график линии тренда:



### Результаты:

Была построена модель линейной регрессии, зависящей от факторов  $z_1$ ,  $z_2$ ,  $z_3$ ,  $z_4$ ,  $z_5$ ,  $z_6$ . В ходе построения было выяснено, что влияние факторов  $z_1$ ,  $z_4$ ,  $z_5$  несущественно. Были найдены оценки вектора тренда и параметра  $\sigma$ :

$$y = -0,107948 + z_2 * 3,003817 + z_3 * 3,002383 + z_6 * 2,999468$$

$$\sigma = 0,003281466$$

Графики экспериментального  $y$  и предсказанным накладываются друг на друга, поскольку для всех 20 экспериментов эти значения отличаются приблизительно на 0,001.

## **ЗАДАНИЕ 2.2**

Дана модель полиномиальной регрессии:

$$Y = x_*^0 + t \cdot x_*^1 + t^2 \cdot x_*^2 + t^3 \cdot x_*^3 + \varepsilon. \quad (2)$$

Для оценки неизвестных вектора тренда  $\mathbf{x}_* = [x_*^0, x_*^1, x_*^2, x_*^3] \in \mathbf{E}^4$  и параметра  $\sigma$  случайной составляющей  $\varepsilon \sim N(0, \sigma)$  модели полиномиальной регрессии (21) проводился эксперимент, в котором получены  $m=20$  значений  $y^1, \dots, y^m \in \mathbf{R}$  (см. Таблицы 5 и 6) регрессора модели (2) для  $m$  попарно различных значений  $t_1, \dots, t_m \in \mathbf{R}$  (см. Таблицу 7) единственного фактора модели (2). Требуется получить оценки вектора тренда  $\mathbf{x}_* = [x_*^0, x_*^1, x_*^2, x_*^3] \in \mathbf{E}^4$  и параметра  $\sigma$  случайной составляющей  $\varepsilon \sim N(0, \sigma)$  модели полиномиальной регрессии (2). Результаты расчётов проиллюстрировать графически, сопроводив их необходимыми комментариями. ►

### **Решение:**

Значения входных данных для 8 варианта задания:

t	t·t	t·t·t	y	y + α
0,05	0,0025	0,000125	-3,91	-3,81
0,1	0,01	0,001	-3,74	-3,64
0,15	0,0225	0,003375	-3,58	-3,48
0,2	0,04	0,008	-3,43	-3,33
0,25	0,0625	0,015625	-3,29	-3,19
0,3	0,09	0,027	-3,15	-3,05
0,35	0,1225	0,042875	-3,02	-2,92
0,4	0,16	0,064	-2,88	-2,78
0,45	0,2025	0,091125	-2,75	-2,65
0,5	0,25	0,125	-2,6	-2,5
0,55	0,3025	0,166375	-2,44	-2,34
0,6	0,36	0,216	-2,28	-2,18
0,65	0,4225	0,274625	-2,09	-1,99
0,7	0,49	0,343	-1,89	-1,79
0,75	0,5625	0,421875	-1,66	-1,56
0,8	0,64	0,512	-1,41	-1,31
0,85	0,7225	0,614125	-1,13	-1,03
0,9	0,81	0,729	-0,82	-0,72
0,95	0,9025	0,857375	-0,48	-0,38
1	1	1	-0,1	0

Вычисления произведем в программе Excel, воспользовавшись пунктом “Анализ данных”→ “Регрессия”.

Получили следующие данные:

	<i>Коэффициенты</i>	<i>Стандартная ошибка</i>	<i>t-статистика</i>	<i>Нижние 95%</i>	<i>Верхние 95%</i>
Y-пересечение	-4,102249742	0,003025101	-1356,07049	-4,108662669	-4,095836815
t	3,999111473	0,02434377	164,2765897	3,947504987	4,05071796
t·t	-3,981811073	0,053189958	-74,86020389	-4,094568746	-3,8690534
t·t·t	3,986180284	0,033351611	119,5198706	3,915478027	4,056882542

Редуцировать модель регрессии нельзя.

Оценки параметров вектора тренда представлены в столбце «Коэффициенты».

