МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ

РОССИЙСКОЙ ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ

УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ

«РОССИЙСКИЙ ГОСУДАРСТВЕННЫЙ

ПЕДАГОГИЧЕСКИЙ УНИВЕРСИТЕТ им. А. И. ГЕРЦЕНА»

**институт информационных технологий и технологического образования**

**кафедра информационных технологий и электронного обучения**

Основная профессиональная образовательная программа

Направление подготовки 09.03.01 Информатика и вычислительная техника

Направленность (профиль) «Технологии разработки программного обеспечения»

форма обучения очная

**Курсовая работа**

по дисциплине «Технологии компьютерного моделирования»

Применение метода наименьших квадратов в экономике

Обучающейся 2 курса

Мельниковой Т. В.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Руководитель:

д.п.н, профессор

\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Власова Е. З.

«\_\_\_\_\_\_\_» \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ 2020 г.

Санкт-Петербург

2020

Оглавление

[ВВЕДЕНИЕ 3](#_Toc42271006)

[ГЛАВА 1 ТЕОРЕТИЧЕСКАЯ ЧАСТЬ 4](#_Toc42271007)

[1 Основные положения регрессионного анализа 4](#_Toc42271008)

[2 Метод наименьших квадратов 4](#_Toc42271009)

[3 Решение системы линейных уравнений 5](#_Toc42271010)

[ГЛАВА 2 ПРАКТИЧЕСКАЯ ЧАСТЬ 8](#_Toc42271011)

[1 Программный код 8](#_Toc42271012)

[2 Задача №1 11](#_Toc42271013)

[3 Задача №2 14](#_Toc42271014)

[ЗАКЛЮЧЕНИЕ 18](#_Toc42271015)

[ЛИТЕРАТУРА 19](#_Toc42271016)

# ВВЕДЕНИЕ

Наш мир не идеален, ни в чем нельзя быть уверенным с абсолютной точностью.

Эконометрика - наука о применении математических методов в экономическом анализе для проверки правильности экономических теоретических моделей и способов решения экономических проблем.

Модели используются в анализе и прогнозировании общих закономерностей и конкретных количественных характеристик рассматриваемых процессов. Но при этом часто возникают ошибки.

Метод наименьших квадратов (МНК), о котором пойдет речь в работе, является одним из способов противостоять ошибкам измерений.

Метод наименьших квадратов является одним из наиболее разработанных и распространенных методов оценки параметров линейных эконометрических моделей. Он не предъявляет жестких требований к закону распределения ошибок моделей. Вследствие этого оценки коэффициентов моделей, полученные на основе МНК, не зависят от фактического (или предполагаемого) закона распределения.

Актуальность темы определена тем, что МНК широко используется в регрессионном анализе, который позволяет спрогнозировать искомые признаки на основе известного.

Цель работы: решение задач экономики с помощью метода наименьших квадратов.

Для достижения поставленной цели необходимо решить следующие задачи:

* рассмотреть применение метода наименьших квадратов для нахождения неизвестных параметров уравнения регрессии на примере модели линейной и параболической парной регрессии;
* разработка программного кода для нахождения коэффициентов полиномиальной регрессии;
* представление результата в графическом виде;
* рассмотреть применение МНК на конкретных примерах в области экономики.

Объектом исследования является метод наименьших квадратов.

Работа состоит из введения, основной части, которая включает в себя рассмотрение теоретических вопросов и практическую часть, состоящую из решений задач, заключения.

При анализе различных источников информации предпочтение отдано книгам, посвященным эконометрики и математическому моделированию.

# ГЛАВА 1 ТЕОРЕТИЧЕСКАЯ ЧАСТЬ

## 1 Основные положения регрессионного анализа

Регрессионный анализ в моделировании представляет собой исследования, которые применяют для оценки взаимосвязи между какими-либо переменными. Этот математический метод состоит из множества других методов для моделирования и анализа нескольких переменных[1].

**Регрессионный анализ** – это один из основных методов в современной математической статистике, позволяющий аналитически представить связь между переменными объекта.

Модель регрессии основывается на известных данных, в роли которых выступают пары чисел. Количество таких пар определено заранее. Если представить, что первое число в паре – значение координаты X, а второе – Y, то можно множество таких пар чисел представить в декартовой системе координат в виде множества точек. На практике, зачастую, второе число зависит от первого. Построить регрессию – значит подобрать такую линию (если быть точнее, функцию), которая как можно точнее приближает к себе (аппроксимирует) множество вышесказанных точек[14].

