Пермский филиал федерального государственного автономного образовательного учреждения высшего образования   
«Национальный исследовательский университет   
«Высшая школа экономики»

*Факультет экономики, менеджмента и бизнес-информатики*

Можегова Анна Сергеевна

**Модульное задание №2 по дисциплине**

**«Прикладной статистический анализ»**

студента образовательной программы «Программная инженерия»

по направлению подготовки *09.03.04 Программная инженерия*

Статистическое исследование: Регрессионный и корреляционный анализ двумерных данных

**Руководитель:**

Арефьева Валерия  
 Александровна

Пермь, 2022 год

Оглавление

[Введение 3](#_Toc121263260)

[Глава 1. Изучение выбранных двумерных данных 5](#_Toc121263261)

[1.1 Экономическая интерпретация 5](#_Toc121263262)

[1.2 Характеристики по шкалам 5](#_Toc121263263)

[1.3 Определение результативного признака и независимой переменной 5](#_Toc121263264)

[Глава 2. Корреляционный анализ 7](#_Toc121263265)

[2.1 Корреляционное облако 7](#_Toc121263266)

[2.2 Определение связи по корреляционному облаку 9](#_Toc121263267)

[2.3 Коэффициент корреляции 10](#_Toc121263268)

[2.4 Гипотеза о наличии связи 11](#_Toc121263269)

[Глава 3. Регрессионный анализ 14](#_Toc121263270)

[3.1 Построение линейной регрессионной модели 14](#_Toc121263271)

[3.2 Описание показателей необходимых для оценки качества полученной модели 16](#_Toc121263272)

[3.3 Коэффициент детерминации 16](#_Toc121263273)

[3.4 Проверка значимости уравнения регрессии и его коэффициентов 17](#_Toc121263274)

[3.5 Выполнение предпосылок МНК 19](#_Toc121263275)

[3.6 Нелинейные модели регрессии 21](#_Toc121263276)

[3.7 Графическое представление данных 22](#_Toc121263277)

[Заключение 26](#_Toc121263278)

[Литература и используемые информационные источники 27](#_Toc121263279)

[Приложение 28](#_Toc121263280)

Введение

В представленной работе производится анализ двумерных данных, они связаны с выявлением зависимости между показателями численностью размещенных лиц в коллективных средствах размещения и оценкой туристского потока по 59 регионам Российской федерации.

Актуальность анализа и поиска зависимости обосновывается тем, что оценка туристического потока показывает количество выехавших российских граждан за границу по цели визита «туризм» и «частная». А численность размещенных лиц указывает на предоставленные туристам места для ночевки в приспособленном для данных целей помещении, отличительной особенностью которой является число номеров, определенное стандартами каждой страны индивидуально. Грубо говоря, одни наблюдения показывают передвижения граждан из страны, другой же внутри страны по регионам. Если же анализ покажет, что данные показатели зависимы друг от друга, еще и имеют положительную прямую тенденцию, то необходимо заострять внимание на эти вещи, если у государства имеется желание повысить передвижения граждан внутри страны с целью туризма, то нужно учитывать и количество их передвижений вне страны.

Целью работы является проведение корреляционного и регрессионного анализа двумерных показателей.

Задачами, которые необходимо выполнить для достижения цели, связанной с корреляционным анализом, являются:

* изучение выбранных данных;
* построить корреляционное облако;
* определить характер связи признаков;
* гипотезу о наличии (отсутствии) связи;
* сделать расчет линейный коэффициент корреляции и сделать анализ по его показателю(проверка);

Задачами, которые необходимо выполнить для достижения цели, связанной с регрессионным анализом, являются:

* построить линейную регрессионную модель;
* оценка качества полученной модели;
* провести описание коэффициентов регрессии;
* сделать интерпретацию коэффициента детерминации (R2);
* найти МНК-оценку для линейной модели;
* проверить гипотезу о значимости уравнения регрессии и коэффициента уравнения регрессии;
* графическое представление исходных данных и расчетных данных, полученных по модели регрессии;
* построить две нелинейные регрессионные модели методом линеаризации и выявить нелинейные уравнения регрессии;
* вычислить МНК-оценки для построенных нелинейных регрессионных моделей;
* сравнить МНК-оценки всех построенных моделей (включая линейную), сделать вывод о том, какая из моделей оптимальна.

# **Глава 1.** **Изучение выбранных двумерных данных**

В данной главе производится описание показателей, выбранных данных.

* 1. Экономическая интерпретация

Для своего исследования в качестве исходных данных я брала двумерные данные с Росстата из разных выборок. Первая выборка отвечала за численность размещенных лиц в коллективных средствах размещения, из нее я взяла данные по 59 регионам на январь 2022 года (см. Вложение 1). Второй кластер данных показывал оценку туристского потока на январь 2022 года, отсюда я также взяла данные по тем же 59 регионам России.

* 1. Характеристики по шкалам

Названия регионов/областей являются номинальной шкалой, ведь это данные, с которыми мы не может делать каких-либо арифметических операций, а вот значения оценки туристического потока или количества размещенных лиц — это интервальная шкала, эта шкала позволяет находить разницу между двумя величинами, таким образом, мы, например, можем рассмотреть разность размещенного количества людей в различных регионах.

* 1. Определение результативного признака и независимой переменной

Прежде, чем приступать к непосредственному определению переменных, нужно понимать что от нас требуется найти. Из курса статистики мне известно, что за зависимую переменную отвечает показатель, который описывает какой-либо процесс. Именно это процесс в дальнейшем мы пытаемся предсказать или понять. Независимая переменная в свою очередь используется для моделирования или прогнозирования значений зависимых переменных. Все переменные представляются в уравнении регрессии, где они располагаются справа от знака равенства и часто называются объяснительными переменными. Кроме того, о зависимой переменная можно сказать, что она является функцией независимых переменных.

В моем случае за независимую переменную отвечает оценка туристического потока, а за зависимую численность размещенных лиц в коллективных средствах размещения

# **Глава 2. Корреляционный анализ**

Первым этапом анализа двумерных данных является корреляционный анализ. Он представляет из себя один из способов обработки данных, в результате которого измеряется теснота связи между двумя или более переменными. Сама по себе корреляция как бы показывает степень согласованности каких-либо процессов. Еще корреляция не обязательно относится к двум элементам, т.е является парной, можно выявлять и множественную корреляцию. Чтобы рассмотреть данные в этом виде анализа нужно уметь строить корреляционное облако и знать, как определять по нему связь, кроме того, можно подкреплять эту связь расчётом коэффициента корреляции. Этим я и займусь в этой главе.

* 1. Корреляционное облако

Итак, диаграмма рассеивания или корреляционное облако применяется, когда нужно представить, что происходит с одной переменной, если друга переменная изменяется и проверить предложение о взаимосвязи двух переменных величин. С помощью диаграммы мы можем лишь наблюдать существующую связь и ее степень, нельзя сказать наверняка о том, что одна переменная является причиной изменения для другой.

Существует два вида зависимости: линейная и нелинейная. В свою очередь они еще делятся на прямую, обратную и отсутствующую связь. По диаграмме мы можем легко определить принадлежность данных к той или иной зависимости. Это необходимо не только для подтверждений своих домыслов, но и для определения уравнения регрессии, но о нем мы поговорим в следующей главе.

