

## СОДЕРЖАНИЕ к ЛР-2

Лабораторная работа 2. Модели статистического моделирования и прогнозирования динамических систем по временному ряду (на основе классического метода МНК) .....	2
Справочные сведения по модели статистического моделирования и прогнозирования .....	2
Рекомендуемая литература для лабораторной работы .....	4
ПРИЛОЖЕНИЕ 2.1. Варианты для моделирования временного ряда к ЛР2 .....	5
ПРИЛОЖЕНИЕ 2.2. Пример оформления ЛР-2 .....	8

## Лабораторная работа 2. Модели статистического моделирования и прогнозирования динамических систем по временному ряду (на основе классического метода МНК)

### Цель работы

Цель настоящей работы – освоить средства моделирования стохастических временных рядов.

### Ход работы

1. Ознакомиться со справочными сведениями.
2. Сформулировать задачу МНК при построении функции регрессии.
3. Разработать программу, моделирующую алгоритм поиска оптимального решения для формализованной задачи, используя математический пакет MatLab или язык программирования Python:
  - a. Самостоятельно реализовать МНК для решения задачи поиска коэффициентов модели, заданной в виде полинома второго порядка  $f_1(x) = a_2x^2 + a_1x + a_0$ .
  - b. С использованием встроенной реализации МНК в MatLab или Python подобрать степень  $p$  полиномиальной модели  $f_2(x) = \sum_{i=0}^p a_i x^i$ , наилучшим образом соответствующей исходным данным при визуальной оценке на графике. Для этого построить график с исходными данными (крестики, точки и т.п.) и различными вариантами полиномиальных моделей степени  $p$ , где  $p \neq 2$ .
  - c. Построить дополнительно функциональную модель вида  $f_3(x) = \sqrt[3]{x+1}$ .
  - d. Используя скорректированный коэффициент детерминации  $R_{adj}^2$  определить наилучшую из трех моделей  $f_1(x)$ ,  $f_2(x)$ ,  $f_3(x)$ .
4. Составить и представить преподавателю отчет о работе.

*Исходные данные:* Варианты задач в Приложении 3.1 по номеру студента в списке.

### Справочные сведения по модели статистического моделирования и прогнозирования

В рамках лабораторной работы рассматривается метод численного моделирования функции по экспериментальным данным с целью аппроксимации фактических данных (с целью прогнозирования, в том числе).

Для сравнения моделей между собой обычно используют оценку погрешностей аппроксимации или коэффициент детерминации. В данной лабораторной работе предлагается использовать последний.

Коэффициент детерминации модели описывает долю дисперсии

зависимой переменной  $y$ , объясняемую моделью. В общем случае коэффициент детерминации можно вычислить по формуле:

$$R^2 = 1 - \frac{D[y|x]}{D[y]} = 1 - \frac{\sigma^2}{\sigma_y^2},$$

где  $D[y|x] = \sigma^2 = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n-k-1}$  – условная дисперсия ошибки модели, исправленная нормирующим коэффициентом,  $D[y] = \sigma_y^2$  – дисперсия случайной величины  $y$ . Здесь  $\hat{y}_i = f_j(x_i)$  – результат  $j$ -ой модели в точке  $x_i$ ,  $n$  – количество наблюдений за переменными  $x$  и  $y$ ,  $k$  – количество параметров  $j$ -ой модели. Чем ближе значение коэффициента детерминации к единице, тем лучше данная модель описывает исходные данные.

Коэффициент детерминации обладает существенным недостатком: при увеличении количества параметров  $k$ , входящих в модель, его величина растет. Поэтому на практике обычно используют скорректированный коэффициент детерминации, лишенный данного недостатка:

$$R_{adj}^2 = 1 - (1 - R^2) \frac{n-1}{n-k}.$$

Реализация МНК для полиномиальных моделей в математических пакетах осуществляется с помощью функций `polyfit` в MatLab и `numpy.polyfit` в Python.

## **Рекомендуемая литература для лабораторной работы**

1. Бокс Дж., Дженкинс Г. Анализ временных рядов. Прогноз и управление. М.: Мир, 1974. Выпуск 1, 2.
2. Воронцов К.В., Егорова Е.В. Динамически адаптируемые композиции алгоритмов прогнозирования // Искусственный Интеллект. – № 10. - 2006. – С. 277–280.
3. Гребенников А.В, Крюков Ю.А, Чернягин Д. В. Моделирование сетевого трафика и прогнозирование с помощью модели ARIMA.
4. Безручко Б.П., Смирнов Д.А. Статистическое моделирование по временным рядам. Учебнометодическое пособие. Саратов: Издательство ГосУНЦ “Колледж”. 2000. 23 с.
5. Афанасьев В.Н., Цыпин А.П. Эконометрика в пакете STATISTICA: учебное пособие по выполнению лабораторных работ. Оренбург: ГОУ ОГУ, 2008. 204 с.
6. Дуброва Т.А. Статистические методы прогнозирования: учеб. пособие для вузов. – М.: ЮНИТИДАНА, 2003. – 206 с.

