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 ФУНДАМЕНТАЛЬНЫХ НАУК
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 №9

Студент: Очкин Н.В.

Преподаватель: Кутыркин В.А.

Оценка:

Москва, 2024

Задание 2.1

Дана модель линейной регрессии:

$$Y = x_*^0 + z_1 x_*^1 + z_2 x_*^2 + z_3 x_*^3 + z_4 x_*^4 + z_5 x_*^5 + z_6 x_*^6 + \varepsilon \tag{1}$$

Для оценки неизвестных вектора тренда $^{>}x_* = \left[x_*^0, x_*^1, \ldots, x_*^k\right\rangle \in {}^{>}\mathbb{E}^{k+1}$ и параметра σ от случайной составляющей $\varepsilon \sim \mathcal{N}(0,\sigma)$ модели линейной регрессии (1). проводился эксперимент, в котором получены $m = 20$ значений $y^1, \ldots, y^m \in \mathbb{R}$ регрессора модели (1) для m различных наборов $^{<}z^1 = \left\langle z_1^1, \ldots, z_6^1\right], \ldots, {}^{<}z^m = \langle z_1^m, \ldots, z_6^m] \in\in {}^{<}\mathbb{R}^6$ шести факторов модели (1).

Требуется получить оценки вектора тренда $^{>}x_* = \left[x_*^0, x_*^1, \ldots, x_*^k\right\rangle \in {}^{>}\mathbb{E}^{k+1}$ и параметра σ от случайной составляющей $\varepsilon \sim \mathcal{N}(0,\sigma)$ модели линейной регрессии (1). Если возможно, редуцировать модель регрессии (1) до приведённой модели. Результаты расчётов проиллюстрировать графически, сопроводив их необходимыми комментариями.

Решение

$\mathrm{N} = 9, \alpha = -0.025$

z^1	z^2	z^3	z^4	z^5	z^6	$y + \alpha$	y
1.158574	1.194067	1.745872	1.566271	1.825556	1.942503	14.77	14.795
1.238868	1.913419	1.182653	1.044649	1.304209	1.924039	13.41	13.435
1.564043	1.561357	1.070589	1.778954	1.226447	1.824122	13.84	13.865
1.737266	1.798975	1.952239	1.752281	1.247871	1.54796	13.6	13.625
1.364544	1.03122	1.380596	1.688101	1.987396	1.058504	13.23	13.255
1.535295	1.742973	1.580401	1.063356	1.999237	1.425459	14.88	14.905
1.780725	1.306711	1.972594	1.68627	1.582629	1.767235	15.39	15.415
1.135044	1.139164	1.686178	1.220069	1.034577	1.019745	9.56	9.585
1.246498	1.114597	1.079653	1.333415	1.054445	1.156743	10.37	10.395
1.416456	1.349223	1.68038	1.003235	1.471908	1.095523	11.95	11.975
1.611866	1.972991	1.443953	1.014008	1.91699	1.182531	14.13	14.155
1.520585	1.427992	1.464156	1.011505	1.108341	1.981536	13.83	13.855
1.229896	1.304392	1.852107	1.705496	1.725639	1.21482	12.51	12.535
1.726829	1.866756	1.074984	1.09888	1.983154	1.256935	14.9	14.925
1.77279	1.363353	1.227454	1.076754	1.656758	1.675253	15.31	15.335
1.418256	1.072481	1.123447	1.438917	1.059481	1.080325	10.67	10.695
1.119724	1.947356	1.372631	1.635578	1.94058	1.112827	12.52	12.545
1.728446	1.802332	1.365001	1.184759	1.119633	1.880032	14.18	14.205
1.161107	1.359294	1.956206	1.143406	1.49144	1.688437	13.02	13.045
1.963561	1.271859	1.250008	1.19367	1.466262	1.624409	15.16	15.185

Разобъем данные на **Z** - матрица регрессоров (с добавлением свободного члена) и **Y** - вектор наблюдений:

