Лабораторная работа №5

Аппроксимация эмпирических данных

1. Аппроксимация эмпирических данных с использованием Excel

В инженерной практике часто возникает необходимость описать в виде функциональной зависимости связь между величинами, заданными таблично, в виде набора n точек с координатами (X_i, y_i) (i = 1,2, … , n), которая позволила бы «сгладить» экспериментальные погрешности, получить промежуточные и экстраполяционные значения функций, изначально не содержащиеся в исходной табличной информации. В связи с чем, определена цель исследования – установить вид эмпирической зависимости $y = f(X; a_1, a_2, .. a_m)$ и определить значения неизвестных параметров a_1, a_2, .. a_m, наилучшим образом описывающих экспериментальные данные.

В задачи исследования входит:

используя метод наименьших квадратов провести аппроксимацию многочленами первой и второй и экспоненциальной зависимостью; построить линии тренда и провести сравнительный анализ полученных результатов.

Подобрать вид функциональной зависимости можно исходя из теоретических соображений или анализируя расположение точек на координатной плоскости. Определение параметров при известном виде зависимости осуществляют методом наименьших квадратов, согласно которому наилучшими коэффициентами $a_{1,}a_2, \ldots, a_m$ считаются те, для которых сумма квадратов отклонений найденной эмпирической функции от заданных значений функции

$$s(a_1, a_2, \ldots, a_m) = \sum_{i=1}^{n} \left[f\left(x_i; a_{1,}a_2, \ldots, a_m\right) - y_i \right]^2 \qquad (1)$$

будет минимальной.

Для того, чтобы найти коэффициенты $a_1, a_2, \ldots, a_m$, доставляющих минимум функции S, определяемой формулой (1), используем необходимое условие экстремума функции нескольких переменных – равенство нулю частных производных первого порядка [3]. В результате получим систему для определения коэффициентов $a_i (i = 1,2, \ldots, m)$:

$$\frac{\partial S}{\partial a_1} = 0; \ \frac{\partial S}{\partial a_2} = 0; \ldots\ldots \ ; \frac{\partial S}{\partial a_m} = 0. \qquad (2)$$

Конкретный вид системы (2) зависит от вида аппроксимирующей функции. В случае линейной зависимости $y = a_1 + a_2 x$ система (2) примет вид:

$$\begin{cases} a_1 n + a_2 \sum_{i=0}^{n} x_i = \sum_{i=1}^{n} y_i\,, \\ a_1 \sum_{i=1}^{n} x_i + a_2 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i\,. \end{cases} \qquad (3)$$

В случае квадратичной зависимости $y = a_1 + a_2 x + a_3 x^2$ будем иметь:

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

В некоторых случаях, например, экспоненциальной зависимости $= a_1 \cdot e^{a_2 x}$, задачу сначала нужно линериазовать т.е. свести к линейной, путем логарифмирования [2].

Пусть функция $y = f(x)$ задана таблицей 1. Требуется выяснить какая из функций линейная, квадратичная или экспоненциальная наилучшим образом аппроксимирует функцию $y = f(x)$.

Таблица 1

Экспериментальные данные

x_i	y_i	x_i	y_i	x_i	y_i	x_i	y_i	x_i	y_i
12,85	154,77	9,65	81,43	7,74	55,86	5,02	24,98	1,86	3,91
12,32	145,59	9,63	80,97	7,32	47,63	4,65	22,87	1,76	3,22
11,43	108,37	9,22	79,04	7,08	48,03	4,53	20,32	1,11	1,22
10,59	100,76	8,44	61,76	6,87	36,85	3,24	9,06	0,99	1,10
10,21	98,32	8,07	60,54	5,23	25,65	2,55	6,23	0,72	0,53

Для проведения расчетов в табличном процессоре Microsoft Excel данные целесообразно расположить в виде таблицы 2 [1].

