РАСЧЕТНО-АНАЛИТИЧЕСКАЯ РАБОТА

По дисциплине: Анализ данных

Москва, 2021

Оглавление

[1. Загрузка исходных данных 4](#_Toc72868907)

[2. Вычисление дополнительных признаков для каждой компании 4](#_Toc72868908)

[3. Исследование цен и объемов торгов 5](#_Toc72868909)

[4. Исследование логарифмических доходностей (логдоходностей) акций 10](#_Toc72868910)

[4.1. Исследование до удаления выбросов 10](#_Toc72868911)

[4.2. Удаление выбросов 13](#_Toc72868912)

[5.2. Интервальные оценки параметров логарифмических доходностей (в предположении нормального распределения) 15](#_Toc72868913)

[5.3. Проверка гипотезы о нормальности логдоходности 17](#_Toc72868914)

[5.3.1. Проверка гипотезы по критерию Пирсона (хи-квадрат) 17](#_Toc72868915)

[5.3.2. Проверка гипотезы по критерию Колмогорова-Смирнова 20](#_Toc72868916)

[5.3.3. Сравнение результатов проверки по двум критериям 22](#_Toc72868917)

[5.4. Построение диаграммы, содержащей гистограмму эмпирической плотности и график теоретической плотности 22](#_Toc72868918)

[5.5. Построение графиков эмпирической функции распределения и график теоретической функции распределения. 23](#_Toc72868919)

[6. Проверка гипотезы о равенстве нулю математического ожидания недельной логдоходности при альтернативной гипотезе о том, что оно больше нуля. 24](#_Toc72868920)

[7. Проверка гипотезы о равенстве дисперсий логарифмической доходности за последний и предпоследний годы (двусторонний F-тест) 26](#_Toc72868921)

[8. Проверка гипотезы о равенстве средних значений логарифмической доходности за последний и предпоследний годы (двусторонний t-тест) 27](#_Toc72868922)

[9. Сравнительный анализ логдоходностей трех компаний. 28](#_Toc72868923)

[9.1. Критерий Бартлетта 28](#_Toc72868924)

[9.2. Гипотеза о значимости зависимости значений логдоходности акций компании «ДВМП» от номера года. 29](#_Toc72868925)

[9.3. Гипотеза о значимости зависимости значений логдоходности акций компаний от названия компании. 30](#_Toc72868926)

[10. Графики временных рядов логдоходностей трех компаний 31](#_Toc72868927)

[11. Исследование тесноты связи между логдоходностями 33](#_Toc72868928)

# Загрузка исходных данных

Для выполнения расчетно-аналитической работы были отобраны акции

трех эмитентов: Акрон, ДЭК ао, Сбербанк.

На финансовом портале MFD.ru в разделе «Мосбиржа Акции и ПИФы» был произведен экспорт данных о котировках трех выбранных эмитентов в MS Excel. Временной интервал был выбран в соответствии с рекомендациями учебного пособия по расчетно-аналитической работе – с 1 января 2015 года до 12 февраля 2020 года. Результат данного шага представлен на рисунке 1.

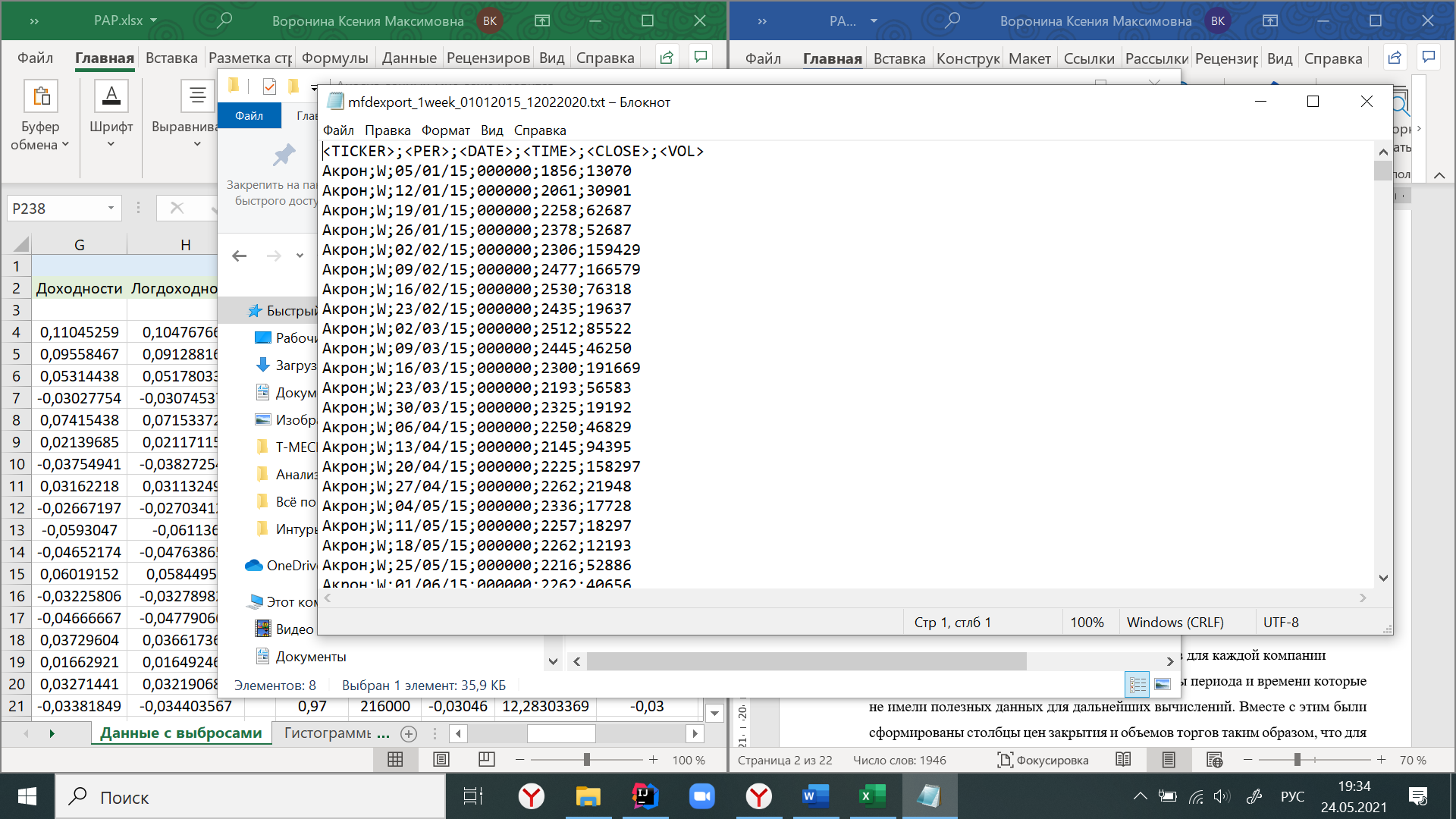


Рисунок 1. Первичные исходные данные

# 2. Вычисление дополнительных признаков для каждой компании

На следующем шаге были удалены столбцы периода и времени. Вместе с этим были сформированы столбцы цен закрытия и объемов торгов таким образом, что для каждого тиккера и каждого параметра был образован персональный столбец. Далее на основе данных о ценах и объёмах были рассчитаны показатели недельной доходности, логарифмической доходности, логарифма цены и логарифма объёма. Данные показатели были рассчитаны для всех пяти акций. Получилась довольно объёмная таблица, более подробно с которой можно ознакомиться в файле Excel на листе «Данные с выбросами».

