**Министерство образования и науки Российской Федерации**

Федеральное государственное бюджетное образовательное учреждение

высшего образования

**«ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ УПРАВЛЕНИЯ»**



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

**Проектная работа**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| по дисциплине | | Эконометрика | | | | |
|  | Оценка спроса на лекарственные препараты под влиянием пандемии | | | | | |
| (тема курсового проекта) | | | | | | |
|  | |  | | | | |
| Направление подготовки | | 01.03.02 | |  | Прикладная математика и информатика | |
| (код) | |  | (наименование) | |
| Образовательная программа  Роли в группе: | | Прикладная математика и информатика | | | | |
| (название образовательной программы) | | | | |
| Руководство и оформление  Сбор данных и анализ  Анализ  Презентация | | Финагеев Максим | | | | |
| Черняков Артём | | | | |
| Сагандыков Ильдар | | | | |
| Черняков Артём | | | | |
| (Фамилия, Имя, Отчество) | | | | |
| 3 курс, 5 группа | |
| (курс, номер группы) | |
| Руководитель работы | | Заведующая кафедрой, к.э.н., доцент, Писарева О.М. | | | | |
| (ученая степень, звание, Фамилия и Инициалы) | | | | |
| Оценка: | | |  | |
| Подпись руководителя: | | |  | |
| Дата защиты: | | | «\_\_\_» \_\_\_\_\_\_\_\_\_\_\_\_\_2021г. | |

Москва – 2021

**СОДЕРЖАНИЕ**

[ВВЕДЕНИЕ 2](#_Toc92722855)

[1. ПОДБОР МАССИВА ДАННЫХ. 3](#_Toc92722856)

[2. ПРЕДВАРИТЕЛЬНЫЙ АНАЛИЗ 4](#_Toc92722857)

[3. РЕГРЕССИОННЫЙ АНАЛИЗ ПО ИСХОДНЫМ ДАННЫМ 8](#_Toc92722858)

[4. ВЕРИФИКАЦИЯ МОДЕЛЕЙ 12](#_Toc92722859)

[5. ВЫБОР ЛУЧШЕЙ МОДЕЛИ 16](#_Toc92722860)

[ЗАКЛЮЧЕНИЕ 17](#_Toc92722861)

[ПРИЛОЖЕНИЕ 1 22](#_Toc92722862)

[ПРИЛОЖЕНИЕ 2 23](#_Toc92722863)

[ПРИЛОЖЕНИЕ 3 28](#_Toc92722864)

# ВВЕДЕНИЕ

**Актуальность работы.**

Под влиянием пандемии произошло изменение спроса на актуальные лекарственные препараты, анализ позволит понять, как именно пандемия сказалась на фармацевтическом рынке

**Цели и задачи работы.**

* Определение необходимой информации и направлений ее поиска;
* Сбор, первичная обработка и анализ информации;
* Ознакомление с базовыми возможностями статистических пакетов;
* Формулировка рабочих гипотез;
* Закрепление теоретических знаний, связанных с вопросами

обоснования необходимости и допустимости применения факторных регрессионных моделей;

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

Для выполнения проектной работы был использован табличный процессор – MS Excel и SPSS Statistics.

Таблица 1 ⎯ Участники работы и их задачи

|  |  |  |
| --- | --- | --- |
| ФИО | Задачи | Оценка |
| Финагеев М.В.  Сагандыков И.Н. | Оформление проектной работы, общение с куратором проектной работы (руководитель группы) |  |
| Сагандыков И.Н. | Выполнение математических расчетов |  |
| Черняков А. | Выполнение математических расчетов |  |

# ПОДБОР МАССИВА ДАННЫХ.

Необходимо проанализировать зависимость одного эконометрического

показателя от других. Для полноты картины использовалось пять показателей.

Введем условные обозначения:

Y – Объем выпущенной продукции, млн уп.

X1 – Объем продаж, тысячи миллионов руб.

X2 – Количество заболевших COVID-19 на 2020 год

X3 – Средняя цена лекарственного препарата, руб.

X4 – Объем продаж, млн. упаковок

Исследование проводится по данным 2020 года, проанализированны 22 области России, таблицы исходных данных представлены в приложении «А».

# ПРЕДВАРИТЕЛЬНЫЙ АНАЛИЗ

Для предварительного анализа каждой переменной имеющегося массива данных был использован инструмент «Описательная статистика» из пакета «Анализ данных» Microsoft Excel:

Таблица 2.1 - Описательная статистика для 22 наблюдений



Проверим Y на наличие аномальных наблюдений:

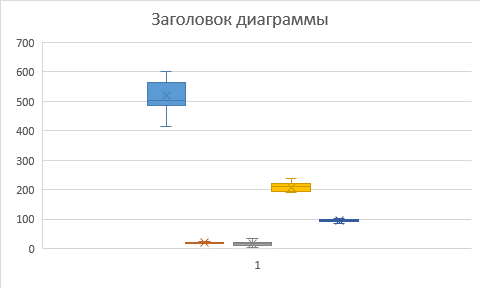


Рисунок 2.1 - Коробчатая диаграмма

По коробчатой диаграмме видно, что аномальные значения отсутствуют

В результате получены значения выборочных характеристик исходных данных:

Выборочное среднее – приближение теоритического среднего распределения, основанное на выборке из него:  :

, , , , 96,22,

Стандартное отклонение :

= 1,70,

**Графический анализ исходной выборки**

Рисунок 2.2 - Графики взаимного разброса

На основании графиков взаимного разброса можно предположить наличие линейной зависимости между результативным признаком Y и факторными признаками X. На основании вышеизложенного, в дальнейшем мы будем рассматривать построение линейной модели.

