**Министерство образования и науки 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 |

**Проектная работа**

|  |  |
| --- | --- |
| по дисциплине | Методы и модели эконометрики |
|  | На тему |
| «Исследование с применением методов и моделей эконометрики влияния санкционных ограничений на динамику товарных ресурсов в розничной торговле» | |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Направление подготовки | 01.03.02 |  | Прикладная математика и информатика | |
| (код) |  | (наименование) | |
| Образовательная программа | Прикладная математика и информатика | | |
| (название образовательной программы) | | |
| Обучающиеся | Бобина Т.С.,  Ходжигорова К.Б.,  Челышев М.С.,  Шилкин С.В.,  Энверова А. | | |
| (Фамилия, Имя, Отчество) | | |
| 3 курс, 2 подгруппа |
| (курс, номер группы) |
| Руководитель работы | Заведующая кафедрой, к.э.н. Писарева О. М. | | |
| (ученая степень, звание, Фамилия и Инициалы) | | |

|  |  |
| --- | --- |
| Оценка: |  |
| Подпись руководителя: |  |
| Дата защиты: | «\_\_\_» \_\_\_\_\_\_\_\_\_\_\_\_\_2022 г. |

Москва – 2022

**СОДЕРЖАНИЕ**

[**СОДЕРЖАНИЕ 2**](#_Toc90223819)

[**ВВЕДЕНИЕ 3**](#_Toc90223820)

[**1. ПОСТАНОВКА ЗАДАЧИ ИССЛЕДОВАНИЯ И СБОР ДАННЫХ 5**](#_Toc90223821)

[**2. ПРЕДВАРИТЕЛЬНЫЙ АНАЛИЗ МАССИВА ДАННЫХ И ЕГО ПРЕДМОДЕЛЬНАЯ ОБРАБОТКА**](#_Toc90223822) **7**

[**2.1. Визуальный анализ………………………………………………………………………….**](#_Toc90223823)**7**

[**2.2. Поиск аномальных значений 1**](#_Toc90223824)**2**

**[2.3. Корреляционный анализ 1](#_Toc90223825)5**

**[2.4. Проверка на нормальность распределения 1](#_Toc90223825)7**

[**3. МНОЖЕСТВЕННЫЙ ЛИНЕЙНЫЙ РЕГРЕССИОННЫЙ АНАЛИЗ 1**](#_Toc90223829)**9**

[**4. ДИСПЕРСИОННЫЙ АНАЛИЗ 2**](#_Toc90223830)**3**

[**5. ПРОВЕРКА ДОПОЛНИТЕЛЬНЫХ ИНФОРМАЦИОННЫХ И ПРОГНОСТИЧЕСКИХ СВОЙСТВ УРАВНЕНИЙ РЕГРЕССИИ 2**](#_Toc90223831)**8**

[**5.1. Оценка остатков модели на гетероскедастичность 2**](#_Toc90223832)**8**

[**5.2. Оценка автокорреляции остатков модели 2**](#_Toc90223833)**9**

[**6. ВЫЯВЛЕНИЕ ЛУЧШИХ МОДЕЛЕЙ**](#_Toc90223834) **31**

[**7. ПРОГНОЗИРОВАНИЕ 3**](#_Toc90223835)**5**

**ЗАКЛЮЧЕНИЕ………………………………………………………………..……………….36**

[**СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ 3**](#_Toc90223839)**7**

[**ПРИЛОЖЕНИЕ 1- Исходные данные**](#_Toc90223842) **41**

[**ПРИЛОЖЕНИЕ 2 – Графики разброса, корреляция, описательные статистики**](#_Toc90223843) **44**

[**ПРИЛОЖЕНИЕ 4 – Регрессионный анализ 4**](#_Toc90223844)**8**

[**ПРИЛОЖЕНИЕ 6 – Графики разброса остатков**](#_Toc90223845) **53**

**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”. Для построения гистограмм и графиков разброса, корреляционного анализа, проверки коэффициентов на значимость был написан код с использованием библиотеки «Seaborn» и «Pandas». Для удобства было принято решение распределить работу таким образом, как представлено в таблице 1:

Таблица 1 – Распределение ролей

|  |  |  |
| --- | --- | --- |
| ФИО | Роль в команде | Оценка |
| Бобина Т.С. | Руководитель команды, сбор данных, расчетно-аналитическая работа, оформление отчета и презентации, защита |  |
| Ходжигорова К.Б. | Сбор данных, расчетно-аналитическая работа, оформление отчета и презентации, защита |  |
| Челышев М.С. | Сбор данных, расчетно-аналитическая работа, оформление отчета, защита |  |
| Шилкин С.В. | Сбор данных, расчетно-аналитическая работа, оформление отчета и презентации, защита |  |
| Энверова А. | Сбор данных, расчетно-аналитическая работа, оформление отчета и презентации, защита |  |

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

Для проведения эконометрического моделирования была сформулирована задача построения факторной регрессионной модели влияния санкционных ограничений на динамику товарных ресурсов в розничной торговле.

Для начала дадим определения терминам:

[Санкционные ограничения](https://mnv.irgups.ru/sites/default/files/articles_pdf_files/tkachenko.pdf) – это экономические меры запретительного характера, которые используются одним участником международной торговли по отношению к другому участнику, которые призваны ухудшить экономическую ситуацию последнего с целью принудить его к изменению политического курса [7].

Товарные ресурсы - товары, формируемые из объемов производства отечественных производителей, импортных поступлений и вовлечения товаров из запасов [8].

[Розничная торговля](https://ru.wikipedia.org/wiki/%D0%A2%D0%BE%D1%80%D0%B3%D0%BE%D0%B2%D0%BB%D1%8F#:~:text=%D0%A0%D0%BE%D0%B7%D0%BD%D0%B8%D1%87%D0%BD%D0%B0%D1%8F%20%D1%82%D0%BE%D1%80%D0%B3%D0%BE%D0%B2%D0%BB%D1%8F%20%E2%80%94%20%D1%8D%D1%82%D0%BE%20%D0%BE%D1%81%D0%BE%D0%B1%D0%B0%D1%8F,%D0%B4%D0%B2%D0%B8%D0%B6%D0%B5%D0%BD%D0%B8%D1%8F%20%D1%82%D0%BE%D0%B2) — деятельность, связанная с осуществлением акта купли-продажи товаров конечным потребителям [9].

Для исследования использовался ресурс «Федеральная служба государственной статистики» [4] и “Единая межведомственная информационно-статистическая система” [10], предоставляющиe официальную статистику по РФ.

В качестве независимых переменных были выбраны следующие экономические показатели страны:

1. среднедушевые денежные доходы населения;
2. индекс потребительских цен;
3. доля оборота розничных торговых сетей в общем объеме оборота розничной торговли;
4. обеспеченность населения торговыми площадями современных форматов;
5. доля населения с доходами ниже прожиточного минимума;
6. уровень безработицы (по методологии МОТ);
7. доля городского населения.

Исходная статистика данных за 2013, 2014, 2015 годы по 83 субъектам Российской Федерации приводится в Приложении 1.

Сформулируем задачу следующим образом. На основе данных об экономическом состоянии 83 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. 2.1).

Таблица 2.1 – Показатели и единицы измерения

|  |  |  |
| --- | --- | --- |
| Показатель | Обозначение | Единицы измерения |
| Оборот розничной торговли на душу населения |  | млн. руб. |
| Среднедушевые денежные доходы населения |  | руб./месяц |
| Индекс потребительских цен |  | % |
| Доля оборота розничных торговых сетей в общем объеме оборота розничной торговли |  | % |
| Обеспеченность населения торговыми площадями современных форматов |  | м2 на 1000 чел. |
| Доля населения с доходами ниже прожиточного минимума |  | % |
| Уровень безработицы (по методологии МОТ) |  | % |
| Доля городского населения |  | % |

**2.1. Визуальный анализ**

Проведем визуальный анализ исходного массива данных. Для этого воспользуемся библиотекой «Seaborn» языка программирования Python. Листинг кода (см. Приложение 2 листинг 2.1), гистограммы и графики разброса (см. рис. 2.1.1 и 2.1.2) представлены ниже для 2013 года. Для 2014 и 2015 годов результаты представлены в Приложении 2.