Со своими данными я построила корреляционное облако двумя способами. В первом случае я просто отложила по оси ординат значения, которые отвечают показатели первого признака, а по оси абсцисс разместила значения, отвечающие за показатели второго признака. Во втором случае уже необходимо было сделать расчеты для линейного уравнения. Сначала нам нужно получить столбцы с такими значениями , . После чего находим сумму и среднее значение всех столбцов, т.е, как исходных данных, так и новых полученных. Переходим к более интересным вещам, нужно рассчитать средние квадратичные отклонения . Мы получим через разность показателей среднего значения по столбцу и среднего значения по столбцу , возведенного в квадрат. Аналогично считается и , только берутся показатели, отвечающие за . Пока что я не буду вдаваться в подробности относительно уравнения регрессии, потому что мы еще не рассматриваем данное понятие подробно в данной главе, но я скажу, что дальше мы восстанавливаем соответствующие этим данным уравнение, для гистограммы мы откладываем значения, получившиеся для уравнения и значения показателя, который характеризует .

Несмотря на то, что корреляционные облака построено разными способами, они имеют одинаковых вид***.***

* 1. Определение связи по корреляционному облаку

Определить тип связи по визуальному преставлению зависимости показателей очень просто. Если все значения на точеном графике располагаются так, что значение и при возрастании/ убывании одной из величин приводят соответственно к возрастанию/ убыванию другой величины, грубо говоря, мы видим, что значения переменных увеличиваются или уменьшаются вместе , то наблюдается положительная корреляция, смотря на вид распределения можно сказать о том какое уравнение соответствует случаю, для линейного распределения значения будут находиться вдоль линии. Кроме того, существует понятие степени корреляции, она зависит от плотности точек. Точки, которые будут сильно отдаляться от основного распределения, являются выбросами. Если же значение, наоборот, меняются таким образом, что при увеличении/уменьшении одной, другая будет реагировать противоположным образом, т.е. уменьшаться/увеличиваться, то это отрицательная корреляция. Но существует случай, когда значения могут быть разбросаны рандомно или же скомкаются в круг, тогда у показателей отсутствует какая-либо связь.

Ниже приведен наглядная классификация для определения связи и ее степени по корреляционному облаку.



Рисунок 2.2 – Определение связи из визуального представления корреляционного облака.

* 1. Коэффициент корреляции

Коэффициент корреляции необходим для определения статистической связи между двумя переменными. Под переменными мы можем понимать два столбца с каким-то заданным набором данных(выборка) или двумя компонентами многомерной случайной величины с известным распределением.

Если корреляционное облако дает нам визуализацию связи, то коэффициент корреляции показывает тесноту взаимосвязи или же качество связи. Данный коэффициент может находиться в промежутке. Где минус один полную (функциональную) обратную взаимосвязь или же строгую отрицательную корреляцию, а один – полную (функциональную) положительную взаимосвязь или же строгую положительную корреляцию. Так же коэффициент может принимать значение 0, которое говорит об отсутствии какой- либо связи.

Мало определить наличие связи, она еще может быть сильной или слабой (речь идет о тесноте связи). Эта сила определяется по шкалам, например, по шкале Чеддока или шкале Е.П Голубкова. В шкале Чеддока говорится следующее: если коэффициент корреляции лежит в промежутке , то наблюдается слабая теснота показателей, в промежутке , теснота интерпретируется как умеренная, в промежутке , теснота будет называться заметной, промежуток является обозначением высокой тесноты между показателями, если же вдруг коэффициент и вовсе попал в , то это весьма высокая связь и теснота.

Для своего случая я рассчитала коэффициент корреляции двумя способами. Первый способ наиболее простой, нужно лишь воспользоваться встроенной функцией excel КОРЕЛЛ(), в качестве параметров вводим 2 массива выбранными для анализа показателями. Второй способ включает в себя использование формулы . В этой формуле легко запутаться, потому что по невнимательности можно перепутать , а также . В качестве я рассматривала второй массив данных, а первый массив данных.

Таким образом в обоих случаях я получила коэффициент корреляции равный 0.93 (с округлением). С помощью этого коэффициента я могу подтвердить графическое представление корреляционного облака. Так как коэффициент получился неотрицательным, то у моих данных имеется положительная корреляция с весьма высокой теснотой.

* 1. Гипотеза о наличии связи

Для проверки гипотезы о наличии линейной корреляционной связи наибольшее распространены расчеты коэффициент линейной корреляции (Пирсона), предполагающий нормальный закон распределения наблюдений, и коэффициент ранговой корреляции Спирмена.

В первом случае для двумерной нормально распределенной случайной величины проводится процедура проверки, а именно расчет выборочной оценки коэффициента корреляции и оценке значимости его отличия от нуля.

А коэффициент ранговой корреляции Спирмена используется для выявления и оценки тесноты связи между двумя рядами сопоставляемых количественных показателей. В том случае, если ранги показателей, упорядоченных по степени возрастания или убывания, в большинстве случаев совпадают (большему значению одного показателя соответствует большее значение другого показателя), делается вывод о наличии прямой корреляционной связи. Если ранги показателей имеют противоположную направленность (большему значению одного показателя соответствует меньшее значение другого), то говорят об обратной связи между показателями.

Грубо говоря, чтобы проверить корреляцию мы либо используем критерий хи-квадрат, который может применяться при анализе таблиц сопряженности, содержащих сведения о частоте исходов в зависимости от наличия фактора риска или же находим корреляцию другим способом, сравнивая ее с первоначальной.

Критерия хи-квадрат Пирсона имеет ограничения и указания, такие как:

1. Можно работать с показателями, которые измеряются в номильной или порядковой шкале. Например, пол пациента - мужской или женский, и степень заболевания, принимающая значения от 0 до 3.
2. Данный метод позволяет проводить анализ как для четырехпольных таблиц, когда и фактор, и исход являются бинарными переменными, так и для многопольных таблиц, когда фактор и (или) исход принимают три и более значений.
3. Показатели должны быть полностью независимыми (т.е. коэффициент корреляции равен 0), так же исключение на применение есть у данных "до-"после". Как я выяснила для проведения анализа над подобными данными работают с тестом Мак-Немара или рассчитывается Q-критерий Кохрена.
4. При анализе четырехпольных таблиц может возникнуть ситуация, когда нужно использовать расчет точного коэффициента Фишера, но это уже нужно вдаваться в детали, для таких исключительных ситуаций.

Для того чтобы рассчитать корреляцию Пирсона нужно сделать следующие этапы. Во-первых, необходимо рассчитать ожидаемое количество наблюдений в каждой строке и столбце таблицы. Каким путем это делается? Все очень просто перемножаются сумм рядов и столбцов с последующим делением полученного произведения на общее число наблюдений. Самое главное, нельзя забывать что все делается при условии нулевой гипотезы об отсутствии Далее считаем коэффициент , мы рассматриваем как бы матрицу, поэтому – номер строки (от 1 до ), – номер столбца (от 1 до ), – фактическое количество наблюдений в , – ожидаемое число наблюдений в . После чего мы сталкиваемся с понятием «степень свободы», она рассчитывается по следующей формуле: . Для четырехпольной таблицы данное значение будет равно 1. Все это считается не просто так, нам нужно сравниваем значение критерия χ2 с критическим значением при числе степеней свободы f, для этого существует таблица, в которой представляется еще и два уровня значимости 0,05 и 0,01. Уровень значимости — это не просто какой-то выдуманная характеристика, на самом деле это показывает вероятность того, что отклонение от нулевой гипотезы ошибочна. Исходя из этого мы и проверяем гипотезу наличия связи.

