**МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ**

**ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИ «РОССИЙСКИЙ ГОСУДАРСТВЕННЫЙ ПЕДАГОГИЧЕСКИЙ УНИВЕРСИТЕТ им. А. И. ГЕРЦЕНА»**

Институт компьютерных наук и технологического образования

Кафедра компьютерных технологий и электронного обучения

**КУРСОВАЯ РАБОТА**

МОДЕЛИРОВАНИЕ ЭКОЛОГИЧЕСКИХ ПРОЦЕССОВ С ИСПОЛЬЗОВАНИЕМ РЕГРЕССИОННОГО АНАЛИЗА

Направление подготовки: “Информатика и вычислительная техника”

Научный руководитель:

Уч. степень, уч. звание,

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

“\_\_” \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_2017 г.

Автор работы:

Студент группы

\_\_\_\_\_\_\_\_\_\_\_\_\_Е.И.Белорукова

“\_\_” \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_2017 г.

Оглавление

[ВВЕДЕНИЕ 3](#_Toc9801997)

[РАЗДЕЛ 1. ТЕОРИЯ И МАТЕМАТИЧЕСКОЕ ИССЛЕДОВАНИЕ ПРЕДМЕТНОЙ ОБЛАСТИ МОДЕЛИРОВАНИЯ, РЕГРЕССИОННОГО АНАЛИЗА И МНК. 4](#_Toc9801998)

[1.1 История развития компьютерного моделирования. 4](#_Toc9801999)

[1.2 Регрессионный анализ. Основные определения и понятия, виды анализа. 6](#_Toc9802000)

[РАЗДЕЛ 2. ПОСТАНОВКА ЭКОЛОГИЧЕСКОЙ ЗАДАЧИ И ЕЕ РЕШЕНИЕ 13](#_Toc9802001)

[2.1 Исследование зависимости количества использования свежей воды по Российской Федерации от количества выбросов наиболее распространенных загрязняющих атмосферу веществ стационарными и передвижными источниками 13](#_Toc9802002)

[2.2 Исследование зависимости количества использования свежей воды по Российской Федерации от объема сброса загрязненных сточных вод по бассейнам отдельных рек и морей Российской Федерации 18](#_Toc9802003)

[ЗАКЛЮЧЕНИЕ 24](#_Toc9802004)

[ЛИТЕРАТУРА 25](#_Toc9802005)

[ПРИЛОЖЕНИЕ А 26](#_Toc9802006)

[ПРИЛОЖЕНИЕ Б 28](#_Toc9802007)

[ПРИЛОЖЕНИЕ В 29](#_Toc9802008)

[ПРИЛОЖЕНИЕ Г 30](#_Toc9802009)

# ВВЕДЕНИЕ

**Актуальность темы исследования** обусловлена тем, что в настоящее время изучение экологических процессов стоит на первом месте в решении проблем планетарного масштаба. Страны всего мира уделяют большое внимание охране окружающей среды и ищут возможности решения проблем загрязнения природной среды. Ученые используют разные методы решения данной проблемы. Однако, помимо работы над текущей задачей нужно уметь «предсказывать», или говоря научным языком, моделировать процессы, происходящие в природе. Именно поэтому я считаю, что тема моего исследования и курсовой работы является очень актуальной на текущий момент времени.

**Объектом исследования** является компьютерное моделирование экологических процессов, а именно моделирование с использованием регрессионного анализа.

**Предмет исследования –** процессы, методы и способы при помощи, которых можно смоделировать природные процессы, используя регрессионный анализ.

**Цель исследования** заключается в составлении модели определенного экологического процесса с обращением к регрессионному анализу.

Для достижения указанной цели в курсовой работе решаются следующие исследовательские задачи:

1. Изучить теоретический материал данной темы в разных источниках.

2. Рассмотреть возможности моделирования в других сферах и процессах.

3. Сделать практическую модель с помощью регрессионного анализа.

**Методы исследования.** В курсовой работе применяются такие общенаучные методы исследования, как наблюдение, измерение, описание, сравнение, анализ, индукция, дедукция, аналогия, и некоторые другие.

**Структура исследования.** Курсовая работа содержит в себе введение, 2 раздела, в которых решаются задачи, указанные в целях работы, так же работа содержит заключение и список литературы и источников.

# РАЗДЕЛ 1. ТЕОРИЯ И МАТЕМАТИЧЕСКОЕ ИССЛЕДОВАНИЕ ПРЕДМЕТНОЙ ОБЛАСТИ МОДЕЛИРОВАНИЯ, РЕГРЕССИОННОГО АНАЛИЗА И МНК.

* 1. История развития компьютерного моделирования.

**История развития компьютерного моделирования**

История развития первых работ по компьютерному моделированию были связаны с физикой, где решался целый ряд задач по гидравлики, фильтрации и так далее. В основном моделирование представляло собой решение сложных задач по математической физике с помощью итерационных систем. После удачных результатов использования моделирования в физике оно распространилось в задачи по химии, биологии и некоторые другие дисциплины. Сложность решения задач с помощью моделирования всегда заключалась только в мощности используемых компьютеров.

**Значимость использования компьютерного моделирования.**

Современные ученые широко используют вычислительные эксперименты для прогнозирования процессов в природной среде. Однако, проведение реального эксперимента затрудняется как за счет использования слишком большого количества ресурсов, то есть возникнут финансовые и физические препятствия, так и за счет результатов эксперимента, которые могут быть непредсказуемыми и губительными для окружающей среды.

Поэтому технологии моделирования активно используются в настоящее время, затрагивая все области научных экспериментов. Современное компьютерное моделирование является средством общения людей (обмен информацией, программами и моделями), обучения людей на тренажерах, оптимизации работы за счет подбора параметров, а также изучения процессов, протекающих в окружающей среде. Таким образом снижая риск повреждения реальных сред обитания. На сегодняшний день целесообразность использования компьютерного моделирования для проведения научных исследований и сложных технических разработок не вызывает никаких колебаний.