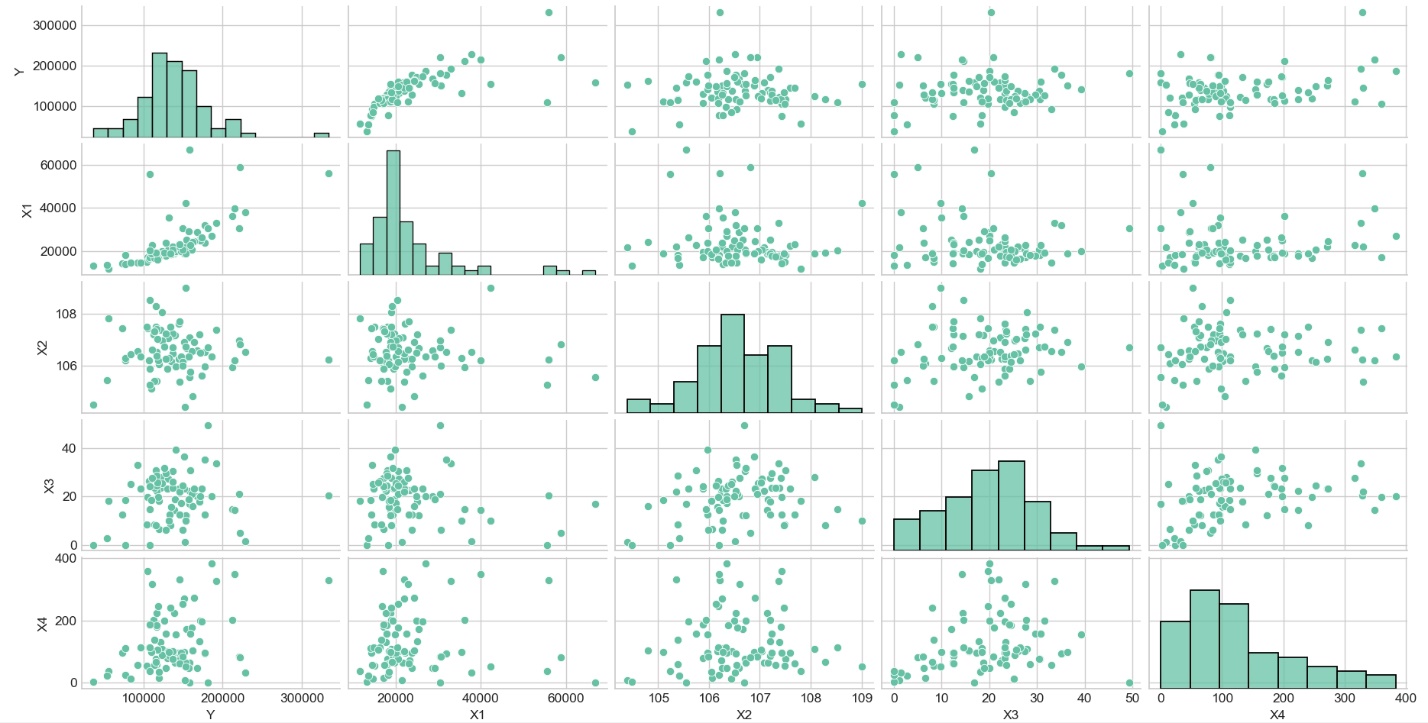


Рисунок 2.1.1 – Гистограммы и графики разброса 2013 год

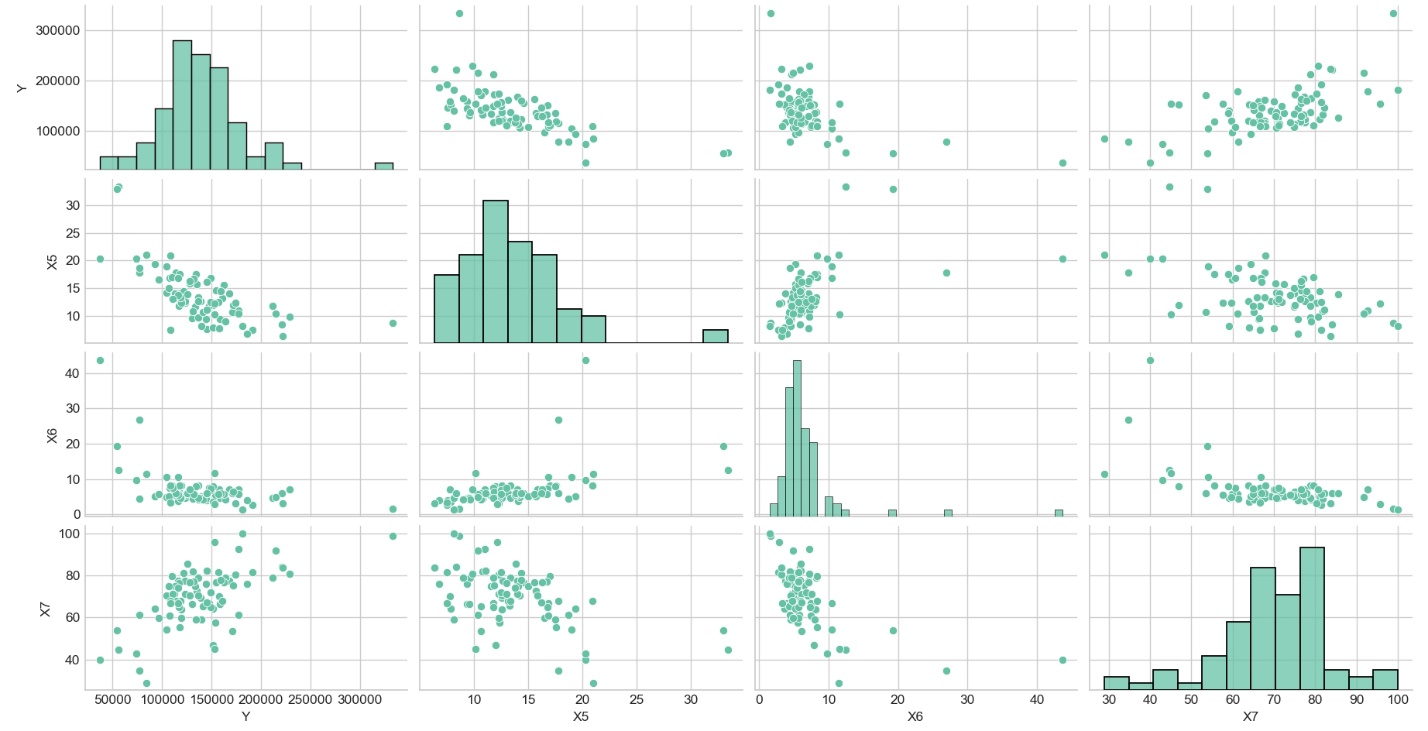


Рисунок 2.1.2 – Гистограммы и графики разброса 2013 год

Сделаем некоторые выводы по визуальному анализу исходных данных:

* по гистограммам видно, что распределение приближено к нормальному;
* между эндогенной и экзогенными переменными существует линейная связь. Наиболее выраженно это прослеживается с среднедушевыми денежными доходами населения, долей населения с доходами ниже прожиточного минимума;
* между экзогенными переменными также прослеживается линейная связь.

При анализе 2014 и 2015 годов выводы получились аналогичными.

Обратимся к статистическому пакету «Gretl». В нем воспользуемся построением графиков «Коробчатая диаграмма».

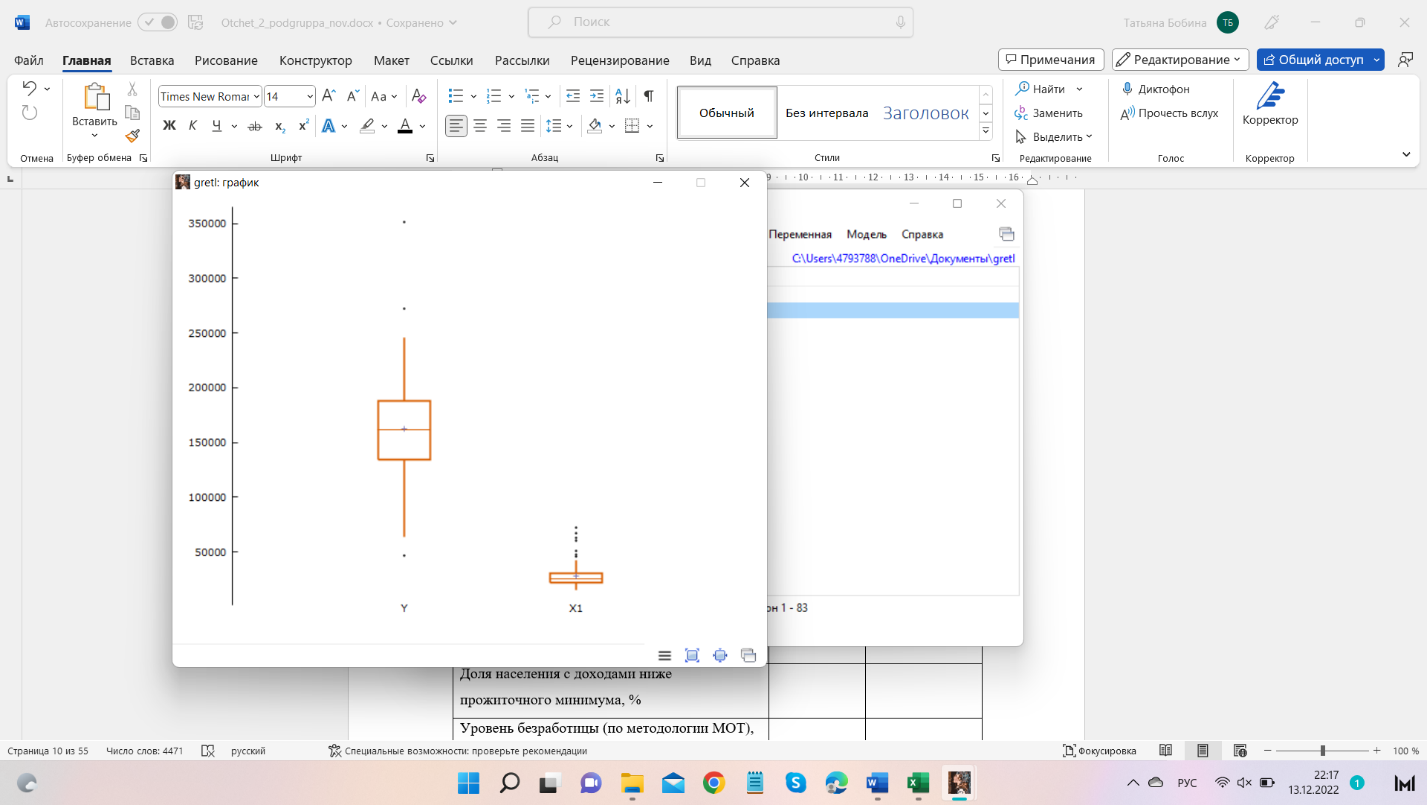
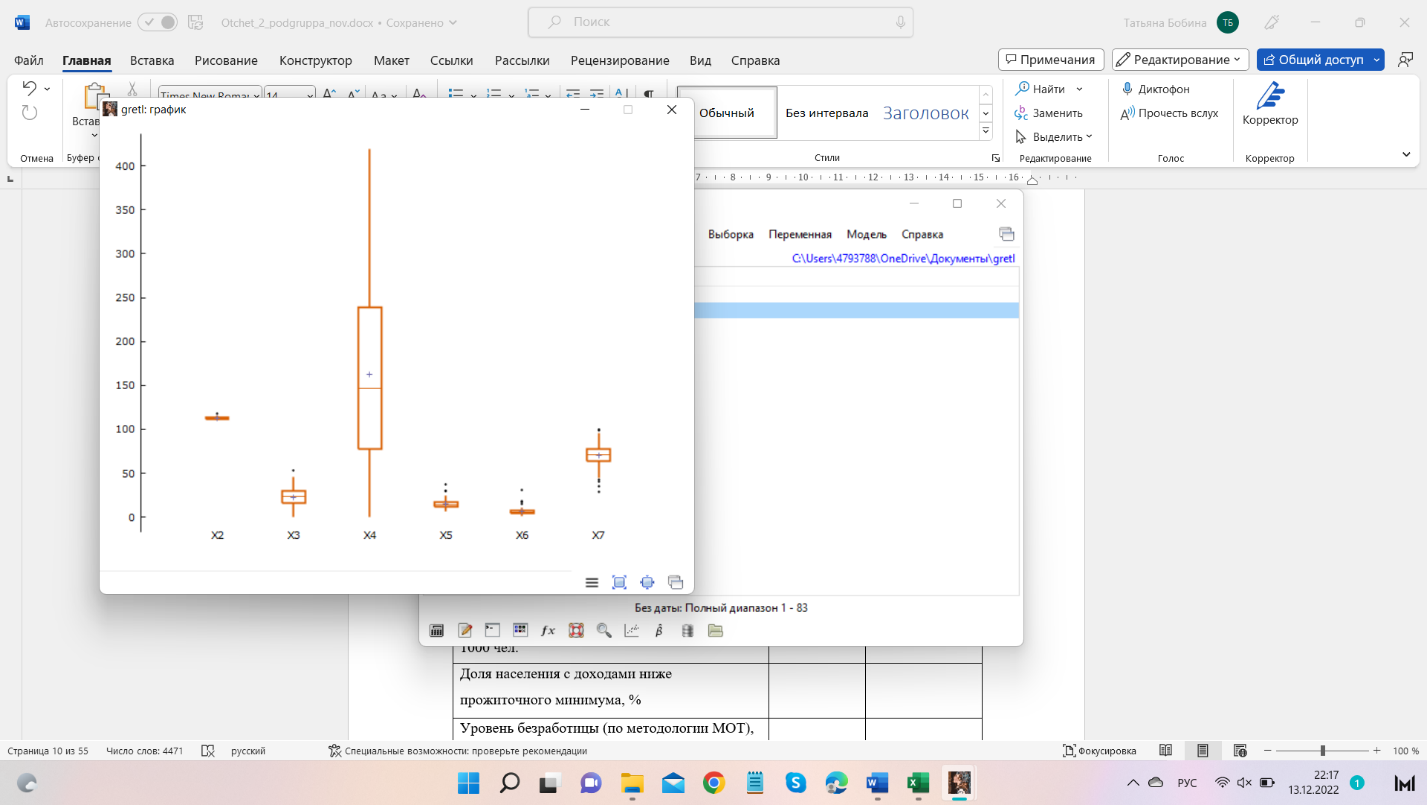
 

