Практический семинар по Математической экономике (17.М18-э + 17.М19-э)

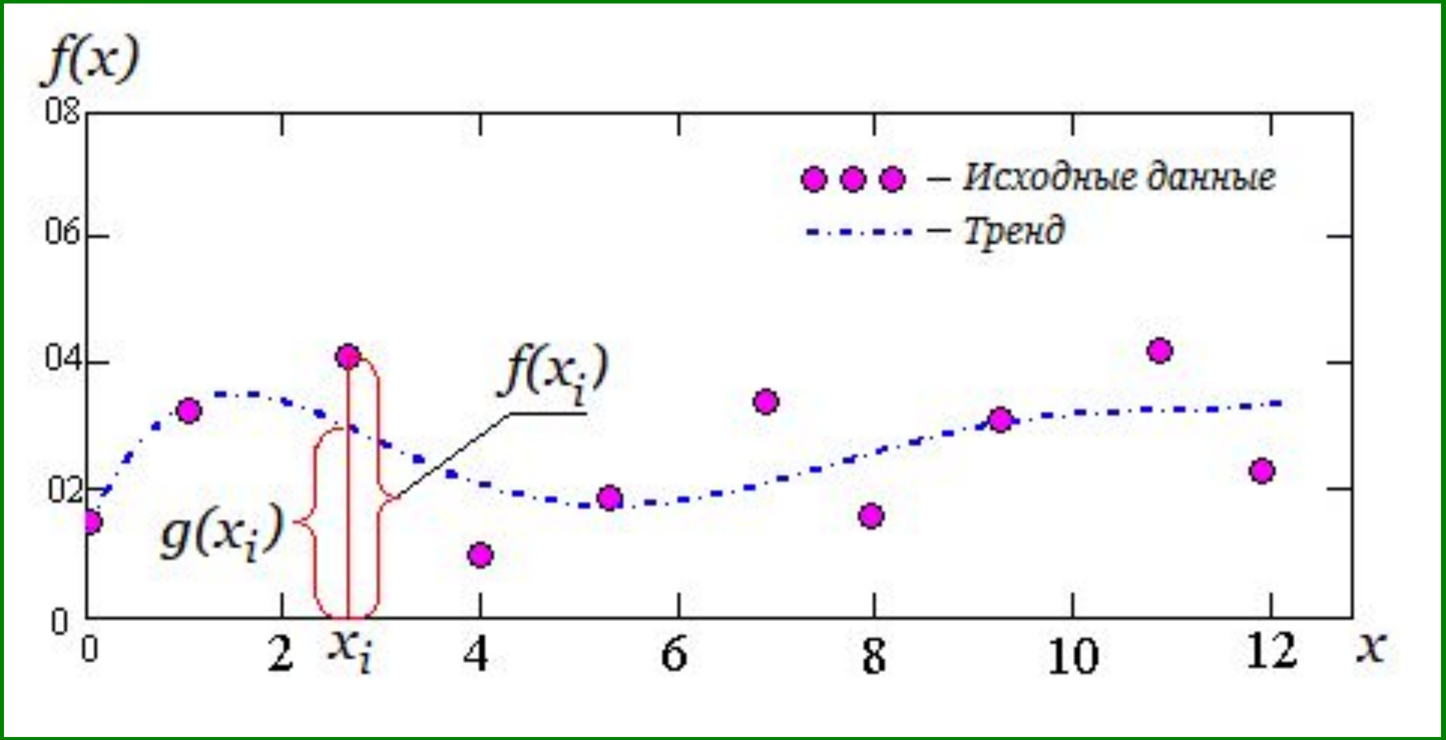
Занятие 3

ОБРАБОТКА ДАННЫХ.

