**Министерство науки и высшего образования Российской Федерации**

Федеральное государственное бюджетное образовательное учреждение

высшего образования

**«ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ УПРАВЛЕНИЯ»**

Изображение выглядит как текст, гаечный ключ, инструмент

Автоматически созданное описание

|  |  |
| --- | --- |
| Институт | Информационных систем |
| Кафедра | Математических методов в экономике и управлении |

**Семестровая работа**

|  |  |  |  |
| --- | --- | --- | --- |
| по дисциплине | Методы и модели эконометрики | | |
|  | На тему | | |
| «Исследование развития розничной торговли в разрезе субъектов Российской Федерации на основе методов и моделей эконометрики» | | | |
|  |  | | |
| Направление подготовки | 01.03.02 |  | Прикладная математика и информатика | |
| (код) |  | (наименование) | |
| Образовательная программа | Прикладная математика и информатика | | |
| (название образовательной программы) | | |
| Обучающийся | Рыбалко Никита Александрович | | |
|
|
|
| (Фамилия, имя, отчество) | | |
| III курс, ПМИ 3-1 |
| (курс, номер группы) |
| Руководитель работы | Заведующая кафедрой, к.э.н. Писарева О. М. | | |
| (ученая степень, звание, Фамилия и Инициалы) | | |

|  |  |
| --- | --- |
| Оценка: |  |
| Подпись руководителя: |  |

Москва – 2023

**СОДЕРЖАНИЕ**

[ВВЕДЕНИЕ 3](#_Toc153894692)

[1. ПОСТАНОВКА ЗАДАЧИ ИССЛЕДОВАНИЯ И СБОР ДАННЫХ 5](#_Toc154409852)

[2.1. Визуальный анализ 11](#_Toc154409853)

[2.2. Поиск аномальных значений 14](#_Toc154409854)

[2.3. Корреляционный анализ 16](#_Toc154409855)

[3. МНОЖЕСТВЕННЫЙ ЛИНЕЙНЫЙ РЕГРЕССИОННЫЙ АНАЛИЗ 19](#_Toc154409858)

[4. ДИСПЕРСИОННЫЙ АНАЛИЗ 21](#_Toc154409859)

[5. ПРОВЕРКА ДОПОЛНИТЕЛЬНЫХ ИНФОРМАЦИОННЫХ И ПРОГНОСТИЧЕСКИХ СВОЙСТВ УРАВНЕНИЙ РЕГРЕССИИ 22](#_Toc154409860)

[5.1. Оценка соответствия остатков нормальному закону 22](#_Toc154409861)

[5.2. Проверка несмещённости МНК-оценок 23](#_Toc154409862)

[5.3. Проверка гетероскедастичности остатков 23](#_Toc154409863)

[5.4. Оценка автокорреляции остатков модели 24](#_Toc154409864)

[5.5. Устранение гетероскедастичности остатков 25](#_Toc154409865)

[6. ВЫЯВЛЕНИЕ ЛУЧШИХ МОДЕЛЕЙ 26](#_Toc154409866)

[7. ПРОГНОЗИРОВАНИЕ 27](#_Toc154409867)

[ЗАКЛЮЧЕНИЕ 29](#_Toc154409868)

[СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ 30](#_Toc154409869)

[ПРИЛОЖЕНИЕ А - Исходные данные 32](#_Toc154409879)

[ПРИЛОЖЕНИЕ Б – Аномальные значения 34](#_Toc154409880)

[ПРИЛОЖЕНИЕ В – Регрессионный анализ 36](#_Toc154409880)

[ПРИЛОЖЕНИЕ Г – Графики разброса остатков 37](#_Toc154409881)

**ВВЕДЕНИЕ**

Цель работы:

– закрепление теоретических знаний, полученных в ходе изучения учебных дисциплин «Эконометрика», «Теория вероятностей и математическая статистика» и др., связанных с вопросами обоснования необходимости и допустимости применения факторных регрессионных моделей, а также отработки методики их построения и использования с целью проверки гипотез относительно возможных свойств и закономерностей проявления связей между показателями социально-экономических систем микро-, мезо- или макроуровня;

– приобретение и закрепление практических навыков в области предмодельной обработки данных, спецификации, калибровки, валидации факторных эконометрических моделей;

– проведение экспериментальных расчетов по обоснованию, оценке и проведению процедуры верификации факторной регрессионной модели, приобретение начального опыта по идентификации и снижению эффектов ошибок спецификации факторных линейных моделей;

– получение начального опыта проектирования технологий информационной поддержки прогнозно-аналитической деятельности в рамках интегрированной информационной среды компании;

– знакомство с базовыми возможностями и получение навыков работы с доступными версиями специальных статистических пакетов общего назначения (например, Gretl и т. п.), реализующих вычислительные алгоритмы обработки данных на основе методов многомерного, корреляционного, регрессионного и дисперсионного анализа, а также методы разведочного анализа, статистического тестирования и др.;

– получение элементарных представлений о процессах информационно-программной интеграции средств обработки и хранения данных в ходе реализации прогнозно-аналитических процедур в системах поддержки принятия управленческих решений регулятора сферы деятельности.

Для достижения поставленных целей предусматривается постановка задачи исследования следующих основных этапов:

1) определение необходимой информации, сбор и построение массива данных, содержащего значения макроэкономических показателей на определенных период времени;

2) анализ статистической информации, предмодельная обработка данных и приведение информации пригодной к обработке;

3) разбиение массива данных на кластеры для дальнейшего построения факторных регрессионных моделей (если это необходимо) и их спецификаций. Характеристики качества этой модели;

4) оценка параметров конкурирующих моделей, их верификация и оценка прогностических свойств;

5) прогнозирование на основе факторной регрессионной модели.

Для выполнения проектной работы использовались пакеты обработки статистических данных – MS Excel: пакет «Анализ данных», Gretl и язык “Python”.

1. **ПОСТАНОВКА ЗАДАЧИ ИССЛЕДОВАНИЯ И СБОР ДАННЫХ**

Для проведения эконометрического моделирования была сформулирована задача построения факторной регрессионной модели развития розничной торговли в разрезе субъектов РФ. Основной целью моделирования является получение количественных оценок параметров модели и проверка статистической значимости этих параметров, чтобы сделать выводы о влиянии различных факторов на изучаемые экономические явления. Такая модель может быть использована для прогнозирования будущих значений зависимой переменной, а также для тестирования экономических гипотез.

Основные термины:

Оборот розничной торговли - выручка от продажи товаров населению для личного потребления или использования в домашнем хозяйстве за наличный расчет или оплаченных по кредитным карточкам, расчетным чекам банков, по перечислениям со счетов вкладчиков, по поручению физического лица без открытия счета, посредством платежных карт. Оборот розничной торговли приводится в фактических продажных ценах, включающих торговую наценку, налог на добавленную стоимость и аналогичные обязательные платежи;

ВРП региона на душу населения (Валовый региональный продукт) - это сумма стоимости всех финальных товаров и услуг, произведенных в данном регионе за определенный период времени (обычно за год). Он служит важной макроэкономической характеристикой, позволяющей измерить экономическую активность и уровень производства в данной территориальной области. ВРП региона и оборот розничной торговли взаимосвязаны, поскольку розничная торговля представляет собой значительную часть экономической активности, которая влияет на общую экономическую производительность региона. Ниже приведены несколько прямых и косвенных связей между ВРП региона и оборотом розничной торговли:

1. Вклад в ВРП: Розничная торговля вносит значительный вклад в общий ВРП региона. Чем больше оборот розничной торговли, тем больше вклад в ВРП региона;

2. Работа и заработная плата: Розничная торговля создает множество рабочих мест в регионе, что способствует увеличению занятости и заработной платы. Это влияет на общий уровень доходов населения, что, в свою очередь, может стимулировать потребительский спрос и, следовательно, ВРП региона;

3. Потребительский спрос: Розничная торговля предоставляет доступ к широкому ассортименту товаров и услуг для потребителей. Потребительский спрос является одним из ключевых факторов, влияющих на экономический рост региона, и сильный оборот розничной торговли может способствовать увеличению потребительского спроса и ВРП региона;

4. Инвестиции: Успех розничной торговли может привлечь инвестиции в регион. Инвестиции в развитие розничной торговли, включая открытие новых магазинов и торговых центров, могут способствовать росту ВРП региона.