Моделирование дает возможность изучать процесс до его осуществления. При этом выявляются возможные отрицательные последствия, что позволяет ликвидировать или ослабить их до реального проявления. Прогнозирование последствий - одна из важнейших целей (задач) моделирования.

**Основные понятия и виды моделирования.**

Модель – это материальный объект или компьютерная программа, которая имитирует структуру, реализуя представление объекта, системы или приближенной к реальной среде формы, которые включают в себя набор данных, характеризующих свойства системы и динамику их изменений со временем.

Моделирование – это представление, с помощью другой системы, различных характеристик поведения физической системы.

Существуют конкретные типы моделирования:

1. Физическое моделирование — это создание уменьшенных копий реальных объектов и систем. Основным недостатком такого типа модели является то, что при обратно масштабном переходе, т.е. при увеличении размеров, некоторые закономерности, которые соблюдались на модели, соблюдаться перестают. Физическое моделирование характеризуется, прежде всего, тем, что исследования проводятся на установках, обладающих физическим подобием, т. е. сохраняющих полностью или хотя бы в основном природу явлений. Если осуществлено полное или неполное физическое моделирование, то по характеристикам модели можно получить все характеристики оригинала пересчетом через масштабные коэффициенты.

2. Концептуальное моделирование — это создание блок-схем, взаимодействие подсистемы процессов в пределах более сложных систем. Примером являются круговороты веществ.

3. Графическое моделирование — это изображение зависимости между переменными в одной из систем координат, чаще всего в прямоугольной декартовой системе. Примером являются графики изменений численности популяций.

4. Математическое моделирование имеет огромное преимущество. Поскольку при этом способе моделирования нет необходимости сохранять размеры сооружении, нагрузки на элементы конструкции, имеется возможность получить существенный выигрыш во времени и стоимости исследования. Под этим видом моделирования понимают способ исследования различных процессов путем изучения явлений, имеющих различное физическое содержание, но описываемых одинаковыми математическими моделями. Например, колебания и волны различной природы (колебания маятника и колебания в электрической цепи аналогичны).

Однако, в моём исследование я прибегала к моделированию с использованием регрессионного анализа, поэтому необходимо обратиться к теоретической части анализа.

* 1. Регрессионный анализ. Основные определения и понятия, виды анализа.

Регрессионный анализ — статистический метод исследования влияния одной или нескольких независимых переменных на зависимую переменную.

Другими словами, независимые переменные обозначают как регрессоры или предикторы, так же зависимые переменные называют критериальными.

Регрессионный анализ необходим для выявления наиболее нужных факторов, влияющих на зависимую переменную

При работе с математическими зависимостями существует две формы взаимодействия между функцией и переменными: функциональная и регрессионная. Однако, в отличие от функциональных связей, которые могут быть точно выражены аналитическими уравнениями, регрессионные связи лишь приблизительно могут быть выражены уравнениями.

Исследователь составляет уравнение регрессии, как правило, поэтапно на основе связей между функцией и аргументами.

Для начала рассмотрим линейную форму связь вида:

(1)

где – факторы хорошо разработанного математического аппарата.

При условии этого уравнения могут решаться такие задачи, как:

1. Установление точности определения коэффициентов уравнения регрессии в виде значений дисперсий или величины доверительных интервалов;
2. Установление значимости коэффициента ;
3. Проверка адекватности установленной формы связи и экспериментальных данных.

При установлении тесноты связи между Y и Х решается задача установления строгости соблюдения функциональной зависимости между изменениями Y и X. Для оценки тесноты связи между случайными переменными величинами используются показатели в случае линейной формы связи:

1. Коэффициент парной корреляции или , характеризующий строгость соблюдения пропорциональности, т.е. близость исследуемой формы связи c линейной;
2. Коэффициент частной корреляции , характеризующий тесноту связи между изучаемыми переменными при условии, что влияние остальных факторов исключается;
3. Коэффициент множественной корреляции , характеризующий суммарное влияние всех факторов на величину Y.

Регрессионный анализ используется в двух направлениях:

1. Для проведения статистического анализа результатов наблюдений пассивных экспериментов, в которых экспериментатор может изменять независимые переменные и тем самые они не регулируются. При таком анализе решение вопроса о виде формы связи не является заключительным. И поэтому можно принять в качестве математической модели большое число регрессий, что могут удовлетворять экспериментальные данные.
2. Для планирования статистических экспериментов и анализа их результатов с использованием метода наименьших квадратов. В таком случае проведение эксперимента проходит в соответствии с общим видом уравнений связи между X и Y.

В соответствии с числом учитываемых независимых переменных и характером связи между X и Y различают:

А) по количеству исследуемых переменных: парный корреляционно-регрессионный анализ и множественный анализ.

Б) в зависимости от формы связи: линейный анализ и нелинейный анализ.

Именно поэтому моё исследование сводится к методу наименьших квадратов в котором планирование экспериментов осуществляется в соответствии с видом уравнения связи X и Y.

**Метод наименьших квадратов.**

В практике математического моделирования распространено уравнения вида:

(2)

где x - величина, рассматриваемая как случайная независимая величина, а y - случайная зависимая величина.

При линейной форме связи эту зависимость можно выразить уравнением прямой:

*(3)*

Для построения этого уравнения требуется проведение экспериментов в объеме n, в каждом из которых должна быть заключена пара значений (). Результаты экспериментов представляются либо в виде таблиц, либо в виде графиков.

Если экспериментальные точки () в прямоугольной системе координат частично лежат на прямой Y=b0+b1X (часть выше и ниже ее), то в данном случае используется линейное уравнение регрессии для построения зависимостей между X и Y. В данной задаче необходимо определить коэффициенты и , которые определяют положение прямой относительно всех экспериментальных точек так, чтобы сумма квадратов отклонений между значениями (полученными экспериментально) и значениями y при (подставляемое в гипотетическое уравнение) было минимально. Для этого необходим метод наименьших квадратов.