Задача о наилучшем среднеквадратическом приближении (задача о тренде)

2018/2019 уч. год

Задача о тренде. Метод наименьших квадратов



Критерий: $$\sum_{i=1}^{n}(f(x_i) - g(x_i))^2 \to min$$

Формальная постановка задачи. Условие минимума среднеквадратического отклонения

Пусть дана таблица значений функции

$$\begin{array}{c|c} x & x_1, x_2, x_3, \ldots, x_n \\ \hline y & y_1, y_2, y_3, \ldots, y_n \end{array}, x_i \neq x_j, i, j = 1,2,..,n$$

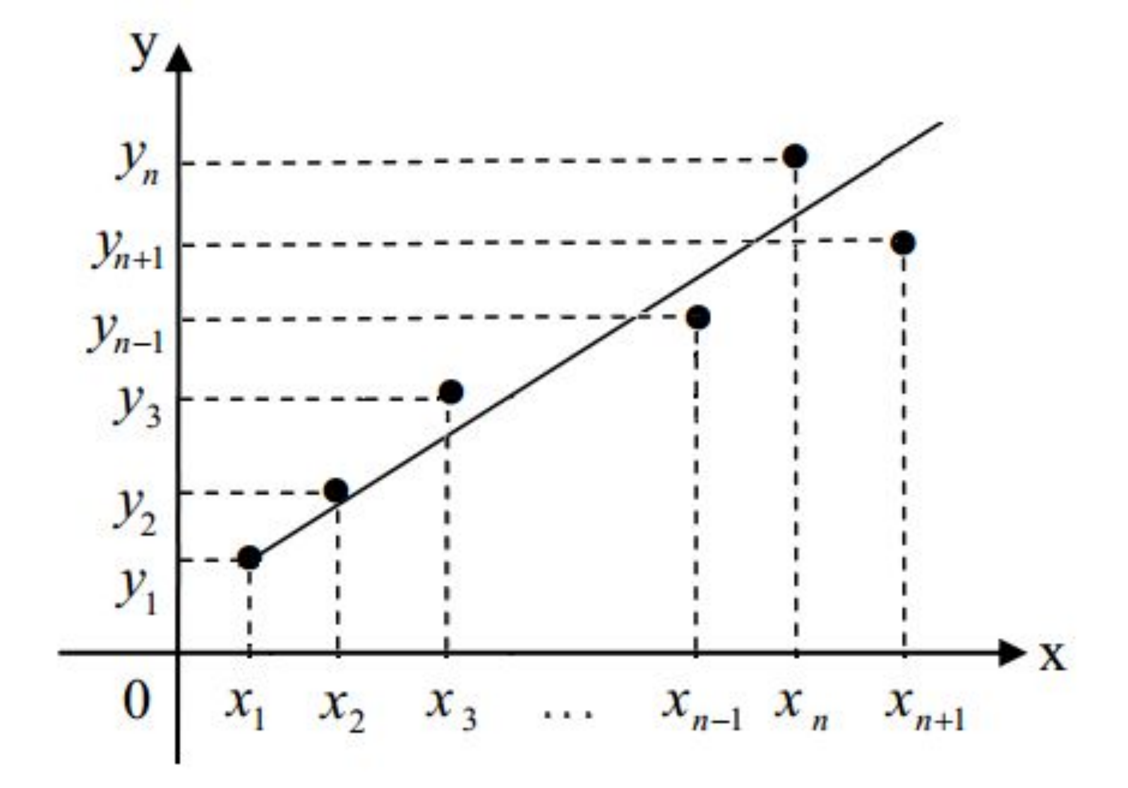
Найти полином заданной степени ***m*** ,

$$P_m(x) = a_0 + a_1 x + a_2 x^2 + \ldots + a_m x^m ,$$

удовлетворяющий условию:

$$\boxed{\sum_{i=1}^{n} (y_i - P_m(x_i))^2 \rightarrow min}$$

Метод наименьших квадратов (МНК) для построения линейного тренда



$$P_1(x) = ax + b, \quad S(a,b) = \sum_{j=1}^{n} \left(y_j - (ax_j + b)\right)^2$$

Нахождение коэффициентов

$$\begin{cases} \dfrac{\partial S}{\partial a} = 0 \\ \dfrac{\partial S}{\partial b} = 0 \end{cases} \Leftrightarrow \begin{cases} -2\sum\limits_{j=1}^{n}\Big(y_j - (ax_j + b)\Big)x_j = 0 \\ -2\sum\limits_{j=1}^{n}\Big(y_j - (ax_j + b)\Big) = 0 \end{cases}$$

$$\begin{cases} a\sum\limits_{j=1}^{n} x_j^2 + b\sum\limits_{j=1}^{n} x_j = \sum\limits_{j=1}^{n} x_j y_j \\ a\sum\limits_{j=1}^{n} x_j + \sum\limits_{j=1}^{n} b = \sum\limits_{j=1}^{n} y_j \end{cases} \Leftrightarrow \begin{cases} a\sum\limits_{j=1}^{n} x_j^2 + b\sum\limits_{j=1}^{n} x_j = \sum\limits_{j=1}^{n} x_j y_j \\ a\sum\limits_{j=1}^{n} x_j + nb = \sum\limits_{j=1}^{n} y_j \end{cases}$$