Таблица 2

Экспериментальные данные для проведения расчетов в табличном процессоре Microsoft Excel

	A	B	C	D	E	F	G	H	I
1	x	y	x^2	xy	x^3	x^4	x^2y	lny	$xlny$
2	12,85	154,77	165,12	1988,79	2121,82	27265,4	25556,01	5,04	64,79
3	12,32	145,59	151,78	1793,66	1869,96	23037,9	22098,00	4,98	61,36
4	11,43	108,37	130,64	1238,66	1493,27	17068,0	14157,99	4,69	53,56
5	10,59	100,75	112,15	1066,94	1187,65	12577,2	11298,92	4,61	48,85
6	10,21	98,32	104,24	1003,84	106433	10866,8	10249,28	4,59	46,85
7	9,55	81,43	91,20	777,656	898,63	8317,90	7426,62	4,40	42,02
8	9,53	80,97	90,82	771,644	893,06	8248,44	7353,77	4,39	41,88
9	9,22	79,04	85,01	728,748	783,78	7226,43	6719,06	4,37	40,29
10	8,44	51,75	71,23	436,77	601,22	5074,23	3686,34	3,95	33,31
11	8,07	50,54	65,12	407,857	525,56	4241,25	3291,41	3,92	31,66
12	7,74	55,85	59,91	432,279	463,68	3588,92	3345,84	4,02	31,14
13	7,32	47,53	53,58	347,919	392,22	2871,07	2546,77	3,86	28,27
14	7,08	48,03	50,13	340,052	354,89	2512,66	2407,57	3,87	27,41
15	5,87	35,85	34,46	210,439	324,24	1187,28	1235,28	3,58	21,01
16	5,23	25,55	27,35	133,626	143,06	748,18	698,87	3,24	16,95
17	5,02	24,98	25,20	125,399	126,51	635,06	629,51	3,22	16,15
18	4,55	22,87	20,70	104,058	100,54	428,59	473,47	3,13	14,24
19	4,53	20,32	20,52	92,0496	92,96	421,11	416,98	3,01	13,64
20	3,24	9,05	10,50	29,322	34,01	110,20	95,00	2,20	7,14
21	2,55	5,23	6,50	13,3365	16,58	42,28	34,01	1,65	4,22

22	1,85	3,91	3,42	7,2335	6,43	11,71	13,38	1,36	2,52
23	1,75	3,22	3,06	5,635	5,45	9,38	9,86	1,17	2,05
24	1,11	1,22	1,23	1,3542	1,37	1,52	1,50	0,20	0,22
25	0,99	1,10	0,98	1,089	0,97	0,96	1,08	0,10	0,09
26	0,72	0,53	0,52	0,3816	0,37	0,27	0,27	-0,63	-0,46
27	163,08	1279,01	1402,97	12289,3	13502,5	12289,3	125964,2	79,49	657,57

Аппроксимируем $y = f(x)$ линейной функцией $y = a_1 + a_2 x$. Для определения коэффициентов a_1 и a_2 воспользуемся системой (3) и итоговыми суммами таблицы 2, расположенными в ячейках A27, B27, C27 и D27:

$$\begin{cases} 25a_1 + \ 163{,}08a_2 = 1279{,}01, \\ 163{,}08a_1 + \ 1402{,}97a_2 = 12289{,}30. \end{cases}$$

Решив систему средствами Microsoft Excel, получим $a_1 = -24{,}74$ и $a_2 = 11{,}63$. Таким образом, линейная аппроксимация имеет вид y= -24,74+11,63x.

Аппроксимируем $y = f(x)$ квадратичной функцией $y = a_1 + +a_2 x + a_3 x^2$. Для определения коэффициентов a_1, a_2 и a_3 воспользуемся системой (4). Используя итоговые суммы таблицы 2, расположенные в ячейках A27, B27, C27, D27, E27, F27, G27 запишем систему в виде:

$$\begin{cases} 25a_1 + 163{,}08a_2 + 1402{,}97a_3 = 1279{,}01, \\ 163{,}08a_1 + 1402{,}97a_2 + 13502{,}57a_3 = 12289{,}30, \\ 1402{,}97a_1 + 13502{,}57a_2 + 12289{,}30a_3 = 125964{,}22. \end{cases}$$

В результате решения средствами Microsoft Excel получим $a_1 = 1,596,\ a_2 = -0,621,\ a_3 = 0,955.$ Квадратичная аппроксимация имеет вид .$y = 1,596 - 0,621 \cdot x + 0,955 \cdot x^2$.