## ПРИЛОЖЕНИЕ 2.1. Варианты для моделирования временного ряда к ЛР2

### Вариант 1

Изучается динамика потребления молока в регионе. Для этого собраны данные об объемах среднедушевого потребления мяса (кг)  $Y(t)$  за 7 месяцев. Обосновать и построить тренд данного ряда. Оценить достоверность уточненной по МНК модели.

$t$	1	2	3	4	5	6	7
$Y(t)$	8,16	8,25	8,41	8,76	9,2	9,78	10,1

### Вариант 2

Банк изучает динамику изменения величины депозитов физических лиц за несколько лет (млн.\$ в сопоставимых ценах). Обосновать и построить тренд данного ряда. Оценить достоверность уточненной по МНК модели.

Время, $t$	1	2	3	4	5	6	7
Размер депозитов, $Y(t)$	2	6	7	3	10	12	13

### Вариант 3

Изучается динамика рождаемости в России. Собраны данные о числе рожденных (млн)  $Y(t)$  за 7 лет (2009-2015). Обосновать и построить тренд данного ряда. Оценить достоверность уточненной по МНК модели.

$t$	2009	2010	2011	2012	2013	2014	2015
$Y(t)$	1,767	1,788	1,796	1,902	1,895	1,947	1,944

### Вариант 4

Изучается динамика потребления сахара в России. Для этого собраны данные об объемах среднедушевого потребления сахара (г/сутки)  $Y(t)$  за 7 десятилетий. Обосновать и построить тренд данного ряда. Оценить

достоверность уточненной по МНК модели.

$t$	1(1950)	2(1960)	3(1970)	4(1980)	5(1990)	6(2000)	7(2015)
$Y(t)$	32	85	115	130	130	96	107

### Вариант 5

Изучается динамика потребления мяса птицы в Европе. Для этого собраны данные об объемах среднедушевого потребления мяса (кг/чел/год)  $Y(t)$  за 10 лет (2000-2009). Обосновать и построить тренд данного ряда. Оценить достоверность уточненной по МНК модели.

$t$	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
$Y(t)$	16,0	17,9	18,6	18,3	19,0	19,3	19,2	20,3	21,1	21,9

### Вариант 6

Изучается динамика потребления мяса птицы в Азии. Для этого собраны данные об объемах среднедушевого потребления мяса (кг/чел/год)  $Y(t)$  за 10 лет (2000-2009). Обосновать и построить тренд данного ряда. Оценить достоверность уточненной по МНК модели.

$t$	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
$Y(t)$	6,7	6,6	6,8	7,0	7,0	7,5	7,7	8,2	8,6	8,8

### Вариант 7

Изучается динамика потребления мяса птицы в Африке. Для этого собраны данные об объемах среднедушевого потребления мяса (кг/чел/год)  $Y(t)$  за 10 лет (2000-2009). Обосновать и построить тренд данного ряда. Оценить достоверность уточненной по МНК модели.

$t$	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
$Y(t)$	4,2	4,3	4,5	4,7	4,6	4,7	4,8	5,2	5,4	5,5

### Вариант 8

Изучается динамика объема депозитов и прочих средств, размещенных в банках в России. Для этого собраны данные об объемах указанных средств (млн.р.)  $Y(t)$  за 12 месяцев 2017г. Обосновать и построить тренд данного ряда. Оценить достоверность уточненной по МНК модели.

$t$	1	2	3	4	5	6	7	8	9	10	11	12
$Y(t)$	10,64	10,61	10,64	10,73	10,84	10,92	11,04	11,19	11,38	11,54	11,69	11,88

### Вариант 9

Изучается динамика объема кредитов, выданных в России. Для этого собраны данные об объемах указанных средств в иностранной валюте (усл.ед.)  $Y(t)$  за 12 месяцев 2017г. Обосновать и построить тренд данного ряда. Оценить достоверность уточненной по МНК модели.

$t$	1	2	3	4	5	6	7	8	9	10	11	12
$Y(t)$	16033	15224	14484	13445	13244	12946	13718	13553	12957	12223	11577	11818
)	0	9	9	1	6	0	6	1	8	9	3	1

### Вариант 10

Изучается динамика производства стали в мире. Для этого собраны данные об объемах ее производства (млн.т.)  $Y(t)$  за первые 7 месяцев 2018 года. Обосновать и построить тренд данного ряда. Оценить достоверность уточненной по МНК модели.