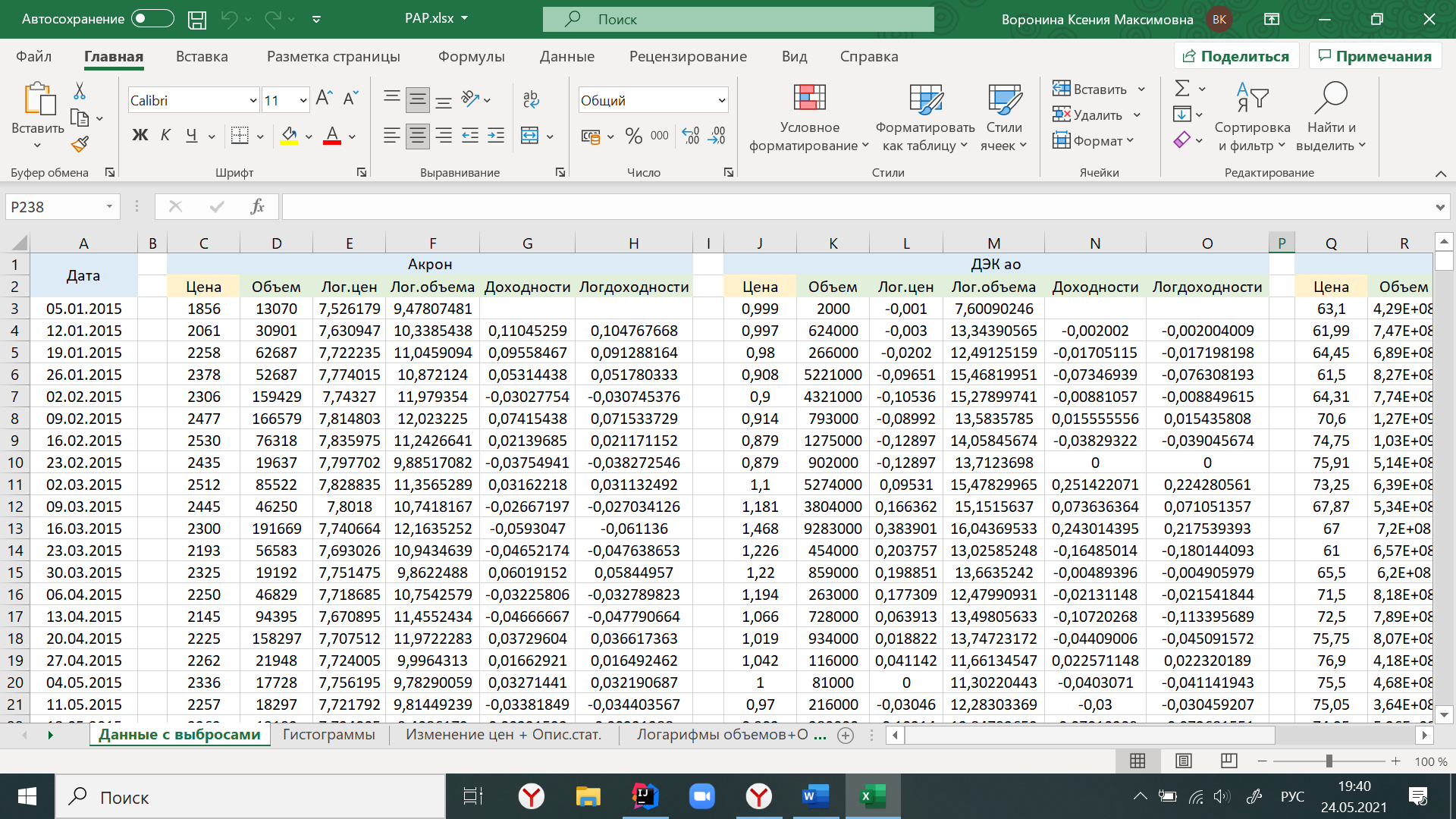
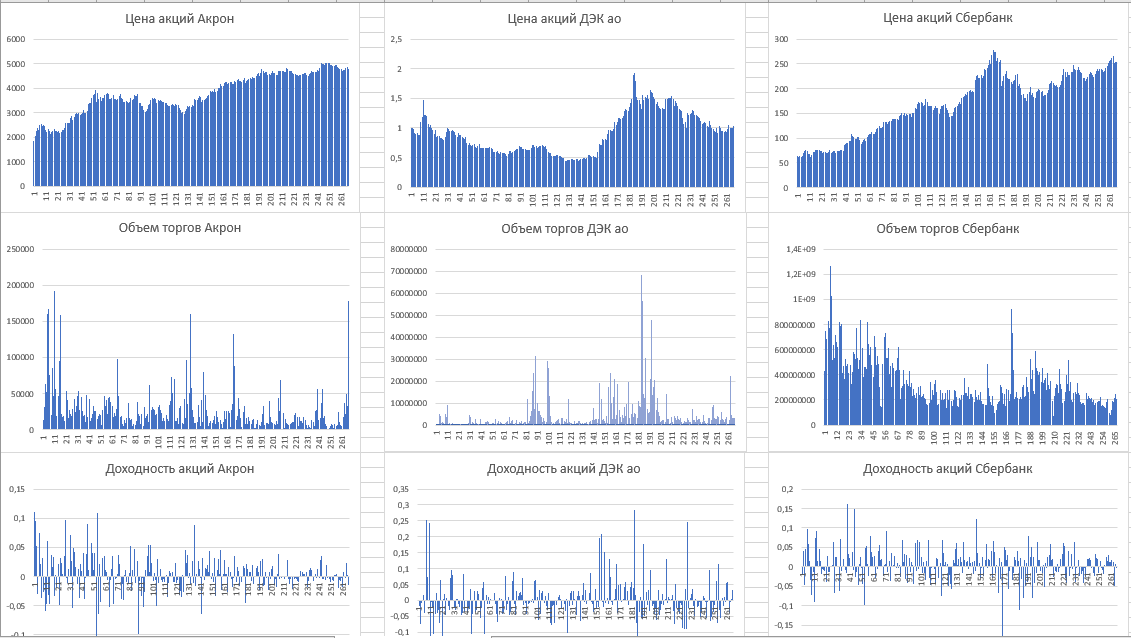


Рисунок 2. Сформированный лист исходных данных

# 3. Исследование цен и объемов торгов

Для получения первоначального представления о поведении исследуемых величин используем средства визуализации, в данном случае воспользуемся сервисом MS Excel под названием «Вставка» для построения графиков.

Первый этап – построение гистограмм для исследуемых величин для каждой компании. Все построенные гистограммы можно найти на листе «Гистограммы» в прилагаемом файле.



