**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 различных факторов на мировое производство сельскохозяйственной продукции» | |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Направление подготовки | 01.03.02 |  | Прикладная математика и информатика | |
| (код) |  | (наименование) | |
| Образовательная программа | Прикладная математика и информатика | | |
| (название образовательной программы) | | |
| Обучающийся | Тарарин А. В. | | |
| (Фамилия, Имя, Отчество) | | |
| 3 курс, 1 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: | «\_\_\_» \_\_\_\_\_\_\_\_\_\_\_\_\_2023 г. |

Москва – 2023

СОДЕРЖАНИЕ

ВВЕДЕНИЕ 3

1 ПОСТАНОВКА ЗАДАЧИ ИССЛЕДОВАНИЯ И СБОР ДАННЫХ………….6

[2 ПРЕДВАРИТЕЛЬНЫЙ АНАЛИЗ МАССИВА ДАННЫХ И ЕГО ПРЕДМОДЕЛЬНАЯ ОБРАБОТКА 8](#_Toc153915340)

2.1 Визуальный анализ 9

[2.2 Поиск аномальных значений 11](#_Toc153915341)

[2.3 Корреляционный анализ 14](#_Toc153915342)

[2.4 Описательная статистика 21](#_Toc153915343)

[2.5 Проверка на нормальность распределения 21](#_Toc153915344)

[3. МНОЖЕСТВЕННЫЙ ЛИНЕЙНЫЙ РЕГРЕССИОННЫЙ АНАЛИЗ 23](#_Toc153915345)

[4. ДИСПЕРСИОННЫЙ АНАЛИЗ 32](#_Toc153915346)

[5. ПРОВЕРКА ДОПОЛНИТЕЛЬНЫХ ИНФОРМАЦИОННЫХ И ПРОГНОСТИЧЕСКИХ СВОЙСТВ УРАВНЕНИЙ РЕГРЕССИИ 36](#_Toc153915347)

[5.1. Оценка остатков модели на гетероскедастичность 36](#_Toc153915348)

[5.2. Оценка автокорреляции остатков модели 37](#_Toc153915349)

[6. ВЫЯВЛЕНИЕ ЛУЧШИХ МОДЕЛЕЙ 40](#_Toc153915350)

7 ПРОГНОЗИРОВАНИЕ……………………………...………………………...43

## 7.1 Оценка прогностических свойств моделей………………………….43

[ИТОГОВАЯ ТАБЛИЦА ЗНАЧИМЫХ МОДЕЛЕЙ 47](#_Toc153915352)

[ПРИЛОЖЕНИЕ 1 - Исходные данные 52](#_Toc153915353)

[ПРИЛОЖЕНИЕ 2 58](#_Toc153915354)

[ПРИЛОЖЕНИЕ 3 63](#_Toc153915355)

[СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ 69](#_Toc153915356)

**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 пред модельной 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;

– 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 решение 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) Оценка параметров конкурирующих моделей, их верификация и оценка прогностических свойств.

4) Прогнозирование на основе факторной регрессионной модели.

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

**Актуальность работы**

Сельское хозяйство остается ключевым компонентом мировой экономики, несмотря на динамику урбанизации и индустриализации. Эконометрическое исследование, направленное на анализ влияния различных факторов на стоимость Валового сельскохозяйственного продукта (ВСП) в 50 странах, представляет собой значимый вклад в понимание динамики сельского хозяйства в глобальном масштабе. В данной секции рассмотрим актуальность темы и определим потенциальных заказчиков исследования.

1. Глобальная Зависимость от Сельского Хозяйства:

Сельское хозяйство остаётся фундаментальным элементом экономики многих стран, особенно развивающихся. Зависимость от сельского сектора сохраняется в силу необходимости обеспечения продовольственной безопасности и поддержания устойчивого сельского образа жизни.

2. Проблемы Продовольственной Безопасности:

В условиях растущего населения и изменяющегося климата обеспечение продовольственной безопасности становится ключевой проблемой. Исследование напрямую касается факторов, влияющих на ВСП, и, таким образом, имеет непосредственное отношение к проблемам продовольственной безопасности в мировом масштабе.

3. Экономическое Развитие и Сельское Хозяйство:

Анализ взаимосвязи стоимости ВСП с факторами, такими как здоровое питание, инвестиции, кредиты, может предоставить важные показатели о роли сельского хозяйства в экономическом развитии страны. Это актуально в свете стремления к созданию устойчивых экономических моделей.

4. Роли Государства и Инвестиций:

Государственные инвестиции и выделение кредитов на сельское хозяйство отражают роль государства в поддержке данного сектора. Выводы могут стать важным инструментом для разработки эффективных политик по развитию сельского хозяйства и использованию ресурсов.

5. Потенциальные Заказчики и Интересующиеся Стороны:

Исследование может заинтересовать правительственные органы, такие как министерства сельского хозяйства и экономического развития. Международные организации, занимающиеся продовольственной безопасностью, а также ассоциации сельскохозяйственных предприятий, будут заинтересованы в результатах вашего исследования.

6. Актуальность Данных для Принятия Решений:

Результаты могут стать ценным источником данных для принятия экономических и политических решений в сфере сельского хозяйства. Политики, ориентированные на устойчивое развитие, могут использовать эти выводы для формирования стратегий развития сельскохозяйственного сектора.

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

Для проведения эконометрического моделирования была сформулирована задача построения факторной регрессионной модели для прогнозирования валовой стоимости сельскохозяйственной продукции в странах по всему миру.

Для начала дадим определения терминам:

Валовая стоимость сельскохозяйственной продукции — стоимость всей произведенной продукции за год.

Стоимость здорового питания — показатель стоимости и доступности здорового питания оцениваются в каждой стране и показывают физический и экономический доступ населения к наименее дорогим местным продуктам питания, отвечающим требованиям здорового питания, как определено в руководящих принципах по питанию на основе пищевых продуктов.

Государственные вложения — это денежных средств государством в следующие отрасли сельского хозяйства — лесное хозяйство, рыболовство, защита природы, растениеводство, животноводство.

Индекс цен производителей — это индекс цен производителей сельскохозяйственной продукции, который измеряет среднегодовое изменение с течением времени отпускных цен, получаемых фермерами (цены на выходе с фермы или в первой торговой точке).

Земля, оборудованная для полива - земельный участок, оснащенный ирригационной инфраструктурой и оборудованием для обеспечения водой сельскохозяйственных культур, которые находятся в рабочем состоянии. Оборудование не обязательно должно использоваться в течение отчетного года.

Сельскохозяйственная площадь – пахотные земли, временные посевы, общая сельскохозяйственная площадь.

Для отбора данных был использован международный ресурс «ФАОСТАТ», предоставляющий статистику по разным странам. В качестве независимых переменных были выбраны:

1. Стоимость здорового питания;

2. Государственные вложения;

3. Индекс цен производителей;

4. Земля, оборудованная для полива;

5. Сельскохозяйственная площадь.

Исходная статистика данных за 2017, 2018, 2019 годы по 49 странам приводится в Приложении 1.

Обоснования выбора для исследования 2017, 2018 и 2019 года.

Стабильность Экономического Климата:

В выбранные года мировая экономика и сельское хозяйство находились в относительно стабильном состоянии. Это обстоятельство позволяет лучше оценить влияние рассматриваемых факторов на Валовой сельскохозяйственный продукт без существенных внешних воздействий, таких как экономические кризисы.