Среднегодовая численность населения - это средняя арифметическая из численностей населения на начало и конец соответствующего года. Это статистический показатель, среднюю численность населения региона и включает всех резидентов, независимо от их возраста, пола или гражданства. Среднегодовая численность населения региона является важным фактором, определяющим социально-экономическое развитие и потребительский спрос, а также оказывает влияние на инфраструктуру, услуги, занятость и другие аспекты жизни в данном регионе численность населения региона может оказывать значительное влияние на оборот розничной торговли в данном регионе;

Причины:

1. Потребительский спрос: Большая среднегодовая численность населения обычно приводит к увеличению потребительского спроса. Большее количество людей требует больше товаров и услуг, что стимулирует развитие розничной торговли. Таким образом, регионы с более многочисленным населением часто имеют более высокий оборот розничной торговли;

2. Разнообразие товаров и услуг: Большая среднегодовая численность населения позволяет развитую розничную торговлю предлагать более разнообразный ассортимент товаров и услуг. Это важно для удовлетворения различных потребностей и предпочтений населения. Разнообразие товаров и услуг повышает привлекательность региона для покупателей и способствует росту оборота в розничной торговле;

3. Работа и заработная плата: Большая среднегодовая численность населения предоставляет больше рабочих мест, связанных с розничной торговлей. Это включает работу в магазинах, ресторанах, торговых центрах и других торговых предприятиях. Увеличение численности населения способствует развитию розничной торговли и увеличению ее оборота;

4. Экономический рост: Рост среднегодовой численности населения часто связан с экономическим ростом региона. Это может включать больше инвестиций, увеличение заработной платы и развитие других отраслей. В результате регионы с большей численностью населения могут испытывать больший экономический рост, что в свою очередь стимулирует рост розничной торговли.

Средний доход на душу населения — показатель экономического благосостояния [страны](https://ru.wikipedia.org/wiki/%D0%A1%D1%82%D1%80%D0%B0%D0%BD%D0%B0), измеряющий среднестатистический [доход](https://ru.wikipedia.org/wiki/%D0%94%D0%BE%D1%85%D0%BE%D0%B4), получаемый отдельно взятым лицом в стране за год. Вычисляется из [национального дохода](https://ru.wikipedia.org/wiki/%D0%9D%D0%B0%D1%86%D0%B8%D0%BE%D0%BD%D0%B0%D0%BB%D1%8C%D0%BD%D1%8B%D0%B9_%D0%B4%D0%BE%D1%85%D0%BE%D0%B4), поделённого на численность [населения](https://ru.wikipedia.org/wiki/%D0%9D%D0%B0%D1%81%D0%B5%D0%BB%D0%B5%D0%BD%D0%B8%D0%B5). Чем выше среднедушевые денежные доходы населения, тем больше люди могут себе позволить потратить на покупки, что влияет на увеличение оборота розничной торговли. Наоборот, снижение доходов может привести к уменьшению потребительского спроса и снижению оборота розничной торговли.

Обеспеченность населения торговыми площадями современных форматов означает наличие современных розничных торговых объектов, таких как супермаркеты, торговые центры, магазины фешн-ритейла и т.д., которые предлагают широкий ассортимент товаров и услуг, а также создают комфортные условия для покупателей.

Наличие таких торговых площадей может положительно влиять на оборот розничной торговли, поскольку:

1. Повышает конкуренцию, что стимулирует розничные предприятия улучшать качество товаров и услуг, а также предлагать более выгодные цены;

2. Создает удобные условия для покупателей, что способствует увеличению количества покупок;

3. Привлекает большее количество покупателей за счет более разнообразного и широкого ассортимента товаров и услуг.

Уровень безработицы - это процентное соотношение числа безработных людей к общему числу активного населения в стране или регионе. Он представляет собой показатель, отражающий масштабы и динамику безработицы в регионе. Уровень безработицы может иметь влияние на оборот розничной торговли следующим образом:

1. Потребительский спрос: При высоком уровне безработицы люди часто ограничивают свои траты на товары и услуги, что приводит к снижению оборота розничной торговли. Низкий уровень безработицы, напротив, способствует росту потребительского спроса и увеличению объема продаж;

2. Покупательская сила: Высокий уровень безработицы может ослабить покупательскую способность населения, так как люди имеют ограниченные средства для траты. Это может снизить общий оборот розничной торговли и привести к ухудшению финансовых показателей торговых предприятий. В то же время, низкая безработица повышает покупательскую силу и стимулирует рост торговли;

3. Предложение рабочей силы: Уровень безработицы может влиять на предложение рабочей силы в розничном секторе. По мере снижения безработицы, предложение рабочей силы может уменьшаться, что может повлечь за собой рост зарплат и стоимости рабочей силы. Это может отразиться на стоимости товаров и услуг, которые могут оказаться выше, что может повлиять на спрос и оборот розничной торговли.

Для отбора данных был использован официальный сайт Федеральной службы государственной статистики (https://rosstat.gov.ru) предоставляющий статистику по субъектам РФ. В качестве переменных были выбраны:

В качестве зависимой переменной (Y) был выбран оборот розничной торговли на душу населения по субъектам РФ.

В качестве независимых переменных были выбраны:

X1 – ВРП региона на душу населения;

X2 – Среднегодовая численность населения региона;

X3 – Средний доход на душу населения в регионе;

X4 – Обеспеченность населения торговыми площадями современных форматов;

X5 – Уровень безработицы.

Исходная статистика данных 2021 год по 79 субъектам Российской Федерации, за исключением городов федерального значения, которые взяты за отдельные субъекты (г. Москва, г. Санкт-Петербург, г. Севастополь) приводится в Приложении А.

На основе данных об экономическом состоянии 79 субъектов Российской Федерации, проверим гипотезу о влиянии независимых факторов, упомянутых выше, на оборот розничной торговли.

1. **ПРЕДВАРИТЕЛЬНЫЙ АНАЛИЗ МАССИВА ДАННЫХ И ЕГО ПРЕДМОДЕЛЬНАЯ ОБРАБОТКА**

Введем обозначения для исследуемых показателей и единицы измерения (см. табл. 2).

Таблица 2 - показатели и единицы измерения

|  |  |  |
| --- | --- | --- |
| Показатель | Обозначение | Единицы измерения |
| Оборот розничной торговли на душу населения | Y | Руб. |
| ВРП региона на душу населения | X1 | Тыс. руб. |
| Среднегодовая численность населения региона | X2 | Тыс. чел. |
| Средний доход на душу населения в регионе | X3 | Руб. |
| Обеспеченность населения торговыми площадями современных форматов | X4 | Кв. метр. |
| Уровень безработицы | X5 | % |

**2.1. Визуальный анализ**

Проведем визуальный анализ исходного массива данных. Для этого воспользуемся библиотекой «Seaborn» языка программирования Python. Гистограммы и графики разброса (см. рис. 2.1.1) представлены ниже для 2021.