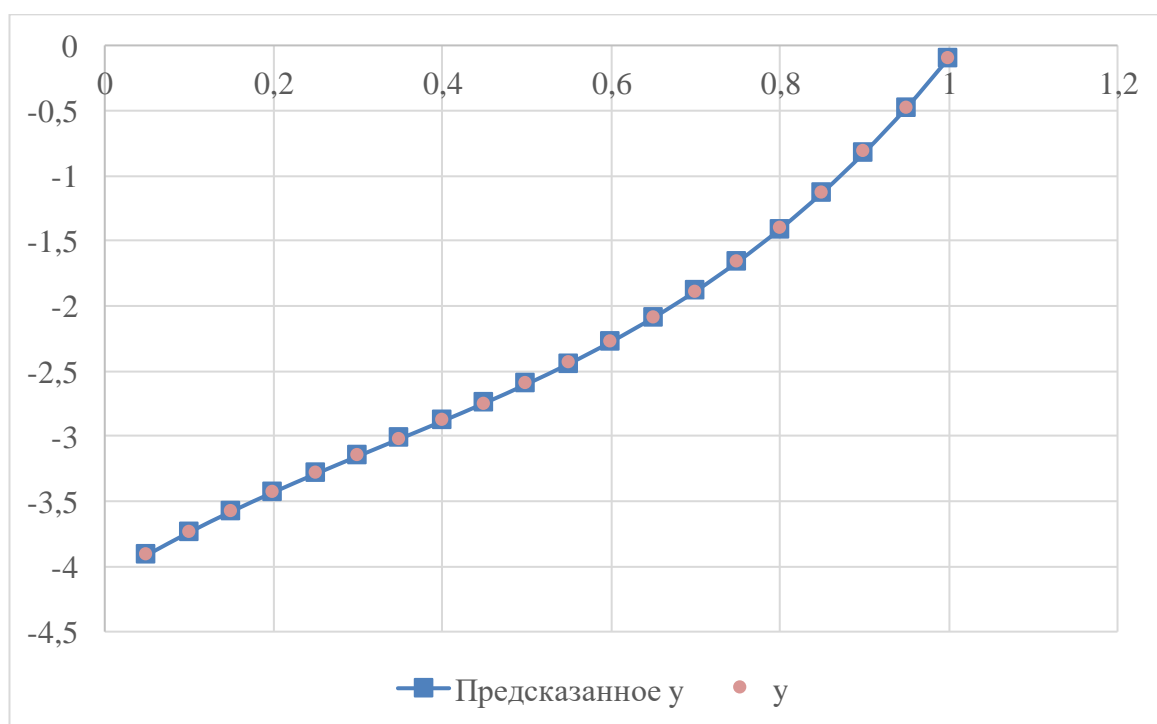
Оценка параметра  $\sigma$ : 0,002769397

<i>Наблюдение</i>	<i>Предсказанное y</i>	<i>Остатки</i>
1	-3,911750423	0,001750423
2	-3,738170525	-0,001829475
3	-3,578520412	-0,001479588
4	-3,429810448	-0,000189552
5	-3,289050999	-0,000949001
6	-3,153252429	0,003252429
7	-3,019425103	-0,000574897
8	-2,884579386	0,004579386
9	-2,745725643	-0,004274357
10	-2,599874238	-0,000125762
11	-2,444035536	0,004035536
12	-2,275219903	-0,004780097
13	-2,090437702	0,000437702
14	-1,886699299	-0,003300701
15	-1,661015058	0,001015058
16	-1,410395344	0,000395344
17	-1,131850523	0,001850523
18	-0,822390958	0,002390958
19	-0,479027014	-0,000972986
20	-0,098769057	-0,001230943



Предсказанное $y$	$y$
-6,766199887	-6,77
-6,462270498	-6,46
-6,182985428	-6,18
-5,923076990	-5,92
-5,677277500	-5,68
-5,440319271	-5,44
-5,206934617	-5,21
-4,971855854	-4,97
-4,729815296	-4,73
-4,475545257	-4,48
-4,203778051	-4,20
-3,909245993	-3,91
-3,586681398	-3,59
-3,230816579	-3,23
-2,836383851	-2,83
-2,398115528	-2,40
-1,910743926	-1,91
-1,369001357	-1,37
-0,767620137	-0,77
-0,101332580	-0,10

Ниже представлен график сравнения экспериментального  $y$  с предсказанным:



## Результаты

Была построена модель полиномиальной регрессии, представленной через модель линейной регрессии, зависящей от факторов  $t^1, t^2, t^3$ . Были найдены оценки вектора тренда и параметра  $\sigma$ :

$$y = -4,10225 + 3,999112 \cdot t - 3,981811 \cdot t \cdot t + 3,98618 \cdot t \cdot t \cdot t$$

$$\sigma = 0,003142723$$

Графики экспериментального  $y$  и предсказанным совпадают друг с другом. Линия тренда представляет собой прямую линию, поскольку регрессия, зависящая от факторов  $t^1, t^2, t^3$  нелинейна.