Аппроксимируем функцию $y = f(x)$ экспоненциальной зависимостью $y = a_1 \cdot e^{a_2 x}$. Для определения коэффициентов a_1 и a_2 прологарифмируем значения y_i и, используя итоговые суммы таблицы 2, расположенные в ячейках A27, C27, H27 и I27 получим систему:

$$\begin{cases} 25c + 163,08a_2 = 79,49, \\ 163,08c + 1402,97a_2 = 657,57, \end{cases}$$

где $c = \ln(a_1)$.

Решив систему, средствами Microsoft Excel, найдем $c = 0,5061, a_2 = 0,40999.$После потенцирования получим $a_1 = 1,6589.$Экспоненциальная аппроксимация имеет вид $y = 1,6589e^{0,4099x}$.

Построим в Excel графики полученных зависимостей и линии их трендов с использованием функции ЛИНЕЙН [1].

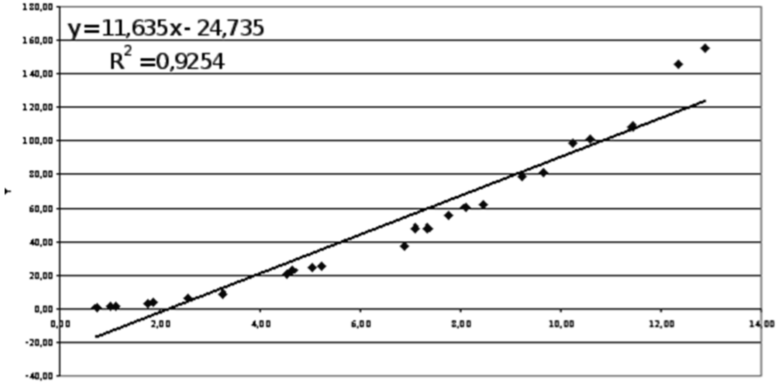

Рис.1 График линейной аппроксимации

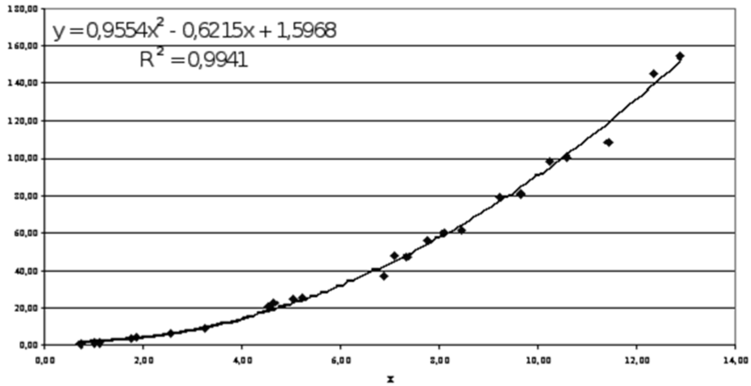

Рис. 2 График квадратичной аппроксимации

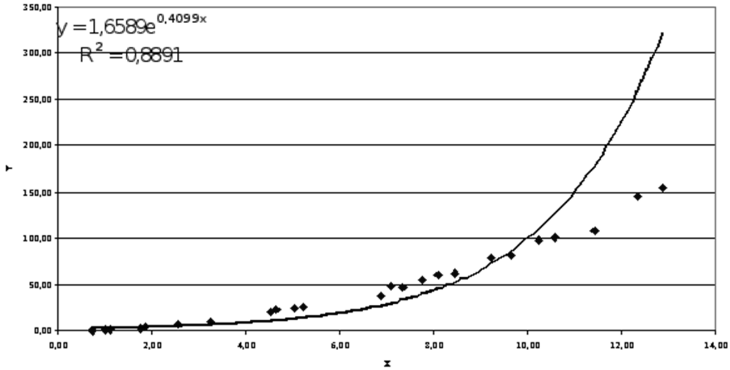

Рис. 3 График экспоненциальной аппроксимации

Анализ результатов расчетов и построенные графики показывают, что квадратичная аппроксимация наилучшим образом описывает экспериментальные данные.

Библиографический список

1. Апроксимация в Excel [Электронный ресурс]. – Режим доступа: http://tgspa.ru/info/education/faculties/ffi/ito/programm/aproksimazia / 1.3.html (дата обращения 28.09.2015)

2. Беришвили, О. Н. Линейная алгебра и аналитическая геометрия: учеб. пособие / О.Н. Беришвили. – Самара: РИЦ СГСХА, 2012. – 301 с.