Пусть разность между экспериментальными и гипотетическими значениями равна при

(4)

или

*(5)*

При изменении величин и меняется . Возьмем функцию:

*(6)*

Являющуюся функцией двух переменных и . Наилучшая прямая, описывающая зависимость Y от X для экспериментальных данных будет такая, где значение

*(7)*

Минимум находится путем приравнивания частных производных от к нулю.

*(8)*

*(9)*

Решение первого уравнения:

*(10)*

*(11)*

Для упрощения вычисления можно вынести 2 за знак суммы и разделить обе части уравнения на 2.

*(12)*

По свойства суммы можно раскрыть скобки:

*(13)*

*(14)*

*(15)*

Решение второго уравнения:

*(16)*

*(17)*

Для упрощения вычисления можно вынести 2 за знак суммы и разделить обе части уравнения на 2.

*(18)*

По свойства суммы можно раскрыть скобки:

*(19)*

Разделим обе части уравнения на n и выполним замену с помощью формулы ():

*(20)*

*(21)*

Коэффициент – это постоянная уравнения, определенная при , а – это угол наклона прямой регрессии Y к оси OX.

Качество уравнений регрессии оценивается с помощью средней ошибки аппроксимации, вычисляемой по следующей формуле:

*(22)*

где - количество измерений;

- заданные значения;

- рассчитанные значения.

Значение уравнения регрессии считается качественным, если ).

Другой критерий для оценки полученной регрессионной модели – это средний коэффициент эластичности, рассчитываемы по формуле:

*(23)*

Средний коэффициент эластичности показывает, что в среднем при увеличении изучаемого фактора на 1% результирующий признак в среднем увеличивается на величину Э, выраженную в процентах.

В качестве меры зависимости между случайными величинами используется коэффициент корреляции, который всегда находится в пределах .

Формула для вычисления коэффициента корреляции:

*(24)*

где среднеквадратичное отклонение по ;

среднеквадратичное отклонение по .

Если случайные величины X и Y независимы, то , а если связь функциональная, то .

В качестве меры адекватности регрессионной модели статистическим данным часто используют коэффициент детерминации. Он показывает на сколько процентов и в какой степени результирующий признак зависит от изучаемого фактора.

Коэффициент детерминации рассчитывается по следующей формуле:

*(25)*

Показатель детерминации показывает на сколько процентов и в какой степени результирующий признак зависит от изучаемого фактора.

Так как исходные данные – выборочные, то необходимо оценить значимость величины коэффициента корреляции. Для этого необходимо выдвинуть гипотезу , называемую нулевой. Проверка нулевой гипотезы производится путем применения к ней критерия Стьюдента, рассчитываемого по следующей формуле:

*(26)*

где – количество степеней свободы.

Также необходимо найти критическое распределения Стьюдента, которое можно узнать, используя таблицу «Критические точки распределения Стьюдента».

При сравнении и , если > , то нулевая гипотеза отвергается и считается, что изучаемый фактор оказывает статистически существенное влияние на результирующий признак. В противоположном случае, нулевая гипотеза считается верной.

# РАЗДЕЛ 2. ПОСТАНОВКА ЭКОЛОГИЧЕСКОЙ ЗАДАЧИ И ЕЕ РЕШЕНИЕ

* 1. Исследование зависимости количества использования свежей воды по Российской Федерации от количества выбросов наиболее распространенных загрязняющих атмосферу веществ стационарными и передвижными источниками

**Постановка задачи:** исследовать зависимость количества использования свежей воды по Российской Федерации (в миллиардах кубических метров) от выбросов наиболее распространенных загрязняющих атмосферу веществ стационарными и передвижными источниками (в тыс. тонн) с 2002 по 2017 года.

Данные, на основе которых будет проводиться исследование, представлены в приложении А.

**Математическое решение поставленной задачи**

Для упрощения вычислений в данном разделе будет использоваться MS Excel. Все расчеты в этой программе приведены в приложении Б.

Математическое решение поставленной задачи будет проводится в несколько этапов:

1. Исходя из целей и задач исследования зависимости установить результативный и факторные признаки.

Из поставленной задачи факторным признаком является выбросы наиболее распространенных загрязняющих атмосферу веществ стационарными и передвижными источниками (в тыс. тонн), в результативным - использование свежей воды по Российской Федерации (в миллиардах кубических метров).

1. Обоснование модели уравнения регрессии.

Так как в исследовании участвуют два признака, факторный и результирующий, то задача сводится к исследованию парной зависимости. Для парной зависимости обычно используют графический метод. По полученному графику подбирается модель регрессии.

На рисунке 1 представлен график, построенный по данным, приведенным в приложении А.

Рисунок 1

Из рисунка 1 можно сделать предположении о наличии линейной зависимости между данными признаками. Таким образом можно использовать линейное уравнение регрессии для построения модели зависимости от .

1. Рассчитать параметры уравнения регрессии методом наименьших квадратов.

Для вычисления параметров уравнения линейной регрессии и используются формулы (21) и (15) соответственно.   
Вычисление средних значений производиться в MS Excel.

*(27)*

*(28)*

Таким образом из формулы 3 следует, что линейное уравнение регрессии для поставленной задачи имеет вид:

*(29)*

1. С учетом вычисленных параметров построить модель, описывающую зависимость использования свежей воды по Российской Федерации от выбросов наиболее распространенных загрязняющих атмосферу веществ стационарными и передвижными источниками.

Расчет новых значений производится по формуле (29). Результаты представлены в приложении Б.

Модель, построенная по новым данным, представлена на рисунке 2.

Рисунок 2

1. Оценить качество уравнения с помощью средней ошибки аппроксимации.

Значение средней ошибки аппроксимации рассчитывается по формуле (22).