Рисунок 2.1.3 – Коробчатая диаграмма 2013 год

Исходя из построенных коробчатых диаграмм можно сделать следующие выводы:

* наибольший разброс данных у оборота розничной торговли на душу населения и обеспеченности населения торговыми площадями современных форматов;
* присутствуют аномальные значения.

По диаграммам за 2014 и 2015 год выводы аналогичны.

**2.2. Поиск аномальных значений**

Выполним поиск аномальных значений по методу Граббса.

Таблица 2.2.1 - Расчетные значения

|  |  |  |
| --- | --- | --- |
| Переменные | Среднее | Стандартное отклонение |
| Оборот розничной торговли на душу населения, млн. руб. | 138506,6 | 42531,25 |
| Среднедушевые денежные доходы населения, руб./месяц | 23400,28 | 10233,4 |
| Индекс потребительских цен, % | 106,5669 | 0,846437 |
| Доля оборота розничных торговых сетей в общем объеме оборота розничной торговли, % | 19,4988 | 9,644454 |
| Обеспеченность населения торговыми площадями современных форматов, м^2 на 1000 чел. | 126,7193 | 93,19114 |
| Доля населения с доходами ниже прожиточного минимума, % | 13,59277 | 4,693529 |
| Уровень безработицы (по методологии МОТ), % | 6,838554 | 5,322223 |
| Доля городского населения, % | 69,80241 | 13,28087 |

Для переменной «Оборот розничной торговли на душу населения» по графику (см. рис. 2.2.1) видим 15 «подозрительных» точек. Проверим их на аномальность.

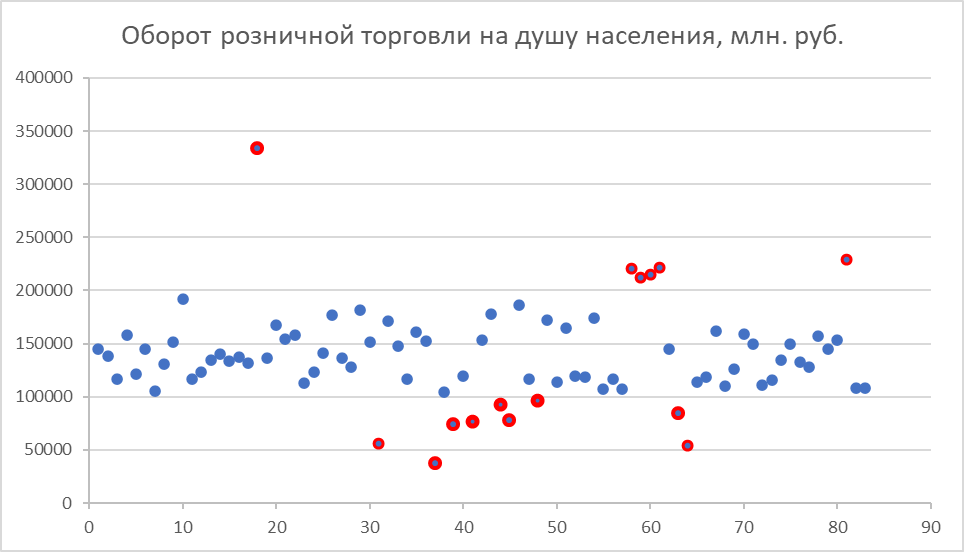


Рисунок 2.2.1 Диаграмма разброса по переменной «Оборот розничной торговли на душу населения»

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

Следовательно наблюдение аномально и нужно удалить из выборки. Аналогично проверяем следующие 14 «подозрительные» наблюдений. Они не аномальны.

Аналогично выявлены аномальные наблюдения по другим переменным. Ниже представлены графики разброса, где присутствуют аномальные значения.

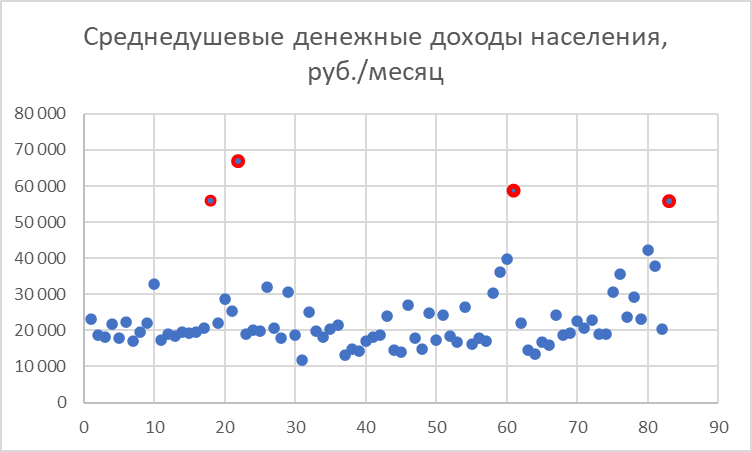


Рисунок 2.2.2 Диаграмма разброса по переменной «Среднедушевые денежные доходы населения»



Рисунок 2.2.3 Диаграмма разброса по переменной «Доля населения с доходами ниже прожиточного минимума»

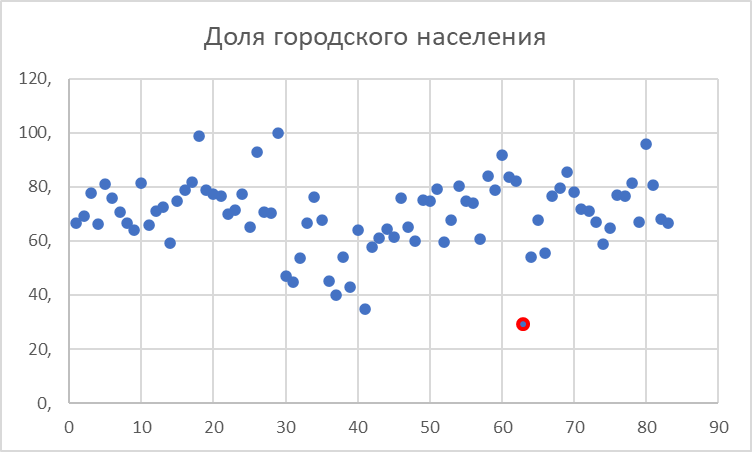


Рисунок 2.2.4 Диаграмма разброса по переменной «Доля городского населения»

Таким образом за 2013, 2014, 2015 годы аномальными признаны значения по следующим субъектам: г. Москва, Ненецкий АО, Ямало-Ненецкий АО, Чукотский АО, Республика Калмыкия, Республика Тыва, Республика Алтай.

**2.3. Корреляционный анализ**

Для исходного массива данных проведем корреляционный анализ. Используем функцию corr() из библиотеки для работы с данными на Python «Pandas».

Изображение выглядит как стол

Автоматически созданное описание

Рисунок 2.3.1 - Карта корреляции для данных за 2013 год

Также проверим коэффициенты корреляции на значимость. 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.

Изображение выглядит как стол

Автоматически созданное описание

Рисунок 2.3.2 Проверка коэффициентов корреляции на значимость 2013 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 все коэффициенты корреляции являются значимыми.

Для данных за 2014 и 2015 годы также проведем корреляционный анализ.

Изображение выглядит как стол

Автоматически созданное описание

Рисунок 2.3.3. Карта корреляции для данных за 2014 год

Изображение выглядит как стол

Автоматически созданное описание

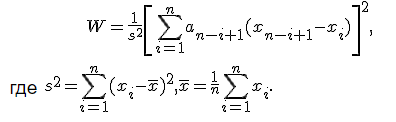
Рисунок 2.3.4. Карта корреляции для данных за 2015 год

## Между переменными не прослеживается такого явления, как мультиколлинеарность. Это значит, что нет линейной зависимости между объясняющими переменными регрессионной модели.

## **2.4. Проверка на нормальность распределения**

Реализуем тест Шапиро-Уилка для определения нормальности распределения.

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 =0,05. Обратимся к статистическому пакету «Gretl». В нем применим для каждого фактора «тест на нормальность».