Доступность Данных:

Данные за выбранный период более легко доступны и полны, что обеспечивает более надежный и полный анализ. Это важно для обеспечения качественного исследования, основанного на достоверных информационных источниках.

Отражение Реальной Ситуации:

2017, 2018 и 2019 годы могут отражать более реалистичные тенденции и состояние сельского хозяйства, поскольку они предшествуют годам 2020 и 2021, которые были затронуты глобальной пандемией COVID-19. Это позволяет избежать искажений данных, вызванных влиянием пандемии.

На основе данных об экономическом состоянии 49 стран проверим гипотезу о влиянии выбранных независимых факторов, упомянутых выше, на уровень валовой стоимости сельскохозяйственной продукции.

**2 ПРЕДВАРИТЕЛЬНЫЙ АНАЛИЗ МАССИВА ДАННЫХ И ЕГО ПРЕДМОДЕЛЬНАЯ ОБРАБОТКА**

Для удобства введем обозначения для исследуемых показателей и единицы измерения (см. табл. 2.1).

Таблица 2.1 - Показатели и единицы измерения

|  |  |  |
| --- | --- | --- |
| Показатель | Обозначение | Единицы измерения |
| Валовая стоимость сельскохозяйственной продукции | GROSS, Y | 1000 долларов США |
| Стоимость здорового питания | HEALTHY, X1 | долларов США на человека в день |
| Государственные вложения | GOVINVEST, X2 | миллионы долларов США |
| Индекс покупательских цен | PPP, X3 | Пункты |
| Площадь земли об. для орошения | EQAREA, X4 | 1000 гектар |
| Сельскохозяйственная площадь | AGRAREA, X5 | 1000 гектар |
| Сумма кредитов | CREDITS, X6 | миллионов долларов США |

Стоимость Здорового Питания:

Обоснование: Здоровое питание является ключевым аспектом человеческого благополучия и продуктивности. Взаимосвязь между стоимостью ВСП и стоимостью здорового питания может подчеркнуть влияние пищевых стандартов и потребительских предпочтений на сельское хозяйство.

Государственные Вложения:

Обоснование: Государственные инвестиции являются важным фактором в развитии сельского хозяйства. Исследование этого влияния может помочь определить, какие виды государственных вложений оказывают наибольший положительный эффект на ВСП.

Индекс Цен Производителей:

Обоснование: Индекс цен производителей отражает изменения в ценах на сельскохозяйственную продукцию и может служить индикатором эффективности сельскохозяйственного сектора. Анализ этого показателя позволяет выявить влияние цен на производство на общую стоимость ВСП.

Земля, Оборудованная для Полива:

Обоснование: Полив обеспечивает устойчивость сельского хозяйства и повышает урожайность. Исследование влияния земель, оборудованных для полива, на стоимость ВСП может выявить связь между доступностью воды и производственной эффективностью.

Сельскохозяйственная Площадь:

Обоснование: Размер сельскохозяйственной площади напрямую связан с объемом производства. Анализ влияния сельскохозяйственной площади на стоимость ВСП может выявить, как объем производства влияет на экономическую производительность.

Общее обоснование выбора этих факторов заключается в их потенциальной способности влиять на различные аспекты сельского хозяйства, от производства до потребления, и, следовательно, на общую стоимость ВСП. Исследование этих переменных может помочь выявить ключевые факторы, формирующие экономику сельского хозяйства в различных странах мира.

* 1. **Визуальный анализ**

Далее будет проделан визуальный анализ данных:

С помощью Microsoft Excel проанализируем разброс переменных и сделаем соответствующие выводы по графикам.

Обратимся к программе для работы с электронными таблицами Excel. В ней воспользуемся построением графиков «Ящик с усами» (см. рис. 2.1.1).

Изображение выглядит как линия, снимок экрана, число, текст

Автоматически созданное описание

Изображение выглядит как текст, число, линия, чек

Автоматически созданное описание

Изображение выглядит как текст, число, линия, снимок экрана

Автоматически созданное описание

Изображение выглядит как текст, снимок экрана, число, Шрифт

Автоматически созданное описание

Рисунок 2.1.1 - «Ящик с усами» (2017 год)

Вывод по коробчатым диаграммам за 2017 год следующий: в данных присутствуют аномальные значения. Эти аномальные значения смещают значения среднего и медианы по выборке. Эти значения нужно проверить под известным критериям.

По диаграммам за 2018 и 2019 год выводы аналогичны.

2.2 Поиск аномальных значений

Поскольку при визуальном анализе возникли предположения о наличии аномальных значений, выполним их поиск по методу Граббса. Для этого нужно посчитать описательную статистику по каждой переменной. Рассчитать значения статистики Граббса и сравнить с критическим.

Таблица 2.2 - Расчетные значения

|  |  |  |
| --- | --- | --- |
|  | Среднее | Стандартное отклонение |
| Валовая стоимость продукции,  тыс. долларов США | 61136688,82 | 149442769,7 |
| Стоимость здорового питания,  долларов на человека в день | 3,170102041 | 0,617527302 |
| Государственные вложения,  миллионы долларов США | 427071,8994 | 1109706,991 |
| Индекс покупательских цен,  пункты | 105,5471429 | 10,36299627 |
| Площадь земли об. для орошения,  1000 гектар | 5377,373927 | 14765,22938 |
| Сельскохозяйственная площадь,  1000 гектар | 58499,92171 | 115359,6225 |
| Сумма кредитов,  миллионов долларов США | 709837,0266 | 2075718,505 |

Для переменной «Валовая стоимость продукции» по графику (см. рис. 2.2.1) видим подозрительную точку для США, Китая и Индии. Проверим их на аномальность.

Рисунок 2.2.1 - Диаграмма разброса по переменной «Валовая стоимость продукции»

Для США

Для Индии , где – значение показателей по рассматриваемым странам, – среднее по выборке, – выборочное среднеквадратическое отклонение.

Cледовательно, принимаем гипотезу . Таким образом, наблюдения аномальны и их нужно удалить из выборки.

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

Рисунок 2.2.2 - Диаграмма разброса по переменной «Стоимость здорового питания»

Рисунок 2.2.3 - Диаграмма разброса по переменной «Сельскохозяйственная площадь»

Таким образом, за 2017 год аномальными признаны значения по следующим странам: Австралия, Ангола, Гренада, США, Индия, Китай.

Аналогично за 2018 год аномальными признаны значения по следующим странам: Австралия, Ангола, Гренада, США, Индия, Китай. За 2019 год аномальными признаны значения по следующим странам: Таиланд, Албания, Индонезия, Сербия, Бразилия, США, Индия.

## **Корреляционный анализ**

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

Листинг 2.2. Корреляция

corr\_mat = df[['GROSS', 'HEALTHY', 'GOVINVEST', 'PPP', 'EQAREA', 'AGRAREA', 'CREDITS']].corr()

plt.figure(figsize=(10, 8))

sns.heatmap(corr\_mat, annot=True, cmap='coolwarm')

plt.title('Correlation Matrix Heatmap')

plt.show()

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:

(1)

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

Изображение выглядит как снимок экрана, Прямоугольник, прямоугольный

Автоматически созданное описание

Рисунок 2.3.1 - Карта корреляции для данных за 2017 год

Листинг 2.3. Проверка на значимость коэффициентов корреляции

corr\_mat = df[['GROSS', 'HEALTHY', 'GOVINVEST', 'PPP', 'EQAREA', 'AGRAREA', 'CREDITS']].corr()

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.index) - 2) / (1 - df['Корреляция'].values \*\* 2))

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

significant\_corr = df.loc[df['P\_значение'] < 0.05].sort\_values('Корреляция', ascending=False)

Изображение выглядит как текст, меню, снимок экрана, Шрифт

Автоматически созданное описание

Рисунок 2.3.2 - Проверка коэффициентов корреляции на значимость (2017 год)

Поскольку уровень значимости равен 0,05, значимыми будут считаться коэффициенты, чьи p-значения окажутся меньше, чем 0,05. Судя по полученным результатам, все коэффициенты корреляции являются значимыми.