Цель регрессионного анализа – при помощи уравнения регрессии предсказать ожидаемое среднее значение результирующей переменной.

Коэффициент регрессии показывает, как в среднем изменится результативный признак, если увеличится на единицу факторный признак [5].

Чтобы узнать коэффициенты уравнения регрессии используют метод наименьших квадратов, основанный на минимизации суммы квадратов отклонений функций от исходных переменных.

Рассмотрим линейную и полиномиальную регрессии.

## 2 Метод наименьших квадратов

Рассмотрим метод наименьших квадратов для линейной регрессии.

Суть метода наименьших квадратов (или МНК) в том, что нужно найти такие коэффициенты линейной зависимости, при которых значение функции двух переменных будет минимальным. Иначе говоря, при определенных значениях *a* и *b* сумма квадратов отклонений представленных данных от получившейся прямой будет иметь наименьшее значение[12].

Чтобы вывести формулы для вычисления коэффициентов, необходимо составить систему уравнений с двумя переменными и решить её. Для этого вычисляются частные производные выражения по *a* и *b* и приравниваются к 0.

⇔⇔⇔

⇔

Рассмотрим метод наименьших квадратов в матричном виде для полиномиальной регрессии.

Искомая функция в общем виде может быть описана многочленом степени (n-1). А именно:

Тогда матричная интерпретация задачи, учитывая, что есть m измерений:

× =

[Матрица факторных признаков] × [Матрица искомых коэффициентов] = [Матрица результирующих признаков]

Несложно заметить, что линейная аппроксимация является частным случаем полиномиальной аппроксимации.

Запишем сокращенно:

X \* A = Y

- матричная форма записи МНК

Введем для обозначения:

Для нахождения неизвестных коэффициентов, необходимо решить матричное уравнение:

## 3 Решение системы линейных уравнений

 Для решения СЛАУ, выведенных ранее, можно использовать разные методы, к примеру, метод Гаусса оптимального исключения. В итоге получают коэффициенты по методу наименьших квадратов [2].

- СЛАУ в общем виде.

Метод Гаусса - метод, который применяется при решении СЛАУ [15].

В методе Гаусса оптимального исключения неизвестных по столбцам исходная матрица преобразуется в треугольную матрицу с диагональными элементами, не равными единицы. Рассмотрим его в общем виде на получении верхней правой треугольной матрицы[10].

В рассматриваемом алгоритме на этапе прямого хода необходимо получить треугольную матрицу. Для этого в цикле существует ведущая i-я строка и ведущий элемент aii. Чтобы получить ноль на месте ведомого элемента aki нужно узнать коэффициент преобразования k-й строки:

Дальше в каждом цикле частичного обнуления i-го столбца из каждой ведомой k-ой строки вычитается ведущая строка, кратная коэффициенту преобразования .

В рекуррентном виде весь алгоритм может быть записан в виде:

где, i = 1 ÷ (n-1),

k = (j+1) ÷ n,

j = i ÷ (n+1).

При j = i под диагональные элементы матрицы равны нулю.

На этапе обратного хода находим неизвестные коэффициенты. Алгоритм обратного хода имеет вид:

;

,

где i = (n-1) ÷ 1 - обратный порядок изменения.

**Вывод**

Уравнение регрессии даёт возможность найти значение зависимых переменных, при условии, что величина независимых переменных известна.

Его применяют для составления прогнозов. Зачастую необходимо узнать Y, зная только X, если он отличается от тех иксов, на основе которых строилась регрессия. Грамотно построив модель регрессии, можно с некоторой достоверностью «предсказать» значение Y [3].

На практике, в зависимости от ситуации, в построении моделей регрессии применяют линейные, параболические и другие полиномы различных степеней. Все эти модели являются частными случаями более сложной модели – полиномиальной.

Построить модель регрессии – значит найти параметры той функции, которая будет в ней фигурировать [4].

Для регрессии – два параметра: коэффициенты и свободный член. Для их нахождения применяют метод наименьших квадратов. С помощью решения системы линейных уравнений, которые получаем в МНК, можно найти необходимые коэффициенты.

