**Федеральное государственное образовательное бюджетное учреждение**

**высшего образования**

**«ФинансоВЫЙ УНИВЕРСИТЕТ**

**при Правительстве Российской Федерации»**

**(Финансовый университет)**

**Департамент анализа данных,  
принятия решений и финансовых технологий**

**О. А. Баюк**

**УЧЕБНОЕ ПОСОБИЕ**

**ПО ВЫПОЛНЕНИЮ РАСЧЕТНО-АНАЛИТИЧЕСКОЙ РАБОТЫ ПО ДИСЦИПЛИНЕ "АНАЛИЗ ДАННЫХ"**

Учебное пособие для студентов

**Москва 2020**

УДК 519.22(073)

ББК 22.16я73

М 47

Рецензент**:**

**С.А. Зададаев**, кандидат физико-математических наук, доцент Департамента «Анализа данных, принятия решения и финансовых технологий»

**О.А. Баюк**

Учебное пособие по выполнению расчетно-аналитической работы по дисциплине "Анализ данных" (Учебное пособие). Для бакалавров направления 38.03.05 «Бизнес-информатика». – М.: ФГОБУ ВО «Финансовый университет при Правительстве Российской Федерации», департамент «Анализа данных, принятия решений и финансовых технологий», 2020 - 63 с.

Данное учебное пособие предназначено для студентов очного отделения, изучающих дисциплину «Анализ данных». Пособие написано в соответствии с программой дисциплины «Анализ данных». Оно предназначено для подготовки бакалавров по направлениям экономика Финуниверситета. В пособии отражены темы: выборочный метод, точечные и интервальные оценки, проверка статистических гипотез, корреляционный анализ, анализ временных рядов. Материал иллюстрируется большим количеством примеров. Пособие может быть использовано как для проведения семинарских занятий, так и для организации самостоятельной работы студентов.

***Учебное издание***

**Баюк Олег Александрович**

УЧЕБНОЕ ПОСОБИЕ ПО ВЫПОЛНЕНИЮ РАСЧЕТНО-АНАЛИТИЧЕСКОЙ РАБОТЫ ПО ДИСЦИПЛИНЕ "АНАЛИЗ ДАННЫХ"

Учебное пособие

Компьютерный набор, верстка: Баюк О.А.

Формат 60х90/16. Гарнитура *TimesNewRoman*

© О. А. Баюк, 2020

© ФГОБУ ВО «Финансовый университет при Правительстве Российской Федерации», 2019

**Оглавление**

[**Введение** 4](#_Toc37032129)

[**Задание по РАР** 4](#_Toc37032130)

[**1. Загрузка исходных данных** 7](#_Toc37032131)

[**2. Вычисление дополнительных признаков для каждой компании** 14](#_Toc37032132)

[**3. Исследование изменения цен и объемов торгов акций** 16](#_Toc37032133)

[**4. Исследование логарифмических доходностей (логдоходностей) акций** 20](#_Toc37032134)

[**4.1. Исследование до удаления выбросов** 20](#_Toc37032135)

[**4.2.Удалениевыбросов** 20](#_Toc37032136)

[**5. Детальный анализ логдоходностей для каждой компании** 28](#_Toc37032137)

[**(с данными без выбросов)** 28](#_Toc37032138)

[**5.1. Точечные оценки параметров нормального распределения** 28](#_Toc37032139)

[**5.3. Проверка гипотезы о нормальности логдоходности** 33](#_Toc37032140)

[**5.3.1. Проверка гипотезы по критерию Пирсона (хи-квадрат)** 33](#_Toc37032141)

[**5.3.2. Проверка гипотезы по критерию Колмогорова-Смирнова** 37](#_Toc37032142)

[**5.3.3. Сравнение результатов проверки по двум критериям** 40](#_Toc37032143)

[**5.4. Построение диаграммы, содержащей гистограмму эмпирической плотности и график теоретической плотности *(это желательный, но не обязательный пункт).*** 40](#_Toc37032144)

[**5.5. Построение графиков эмпирической функции распределения и график теоретической функции распределения.** 41](#_Toc37032145)

[**6. Проверка гипотезы о равенстве нулю математического ожидания недельной логдоходности при альтернативной гипотезе о том, что оно больше нуля.** 41](#_Toc37032146)

[**7. Проверка гипотезы о равенстве средних значений логарифмической доходности за последний и предпоследний годы (двусторонний t-тест без каких-либо предположений об однородности дисперсий)** 47](#_Toc37032147)

[**8. Проверка гипотезы о равенстве дисперсий логарифмической доходности за последний и предпоследний годы (двусторонний F-тест)** 52](#_Toc37032148)

[**9. Сравнительный анализ логдоходностей трех компаний.** 55](#_Toc37032149)

[**10. Графики временных рядов логдоходностей трех компаний** 65](#_Toc37032150)

[**11. Исследование тесноты связи между логдоходностями** 68](#_Toc37032151)

[**Заключение** 70](#_Toc37032152)

[**Список литературы** 70](#_Toc37032153)

# **Введение**

В предлагаемом пособии сформулировано задание по расчетно-аналитической работы РАР по курсу «Анализ данных» и приведены рекомендации по выполнению этого задания с использованием средства программирования **MSExcel**.

# **Задание по РАР**

1. Скачать дневные цены закрытия акций в течение 4 лет (2015 г.-2019 г.).