*(30)*

Из данного результата можно сделать вывод о качественности уравнения регрессии, так как

1. Найти средний коэффициент эластичности по формуле (23):

*(31)*

Так как средний коэффициент эластичности меньше 1, то при увеличении среднего количества выбросов наиболее распространенных загрязняющих атмосферу веществ стационарными и передвижными источниками на 1% использование свежей воды по Российской Федерации измениться меньше, чем на 1%.

1. Оценить тесноту связи между переменными с помощью показателей корреляции и детерминации.

Показатель корреляции рассчитывается по формуле (24)

*(32)*

По таблице В.1 (Приложение В) можно сделать вывод о том, что между изучаемыми факторами средняя статистическая взаимосвязь.

Показатель детерминации рассчитывается по формуле (25):

*(33)*

Значение коэффициента детерминации в процентах показывает, что использование свежей воды по Российской Федерации зависит от выбросов наиболее распространенных загрязняющих атмосферу веществ стационарными и передвижными источниками только на 34,593% и на 65,407% от других факторов.

1. Оценить значимость коэффициентов корреляции и регрессии по критерию t-Стьюдента при уровне значимости α = 0.05.

Исходные данные выборочные, поэтому значимость коэффициентов корреляции и регрессии производится с помощью гипотез.

Пусть коэффициент корреляции совокупности равен 0 и количество использования свежей воды по Российской Федерации не зависит от количества выбросов наиболее распространенных загрязняющих атмосферу веществ стационарными и передвижными источниками.

Таким образом,

*(35)*

*(36)*

Проверка нулевой гипотезы производится по t- критерию Стьюдента, рассчитываемого по формуле (26):

*(34)*

Критическое значение t- критерия Стьюдента находится из таблицы В.2 «Критические точки t-распределения Стьюдента». (Приложение В)

Для данной задачи использует уровень значимости α = 0.05 и степень свободы равную 16.

*(35)*

Из результатов (34) и (35) следует, что расчетное t больше критического. А следовательно нулевая гипотеза отвергается, за верную принимается гипотеза Из этого можно сделать вывод о том, что количество выбросов наиболее распространенных загрязняющих атмосферу веществ стационарными и передвижными источниками оказывает статистически существенное влияние на количество использования свежей воды по Российской Федерации.

Результаты математического решения показывают, что количество использования свежей воды по Российской Федерации всего лишь на 34% напрямую зависит от количества выбросов наиболее распространенных загрязняющих атмосферу веществ стационарными и передвижными источниками. При этом построенная модель регрессии имеет вид (29) и является качественной, исходя из вычисленного значения средней ошибки аппроксимации.

**Программная реализация**

Программа для построения регрессионной модели и оценки ее качества была написана на языке программирования Python 3. Программный код представлен в Приложении Г.

При вводе данных для исследования в виде двух массивов, названий исследуемого и результирующего фактора, названия осей графика, пользователь получает полный отчет, который включает в себя уравнение регрессии и выводы, основанные на критериях оценки ее качества, описанные в теоретической части данной курсовой работы. Так же на экран выводится график, с нанесенными на него изначальными данными и график полученной регрессионной модели.

* 1. Исследование зависимости количества использования свежей воды по Российской Федерации от объема сброса загрязненных сточных вод по бассейнам отдельных рек и морей Российской Федерации

**Постановка задачи:** исследовать зависимость количества использования свежей воды по Российской Федерации (в миллиардах кубических метров) от объема сброса загрязненных сточных вод по бассейнам отдельных рек и морей Российской Федерации (в миллиардах кубических метров) с 2002 по 2017 года.

Данные, на основе которых будет проводиться исследование, представлены в приложении А.

**Математическое решение поставленной задачи**

Для упрощения вычислений в данном разделе будет использоваться MS Excel. Все расчеты в этой программе приведены в приложении Г.

Математическое решение поставленной задачи будет проводится в несколько этапов:

1. Исходя из целей и задач исследования зависимости установить результативный и факторные признаки.

Из поставленной задачи факторным признаком является объема сброса загрязненных сточных вод по бассейнам отдельных рек и морей Российской Федерации (в миллиардах кубических метров), а результативным - использование свежей воды по Российской Федерации (в миллиардах кубических метров).

1. Обоснование модели уравнения регрессии.

Так как в исследовании участвуют два признака, факторный и результирующий, то задача сводится к исследованию парной зависимости. Для парной зависимости обычно используют графический метод. По полученному графику подбирается модель регрессии.

На рисунке 1 представлен график, построенный по данным, приведенным в приложении А.

Рисунок 3

Из рисунка 3 можно сделать предположении о наличии линейной зависимости между данными признаками. Таким образом можно использовать линейное уравнение регрессии для построение модели зависимости от .

1. Рассчитать параметры уравнения регрессии методом наименьших квадратов.

Для вычисления параметров уравнения линейной регрессии и используются формулы (21) и (15) соответственно.   
Вычисление средних значений производится в MS Excel.

*(27)*

*(28)*

Таким образом из формулы 3 следует, что линейное уравнение регрессии для поставленной задачи имеет вид:

*(29)*

1. С учетом вычисленных параметров построить модель, описывающую зависимость использования свежей воды по Российской Федерации от объема сброса загрязненных сточных вод по бассейнам отдельных рек и морей Российской Федерации.

Расчет новых значений производится по формуле (29). Результаты представлены в приложении Б.

Модель, построенная по новым данным, представлена на рисунке 4.

Рисунок 4

1. Оценить качество уравнения с помощью средней ошибки аппроксимации.

Значение средней ошибки аппроксимации рассчитывается по формуле (22).

*(30)*

Из данного результата можно сделать вывод о качественности уравнения регрессии, так как

1. Найти средний коэффициент эластичности по формуле (23):

*(31)*

Так как средний коэффициент эластичности меньше 1, то при увеличении объема сброса загрязненных сточных вод по бассейнам отдельных рек и морей Российской на 1% использование свежей воды по Российской Федерации измениться меньше, чем на 1%.