# ГЛАВА 2 ПРАКТИЧЕСКАЯ ЧАСТЬ

## 1 Программный код

В ходе выполнения курсовой работы был написан код на языке программирования python в онлайн компиляторе repl.it. Он находит коэффициенты и строит уравнение регрессии. Программа может считать полиномы n-ой степени. Программная реализация:

import matplotlib.pyplot as plt

import numpy as np

import os

def printMatrix(matrix1):

for i in range(len(matrix1)):

for j in range(len(matrix1[i])):

print("{:4f}".format(matrix1[i][j]), end = " ")

print()

def graph(xx, yy, xp1, pp):

x1 = np.array(xx)

y1 = np.array(yy)

plt.plot(x1, y1, 'o', label='Original data', markersize=10)

y2 = 0

for i in range(pp+1):

y2 += xp1[i] \* (x1 \*\* i)

plt.plot(x1, y2, 'r', label='Fitted line')

plt.legend()

plt.show()

def toFixed(numObj, digits=0):

return f"{numObj:.{digits}f}"

def main():

p = int(input("Введите степень полинома: "))

choice = None

print('1 - Считывание данных из файла')

print('2 - Ввод вручную')

choice = input("Выберите способ считывания ")

if choice == '1':

with open("x\_data.txt") as f:

for line in f:

x=[float(i) for i in line.split()]

with open("y\_data.txt") as f:

for line in f:

y=[float(i) for i in line.split()]

n = len(x)

if choice == '2':

n = int(input("Введите количество столбцов: "))

x = []

y = []

print("X: ")

for i in range(n):

x.append(float(input()))

print("Y: ")

for i in range(n):

y.append(float(input()))

print(n)

x\_copy = x.copy()

matrix = []

matrix.append(x)

matrix.append(y)

# printMatrix(matrix)

xx = []

for i in range(n):

pr = 1

xx.append([])

for j in range(p+1):

xx[i].append(pr)

pr \*= matrix[0][i]

print("\nМатрица факторных признаков (X)\n")

printMatrix(xx)

xt = []

for i in range(p+1):

xt.append([])

for j in range(n):

xt[i].append(xx[j][i])

print("\nТранспонированная матрица факторных признаков (XT)\n")

printMatrix(xt)

cx = []

for i in range(p+1):

cx.append([])

for j in range(p+1):

cx[i].append(0)

for k in range(n):

cx[i][j] += xt[i][k] \* xx[k][j]

print("\nC = XT \* X \n")

printMatrix(cx)

y1 = []

for i in range(p+1):

y1.append([])

for j in range(1):

y1[i].append(0)

for k in range(n):

y1[i][j] += xt[i][k] \* matrix[1][k]

print("\nY1 = XT \* Y\n")

printMatrix(y1)

for i in range(p+1):

cx[i].append(y1[i][0])

print("\nРасширенная матрица C\n")

printMatrix(cx)

#метод Гаусса

for i in range(p+1):

for k in range(i+1, p+1):

if(cx[k][i] != 0):

c = cx[k][i]/cx[i][i]

for j in range(i, p+2):

cx[k][j] = cx[k][j] - c \* cx[i][j]

print("\nПреобразованная расширенная матрица C\n")

printMatrix(cx)

x[p] = cx[p][p+1] / cx[p][p]

for i in range(p, -1, -1):

s = 0

for j in range(i+1, p+1):

s += cx[i][j]\*x[j]

x[i] = (cx[i][p+1]-s) / cx[i][i]

print("\nКоэффициенты уравнения регрессии\n")

for i in range(p+1):

print("A{} = {}".format(i+1,x[i]))

yr = 'y = '

for i in range(p, -1, -1):

if(i == 0):

if(x[i] > 0):

yr += " + " + str(toFixed(x[i], 4))

if(x[i] < 0):

yr += str(toFixed(x[i], 4))

elif(i == p):

if(i == 1):

yr += str(toFixed(x[i], 4)) + "\*x"

else:

yr += str(toFixed(x[i], 4)) + "\*x^" + str(i)

elif(i == 1):

if(x[i] > 0):

yr += " + " + str(toFixed(x[i], 4)) + " \*x"

if(x[i] < 0):

yr += str(toFixed(x[i], 4)) + "\*x"

elif(x[i] > 0):

yr += " + " + str(toFixed(x[i], 4)) + "\*x^" + str(i)

elif(x[i] < 0):

yr += str(toFixed(x[i], 4)) + "\*x^" + str(i)

print("\nУравнение регрессии\n")

print(yr)

graph(x\_copy, y, x, p)

os.system("cls")

main()

## 2 Задача №1

Имеются данные, характеризующие уровень безработицы в регионе, % (таблица 1). Нанесите полученные результаты на координатное поле и рассчитайте коэффициенты регрессии.