Поскольку цены акций выбранных компаний отличаются на несколько порядков, их изображение на одном графике нецелесообразно, поэтому воспользуемся инструментом относительных цен, которые определяются по формуле:

где

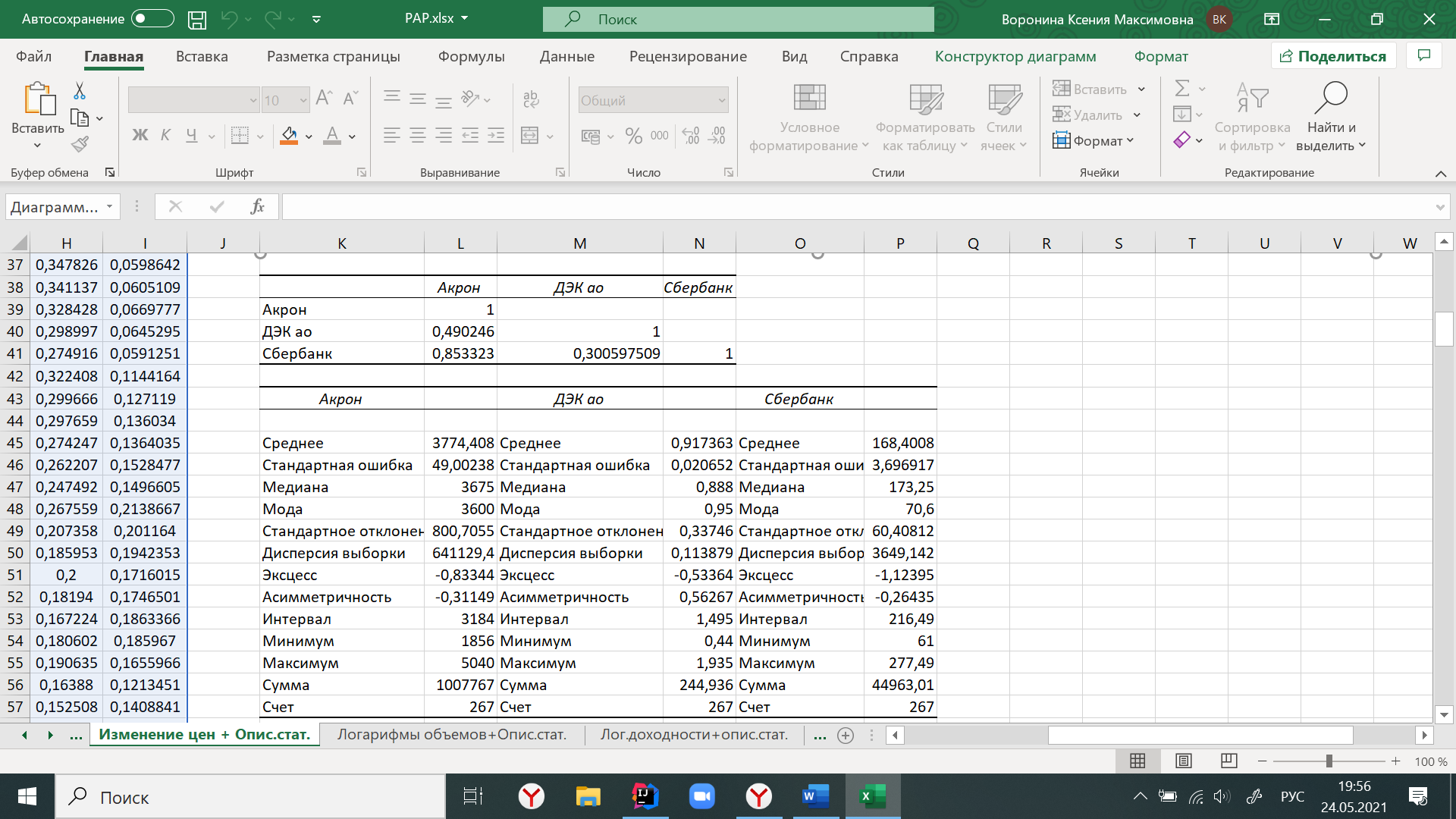
- соответственно минимальное и максимальное значения цены.

Рассчитанные данные об относительных ценах можно найти на листе Excel под названием «Изменение цен+опис.стат.». Далее по ним был построен линейный график, иллюстрирующий динамику относительных цен компаний, который изображен на рисунке 3.

Рисунок 3. Динамика относительных цен

На нем мы можем наблюдать рост акций «Акрон» за весь период, резкий рост акций «Сбербанка» с конца 2017 года, а также стремительный рост акций «ДЭК ао», после – спад.

Также была сформирована описательная статистика и корреляционная матрица для цен трех компаний (рисунок ниже).



Далее для изучения корреляции между объема торгов акций трех компаний были построены графики рассеяния и сформирована корреляционная матрица с помощью сервиса «Анализ данных». Данные графики можно найти, обратившись к листу «Логарифмы объемов+опис.стат.» в расчетном файле Excel.

Рисунки 4. Диаграмма рассеяния значений признаков «логарифмов объемов торгов» акций.