В своем случае для исходных данных я применила коэффициент Спирмена. Для того, чтобы найти его значение нужно воспользоваться формулой , где , .

Таким образом, коэффициент корреляции и коэффициент Спирмена совпали.

Глава 3. Регрессионный анализ

В данной разделе мы рассмотрим уже регрессионный анализ. Прежде, чем приступить к анализу разберемся с определениями.

Основными являются понятия: регрессия, регрессионный анализ и уравнение регрессии. Под регрессией в статистике понимается зависимость среднего значения какой-либо величины от некоторой другой величины или от нескольких других величин. Здесь и выявляется главное отличие от функциональной зависимости, ведь возможен вариант, что одному и тому же значению независимой переменной х могут соответствовать в зависимости от конкретного случая различные значения зависимой переменной у. Регрессионный анализ – это раздел математической статистики, изучающий регрессионную зависимость между случайными величинами. Это делается для определения общего вида уравнения регрессии. Уравнением регрессии называется числовое соотношение между величинами, его представляют в виде тенденции к возрастанию/ убыванию одной переменной величины при изменении другой. По уравнению мы оцениваем влияние каждой независимой переменной на прогнозируемые значения, включая коэффициент регрессии для каждой независимой переменной. Можно сравнить величины уклона для определения влияния каждой независимой переменной на зависимую переменную; чем дальше от нуля значение уклона – тем больше влияние.

Для линейных уравнений применяется функция . Парабола, гипербола, показательная и степенная функция относятся к нелинейным уравнениям.

* 1. Построение линейной регрессионной модели

В качестве уравнения для построения линейной регрессионной модели используется уравнение вида . Чтобы найти коэффициенты нужно решить систему уравнений, где первое уравнение имеет вид , а второе , где - число наблюдений.

После строится таблица расчётных данных, где необходимо представить такие значения, как . В конце каждого столбца считаем среднее значение по всему столбцу. Далее нам потребуется определить среднее квадратичное отклонение по формуле , возведением этого в квадрат мы получаем дисперсию. Параметр находится по формуле . Второй параметр находится по формуле . Таким образом я получила уравнение =-2332-0.55\*x. Следовательно, с повышением оценки туристического потока на одну единицу измерения, численность размещенных лиц в коллективных средствах размещения увеличивается в среднем на 0,2085.

* 1. Описание показателей необходимых для оценки качества полученной модели

Качество модели регрессии оценивается по нескольким факторам, а именно по проверке качества всего уравнения регрессии, проверке значимости всего уравнения регрессии, производится так же проверка статистической значимости коэффициентов уравнения регрессии и выполнения предпосылок МНК. Так же для оценки качества модели регрессии вычисляют коэффициент множественной корреляции (индекс корреляции) и коэффициент детерминации.

* 1. Коэффициент детерминации

Чтобы провести оценку адекватности модели нужно иметь показатель, отражающий, в какой мере функция регрессии определяется факторными (объясняющими) переменными. Изначально хотелось использовать в качестве оценивания сумма квадратов отклонений фактического значения зависимой. Но эта величина не подходит в силу того, что зависит от единицы измерения зависимой переменной Y и от числа наблюдений в выборке. Поэтому используется расчет коэффициента детерминации.

Однако не нужно полностью полагаться на данный показатель, так как коэффициент детерминации может быть близким к единице в силу того, что обе исследуемые величины Y и X имеют выраженный временной тренд, не связанный с их причинно-следственной зависимостью, в этом случае никто не говорит о гарантиях о наличии значимой линейной связи между исследуемыми показателями. Для этого существует и несколько параметров для оценки модели.

Найдём индекс детерминации для данной модели, используем следующую формулу или же можно ее заменить на функцию Excel КВПИРСОН() в качестве параметров принимается массивы данных по и . Если показатель равен 1, то эмпирические значения лежат на регрессионной прямой. Если коэффициент детерминации равен нулю, то между нет никакой корреляционной связи и линия регрессии параллельна оси . Таким образом, если существует статистически значимая линейная связь величин , то коэффициент детерминации должен быть близок к единице.

Далее было бы полезно определить линейный коэффициент парной корреляции, возвести его в квадрат и узнать коэффициент детерминации. Для этого воспользуемся формулой . Итак, далее мы сравниваем между собой и . Так мы дополнительно сможем узнать, как повлияло добавление фиктивной переменной в уравнение. В моем случае , а , из этого следует, что доля объяснённой вариации не изменилась, несмотря на это, связь между признаками остаётся сильной.

* 1. Проверка значимости уравнения регрессии и его коэффициентов

Для того чтобы провести оценку качества уравнения регрессии необходимо воспользоваться F-критерием Фишера. F-критерий помогает сделать проверку гипотезы статистической незначимости уравнения регрессии и показателя тесноты связи. Проведем сравнение фактического и критического (табличного) значений F-критерия Фишера.

Фактическое значение считается по формуле , где – число единиц совокупности, – число параметров при переменных х. Таким образом я получила значение 5.58, которое превышает табличное значение критерия. Это говорит о что гипотеза о случайной природе оцениваемых характеристик не принимается и признаётся их статистическая значимость и надёжность.

Оценку статистической значимости коэффициентов регрессии проводится с помощью t-статистики Стьюдента и путём расчёта доверительного интервала каждого из показателей. Выдвигаем гипотезу Н0 о статистически незначимом отличии показателей от нуля: . Табличное значение равно 2,002 при степени свободы равной .

Переходим к расчету случайной ошибки . Воспользуемся формулами и . Из полученных значений находятся фактические значения t-статистики по формулам и . После этого находим критерий для коэффициента корреляции, его можно найти либо по аналогичным первым двум формулам или же взяв корень из значения . Далее мы сравниваем все критерии с табличным значением, если значения будут меньше, то полученная линейная зависимость является недостоверной. В моем случае значения получились таким образом, что они превышают табличное. Из этого следует, что для коэффициентов нарушается гипотеза о их несущественности, а зависимость достоверна.

* 1. Выполнение предпосылок МНК

Сейчас мы проверим предпосылки относительно случайной составляющей ε. Доказано, что для получения по МНК наилучших результатов необходимо, чтобы выполнялся ряд предпосылок относительно случайного отклонения.

Рассмотрим условия, которые выдвигаются при теореме Гаусса-Маркова. Во-первых, математическое ожидание случайного отклонения обязательно должно быть равно нулю для всех наблюдений. Из этого высказывания следует, что случайное отклонение не имеет влияния над зависимой переменной. В каждом конкретном наблюдении случайный член может быть либо положительным, либо отрицательным, но он не должен иметь систематического смещения. В случае с парной корреляцией данная предпосылка не нуждается в подтверждение, потому что выполняется автоматически. Во-вторых, из первого условия следует, что дисперсия случайных отклонений постоянна для любых наблюдений i и j. Выполнимость данной предпосылки называется гомоскедастичностью (постоянством дисперсии отклонений). Обратное явление называется гетероскедастичностью. В-третьих, случайные отклонения параметров независимы друг от друга, при . Если эта предпосылка будет выполняться, то отсутствует систематическая связь между любыми случайными отклонениями. Еще в статистике говорят, что если условие выполняется, то мы имеем отсутствие автокорреляции. В-четвертых, случайное отклонение должно быть независимо от объясняющих переменных. Этому условию существенно выполняться автоматически, если переменные не являются случайными в данной модели.