1. Оценить тесноту связи между переменными с помощью показателей корреляции и детерминации.

Показатель корреляции рассчитывается по формуле (24)

*(32)*

По таблице 1 можно сделать вывод о том, что между изучаемыми факторами средняя статистическая взаимосвязь.

Показатель детерминации рассчитывается по формуле (25):

*(33)*

Значение коэффициента детерминации в процентах показывает, что использование свежей воды по Российской Федерации зависит от объема сброса загрязненных сточных вод по бассейнам отдельных рек и морей Российской только на % и на 8,327% от других факторов.

1. Оценить значимость коэффициентов корреляции и регрессии по критерию t-Стьюдента при уровне значимости α = 0.05.

Исходные данные выборочные, поэтому значимость коэффициентов корреляции и регрессии производится с помощью гипотез.

Пусть коэффициент корреляции совокупности равен 0 и количество использования свежей воды по Российской Федерации не зависит от объема сброса загрязненных сточных вод по бассейнам отдельных рек и морей Российской Федерации.

Таким образом,

*(35)*

*(36)*

Проверка нулевой гипотезы производится по t- критерию Стьюдента, рассчитываемого по формуле (26):

*(34)*

Критическое значение t- критерия Стьюдента находится из таблицы В.2 «Критические точки t-распределения Стьюдента». (Приложение В)

Для данной задачи использует уровень значимости α = 0.05 и степень свободы равную 16.

*(35)*

Из результатов (34) и (35) следует, что расчетное t больше критического. А следовательно нулевая гипотеза отвергается, за верную принимается гипотеза Из этого можно сделать вывод о том, что объем сброса загрязненных сточных вод по бассейнам отдельных рек и морей Российской Федерации оказывает статистически существенное влияние на количество использования свежей воды по Российской Федерации.

Результаты математического решения показывают, что количество использования свежей воды по Российской Федерации всего лишь на 34% напрямую зависит от объема сброса загрязненных сточных вод по бассейнам отдельных рек и морей Российской Федерации. При этом построенная модель регрессии имеет вид (29) и является качественной, исходя из вычисленного значения средней ошибки аппроксимации.

**Программная реализация**

Программа для построения регрессионной модели и оценки ее качества была написана на языке программирования Python 3. Программный код представлен в Приложении Г.

При вводе данных для исследования в виде двух массивов, названий исследуемого и результирующего фактора, названия осей графика, пользователь получает полный отчет, который включает в себя уравнение регрессии и выводы, основанные на критериях оценки ее качества, описанные в теоретической части данной курсовой работы. Так же на экран выводится график, с нанесенными на него изначальными данными и график полученной регрессионной модели.

# ЗАКЛЮЧЕНИЕ

В заключение хочется еще раз подчеркнуть значимость моделирования в современном мире, особенно при работе с экологическими процессами. При помощи моделирования мы можем создать модель и предсказать последствия эксперимента в реальных условиях.

Цель исследования заключалась в составлении модели определенного экологического процесса с обращением к регрессионному анализу. Для этого я обратилась к официальной статистике Российской Федерации. В процессе исследования мы выяснили, что для построения правильной модели нам нужно обратиться к линейному регрессионному анализу, а точнее к методу наименьших квадратов.

Для сравнения, я составила две модели, в одной из них после просчета выяснилось, что зависимость взятых мной статистик минимальна, а значит не имеет большого значения. Однако, во второй модели зависимость была достаточно велика, чтобы исследование имело место быть.

# ЛИТЕРАТУРА

1) Богданова, Л.Н. Факторный анализ преступности: корреляционный и регрессионный методы: Монография / С.М. Иншаков, Л.Н. Богданова, А.Д. Виноградова; под ред. С.М. Иншаков. - М.: ЮНИТИ-ДАНА, 2012.

2) Голицын, А.Н. Промышленная экология и мониторинг загрязнения природной среды: Учебник / А.Н. Голицын. - М.: Оникс, 2010.

3) Дрейпер, Н. Прикладной регрессионный анализ / Н. Дрейпер, Г. Смит. - М.: Вильямс, 2016.

4) Дрейпер Н., Смит Г. Прикладной регрессионный анализ. Множественная регрессия = Applied Regression Analysis. — 3-е изд. — М.: «Диалектика», 2007.

5) Королев, А.Л. Компьютерное моделирование / А.Л. Королев. - М.: БИНОМ. ЛЗ, 2013.

6) Радченко С. Г. Методология регрессионного анализа: Монография. — К.: «Корнийчук», 2011

7) Сирота, А.А. Анализ и компьютерное моделирование информационных процессов и систем / Э.К. Алгазинов, А.А. Сирота; под общ. ред. проф. д.т.н. Э.К. Алгазинов. - М.: ДИАЛОГ-МИФИ, 2009.

8) Тарасевич, Ю.Ю. Математическое и компьютерное моделирование. Вводный курс: Учебное пособие / Ю.Ю. Тарасевич. - М.: ЛИБРОКОМ, 2013.

9) Электронные библиотеки как электронные ресурсы.