Изучим линейную зависимость результативного признака Y – уровень занятости населения (в процентах) от пяти факторных признаков — регрессоров:

* X1 – Объем продаж, тыс. млн. рублей.
* X2 – Количество заболевших.
* X3 – Средняя цена препарата, руб.
* X4 – Объем продаж, млн. уп.

Для расчета коррелироемости данных, воспользуемся пакетом «Анализ данных» от Excel. Программа «Корреляция» выведет матрицу оценок коэффицентов корреляции. Результаты работы программы представлены в таблице 2.5.

Таблица 2.5 - Корреляция

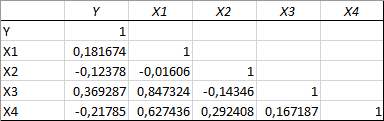


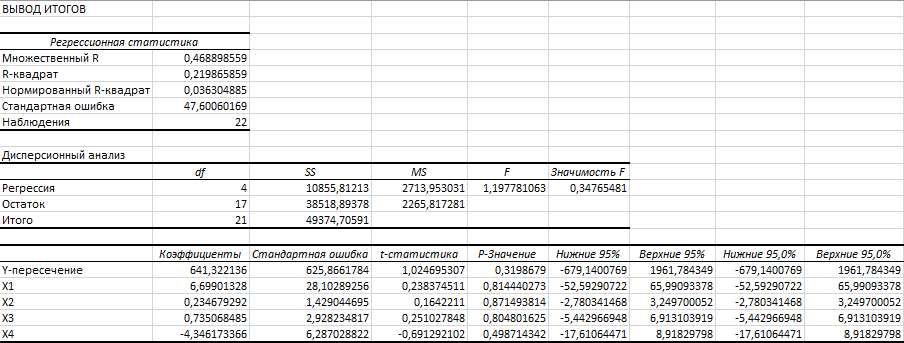
Таблица 2.5 показывает, что результативный признак Y значимо коррелирует с факторными признаками X1, X2, X3

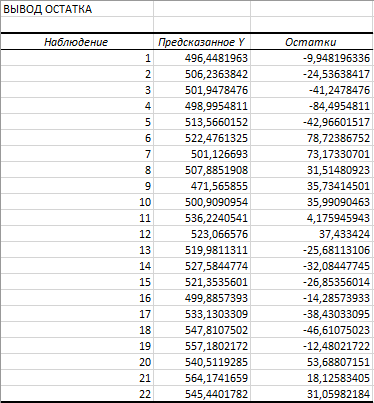
# **РЕГРЕССИОННЫЙ АНАЛИЗ ПО ИСХОДНЫМ ДАННЫМ**

1. Рассчитаем оценки  *и* параметров модели линейной регрессии. Воспользуемся программой «Регрессия», выбрав этот пункт в меню надстройки «Анализа данных» Microsoft Excel.

Результаты работы программы «Регрессия» представлены в таблице 3.1.

Таблица 3.1 -Результаты работы программы «Регрессия»





Оценки = 641,322, = 6,70, = 0,23, = 0,73, = -4,35, находятся в выделенной таблице в столбце «Коэффициенты» под заголовками «Y-пересечение», «X(1)», «X(2)», «X(3)», «X(4)», «X(5)» соответственно. Оценка линейной функции регрессии такова:

В таблице «Вывод остатка» содержится пункт «Предсказанное Y» — это , рассчитанные по построенному уравнению регрессии. Пункт «Остатки» — это регрессионные остатки, или же разности (*yi* −).   
Зная эти остатки, можно рассчитать среднюю относительную ошибку аппроксимации: .

В таблице 3.1 приведено:

(R-квадрат) — судя по наблюдениям, 21% вариации объема выпущенной продукции зависит от объема продаж, количества заболевших, средней цены препаратов и объема продаж

Множественный коэффицент корреляции характеризует тесноту линейной связи между главным фактором Y и совокурностью X1, X2, X3, X4, X5. В нашем случае он равен:

Нормированный R-квадрат:

В отличие от коэффициента , который всегда увеличивается при добавлении новых объясняющих переменных, нормированный коэффициент детерминации может и уменьшаться; чем больше , тем уравнение регрессии более адекватно; Стандартная ошибка регрессии = 46,33.

2. а)В таблице «Дисперсионный анализ» в столбце «df» приводятся количества степеней свободы *m* = 4, *n* - *m - 1* = 17, *n* – 1 = 21 соответственно случайных величин приводятся в столбце «SS».

В столбце «MS», который показывает дисперсию на одну степень свободы, приведены значения величин:

,  
Отсюда

Проверка гипотезы

производится на основе анализа статистики

имеющей распределение Фишера с *m* и (*n* – *m* – 1) степенями свободы.

Гипотезу можно проверить и так: если значимость F оказывается больше принятого уровня значимости α = 0,05, то гипотезу принимают, это означает что уравнение регрессии статистически незначимо, не адекватно, а если значимость F оказывается меньше α, гипотезу отвергают, и говорят, что уравнение значимо, адекватно. Для данной модели значимость F = 0,34765. Таким, образом гипотеза принимается

Наблюдаемое значение статистики *F*α; *m*; *n*−*m*−1и рассчитанный уровень значимости гипотезы приводятся в столбцах «F» и «Значимость F» таблицы «Дисперсионный анализ».

б)Проверим теперь гипотезы : = 0 при альтернативах : ≠ 0*, j*=1,2,3,4,5.