Таблица 1

|  |  |  |
| --- | --- | --- |
| Месяц | Условное обозначение времени, X | Уровень безработицы, Y, % |
| Январь | 1 | 2,99 |
| Февраль | 2 | 2,66 |
| Март | 3 | 2,63 |
| Апрель | 4 | 2,56 |
| Май | 5 | 2,4 |
| Июнь | 6 | 2,22 |
| Июль | 7 | 1,97 |
| Август | 8 | 1,72 |
| Сентябрь | 9 | 1,56 |
| Октябрь | 10 | 1,42 |

Рассмотрим решение задачи с помощью электронных таблиц и с помощью написанного кода.

**Решение**

Моделирование в Excel

Для того чтобы построить уравнение регрессии в Excel необходимо выполнить следующие действиях [10, с. 70]:

* Внести все данные в Excel;
* Построить пары точек (x,y);
* На вкладке «Работа с диаграммами» выбрать «Макет»;
* Далее находим кнопку «Линия тренда» и открываем дополнительные параметры;
* Выбираем линейную линию тренда и ставим галочку в поле «показывать уравнение на диаграмме».

После выполнения всех действия мы получили график, на котором изображено уравнение регрессии и написаны его коэффициенты (рисунок 1).

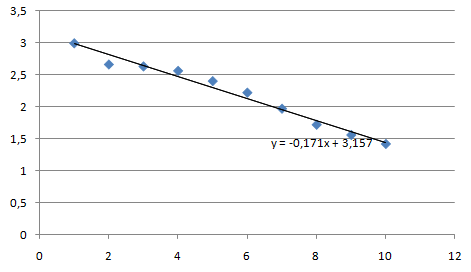


Рисунок 1

Коэффициенты: -0.171 и 3.157

Код Python

Реализация данной задачи представлена по ссылке:

<https://repl.it/@taniamelnikova/zadacha1#main.py>

Чтобы решить эту задачу, сделаем несколько действий:

* внесем в файл “x\_data.txt” значения независимой переменной, а в файл “y\_data.txt” – значения зависимой.
* Далее запустим программу, введем то, что просит программа
* Степень полинома в данном случае 1, так как мы ищем линейную зависимость.
* Способ считывания: 1, потому что мы ввели необходимые значения в нужные файлы.
* Затем получаем решение.

Промежуточные вычисления показаны на рисунках 2 - 4

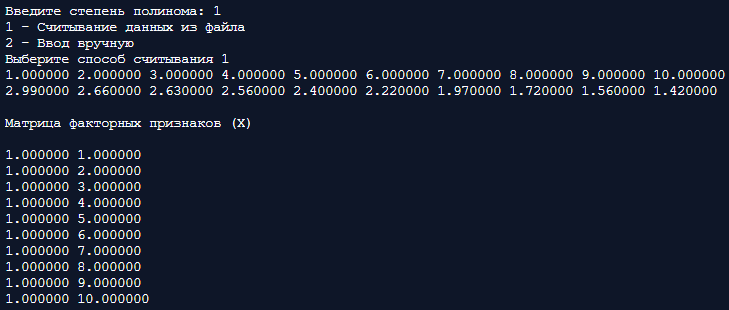


Рисунок 2

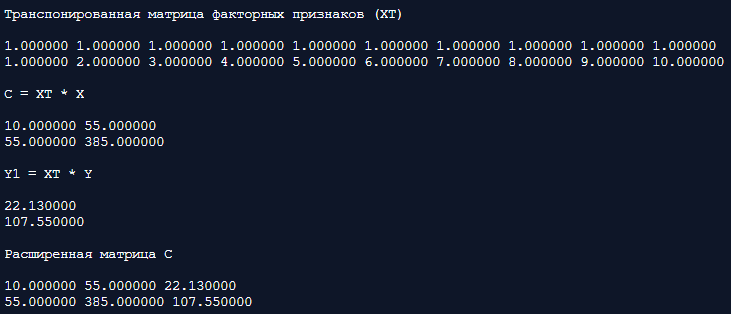


Рисунок 3

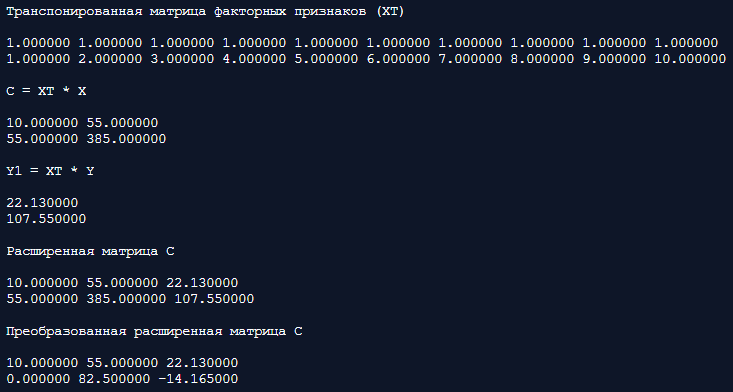


Рисунок 4

На рисунке 5 показаны вычисленные значения коэффициентов, а на рисунке 6 построенное уравнение регрессии.