3. Беришвили, О. Н. Методы оптимальных решений: учеб. пособие / О. Н. Беришвили, С. В. Плотникова. – Самара: РИЦ СГСХА, 2013. – 180 с.

Проверка адекватности модели по особенностям остаточных величин

Представлены данные о доходах по акциям x и балансовой прибыли y по 11 предприятиям одной отрасли, ден. ед.

x	3	4	5	7	8	10	11	12	15	20	30
y	12	13	20	19	31	24	41	28	52	55	103

Задание

Проверить выполнение следующих требований:

1. Уровни ε_i яда остатков имеют случайный характер.
2. Математическое ожидание уровней ряда остатков равно нулю.
3. Значения ε_i независимы друг от друга, т.е. отсутствует автокорреляция.

1. Для проверки случайности ряда остатков можно использовать *критерий поворотных точек (пиков)*. Предварительно составляют таблицу данных:

$\tilde{y}_i$	$\tilde{y}_1$	$\dots$	$\tilde{y}_n$
ε_i	ε_1	$\dots$	ε_n

Точка ε_i считается *поворотной*, если выполняются следующие условия

$\varepsilon_{i-1} < \varepsilon_i > \varepsilon_{i+1}$ или $\varepsilon_{i-1} > \varepsilon_i < \varepsilon_{i+1}$.

(1.1)

Далее подсчитывается число поворотных точек *p*. Критерием случайности с 5%-ным уровнем значимости, т.е. с доверительной вероятностью 95%, является выполнение равенства

$$p > \left[\frac{2}{3}(n-2) - 1{,}96 \cdot \sqrt{\frac{16n-29}{90}} \right], \qquad (1.2)$$

где [...]– целая часть числа. Если неравенство выполняется, то модель считается адекватной.

Пусть расчет регрессии дал следующие результаты

$$\tilde{y} = -0{,}52544 + 3{,}230239x,$$

x	y	$\tilde{y}$	Остатки $\varepsilon = y - \tilde{y}$
3	12	9,165277	2,834723
4	13	12,39552	0,604484
5	20	15,62576	4,374245
7	19	22,08623	-3,086233
8	31	25,31647	5,683528
10	24	31,77695	-7,77695
11	41	35,00719	5,992811

12	28	38,23743	-10,237428
15	52	47,92815	4,071855
20	55	64,07934	-9,07934
30	103	96,38173	6,61827
Среднее			-3,18182E-06

Цветом выделены поворотные точки. Их всего 9, в этом легко убедиться, если просмотреть пики графика (значения фактора *x* должны быть отсортированы по возрастанию)

$$9 > \left[\frac{2}{3}(11-2) - 1{,}96 \cdot \sqrt{\frac{16 \cdot 11 - 29}{90}} \right] = [3{,}45].$$

$$9 > 3$$

Неравенство верное, остатки признаем случайными.

2. Для проверки равенства математического ожидания остаточной последовательности нулю вычисляется среднее значение ряда остатков

$$\bar{\varepsilon} = \frac{1}{n} \cdot \sum \varepsilon_i \ . \qquad (1.3)$$

Если $\bar{\varepsilon} \approx 0$, то считается, что модель не содержит постоянной систематической ошибки и адекватна по критерию нулевого среднего. Если $\bar{\varepsilon} \neq 0$, то проверяется гипотеза о равенстве нулю математического ожидания. Для этого вычисляют t-критерий Стъюдента по формуле

$$t = \frac{|\bar{\varepsilon}| - 0}{S_\varepsilon} \cdot \sqrt{n} \ , \qquad (1.4)$$

где S_ε – среднее квадратическое отклонение ряда остатков, $S_\varepsilon = \sqrt{\frac{\sum \varepsilon^2}{n - m - 1}}$, m – число параметров при переменной x.

Значение t-критерия сравнивают с табличным при заданном уровне значимости. Если выполняется неравенство $t > t_{табл}$, то модель неадекватна по данному критерию.

По расчетам $\bar{\varepsilon} = -3{,}18 \cdot 10^{-6} \approx 0$, то есть по данному пункту модель признаем адекватной.