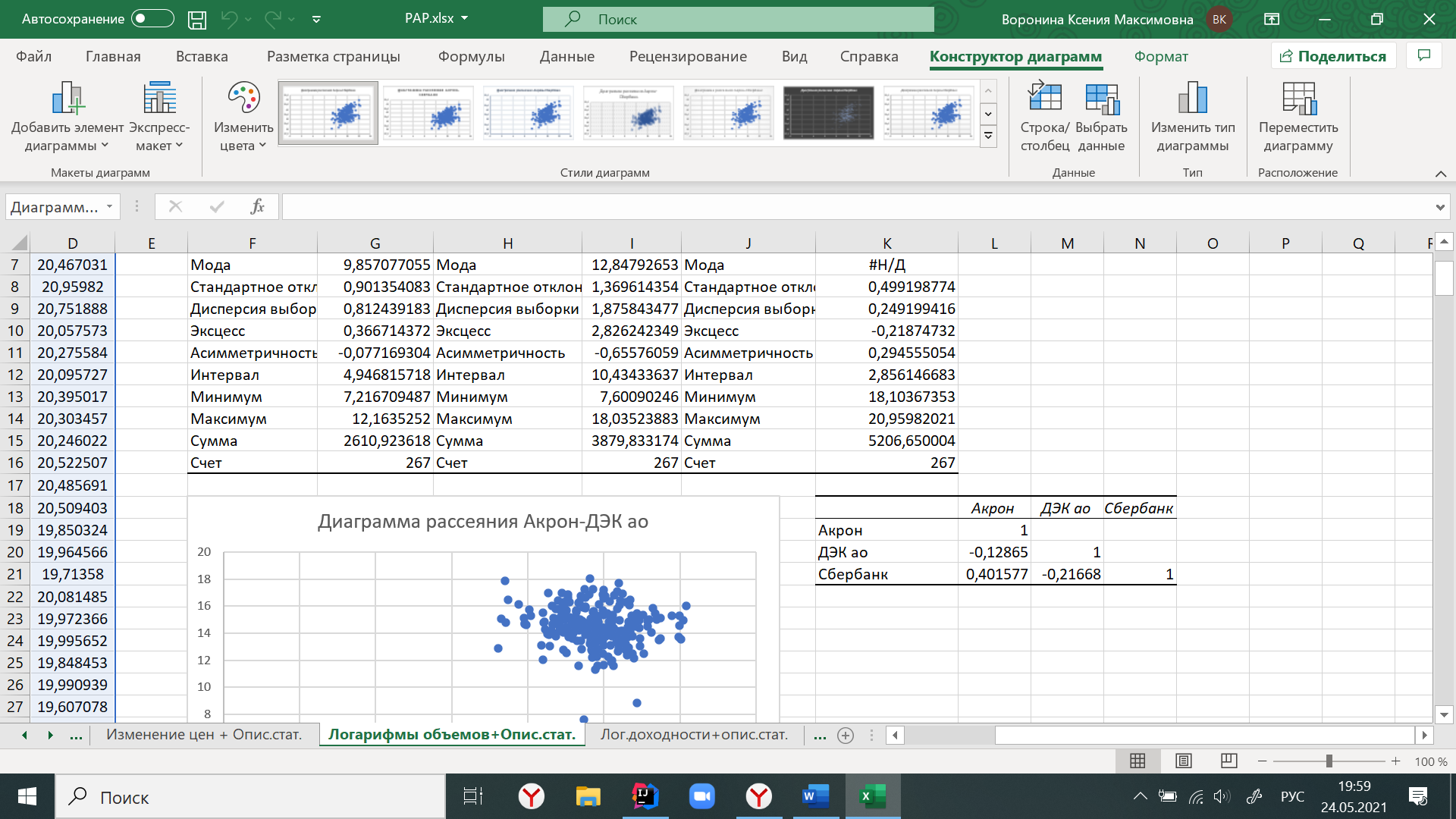
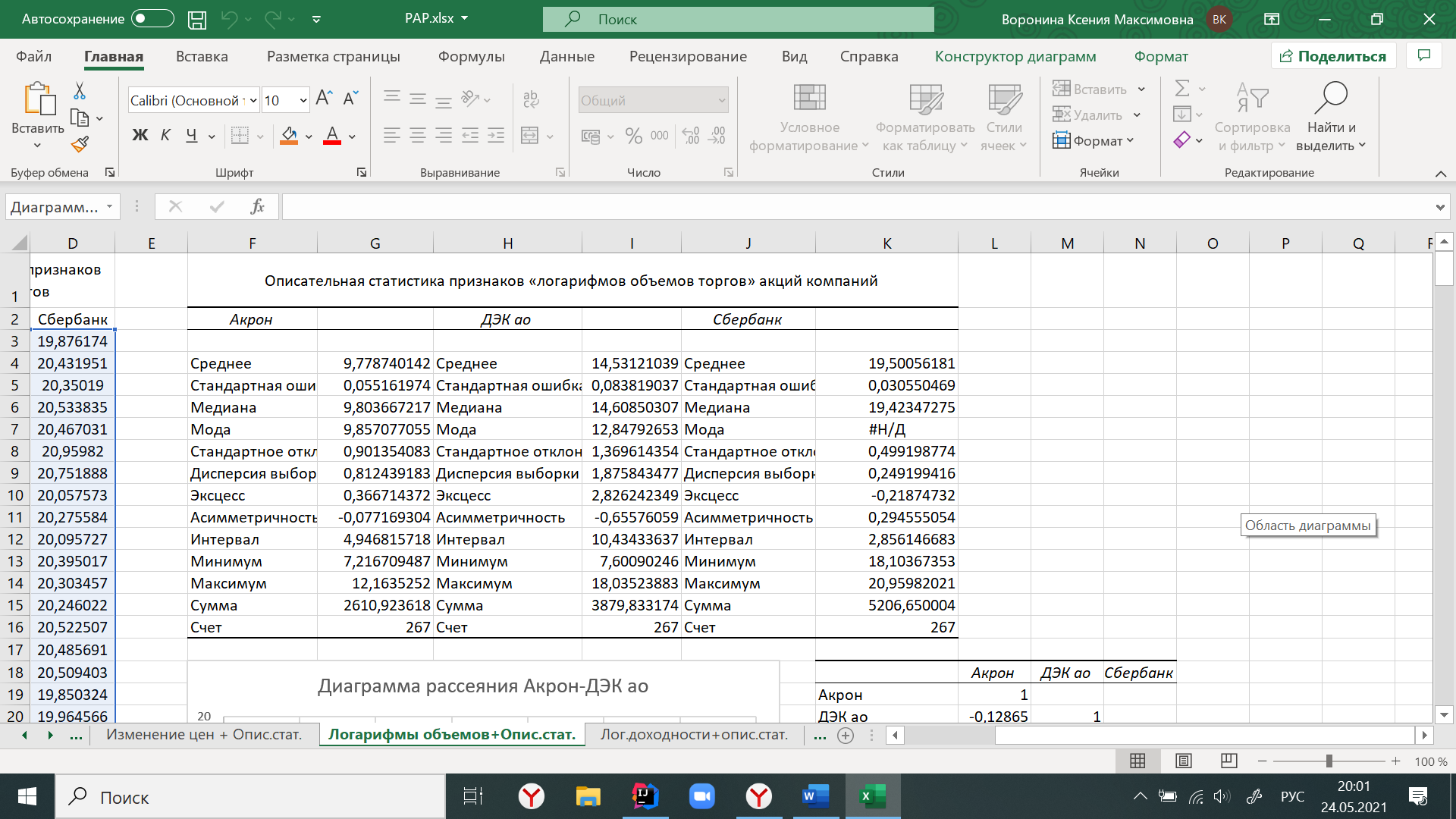


Рисунок 5. Корреляционная матрица значений признаков «логарифмов объемов торгов»

Проанализировав три графика рассеивания логарифмов объемов и обратившись к корреляционной матрице, можно сказать, что какая-либо связь между логарифмами объемов торгов компаний практически отсутствует. Также была сформирована описательная статистика для логарифмов объемов.



# Исследование логарифмических доходностей (логдоходностей) акций

## **Исследование до удаления выбросов**

Сгенерируем корреляционную матрицу логарифмических доходностей компаний до удаления выбросов с помощью сервиса «Анализ данных». Полученные расчеты можно увидеть, обратившись к листу «Логарифм доходности+опис.стат.» в расчетном документе Excel.

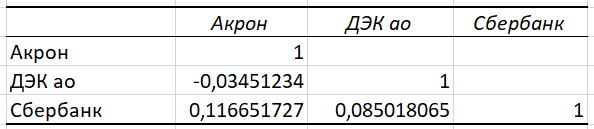
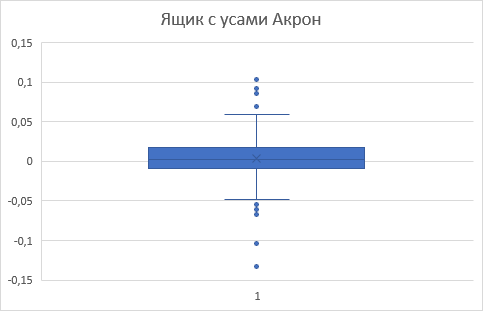
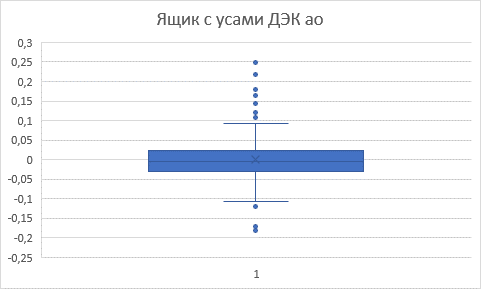
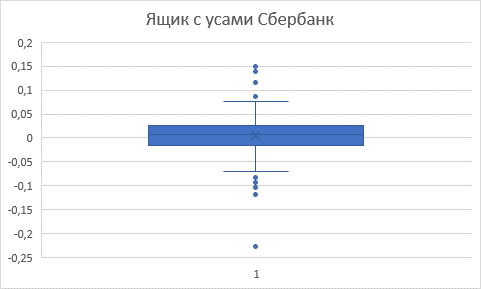


Рисунок 8. Корреляционная матрица значений признаков «логарифмические доходности»









На рисунках выше представлены диаграммы рассеяния и диаграммы «Ящик с усами» для логарифмических доходностей компаний до удаления выбросов, а также описательная статистика.