Для данных за 2018 и 2019 годы также проведем корреляционный анализ.

Между переменными 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. Уберем переменные, где корреляция больше 0,7. Таким образом, для 2017 года оставляем переменные:

* Стоимость здорового питания, долларов на человека в день (Х1);
* Государственные вложения, миллионы долларов США (Х2);
* Индекс покупательских цен,  
   пункты (Х3);
* Площадь земли об. для орошения,  
   1000 гектар (Х4);
* Сельскохозяйственная площадь,  
   1000 гектар(Х5).

Выделим также независимые между собой переменные для 2018 и 2019 годов.

В итоге получаем переменные для моделей.

Таблица 2.3 - Переменные для моделей по годам

|  |  |
| --- | --- |
| Год | Независимые переменные |
| 2017 | Стоимость здорового питания (Х1), Государственные вложения (Х2), Индекс покупательских цен (Х3), Площадь земли об. для орошения (Х4), Сельскохозяйственная площадь |

Продолжение таблицы 2.3

|  |  |
| --- | --- |
| 2018 | Стоимость здорового питания (Х1), Государственные вложения (Х2), Индекс покупательских цен (Х3), Площадь земли об. для орошения (Х4), Сельскохозяйственная площадь |
| 2019 | Стоимость здорового питания (Х1), Государственные вложения (Х2), Индекс покупательских цен (Х3), Площадь земли об. для орошения (Х4), Сельскохозяйственная площадь |

Изображение выглядит как снимок экрана, прямоугольный, Прямоугольник

Автоматически созданное описание

Рисунок 2.3.3 - Карта корреляции для данных за 2018 год

Изображение выглядит как снимок экрана, прямоугольный, Прямоугольник, шаблон

Автоматически созданное описание

Рисунок 2.3.4 - Карта корреляции для данных за 2019 год

## **2.4. Описательная статистика**

С помощью статистического пакета «Gretl» выведем описательную статистику по исходным данным. Для этого воспользуемся вкладкой «Вид – Описательная статистика».

Изображение выглядит как текст, меню, снимок экрана, Шрифт

Автоматически созданное описание

Рисунок 2.4 – Описательная статистика (2017 год)

Рассмотрим показатель вариации и сделаем вывод об однородности исходных данных. Коэффициент вариации не должен превышать , однако, почти все показатели превышают это значение. Можно сделать вывод о том, что данные неоднородны. Такие же выводы у данных за 2018 и 2019 годы.

# **МНОЖЕСТВЕННЫЙ ЛИНЕЙНЫЙ РЕГРЕССИОННЫЙ АНАЛИЗ**

Модель множественного линейного регрессионного анализа признака записывается следующим образом:

;

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:

В качестве примера рассмотрим функцию:

называется линейной функцией множественной регрессии.

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

Оценки = , =0,164, = 0,1792, = 0,00491, 0,3347, 0,513, 0,223, содержатся в результатах работы программы «Регрессия» в выделенной таблице в столбце «Коэффициенты» под заголовками «Y-пересечение», «», «», «», «», «», «», соответственно.

Таким образом, оценка линейной функции регрессии такова:



Рисунок 3.1 – Результат регрессионного анализа



Рисунок 3.2 – Вывод остатков

В таблице «Вывод остатка», фрагмент которой приведен на рисунке 3.2, содержится предсказанное — это , рассчитанные по построенному уравнению регрессии, и остатки — это разности ().

В таблице «Регрессионная статистика» (рисунок 3.1) приведены:

– оценка коэффициента множественной линейной детерминации

0,7795

– оценка коэффициента множественной линейной корреляции

0,8829

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

– оценка нормированного коэффициента линейной детерминации

0,7427

(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 0,50717898 (Стандартная ошибка).

1) Проверим гипотезы при альтернативах ,

Эти гипотезы можно проверить при помощи интервальных оценок параметров уравнения регрессии. Все в той же таблице в столбцах «Нижние 95%» и «Верхние 95%» приводятся нижние и верхние границы интервальных оценок параметров . Поскольку в интервалы (0,1587076493361; 0,510606005487749), (0,690092644683387; 0,334923765239142) не попадает нуль, то только гипотезы отвергаются, а остальные не отвергаются.

Таким образом, в построенном уравнении регрессии большинство коэффициентов оказались незначимы, и такое уравнение нельзя считать приемлемым.

Исключим из уравнения регрессор , при котором коэффициент незначим, а соответствующая этому коэффициенту абсолютная величина значения статистики T, равного 0,39627871561685, является наименьшей (рассчитанный уровень значимости p = 0,69423859722719 является наибольшим).

Оценка линейной функции регрессии будет такой:

,

гипотезы отвергаются, а остальные не отвергаются.

2) Исключим из уравнения регрессор , при котором коэффициент незначим, а соответствующая этому коэффициенту абсолютная величина значения статистики T, равного 1,13800174430642, является наименьшей (рассчитанный уровень значимости p = 0,26243527 является наибольшим).

Оценка линейной функции регрессии будет такой:

,

гипотезы отвергаются, а остальные не отвергаются.

3) Исключим из уравнения регрессор , при котором коэффициент незначим, а соответствующая этому коэффициенту абсолютная величина значения статистики T, равного 1,87801773008777, является наименьшей (рассчитанный уровень значимости p = 0,0680686053830424 является наибольшим).

Оценка линейной функции регрессии будет такой:

.

Гипотезы отвергаются, а остальные не отвергаются.

Все гипотезы отвергаются, наилучшим уравнением является полученное на четвертом шаге, поскольку и само уравнение, и все его коэффициенты значимы.

Повторим то же самое для наших моделей.

**До исключения аномальных значений 2017 год**

Новая оценка линейной функции регрессии после исключения регрессоров:

**После исключения аномальных значений 2017 год**

Новая оценка линейной функции регрессии после исключения регрессоров:

**До исключения аномальных значений 2018 год**

Новая оценка линейной функции регрессии после исключения регрессоров:

**После исключения аномальных значений 2018 год**

Новая оценка линейной функции регрессии после исключения регрессоров:

**До исключения аномальных значений 2019 год**

Новая оценка линейной функции регрессии после исключения регрессоров:

**После исключения аномальных значений 2019 год**

Новая оценка линейной функции регрессии после исключения регрессоров:

# 

# **ДИСПЕРСИОННЫЙ АНАЛИЗ**

Проведем дисперсионный анализ для определения значимости уравнений регрессии, приведенных в пример. 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 31,48, 10,519 и 42, приводятся в столбце «SS»; а в столбце «MS» приведены значения величин , , равные соответственно 10,4934 и 0,2697.

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 (2017 год)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | *df* | *SS* | *MS* | *F* |
| Регрессия | 3 | 31,48031672 | 10,49343891 | 38,9027033 |
| Остаток | 39 | 10,51968328 | 0,269735469 |  |
| Итого | 42 | 42 |  |  |

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.