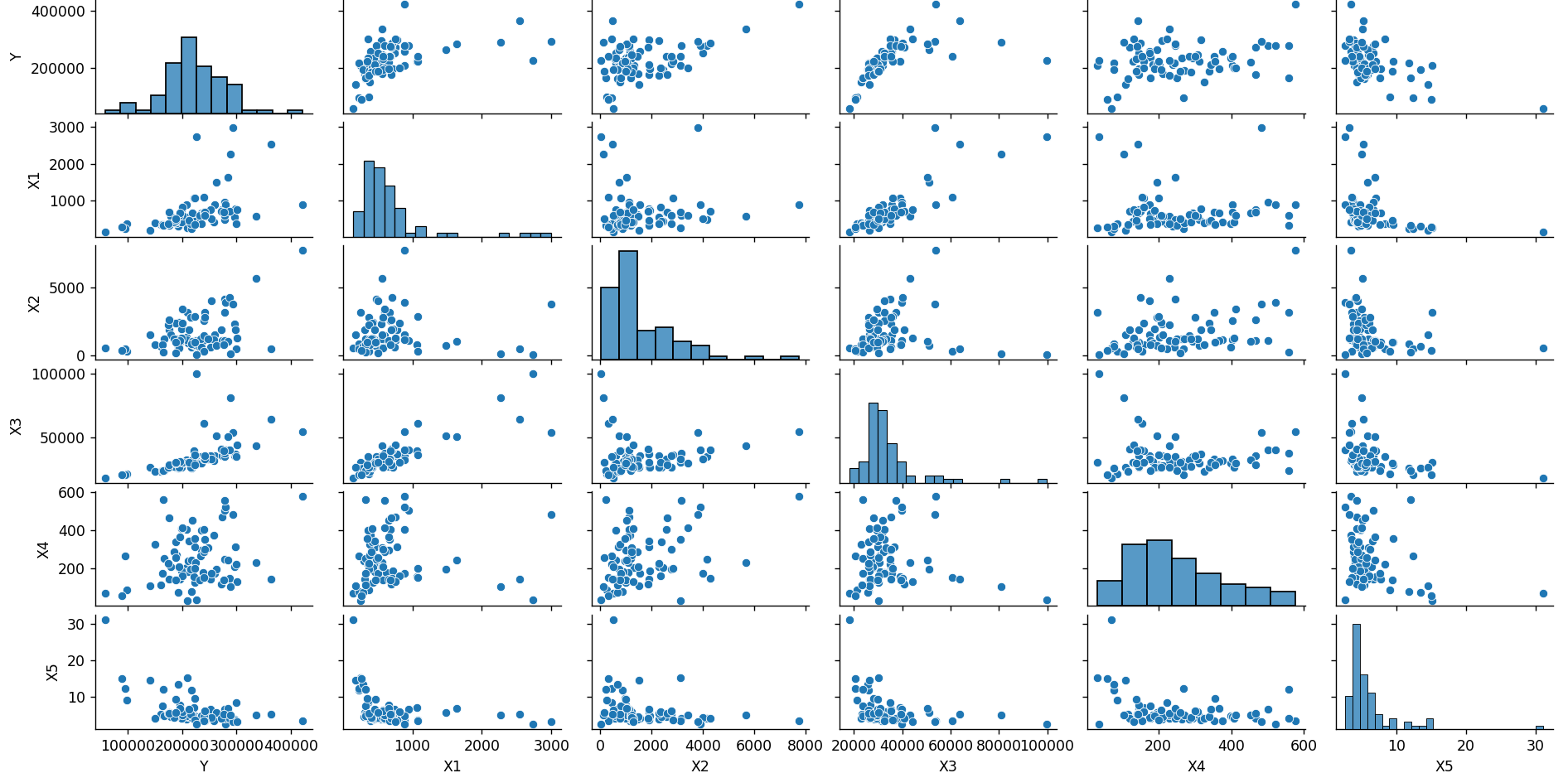


Рисунок 2.1.1 – Гистограммы и графики разброса 2021 год

Сделаем некоторые выводы по визуальному анализу исходных данных:

* между показателями оборот розничной торговли на душу населения (Y) и ВРП на душу населения (X1), существует средняя положительная линейная связь;
* между показателями оборот розничной торговли (Y) и среднегодовая численность населения региона (X2), существует средняя положительная линейная связь;
* между показателями оборот розничной торговли (Y) средний доход на душу населения в регионе (X3), существует сильная положительная линейная связь;
* между показателями оборот розничной торговли (Y) и обеспеченность населения торговыми площадями современных форматов (X4), существует слабая положительная линейная связь;
* между показателями оборот розничной торговли (Y) и уровень безработицы в регионе (X5), существует слабая отрицательная линейная связь;
* между экзогенными показателями также присутствует сильная или средняя линейная связь.

При помощи EXCEL построим коробчатые диаграммы для каждой переменной (см. рис. 2.1.3).

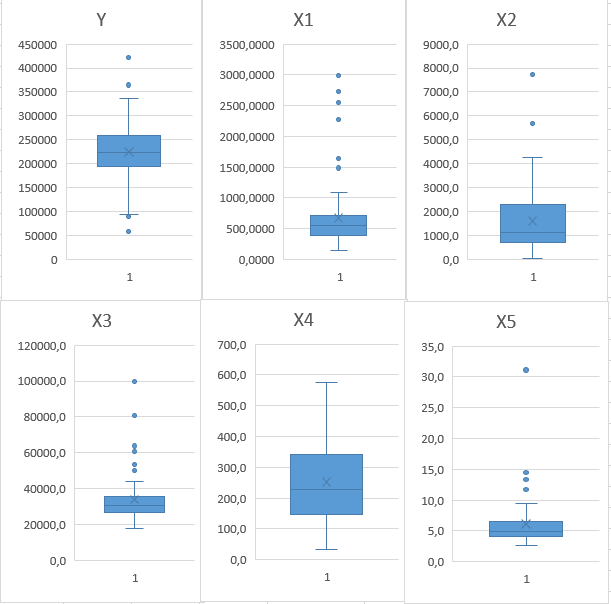


Рисунок 2.1.3 – Коробчатая диаграмма 2021 год

Исходя из построенных коробчатых диаграмм можно сделать следующий вывод: почти в каждом показатели присутствуют выбросы от 2 до 6 подозрительных точек.

* 1. **Поиск аномальных значений**

Выполним поиск аномальных значений по методу Граббса.

Расчётные значения приведены в таблице 2.2.

Таблица 2.2 - Расчетные значения

|  |  |  |
| --- | --- | --- |
| Переменные | Среднее | Стандартное отклонение |
| Оборот розничной торговли на душу населения (руб). | 224289,3 | 58463,8 |
| ВРП региона на душу населения (тыс. руб). | 674,871 | 529,731 |
| Среднегодовая численность населения региона (тыс. чел). | 1611,672 | 1340,647 |
| Средний доход на душу населения в регионе (руб). | 33982 | 12660 |
| Обеспеченность населения торговыми площадями современных форматов (кв. метр). | 251,64 | 131,64 |
| Уровень безработицы (%) | 6,09 | 3,99 |

Для переменной Y по графику (см. рис. 2.2.1) видим 6 «подозрительных» точек. Проверим их на аномальность.

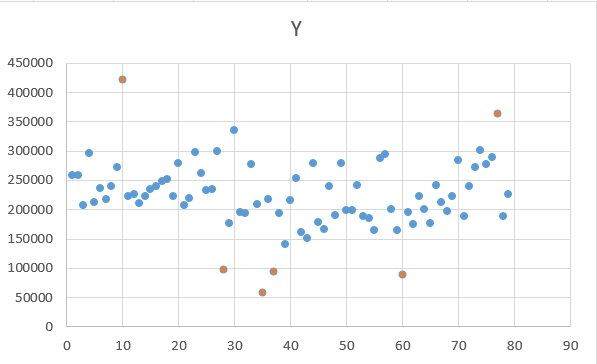


Рисунок 2.2.1 Диаграмма разброса по переменной Y

, где – значение показателя по рассматриваемому региону, – среднее по выборке, – выборочное среднеквадратическое отклонение.

Следовательно, наблюдение аномально и его нужно удалить из выборки. Аналогично проверяем следующие 5 «подозрительных» наблюдений. Они не аномальны.

Аналогично выявлены аномальные наблюдения по другим переменным приводится в Приложении Б.

Заключение: таким образом за 2021 год аномальными признаны значения по следующим субъектам: Московская область, Тюменская область, Чукотский авт. округ, республика Ингушетия. Исключим их из выборки.