Как мы можем видеть, корреляция между логдоходностями выбранных компаний почти отсутствует, либо она очень слабая, что может быть обусловлено разными финансовыми показателями, имеющимися у компаний за выбранный период.

## **Удаление выбросов**

Следующим шагом было удаление выбросов из выборки. Были рассчитаны нижние и верхние границы для логдоходностей каждой компании.

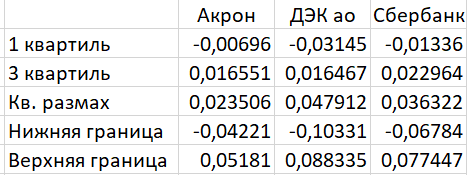
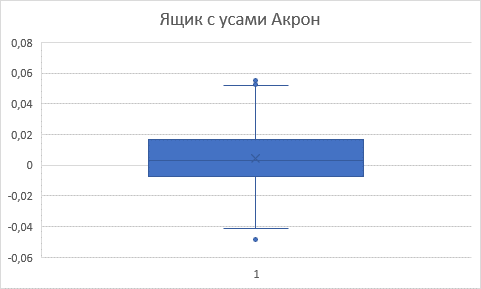
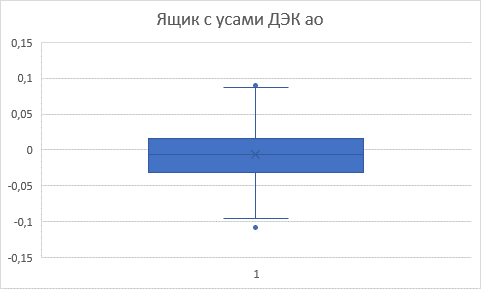
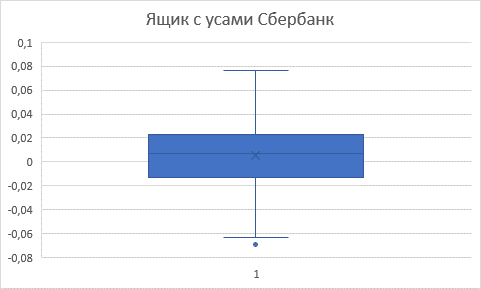


Рисунок 10. Расчет границ для удаления выбросов

Далее с помощью сортировки были найдены выбросы и удалены из выборки. По данным без выбросов были построены требуемые диаграммы, а также сформированы таблицы с помощью сервиса «Описательная статистика». Все расчеты можно найти, обратившись к Excel.







# 5.2. Интервальные оценки параметров логарифмических доходностей (в предположении нормального распределения)

Для расчета доверительного интервала для оценки математического ожидания, дисперсии и квадратичного отклонения был использован сервис «Описательная статистика» для каждой компании, а также статистические формулы ДОВЕРИТ.СТЬЮДЕНТ и ХИ2.ОБР. Результаты расчетов представлены ниже, их также можно найти на листе «Доверительные интервалы» в рабочем файле Excel.



Рисунок 12. Результат процедуры «Описательная статистика».

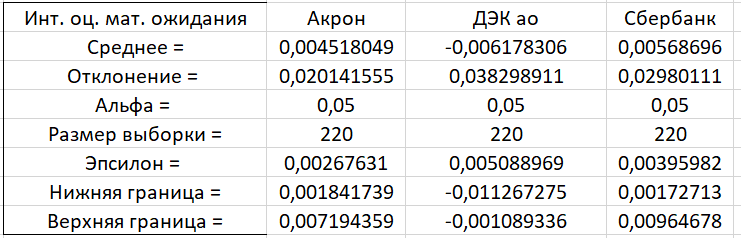


Рисунок 13. Интервальное оценивание математического ожидания.

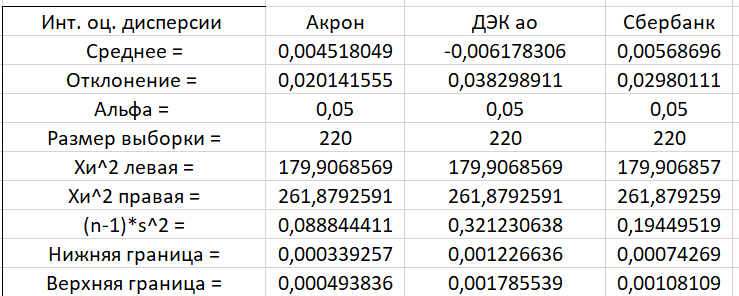


Рисунок 14. Интервальное оценивание дисперсии.

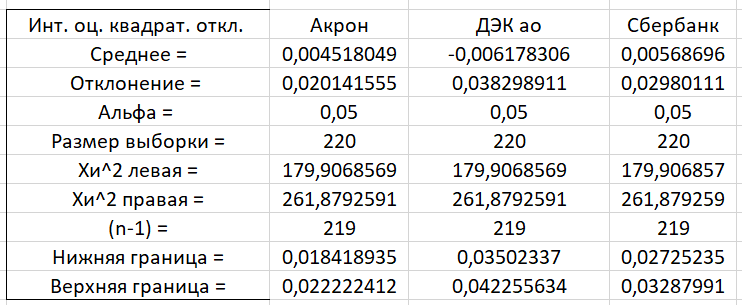


Рисунок 15. Интервальная оценка квадратичного отклонения.

Данный анализ показывает, что верхняя граница логарифмической доходности компании «Сбербанк» выше, чем у остальных компаний. Также он показывает, что нижняя граница для «ДЭК ао» является отрицательной величиной, что свидетельствует о возможной вероятности убытков у данной компании.

# 5.3. Проверка гипотезы о нормальности логдоходности

## **5.3.1. Проверка гипотезы по критерию Пирсона (хи-квадрат)**

Используем критерий согласия Пирсона (χ2) для проверки гипотезы о соответствии эмпирического распределения данных о котировках акций предполагаемому теоретическому нормальному распределению F(x) при объеме выборки n> 100.

Н0: данные соответствуют нормальному распределению (логдоходность)

Н1: данные не соответствуют нормальному распределению (логдоходность)

С помощью сервиса «Описательная статистика» и правила Стерджеса были рассчитаны количество интервалов и их длина. Далее были рассчитаны границы сформированных интервалов, эмпирические и теоретические интервальные частоты. В итоге были получены значения критерия для каждой отдельной компании и произведено сравнение с критическими значениями. Все расчеты представлены ниже (с ними также можно ознакомиться на листе «критерий Пирсона» в расчетном документе Excel).

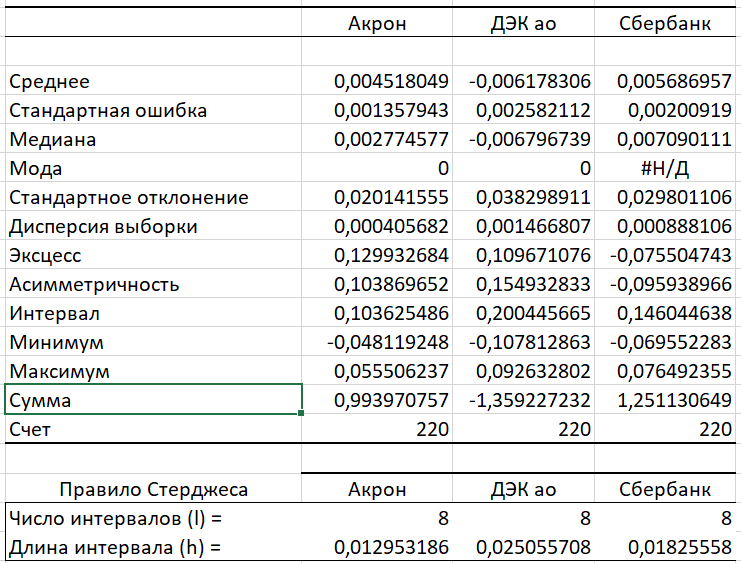


Рисунок 16. Расчет количества и длины интервалов



Рисунок 17. Расчет границ интервалов и эмпирических интервальных частот.