В результатах работы программы «Регрессия» в столбце «t-статистика» приводятся значения статистики, которая при выполнении гипотезы имеет распределение Стьюдента с (*n* – *m* – 1) степенью свободы. Область отклонения гипотезы такова: >. В задаче значение статистики  *=* 0,23, = 0,16, = 0,25, = -0,69,

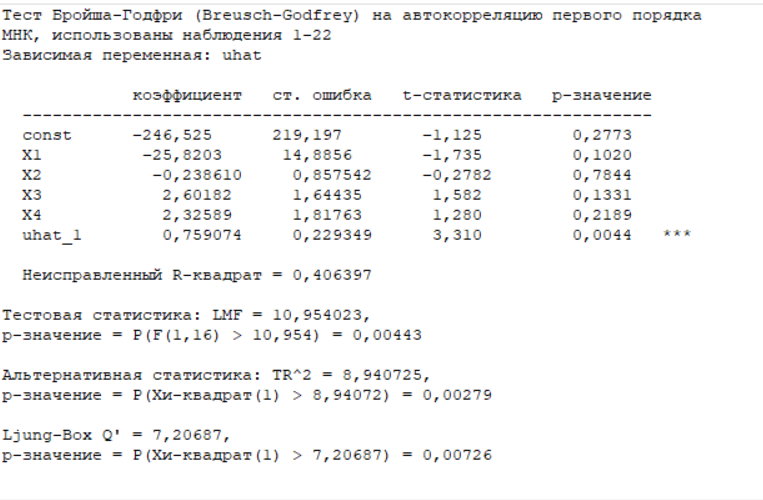
3. В построенном уравнении регрессии многие коэффициенты оказались незначимы. Такое уравнение нельзя считать приемлемым. Исключим из уравнения регрессор , P-значение которого равно 0,83 и является наибольшим. Проведем ту же последовательность действий для полученных моделей, записав их в таблицу N, в конце проектной работы. Результаты работы программы «Регрессия» к каждой модели представлены в приложении 2.

4. Наилучшим уравнением является полученное на третьем шаге, так как само уравнение и все его коэффициенты – значимы. Впишем все уравнения и их характеристики в Таблицу N, в приложении к проектной работе. Проверим полученные модели на автокорреляцию и гетероскедастичность, для этого перейдем в пункт 4.

# ВЕРИФИКАЦИЯ МОДЕЛЕЙ

a) Проверим нашу значимую модель через тест Бройша-Годфри. Для этого обратимся к выводу работы программы «Автокорреляция», пункту «Тесты» прикладного пакетного обеспечения Gretl (см. таблицу 4.1).

Таблица 4.1 - Проверка на автокорреляцию через тест Бройша-Годфри



Для Х1234 наблюдений по 5-м переменным р-значение составляет 0,00443.

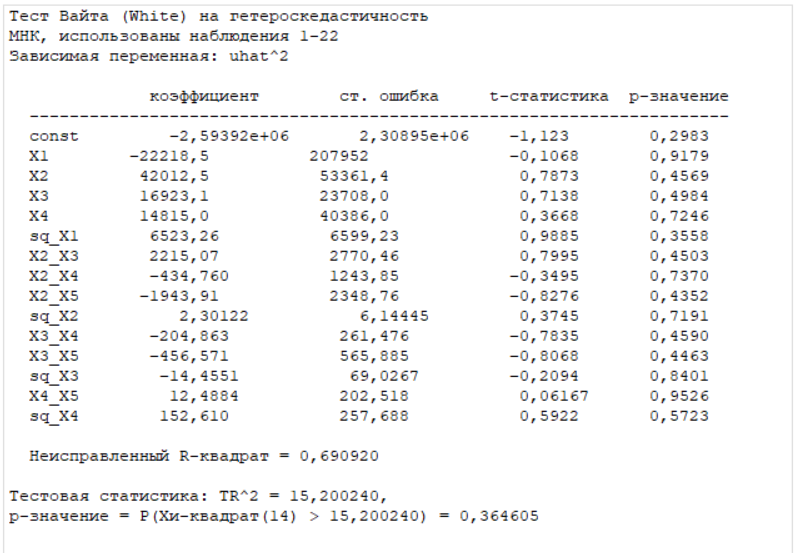
Если наблюдаемое р-значение меньше значения критической точки 0,05, можно сделать вывод, что в данных есть автокорреляции.

Если наблюдаемое значение меньше критической точки, говорят, что присутствует автокорреляция в остатках, нулевая гипотеза отклоняется и принимается альтернативная.

Проверим каждую модель на наличие автокорреляции и внесем результаты в таблицу N в конце.

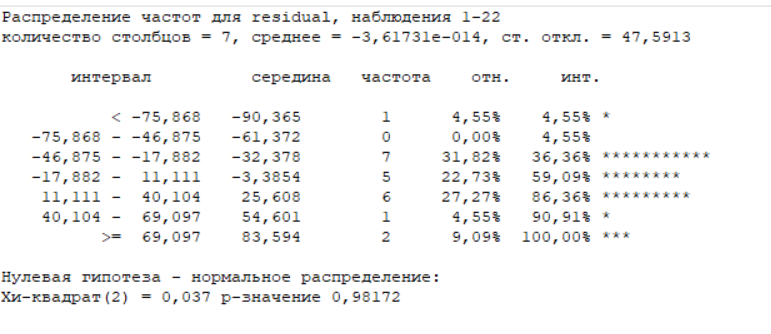
б) Проверим модель на гетероскедастичность, используя тест Вайта. Для этого обратимся к выводу работы программы «Гетероскедастичность», пункту «Тесты» прикладного пакетного обеспечения Gretl и выберем «Тест Вайта» (см. таблицу 4.2).