**До исключения аномальных значений 2017 год**

Рисунок 4.2 – Дисперсионный анализ до исключения (2017 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.

**До исключения аномальных значений 2018 год**

Таблица 4.3 – Дисперсионный анализ до исключения (2018 год)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Дисперсионный анализ |  |  |  |  |  |
|  | df | SS | MS | F | Значимость F |
| Регрессия | 4 | 1,06794E+18 | 2,66985E+17 | 328,4857996 | 3,80731E-32 |
| Остаток | 44 | 3,57621E+16 | 8,12774E+14 |  |  |
| Итого | 48 | 1,1037E+18 |  |  |  |

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 исключения аномальных значений 2018 год**

Таблица 4.4 – Дисперсионный анализ после исключения (2018 год)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | df | SS | MS | F | Значимость F |
| Регрессия | 6 | 6,01041E+16 | 1,00174E+16 | 18,78718085 | 9,18844E-10 |
| Остаток | 36 | 1,91953E+16 | 5,33202E+14 |  |  |
| Итого | 42 | 7,92994E+16 |  |  |  |

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.

**До исключения аномальных значений 2019 год**

Таблица 4.5 – Дисперсионный анализ до исключения (2019 год)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | df | SS | MS | F | Значимость F |
| Регрессия | 3 | 3,27648E+17 | 1,09216E+17 | 135,5627185 | 5,40422E-21 |
| Остаток | 40 | 3,2226E+16 | 8,0565E+14 |  |  |
| Итого | 43 | 3,59874E+17 |  |  |  |

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 исключения аномальных значений 2019 год**

Таблица 4.6 – Дисперсионный анализ после исключения (2019 год)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | df | SS | MS | F | Значимость F |
| Регрессия | 4 | 1,89849E+16 | 4,74623E+15 | 22,34866114 | 6,94733E-09 |

Продолжение таблицы 4.6

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Остаток | 32 | 6,7959E+15 | 2,12372E+14 |  |  |
| Итого | 36 | 2,57808E+16 |  |  |  |

Таким образом, 22,34866114, а из таблиц распределения Фишера определяем . Следовательно, так как F>, то с вероятность 95% можно говорить о значимости в целом построенной модели регрессии.

# **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 2017 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 – Результаты работы программы «Регрессия» для 2017 года

Результаты теста Уайта показывают присутствие гетероскедастичности, так как при 5% уровне значимости . Р-вероятность принятия гипотезы о гетероскедастичности равна 0,00, что меньше 0,05.

Проверим остальные значимые модели на гетероскедастичность в MS Excel. Результаты будут представлены в итоговой таблице значимых моделей, графики разброса остатков в приложении 6 (рис. 6.1.2–6.1.). Таким образом, все модели гетероскедастичны.

## **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 из предыдущего пункта, их квадрат и произведение предыдущего и нынешнего значения остатков. Посчитаем суммы, начиная от второго значения (см. рис. 5.2.1)



Рисунок 5.2.1 - Вспомогательные расчеты для статистики Дарбина-Уотсона

Посчитаем по формуле значение статистики Дарбина-Уотсона:

(2)

Определим интервалы неопределённости:

Из таблицы значений Дарбина-Уотсона

.

Интервалы равны:

.

Обратимся к интервалу определения автокорреляции (см. рис. 5.2.2)

Изображение выглядит как текст, линия, снимок экрана, Параллельный

Автоматически созданное описание

Рисунок 5.2.2 – Интервалы автокорреляции

По модели за 2017 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.

По итогу верификации моделей у нас было обнаружено несколько гомоскедастичных моделей с отсутствием автокорреляции.

# **ВЫЯВЛЕНИЕ ЛУЧШИХ МОДЕЛЕЙ**

Исходные данные представляли собой 49 наблюдений странам мира, где в качестве зависимой переменной был представлена валовая стоимость сельскохозяйственной продукции (1000 долларов США) с независимыми- Стоимость здорового питания (долларов США на человека в день.), - Государственные вложения (миллионы долларов США.), - Индекс покупательских цен (Пункты), - Площадь земли об. для орошения (1000 гектар), Сельскохозяйственная площадь (1000 гектар), - Сумма кредитов (миллионов долларов США).

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 2017 года подходит модель №5 из итоговой таблицы значимых уравнений. Уравнение регрессии модели:

Характеристики модели:

|  |  |  |  |
| --- | --- | --- | --- |
|  | F-значимость | Ст. ошибка | DW-статистика |
| 0,74 | 0,00 | 21573499,68 | 2,27 |

Данная модель гетероскедастична, но в ней отсутствует автокорреляция.

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

Для модели №5 . Разделим исходные векторы значений и на и построим новое уравнение регрессии:

Условия гомоскедастичности и выполняются, следовательно, остатки гомоскедастичны.

Рассмотрим модели за 2018 год. Наилучшей моделью становится модель №10, она имеет уравнение:

Характеристики модели:

|  |  |  |  |
| --- | --- | --- | --- |
|  | F-значимость | Ст. ошибка | DW-статистика |
| 0,96 | 0,00 | 28509191,33 | 2,16 |

У этой модели автокорреляция отсутствует, присутствует гетероскедастичность, от которой необходимо избавиться. В ней не удалялись аномальные значения, n = 49. Т. к. модель №9 имеет такие же характеристики качества, выбор осуществлялся по наименьшей стандартной ошибки модели.

Для 2019 года лучшая модель №19. Её уравнение регрессии:

Характеристики модели:

|  |  |  |  |
| --- | --- | --- | --- |
|  | F-значимость | Ст. ошибка | DW-статистика |
| 0,9 | 0,00 | 28383976,96 | 1,96 |

Эта модель имеет наименьшую стандартную ошибку у уравнений за 2019 год, где нет мультиколлинеарности и автокорреляции, но у неё присутствует гетероскедастичность, от которой необходимо избавиться.

Эти модели были выбраны на основании сравнений характеристик качества, представленных в итоговой таблице. После использования метода взвешенных наименьших квадратов, были получены следующие уравнения регрессии:

2017 год:

2018 год:

2019 год:

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

Проинтерпретируем выбранные модели.

Модель 2017 года:

* Значение нормированного коэффициента регрессии для (0,619) говорит о том, что увеличение стоимости здорового питания на 1 доллар США в день приводит к увеличению валовой стоимости сельскохозяйственной продукции на 0,619 тысячи долларов США.
* Значение нормированного коэффициента регрессии для (0,0000009) означает, что увеличение государственных вложений на 1 миллион долларов США приводит к увеличению валовой стоимости сельскохозяйственной продукции на 0,0000009 тысячи долларов США.
* Значение нормированного коэффициента регрессии для (0,0001888) указывает на то, что увеличение площади земли для орошения на 1000 гектар приводит к увеличению валовой стоимости сельскохозяйственной продукции на 0,0001888 тысячи долларов США.
* Значение нормированного коэффициента регрессии для (0,0000176) говорит о том, что увеличение сельскохозяйственной площади на 1000 гектар приводит к увеличению валовой стоимости сельскохозяйственной продукции на 0,0000176 тысячи долларов США.
* Значение нормированного коэффициента регрессии для (0,000001) означает, что увеличение суммы кредитов на 1 миллион долларов США приводит к увеличению валовой стоимости сельскохозяйственной продукции на 0,000001 тысячи долларов США.
* Значение нормированного коэффициента детерминации (R^2) равное 0,9, говорит о том, что 90% изменчивости зависимой переменной Y объясняется независимыми переменными в модели.