Выше описаны базовые предпосылки, но в учебниках можно найти еще 3 дополнительные условия, к ним относятся то, что наша модель должна быть линейной относительно параметров. Если же мы работаем с множественной линейной регрессией тогда рассматриваются еще две предпосылки. В нашей модели отсутствует мультиколлинеарности. Простыми словами это значит, что между переменными отсутствует сильная линейная зависимость. И самое последнее условие заключается в том, что случайные отклонения имеют нормальное распределение. Это имеет весомое значение для проверки статистических гипотез и построения интервальных оценок.

Существует множество факторов, которые рассматривают при построении линейных регрессионных моделей, к ним относятся:

* объясняющие переменные не могут являться случайными величинами;
* число наблюдений существенно больше числа объясняющих переменных (числа параметров уравнения);
* правильно выбран вид уравнения и в него включены все необходимые переменные;

Если мы гарантируем выполнение четвертого условия, то мы обеспечиваем несмещенность оценки параметра b, а за параметр a отвечает выполнение первого условия. Если предпосылки второго и третьего условия были нарушены, т. е. дисперсия отклонений непостоянна и (или) значения связаны друг с другом, то свойства несмещенности и состоятельности сохраняются, но свойство эффективности – нет. Нарушение одного из условий Гаусса-Маркова приводит к нарушению эффективности оценок.

Итак, как уже известно первое условие нет необходимости проверять при наличии парной корреляции, а из первого условия вытекает и второе. Чтобы проверить третье условие достаточно найти коэффициент автокорреляцию по формуле . Таким образом, мы узнаем тесноту связи между уровнями ряда, сдвинутыми на 1 или более шаг и находим влияние предыдущих уровней ряда на последующие ряды. Самое интересное, что этот коэффициент со своими данными можно просчитать не только в Excel, но и на различных сайтах, таким образов я выяснила, что у моих данных отсутствует автокорреляция. Кроме того, переменные в данной модели не являются случайными, из этого следует выполнение 4 условия.

* 1. Нелинейные модели регрессии

В статистике рассматриваются два класса нелинейных регрессий. Первая это нелинейные относительно включенных в анализ объясняющих переменных, но линейные по оцениваемым параметрам и ко второй относятся нелинейные по оцениваемым параметрам. В первом классе строят следующие функции: полиномы различных степеней, равносторонняя гипербола, полулогарифмическая. Во-втором случае рассматриваются степенная, показательная, экспоненциальная функции.

Когда результатом анализа является нелинейная зависимость, то необходимо привести уравнение к линейному виду или, другими словами, линеаризовать. Это можно сделать либо с помощью подходящих преобразований исходных переменных, либо использовать методы оптимизации, но применяется это когда не удалось найти каких-либо адекватных преобразований. В математике самые популярные преобразования это замена переменных, комбинированные методы и логарифмирование. А оценка в моделях нелинейных требований МНК применяется не к исходным данным результативного признака, а к их преобразованным величинам.

Несмотря на то, что мои рассмотренные данные являются линейными, я строила функции вида: гипербола, экспонента и степенная.

Сравнивая все эти модели, можно сказать, что график степенной функции и линейной функции совпадают. Гипербола и экспонента функция повели себя несколько странно. Я сделала свертку итогов по исходным и получившимся данным, их я предоставлю в приложении (см. Приложение 2-5). Как можно заметить, что имеются сильные отличия. Я предпочла бы довериться простой линейной функции без преобразований.

* 1. Графическое представление данных

Первоначальный вид данных, приведенных к уравнению линейного вида представлен ниже(рис.3.7.1)

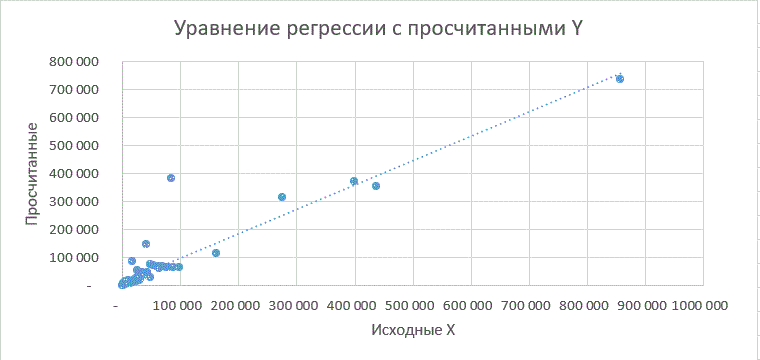


Рисунок 3.7.1 – Определение связи из визуального представления корреляционного облака.

Чтобы увидеть разницу между нелинейной и линейной функцией для уравнения, конечно же нагляднее представить из на одном графике. Таким образом, я визуализировала разницу между линейной и степенной функцией(рис.3.7.2).

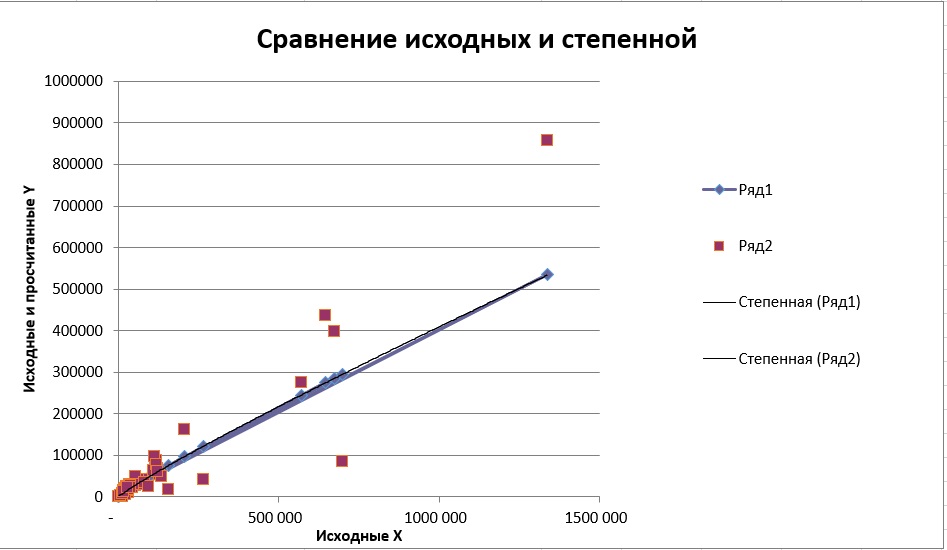


Рисунок 3.7.2 – Определение связи из визуального представления корреляционного облака.