# ПРИЛОЖЕНИЕ А

1. [Выбросы наиболее распространенных загрязняющих атмосферу веществ стационарными и передвижными источниками: https://docs.google.com/spreadsheets/d/1FmgPRZLfFkpV82ZegBItRbTApBLuX\_8gJKNv\_UiJRSs/edit?usp=sharing](https://docs.google.com/spreadsheets/d/1FmgPRZLfFkpV82ZegBItRbTApBLuX_8gJKNv_UiJRSs/edit?usp=sharing)
2. Использование свежей воды по Российской Федерации(рис 1):

<https://docs.google.com/spreadsheets/d/1rpI2AVMxlkBzIszSbtuMRvVFoD5s1RM29ueSWaBG4Kk/edit?usp=sharing>

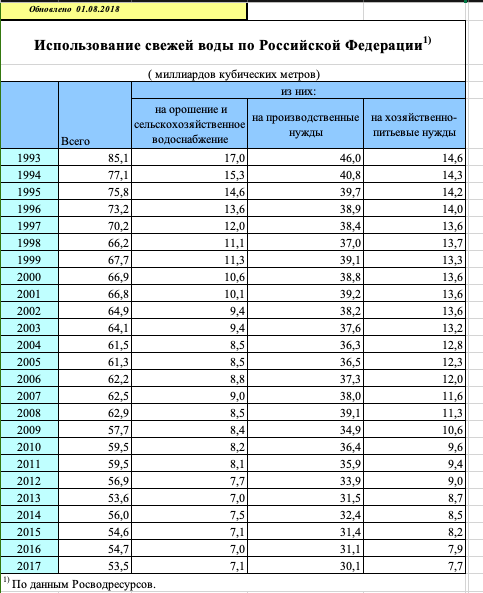


Рис 1.

1. Объем сброса загрязненных сточных вод по бассейнам отдельных рек и морей Российской Федерации(рис 2):

<https://docs.google.com/spreadsheets/d/1HmSR-vFYeqGRgE6UXIYt7CdTRC2oqdXEfFw0jgYS4Fg/edit#gid=1379284202>

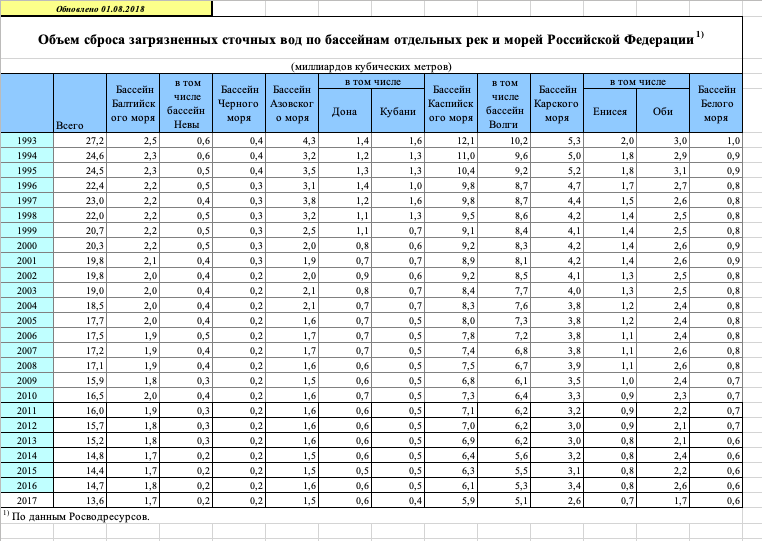


Рис 2.

# ПРИЛОЖЕНИЕ Б

1. Ссылка на исследование №1(решение в среде Excel) (рис 3):

[https://docs.google.com/spreadsheets/d/16Nk4RBle6eV9mOfJEOvVJPx04zD7Iy6Jt81fRd5klAw/edit#gid=1796815638](https://docs.google.com/spreadsheets/d/16Nk4RBle6eV9mOfJEOvVJPx04zD7Iy6Jt81fRd5klAw/edit#gid=1796815638#)

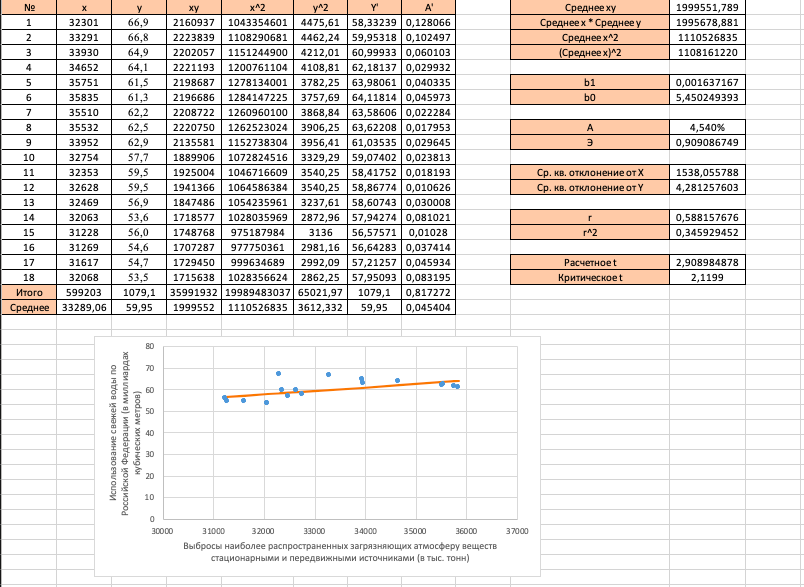


Рис 3.

1. Ссылка на исследование №2(решение в среде Excel) (рис 4):

<https://docs.google.com/spreadsheets/d/1-W8mmacVL4Of-OFpfj54UuQ76eyzdlAAKAWrB7kD8k0/edit#gid=1310835248>