Модель 2018 года:

* Государственные вложения (x2): Увеличение государственных вложений на 1 миллион долларов США приводит к увеличению валовой стоимости сельскохозяйственной продукции на 0,000000748 тысячи долларов США.
* Индекс покупательских цен (x3): В данной модели индекс покупательских цен не включен в уравнение регрессии, поэтому мы не можем дать интерпретацию для этой переменной.
* Площадь земли для орошения (x4): Увеличение площади земли для орошения на 1000 гектар приводит к увеличению валовой стоимости сельскохозяйственной продукции на 0,0001895 тысячи долларов США.
* Сельскохозяйственная площадь (x5): Увеличение сельскохозяйственной площади на 1000 гектар приводит к увеличению валовой стоимости сельскохозяйственной продукции на 0,00000686 тысячи долларов США.
* Сумма кредитов (x6): Увеличение суммы кредитов на 1 миллион долларов США приводит к увеличению валовой стоимости сельскохозяйственной продукции на 0,00000123 тысячи долларов США.

Значение нормированного коэффициента детерминации (R^2) равное 0,96 говорит о том, что 96% изменчивости зависимой переменной Y объясняется независимыми переменными в модели.

Значимость F-теста на уровне значимости 0,000001 указывает на общую статистическую значимость модели в целом.

Модель 2019 года:

* Значение нормированного коэффициента регрессии для () означает, что увеличение государственных вложений на 1 миллион долларов США приводит к увеличению валовой стоимости сельскохозяйственной продукции на тысячи долларов США.
* Значение нормированного коэффициента регрессии для (0,0001888) указывает на то, что увеличение площади земли для орошения на 1000 гектар приводит к увеличению валовой стоимости сельскохозяйственной продукции на 0, тысячи долларов США.
* Значение нормированного коэффициента регрессии для (0,0000176) говорит о том, что увеличение сельскохозяйственной площади на 1000 гектар приводит к увеличению валовой стоимости сельскохозяйственной продукции на тысячи долларов США

**7 ПРОГНОЗИРОВАНИЕ**

## **7.1 Оценка прогностических свойств моделей**

На основе выбранных моделей проведем прогнозирование и оценим его качество. Выберем лучшую модель из Таблицы 7.1.1

Таблица 7.1.1 – Значения информационных и прогностических характеристик конкурирующих регрессионных моделей

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Модель |  |  |  |  |  |  |
|  | 74% | 96% | 64% | 89% | 88% | 88% |
|  | 2624855,91 | 2936942,60 | 295998.59 | 391,9 | 418,66 | 218,66 |
|  | 19327637,84 | 14297058,26 | 513378114707714,86 | 464,61 | 459,84 | 3459,84 |
| MSE | 6888155153965,60 | 8623613842194,47 | 206188417195395,46 | 397581,2 | 384766,5 | 2354766,5 |
|  | 40% | 21% | 43% | 57% | 52% | 47% |

Лучшей оказалась модель

,

т.к. она имеет наименьшую процентную ошибку прогноза.

По итогам исследования моделей можно сделать следующие выводы:

1. Показатели информационной пригодности моделей говорят о плохом качестве прогнозирования;

Формально лучшей можно выбрать модель 2018 года

,

т. к. она имеет наименьшую ошибку прогноза, но как таковых высококачественных моделей получить не удалось.

**ИТОГОВАЯ ТАБЛИЦА ЗНАЧИМЫХ МОДЕЛЕЙ**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Год | № | Уравнение регрессии | Характеристики качества | | | | | Описание модели |
|  | F-значимость | Ст. ошибка | DW-статистика | Выводы тестирования |
| 2017 | 1 |  | 0,96 | 0,00 | 28969943,92 | 2,13 | * Гетероскедастичность присутствует * Автокорреляция присутсвует * Мультиколлинеарность присутствует | Исходные данные за 2017 год, до удаления аномальных значений, n = 50 |
| 2 |  | 0,96 | 0,00 | 28635997,24 | 2,13 | * Гетероскедастичность присутствует * Автокорреляция присутствует * Мультиколлинеарность присутствует | Исходные данные за 2017 год, до удаления аномальных значений, n = 50 |
| 3 |  | 0,96 | 0,00 | 28434265,72 | 2,09 | * Гетероскедастичность присутствует * Автокорреляция отсутствует * Мультиколлинеарность присутствует | Исходные данные за 2017 год, до удаления аномальных значений, n = 50 |
| 4 |  | 0,74 | 0,00 | 21823532,46 | 2,27 | * Гетероскедастичность присутствует * Автокорреляция присутствует * Мультиколлинеарность присутсвует | Исходные данные за 2017 год, удалены аномальные значения, n = 44 |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 2017 | | 5 | |  | 0,74 | | 0,00 | | 21573499,68 | | 2,27 | | * Гетероскедастичность присутствует * Автокорреляция отсутствует * Мультиколлинеарность присутствует | | Исходные данные за 2017 год, удалены аномальные значения, n = 44 | |
| 6 | |  | 0,74 | | 0,00 | | 21657090,60 | | 2,34 | | * Гетероскедастичность присутствует * Автокорреляция не определена * Мультиколлинеарность присутствует | | Исходные данные за 2017 год, удалены аномальные значения, n = 44 | |
| 7 | |  | 0,73 | | 0,00 | | 22347699,88 | | 2,43 | | * Гетероскедастичность присутствует * Автокорреляция не определена * Мультиколлинеарность отсутствует | | Исходные данные за 2017 год, удалены аномальные значения, n = 44 | |
| 2018 | | 8 | |  | 0,96 | | 0,00 | | 28938302,81 | | 2,24 | | * Гетероскедастичность присутствует * Автокорреляция не определена * Мультиколлинеарность присутствует | | Исходные данные за 2018 год, до удаления аномальных значений, n = 49 | |
| 9 | |  | 0,96 | | 0,00 | | 28653308,28 | | 2,2 | | * Гетероскедастичность присутствует * Автокорреляция отсутствует * Мультиколлинеарность присутствует | | Исходные данные за 2018 год, до удаления аномальных значений, n = 49 | |
| 10 | |  | 0,96 | | 0,00 | | 28509191,33 | | 2,16 | | * Гетероскедастичность присутствует * Автокорреляция отсутствует * Мультиколлинеарность отсутствует | | Исходные данные за 2018 год, до удаления аномальных значений, n = 49 | |
| 11 | |  | 0,71 | | 0,00 | | 23091161,46 | | 2,38 | | * Гетероскедастичность присутствует * Автокорреляция положительная * Мультиколлинеарность присутствует | | Исходные данные за 2018 год, удалены аномальные значения, n = 44 | |
| 12 | |  | 0,72 | | 0,00 | | 22875815,96 | | 2,37 | | * Гетероскедастичность присутствует * Автокорреляция положительная * Мультиколлинеарность присутсвует | | Исходные данные за 2018 год, удалены аномальные значения, n = 44 | |
|  | | 13 | |  | 0,72 | | 0,00 | | 22865777,44 | | 2,43 | | * Гетероскедастичность присутствует * Автокорреляция положительная * Мультиколлинеарность присутсвует | | Исходные данные за 2018 год, удалены аномальные значения, n = 44 | |
|  | | 14 | |  | 0,71 | | 0,00 | | 23362884,95 | | 2,47 | | * Гетероскедастичность присутствует * Автокорреляция положительная * Мультиколлинеарность отсутсвует | | Исходные данные за 2018 год, удалены аномальные значения, n = 44 | |
| 2019 | 15 | |  | | | 0,9 | | 0,00 | | 28898284,81 | | 2,04 | | * Гетероскедастичность присутствует * Автокорреляция отсутсвует * Мультиколлинеарность присутсвует | | Исходные данные за 2019 год, до удаления аномальных значений, n = 44 | |
| 17 | |  | | | 0,9 | | 0,00 | | 28572816,55 | | 2,05 | | * Гетероскедастичность присутствует * Автокорреляция отсутствует * Мультиколлинеарность отсутствует | | Исходные данные за 2019 год, до удаления аномальных значений, n = 44 | |
| 18 | |  | | | 0,9 | | 0,00 | | 28260992,6 | | 2,05 | | * Гетероскедастичность присутствует * Автокорреляция отсутствует * Мультиколлинеарность отсутствует | | Исходные данные за 2019 год, до удаления аномальных значений, n = 44 | |
| 19 | |  | | | 0,9 | | 0,00 | | 28383976,96 | | 1,96 | | * Гетероскедастичность присутствует * Автокорреляция положительная * Мультиколлинеарность отсутствует | | Исходные данные за 2019 год, до удаления аномальных значений, n = 44 | |
|  | 20 | |  | | | 0,7 | | 0,00 | | 14492379,37 | | 1,72 | | * Гетероскедастичность присутствует * Автокорреляция отсутсвует * Мультиколлинеарность отсутствует | | Исходные данные за 2019 год, аномальные значения удалены, n = 37 | |
|  | 21 | |  | | | 0,7 | | 0,00 | | 14538915,00 | | 1,54 | | * Гетероскедастичность присутствует * Автокорреляция отсутсвует * Мультиколлинеарность отсутствует | | Исходные данные за 2019 год, аномальные значения удалены, n = 37 | |
|  | 22 | |  | | | 0,7 | | 0,00 | | 14572982,04 | | 1,71 | | * Гетероскедастичность присутствует * Автокорреляция отсутсвует * Мультиколлинеарность отсутствует | | Исходные данные за 2019 год, аномальные значения удалены, n = 37 | |