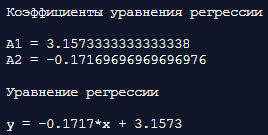
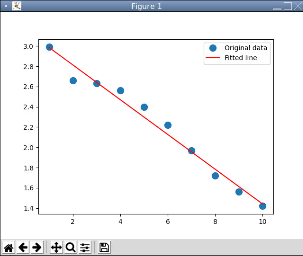
 

Рисунок 5 Рисунок 6

Коэффициенты: -0,1717 и 3,1573

Вывод: решение в программе с кодом на python привело к тому же результату, который был получен с помощью Excel.

## 3 Задача №2

Считая, что зависимость между переменными *x* и *y* имеет вид , найти оценки параметров a, b и c методом наименьших квадратов по выборке (таблица 2) и построить уравнение регрессии.

Таблица 2

|  |  |  |
| --- | --- | --- |
| i | xi | yi |
| 1 | -3,2 | -6,7 |
| 2 | -2,1 | -2,9 |
| 3 | -1,1 | 1,1 |
| 4 | 0,4 | 4,3 |
| 5 | 1,2 | 4,2 |
| 6 | 2,1 | 4,5 |
| 7 | 3,4 | 1,2 |
| 8 | 4,2 | -2,4 |
| 9 | 5,5 | -6,6 |
| 10 | 6,6 | -10 |

Решим эту задачу также двумя способами

**Решение**

Моделирование в Excel

Используем алгоритм, который был написан к задаче 1, но в параметрах линии тренда укажем не линейную, а полиномиальную линию тренда 2 порядка. Тогда получим результат, который показан на рисунке 7.

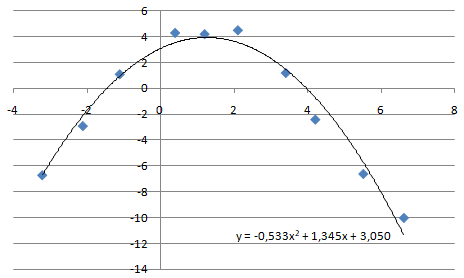


Рисунок 7

Коэффициенты: -0.533, 1.345 и 3.05.

Код Python

Реализация данной задачи представлена по ссылке:

<https://repl.it/@taniamelnikova/zadacha2#main.py>

Воспользуемся алгоритмом, представленным в задаче 1, но теперь будем вводить степень полинома 2, так как зависимость, по условию, квадратичная.

Тогда получим на рисунках 8 – 9 промежуточные вычисления.

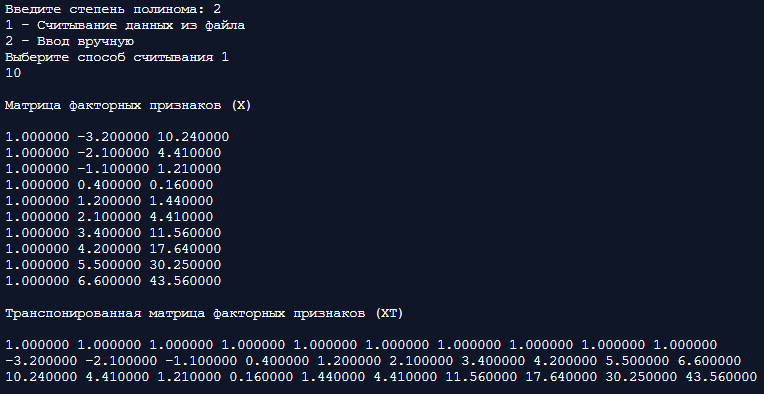


Рисунок 8

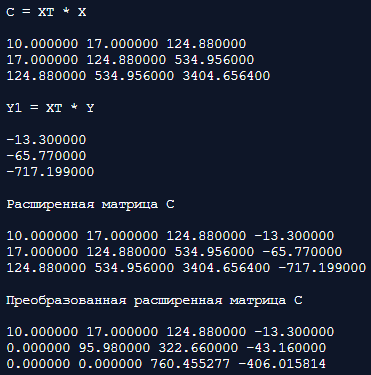


Рисунок 9

На рисунке 10 показаны вычисленные значения коэффициентов, а на рисунке 11 построенное уравнение регрессии.