3. Проверку независимости последовательности остатков (отсутствие автокорреляции) осуществляют с помощью d-критерия Дарбина-Уотсона. Расчетное значение критерия определяется по формуле

$$d = \frac{\sum (\varepsilon_i - \varepsilon_{i-1})^2}{\sum \varepsilon_i^2} \qquad (1.5)$$

и сравнивается с нижним $(d_1$ или $d_L)$ и верхним $(d_2$ или $d_U)$ критическими значениями статистики Дарбина-Уотсона.

Возможны следующие случаи:

1) Если $d < d_1$, то гипотеза о независимости остатков отвергается, и модель признается неадекватной по критерию независимости остатков.

2) Если $d_1 < d < d_2$, включая сами эти значения, то считается, что нет достаточных оснований делать тот или иной вывод.

3) Если $d_2 < d < 2$, то гипотеза о независимости

остатков принимается и модель признается адекватной по данному критерию.

4) Если $d > 2$, то это свидетельствует об отрицательной автокорреляции остатков. В этом случае расчетное значение критерия необходимо преобразовать по формуле $d' = 4 - d$ и сравнивать $\boxed{(\varepsilon_i - \varepsilon_{i-1})^2}$ критическим значением не d, а d'.

Составляем вспомогательную таблицу:

x	y	ỹ	остатки ε	$(\varepsilon_i - \varepsilon_{i-1})^2$	ε_i^2
3	12	9,165277	2,834723		8,035654
4	13	12,39552	0,604484	4,973965997	0,365400906
5	20	15,62576	4,374245	14,211098	19,13401932
7	19	22,08623	-3,086233	55,65873199	9,52483413
8	31	25,31647	5,683528	76,908708	32,30249053
10	24	31,77695	-7,77695	181,184468	60,4809513
11	41	35,00719	5,992811	189,606318	35,91378368
12	28	38,23743	-10,237428	263,420658	104,8049321
15	52	47,92815	4,071855	204,75558	16,58000314
20	55	64,07934	-9,07934	172,9539299	82,43441484

30	103	96,38173	6,61827	246,4149597	43,80149779
Сумма				1410,088418	413,3779817

Определяем значение

$$d = \frac{\sum(\varepsilon_i - \varepsilon_{i-1})^2}{\sum \varepsilon_i^2} = \frac{1410{,}09}{413{,}38} \approx 3{,}41.$$

Критические значения критерия Дарбина-Уотсона находят по специальным таблицам для заданных объема наблюдений n и числа независимых переменных модели k.

В нашем случае $d_1 = 0{,}93$; $d_2 = 1{,}32$. Имеем отрицательную автокорреляцию остатков. Переходим к $d' = 4 - d$, $d' = 4 - 3{,}41 = 0{,}59$.

Так как $d' < d_1$, модель признается неадекватной, остатки регрессии взаимозависимы. Уравнение регрессии $\tilde{y} = -0{,}52544 + 3{,}230239x$ не может быть использовано для прогнозирования. Автокорреляция в остатках может иметь разные причины. Возможно, форма связи неточна, или в уравнение не включен какой-либо существенный фактор.

Значения статистики Дарбина-Уотсона $d_1 d_2$

на 5%-ном уровне значимости

n	$k=1$		$k=2$		$k=3$		$k=4$		$k=5$	
	d_1	d_2	d_1	d_2	d_1	d_2	d_1	d_2	d_1	d_2
6	0,61	1,40								
7	0,70	1,36	0,47	1,90						
8	0,76	1,33	0,56	1,78	0,37	2,29				
9	0,82	1,32	0,63	1,70	0,46	2,13				
10	0,88	1,32	0,70	1,64	0,53	2,02				
11	0,93	1,32	0,66	1,60	0,60	1,93				
12	0,97	1,33	0,81	1,58	0,66	1,86				
13	1,01	1,34	0,86	1,56	0,72	1,82				
14	1,05	1,35	0,91	1,55	0,77	1/78				
15	1,08	1,36	0,95	1,54	0,82	0,75	0,69	1,97	0,56	2,21
16	1,10	1,37	0,98	1,54	0,86	1,73	0,74	1,93	0,62	2,15
17	1,13	1,38	1,02	1,54	0,90	1,71	0,78	1,90	0,67	2,10
18	1,16	1,39	1,05	1,53	0,93	1,69	0,82	1,87	0,71	2,06
19	1,18	1,40	1,08	1,53	0,97	1,68	0,86	1,85	0,75	2,02
20	1,20	1,41	1,10	1,54	1,00	1,68	0,90	1,83	0,79	1,99
21	1,22	1,42	1,13	1,54	1,03	1,67	0,93	1,81	0,83	1,96
22	1,24	1,43	1,15	1,54	1,05	1,66	0,96	1,80	0,86	1,94
23	1,26	1,44	1,17	1,54	1,08	1,66	0,99	1,79	0,90	1,92
24	1,27	1,45	1,19	1,55	1,10	1,66	1,01	1,78	0,93	1,90
25	1,29	1,45	1,21	1,55	1,12	1,66	1,04	1,77	0,95	1,89
26	1,30	1,46	1,22	1,55	1,14	1,65	1,06	1,76	0,98	1,88
27	1,32	1,47	1,24	1,56	1,16	1,65	1,08	1,76	1,01	1,86
28	1,33	1,48	1,26	1,56	1,18	1,65	1,10	1,75	1,03	1,85
29	1,34	1,48	1,27	1,56	1,20	1,65	1,12	1,74	1,05	1,84
30	1,35	1,49	1,28	1,57	1,21	1,65	1,14	1,74	1,07	1,83