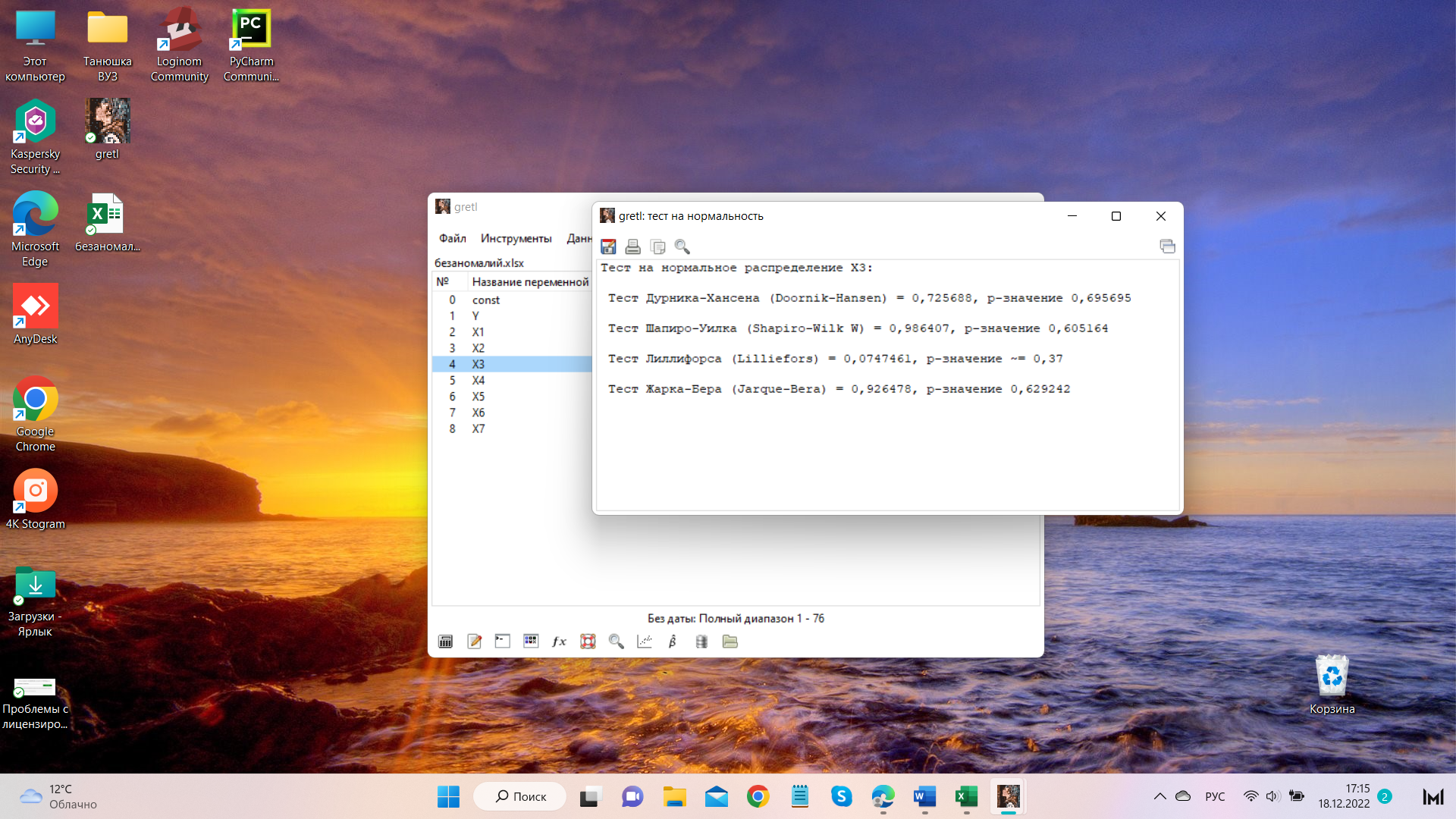


Рисунок 2.4. Тест на нормальное распределение для фактора Х3

Таким образом, нормальное распределение имеет только переменная Х3.

Ненормальность распределения может привести к неудовлетворительным результатам при построении регрессионных моделей. В таком случае необходимо прибегнуть к нормализации данных.

1. **МНОЖЕСТВЕННЫЙ ЛИНЕЙНЫЙ РЕГРЕССИОННЫЙ АНАЛИЗ**

Модель множественного линейного регрессионного анализа признака : ; где случайные величины (случайные эффекты влияния на результативный признак неконтролируемых факторов) независимы и имеют одинаковое нормальное распределение , или, иначе, наблюдения независимы и имеют нормальное распределение:

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



Рисунок 3.1 – Результат регрессионного анализа 2013 год

Оценки = 141151, =, = -424,4, = -224,6, , ,  -768,5, содержатся в результатах работы программы «Регрессия» в столбце «Коэффициенты» под заголовками «Y-пересечение», «», «», «», «», «», «», «» соответственно (см. рис. 3.1).

Таким образом, оценка линейной функции регрессии такова:



Рисунок 3.2 – Вывод остатков 2013 год

В таблице «Вывод остатка» содержится предсказанное — это , рассчитанные по построенному уравнению регрессии, и остатки — это разности ().

В «Регрессионная статистика» приведены:

– оценка коэффициента множественной линейной детерминации

0,665

– оценка коэффициента множественной линейной корреляции

0,816 (множественный R) — такова, судя по наблюдениям, степень линейной зависимости от ), , … , ;

– оценка нормированного коэффициента линейной детерминации

0,634 (нормированный R-квадрат) — в отличие от коэффициента , который при включении в имеющуюся линейную регрессионную модель дополнительного регрессора всегда увеличивается, нормированный коэффициент детерминации может и увеличиваться, и уменьшаться; чем больше , тем более адекватно уравнение регрессии.

– стандартная ошибка регрессии 25724,62 (Стандартная ошибка).

1) Проверим гипотезы при альтернативах ,

В построенном уравнении регрессии некоторые коэффициенты оказались незначимы, и такое уравнение нельзя считать приемлемым.

1. Исключим из уравнения регрессор , при котором коэффициент

незначим, а соответствующий этому коэффициенту, рассчитанный уровень значимости p = 0,9 является наибольшим.

Оценка линейной функции регрессии будет такой: .

2) Исключим из уравнения регрессор , при котором коэффициент незначим, а соответствующий этому коэффициенту, рассчитанный уровень значимости p = 0,5 является наибольшим.

Оценка линейной функции регрессии будет такой:

3) Исключим из уравнения регрессор , при котором коэффициент незначим, а соответствующий этому коэффициенту, рассчитанный уровень значимости p = 0,4 является наибольшим.

Оценка линейной функции регрессии будет такой:

Наилучшим уравнением является полученное на третьем шаге, поскольку и само уравнение, и все его коэффициенты значимы.

Таблица 3 – Модели до и после исключения аномалий

|  |  |  |  |
| --- | --- | --- | --- |
| Год | № модели | До/После исключения аномалий | Модель |
| 2013 | 1 | До |  |
|  | 2 | После |  |
|  | 3 | После |  |
|  | 4 | После |  |
|  | 5 | После |  |
| 2014 | 6 | До |  |
|  | 7 | После |  |
|  | 8 | После |  |
|  | 9 | После |  |
| 2015 | 10 | До |  |
|  | 11 | После |  |
|  | 12 | После |  |
|  | 13 | После |  |

1. **ДИСПЕРСИОННЫЙ АНАЛИЗ**

Проведем дисперсионный анализ для определения значимости уравнений регрессии. Обратимся к таблице «Дисперсионный анализ», которая выводится при построении уравнений регрессии в MS Excel.



Рисунок 4.1 – Дисперсионный анализ

Основные составляющие полной вариации объясняемой переменной:

значения которых, равные соответственно 9,8Е+10, 5,04Е+10 и 1,5Е+11 приводятся в столбце «SS»; а в столбце «MS» приведены значения величин , , равные соответственно 2,5Е+10 и 6,5Е+08.

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).

Таблица 4.1 – Дисперсионный анализ до исключения 2013 год

|  |  |  |  |
| --- | --- | --- | --- |
| Источник вариации | Показатели вариации | Число степеней свободы | F-статистика |
| Модельная |  |  | F=37,87 |
| Остаточная |  |  |
| Общая | 1,5Е+11 | n-1=82 |  |

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.



Рисунок 4.2 – Дисперсионный анализ после исключения аномальных значений 2013 год

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.



Рисунок 4.3 – Дисперсионный анализ после исключения аномальных значений 2013 год

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.



Рисунок 4.4 – Дисперсионный анализ после исключения аномальных значений 2013 год

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.



Рисунок 4.5 – Дисперсионный анализ до исключения аномальных значений 2014 год

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.



Рисунок 4.6 – Дисперсионный анализ после исключения аномальных значений 2014 год

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.



Рисунок 4.7 – Дисперсионный анализ после исключения аномальных значений 2014 год

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.



Рисунок 4.8 – Дисперсионный анализ после исключения аномальных значений 2014 год

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.



Рисунок 4.9 – Дисперсионный анализ до исключения аномальных значений 2015 год

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.



Рисунок 4.10 – Дисперсионный анализ после исключения аномальных значений 2015 год

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.



Рисунок 4.11 – Дисперсионный анализ после исключения аномальных значений 2015 год

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.



Рисунок 4.12 – Дисперсионный анализ после исключения аномальных значений 2015 год

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.

После построения уравнений регрессии, 13 моделей оказались значимыми. Для удобства они были занесены в таблицу с итоговыми значимыми моделями (см. табл. «Итоговая таблица значимых моделей»).

1. **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:

**5.1. Оценка остатков модели на гетероскедастичность**

Проверим свойство гетероскедастичности остатков значимых моделей. Выдвинем следующие гипотезы:

,

Для проверки используем тест Уайта. Построим вспомогательную регрессию на переменных значимой модели за 2013 год. 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. 5.1.



Рисунок 5.1 – Дисперсионный анализ для 2013 года

Результаты теста Уайта показывают отсутствие гетероскедастичности, так как при 5% уровне значимости Р-вероятность принятия гипотезы о наличии гетероскедастичности в остатках равна 0,81, что больше 0,05.

Проверим остальные значимые модели на гетероскедастичность в MS Excel. Выводы по результатам теста Уайта (см. Итоговая таблица значимых моделей), а графики разброса остатков (см. Приложение 5).