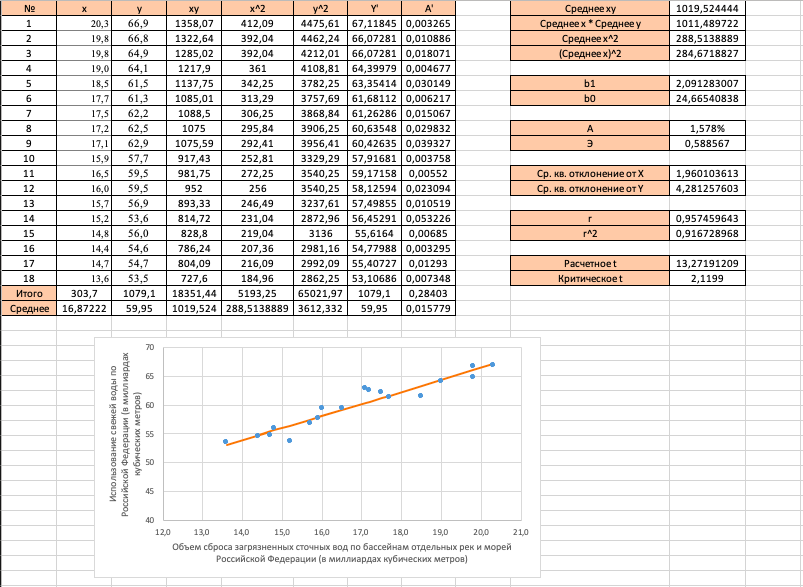


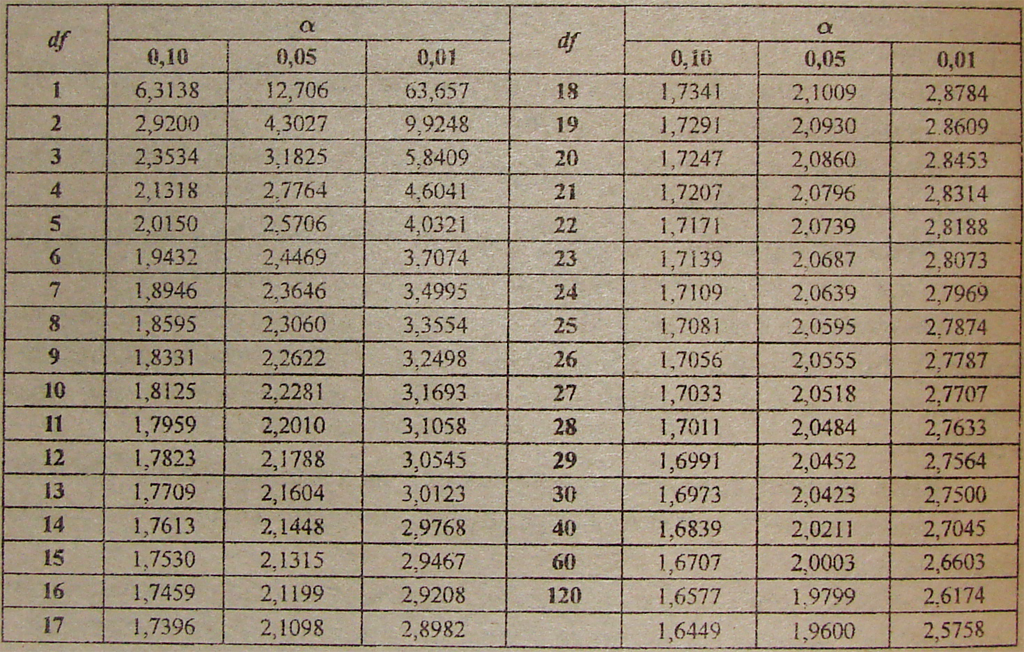
Рис 4.

# ПРИЛОЖЕНИЕ В

Таблица В.1

|  |  |
| --- | --- |
|  | функциональная зависимость |
|  | сильная статистическая взаимосвязь |
|  | средняя статистическая взаимосвязь |
|  | слабая статистическая взаимосвязь |
|  | очень слабая статистическая взаимосвязь |
|  | корреляции нет (линейной) |

Таблица В.2 Таблица критических значений t-критерия Стьюдента



# ПРИЛОЖЕНИЕ Г

Программная реализация на языке программирования Python 3:

import matplotlib.pyplot as plt

import numpy as np

def tCriteria(x):

selectX = {

x == 1: 12.70, x == 2: 4.303, x == 3: 3.182, x == 4: 2.776, x == 5: 2.571, x == 6: 2.447, x == 7: 2.365, x == 8: 2.306,

x == 9: 2.262, x == 10: 2.228, x == 11: 2.201, x == 12: 2.179, x == 13: 2.160, x == 14: 2.145, x == 15: 2.131, x == 16: 2.120,

x == 17: 2.110, x == 18: 2.101, x == 19: 2.093, x == 20: 2.086, x == 21: 2.080, x == 22: 2.074, x == 23: 2.069, x == 24: 2.064,

x == 25: 2.060, x == 26: 2.056, x == 27: 2.052, x == 28: 2.049, x == 29: 2.045, x == 30: 2.042, x == 31: 2.040, x == 32: 2.037,

x == 33: 2.035, x == 34: 2.032, x == 35: 2.030, x == 36: 2.028, x == 37: 2.026, x == 38: 2.024, x == 39: 2.023, x == 40: 2.021,

x == 41: 2.020, x == 42: 2.018, x == 43: 2.017, x == 44: 2.015, x == 45: 2.014, x == 46: 2.013, x == 47: 2.012, x == 48: 2.011,

x == 49: 2.010, x == 50: 2.009, x == 51: 2.008, x == 52: 2.007, x == 53: 2.006, x == 54: 2.005, x == 55: 2.004,x == 56: 2.003,

x == 57: 2.002, x == 58: 2.002, x == 59: 2.001, x == 60: 2.000, x == 61: 2.000, x == 62: 1.999, x == 63: 1.998, x == 64: 1.998,

x == 65: 1.997, x == 66: 1.997, x == 67: 1.996, x == 68: 1.995, x == 69: 1.995, x == 70: 1.994, x == 71: 1.994, x == 72: 1.993,

x == 73: 1.993, x == 74: 1.993, x == 75: 1.992, x == 76: 1.992, x == 77: 1.991, x == 78: 1.991, x == 79: 1.990, x == 80: 1.990,

81 <= x <= 90: 1.987, 91 <= x <= 100: 1.984, 101 <= x <= 110: 1.982, 111 <= x <= 120: 1.980, 121<= x <= 130: 1.978,

131 <= x <= 140: 1.977, 141 <= x <= 150: 1.976, 151 <= x <= 200: 1.972, 201 <= x == 250: 1.969,