**ПРИЛОЖЕНИЕ 1 - Исходные данные**

Таблица 1.1 - Исходные данные за 2017 год

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Валовая стоимость сельскохозяйственной продукции,    1000 долларов США | Стоимость здорового питания,  долларов США на человека в день | Государственные вложения,  миллионы долларов США | Индекс покупательских цен,  Пункты | Площадь земли об. для орошения,  1000 гектар | Сельскохозяйственная площадь,  1000 гектар | Сумма кредитов,  миллионов долларов США |
| 2137891 | 3,95 | 3805,99 | 101 | 354,4 | 1175,861 | 3058,598 |
| 5858647 | 4,33 | 41011,56 | 132,5 | 85,53 | 45772 | 21801,54 |
| 1801096 | 3,1 | 3193,61 | 95,41 | 213 | 2043,8 | 4917,1975 |
| 45435014 | 2,26 | 496613,1 | 96,34 | 2546 | 393797 | 728367,21 |
| 6162318 | 2,77 | 205500,4 | 106,3 | 99,8 | 2654,62 | 196171,4 |
| 4786817 | 2,35 | 13677,76 | 103,3 | 1445,8 | 4777,5 | 6831,3437 |
| 11603043 | 3,18 | 22651,63 | 123,7 | 30,3 | 8502 | 15385,788 |
| 9436588 | 2,86 | 261590,1 | 103,1 | 24 | 1327 | 252126,76 |
| 6415141 | 3,55 | 12871,37 | 71,88 | 441,74 | 37671,547 | 20772,617 |
| 1328351 | 3,85 | 7239,61 | 97,16 | 3 | 2228 | 4280,8018 |
| 238734947 | 2,81 | 994282,6 | 104,8 | 7577,35 | 236878,6 | 973017,26 |
| 5348453 | 3,78 | 19673,1 | 97,17 | 135,9 | 5029,52 | 29288,266 |
| 49461024 | 2,86 | 603655,1 | 101,8 | 1218 | 59404,2 | 1052806,9 |
| 904055036 | 2,57 | 3004777 | 99,96 | 73946 | 522199,2 | 11797906 |
| 4714311 | 3,96 | 17067,72 | 108,7 | 159,6 | 1781 | 27609,048 |
| 5566209 | 2,9 | 77007,97 | 98,13 | 46 | 3521 | 131992,23 |
| 34843904 | 3,46 | 54817,5 | 140,8 | 3823 | 3836,9688 | 61270,741 |

Продолжение таблицы 1.1

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 954440 | 3,13 | 10564,1 | 104,4 | 3 | 982 | 7995,9963 |
| 3118664 | 2,55 | 137115,1 | 98,9 | 55 | 2270 | 86003,005 |
| 60848857 | 2,94 | 1466352 | 102,7 | 2734,7 | 28697,499 | 1072820,9 |
| 51987177 | 2,79 | 1630438 | 106,6 | 676,4 | 18097 | 1585044,4 |
| 11233362 | 3,04 | 97006,85 | 103 | 1799,143 | 6059,9098 | 92762,24 |
| 25027 | 5,38 | 262,33 | 106,7 | 2 | 8 | 605,50185 |
| 9379041 | 3,3 | 66858,77 | 109,7 | 197,8 | 5564 | 23668,918 |
| 428963425 | 2,82 | 593750,1 | 110,1 | 70400 | 179235,6 | 1238126,9 |
| 100233594 | 4,13 | 169401,7 | 84,77 | 6722 | 63300 | 356006,75 |
| 4429026 | 2,44 | 137135,5 | 98,09 | 306 | 644 | 293170,45 |
| 38492626 | 2,89 | 956637,7 | 102,6 | 4124 | 12826,57 | 918390,13 |
| 16375727 | 2,39 | 38160,58 | 109,5 | 2204,8 | 216213,7 | 38973,35 |
| 3306576 | 2,97 | 3255,2 | 115,9 | 1023,3 | 10540,1 | 1572,9675 |
| 61624817 | 2,99 | 263712,4 | 113,1 | 6811 | 96946 | 244439,53 |
| 541034 | 3,26 | 5452,23 | 112,4 | 8 | 38811 | 6884,4023 |
| 18525450 | 2,74 | 354012,7 | 106,7 | 522 | 1819 | 345319,81 |
| 16875492 | 2,67 | 70531,77 | 108,2 | 747 | 10644 | 489553,02 |
| 58962900 | 3,57 | 39039,85 | 111,1 | 331,2 | 68039 | 51475,159 |
| 1032159 | 2,82 | 31921,2 | 118,3 | 100,92 | 1451,92 | 53345,995 |
| 71356748 | 3,41 | 89020,43 | 110,6 | 19990 | 37003 | 39619,903 |
| 2902962 | 2,46 | 2947,31 | 109 | 228 | 2335,3 | 1809,4614 |
| 95741759 | 3,15 | 576744,9 | 101,9 | 4300 | 215494 | 500822,63 |
| 5882860 | 4,07 | 17827,77 | 106,3 | 70 | 3438 | 9641,9486 |
| 5779 | 4,01 | 546,44 | 108,3 | 0,3 | 1,55 | 507,02523 |
| 2051612 | 3,01 | 37776,54 | 101,8 | 62,8 | 1942 | 19912,794 |
| 50106175 | 2,7 | 533905,4 | 104,5 | 3923 | 26295,53 | 668336,33 |
| 3847312 | 2,52 | 225616,3 | 99,1 | 52 | 1540,4111 | 1151365,5 |
| 4160496 | 3,03 | 2595,51 | 111,1 | 817 | 4907,1 | 808,56724 |
| 46641082 | 3,97 | 96675,37 | 101,6 | 6415 | 22847 | 424286,69 |
| 69279274 | 2,87 | 288849,5 | 109 | 5215 | 37964 | 473931,84 |
| 371774193 | 3,23 | 7082400 | 98,46 | 26915,54 | 405810,35 | 8966105 |
| 47349316 | 3,59 | 60573,48 | 105,3 | 4585 | 12168,8 | 291103,83 |