**5.2. 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. Посчитаем суммы, начиная от второго значения (см. табл. 5.1)

Таблица 5.1 - Вспомогательные расчеты для статистики Дарбина-Уотсона

|  |  |  |
| --- | --- | --- |
| e | (e(t)-e(t-1))^2 | e^2 |
| -24306,9 |  | 5,91E+08 |
| 8349,622 | 1066446953 | 69716189 |
| -2400,29 | 115560513,9 | 5761371 |
| 3308,018 | 32584724,72 | 10942981 |
| 5932,991 | 6890482,638 | 35200376 |
| -13096,1 | 362105131,6 | 1,72E+08 |
| 4562,15 | 311812735,1 | 20813216 |
| … | … | … |
| -4981,52 | 2219983232 | 24815548 |
| -63874,6 | 3468400180 | 4,08E+09 |

Посчитаем по формуле значение статистики Дарбина-Уотсона:

Определим интервалы неопределённости:

Из таблицы значений Дарбина-Уотсона

.

Интервалы равны:

.

Обратимся к интервалу определения автокорреляции (см. рис. 5.2)

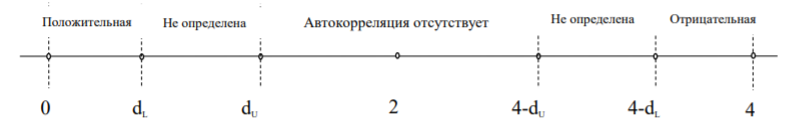


Рисунок 5.2 – Интервалы автокорреляции

По модели за 2013 год на исходных данных без аномальных значений можно сделать вывод, что автокорреляция отсутствует, так как попадает в промежуток . Остальные результаты будут занесены в итоговую таблицу значимых моделей.

1. **ВЫЯВЛЕНИЕ ЛУЧШИХ МОДЕЛЕЙ**

Исходные данные представляли собой наблюдения по федеральным округам РФ, где в качестве зависимой переменной был представлен оборот розничной торговли на душу населения (млн.руб.) с независимыми- среднедушевые денежные доходы населения (руб./месяц), - индекс потребительских цен (%), - доля оборота розничных торговых сетей в общем объеме оборота розничной торговли (%), - обеспеченность населения торговыми площадями современных форматов (м^2 на 1000 чел.), - доля населения с доходами ниже прожиточного минимума (%), - уровень безработицы (по методологии МОТ) (%), - доля городского населения (%).

Оценим параметры линейных регрессий всех моделей итоговой таблицы для каждого года. В качестве наилучшей модели будем искать ту, где отсутствует гетероскедастичность и автокорреляция, поскольку все модели имеют одинаково высокий коэффициент детерминации и F-значимость. Таким образом, для 2013 года подходит модель №4 из итоговой таблицы значимых уравнений. Уравнение регрессии модели:

Характеристики модели:

|  |  |  |  |
| --- | --- | --- | --- |
|  | F-значимость | Ст. ошибка | DW-статистика |
| 0,80 | 0,04 | 26020 | 2,15 |

Эта модель не имеет мультиколлинеарных переменных, автокорреляция отсутствует, присутствует гетероскедастичность, от которой необходимо избавиться.

Для устранения гетероскедастичности воспользуемся методом взвешенных наименьших квадратов. Необходимо рассчитать , , . Получим следующее преобразованное представление исходной однофакторной регрессии:

Для модели №4 разделим исходные векторы значений и на = 26013 и построим новое уравнение регрессии:

Условия гомоскедастичности и выполняются, следовательно, остатки гомоскедастичны.

Рассмотрим модели за 2014 год. Наилучшей моделью становится модель №8, она имеет уравнение:

Характеристики модели:

|  |  |  |  |
| --- | --- | --- | --- |
|  | F-значимость | Ст. ошибка | DW-статистика |
| 0,86 | 0,01 | 17510 | 1,44 |

Эта модель не имеет мультиколлинеарных переменных, автокорреляция отсутствует, присутствует гетероскедастичность, от которой необходимо избавиться.

Для модели №8 разделим исходные векторы значений и на = 17503 и построим новое уравнение регрессии:

Условия гомоскедастичности и выполняются, следовательно, остатки гомоскедастичны.

Для 2015 года лучшая модель №12. Её уравнение регрессии:

Характеристики модели:

|  |  |  |  |
| --- | --- | --- | --- |
|  | F-значимость | Ст. ошибка | DW-статистика |
| 0,80 | 0,04 | 21666 | 1,93 |

Эта модель не имеет мультиколлинеарных переменных, автокорреляция отсутствует, присутствует гетероскедастичность.

Для модели №8 разделим исходные векторы значений и на = 21659 и построим новое уравнение регрессии:

Условия гомоскедастичности и выполняются, следовательно, остатки гомоскедастичны.

Теперь все модели гомоскедастичны и автокорреляция в остатках отсутствует. Остатки отобранных моделей распределены по нормальному закону.

Проинтерпретируем выбранные модели.

Модель 2013 года:

* свободный член уравнения не может быть проинтерпретирован в условиях данной задачи;
* оборот розничной торговли зависит от среднеденежных доходов на душу населения. При росте среднеденежных доходов на душу населения на единицу, оборот розничной торговли вырастет на млн. руб.

Модель 2014 года:

* свободный член уравнения не может быть проинтерпретирован в условиях данной задачи;
* оборот розничной торговли зависит от среднеденежных доходов на душу населения. При росте среднеденежных доходов на душу населения на единицу, оборот розничной торговли вырастет на млн. руб.

Модель 2015 года:

* свободный член уравнения не может быть проинтерпретирован в условиях данной задачи;
* оборот розничной торговли зависит от среднеденежных доходов на душу населения. При росте среднеденежных доходов на душу населения на единицу, оборот розничной торговли вырастет на млн. руб.

Проведём анализ результатов эконометрического моделирования и попробуем осуществить прогнозирование по полученным моделям, сравним результаты предсказанного оборота розничной торговли с данными за следующий год.

1. **ПРОГНОЗИРОВАНИЕ**

**Оценка прогностических свойств моделей**

На основе выбранных моделей проведем прогнозирование и оценим его качество. Выберем лучшую модель из Таблицы 7.1.

Таблица 7.1 – Значения информационных и прогностических характеристик конкурирующих регрессионных моделей

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Модель |  |  |  |  |  |  |
|  | 80% | 82% | 80% | 87% | 88% | 81% |
|  | -64,03 | -43,78 | 42,12 | 391,9 | 418,66 | 432,78 |
|  | 320,46 | 312,23 | 292,45 | 464,61 | 459,84 | 454,92 |
| MSE | 248090,55 | 211771,04 | 223335,36 | 397581,2 | 384766,5 | 375678,4 |
|  | 56% | 47% | 35% | 59% | 169% | 197% |

Лучшей оказалась модель , т.к. она имеет наименьшую процентную ошибку прогноза.

**ЗАКЛЮЧЕНИЕ**

В процессе выполнения данной проектной работы была проделана работа по анализу и прогнозированию влияния санкционных ограничений на динамику товарных ресурсов в розничной торговле.

Были рассмотрены 83 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:

В результате санкционных ограничений товары стали дорожать из-за резко выросшего курса доллара, что повлекло уменьшение спроса на продукцию, следовательно оборот розничной торговли начал снижаться после 2014 года. Однако благодаря политики, проводимой правительством РФ по импортозамещению (производство товаров у нас в стране), себестоимость товаров начала уменьшаться, следовательно стал повышаться товарооборот.

**СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ**

1. Писарева О.М. Эконометрика: Начальный курс. Москва: Государственный университет управления, Институт информационных систем ГУУ, 2014.
2. Магнус Я.Р., Катышев П.К., Пересецкий А.А. Эконометрика: начальный курс. Москва: Дело, 2004.
3. Галочкин В. Т. Эконометрика: учебник и практикум для вузов. Москва: Издательство Юрайт, 2022.
4. Исходные данные // Федеральная служба государственной статистики URL: <https://rosstat.gov.ru> (дата обращения: 01.12.2022)
5. Колемаев, В.А. Эконометрика. Москва: Инфра-М, 2005.
6. Frank E. Grubbs. Extension of Sample Sizes and Percentage Points for Significance Tests of Outlying Observations / Frank E. Grubbs, Glenn Beck. // Technometrics, 1972.- Vol. 14.- № 4.- P. 847-854.
7. <URL:https://mnv.irgups.ru/sites/default/files/articles_pdf_files/tkachenko.pdf> (дата обращения: 04.12.2022)
8. <URL:https://rosstat.gov.ru/storage/mediabank/Ponyatiya_i_opredeleniya_Tovarnye_resursy.pdf> (дата обращения: 04.12.2022)
9. <URL:https://www.booksite.ru/localtxt/vei/tsl/evy/osn/ovy/text.pdf?ysclid=lbl9frmvnt218267020> (дата обращения: 04.12.2022)
10. URL: <https://www.fedstat.ru/?ysclid=lbl9tyk815623429637> (дата обращения: 04.12.2022)

Таблица 8 – Итоговая таблица значимых моделей

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Год | № | Уравнение регрессии | Характеристики качества | | | | | Описание модели |
|  |  |  |  | F-значимость | Ст. ошибка | DW-статистика | Выводы тестирования |  |
| 2013 | 1 |  | 0,91 | 0,00 | 18698 | 2,13 | * Гетероскедастичность отсутствует * Автокорреляция отсутствует | Исходные данные за 2013 год, n=83, до удаления аномальных значений |
|  | 2 |  | 0,85 | 0,00 | 23345 | 1,80 | * Гетероскедастичность присутствует * Автокорреляция отсутствует | Исходные данные за 2013 год, n=73, после удаления аномальных значений |
|  | 3 |  | 0,84 | 0,04 | 24193 | 1,88 | * Гетероскедастичность присутствует * Автокорреляция отсутствует | Исходные данные за 2013 год, n=73, после удаления аномальных значений |
|  | 4 |  | 0,80 | 0,04 | 26020 | 2,15 | * Гетероскедастичность присутствует * Автокорреляция отсутствует | Исходные данные за 2013 год, n=73, после удаления аномальных значений |