**2.3. Корреляционный анализ**

Для массива данных без аномальны значений проведем корреляционный анализ, используя пакет анализ данных Microsoft EXCEL (см рис 2.3.1)

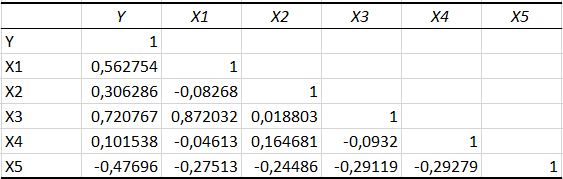
****

Рисунок 2.3.1 - Карта корреляции с исключёнными аномальными значениями для данных за 2021 год

Также проверим коэффициенты корреляции на значимость (см. рис 2.3.2). Для оценки значимости коэффициента корреляции используем t-критерий Стьюдента (t-статистику). При этом выдвигается и проверяется гипотеза о равенстве нулю коэффициента корреляции. Если эта гипотеза отвергается, то коэффициент корреляции признается значимым, а связь между переменными существенной. Найдем расчетные значения по формуле:

Полученные значения будут сравниваться с критическим, если расчетное значение t-критерия оказывается больше табличного, коэффициент корреляции признается значим, в ином случае — нет.



Рисунок 2.3.2 - Проверка коэффициентов корреляции на значимость 2021 год

Поскольку уровень значимости равен 0,05, значимыми будут считаться коэффициенты, чьи p-значения окажутся меньше, чем 0,05. По полученным результатам можно сделать вывод, что не все коэффициенты корреляции являются значимыми.

Между переменными X1 и X3 прослеживается такое явление, как мультиколлинеарность. Это значит, что есть наличие линейной зависимости между объясняющими переменными (факторами) регрессионной модели. Мультиколлинеарность может искажать результаты регрессионной модели, поэтому от нее необходимо избавиться.

Заключение: по итогам корреляционного анализа за 2021 год исключаем фактор X1, чтобы избавиться от мультиколлинеарности, так как значение коэффициента корреляции между Y и X1 меньше, чем между Y и X3 .

1. **МНОЖЕСТВЕННЫЙ ЛИНЕЙНЫЙ РЕГРЕССИОННЫЙ АНАЛИЗ**

Рассчитаем оценки параметров модели линейной регрессии. Для этого воспользуемся программой «Регрессия», выбрав соответствующий пункт меню надстройки «Анализ данных» Microsoft Excel.

Проведём тестирование статистической значимости параметров модели посредством -статистики Стьюдента, -значений и доверительных интервалов.

В случае, если абсолютное расчетное значение статистики Стьюдента превышает критическое значение, нулевая гипотеза отвергается, что указывает на статистическую значимость соответствующей оценки параметра регрессии.

Если -значение, связанное с оценкой параметра регрессии, оказывается меньше установленного уровня значимости, то нулевая гипотеза отвергается, что свидетельствует о статистической значимости соответствующей оценки параметра регрессии.

Если доверительный интервал не включает значение «0», то нулевая гипотеза отклоняется, и соответствующий параметр регрессии статистически значим.

В таблице 3 продемонстрирован процесс пошагового исключения незначимых регрессоров. В приложении В продемонстрированы рисунки из надстройки «Анализ данных» Microsoft Excel.

Заключение: наилучшим является уравнение, полученное на втором шаге, так как она не имеет незначимых факторов, отсутствует мультиколлинеарность между переменными.

Таблица 3 – Пошаговое исключение регрессоров и характеристики моделей регрессии

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Год | № модели | Модель | Набл. знач. t | Значение p | Доверительные интервалы | Характеристики |
| 2021 | 1 |  |  |  | (3,95; 17,21)  (2,69; 4,27)  (-36,1; 89,46)  (-6652,8; -897,85) | 31036 |
| 2 |  |  |  | (4,25; 17,42)  (2,65; 4,19)  (-6883; -1421) | 30975 |

**4.** **ДИСПЕРСИОННЫЙ АНАЛИЗ**

Проведем дисперсионный анализ для определения значимости уравнений регрессии. Будем проводить его для наилучшего уравнения, которое получили при регрессионном анализе. Обратимся к таблице «Дисперсионный анализ», которая выводится при построении уравнений регрессии в MS Excel (см. рис. 4.1.1).

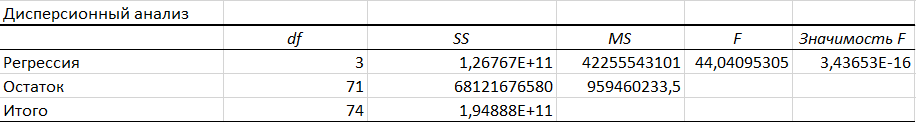


Рисунок 4.4.1 – Дисперсионный анализ наилучшей модели 2021

Основные составляющие полной вариации объясняемой переменной:

значения которых приводятся в столбце «SS»; а в столбце «MS» приведены значения величин , .

Проверка проводится на основе расчета критерия Фишера. Вычисляется F-статистика, которая далее сравнивается с табличным значением распределения Фишера с степенями свободы, где n – количество наблюдений, m – количество независимых переменных (X).

Таким образом, , а из таблиц распределения Фишера определяем . Следовательно, так как F, то с вероятность 95% можно говорить о значимости в целом построенной модели регрессии.

Заключение: построенная нами модель регрессии являются значимой с вероятностью 95%.

**5. ПРОВЕРКА ДОПОЛНИТЕЛЬНЫХ ИНФОРМАЦИОННЫХ И ПРОГНОСТИЧЕСКИХ СВОЙСТВ УРАВНЕНИЙ РЕГРЕССИИ**

На этом этапе исследования необходимо изучить специфические характеристики свойств остаточной компоненты ряда. Это выражается в тестировании остатков на соответствие исходным гипотезам Гаусса-Маркова:

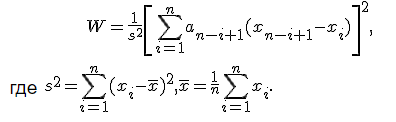
1. Нормальность остатков: .
2. Несмещённость МНК-оценок: .
3. Гомоскедастичность: *.*
4. Отсутствие автокорреляции ошибок:

.

**5.1. Оценка соответствия остатков нормальному закону**

Реализуем тест Шапиро-Уилка для определения нормальности распределения остатков.

Критерий Шапиро-Уилка основан на оптимальной линейной несмещённой оценке дисперсии к её обычной оценке методом максимального правдоподобия. Статистика критерия имеет вид:



Критические значения статистики также находятся таблично.

Если , то нулевая гипотеза о нормальности распределения отклоняется при уровне значимости. Зададим равным 0,05. Обратимся к статистическому пакету «Gretl». В нем применим «тест на нормальность». Мы получили значение p = 0,585462, так как оно больше уровня значимости можно сделать вывод, что остатки распределены нормально.

**5.2 Проверка несмещённости МНК-оценок**

Для доказательства свойства несмещённости будем проверять гипотезу о равенстве математического ожидания остатков нулю. Проверка осуществляется на основе t-критерия Стьюдента. Расчетное значение этого критерия , в свою очередь . Так как расчётное значение меньше критического то гипотеза о равенстве нулю математического ожидания случайной последовательности принимается. Условие несмещённости выполнено.

**5.3 Проверка гетероскедастичности остатков**

Гетероскедастичностью называется нарушение условия постоянства дисперсии ошибок в модели регрессии

.

Непостоянство ошибок регрессии означает, что для одних значений объясняющих переменных разброс значений зависимой переменной будет больше, а для других значений – меньше. Часто гетероскедастичность ошибок регрессии возникает при построении регрессионных моделей для неоднородных данных. Графики разброса остатков приведены в Приложении Г.

Для проверки используем тест Уайта. Построим вспомогательную регрессию на переменных значимой модели №2. Регрессия модели имеет вид:

Вспомогательная регрессия теста Уайта будет иметь вид:

,

Выполним её построение в MS Excel с помощью программы «Анализ данных – Регрессия». Результаты выполнения программы представлены на рис. 5.3.1.