$t$	1	2	3	4	5	6	7
$Y(t)$	145	132	149	149	155	152	155

## ПРИЛОЖЕНИЕ 2.2. Пример оформления ЛР-2

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ  
федеральное государственное автономное образовательное учреждение высшего образования  
«САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ  
АЭРОКОСМИЧЕСКОГО ПРИБОРОСТРОЕНИЯ»

ИНСТИТУТ НЕПРЕРЫВНОГО И ДИСТАНЦИОННОГО ОБРАЗОВАНИЯ

КАФЕДРА КОМПЬЮТЕРНЫХ ТЕХНОЛОГИЙ  
И ПРОГРАММНОЙ ИНЖЕНЕРИИ (КАФЕДРА №43)

ОТЧЕТ  
ЗАЩИЩЕН С ОЦЕНКОЙ

ПРЕПОДАВАТЕЛЬ:

_____	_____	_____
должность, уч. степень, звание	подпись, дата	инициалы, фамилия

### ОТЧЕТ О ЛАБОРАТОРНЫХ РАБОТЕ №2 Моделирование временных рядов

по дисциплине: «Компьютерное моделирование»

РАБОТУ ВЫПОЛНИЛ

СТУДЕНТ ГР. №	B5431		В.А. Захаров
	номер группы	подпись, дата	инициалы, фамилия

Студенческий билет №	2015/1021
----------------------	-----------

Санкт-Петербург 2020



## Вариант 1\*

- 1) Используя метод наименьших квадратов, аппроксимировать исходные данные (см. варианты) линейной моделью  $y=ax+b$  (найти параметры  $a$  и  $b$ ).
- 2) Построить тренд с минимальным СКО из предложенных в пакете Excel.
- 3) Выяснить, какая из двух моделей лучше (адекватность в смысле метода наименьших квадратов) моделирует экспериментальные данные.
- 4) Сделать чертеж (на одних осях).
- 5) Сделать прогноз в последующий момент  $i=n+1$  по «лучшей» модели.

## Вариант 6

Изучается динамика потребления мяса птицы в Азии. Для этого собраны данные об объемах среднечеловеческого потребления мяса (кг/чел/год)  $Y(t)$  за 10 лет (2000-2009). Обосновать и построить тренд данного ряда. Оценить достоверность уточненной по МНК модели.

t	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
Y(t)	6.7	6.6	6.8	7.0	7.0	7.5	7.7	8.2	8.6	8.8

### Выполнение работы

#### Ход выполнения задания

1. Согласно методу наименьших квадратов (МНК) задача заключается в нахождении коэффициентов линейной зависимости, при которых функция двух переменных,  $a$  и  $b$ .

$$F(a, b) = \sum_{i=1}^n (y_i - (ax_i + b))^2 \xrightarrow{a, b} \min$$

Решение примера сводится к нахождению экстремума функции двух переменных.

2. Вывод формул для нахождения коэффициентов,  $a$  и  $b$ . Составляется и решается система из двух уравнений с двумя неизвестными. Находим частные производные функции  $F(a, b)$  по переменным  $a$  и  $b$ , приравниваем эти производные к нулю:

$$\frac{\partial F}{\partial a} = -2 \sum_{i=1}^n (y_i - (ax_i + b)) x_i = 0,$$

$$\frac{\partial F}{\partial b} = -2 \sum_{i=1}^n (y_i - (ax_i + b)) = 0.$$

$$a = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - \left( \sum_{i=1}^n x_i \right)^2}, \quad b = \frac{\sum_{i=1}^n y_i - a \sum_{i=1}^n x_i}{n}.$$

Убедимся, что в найденной стационарной точке (a,b) функция F (a,b) принимает минимум. Дифференциал второго порядка должен быть положительно определенным, или матрица квадратичной формы дифференциала второго порядка для функции F (a,b) , в точке (a,b) должна быть положительно определенной (по критерию Сильвестра).

Матрица квадратичной формы имеет вид:

$$D = \begin{bmatrix} \frac{\partial^2 F}{\partial a^2} & \frac{\partial^2 F}{\partial a \partial b} \\ \frac{\partial^2 F}{\partial a \partial b} & \frac{\partial^2 F}{\partial b^2} \end{bmatrix} = \begin{bmatrix} 2 \sum_{i=1}^n x_i^2 & 2 \sum_{i=1}^n x_i \\ 2 \sum_{i=1}^n x_i & 2n \end{bmatrix}.$$

Значения элементов не зависят от, а и b, главные миноры положительны (докажите методом математической индукции), следовательно, характер экстремума определен по критерию Сильвестра.