Продолжение итоговой таблицы значимых моделей

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 2013 | 5 |  | 0,69 | 0,02 | 31362 | 1,69 | * Гетероскедастичность присутствует * Автокорреляция отсутствует | Исходные данные за 2014 год, n=73, после удаления аномальных значений |
| 2014 | 6 |  | 0,91 | 0,00 | 20771 | 1,32 | * Гетероскедастичность присутствует * Автокорреляция не определена | Исходные данные за 2014 год, n=83, до удаления аномальных значений |
| 7 |  | 0,92 | 0,03 | 13477 | 2,01 | * Гетероскедастичность присутствует * Автокорреляция отсутствует | Исходные данные за 2014 год, n=73, после удаления аномальных значений |
| 8 |  | 0,86 | 0,01 | 17510 | 1,44 | * Гетероскедастичность присутствует * Автокорреляция отсутствует | Исходные данные за 2015 год, n=73, после удаления аномальных значений |
| 9 |  | 0,62 | 0,00 | 26660 | 1,8 | * Гетероскедастичность присутствует * Автокорреляция не определена | Исходные данные за 2015 год, n=73, после удаления аномальных значений |

Продолжение итоговой таблицы значимых моделей

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 2015 | 10 |  | 0,80 | 0,02 | 28378 | 2,44 | * Гетероскедастичность присутствует * Автокорреляция не определена | Исходные данные за 2015 год, n=83, до удаления аномальных значений |
| 11 |  | 0,85 | 0,01 | 18667 | 2,00 | * Гетероскедастичность присутствует * Автокорреляция отсутствует | Исходные данные за 2015 год, n=73, после удаления аномальных значений |
| 12 |  | 0,79 | 0,04 | 21666 | 1,93 | * Гетероскедастичность присутствует * Автокорреляция отсутствует | Исходные данные за 2015 год, n=73, после удаления аномальных значений |
| 13 |  | 0,63 | 0,00 | 27579 | 1,36 | * Гетероскедастичность присутствует * Автокорреляция положительная | Исходные данные за 2015 год, n=73, после удаления аномальных значений |

**ПРИЛОЖЕНИЕ 1 - Исходные данные**

Таблица 1.1 - Исходные данные за 2013 год



Таблица 1.2 - Исходные данные за 2014 год



Таблица 1.3 - Исходные данные за 2015 год



**ПРИЛОЖЕНИЕ 2 – Графики разброса, корреляция, описательные статистики**

Листинг 2.1 – Код для построения графиков:

sns.pairplot(DF\_2013[['Y', 'X1', 'X2’, 'X3 ']])

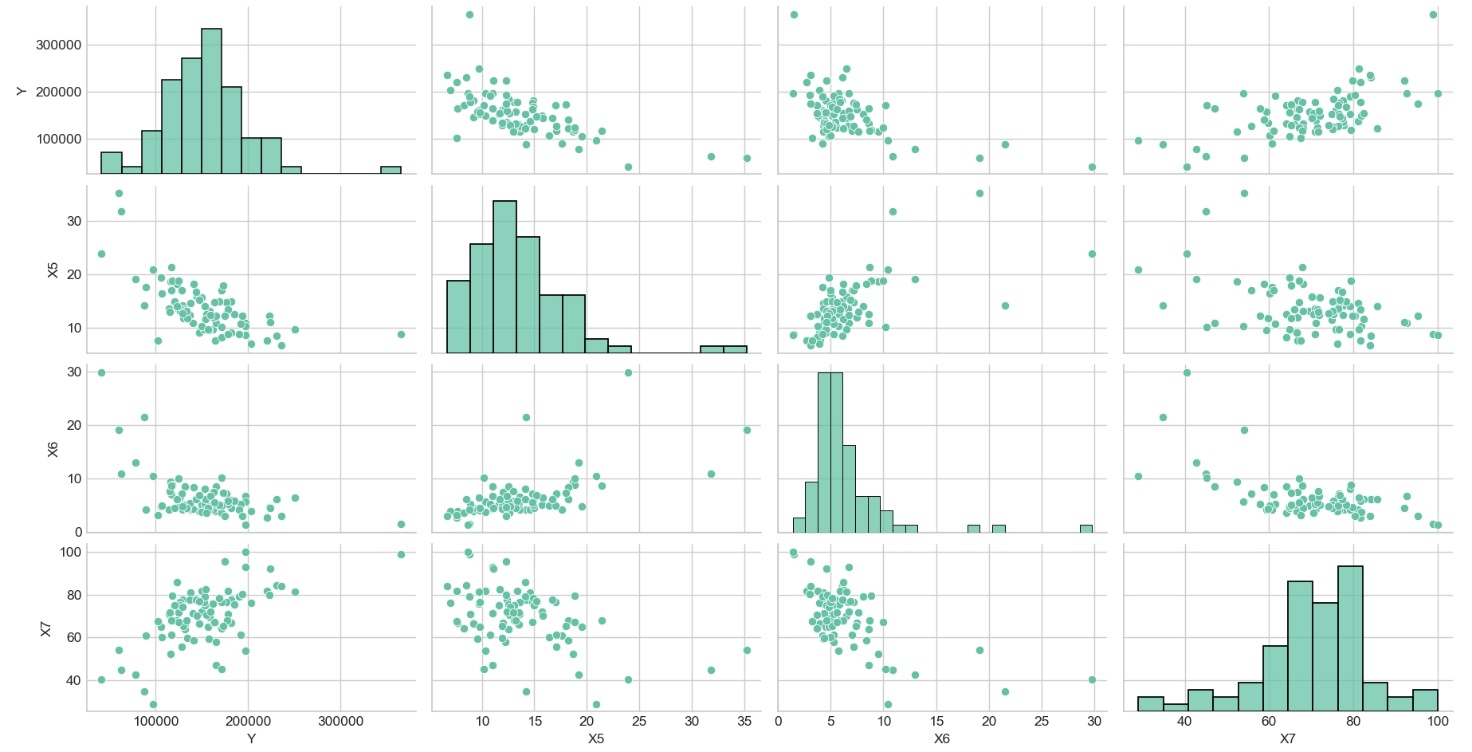
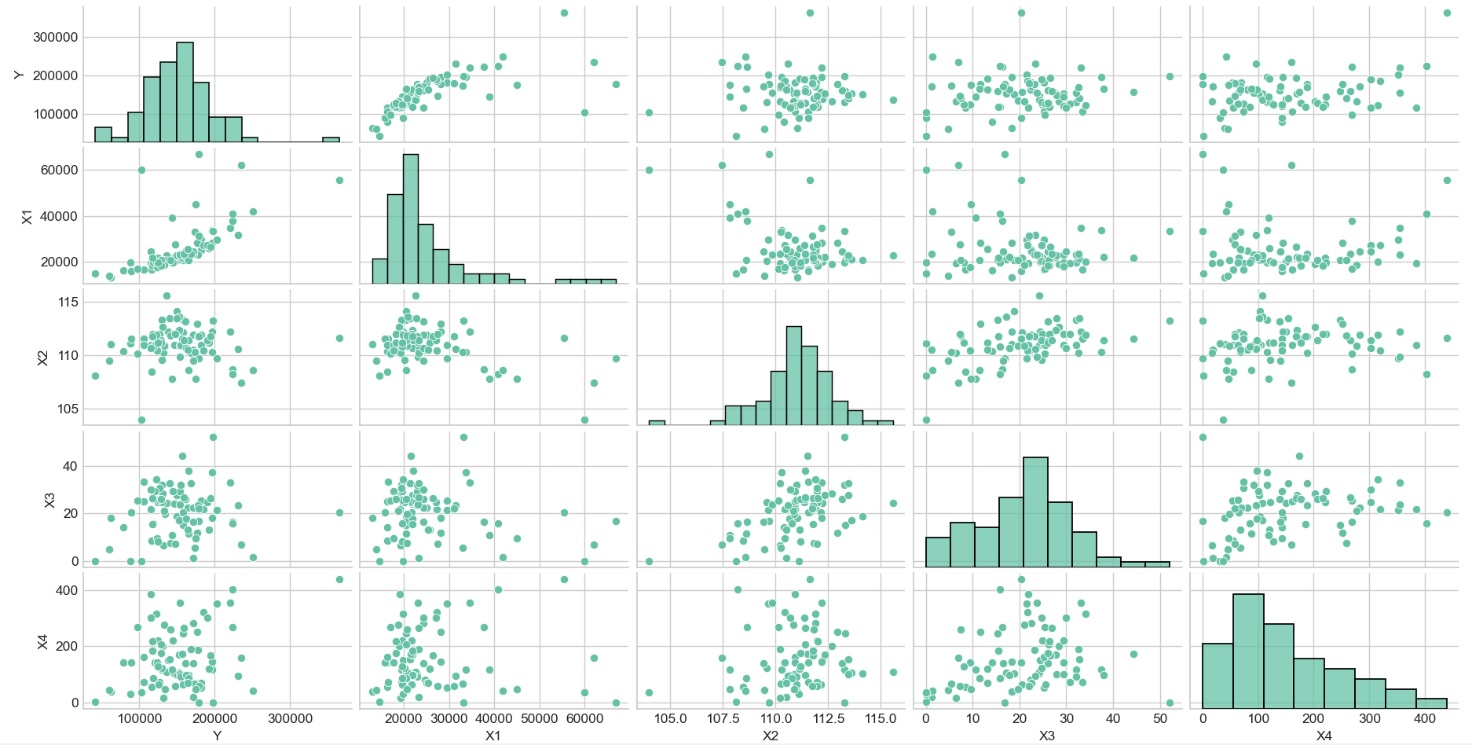
****