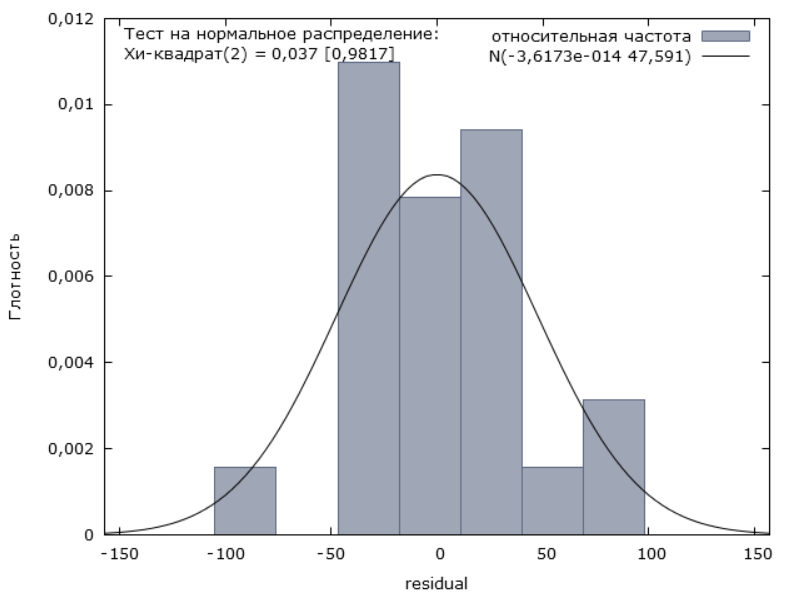
Таблица 4.2 - Проверка на Гетероскедастичность



Результаты теста Уайта показывают отсутствие гетероскедастичности, так как Р-вероятность принятия гипотезы о гетероскедастичности составляет 0,364, что больше 0,05. Результаты работы запишем в таблицу N в конце.

Воспользуемся критерием и рассчитаем его наблюдаемое значение. Построим график распределения остатков с помощью программного обеспечения Gretl (см. Рисунок 4.1 и Рисунок 4.2).

Рисунок 4.1 - Проверка остатков на нормальность

 Рисунок 4.2 - График распределения

Наблюдаемое значение статистики рассчитывается по формуле:

где n число наблюдений, интервальные частости, интервальная частота, v⃰ - число интервалов после их объединения, *l* — число неизвестных параметров нормального закона распределения.

По таблице для распределения Хи-квадрат найдем значение критической точки = 7,8. (по формуле выше считаеться я тупой не ъочу думать) Наблюдаемое значение статистики оказалось меньше критической точки, поэтому нет оснований отвергнуть нулевую гипотезу

о нормальном законе распределения остатков модели. Графики остатков модели находятся в Приложении 3 (см. Рисунок 1,2,3,4,5).

# ВЫБОР ЛУЧШЕЙ МОДЕЛИ

В полученных данных было 22 наблюдения и 5 факторов, 4 из которых независимые, и 1 зависящая.

Необходимо было выявить зависимость объема продаж(млн,руб)(X1), количества заболевших COVID-19(X2), средней цены препарата(X3), объема продаж(млн. уп)(X4) от объема выпущенных медицинских препаратов(Y)

Как видно из таблицы N, модель №3 является лучшей, так как коэффициент детерминации и нормированный коэффициент детерминации равны 22% и 5% соответственно, уровень значимости самого уравнения равен 0,6, гетероскедастичность и автокорреляция модели отсутствуют.

Модель №3:

y = 715,65-0,393 -1,986

Интерпретация лучшей модели:

1) Увеличение на 1 условную единицу числа заболевших обуславливает увеличение на 0,393 повышения объема выпущенной продукции.

2) Увеличение на 1 условную единицу объема продаж в млн. упаковок обуславливает уменьшение на 0,1986 объем выпущенной продукции

Полученная модель может рассматриваться в качестве генератора прогноза. Но для начала удостоверимся в правильности спецификации полученной модели

# **ЗА**КЛЮЧЕНИЕ

В данной работе были рассмотрены 3 федеральных округа РФ: Уральский, Сибирский и Дальневосточный. В ходе работы был выбран показатель экономической эффективности развития регионов и определен перечень факторов, потенциально оказывающих влияние на него.

Было получено 5 эконометрических моделей, из которых была выбрана одна наилучшая. Данную модель мы использовали в качестве генератора прогноза, в ходе чего были получены удовлетворительные результаты.

Также частью нашей работы была оценка тенденций и эффективности развития регионов. В ходе исследования регионы в рассматриваемых федеральных округах были разделены на 3 группы: наиболее преуспевающие, регионы со средними показателями эффективности, регионы с самыми низкими показателями эффективности. Также были оценены тенденции экономического развития и были выявлены регионы с положительной и отрицательной динамикой.

Безусловно, в дальнейшем полученная нами модель требует доработки, поэтому по мере изучения методов математического моделирования мы планируем повышать качество и прогностические возможности нашей модели.

**СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ**

1 Айвазян, С. А. Прикладная статистика: Основы моделирования и первичная обработка данных. Справочное изд. / С. А. Айвазян, И. С. Енюков, Л. Д. Мешалкин. — М.: Финансы и статистика, 1983. — 471 с.

2 Писарева, О. М. Методы прогнозирования и развития социально экономических систем: Учеб. Пособие/ О. М. Писарева. – М.: Высш. шк., 2007. – 591 с.

3 Методические указания к выполнению домашних заданий по дисциплине «Многомерные статистические методы» / Сост.: В. Н. Калинина, В. И. Соловьев; ГУУ. – М., 2005. – 47 с.

4 Куфель Т. Эконометрика. Решение задач с применением пакета программ Gretl. – M.: Горячая линия – Телеком, 2007. – 200 с.