251 <= x <= 300: 1.968, 301 <= x <= 350: 1.967

}

return selectX[True]

def kReg(r):

selectR = {

r == 1: 'функциональная зависимость',

0.7 <= r <= 0.99: 'сильная статистическая взаимосвязь, близкая к линейной',

0.5 <= r <= 0.69: 'средняя статистическая взаимосвязь',

0.2 <= r <= 0.49: 'слабая статистическая взаимосвязь',

0.09 <= r <= 0.19: 'очень слабая статистическая взаимосвязь',

r == 0: 'корреляции нет (линейной)'

}

return selectR[True]

def printTest(args):

[A, E, r, r2, t, n] = args

# средняя ошибка аппроксимации

if A > 10:

print('Средняя ошибка аппроксимации равна {:.3f}%, что больше 10%,'.format(A))

print('а значит модель не является качественной.\n')

else:

print('Средняя ошибка аппроксимации равна {:.3f}%, что меньше или равно 10%,'.format(A))

print('а значит модель является качественной.\n')

# средний коэффициент эластичности

print('Коэффициент эластичности показывает, что в среднем по совокупности результативный признак изменится')

print(' на {:.3f}% при изменении изучаемого фактора на 1% от своего среднего значения.\n'.format(E))

# коэффициент корреляции

if r == 0:

print('Между признаками {text}.\n'.format(text = kReg(abs(r))))

elif (r > 0):

print('Коэффициет корреляции равен {:.3f}. Связь между признаками прямая.'.format(r))

print('Между факторами {text}.\n'.format(text = kReg(abs(r))))

else:

print('Коэффициет корреляции равен {:.3f}. Связь между признаками обратная.'.format(r))

print('Между факторами {text}.\n'.format(text = kReg(abs(r))))

# коэффициент детерминации

print('Коэффициет детерминации равен {:.4f}. Он показывает что исследуемый фактор влияет'.format(r2))

print('на результирующий признак только на {:.2f}%.'.format(r2\*100))

print('Остальные {:.2f}% приходятся на влияние других неучтенных признаков.\n'.format(100 - r2\*100))

# t-критерий Стьюдента

tKr = tCriteria(n - 2)

if (t > tKr):

print('Так как {:.3f} > {:.3f}, то нулевая гипотеза о равенстве нулю коэффициента '.format(t, tKr))

print('корреляции генеральной совокупности отвергается. Таким образом верна гипотеза H1 и')

print('изучаемый фактор оказывает статистически существенное влияние на результирующий признак.\n')

else:

print('Так как {:.3f} < {:.3f}, то нулевая гипотеза о равенстве нулю коэффициента корреляции '.format(t, tKr))

print('генеральной совокупности считается верной. Таким образом зучаемый фактор не оказывает ')

print('статистически существенное влияние на результирующий признак.')

return 0

def tests(y, newY, b1, avgX, avgY, d1, d2):

# средняя ошибка аппроксимации

A = np.mean(list(map(lambda y1, y2: abs(y1-y2)/y1, y, newY)))\*100

# средний коэффициент эластичности

E = b1\*avgX/avgY

# коэффициент корреляции

avgY2 = np.mean(list(map(lambda a: a \* a, y)))

r = d1/((np.sqrt(d2))\*(np.sqrt(avgY2 - avgY\*avgY)))

# коэффициент детерминации

r2 = np.sqrt(r)

# t-критерий Стьюдента

t = abs(r)/(np.sqrt((1-r\*r)/(len(y)-2)))

printTest([A, E, r, r2, t, len(y)])

return 0

def equation(x, y):

avgX = np.mean(x) # среднее значение x

avgY = np.mean(y) # среднее значение y

avgXY = np.mean(list(map(lambda a, b: a \* b, x, y))) # среднее значение xy

avgX2 = np.mean(list(map(lambda a: a \* a, x))) # среднее значение x^2

d1 = (avgXY - avgX\*avgY)

d2 = (avgX2 - avgX\*avgX)

b1 = d1/d2

b0 = avgY - b1 \* avgX

print('\nУравнение регрессии: y = {:.5f} {:+.5f}x\n'.format(b0, b1))

newY = list(map(lambda a: b0 + b1\*a, x)) # рассчитанные значения y

tests(y, newY, b1, avgX, avgY, d1, d2)

return newY

def drawPlot(x, y, textXlbl, textYlbk, xtit, ytit):

print('Исследуемый фактор x: ', xtit)

print('Результативный признак y: ', ytit)

newEq = equation(x, y)

for i in range(len(x)):

plt.scatter(x[i], y[i])

plt.plot(x, newEq)

plt.grid(True)

# plt.title(textTitle)

plt.xlabel(textXlbl)

plt.ylabel(textYlbk)

plt.show()

xtit = 'Выбросы наиболее распространенных загрязняющих атмосферу веществ стационарными и передвижными источниками (в тыс. тонн)'

ytit = 'Использование свежей воды по Российской Федерации (в миллиардах кубических метров)'

x = [32301, 33291, 33930, 34652, 35751, 35835, 35510, 35532, 33952, 32754, 32353, 32628, 32469, 32063, 31228, 31269, 31617, 32068]

y = [66.9, 66.8, 64.9, 64.1, 61.5, 61.3, 62.2, 62.5, 62.9, 57.7, 59.5, 59.5, 56.9, 53.6, 56.0, 54.6, 54.7, 53.5]

x2 = [20.3, 19.8, 19.8, 19.0, 18.5, 17.7, 17.5, 17.2, 17.1, 15.9, 16.5, 16.0, 15.7, 15.2, 14.8, 14.4, 14.7, 13.6]

xtit2 = 'Объем сброса загрязненных сточных вод по бассейнам отдельных рек и морей Российской Федерации (в миллиардах кубических метров)'

# drawPlot(x, y, 'Исследуемый праметр x','Результативный признак y', xtit, ytit)

# drawPlot(x2, y, 'Исследуемый праметр x','Результативный признак y', xtit2, ytit)