Задание

1. *Построить по данным из своего задания график.*
2. *Подобрать модели, аппроксимирующие данные.*
3. *Из всех моделей выбрать две и проверить их на адекватность.*
4. *Проверить точность адекватных моделей.*
5. *Выбрать наилучшую модель (выбор обосновать).*

Варианты для задания

Номер варианта	x_i	2	4	5	6	8	9	11	12	13	15	16	18	Значение x для прогноза
1, 16	y_i	3	6	6	6	10	9	12	14	15	16	17	19	21
2, 17	y_i	5	9	10	13	17	19	23	26	28	28	29	30	20
3, 18	y_i	7	13	16	19	25	28	34	38	41	45	47	50	22
4, 19	y_i	9	15	20	22	33	37	45	50	54	56	60	63	25
5, 20	y_i	11	18	20	31	40	44	56	62	67	70	76	78	24
6, 21	y_i	11	26	31	40	49	58	68	74	80	85	89	93	23
7, 22	y_i	15	29	36	43	57	64	78	86	92	98	103	106	20
8, 23	y_i	17	33	40	48	65	73	88	104	106	114	116	122	21
9, 24	y_i	6	10	12	17	18	20	22	28	29	30	33	35	22
10, 25	y_i	8	14	17	20	26	29	36	39	42	46	49	55	23
11, 26	y_i	9	18	20	22	26	38	47	50	52	54	64	75	24
12, 27	y_i	10	22	32	30	35	48	58	60	70	72	79	80	25
13, 28	y_i	14	20	30	40	42	50	65	70	78	88	100	110	21
14, 29	y_i	15	30	37	44	58	66	80	87	94	98	105	110	22

Часть 2. Аппроксимация эмпирических данных с использованием Scilab

Для определения значений неизвестных параметров a_1 , a_2 , .. a_m, наилучшим образом описывающих экспериментальные данные задачи в Scilab предусмотрена функция:

[a,S]=datafit(F,z,c)

где *F* -_функция, параметры которой необходимо подобрать; *z*-матрица исходных данных; *c*-вектор начальных приближений; *a*-вектор коэффициентов; *S*-сумма квадратов отклонений измеренных значений от расчетных.

Рассмотрим использование функции *datafit* на примере.

Задача 1.

В результате опыта холостого хода определена зависимость потребляемой из сети мощности (P, Вт) от входного напряжения (U, В) для асинхронного двигателя (табл. 1).

Таблица 1. Зависимость потребляемой из сети мощности от входного напряжения.

U, В	132	140	150	162	170	180	190	200	211	220	232	240	251
P, Вт	330	350	385	425	450	485	540	600	660	730	920	1020	1350