Z =

z^0	z^1	z^2	z^3	z^4	z^5	z^6
1.0	1.158574	1.194067	1.745872	1.566271	1.825556	1.942503
1.0	1.238868	1.913419	1.182653	1.044649	1.304209	1.924039
1.0	1.564043	1.561357	1.070589	1.778954	1.226447	1.824122
1.0	1.737266	1.798975	1.952239	1.752281	1.247871	1.54796
1.0	1.364544	1.03122	1.380596	1.688101	1.987396	1.058504
1.0	1.535295	1.742973	1.580401	1.063356	1.999237	1.425459
1.0	1.780725	1.306711	1.972594	1.68627	1.582629	1.767235
1.0	1.135044	1.139164	1.686178	1.220069	1.034577	1.019745
1.0	1.246498	1.114597	1.079653	1.333415	1.054445	1.156743
1.0	1.416456	1.349223	1.68038	1.003235	1.471908	1.095523
1.0	1.611866	1.972991	1.443953	1.014008	1.91699	1.182531
1.0	1.520585	1.427992	1.464156	1.011505	1.108341	1.981536
1.0	1.229896	1.304392	1.852107	1.705496	1.725639	1.21482
1.0	1.726829	1.866756	1.074984	1.09888	1.983154	1.256935
1.0	1.77279	1.363353	1.227454	1.076754	1.656758	1.675253
1.0	1.418256	1.072481	1.123447	1.438917	1.059481	1.080325
1.0	1.119724	1.947356	1.372631	1.635578	1.94058	1.112827
1.0	1.728446	1.802332	1.365001	1.184759	1.119633	1.880032
1.0	1.161107	1.359294	1.956206	1.143406	1.49144	1.688437
1.0	1.963561	1.271859	1.250008	1.19367	1.466262	1.624409

Y =

y
14.795
13.435
13.865
13.625
13.255
14.905
15.415
9.585
10.395
11.975
14.155
13.855
12.535
14.925
15.335
10.695
12.545
14.205
13.045
15.185

Найдем число наблюдений n и количество регрессоров k:

$$n = 20;\ k = 6$$

Для оценки параметров $\hat{x}$ применим метод наименьших квадратов (МНК):

$$\hat{x} = (Z^T Z)^{-1} Z^T Y$$

$$\hat{x} = [0.01702005, 2.99863545, 0.00618907, -0.00102846, 0.00046174, 2.99956383, 3.00008523]$$

Вычислим предсказания:

$$\hat{\mathrm{y}} = [14.80102522, 13.42738099, 13.86772605, 13.62346337, 13.25145528,$$
$$14.90380052, 15.41265467, 9.58908442, 10.39441943, 11.97328504,$$
$$14.15944035, 13.85381575, 12.53271329, 14.92561512, 15.3360968,$$
$$10.69504736, 12.54552937, 14.20896433, 13.04480987, 15.18367277]$$

Найдем t-статистику и доверительный интервал:

$$t_i = \frac{\hat{x_i}}{SE(\hat{x_i})} \text{ -- t-статистика}$$

$$\hat{x_i} \pm t_{\alpha/2} \cdot SE(\hat{x_i}) \text{ -- доверительный интервал}$$

где:

$$SE(\hat{x_i}) = \sqrt{\sigma^2 \cdot (Z^{\mathrm{T}}Z)^{-1}_{ii}} \text{ -- стандартная ошибка}$$
$$\sigma^2 = \frac{1}{n-k-1}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \text{ -- оценка дисперсии}$$

Для нахождения $t_{\alpha/2}$ воспользуемся модулем «`stats`» библиотеки «`scipy`», которая динамически определит нужное критическое значение из таблицы Стьюдента в зависимости от заданных уровня значимости теста α (в данном случае 95% для двустороннего интервала) и количества степеней свободы k (в данном случае n - k - 1 = 13)

	Y-пересечение	z^1	z^2	z^3	z^4	z^5	z^6
Коэффициенты	0.017	2.9986	0.0062	-0.001	0.0005	2.9996	3.0001
t-статистика	1.7929	820.6219	2.0042	-0.3436	0.1451	1091.222	1062.342
Нижние 95%	-0.0035	2.9907	-0.0005	-0.0075	-0.0064	2.9936	2.994
Верхние 95%	0.0375	3.0065	0.0129	0.0054	0.0073	3.0055	3.0062

Цветом выделен коэффициент z^4, т.к. модуль его t-статистики наименьший среди факторов, в доверительный интервал которых входит 0. Исключим z^4 и вычислим регрессию заново.