Нахождение параметров линейной модели

$$na_0 + \left(\sum_{i=1}^{n} x_i\right) a_1 = \sum_{i=1}^{n} y_i$$

$$\left(\sum_{i=1}^{n} x_i\right) a_0 + \left(\sum_{i=1}^{n} x_i^2\right) a_1 = \sum_{i=1}^{n} x_i y_i$$

$$\delta_1 = \sum_{i=1}^{n} \left(y_i - (a_0 + a_1 x_i)\right)^2$$

$$y = ax + b$$

$$S(a,b) = \sum_{i=1}^{n} {e_i}^2 = \sum_{i=1}^{n} \left(y_i - (ax_i + b)\right)^2$$

$$\begin{cases} \dfrac{\partial S}{\partial a} = 0 \\ \dfrac{\partial S}{\partial b} = 0 \end{cases} \qquad \begin{cases} a\sum x_i^2 + b\sum x_i = \sum x_i y_i \\ a\sum x_i + bn = \sum y_i \end{cases}$$

Пример 1 (линейный тренд)

Дана таблица значений функции $y=f(x)$:

x	0	1	2	4	5
y	2.1	2.4	2.6	2.8	3

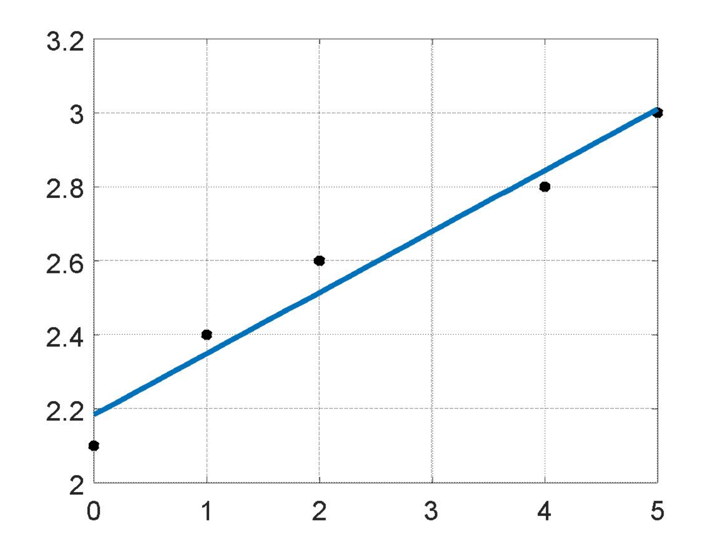
Используя МНК, аппроксимировать эти данные линейной зависимостью $g(x) = ax + b$

x	y	xy	x^2	g
0	2.1	0	0	2.18
1	2.4	2.4	1	2.35
2	2.6	5.2	4	2.51
4	2.8	11.2	16	2.84
5	3.0	15.0	25	3.01
12	**12.9**	**33.8**	**46**	**12.9**

$$\begin{cases} 46a + 12b = 33.8 \\ 12a + 5b = 12.9 \end{cases} \qquad \begin{cases} a \approx 0.165 \\ b \approx 2.184 \end{cases}$$

$$y = 0.165x + 2.184$$

$$\delta = 0.0618$$



Пример 2

Дана таблица значений функции $y=f(x)$:

x	1	2	3	4	5
y	1	2	2	1	1

Используя МНК, аппроксимировать эти данные линейной зависимостью $g(x) = a_0 + a_1 x$