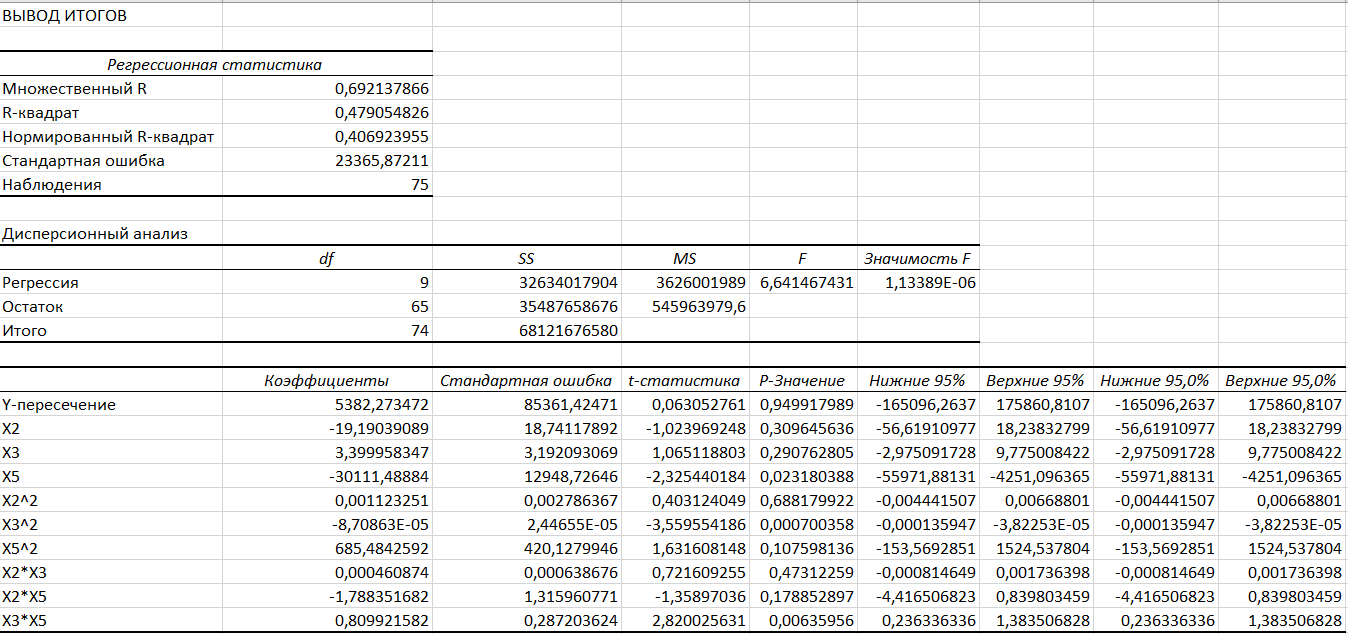


Рисунок 5.3.1 – Результаты работы программы «Регрессия»

Результаты теста Уайта показывают присутствие гетероскедастичности, так как при 5% уровне значимости Р-вероятность принятия гипотезы о наличии гетероскедастичности в остатках равна 1,13E-06, что меньше 0,05.

**5.4 Оценка автокорреляции остатков модели**

Проверим модели на автокорреляцию остатков. Выдвинем гипотезы:

При автокорреляции нарушается условие теоремы Гаусса – Маркова о некоррелированности возмущений для различных моментов времени.

Для определения автокорреляции воспользуемся статистикой Дарбина-Уотсона.

Посчитаем остатки модели, их квадрат и квадрат разницы предыдущего и нынешнего значения остатков модели №2. Посчитаем суммы, начиная от второго значения (см.рис. 5.4.1)

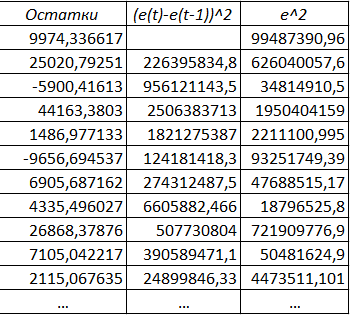


Рисунок 5.4.1 – Вспомогательные расчеты для статистики Дарбина-Уотсона

Посчитаем по формуле значение статистики Дарбина-Уотсона:

Определим интервалы неопределённости:

Из таблицы значений Дарбина-Уотсона

.

Интервалы равны:

.

Обратимся к интервалу определения автокорреляции (см. рис. 5.4.2)

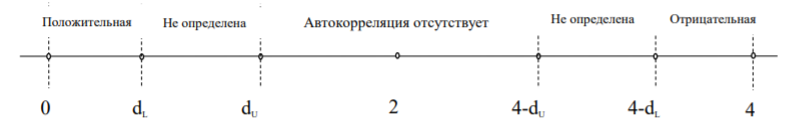


Рисунок 5.4.2 – Интервалы автокорреляции

По модели №2 можно сделать вывод, что автокорреляция отсутствует, так как попадает в промежуток .

## **5.5 Устранение гетероскедастичности остатков**

Для устранения гетероскедастичности воспользуемся методом взвешенных наименьших квадратов. Необходимо рассчитать , , . Получим следующее преобразование:

Для модели №3 разделим исходные векторы значений на = 30975,1551 и построим новое уравнение регрессии:

Условия гомоскедастичности и выполняются, следовательно, остатки гомоскедастичны.

**6.** **ВЫЯВЛЕНИЕ ЛУЧШИХ МОДЕЛЕЙ**

В ходе эконометрического исследования были получены следующие статистически значимые факторные регрессионные модели:

Таблица 6 – Полученные модели и прогностические характеристики

|  |  |  |
| --- | --- | --- |
| Модель | *(\*)* |  |
|  | 65% | 65% |
|  | 30975 | 1 |
|  | 7,9% | 11,4% |
| Автокорреляция |  |  |
| Гетероскедастичность |  |  |

Наилучшей моделью будет являться (\*\*), пусть она и не имеет минимальную ошибку аппроксимации, но у этого уравнения отсутствует автокорреляция и гетероскедостичность в остатках, и она имеет минимальную стандартную ошибку.

**7. ПРОГНОЗИРОВАНИЕ**

На данном этапе попробуем построить прогноз выбранной модели регрессии.

Доверительный интервал прогноза вычисляется согласно формуле:

|  |  |  |
| --- | --- | --- |
|  | |  |
| где: | – стандартная ошибка прогноза. | |

Стандартная ошибка прогноза вычисляется согласно формуле:

|  |
| --- |
|  |

Будем строить доверительные интервалы для каждого региона. Для прогнозирования был использован пакет анализа «Gretl». Мы получили таблицу и график содержащие значение Y, спрогнозированное значение Y и доверительный интервал по каждому региону (см. рис. 7.1 и 7.2).

По графику и таблице видно, что все спрогнозированные значения попали в доверительные интервалы прогноза. Прогнозирование можно считать адекватным.