Также, как и с функцией представленной выше, я поступила с экспонентой и линейно функцией(рис.3.7.3)

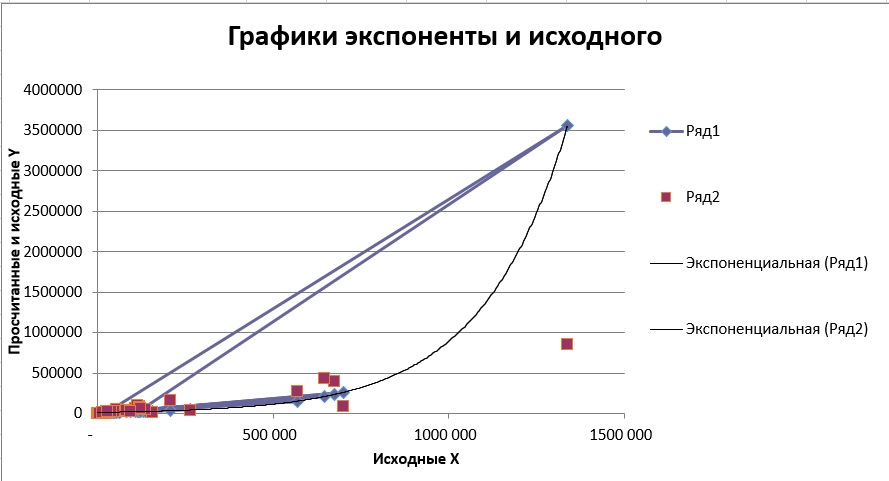


Рисунок 3.7.3 – Определение связи из визуального представления корреляционного облака.

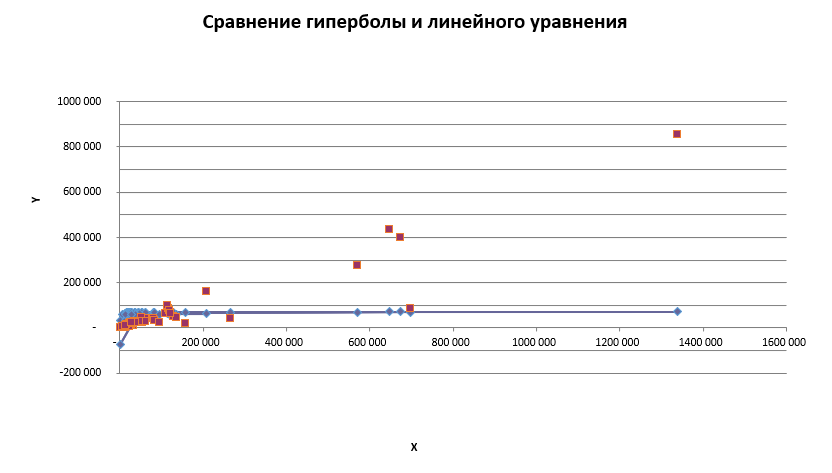


Рисунок 3.7.4 – Определение связи из визуального представления корреляционного облака.

Заключение

По итогам исследовательской работы были достигнуты все цели и выполнены все задачи. Я смогла продемонстрировать графическое представление линейных и нелинейных функций. Научилась находить регрессию и корреляцию, кроме того важным умением стала проверка значимости и верности текущих моделей.

В качестве вывода можно сказать, что данный вид анализа достаточно тяжелый, потому что, на первый взгляд нам может мнимо казаться, что у показателей уж точно есть зависимость, которая будет достоверной, но по окончанию анализа можно понять, что на самом деле эта зависимость вызвана автокорреляцией или другими факторами. Таким образом, в одну секунду весь ваш анализ может потерять смысл. Нужно тщательно продумывать свои цели и хорошо находить данные. Как и в любой работе нужно полагаться не только на то, что ты видишь (т.е. на графики), надо еще и делать точные расчеты, если есть возможность, то делать их при помощи формул, встроенных функций и в конце концов каких-либо сайтов с онлайн калькуляторами.

Что касается заключения экономического характера, то можно смело говорить, что зависимость данных показателей прямая, но я думаю, что в какой-то степени данные значения получись из-за следующей причины. Как мы знаем в первом квартале 2022 года возникли такие ситуации, которые послужили причиной эмиграции, но этот процесс проходил очень болезненно поскольку в некоторые страны не было возможности попасть на прямую, например, одним рейсом на самолете. Поэтому люди были вынуждены передвигаться по регионам страны, а потом уже выезжать за границу. Считаю, что это явление нужно рассматривать еще глубже.

Литература и используемые информационные источники

Литература:

Анализ данных: учебник для вузов / В. С. Мхитарян [и др.]; под редакцией В. С. Мхитаряна. — Москва: Издательство Юрайт, 2020. – ISBN 978-5-534-00616-2.

https://urait.ru/book/analiz-dannyh-489100

Интернет источники данных:

Данные о количестве размещенных лицах и оценке туристического потока:

<https://rosstat.gov.ru/about>

Приложение

Приложение 1

Прибыль различных банков на момент 01.01.2022, тыс. руб

|  |  |  |
| --- | --- | --- |
|  | **Численность размещенных лиц в коллективных средствах размещения(Y)** | **Оценка туристского потока(X)** |
| Белгородская область | 14 333 | 27 553 |
| Брянская область | 12 721 | 20 195 |
| Владимирская область | 40 995 | 265 626 |
| Воронежская область | 42 508 | 81 828 |
| Ивановская область | 19 001 | 23 843 |
| Калужская область | 42 237 | 62 264 |
| Костромская область | 17 072 | 19 301 |
| Курская область | 10 323 | 25 064 |
| Липецкая область | 20 044 | 26 908 |
| Московская область | 398 281 | 673 985 |
| Орловская область | 10 837 | 31 714 |
| Рязанская область | 26 436 | 54 851 |
| Смоленская область | 11 158 | 17 022 |
| Тамбовская область | 8 620 | 13 395 |
| Тверская область | 52 744 | 129 519 |
| Тульская область | 28 361 | 45 637 |
| Ярославская область | 75 260 | 114 986 |
| г.Москва | 857 353 | 1 338 665 |
| Республика Карелия | 24 522 | 26 099 |
| Республика Коми | 19 759 | 27 720 |
| Архангельская область | 23 761 | 45 105 |
| Ненецкий автономный округ | 376 | 714 |
| Архангельская область (кроме Ненецкого автономного округа) | 23 385 | 44 391 |
| Вологодская область | 63 544 | 108 601 |
| Калининградская область | 48 300 | 53 044 |
| Ленинградская область | 84 247 | 698 385 |
| Мурманская область | 30 212 | 38 711 |
| Новгородская область | 15 865 | 28 119 |
| Псковская область | 28 946 | 62 418 |
| г.Санкт-Петербург | 274 930 | 569 559 |
| Республика Адыгея | 8 367 | 19 230 |
| Республика Калмыкия | 1 468 | 1 632 |
| Республика Крым | 87 902 | 117 609 |
| Краснодарский край | 436 894 | 646 352 |
| Астраханская область | 14 571 | 16 383 |
| Волгоградская область | 29 374 | 34 144 |
| Ростовская область | 34 223 | 82 250 |
| г.Севастополь | 3 799 | 13 241 |
| Республика Дагестан | 5 416 | 22 640 |
| Республика Ингушетия | 1 389 | 13 918 |
| Кабардино-Балкарская Республика | 9 109 | 13 420 |
| Карачаево-Черкесская Республика | 17 653 | 157 026 |
| Республика Северная Осетия - Алания | 9 211 | 14 489 |
| Чеченская Республика | 4 444 | 6 473 |
| Ставропольский край | 68 552 | 121 085 |
| Республика Башкортостан | 97 347 | 114 294 |
| Республика Марий Эл | 14 090 | 15 987 |
| Республика Мордовия | 6 958 | 12 893 |
| Республика Татарстан | 161 112 | 207 263 |
| Удмуртская Республика | 29 171 | 47 547 |
| Чувашская Республика | 18 737 | 32 604 |
| Пермский край | 48 196 | 135 910 |
| Кировская область | 26 428 | 38 095 |
| Нижегородская область | 79 021 | 119 046 |
| Оренбургская область | 22 379 | 38 720 |
| Пензенская область | 13 174 | 14 609 |
| Самарская область | 61 911 | 121 501 |
| Саратовская область | 25 780 | 94 081 |
| Ульяновская область | 22 799 | 28 803 |