Рисунок 2.1 – Гистограммы и графики разброса 2014 год

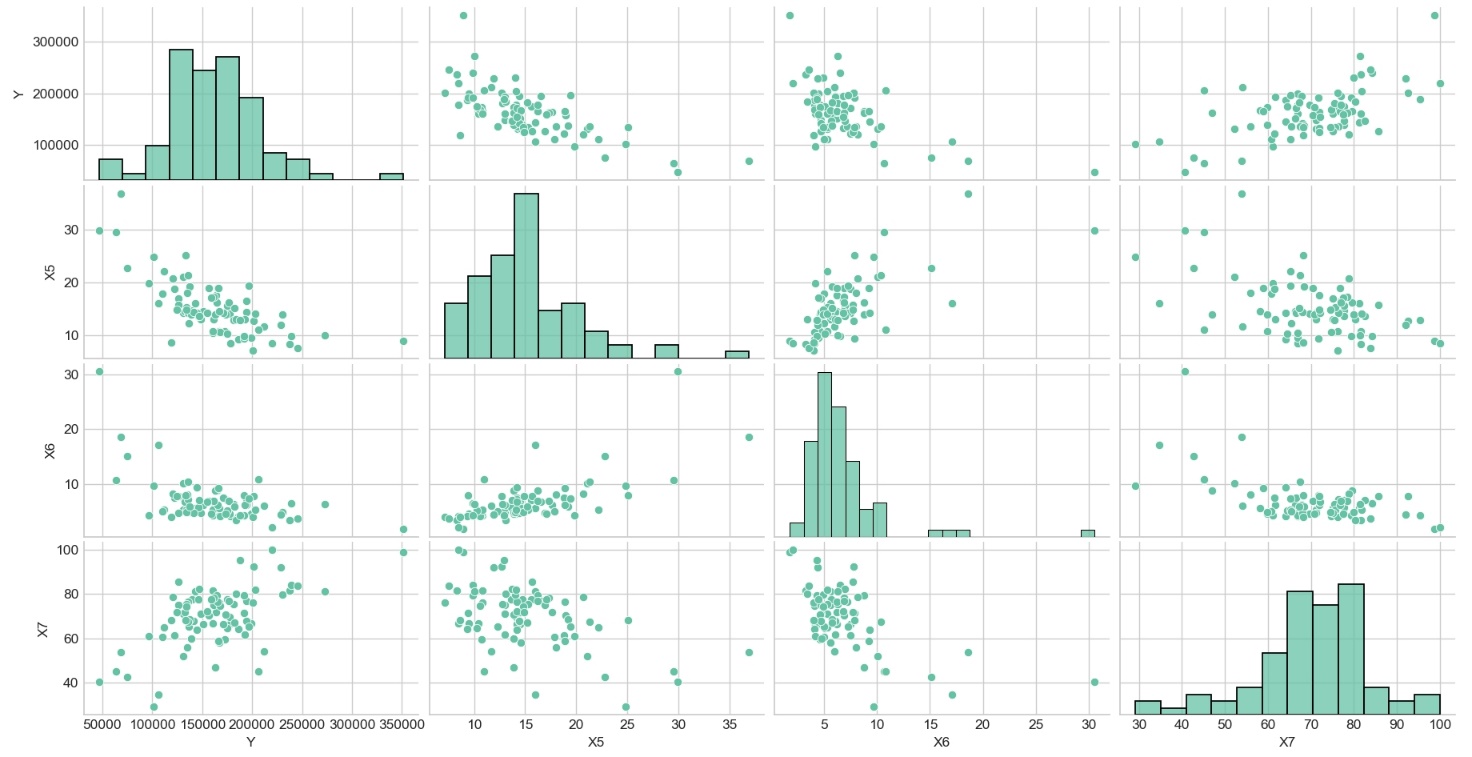
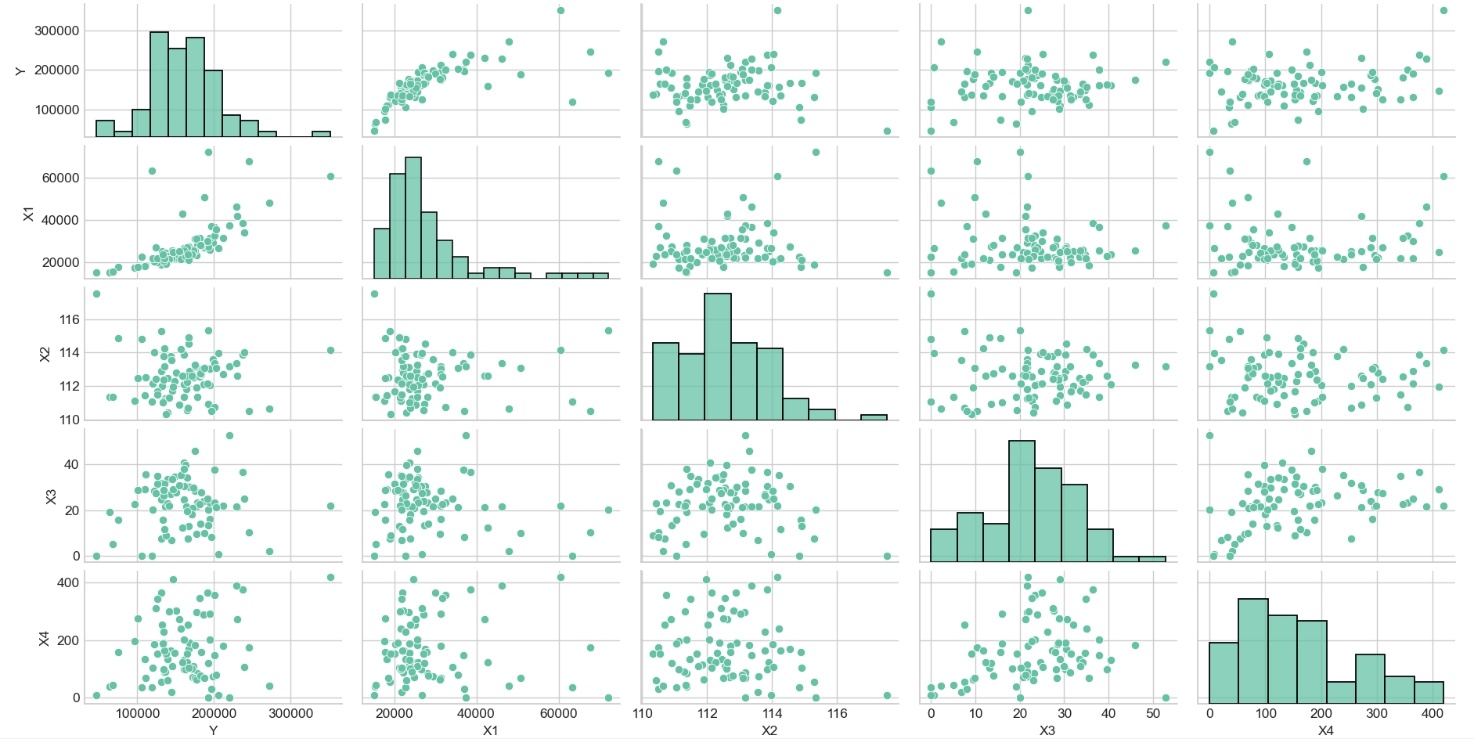


Рисунок 2.2 – Гистограммы и графики разброса 2015 год

Листинг 2.2. Корреляция

corr\_mat = DF\_2013[['Y', 'X1', 'X2’, 'X3 ', ' X4 ', 'X5 ', ' X6 ', ' X7 ']].corr()

corr\_mat.style.background\_gradient().set\_precision(3)

Листинг 2.3. Проверка на значимость коэффициентов корреляции

df = corr\_mat

df = df.where(np.triu(np.ones(df.shape), k=1).astype(bool))

df = df.stack().reset\_index()

df.columns = ['Показатель\_1', 'Показатель\_2', 'Корреляция']

t\_stat = np.sqrt((len(DF\_2013.index) - 2) / (1 - df.Корреляция.values \*\* 2))

df['P\_значение'] = 2\*(1 - t.cdf(abs(t\_stat), len(DF\_2013.index) - 2))

df.loc[df['P\_значение'] < 0.05].sort\_values('Корреляция')

**Изображение выглядит как стол

Автоматически созданное описание**

Рисунок 2.3 - Проверка коэффициентов корреляции на значимость 2014 год

Изображение выглядит как стол

Автоматически созданное описание

Рисунок 2.4 - Проверка коэффициентов корреляции на значимость 2015 год

**ПРИЛОЖЕНИЕ 4 – Регрессионный анализ**



Рисунок 4.1 – Регрессионный анализ до исключения аномальных значений 2013 год



Рисунок 4.2 – Регрессионный анализ после исключения аномальных значений 2013 год



Рисунок 4.3– Регрессионный анализ после исключения аномальных значений 2013 год



Рисунок 4.4 – Регрессионный анализ после исключения аномальных значений 2013 год



Рисунок 4.5 – Регрессионный анализ после исключения аномальных значений 2013 год



Рисунок 4.6 – Регрессионный анализ до исключения аномальных значений 2014 год



Рисунок 4.7 – Регрессионный анализ после исключения аномальных значений 2014 год



Рисунок 4.8 – Регрессионный анализ после исключения аномальных значений 2014 год



Рисунок 4.9 - Регрессионный анализ после исключения аномальных значений 2014 год



Рисунок 4.10 – Регрессионный анализ до исключения аномальных значений 2015 год



Рисунок 4.11 – Регрессионный анализ после исключения аномальных значений 2015 год



Рисунок 4.12 – Регрессионный анализ после исключения аномальных значений 2015 год



Рисунок 4.13 – Регрессионный анализ после исключения аномальных значений 2015 год

**ПРИЛОЖЕНИЕ 6 – Графики разброса остатков**

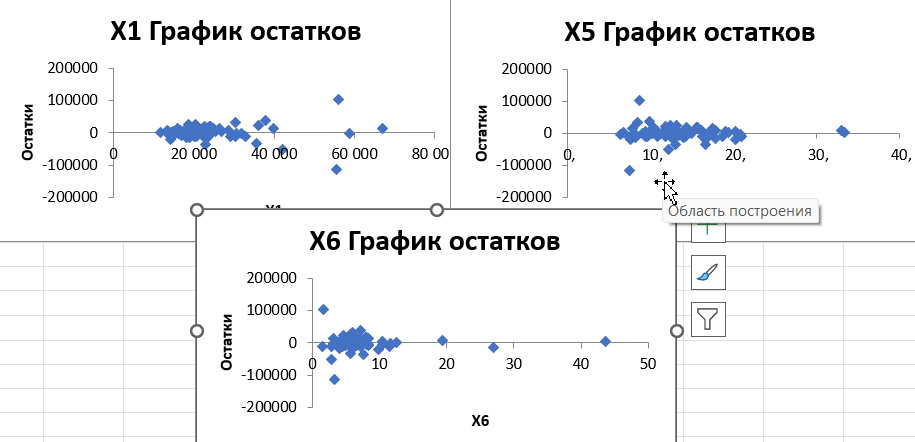


Рисунок 6.1.1 – Графики разброса остатков для модели №1 (итоговая таблица значимых моделей) 2013 года.