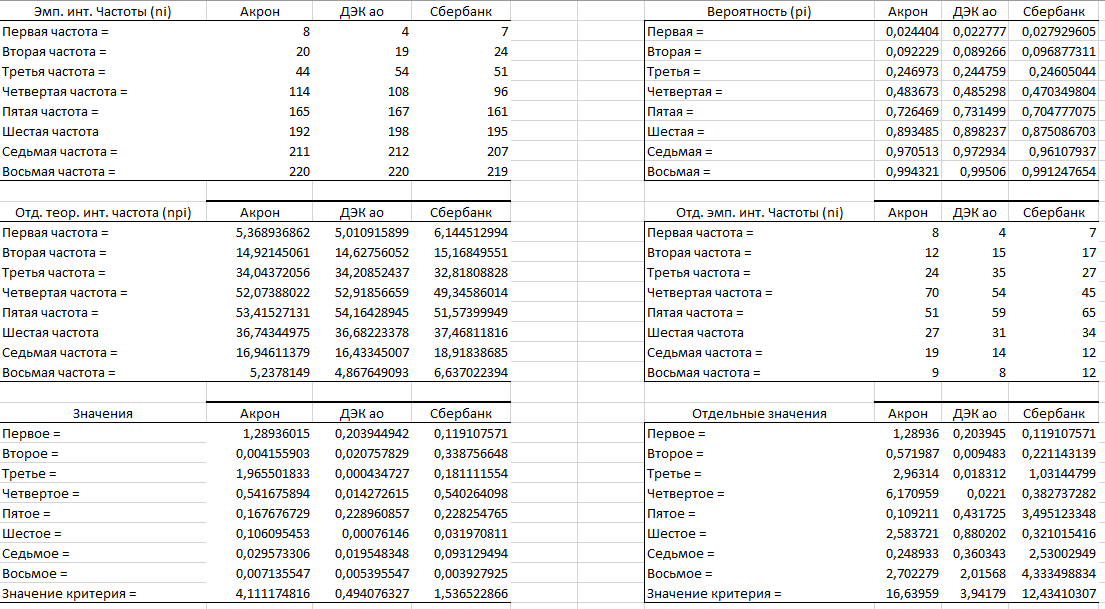


Рисунок 18. Расчет значений критерия согласия Пирсона для каждой компании



Рисунок 19. Расчет критических значений критерия согласия Пирсона для каждой компании

Таким образом, мы получили, что для «ДЭК ао» на любом уровне принимается гипотеза Н0 о нормальном распределении. Для «Акрона» на любом уровне гипотезу Н0 о нормальном распределении отклоняем. Для «Сбербанка» Н0 принимается на уровне значимости равном 0.01, но отклоняется на уровне значимости 0.05.

## **5.3.2. Проверка гипотезы по критерию Колмогорова-Смирнова**

Также в работе использовался критерий Колмогорова-Смирнова, который наряду с критерием Пирсона применяется для проверки гипотезы о том, что генеральная совокупность подчинена заданной непрерывной функции распределения .

Н0: данные соответствуют нормальному распределению (логдоходность)

Н1: данные не соответствуют нормальному распределению (логдоходность)

Все расчеты представлены в расчетном документе Excel на листах «Критерий Колмогорова» для каждой соответствующей компании. Ниже приведем пример расчета значений критерия для компании «Сбербанк».

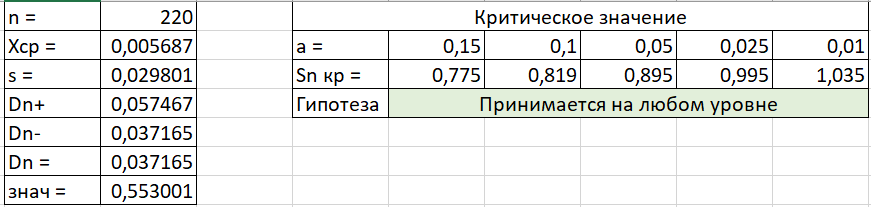
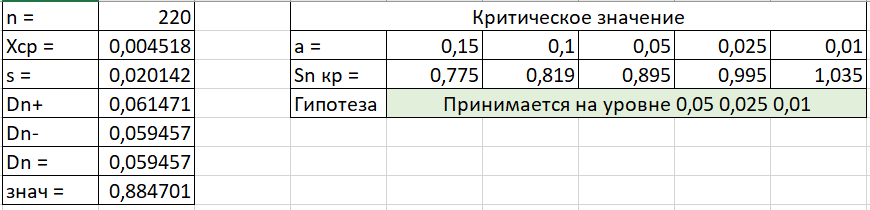


Рисунок 20. Расчет значений критерия Колмогорова-Смирнова для компании «Сбербанк».

Расчет показателей для остальных компаний показал аналогичный результат. Рассчитанные значения критерия оказались ниже критических, кроме компании «Акрон» (на уровнях значимости 0.15 и 0.1 отклоняется), поэтому мы можем сделать вывод, что у нас нет оснований отвергнуть ранее выдвинутую гипотезу.



## **5.3.3. Сравнение результатов проверки по двум критериям**

Таким образом, расчеты критерия согласия Пирсона и критерия Колмогорова-Смирнова дали одинаковые результаты для компании «ДЭК ао», что позволяет принять гипотезу о принадлежности выборки логарифмических доходностей трех компаний к нормальному распределению. Но результаты расчетов для «Акрон» и «Сбербанк» различаются, что говорит о неоднозначности, и не дает уверенности отнести распределения к нормальному.

# 5.4. Построение диаграммы, содержащей гистограмму эмпирической плотности и график теоретической плотности

# 5.5. Построение графиков эмпирической функции распределения и график теоретической функции распределения.

# 6. Проверка гипотезы о равенстве нулю математического ожидания недельной логдоходности при альтернативной гипотезе о том, что оно больше нуля.

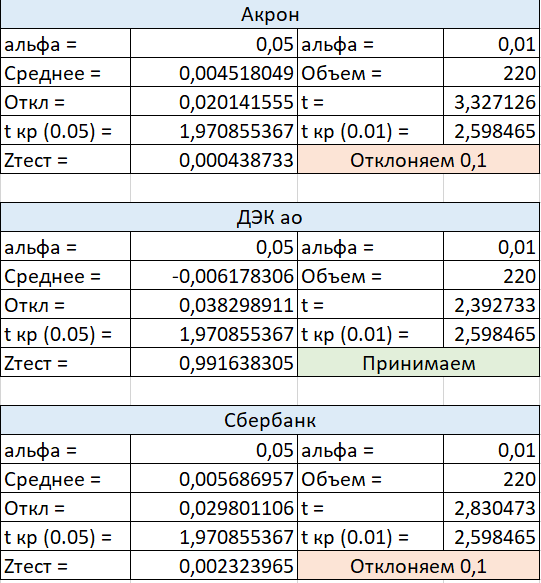
Далее выдвигается гипотеза о равенстве нулю математического ожидания недельной логдоходности при альтернативной гипотезе о том, что оно больше нуля для каждой выбранной компании.