Приложение 2

Расчет ширины интервала по формуле Стержеса

|  |
| --- |
| *Регрессионная статистика* |
| Множественный R | 0,931642208 |
| R-квадрат | 0,867957205 |
| Нормированный R-квадрат | 0,865640664 |
| Стандартная ошибка | 26790,56057 |
| Наблюдения | 59 |
|  |  |
| Дисперсионный анализ |  |  |  |  |
|  | *df* | *SS* | *MS* | *F* | *Значимость F* |
| Регрессия | 1 | 2,69E+11 | 2,69E+11 | 374,6782 | 9,81985E-27 |
| Остаток | 57 | 4,09E+10 | 7,18E+08 |  |  |
| Итого | 58 | 3,1E+11 |  |  |  |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | *Коэффициенты* | *Стандартная ошибка* | *t-статистика* | *P-Значение* | *Нижние 95%* | *Верхние 95%* | *Нижние 95,0%* |
| Y-пересечение | 1053,927335 | 3940,889 | 0,267434 | 0,7901 | -6837,566742 | 8945,421 | -6837,57 |
| Переменная X 1 | 0,300315868 | 0,015515 | 19,35661 | 9,82E-27 | 0,269247811 | 0,331384 | 0,269248 |

|  |  |  |  |
| --- | --- | --- | --- |
| *Наблюдение* | *Предсказанное Y* | *Остатки* | *Стандартные остатки* |
| 1 | 9328,530433 | 857,8623 | 0,032301 |
| 2 | 7118,80628 | 2184,193 | 0,082241 |
| 3 | 80825,62996 | -56028,2 | -2,10961 |
| 4 | 25628,17414 | -1,57637 | -5,9E-05 |
| 5 | 8214,358565 | 4530,149 | 0,170572 |
| 6 | 19752,79451 | 5725,292 | 0,215572 |
| 7 | 6850,323894 | 4837,07 | 0,182128 |
| 8 | 8581,044239 | -592,175 | -0,0223 |
| 9 | 9134,826698 | 4181,256 | 0,157435 |
| 10 | 203462,3173 | 17131,74 | 0,645054 |
| 11 | 10578,14476 | -2307,6 | -0,08689 |
| 12 | 17526,55298 | -707,585 | -0,02664 |
| 13 | 6165,904032 | 2280,554 | 0,085869 |
| 14 | 5076,658381 | 1978,948 | 0,074513 |
| 15 | 39950,53818 | -8714,5 | -0,32812 |
| 16 | 14759,44258 | 3114,446 | 0,117267 |
| 17 | 35586,04768 | 7989,001 | 0,300807 |
| 18 | 403076,2681 | 69094,21 | 2,601576 |
| 19 | 8891,871162 | 6878,204 | 0,258982 |
| 20 | 9378,683183 | 3781,216 | 0,142373 |
| 21 | 14599,67454 | 753,3642 | 0,028366 |
| 22 | 1268,352865 | 1269,453 | 0,047798 |
| 23 | 14385,24901 | 761,7377 | 0,028681 |
| 24 | 33668,53086 | 3486,023 | 0,131258 |
| 25 | 16983,88221 | 11816,79 | 0,444933 |
| 26 | 210790,0245 | -162290 | -6,11064 |
| 27 | 12679,45488 | 6208,802 | 0,233777 |
| 28 | 9498,509214 | 1527,436 | 0,057512 |
| 29 | 19799,04315 | -1604,57 | -0,06042 |
| 30 | 172101,5325 | -19105,2 | -0,71936 |
| 31 | 6829,001468 | 87,95863 | 0,003312 |
| 32 | 1544,042831 | 1592,191 | 0,05995 |
| 33 | 36373,7762 | 14129,23 | 0,532002 |
| 34 | 195163,6889 | 46590,71 | 1,75426 |
| 35 | 5974,002193 | 4342,817 | 0,163518 |
| 36 | 11307,91232 | 7121,111 | 0,268128 |
| 37 | 25754,90744 | -4668,58 | -0,17578 |
| 38 | 5030,409737 | -616,763 | -0,02322 |
| 39 | 7853,078576 | -2553,3 | -0,09614 |
| 40 | 5233,723579 | -2140,78 | -0,08061 |
| 41 | 5084,166277 | 2239,418 | 0,08432 |
| 42 | 48211,32675 | -36205,5 | -1,36323 |
| 43 | 5405,20394 | 1974,277 | 0,074337 |
| 44 | 2997,871946 | 1769,242 | 0,066617 |
| 45 | 37417,67415 | 2481,318 | 0,093428 |
| 46 | 35378,2291 | 20300,73 | 0,764375 |
| 47 | 5855,077109 | 4198,149 | 0,158071 |
| 48 | 4925,899815 | 1218,913 | 0,045895 |
| 49 | 63298,29499 | 27324,58 | 1,028841 |
| 50 | 15333,04589 | 2984,731 | 0,112383 |
| 51 | 10845,42588 | 1754,407 | 0,066058 |
| 52 | 41869,85689 | -13126,2 | -0,49423 |
| 53 | 12494,46031 | 4320,123 | 0,162664 |
| 54 | 36805,3301 | 8830,787 | 0,332502 |
| 55 | 12682,15773 | 1913,53 | 0,072049 |
| 56 | 5441,241844 | 4110,006 | 0,154752 |
| 57 | 37542,60556 | -1282,95 | -0,04831 |
| 58 | 29307,94447 | -12848,5 | -0,48378 |

Приложение 3

Расчет показателей для графического представления данных

|  |
| --- |
| *Регрессионная статистика* |
| Множественный R | 1 |
| R-квадрат | 1 |
| Нормированный R-квадрат | 1 |
| Стандартная ошибка | 3,67E-11 |
| Наблюдения | 59 |
|  |  |
| Дисперсионный анализ |  |  |  |  |
|  | *df* | *SS* | *MS* | *F* | *Значимость F* |
| Регрессия | 1 | 8,95E+11 | 8,95E+11 | 6,64E+32 | 0 |
| Остаток | 57 | 7,69E-20 | 1,35E-21 |  |  |
| Итого | 58 | 8,95E+11 |  |  |  |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | *Коэффициенты* | *Стандартная ошибка* | *t-статистика* | *P-Значение* | *Нижние 95%* | *Верхние 95%* | *Нижние 95,0%* |
| Y-пересечение | 2331,754 | 5,4E-12 | 4,32E+14 | 0 | 2331,754 | 2331,754 | 2331,754 |
| Переменная X 1 | 0,548011 | 2,13E-17 | 2,58E+16 | 0 | 0,548011 | 0,548011 | 0,548011 |