Методом наименьших квадратов подобрать модель вида:

$$P = a_1 + a_2U + a_3U^2 + a_4U^3.$$

Решение задачи 1

```
//Функция, вычисляющая разность между экспериментальными
//и теоретическими значениями.
//Перед использованием необходимо определить
//z=[x;y] - матрицу исходных данных - и
//c - вектор начальных значений коэффициентов,
//размерность вектора должна совпадать
//с количеством искомых коэффициентов.
function [zr]=G(c,z)
zr=z(2)-c(1)-c(2)*z(1)-c(3)*z(1)^2-c(4)*z(1)^3
endfunction
//Исходные данные
x=[1.32 1.40 1.50 1.62 1.70 1.80 1.90...
2.00,2.11,2.20,2.32,2.40,2.51];
y=[3.30 3.50 3.85 4.25 4.50 4.85 5.40...
6.00 6.60 7.30 9.20 10.20 13.50];
//Формирование матрицы исходных данных
z=[x;y];
//Вектор начальных приближений
c=[0;0;0;0];
//Решение задачи
[a,err]=datafit(G,z,c)
S =
0.5287901
```

a =
- 51.576664
95.594671
- 55.695312
11.111453
Итак, в результате работы функции *datafit* была подобрана аналитическая зависимость вида
$P = -51.577+95.595U-55.695U^2+11.111U^3$,
а сумма квадратов отклонений измеренных значений от расчетных составила 0.529.
Геометрическая интерпретация задачи:
Построение графика подобранной зависимости

```
//Построение графика экспериментальных данных
plot2d(x,y,-4);
//Построение графика подобранной функции
t=1.32:0.01:2.51;
Ptc=a(1)+a(2)*t+a(3)*t^2+a(4)*t^3;
plot2d(t,Ptc);
```

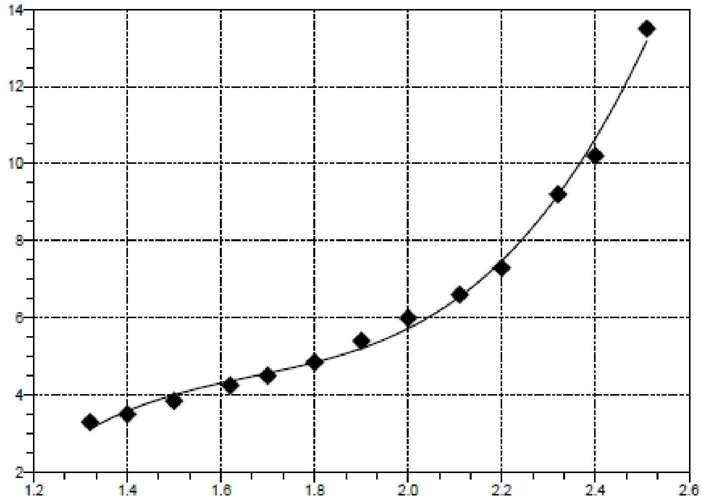

Рис. 1. Графическая интерпретация задачи 1

Задача 2.
К некоторым экспериментальным данным подобрать зависимость вида
$$Y = a_1 \cdot x^{a2} + a_3.$$
Далее представлено решение задачи и ее графическая интерпретация
Решение задачи 2

```
function [zr]=F(c,z)
zr=z(2)-c(1)*z(1)^c(2)-c(3);
endfunction
x=[10.1,10.2,10.3,10.8,10.9,11,11.1,11.4,12.2,13.3,13.8,...
14,14.4,14.5,15,15.6,15.8,17,18.1,19];
y=[24,36,26,45,34,37,55,51,75,84,74,91,85,87,...
```

```
94,92,96,97,98,99];
z=[x;y];
c=[0;0;0];
[a,S]=datafit(F,z,c);
t=10:0.01:19;
Yt=a(1)*t^a(2)+*a*(3);
plot2d(x,y,-3);
plot2d(t,Yt);
```

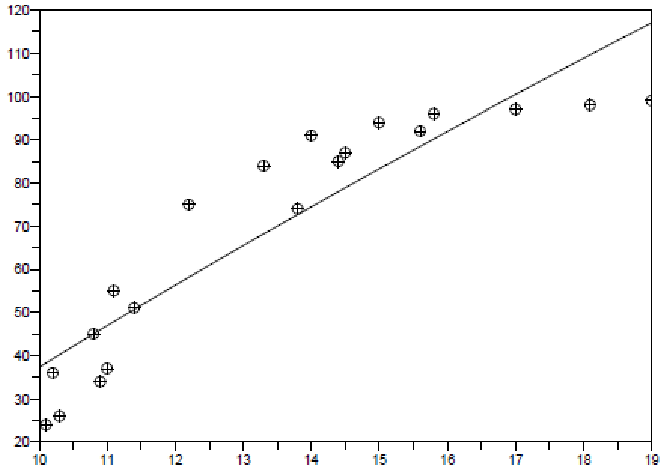

Рис. 2. Графическое решение задачи 2

Одной из наиболее часто используемых в методе наименьших квадратов функций является прямая, описываемая уравнением вида $y = a_1 + a_2x$, которая называется линией регрессии *y* на *x*. Параметры a_1 и a_2 являются коэффициентами регрессии. Показатель, характеризующий тесноту линейной связи между *x* и *y*, называемый коэффициентом корреляции, рассчитывается по формуле:

$$r = \frac{\sum_{i=1}^{n}(x_i - M_x)\cdot(y_i - M_y)}{\sqrt{\sum_{i=1}^{n}(x_i - M_x)^2\sum_{i=1}^{n}(y_i - M_y)^2}}, \qquad Mx = \frac{\sum_{i=1}^{n} x_i}{n}, \qquad My = \frac{\sum_{i=1}^{n} y_i}{n}$$

Значение коэффициента корреляции удовлетворяет соотношению $-1 \le r \le 1$. Чем меньше отличается абсолютная величина *r* от единицы, тем ближе к линии регрессии располагаются экспериментальные точки. Если коэффициент корреляции близок к нулю, то это означает, что между *x* и *y* отсутствует линейная связь, но может существовать другая, нелинейная, зависимость.

Аналогом коэффициента корреляции *r* для нелинейных зависимостей является индекс корреляции, рассчитываемый по формуле:

$$R = \sqrt{1 - \frac{\sum_{i=1}^{n}(y_i - Y_i)^2}{\sum_{i=1}^{n}(y_i - M_y)^2}}$$

где *y* -экспериментальные значения, *Y* -значения, найденные методом наименьших квадратов, *My* - среднее значение *y*. Индекс корреляции по своему абсолютному значению колеблется в пределах от *0* до *1*. При функциональной зависимости индекс корреляции равен *1*. В случае отсутствии связи *R = 0*. Если коэффициент корреляции *r* является мерой тесноты связи только для линейной формы, то индекс корреляции *R* -и для линейной, и для криволинейной. При прямолинейной связи коэффициент корреляции по своей абсолютной величине равен индексу корреляции.
Для расчета коэффициентов регрессии в Scilab предназначена функция
a=regress(x,y)
где *x* и *y*_экспериментальные данные, *a*-вектор коэффициентов линии регрессии a_1 и a_2.
Рассмотрим работу этой функции на примере.
Задача 3.
Число условных частей NaNO3, растворяющихся в 100 частях воды при соответствующих температурах, представлено в табл. 3. Требуется определить растворимость азотнокислого натрия при температуре 32 градуса в случае линейной зависимости и найти коэффициент и индекс корреляции.
Таблица 3. Данные о растворимости NaNO3 в зависимости от температуры воды.

0	4	10	15	21	29	36	51	68
66.7	71.0	76.3	80.6	85.7	92.9	99.4	113.6	125.1

Решение задачи 3

```
//Экспериментальные данные
x=[0 4 10 15 21 29 36 51 68];
y=[66.7 71 76.3 80.6 85.7 92.9 99.4 113.6 125.1];
//Расчет коэффициентов регрессии
a=regress(x,y)
a =
67.507794
0.8706404
//Растворимость азотного натрия при температуре 32 градуса (прогноз)
-->t=32;a(1)+a(2)*t
ans = 95.368287
//Коэффициент корреляции
r=sum((x-mean(x)).*(y-mean(y)))/...
sqrt(sum((x-mean(x))^2)*sum((y-mean(y))^2))
r = 0.9989549
//Индекс корреляции
R=sqrt(1-sum((y-(a(1)+a(2)*x))^2)/sum((y-mean(y))^2))
R = 0.9989549
```

Построение графика экспериментальных данных и линии регрессии:
График подобранной линии регрессии

```
t=0:70; Yt=a(1)+a(2)*t;
plot2d(x,y,-5); plot2d(t,Yt);
```

Для расчета коэффициента корреляции в Scilab также предназначена встроенная функция *a=corr(x,y)*, где *x* и *y* — экспериментальные данные.

Задание.

1. Выполнить подбор параметров модели для своего варианта (задание в Часть 1)