Н0: математическое ожидание равно нулю для каждой компании

Н1: математическое ожидание больше нулю для каждой компании

В расчетном документе Excel был создан новый лист «Z-test». На этот лист скопированы данные о логдоходностях акций трех компаний, оставшиеся после удаления выбросов.

Для вычисления необходимых значений были использованы функции СРЗНАЧ и СТЬЮДЕНТ.ОБР.



Поскольку вычисленные значения критерия для «ДЭК ао» оказались менее критических значений, как при , так и при (см. Рисунок 19), то нет оснований отвергнуть гипотезу Н0 о равенстве нулю математического ожидания недельной логдоходности данного актива. Для компаний «Акрон» и «Сбербанк» мы вынуждены отклонить гипотезу Н0 на уровне значимости 0.01.

Также была использована статистическая функция Z.ТЕСТ. Рассчитанные с помощью нее значения оказались ниже уровней значимости, равных 0,05 и 0,01, для компаний «Акрон» и «Сбербанк». Для компании «ДЭК ао» проводимый тест показал результат выше уровней значимости.

Таким образом, можно сделать вывод, что нет оснований отвергнуть гипотезу Н0 о равенстве нулю математического ожидания недельной логдоходности активов только для компании «ДЭК ао».

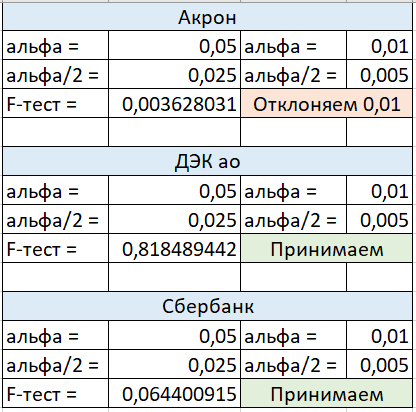
# 7. Проверк**а** гипотезы о равенстве дисперсий логарифмической доходности за последний и предпоследний годы (двусторонний F-тест)

Выдвигается гипотеза о равенстве дисперсий рассматриваемых генеральных совокупностей.

Н0: дисперсии рассматриваемых генеральных совокупностей равны

Н1: дисперсии рассматриваемых генеральных совокупностей не равны

Обратимся к статистической функции F.ТЕСТ, которая возвращает вероятность сходства дисперсий тестируемых выборок. Рассчитанные показатели оказались больше уровней значимости, кроме компании «Акрон». Это позволяет сделать вывод, что нет оснований отвергнуть гипотезу о равенстве дисперсий логарифмической доходности за 2018 и 2019 годы для «ДЭК ао» и «Сбербанка», но есть основания отвергнуть эту гипотезу для «Акрона».



# 8. Проверка гипотезы о равенстве средних значений логарифмической доходности за последний и предпоследний годы (двусторонний t-тест)

Для решения данной применим критерий для проверки гипотезы о сравнении двух математических ожиданий.

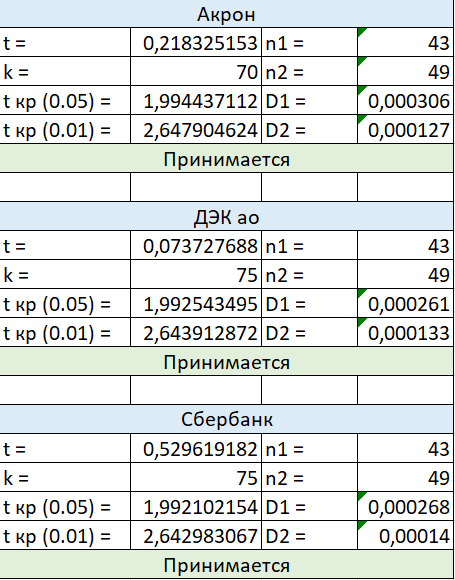
Н0: средние значения логдоходности за последний и предпоследний годы равны

Н1: средние значения логдоходности за последний и предпоследний годы не равны

Рассмотрим решение указанной задачи на примере логдоходности акций выбранных компаний.

С помощью статистической формулы ТТЕСТ рассчитаем показатель T для каждой из компаний. Затем рассчитаем количество степеней свободы k с помощью следующей формулы:

Полученное значение используем в расчете критических значений критерия t с помощью функции СТЬЮДЕНТ.ОБР.

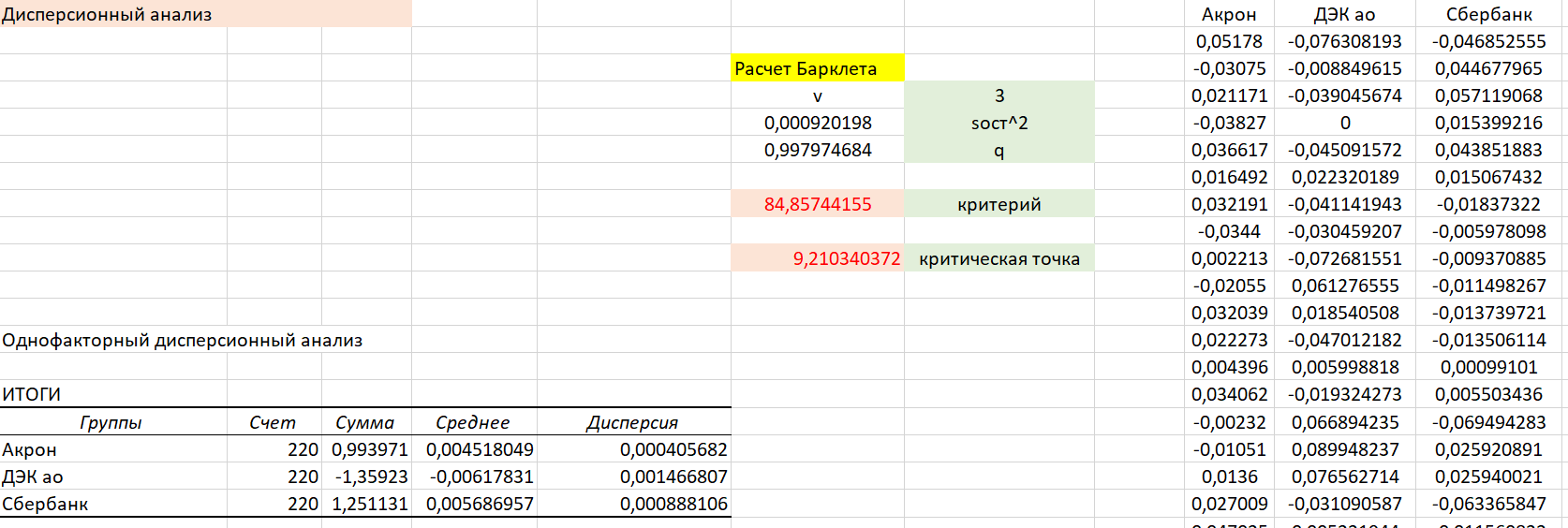


Значения критериев ниже критических значений на уровнях значимости 0,05 и 0,01, значит, нет оснований отвергать гипотезу Н0 о равенстве средних значений логдоходностей за 2018 и 2019 годы для всех компаний.

# 9. Сравнительный анализ логдоходностей трех компаний.

## **9.1. Критерий Бартлетта**

Перед проверкой следующих гипотез необходимо провести проверку на равенство дисперсий критерием Бартлетта.