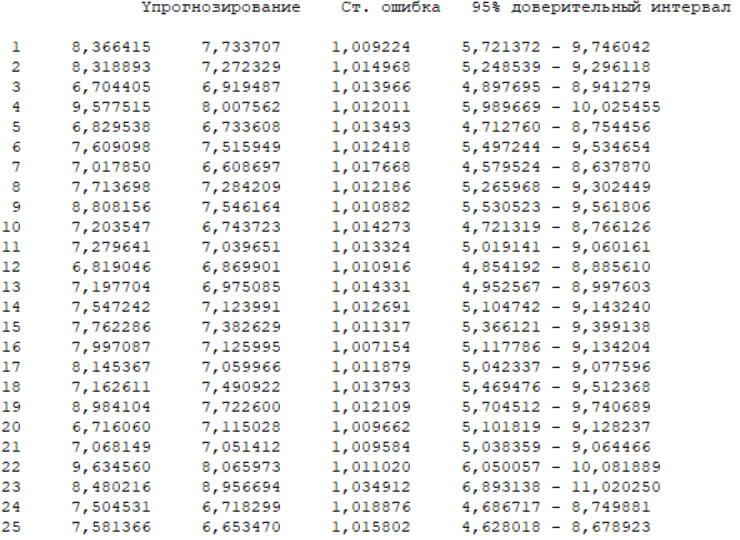


Рисунок 7.1 – Часть таблицы прогноза

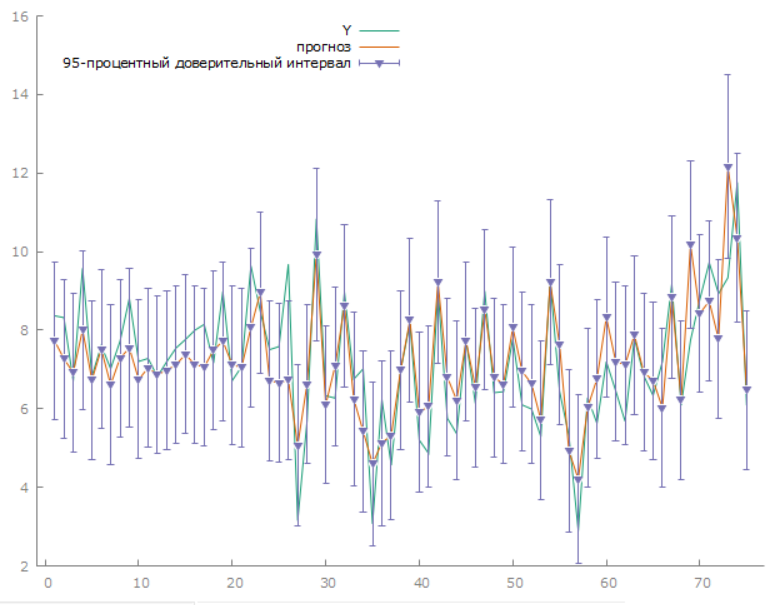


Рисунок 7.2 – График прогноза

**ЗАКЛЮЧЕНИЕ**

В процессе выполнения данной проектной работы была проделана работа по анализу розничной торговли в разрезе субъектов РФ.

Были рассмотрены 79 субъекта РФ. В ходе работы был выбран показатель оборот розничной торговли на душу населения и определены факторы, оказывающие влияние на него.

Проведя ряд исследований и тестов, построив несколько экономических моделей, анализируя которые была выбрана модель для прогнозирования:

Благодаря проведенным исследованиям нами так же были получены следующие навыки:

– закрепились теоретические знания, полученные в ходе изучения учебных дисциплин, связанных с вопросами обоснования необходимости и допустимости применения факторных регрессионных моделей;

– закрепились практические навыки в области предмодельной обработки данных, спецификации, калибровки, валидации факторных эконометрических моделей;

– провели экспериментальные расчеты по обоснованию, оценке и проведению процедуры верификации факторной регрессионной модели;

– приобрели начального опыт по идентификации и снижению эффектов ошибок спецификации факторных линейных моделей.

**СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ**

1. Писарева О.М. Эконометрика: Начальный курс. Москва: Государственный университет управления, Институт информационных систем ГУУ.
2. Елисеева И.И. Эконометрика. Москва: “Финансы и статистика”.
3. Исходные данные по: оборот розничной торговли на душу населения, ВРП региона на душу населения, средний доход на душу населения, уровень безработицы // Федеральная служба государственной статистики URL: <https://rosstat.gov.ru> (дата обращения: 03.10.2023).
4. Исходные данные по: обеспеченность населения торговыми площадями современных форматов: // [Единая межведомственная информационно – статистическая система (ЕМИСС)](https://fedstat.ru/) URL: <https://www.fedstat.ru> (дата обращения: 19.12.2023).
5. Определение оборота розничной торговли // Федеральная служба государственной статистики URL: <https://rosstat.gov.ru/storage/mediabank/Оборот%20розничной%20торговли.htmlа> (дата обращения: 03.10.2023)
6. Определение ВРП // Федеральная служба государственной статистики URL: <https://rosstat.gov.ru/storage/mediabank/WXWTVX2I/Понятия%20и%20определения%20(ВРП).pdf#:~:text=Валовой%20региональный%20продукт%20(ВРП)> (дата обращения: 03.10.2023)
7. Пособие по GRETL URL: <https://gretl.sourceforge.net/gretl-help/gretl-guide.pdf> (дата обращения: 19.12.2023)
8. Статистическое исследование оборота розничной торговли на душу населения РФ URL: <http://www.vectoreconomy.ru/images/publications/2022/1/mathematicalmethods/Spirin.pdf> (дата обращения: 15.11.2023)
9. Оценка влияния региональных факторов на распространение торговых сетей в РФ URL: <http://pe.cemi.rssi.ru/pe_2011_4_71-84.pdf> (дата обращения: 15.11.2023)