Таблица 1.2 - Исходные данные за 2018 год

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Валовая стоимость сельскохозяйственной продукции,  1000 долларов США | Стоимость здорового питания,  долларов США на человека в день | Государственные вложения,  миллионы долларов США | Индекс покупательских цен,  Пункты | Площадь земли об. для орошения,  1000 гектар | Сельскохозяйственная площадь,  1000 гектар | Сумма кредитов,  миллионов долларов США |
| 2183282 | 4,069 | 4355,9 | 93,58 | 355,5 | 1175,7 | 3128,536 |
| 5992427 | 4,293 | 35813,6 | 117,24 | 85,53 | 45792 | 16447,42 |
| 1578014 | 3,166 | 3065,83 | 103,42 | 208,4 | 2044,5 | 5761,237 |
| 41612099 | 2,283 | 508691 | 103,89 | 2546 | 378082 | 742627 |
| 6286885 | 2,848 | 221584 | 109,47 | 99,8 | 2652,8 | 206872,5 |
| 5048094 | 2,399 | 14226,7 | 102,9 | 1449,4 | 4779,5 | 7658,925 |
| 11159712 | 3,228 | 24354,1 | 132,64 | 30,3 | 8460 | 15932,7 |
| 9062006 | 2,962 | 283799 | 122,06 | 24 | 1354 | 271856,9 |
| 6906215 | 3,648 | 7912,85 | 74,63 | 473,32 | 37877 | 25021,71 |
| 241223471 | 2,8 | 908691 | 94,86 | 7888,7 | 237502 | 893565,7 |
| 5302067 | 3,876 | 23607,4 | 99,56 | 135,9 | 5030 | 33034,38 |
| 49789733 | 2,911 | 635910 | 101,27 | 1218 | 59401 | 1264494 |
| 913535625 | 2,63 | 3338959 | 104,54 | 74542 | 521447 | 13126778 |
| 4694431 | 4 | 19180,6 | 88,37 | 159,6 | 1782,5 | 28663,55 |
| 5445258 | 2,921 | 99578,8 | 100,75 | 46 | 3523 | 152159,4 |
| 33405984 | 3,507 | 67941,6 | 146,67 | 3823 | 3863 | 70019,74 |
| 811682 | 3,188 | 12027,2 | 110,94 | 3 | 985 | 8656,515 |
| 2957591 | 2,624 | 147018 | 109,31 | 55 | 2272 | 93186,51 |
| 58607646 | 3,019 | 1553015 | 110,33 | 2778,7 | 28660 | 1188350 |

Продолжение таблицы 1.2

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 48914890 | 2,917 | 1761513 | 111,08 | 676,4 | 18295 | 1752040 |
| 11081712 | 3,102 | 102905 | 102,57 | 1866,4 | 5949,9 | 90199,45 |
| 25642 | 5,536 | 264,26 | 105,06 | 2 | 8 | 630,3104 |
| 9543289 | 3,383 | 74103,1 | 104,12 | 201,7 | 5344 | 27586,11 |
| 445979299 | 2,83 | 618636 | 109,77 | 70400 | 179116 | 1346400 |
| 104162878 | 4,273 | 171998 | 85,63 | 6722 | 63900 | 373817,2 |
| 4322553 | 2,5 | 147342 | 106,45 | 306 | 641,3 | 314786 |
| 38923256 | 2,979 | 1012435 | 107,49 | 4124 | 12451 | 896220,5 |
| 16848034 | 2,426 | 32597,9 | 119,29 | 2204,8 | 216036 | 37979,53 |
| 3366443 | 2,931 | 3168,67 | 106,88 | 1023,3 | 10371 | 1857,662 |
| 63143836 | 3,071 | 279952 | 120,56 | 7056,5 | 96516 | 262958 |
| 527797 | 3,3 | 6244,07 | 123,31 | 8 | 38811 | 7322,585 |
| 17365379 | 2,821 | 386622 | 107,68 | 522 | 1823 | 358237,8 |
| 17511595 | 2,589 | 72470,8 | 118,13 | 747 | 10498 | 503007,7 |
| 61427289 | 3,724 | 45631,3 | 113,45 | 331,2 | 68192 | 49444,66 |
| 1109987 | 2,838 | 35368,5 | 118,31 | 108,89 | 1459,9 | 55880,39 |
| 69649945 | 3,395 | 86276,5 | 115,78 | 19990 | 36953 | 41875,43 |
| 3043164 | 2,571 | 3547,72 | 101,31 | 222,5 | 2276,6 | 2110,025 |
| 93297888 | 3,197 | 568793 | 106,48 | 4300 | 215494 | 514276,8 |
| 3112490 | 2,483 | 2305,79 | 106,8 | 8,5 | 1949 | 1908,971 |
| 6981748 | 4,166 | 20694,2 | 101,58 | 68 | 3487 | 11136,56 |
| 6623 | 3,959 | 602,82 | 110,57 | 0,3 | 1,55 | 547,8496 |
| 2181814 | 3,102 | 42059,7 | 104,9 | 65,5 | 1947 | 21887,33 |
| 57890701 | 2,741 | 585228 | 104,91 | 3923 | 26183 | 644327,6 |
| 4001809 | 2,591 | 230395 | 93,57 | 52 | 1537,5 | 1199659 |
| 4473583 | 3,03 | 2643,08 | 99,28 | 817 | 4911,6 | 732,2234 |
| 47991354 | 4,042 | 106873 | 99,51 | 6415 | 22846 | 457004,1 |
| 69787254 | 2,997 | 266671 | 128,49 | 5215 | 37802 | 391475,3 |
| 375198125 | 3,24 | 7447700 | 97,98 | 26916 | 405810 | 9419544 |
| 48931621 | 3,663 | 63509,1 | 102,56 | 4585 | 12141 | 319049,6 |