5 Малова А. С. Основы эконометрики в среде Gretl: учебное пособие. – М.: Проспект, 2016. – 112 с.

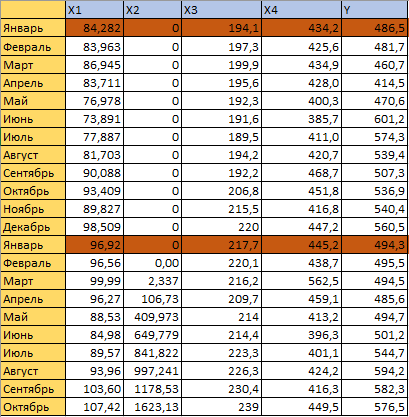
6 Wooldridge J. M. Introductory Econometrics. A modern approach, 5th edition. Michigan State University: South-Western Cengage Learning, 2013. 909p.

Таблица N - Модели

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Номер модели | Вид модели | Показатели качества | | | | | Спецификации |
|  |  | Значимость F | P – значение | Остатки |
| 1 |  |  |  | 0,3476 | P1 = 0,1060  P2= 0,8270  P3= 0,3349  P4= 0,6767  P5 = 0,7261 | Автокорреляция присутствует, гетероскедастичность  отсутствует | Исходные данные 2020 года.  Линейная регрессионная модель, построенная по 22 наблюдениям, по 5 факторам, 4 из которых - независимые |
| 2 | y = 503,1167 + 0,1380 + 1,4103 -2,9510 | 0,22 | 0,086 | 0,2099 | P1 = 0,0417  P2 = 0,9187  P3 = 0,0665  P4 = 0,2033 | Автокорреляция присутствует, гетероскедастичность  отсутствует | Исходные данные 2020 года.  Линейная регрессионная модель, построенная по 22 наблюдениям, по 4 факторам, 3 из которых являются независимыми. |
| 3 | y = 715,65-0,393 -1,986 | 0,2267 | 0,0514 | 0,60 | P1 = 0,0037  P2 = 0,7816  P3 = 0,4058 | Автокорреляция присутствует, гетероскедастичность  отсутствует | Исходные данные 2020 года.  Линейная регрессионная модель, построенная по 22 наблюдениям, по 3 факторам, 2 из которых являются независимыми. |
| 4 | y = 611,960-5,3902 + 0,843−3,956 | 0,4675 | 0,2186 | 0,2071 | P1 = 0,308  P2 = 0,839  P3 = 0,764  Р4= 0,493 | Автокорреляция присутствует, гетероскедастичность  отсутствует | Исходные данные 2020 года.  Линейная регрессионная модель, построенная по 22 наблюдениям, по 4 факторам, 3 из которых являются независимыми. |
| 5 | y = 223,7197-11,385 -0,138+ 2,516 | 0,4448 | 0,1979 | 0,2533 | P1 = 0,1823  P2 = 0,2751  P3= 0,9166  Р4= 0,0828 | Автокорреляция присутствует, гетероскедастичность  отсутствует | Исходные данные 2020 года.  Линейная регрессионная модель, построенная по 22 наблюдениям, по 4 факторам, 3 из которых являются независимыми. |

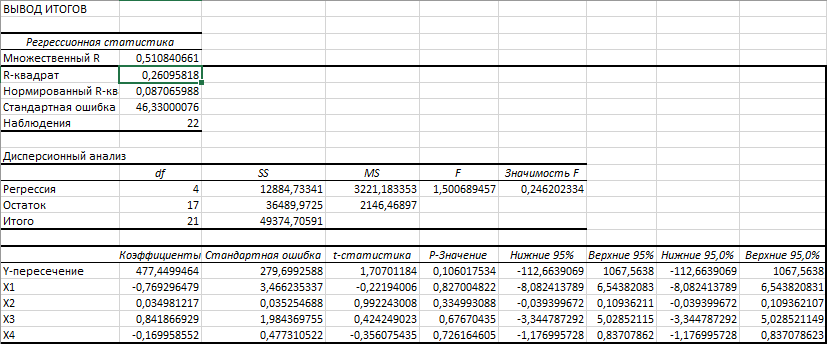
# ПРИЛОЖЕНИЕ 1

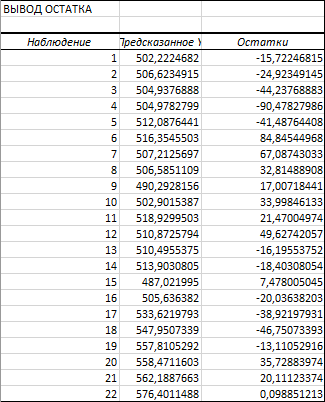
Рисунок А.1 - Таблица исходных данных по 2019-2020 году