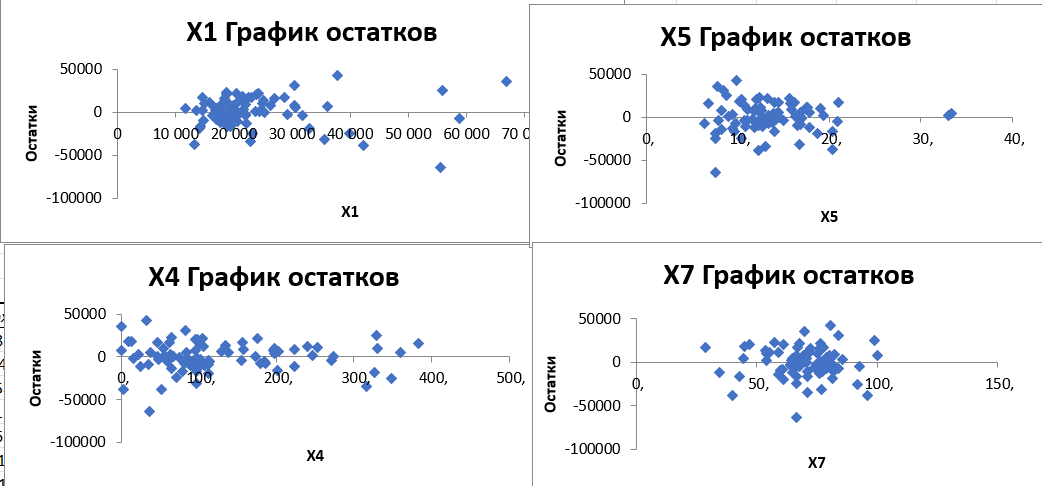


Рисунок 6.1.2 – Графики разброса остатков для модели №2 (итоговая таблица значимых моделей) 2013 года.

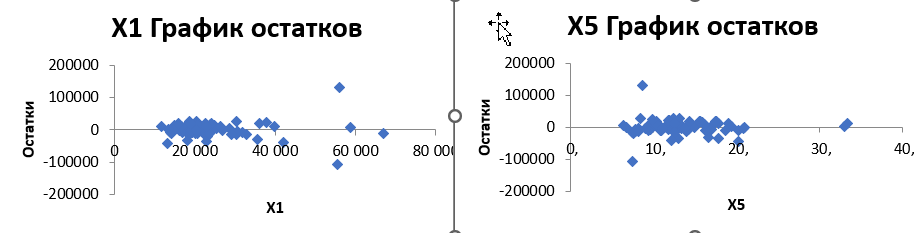


Рисунок 6.1.3 – Графики разброса остатков для модели №3 (итоговая таблица значимых моделей) 2013 года.

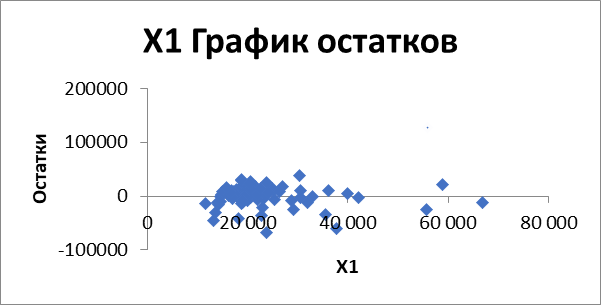


Рисунок 6.1.4 – Графики разброса остатков для модели №4 (итоговая таблица значимых моделей) 2013 года.



Рисунок 6.1.5 – Графики разброса остатков для модели №5 (итоговая таблица значимых моделей) 2013 года.

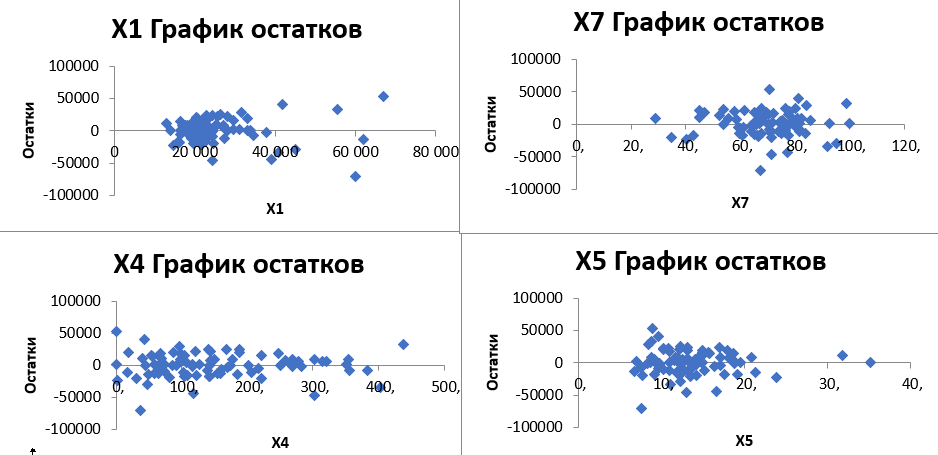


Рисунок 6.1.6 – Графики разброса остатков для модели №6 (итоговая таблица значимых моделей) 2014 года.

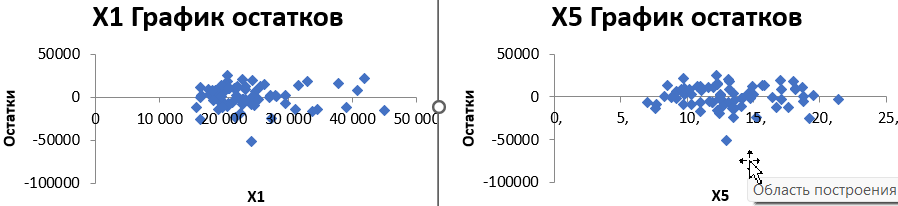


Рисунок 6.1.7 – Графики разброса остатков для модели №7 (итоговая таблица значимых моделей) 2014 года.

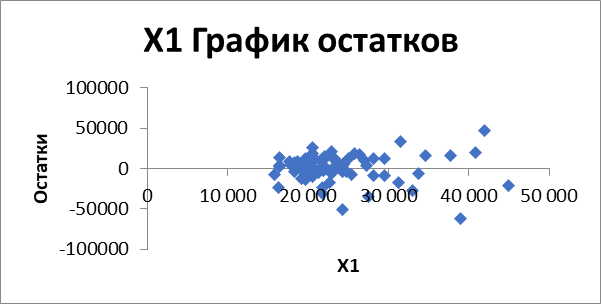


Рисунок 6.1.8 – Графики разброса остатков для модели №8 (итоговая таблица значимых моделей) 2014 года.

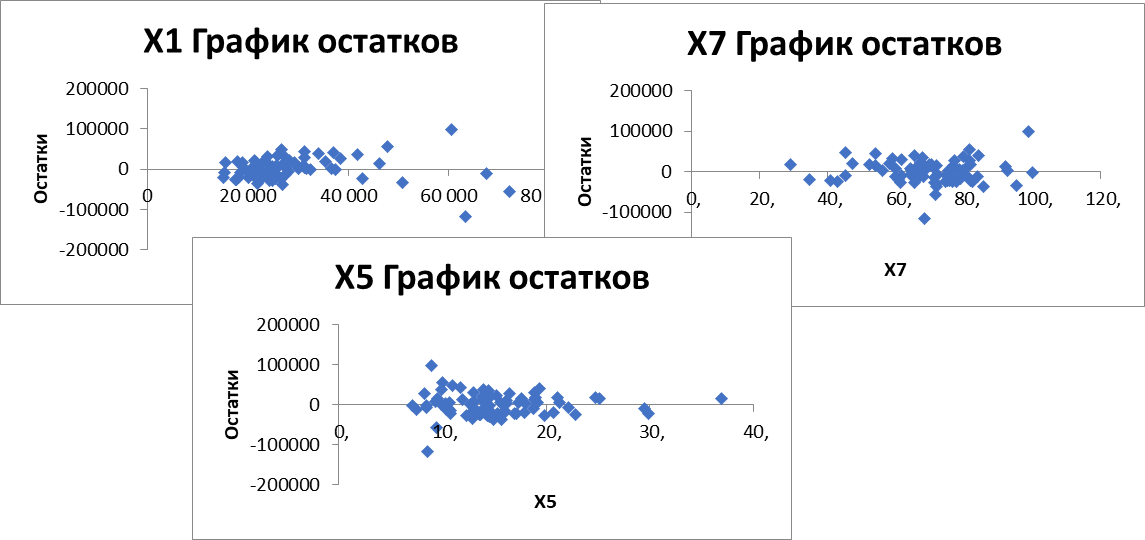


Рисунок 6.1.10 – Графики разброса остатков для модели №10 (итоговая таблица значимых моделей) 2014 года.

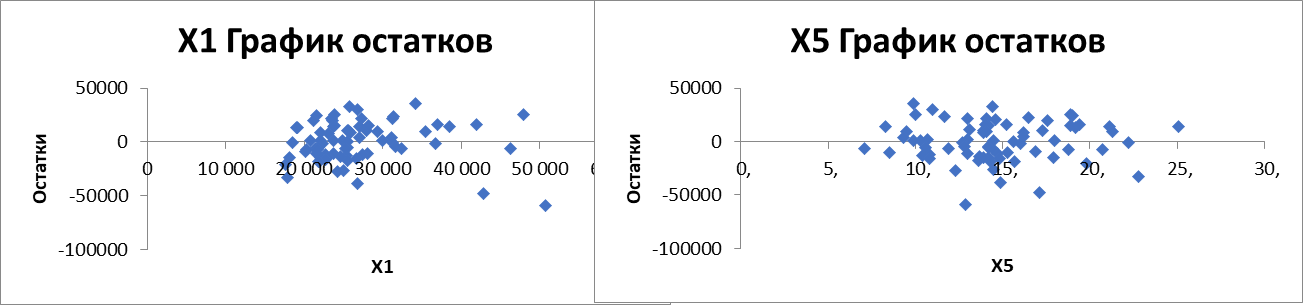


Рисунок 6.1.11 – Графики разброса остатков для модели №11 (итоговая таблица значимых моделей) 2015 года.



Рисунок 6.1.12 – Графики разброса остатков для модели №12 (итоговая таблица значимых моделей) 2015 года.



Рисунок 6.1.13 – Графики разброса остатков для модели №13 (итоговая таблица значимых моделей) 2015 года.