	Y-пересечение	z^1	z^2	z^3	z^5	z^6
Коэффициенты	0.0176	2.9986	0.0061	-0.001	2.9996	3.0001
t-статистика	2.1664	852.139	2.0908	-0.3347	1135.6755	1101.5763
Нижние 95%	0.0002	2.9911	-0.0002	-0.007	2.9939	2.9942
Верхние 95%	0.0351	3.0062	0.0124	0.0051	3.0053	3.0059

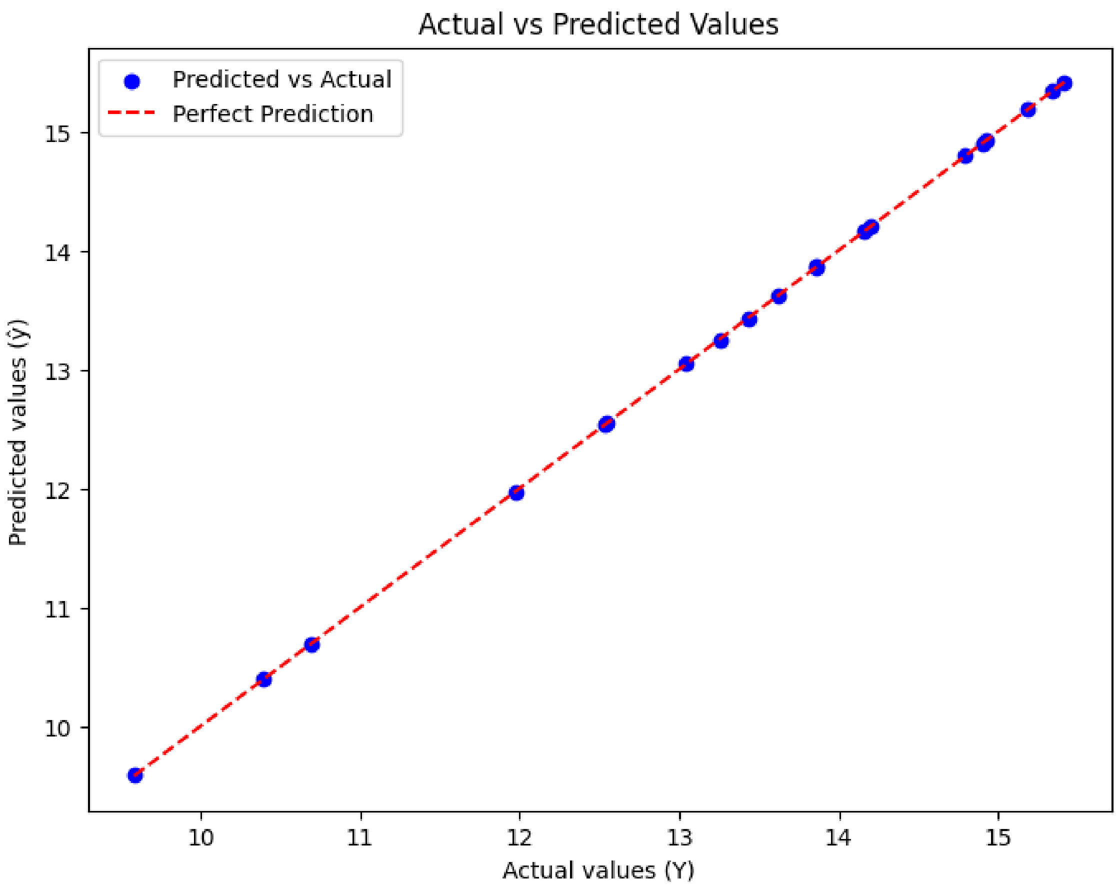
Теперь исключим z^3.

	Y-пересечение	z^1	z^2	z^5	z^6
Коэффициенты	0.0162	2.9988	0.0063	2.9994	2.9999
t-статистика	2.4046	894.7026	2.2389	1195.9259	1158.7859
Нижние 95%	0.0018	2.9917	0.0003	2.9941	2.9944
Верхние 95%	0.0306	3.006	0.0122	3.0048	3.0054

Оценка дисперсии σ^2 = 1e-05; σ = 0.00357

Построим график зависимости $\hat{y}$ от y:

$\hat{y}$	y
14.801	14.795
13.4271	13.435
13.8671	13.865
13.6239	13.625
13.2512	13.255
14.904	14.905
15.413	15.415
9.5894	9.585
10.3941	10.395
11.9737	11.975
14.1596	14.155
13.8539	13.855
12.5329	12.535
14.9254	14.925
15.336	15.335
10.6947	10.695
12.5453	12.545
14.209	14.205
13.0453	13.045
15.1836	15.185



Вывод: предсказанный y ($\hat{y}$) и заданный y близки по значениями, так как их расхождение крайне мало.

$$\mathbf{Y} = 0.0162 + 2.9988 \cdot \mathbf{z}^1 + 0.0063 \cdot \mathbf{z}^2 + 2.9994 \cdot \mathbf{z}^5 + 2.9999 \cdot \mathbf{z}^6$$