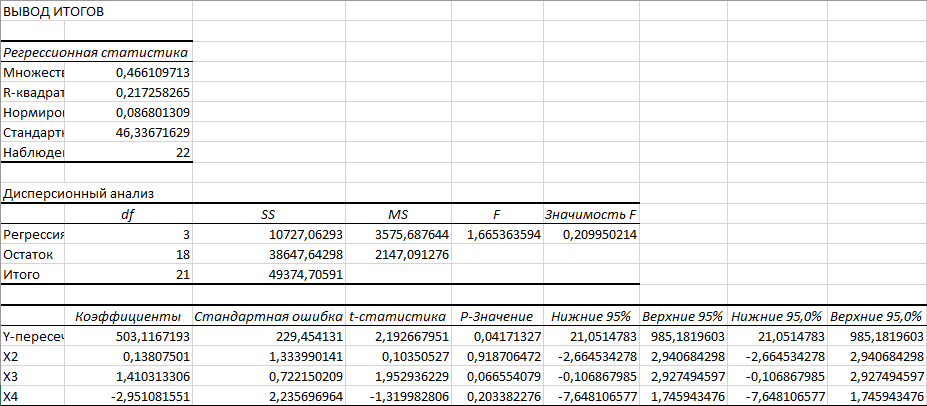
# ПРИЛОЖЕНИЕ 2

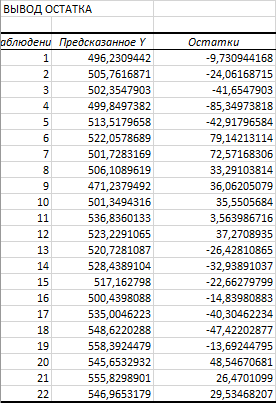
Общий регрессионный анализ



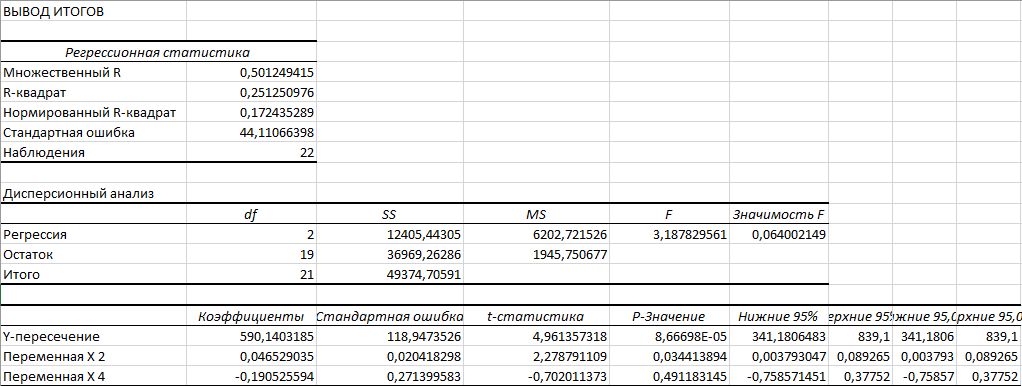


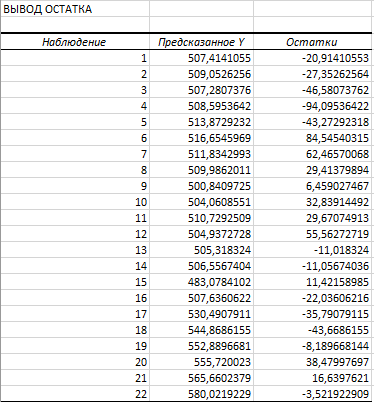
Регрессионный анализ без Х1



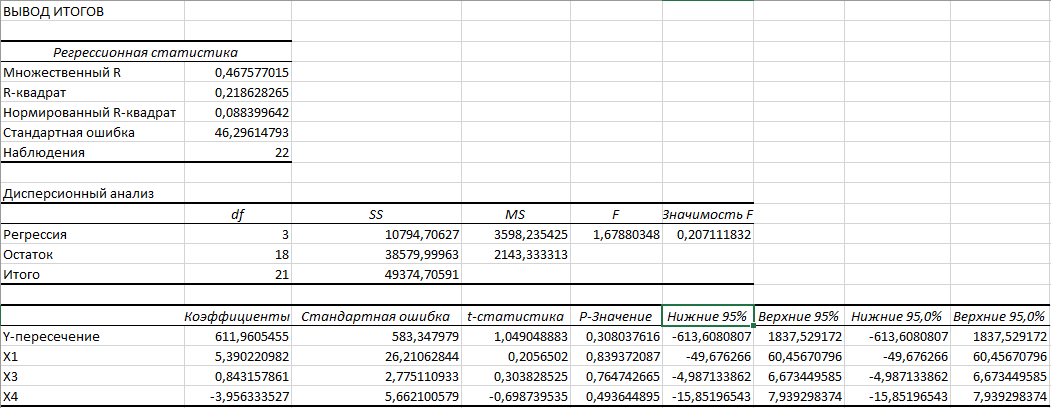


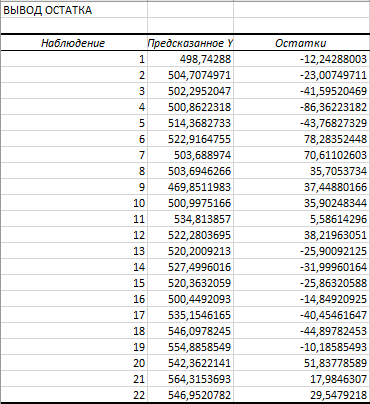
Регрессионный анализ без Х1 и Х3