**ПРИЛОЖЕНИЕ А - Исходные данные**

Таблица 1 - Исходные данные.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Y1 | X1 | X2 | X3 | X4 | X5 |
| Белгородская область | 259151 | 881,7006 | 1536,6 | 35612,0 | 175,3 | 4,2 |
| Брянская область | 257679 | 398,6186 | 1175,7 | 31608,0 | 375,2 | 3,4 |
| Владимирская область | 207670 | 552,8112 | 1332,9 | 28523,0 | 203,5 | 3,9 |
| Воронежская область | 296665 | 546,3288 | 2296,6 | 35198,0 | 210,4 | 3,8 |
| Ивановская область | 211546 | 306,1445 | 982,0 | 28680,0 | 240,0 | 4,5 |
| Калужская область | 235693 | 659,5909 | 1006,9 | 35076,0 | 295,0 | 4,0 |
| Костромская область | 217379 | 386,6948 | 624,6 | 28560,0 | 243,9 | 4,4 |
| Курская область | 238933 | 627,3206 | 1090,0 | 32715,0 | 178,4 | 4,0 |
| Липецкая область | 272834 | 752,9260 | 1120,9 | 35231,0 | 467,3 | 4,2 |
| Московская область | 422211 | 882,8756 | 7738,7 | 53989,0 | 575,4 | 3,4 |
| Орловская область | 223131 | 468,0185 | 719,4 | 29846,0 | 311,6 | 4,7 |
| Рязанская область | 225488 | 487,2764 | 1091,7 | 30496,0 | 229,9 | 4,0 |
| Смоленская область | 211221 | 460,5970 | 915,5 | 30731,0 | 185,1 | 5,0 |
| Тамбовская область | 222950 | 434,6128 | 987,7 | 30241,0 | 129,2 | 4,0 |
| Тверская область | 233777 | 448,4176 | 1237,9 | 30675,0 | 264,5 | 3,9 |
| Тульская область | 240438 | 602,2981 | 1440,8 | 32252,0 | 285,5 | 3,8 |
| Ярославская область | 247711 | 559,1798 | 1234,4 | 33131,0 | 308,6 | 5,9 |
| Республика Карелия | 252304 | 737,7819 | 606,1 | 35373,0 | 145,6 | 6,6 |
| Республика Коми | 221863 | 1059,9604 | 808,5 | 39118,0 | 158,8 | 7,0 |
| Архангельская область | 278284 | 941,7625 | 1120,7 | 39741,0 | 502,0 | 6,6 |
| Вологодская область | 208031 | 881,8155 | 1145,3 | 31858,0 | 402,1 | 4,7 |
| Калининградская область | 218937 | 659,7273 | 1023,2 | 32033,0 | 453,3 | 5,0 |
| Ленинградская область | 298432 | 778,6922 | 1902,1 | 36855,0 | 315,6 | 3,7 |
| Мурманская область | 262676 | 1487,3636 | 728,7 | 51183,0 | 197,2 | 5,8 |
| Новгородская область | 232454 | 580,4950 | 589,3 | 29300,0 | 139,7 | 4,1 |
| Псковская область | 234834 | 356,5946 | 616,8 | 29354,0 | 399,1 | 4,7 |
| Республика Адыгея | 299840 | 366,7017 | 465,8 | 34901,0 | 223,2 | 8,3 |
| Республика Адыгея | 97566 | 371,9560 | 268,9 | 21319,0 | 87,4 | 9,0 |
| Республика Калмыкия | 176044 | 308,8479 | 1899,0 | 26381,0 | 145,0 | 5,8 |
| Республика Крым | 335663 | 562,9260 | 5685,7 | 43255,0 | 230,7 | 5,0 |
| Краснодарский край | 196096 | 661,2449 | 993,6 | 26833,0 | 145,2 | 7,7 |
| Астраханская область | 194191 | 427,0688 | 2462,2 | 27680,0 | 207,5 | 4,9 |
| Волгоградская область | 277842 | 483,9704 | 4167,6 | 35050,0 | 247,2 | 4,0 |
| Ростовская область | 209177 | 259,0764 | 3143,6 | 30260,0 | 32,5 | 15,1 |
| Республика Дагестан | 58176 | 148,5868 | 519,8 | 18139,0 | 70,4 | 31,1 |
| Республика Ингушетия | 217260 | 229,1531 | 869,8 | 25929,0 | 78,8 | 11,7 |
| Кабардино-Балкарская Республика | 95066 | 235,3547 | 464,8 | 20470,0 | 268,2 | 12,3 |
| Карачаево-Черкесская Республика | 193290 | 293,3661 | 690,6 | 25885,0 | 77,1 | 13,4 |
| Республика Северная Осетия Алания | 141300 | 177,8599 | 1507,2 | 26397,0 | 109,6 | 14,5 |
| Чеченская Республика | 215801 | 367,6871 | 2786,5 | 26186,0 | 195,4 | 5,3 |
| Ставропольский край | 253472 | 499,0448 | 4007,7 | 32621,0 | 174,7 | 4,3 |
| Республика Башкортостан | 161110 | 329,6598 | 673,4 | 23184,0 | 115,8 | 5,3 |
| Республика Марий Эл | 150613 | 384,6357 | 774,8 | 22906,0 | 325,8 | 4,2 |
| Республика Мордовия | 279262 | 888,0390 | 3890,3 | 39679,0 | 522,6 | 2,6 |
| Республика Татарстан | 178357 | 565,4723 | 1488,9 | 27700,0 | 211,2 | 4,5 |
| Удмуртская Республика | 166150 | 326,6070 | 1203,2 | 23651,0 | 251,7 | 4,8 |
| Чувашская Республика | 239837 | 677,7599 | 2568,1 | 32750,0 | 402,4 | 4,6 |
| Пермский край | 190042 | 387,4575 | 1242,5 | 26649,0 | 272,2 | 4,9 |
| Кировская область | 278659 | 597,4306 | 3160,4 | 37525,0 | 557,0 | 4,2 |
| Нижегородская область | 198494 | 721,0252 | 1933,7 | 26514,0 | 188,4 | 4,7 |
| Оренбургская область | 199270 | 418,9461 | 1282,5 | 26416,0 | 408,7 | 4,2 |
| Пензенская область | 241042 | 675,3345 | 3142,9 | 32664,0 | 353,8 | 3,5 |
| Самарская область | 188844 | 422,9546 | 2378,0 | 26228,0 | 340,5 | 4,5 |
| Саратовская область | 185625 | 411,8472 | 1211,1 | 26849,0 | 289,6 | 4,3 |
| Ульяновская область | 163801 | 330,6422 | 812,0 | 23746,0 | 177,5 | 7,6 |
| Курганская область | 287048 | 710,3808 | 4277,2 | 40275,0 | 150,9 | 4,1 |
| Свердловская область | 293631 | 2992,7754 | 3792,3 | 53382,0 | 481,8 | 3,2 |
| Тюменская область | 199595 | 595,3854 | 3430,7 | 29498,0 | 413,0 | 4,9 |
| Челябинская область | 164472 | 322,4134 | 221,3 | 23798,0 | 558,6 | 12,0 |
| Республика Алтай | 89480 | 267,7947 | 331,5 | 20651,0 | 60,2 | 15,0 |
| Республика Тыва | 194578 | 580,0160 | 530,2 | 26068,0 | 208,0 | 6,4 |
| Республика Хакасия | 174321 | 370,4343 | 2282,3 | 26010,0 | 229,1 | 5,5 |
| Алтайский край | 222969 | 1074,4242 | 2852,5 | 36090,0 | 201,4 | 3,6 |
| Красноярский край | 200034 | 813,3124 | 2366,1 | 30347,0 | 160,5 | 6,0 |
| Иркутская область | 176024 | 690,1430 | 2618,9 | 28048,0 | 466,8 | 5,4 |
| Кемеровская область - Кузбасс | 241409 | 581,0183 | 2783,1 | 35277,0 | 301,9 | 6,1 |
| Новосибирская область | 211880 | 451,5368 | 1891,6 | 29976,0 | 344,0 | 6,5 |
| Омская область | 196290 | 660,5981 | 1069,3 | 31054,0 | 366,2 | 6,8 |
| Томская область | 223010 | 347,7382 | 984,0 | 28314,0 | 356,3 | 9,5 |
| Республика Бурятия | 284035 | 1636,7342 | 1048,5 | 50369,0 | 245,8 | 6,9 |
| Республика Саха (Якутия) | 188623 | 464,8870 | 987,0 | 29833,0 | 139,1 | 9,3 |
| Забайкальский край | 239715 | 1081,1020 | 312,2 | 60791,0 | 154,3 | 3,5 |
| Камчатский край | 272134 | 699,7778 | 1870,4 | 40843,0 | 120,7 | 4,2 |
| Приморский край | 301099 | 759,3439 | 1300,1 | 44101,0 | 131,3 | 3,1 |
| Хабаровский край | 276635 | 683,1679 | 777,2 | 39626,0 | 141,5 | 5,2 |
| Амурская область | 289171 | 2273,8822 | 138,4 | 80979,0 | 105,7 | 4,9 |
| Магаданская область | 364356 | 2545,5925 | 484,9 | 63854,0 | 144,0 | 5,2 |
| Сахалинская область | 188492 | 507,2122 | 155,2 | 30297,0 | 258,9 | 5,6 |
| Еврейская авт.область | 225671 | 2734,8627 | 49,8 | 99912,0 | 37,2 | 2,6 |
| Чукотский авт.округ | 259151 | 881,7006 | 1536,6 | 35612,0 | 175,3 | 4,2 |