Нахождение параметров квадратичной модели

$$na_0 + \left(\sum_{i=1}^{n} x_i\right) a_1 = \sum_{i=1}^{n} y_i$$

$$\left(\sum_{i=1}^{n} x_i\right) a_0 + \left(\sum_{i=1}^{n} x_i^2\right) a_1 + \left(\sum_{i=1}^{n} x_i^2\right) a_2 = \sum_{i=1}^{n} x_i y_i$$

$$\left(\sum_{i=1}^{n} x_i\right) a_0 + \left(\sum_{i=1}^{n} x_i^2\right) a_1 + \left(\sum_{i=1}^{n} x_i^2\right) a_1 = \sum_{i=1}^{n} x_i y_i$$

$$\delta_2 = \sum_{i=1}^{n} \left(y_i - (a_0 + a_1 x_i + a_2 {x_i}^2)\right)^2$$

$$y = ax^2 + bx + c$$

$$S(a,b,c) = \sum_{i=1}^{n} \left(y_i - (ax_i{}^2 + bx_i + c\,)\right)^2$$

$$\begin{cases} \dfrac{\partial S}{\partial a} = 0 \\ \dfrac{\partial S}{\partial b} = 0 \\ \dfrac{\partial S}{\partial c} = 0 \end{cases} \qquad \begin{cases} a\sum x_i^4 + b\sum x_i^3 + c\sum x_i^2 = \sum x_i^2 y_i \\ a\sum x_i^3 + b\sum x_i^2 + c\sum x_i = \sum x_i y_i \\ a\sum x_i^2 + b\sum x_i + cn = \sum y_i \end{cases}$$

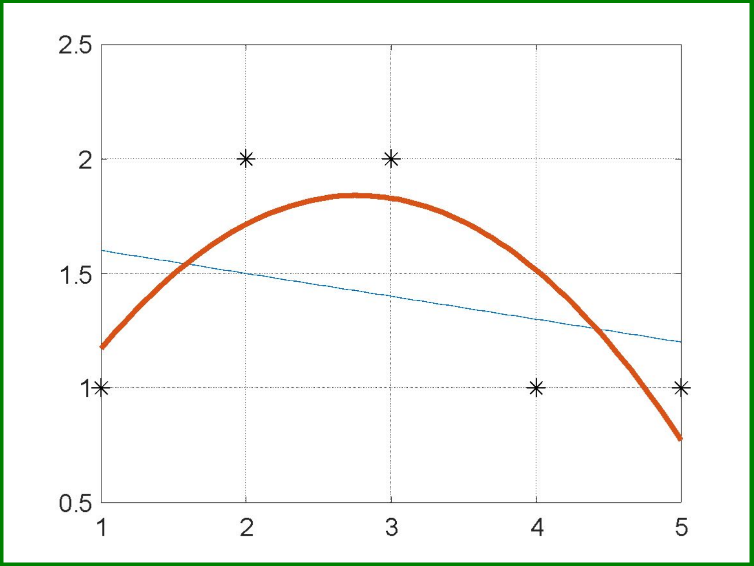
Пример 2 (квадратичный тренд)

X	Y	X^2	Y^2	XY	X^3	X^4	X^2Y
1	1	1	1	1	1	1	1
2	2	4	4	4	8	16	8
3	2	9	4	6	27	81	18
4	1	16	1	4	64	256	16
5	1	25	1	5	125	625	25
15	**7**	**55**	**11**	**20**	**225**	**979**	**68**

$$\begin{cases} 979a + 225b + 55c = 68 \\ 225a + 55b + 15c = 20 \\ 55a + 15b + 5c = 7 \end{cases}$$

a = -0.2143 b = 1.1857 c = 0.2

$$\delta_2 \approx 0.3024$$



Нелинейный тренд. Пример 1

$$y = \frac{a}{x} + b$$

$$S(a,b) = \sum_{i=1}^{n} {e_i}^2 = \sum_{i=1}^{n} \left(y_i - (a/x_i + b)\right)^2$$

$$\begin{cases} \dfrac{\partial S}{\partial a} = 0 \\ \dfrac{\partial S}{\partial b} = 0 \end{cases} \qquad \begin{cases} a\sum 1/x_i^2 + b\sum 1/x_i = \sum y_i/x_i \\ a\sum 1/x_i + bn = \sum y_i \end{cases}$$