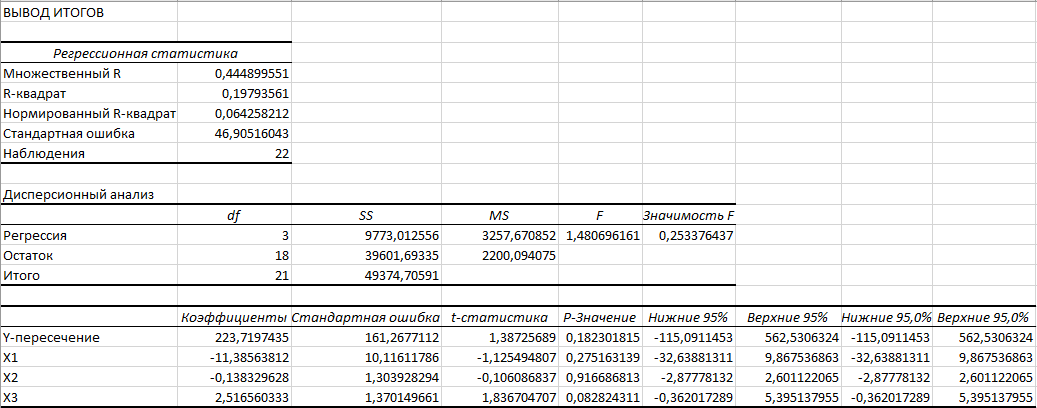


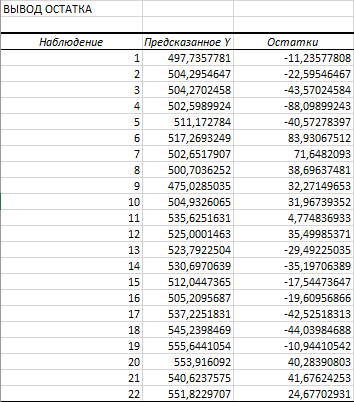
Регрессионный анализ без X2





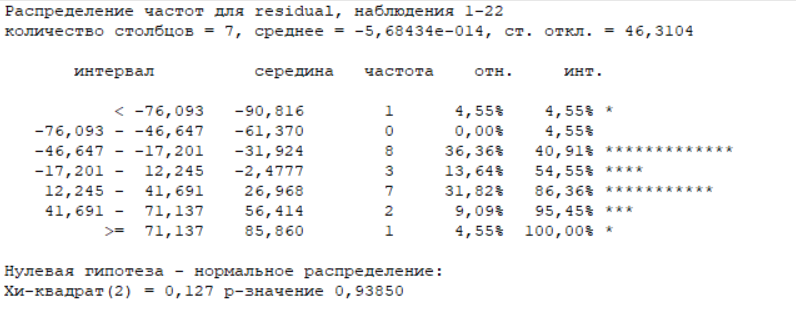
Регрессионный анализ без X4

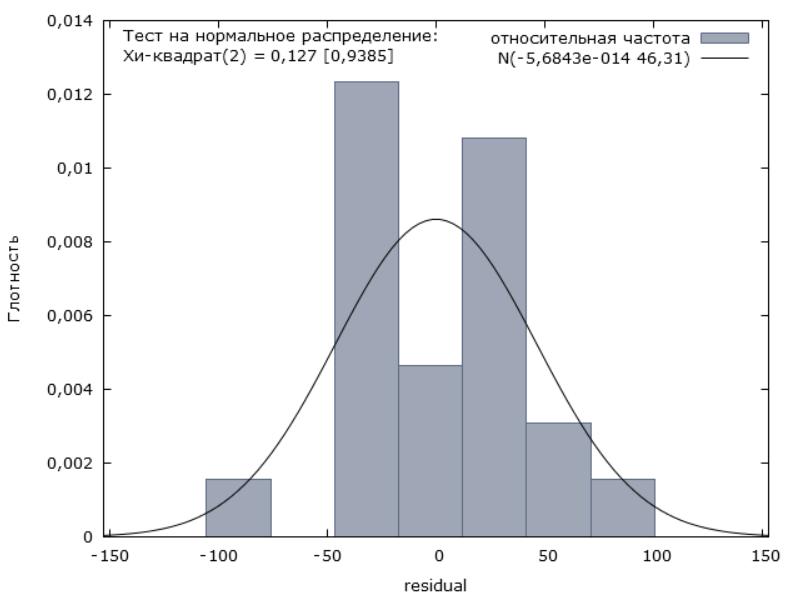


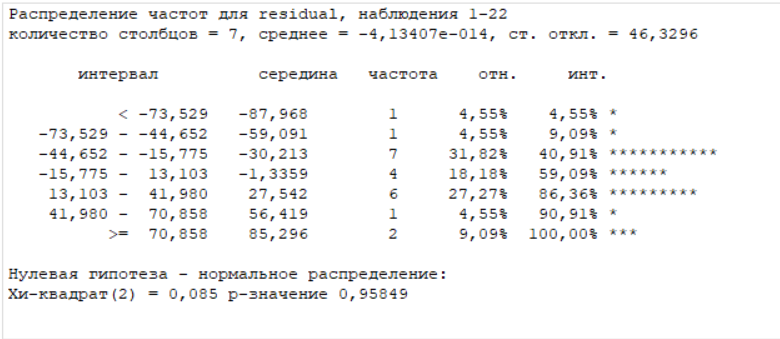


# ПРИЛОЖЕНИЕ 3

**Графики остатков каждой модели.**



 Рисунок 3.1 - Распределение остатков модели 2 (Без Х1)