**ПРИЛОЖЕНИЕ Б – Аномальные значения**

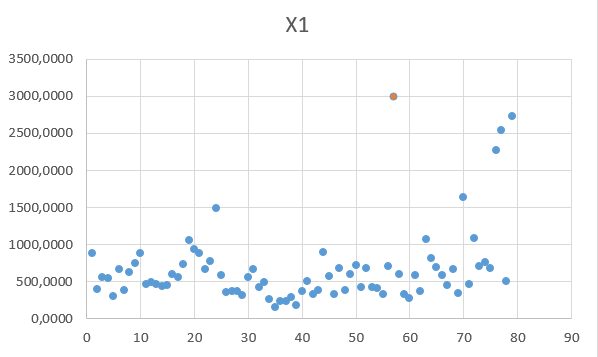


Рисунок 2.2.2 Диаграмма разброса по переменной X1

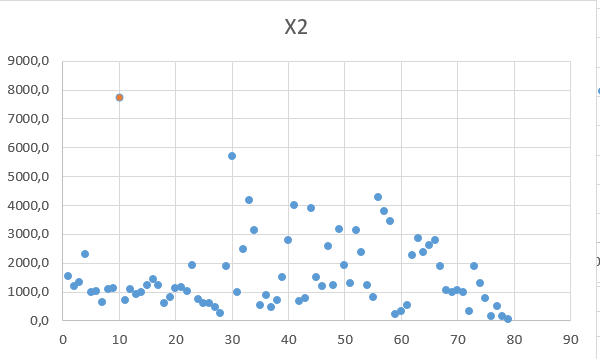


Рисунок 2.2.3 Диаграмма разброса по переменной X2

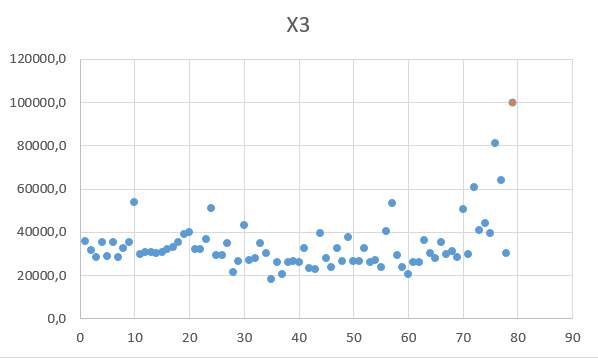


Рисунок 2.2.4 Диаграмма разброса по переменной X3

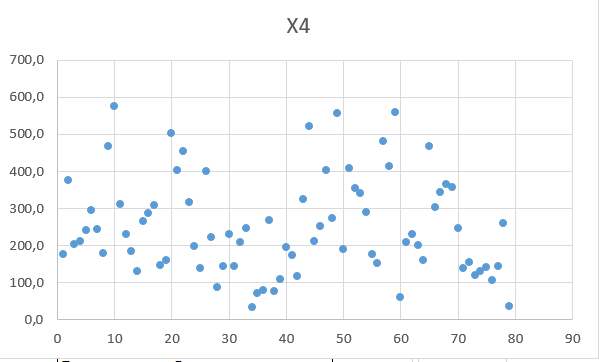


Рисунок 2.2.5 Диаграмма разброса по переменной X4

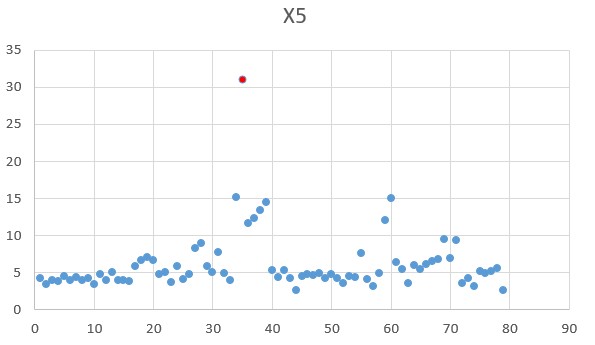


Рисунок 2.2.6 Диаграмма разброса по переменной X5

**ПРИЛОЖЕНИЕ В – Регрессионный анализ**

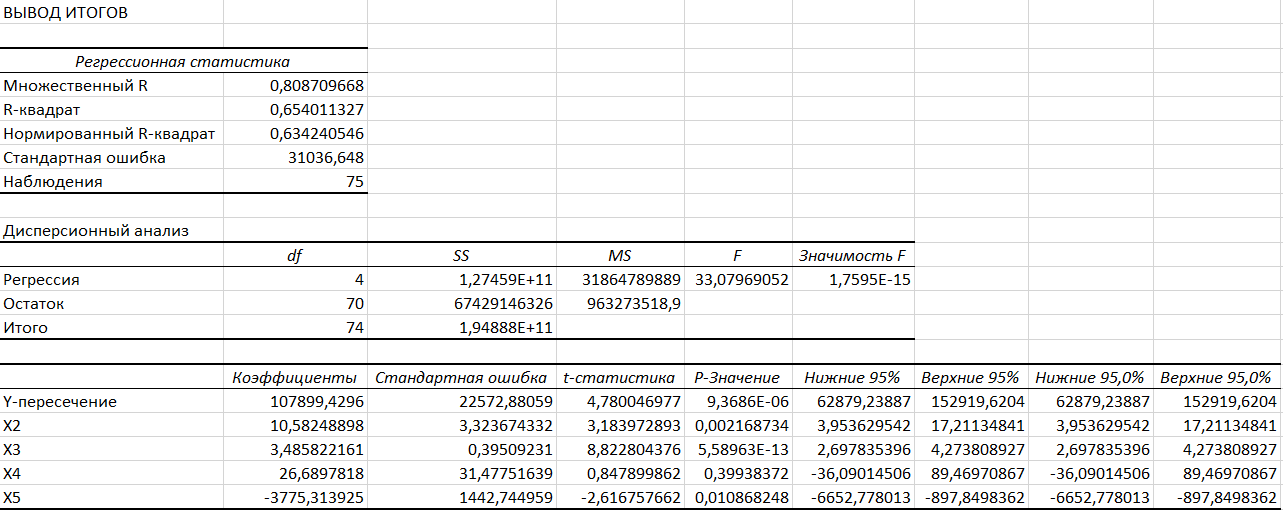


Рисунок 2.1 - Регрессионный анализ после исключения X1

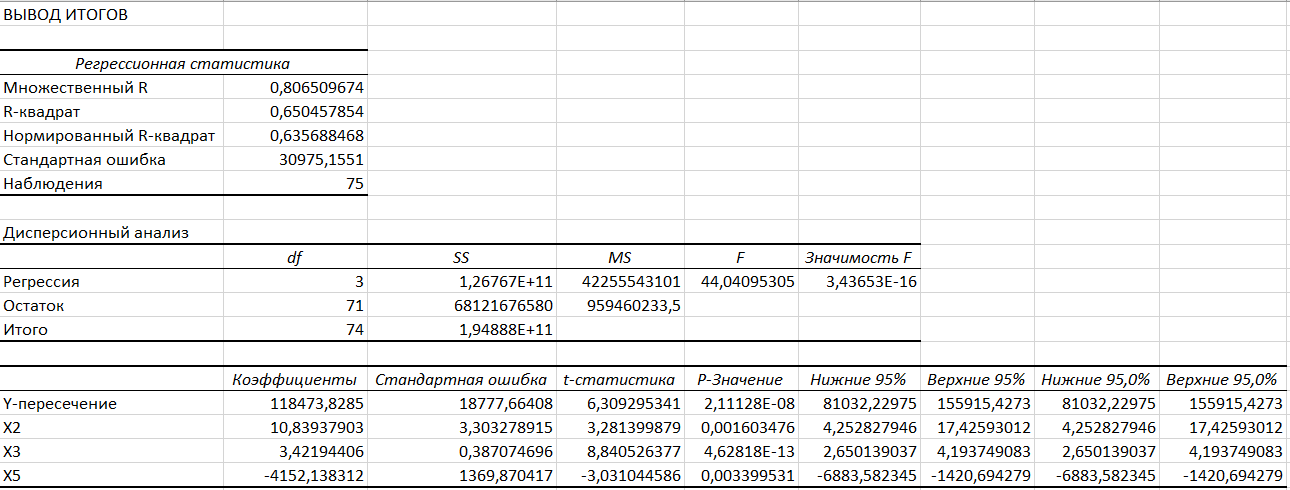


Рисунок 2.2 - Регрессионный анализ после исключения X1, X4

**ПРИЛОЖЕНИЕ Г – Графики разброса остатков**

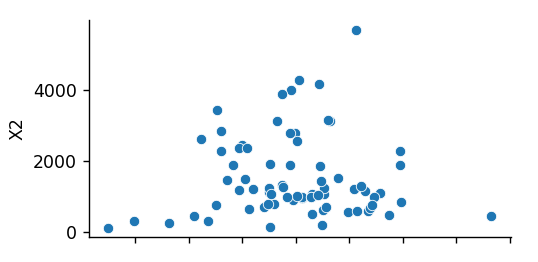


Рисунок 5.3.2 – График разброса остатков относительно X2

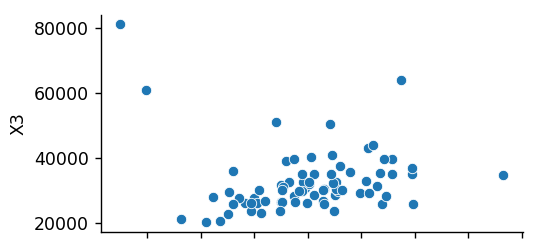


Рисунок 5.3.3 – График разброса остатков относительно X3

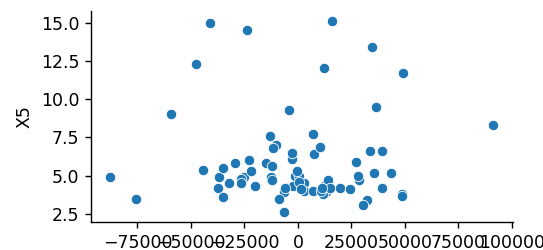


Рисунок 5.3.4 – График разброса остатков относительно X5