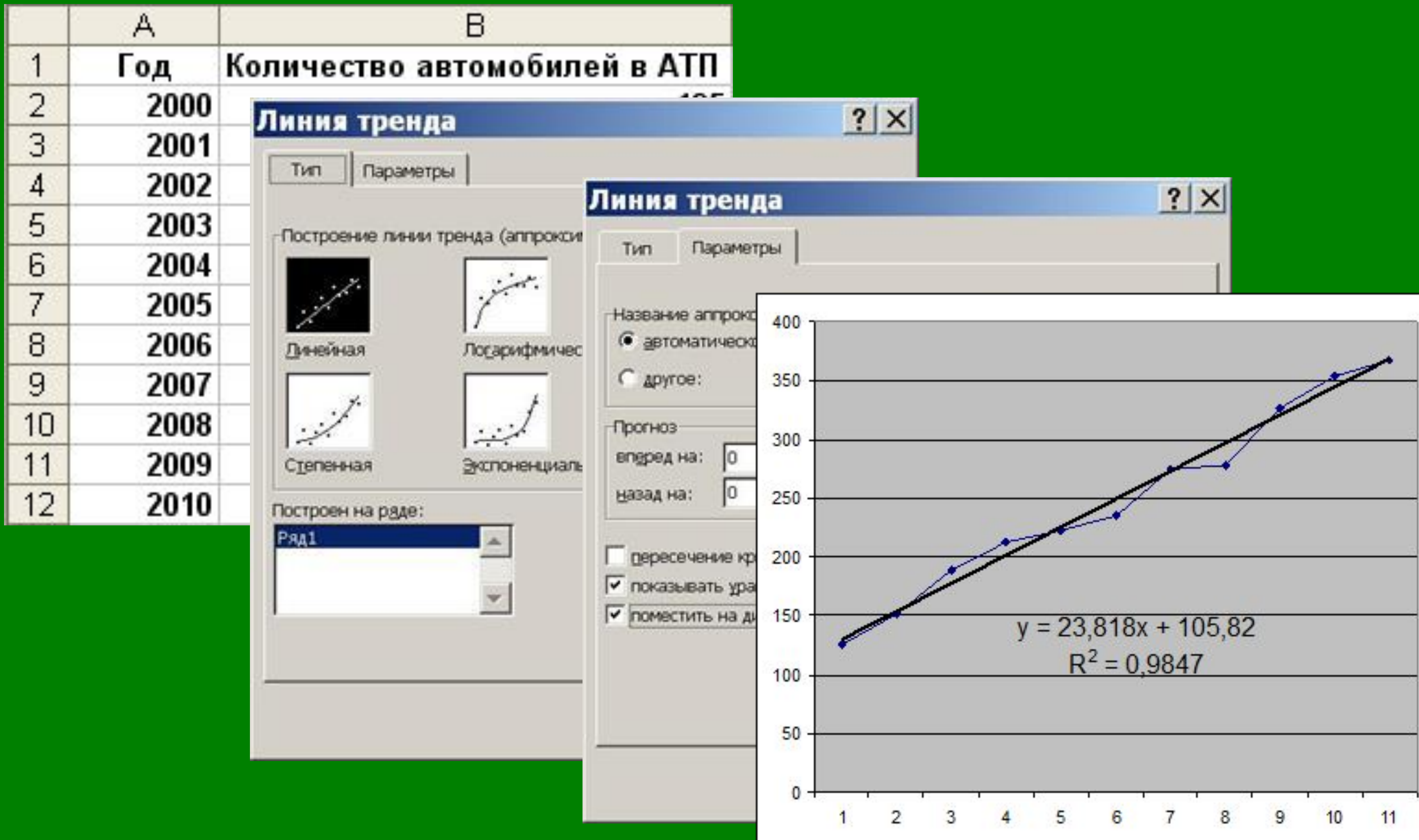
Нелинейный тренд. Пример 2

$$y = a\ln x + b$$

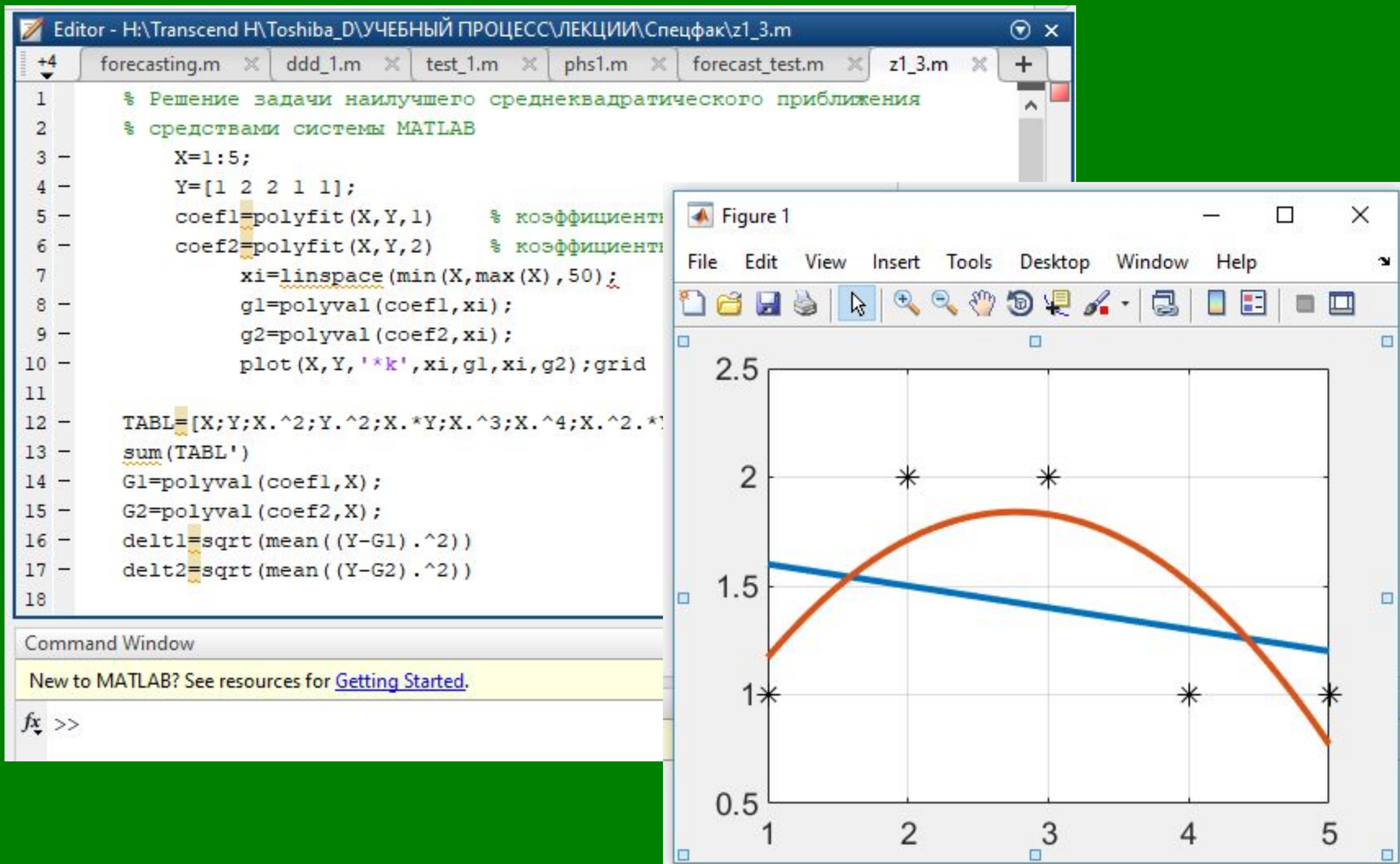
$$S(a,b) = \sum_{i=1}^{n} {e_i}^2 = \sum_{i=1}^{n} \left(y_i - (a\ln x_i + b)\right)^2$$

$$\begin{cases} \dfrac{\partial S}{\partial a} = 0 \\ \dfrac{\partial S}{\partial b} = 0 \end{cases} \qquad \begin{cases} a\sum(\ln^2 x_i + b\sum \ln x_i = \sum y_i \ln x_i \\ a\sum \ln x_i + bn = \sum y_i \end{cases}$$

Решение задачи средствами EXCEL



Решение задачи средствами MATLAB



Популярные метрики

e_k – «ошибка», P_k – процентная «ошибка»

Mean Absolute Percentage Error (*MAPE*)

Root Mean Square Error (*RMSE*)

$$MAPE = \frac{1}{n}\sum_{k=1}^{n}|P_k| ,$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{k=1}^{n} e_k^2} .$$