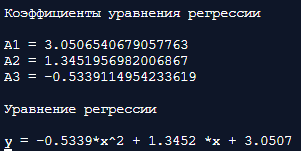


Рисунок 10

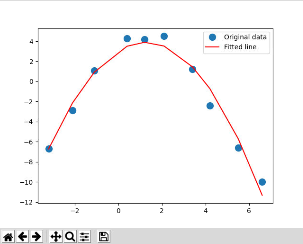


Рисунок 11

Коэффициенты: -0.5339, 1.3452 и 3.0507.

Вывод: решение в программе с кодом на python привело к тому же результату, который был получен с помощью Excel.

Таким образом, вид уравнения парной регрессии будет определяться исходными значениями. Исследуя полученные данные при решении задач, можно сделать вывод, что программный код, написанный на python, работает правильно. Это следует из того, что коэффициенты, найденные при двух разных подходах, равны, и графики уравнений регрессий совпадают.

Реализация метода наименьших квадратов на python удобна тем, что можно один раз ввести исходные значения, а потом строить разные степени полиномов, сравнивать графики и искать наиболее приближенные к парам точек (x,y).

# ЗАКЛЮЧЕНИЕ

В результате проделанной работы были изучены способы аппроксимации функции методом наименьших квадратов. В качестве аппроксимирующих кривых были выбраны: полином и прямая. Для каждого различного набора точек, наилучшим образом сглаживать его будет различная функция. Был составлен алгоритм, а также разработан программный код на Python, позволяющий находить коэффициенты уравнения регрессии и строить его.

В практической части были рассмотрены две задачи из исследуемой области. Для решения каждой из них использовались две информационных технологии: разработанный код и электронные таблицы. Полученные результаты сравнили между собой, коэффициенты оказались равны, графики идентичны, из чего можно сделать вывод, что программа готова для дальнейшего использования.

Данная работа может иметь дальнейшее развитие. В ней не были рассмотренные другие виды регрессии, например, множественная регрессия, показательная регрессия и т.д.

# ЛИТЕРАТУРА

1. Соколов Г. А. Эконометрика: теоретические основы: учеб. пособие. – М. : ИНФРА-М, 2018. – с. 119 – 147.
2. Шумак О. А., Гераськин А. В. Статистика: учеб. пособие. – М. :РИОР: ИНФРА-М, 2019.– с. 113 – 116.
3. Айвазян С. А. Методы эконометрики: учебник – Москва : Магистр : ИНФРА-М, 2020. – с. 33 – 65.
4. Артамонов Н. В. Введение в эконометрику. Электронное издание М.:МЦНМО, 2014. – с. 40 – 53.
5. Бабешко Л. О. Эконометрика и эконометрическое моделирование : учебник – М. : Вузовский учебник : ИНФРА – М, 2019. – с. 13 – 20.
6. Улитина Е. В., Леднева О. В., Жирнова О. Л. Статистика: учеб. пособие – М. : Московский финансово – промышленный университет «Синергия», 2013. – с. 206 – 215.
7. Коломейченко А. С. Математическое моделирование и проектирование: учеб. пособие – М. : ИНФРА-М, 2018. – с. 105 – 110.
8. Шпаков П. С. Математическая обработка результатов измерений: учеб. пособие – Красноярск: Сиб. федер. ун-т, 2014. – с. 140 – 146.
9. Кильдишов В. Д. Использование приложения MSExcelдля моделирования различных задач – Москва: СОЛОН-Пр, 2015. – с. 69 – 73.
10. Гантмахер Ф. Р. Теория матриц. – 5-е издание. – М., 2004. – с. 43 – 45.
11. Невежин В. П. Практическая эконометрика в кейсах: учеб. пособие – М. : ИД «ФОРУМ» : ИНФРА – М, 2019. – с. 75 – 100.
12. Метод наименьших квадратов. URL: <https://zaochnik.com/spravochnik/matematika/stati/metod-naimenshih-kvadratov/>
13. Метод наименьших квадратов для линейно зависимости. URL:<http://метод-наименьших-квадратов.рф/>
14. Модель полиномиальной регрессии: <https://habr.com/ru/post/414245/>
15. Метод Гаусса: <https://zaochnik.com/spravochnik/matematika/issledovanie-slau/metod-gaussa/>