|  |  |  |  |
| --- | --- | --- | --- |
| *Наблюдение* | *Предсказанное Y* | *Остатки* | *Стандартные остатки* |
| 1 | 17431,1 | 2,91E-11 | 0,848238 |
| 2 | 13398,83 | 2,73E-11 | 0,795224 |
| 3 | 147897,7 | 0 | 0 |
| 4 | 47174,38 | 2,18E-11 | 0,636179 |
| 5 | 15397,98 | 2,73E-11 | 0,795224 |
| 6 | 36453,1 | 2,18E-11 | 0,636179 |
| 7 | 12908,91 | 2,73E-11 | 0,795224 |
| 8 | 16067,1 | 2,73E-11 | 0,795224 |
| 9 | 17077,63 | 2,55E-11 | 0,742209 |
| 10 | 371682,8 | -5,8E-11 | -1,69648 |
| 11 | 19711,37 | 2,55E-11 | 0,742209 |
| 12 | 32390,7 | 2,18E-11 | 0,636179 |
| 13 | 11659,99 | 2,73E-11 | 0,795224 |
| 14 | 9672,359 | 2,91E-11 | 0,848238 |
| 15 | 73309,57 | 1,46E-11 | 0,424119 |
| 16 | 27341,32 | 2,18E-11 | 0,636179 |
| 17 | 65345,33 | 1,46E-11 | 0,424119 |
| 18 | 735934,7 | -1,2E-10 | -3,39295 |
| 19 | 16634,29 | 2,91E-11 | 0,848238 |
| 20 | 17522,61 | 2,18E-11 | 0,636179 |
| 21 | 27049,78 | 2,55E-11 | 0,742209 |
| 22 | 2723,033 | 2,91E-11 | 0,848238 |
| 23 | 26658,5 | 2,55E-11 | 0,742209 |
| 24 | 61846,28 | 1,46E-11 | 0,424119 |
| 25 | 31400,44 | 2,18E-11 | 0,636179 |
| 26 | 385054,3 | -5,8E-11 | -1,69648 |
| 27 | 23545,8 | 2,55E-11 | 0,742209 |
| 28 | 17741,27 | 2,91E-11 | 0,848238 |
| 29 | 36537,49 | 2,18E-11 | 0,636179 |
| 30 | 314456,3 | -5,8E-11 | -1,69648 |
| 31 | 12870 | 2,73E-11 | 0,795224 |
| 32 | 3226,107 | 2,86E-11 | 0,834985 |
| 33 | 66782,76 | 1,46E-11 | 0,424119 |
| 34 | 356539,6 | -1,2E-10 | -3,39295 |
| 35 | 11309,81 | 2,73E-11 | 0,795224 |
| 36 | 21043,04 | 2,18E-11 | 0,636179 |
| 37 | 47405,64 | 2,18E-11 | 0,636179 |
| 38 | 9587,965 | 2,73E-11 | 0,795224 |
| 39 | 14738,72 | 2,73E-11 | 0,795224 |
| 40 | 9958,968 | 2,91E-11 | 0,848238 |
| 41 | 9686,059 | 2,73E-11 | 0,795224 |
| 42 | 88383,7 | 1,46E-11 | 0,424119 |
| 43 | 10271,88 | 2,91E-11 | 0,848238 |
| 44 | 5879,028 | 2,82E-11 | 0,821731 |
| 45 | 68687,64 | 2,91E-11 | 0,848238 |
| 46 | 64966,1 | 2,18E-11 | 0,636179 |
| 47 | 11092,8 | 2,73E-11 | 0,795224 |
| 48 | 9397,257 | 2,73E-11 | 0,795224 |
| 49 | 115914,1 | 0 | 0 |
| 50 | 28388,02 | 2,55E-11 | 0,742209 |
| 51 | 20199,1 | 2,55E-11 | 0,742209 |
| 52 | 76811,91 | 1,46E-11 | 0,424119 |
| 53 | 23208,23 | 2,18E-11 | 0,636179 |
| 54 | 67570,25 | 0 | 0 |
| 55 | 23550,73 | 2,55E-11 | 0,742209 |
| 56 | 10337,64 | 2,55E-11 | 0,742209 |
| 57 | 68915,62 | 1,46E-11 | 0,424119 |
| 58 | 53889,16 | 2,18E-11 | 0,636179 |

Приложение 4

Расчет основных числовых характеристик одномерных данных

|  |
| --- |
| *Регрессионная статистика* |
| Множественный R | 0,925074 |
| R-квадрат | 0,855762 |
| Нормированный R-квадрат | 0,853231 |
| Стандартная ошибка | 35429,81 |
| Наблюдения | 59 |
|  |  |
| Дисперсионный анализ |  |  |  |  |
|  | *df* | *SS* | *MS* | *F* | *Значимость F* |
| Регрессия | 1 | 4,25E+11 | 4,25E+11 | 338,1791 | 1,23E-25 |
| Остаток | 57 | 7,16E+10 | 1,26E+09 |  |  |
| Итого | 58 | 4,96E+11 |  |  |  |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | *Коэффициенты* | *Стандартная ошибка* | *t-статистика* | *P-Значение* | *Нижние 95%* | *Верхние 95%* | *Нижние 95,0%* |
| Y-пересечение | 13920,83 | 5101,346 | 2,728854 | 0,008437 | 3705,559 | 24136,1 | 3705,559 |
| Переменная X 1 | 0,64146 | 0,034882 | 18,38965 | 1,23E-25 | 0,571611 | 0,711309 | 0,571611 |

|  |  |  |  |
| --- | --- | --- | --- |
| *Наблюдение* | *Предсказанное Y* | *Остатки* | *Стандартные остатки* |
| 1 | 23114,87 | -7730,44 | -0,2201 |
| 2 | 22080,84 | -10497,8 | -0,29889 |
| 3 | 40217,47 | 81710,25 | 2,3264 |
| 4 | 41188 | 397,7262 | 0,011324 |
| 5 | 26109,2 | -12628,8 | -0,35956 |
| 6 | 41014,16 | -8614,64 | -0,24527 |
| 7 | 24871,83 | -13758,2 | -0,39171 |
| 8 | 20542,62 | -6432,91 | -0,18315 |
| 9 | 26778,25 | -11723,2 | -0,33377 |
| 10 | 269402,1 | 16042,26 | 0,456745 |
| 11 | 20872,33 | -3378,59 | -0,09619 |
| 12 | 30878,46 | -2021,82 | -0,05756 |
| 13 | 21078,24 | -11169,8 | -0,31802 |
| 14 | 19450,21 | -11489,8 | -0,32713 |
| 15 | 47753,98 | 15506,71 | 0,441497 |
| 16 | 32113,27 | -7719,26 | -0,21978 |
| 17 | 62197,09 | -5453,84 | -0,15528 |
| 18 | 563878,3 | -29587,6 | -0,8424 |
| 19 | 29650,7 | -15009,7 | -0,42735 |
| 20 | 26595,43 | -11125,8 | -0,31677 |
| 21 | 29162,55 | -5028,46 | -0,14317 |
| 22 | 14162,02 | -13615,3 | -0,38765 |
| 23 | 28921,36 | -5136,52 | -0,14624 |
| 24 | 54681,75 | -824,013 | -0,02346 |
| 25 | 44903,33 | -16916,4 | -0,48163 |
| 26 | 67961,89 | 226908,3 | 6,460381 |
| 27 | 33300,61 | -12312,2 | -0,35054 |
| 28 | 24097,59 | -8424,7 | -0,23986 |
| 29 | 32488,52 | -15,8043 | -0,00045 |
| 30 | 190277,4 | 54477,19 | 1,551038 |
| 31 | 19287,92 | -8211,61 | -0,2338 |
| 32 | 14862,49 | -13699 | -0,39003 |
| 33 | 70306,42 | -12381,8 | -0,35253 |
| 34 | 294170,7 | -19436,9 | -0,5534 |
| 35 | 23267,54 | -13699,4 | -0,39004 |
| 36 | 32763,07 | -14048,7 | -0,39999 |
| 37 | 35873,5 | 5908,099 | 0,168211 |
| 38 | 16357,73 | -8480,97 | -0,24146 |
| 39 | 17394,97 | -4537,28 | -0,12918 |
| 40 | 14811,81 | -6567,94 | -0,187 |
| 41 | 19763,88 | -11789,9 | -0,33567 |
| 42 | 25244,52 | 50184,96 | 1,428833 |
| 43 | 19829,31 | -11277 | -0,32107 |
| 44 | 16771,47 | -12675,1 | -0,36088 |
| 45 | 57894,18 | 1592,406 | 0,045338 |
| 46 | 76365,01 | -19933,8 | -0,56754 |
| 47 | 22959 | -13602,4 | -0,38728 |
| 48 | 18384,1 | -10696,7 | -0,30455 |
| 49 | 117267,7 | -20067,1 | -0,57134 |
| 50 | 32632,85 | -7307,83 | -0,20806 |
| 51 | 25939,86 | -7998,17 | -0,22772 |
| 52 | 44836,62 | 21269,76 | 0,605578 |
| 53 | 30873,33 | -10190,2 | -0,29013 |
| 54 | 64609,62 | -6038,82 | -0,17193 |
| 55 | 28276,06 | -7283,14 | -0,20736 |
| 56 | 22371,42 | -13754,4 | -0,39161 |
| 57 | 53634,24 | 6039,016 | 0,171939 |
| 58 | 30457,66 | 16781,82 | 0,477801 |