3. Заполняем таблицу для удобства нахождения коэффициентов, а и b:

X,Y	i=1	i=2	i=3	i=4	i=5	i=6	i=7	i=8	i=9	i=10	SUM
Xi	1	2	3	4	5	6	7	8	9	10	55
Yi	6,7	6,6	6,8	7	7	7,5	7,7	8,2	8,6	8,8	74,9
XiYi	6,7	13,2	20,4	28	35	45	53,9	65,6	77,4	88	433,2
Xi*Xi	1	4	9	16	25	36	49	64	81	100	385

Решив систему уравнений получаем:

$$a = 0.258 \quad b = 6.073$$

4. Записываем итоговое выражение для линейной модели

$$Y(x) = ax + b = 0,258 * x + 6.073$$

5. Для ответа на вопрос: какая из кривых лучше аппроксимирует исходные данные, следует оценить погрешности аппроксимаций по формуле:

$$\sigma_1 = \sum_{i=1}^n (y_i - (ax_i + b))^2$$

Сравнивая, делаем вывод:

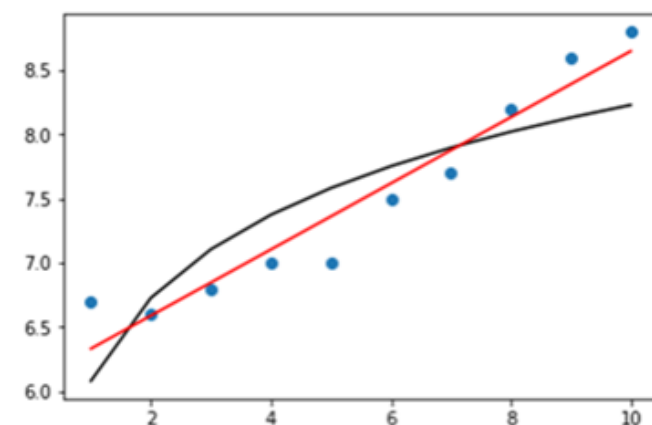
линейная модель на базе МНК - более удачная модель по сравнению с логарифмической.

s1	float	1	1.652843371310869
s2	float	1	0.3955151515151517

6. Графическая интерпретация степени близости кривых к исходным данным:

кривая  $y(x) = 0.258 * x + 6.073$  (красное начертание),  $y(x) = 0,9937 * \ln(x) + 6.0797$  (черное начертание), исходные данные – набор точек (синий цвет начертания).

$$y = 0.257575757576x + 6.07333333333$$



МНК имеет лучшее приближение

### Выводы по работе

В ходе работы освоены средства моделирования МНК для временных рядов, разработана программа для моделирования исходных данных с обоснованием выбора "наилучшей модели".

Исходные данные были аппроксимированы линейным уравнением, для которого были посчитаны коэффициенты. Итоговое линейное уравнение

имеет вид –  $y(x) = 0.258 * x + 6.073$ .

Получено функциональное приближение в форме логарифмической функции:

$$y(x) = 0,9937 * \ln(x) + 6.0797.$$

Сделан сравнительный прогноз для следующего временного промежутка

$$i=11(x = 2010).$$

	2010
линейное	8,911
логарифмическое	8,462489

### Листинг программы

```
# -*- coding: utf-8 -*-
import matplotlib.pyplot as plt
import math
X = [1,2,3,4,5,6,7,8,9,10]
Y = [6.7, 6.6, 6.8, 7.0, 7.0, 7.5, 7.7, 8.2, 8.6, 8.8]
# тренд
def y(x):
    return 0.9337*math.log(x, math.e) + 6.0797
def A(x, y):
    n = len(x)
    xy = []
    x2 = []
    for i in range(n):
        xy.append(x[i] * y[i])
    for i in x:
        x2.append(i ** 2)
    return (n * sum(xy) - sum(x) * sum(y)) / (n * sum(x2) - sum(x) ** 2)
def B(a, x, y):
    return (sum(y) - a * sum(x)) / len(x)
# погрешности аппроксимаций
def sigma(Y, F):
    delta = []
    for i in range(len(Y)):
        delta.append((Y[i] - F[i]) ** 2)
    return sum(delta)
# y = ax+b
a = A(X, Y)
b = B(a, X, Y)
F = [] # аналитическая
F1 = [] # по мнк
for i in range(len(X)):
    F.append(y(X[i]))
    F1.append(a * X[i] + b)
print('y = ' + str(a) + 'x + ' + str(b))
s1 = sigma(Y, F) # аналитическая
s2 = sigma(Y, F1) # мнк
plt.figure(1)
plt.plot(X, Y, 'o')
plt.plot(X, F, 'k')
plt.plot(X, F1, 'r')
```

```
plt.show()
if (s1 < s2):
    print("Логарифмическое уравнение имеет лучшее приближение")
else:
    print("МНК имеет лучшее приближение")
```