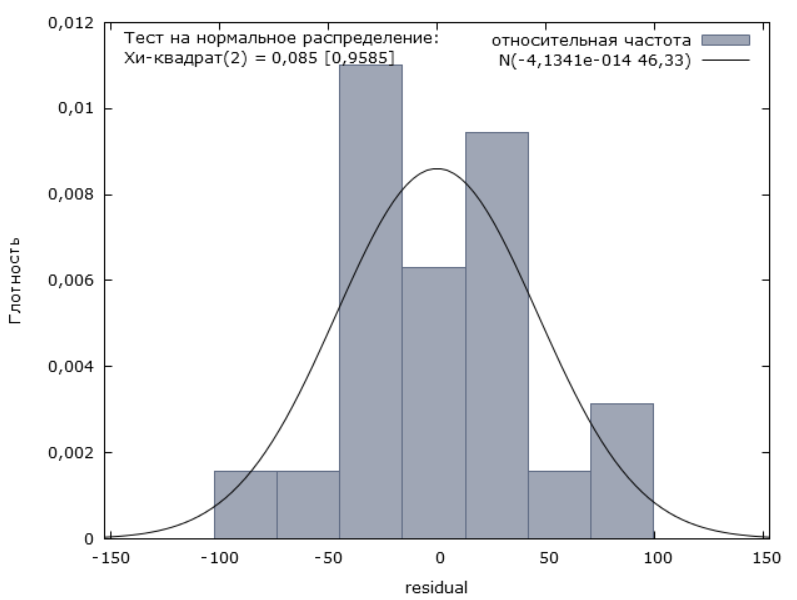
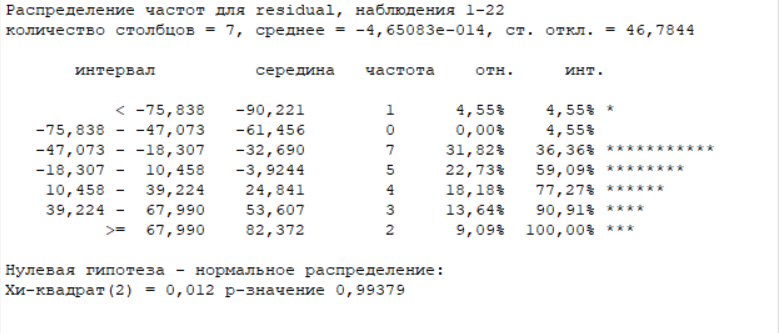


Рисунок 3.2 - Распределение остатков модели 4 (Без Х2)

**

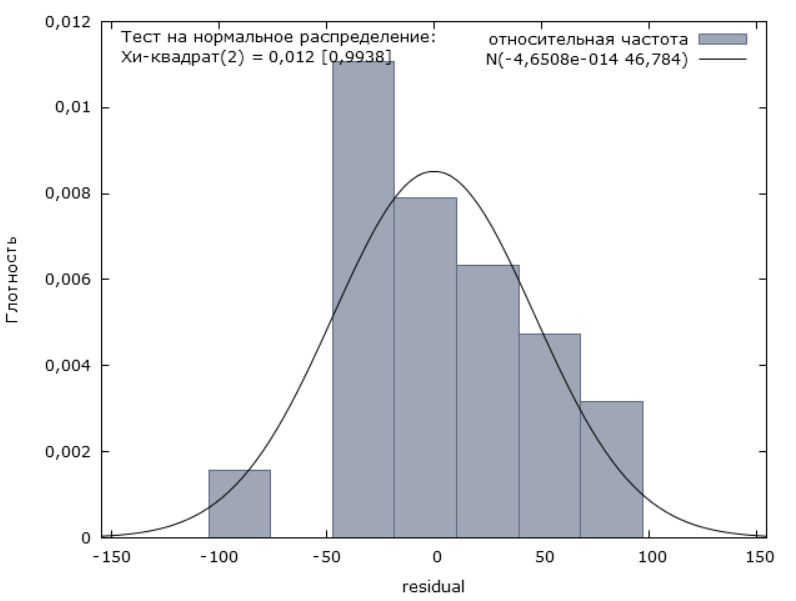
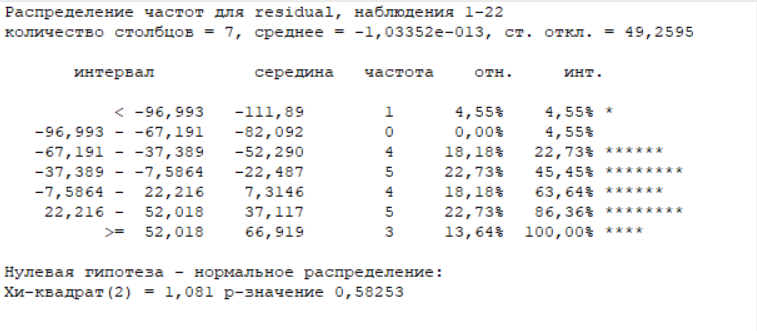
**

Рисунок 3.3 - Распределение остатков модели 5 (Без Х4)

**

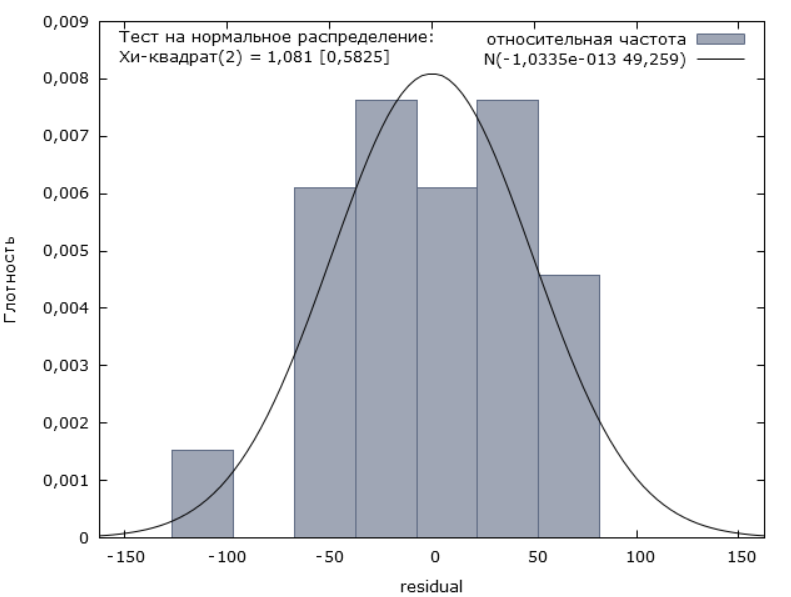
**

Рисунок 3.4 - Распределение остатков модели 3 (Без Х1 и Х3)