Приложение 5

Исходные временные данные, в руб. на душу населения

|  |
| --- |
| *Регрессионная статистика* |
| Множественный R | 0,838947 |
| R-квадрат | 0,703832 |
| Нормированный R-квадрат | 0,698636 |
| Стандартная ошибка | 253816,8 |
| Наблюдения | 59 |
|  |  |
| Дисперсионный анализ |  |  |  |  |
|  | *df* | *SS* | *MS* | *F* | *Значимость F* |
| Регрессия | 1 | 8,72662E+12 | 8,73E+12 | 135,4582 | 1,08E-16 |
| Остаток | 57 | 3,67211E+12 | 6,44E+10 |  |  |
| Итого | 58 | 1,23987E+13 |  |  |  |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | *Коэффициенты* | *Стандартная ошибка* | *t-статистика* | *P-Значение* | *Нижние 95%* | *Верхние 95%* | *Нижние 95,0%* |
| Y-пересечение | -88814 | 36545,70682 | -2,43022 | 0,018261 | -161996 | -15632,5 | -161996 |
| Переменная X 1 | 2,908374 | 0,249889288 | 11,63865 | 1,08E-16 | 2,407979 | 3,408769 | 2,407979 |

|  |  |  |  |
| --- | --- | --- | --- |
| *Наблюдение* | *Предсказанное Y* | *Остатки* | *Стандартные остатки* |
| 1 | -47128,3 | 63998,0817 | 0,254345 |
| 2 | -51816,6 | 68187,14966 | 0,270993 |
| 3 | 30414,78 | 14174,13934 | 0,056332 |
| 4 | 34815,15 | -13760,76582 | -0,05469 |
| 5 | -33552 | 50168,1993 | 0,199381 |
| 6 | 34026,98 | -14588,85211 | -0,05798 |
| 7 | -39162,3 | 55473,1735 | 0,220465 |
| 8 | -58790,9 | 75490,10444 | 0,300017 |
| 9 | -30518,6 | 47343,99379 | 0,188157 |
| 10 | 1069536 | -833340,9117 | -3,31191 |
| 11 | -57296 | 74454,78496 | 0,295903 |
| 12 | -11928,2 | 30786,89294 | 0,122355 |
| 13 | -56362,4 | 72522,23915 | 0,288222 |
| 14 | -63743,8 | 79666,16547 | 0,316614 |
| 15 | 64585,26 | -39005,21334 | -0,15502 |
| 16 | -6329,62 | 24492,04024 | 0,097338 |
| 17 | 130070,2 | -105963,7442 | -0,42113 |
| 18 | 2404689 | 1158073,797 | 4,602486 |
| 19 | -17494,9 | 34264,81551 | 0,136177 |
| 20 | -31347,5 | 48228,75062 | 0,191674 |
| 21 | -19708,1 | 37831,15508 | 0,150351 |
| 22 | -87720,5 | 102839,449 | 0,408711 |
| 23 | -20801,7 | 38871,95222 | 0,154487 |
| 24 | 95995,7 | -72509,51131 | -0,28817 |
| 25 | 51660,45 | -32940,40775 | -0,13091 |
| 26 | 156207,8 | 104727,9748 | 0,416216 |
| 27 | -946,219 | 18602,2688 | 0,07393 |
| 28 | -42672,7 | 59581,48002 | 0,236792 |
| 29 | -4628,22 | 24078,57041 | 0,095694 |
| 30 | 710785,2 | -556573,4117 | -2,21197 |
| 31 | -64479,6 | 80785,84099 | 0,321064 |
| 32 | -84544,5 | 99720,27457 | 0,396314 |
| 33 | 166837,9 | -142471,8709 | -0,56622 |
| 34 | 1181837 | -970840,2025 | -3,85837 |
| 35 | -46436,1 | 62553,85626 | 0,248605 |
| 36 | -3383,44 | 20713,3337 | 0,08232 |
| 37 | 10719,27 | 10371,41702 | 0,041219 |
| 38 | -77765,1 | 93677,42844 | 0,372298 |
| 39 | -73062,3 | 89597,05019 | 0,356082 |
| 40 | -84774,3 | 100730,651 | 0,40033 |
| 41 | -62321,6 | 78245,59584 | 0,310968 |
| 42 | -37472,5 | 66092,69947 | 0,26267 |
| 43 | -62025 | 78018,59079 | 0,310066 |
| 44 | -75889,2 | 91367,8702 | 0,36312 |
| 45 | 110560,8 | -85846,58869 | -0,34118 |
| 46 | 194307,5 | -170269,0073 | -0,67669 |
| 47 | -47835 | 63926,74732 | 0,254061 |
| 48 | -68577,5 | 84467,2839 | 0,335695 |
| 49 | 379759,9 | -344624,4393 | -1,36963 |
| 50 | -3973,84 | 22278,43788 | 0,08854 |
| 51 | -34319,8 | 51541,09163 | 0,204838 |
| 52 | 51357,98 | -25101,71067 | -0,09976 |
| 53 | -11951,5 | 29563,20837 | 0,117492 |
| 54 | 141008,6 | -116499,2337 | -0,463 |
| 55 | -23727,5 | 41384,2103 | 0,164472 |
| 56 | -50499,1 | 66500,54244 | 0,26429 |
| 57 | 91246,32 | -66490,06807 | -0,26425 |
| 58 | -13836,1 | 35970,53166 | 0,142956 |