Таблица 1.3 - Исходные данные за 2019 год

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Валовая стоимость сельскохозяйственной продукции,  1000 долларов США | Стоимость здорового питания,  долларов США на человека в день | Государственные вложения,  миллионы долларов США | Индекс покупательских цен,  Пункты | Площадь земли об. для орошения,  1000 гектар | Сельскохозяйственная площадь,  1000 гектар | Сумма кредитов,  миллионов долларов США |
| 2142931 | 4,262 | 4446,43 | 98,88 | 356,7 | 1175,7 | 3459,89 |
| 1536229 | 3,237 | 3470,3 | 103,06 | 208,1 | 2044,2 | 6760,42 |
| 37895757 | 2,296 | 500817 | 115,04 | 2546 | 383801 | 648176 |
| 6253904 | 2,915 | 215921 | 104,22 | 99,8 | 2651,2 | 688589 |
| 5446083 | 2,459 | 17842,8 | 108,76 | 1450,2 | 4779,7 | 8998,93 |
| 11425651 | 3,31 | 26216,5 | 146,9 | 30,3 | 8391 | 16416,7 |
| 9330668 | 3,047 | 277516 | 110,85 | 24 | 1356,4 | 259861 |
| 246514517 | 2,882 | 878532 | 99,21 | 8200 | 238124 | 885361 |
| 5395318 | 4,036 | 25675,5 | 100,27 | 135,9 | 5037 | 33105,8 |
| 50270350 | 2,927 | 654583 | 101,62 | 1218 | 59176 | 1334055 |
| 4532787 | 4,048 | 20494,5 | 88,84 | 159,6 | 1776 | 27085,5 |
| 5446250 | 3,025 | 93428,2 | 105,47 | 46 | 3523,7 | 150465 |
| 34744687 | 3,503 | 84911,8 | 260,6 | 3880 | 3922 | 81525,9 |
| 1022718 | 3,284 | 12242,6 | 105,55 | 3 | 988,4 | 8156,6 |
| 3250508 | 2,704 | 143079 | 106,59 | 55 | 2274 | 96143,7 |
| 59366462 | 3,177 | 1510485 | 108,73 | 2822,7 | 28621 | 1175261 |
| 50270728 | 2,984 | 1749343 | 112,13 | 676,4 | 18311 | 1746989 |
| 11354265 | 3,167 | 98242,8 | 104,85 | 1933,7 | 5874,4 | 75393,5 |
| 9408799 | 3,49 | 74786 | 110,36 | 190 | 5319 | 28561,4 |

Продолжение таблицы 1.3

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 456824177 | 2,877 | 668317 | 115,7 | 72504 | 179045 | 1404685 |
| 100200874 | 4,268 | 183152 | 78,93 | 6722 | 64100 | 398219 |
| 4370957 | 2,482 | 154851 | 104,22 | 306 | 638,4 | 323152 |
| 38630323 | 3,121 | 975066 | 108,95 | 4124 | 13150 | 792849 |
| 16396287 | 2,537 | 35498,8 | 162,79 | 2224,6 | 213974 | 36224,7 |
| 3430423 | 2,991 | 3360,23 | 99,85 | 1023,3 | 10368 | 2098,29 |
| 63765368 | 3,039 | 285738 | 125,42 | 7301,9 | 96106 | 275509 |
| 18225550 | 2,932 | 382439 | 110,08 | 522 | 1816,8 | 329620 |
| 17437041 | 2,722 | 73803,8 | 112,73 | 735 | 10345 | 507721 |
| 59145437 | 3,87 | 51937,1 | 110,21 | 331,2 | 68342 | 56000,6 |
| 1058364 | 2,921 | 34358,9 | 118,31 | 110,4 | 1461,4 | 56839,4 |
| 73263018 | 3,46 | 83028,2 | 104,7 | 19990 | 36323 | 35227,4 |
| 2973177 | 2,687 | 3414,89 | 104,83 | 222,1 | 2281,9 | 2297,52 |
| 98017408 | 3,264 | 639876 | 117,12 | 4300 | 215494 | 510600 |
| 3459606 | 2,537 | 1947,39 | 111,69 | 8,5 | 1904 | 2146,01 |
| 6990736 | 4,334 | 21754,7 | 101,91 | 63 | 3482 | 11610,6 |
| 6545 | 3,948 | 586,49 | 109,16 | 0,3 | 1,55 | 646,878 |
| 2182471 | 3,242 | 42809,8 | 106,94 | 65,5 | 1941 | 21484,7 |
| 52467498 | 2,845 | 580387 | 100,78 | 3923 | 26207 | 598665 |
| 3831064 | 2,654 | 230723 | 94,56 | 52 | 1534,6 | 1219040 |
| 4789554 | 3,194 | 2497,97 | 101,57 | 817 | 4918,2 | 804,239 |
| 46610677 | 4,181 | 111327 | 100,96 | 6415 | 23000 | 492984 |
| 72789552 | 3,189 | 264622 | 157,36 | 5215 | 37716 | 365737 |
| 364097681 | 3,268 | 7862300 | 102,12 | 26916 | 405810 | 9822961 |
| 48831109 | 3,776 | 66242 | 101,27 | 4585 | 12388 | 355545 |

**ПРИЛОЖЕНИЕ 2**

Изображение выглядит как текст, снимок экрана, число, Шрифт

Автоматически созданное описание

Изображение выглядит как текст, снимок экрана, число, линия

Автоматически созданное описание

Рисунок 2.1 – Гистограммы и графики разброса (2018 год)

Изображение выглядит как текст, снимок экрана, число, Шрифт

Автоматически созданное описание

Изображение выглядит как текст, снимок экрана, число, Шрифт

Автоматически созданное описание

Рисунок 2.2 – Гистограммы и графики разброса (2019 год)

Изображение выглядит как текст, меню, снимок экрана, Шрифт

Автоматически созданное описание

Рисунок 2.3 - Проверка коэффициентов корреляции на значимость (2018 год)

Изображение выглядит как текст, меню, снимок экрана, Шрифт

Автоматически созданное описание

Рисунок 2.4 - Проверка коэффициентов корреляции на значимость (2019 год)

Изображение выглядит как текст, меню, снимок экрана, Шрифт

Автоматически созданное описание

Рисунок 2.5 – Описательная статистика (2018 год)

Изображение выглядит как текст, меню, снимок экрана, Шрифт

Автоматически созданное описание

Рисунок 2.6 – Описательная статистика (2019 год)

**ПРИЛОЖЕНИЕ 3**



Рисунок 3.1 – Регрессионный анализ до исключения аномальных значений (2017 год)



Рисунок 3.2 – Регрессионный анализ до исключения аномальных значений (2017 год)



Рисунок 3.3 – Регрессионный анализ до исключения аномальных значений (2017 год)



Рисунок 3.4 – Регрессионный анализ после исключения аномальных значений (2017 год)



Рисунок 3.5 – Регрессионный анализ после исключения аномальных значений (2017 год)



Рисунок 3.6 – Регрессионный анализ после исключения аномальных значений (2017 год)



Рисунок 3.7 – Регрессионный анализ после исключения аномальных значений (2017 год)



Рисунок 3.8 – Регрессионный анализ до исключения аномальных значений (2018 год)



Рисунок 3.9 – Регрессионный анализ до исключения аномальных значений (2018 год)



Рисунок 3.10 – Регрессионный анализ после исключения аномальных значений (2018 год)



Рисунок 3.11 – Регрессионный анализ после исключения аномальных значений (2018 год)



Рисунок 3.12 – Регрессионный анализ до исключения аномальных значений (2019 год)



Рисунок 3.13 – Регрессионный анализ до исключения аномальных значений (2019 год)



Рисунок 3.14 – Регрессионный анализ после исключения аномальных значений (2019 год)



Рисунок 3.15 – Регрессионный анализ после исключения аномальных значений (2019 год)

# **СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ**

1. Корпоративная статистическая база данных Продовольственной и сельскохозяйственной организации (FAOSTAT). (Дата обращения: 20.09.2023)

– <https://www.fao.org/faostat/en/#home>

1. Писарева, О.М., 2014. Эконометрика: Начальный курс. Москва: Государственный университет управления, Институт информационных систем ГУУ.
2. Ершова, И.Г., 2010. Анализ и прогнозирование динамики занятости на уровне региона. Экономика и управление, 30 (165).