Наблюдаемое числовое значение статистики Бартлетта оказалось больше критической точки, поэтому есть основания отвергнуть проверяемую гипотезу, таким образом, применение дисперсионного анализа необосновано. Это следует учитывать при дальнейшей проверке гипотез дисперсионным анализом.

## **9.2. Гипотеза о значимости зависимости значений логдоходности акций компании «ДВМП» от номера года.**

Выдвигается гипотеза об отсутствии зависимости значений логдоходности акций компании.

Н0: зависимость значений логдоходности акций компании отсутствует

Н1: присутствует зависимость значений логдоходности акций компании

Чтобы сравнить логдоходности за несколько лет обратимся к сервису «Однофакторный дисперсионный анализ» в разделе «Анализ данных». Результаты процедуры представлены с помощью двух таблиц (см. Рисунок 29). Все расчеты можно найти на рабочем листе «Дисп анализ Акрон».

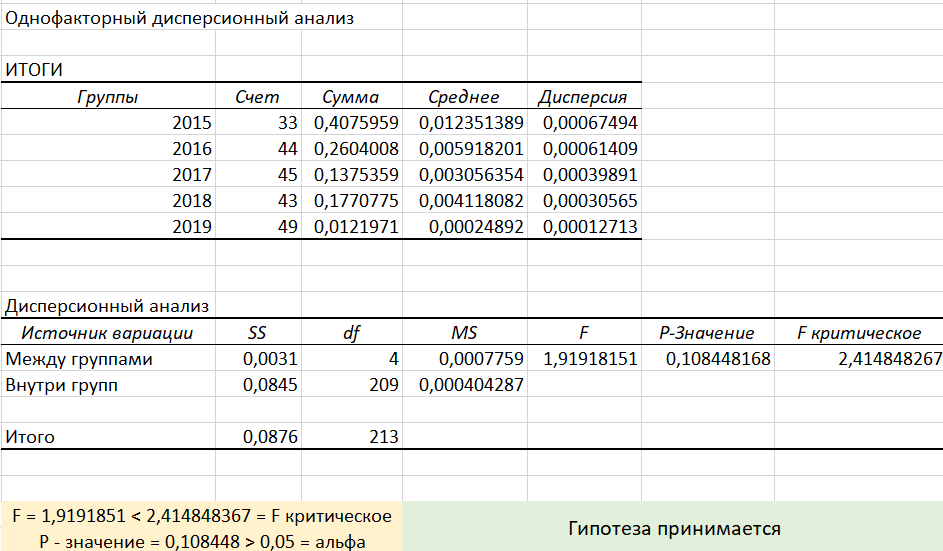


Рисунок 29. Результаты процедуры «Однофакторный дисперсионный анализ» для компании «Акрон».

В результате можно сравнить полученные показатели. Критерий фактического F – распределения оказался меньше критического, а Р-значение превысило заданный уровень значимости. Таким образом, отсутствуют основания отвергнуть выдвинутую гипотезу Н0.

## **9.3. Гипотеза о значимости зависимости значений логдоходности акций компаний от названия компании.**

Данная гипотеза также рассматривается с использованием сервиса «Однофакторный дисперсионный анализ».

Критерий фактического F – распределения оказался больше критического, а Р-значение не превысило заданный уровень значимости. Таким образом, есть основания отвергнуть выдвинутую гипотезу (см. Рисунок 30). Расчеты представлены на листе «Дисп анализ 2» в расчетном файле Excel.

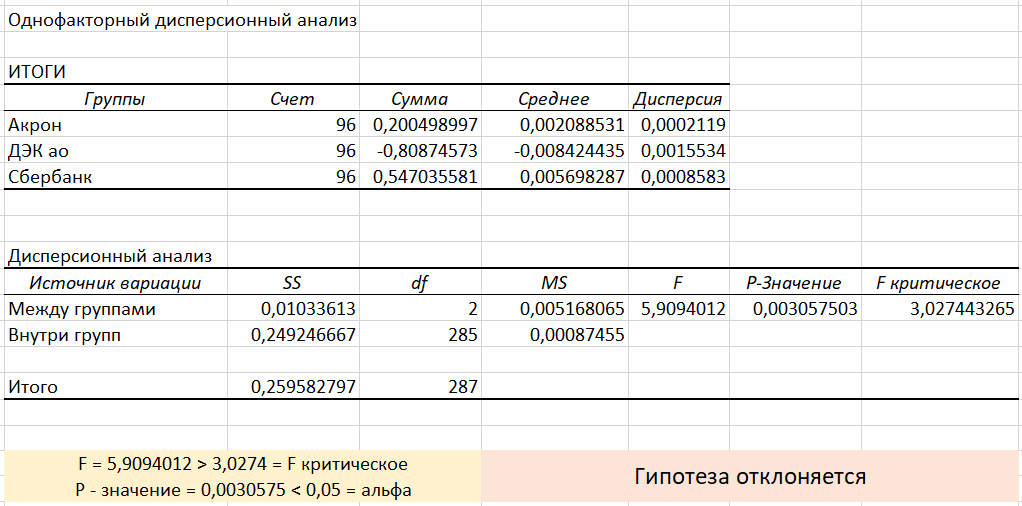


Рисунок 30. Результаты процедуры «Однофакторный дисперсионный анализ» для трех компаний.

# 10. Графики временных рядов логдоходностей трех компаний

Для визуализации имеющихся данных о динамике показателей логдоходности компаний используем сервис «Вставка» для построения графиком (см. Рисунок 31, 32 и 33). Данные графики представлены на рабочем листе «Зависимость ЛД от Т» в расчетном файле Excel.

Рисунок 31. Зависимость показателей логдоходности компании «Акрон» от времени.

Рисунок 32. Зависимость показателей логдоходности компании «ДЭК ао» от времени.

Рисунок 33. Зависимость показателей логдоходности компании «Сбербанк» от времени.

С помощью представленных графиком можно сделать вывод, что динамика логдоходностей компании «ДЭК ао» за рассматриваемый период является относительно «спокойным», а акции «Сбербанк» отличается более высокой волатильностью показателей логдоходности за весь период.

# 11. Исследование тесноты связи между логдоходностями

С помощью сервиса «Корреляция» в разделе «Анализ данных» сгенерируем корреляционную матрицу для логарифмических доходностей трех компаний.

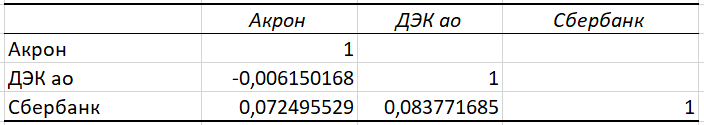


Рисунок 34. Корреляционная матрица

Для каждого значения статистического коэффициента корреляции вычисляем значение:

,

далее находим критические значения распределения Стьюдента с помощью статистической функции СТЬЮДЕНТ.ОБР.2Х. Результаты вычислений приведены в следующей таблице:

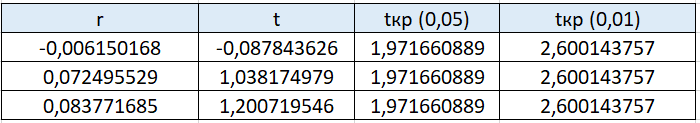


Рисунок 35. Расчет фактического и критических значений критерия T.

Из представленной таблицы следует, что гипотезу о наличии корреляции между логдоходностями компаний нет оснований отклонить на уровне значимости α = 0,05 и на уровне значимости α = 0,01.