Варианты по 3 компании для каждого студента формировать с указанием следующих параметров (тиккеры для каждого варианта должен сообщить преподаватель):

тиккер компании, начальная дата (208 дат с недельным шагом с января 2015 г.)

2. Вычислить следующие признаки для каждой компании:

Логарифм цены

Логдоходность

Логарифм объёма (только для диаграмм рассеяния)

3. Выполнить следующие исследования:

предварительная обработка данных

описательная статистика

выбросы и диаграммы «Ящик с усами»

диаграммы рассеяния (логарифм цены vs логарифм объёма) с вычислением корреляций

гистограммы интервальных частот

4. Удалить строки с выбросами в логдоходностях всех трех компаний для возможного сравнения их между собой

5. Выполнить повторение предварительного исследования с данными без выбросов

описательная статистика

выбросы и диаграммы «Ящик с усами»

диаграммы рассеяния (логарифм цены vs логарифм объёма) с вычислением корреляций

гистограммы интервальных частот

эмпирическая функция распределения

6. Выполнить детальный анализ для каждой компании логдоходностей с данными без выбросов

6.1. точечные оценки параметров нормального распределения

6.2. интервальные оценки параметров нормального распределения

6.3. построить диаграммы, содержащей гистограмму эмпирической плотности и график теоретической плотности *(это желательный, но не обязательный пункт).*

6.4. Построить графики эмпирической функции распределения и график теоретической функции распределения. Проверить гипотезу о нормальности логдоходности:

по критерию хи-квадрат

По критерию Колмогорова-Смирнова

6.5. Сравнить результаты проверки по двум критериям

6.6. Для каждой акции проверьте на 5%-ном и 1%-ном уровне значимости гипотезу о том, что математическое ожидание недельной доходности равно нулю при альтернативной гипотезе о том, что оно больше нуля. В каждом случае вычислите также наблюдаемый уровень значимости (p-value).

6.7. Проверка гипотезы о равенстве средних значений логарифмической доходности за последний и предпоследний годы (двусторонний t-тест без каких-либо предположений об однородности дисперсий)

6.8. Проверка гипотезы о равенстве дисперсий логарифмической доходности за последний и предпоследний годы (двусторонний F-тест)

7. Сравнительный анализ логдоходностей трех компаний:

7.1. Графики временных рядов логдоходностей трех компаний на одном рисунке

7.2. Матрица корреляций логдоходностей

Условное форматирование матрицы корреляций для значимых и не значимых коэффициентов корреляций

8. Собрать всё исследование в один Excel-файл, содержащий все пункты задания с подробными комментариями (возможно с включениями скринов работы программ в RStudio).

9. Оформить соответствующий выполненной работе отчёт в MSWord с формулами и с подробным описанием всех процедур, а также с таблицами и рисунками, на которых должны присутствовать все необходимые обозначения. Таблицы и рисунки должны быть пронумерованы и подробно описаны.

# **1. Загрузка исходных данных**

Данные для исследования считываются из базы данных Московской биржи. Для этого следует перейти по адресу

<https://mfd.ru/export/>

При обращении появится окно (см. Рисунок 1).

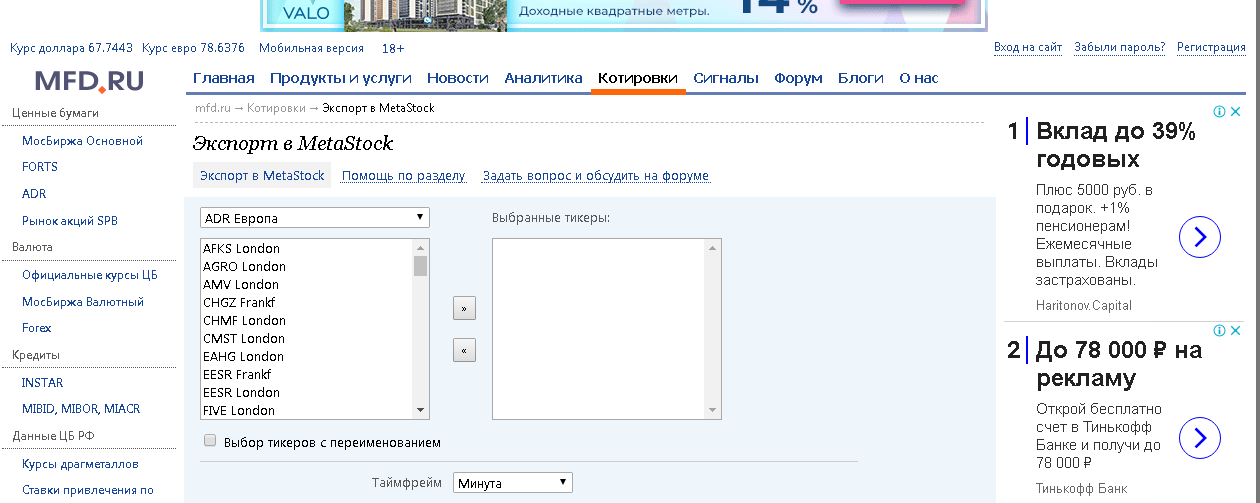


Рисунок 1. Окно сервисной программы московской биржи.

В предлагаемом меню следует выбрать раздел «Мосбиржа Акции и ПИФы» и последовательно выбрать тиккеры нужных компаний.

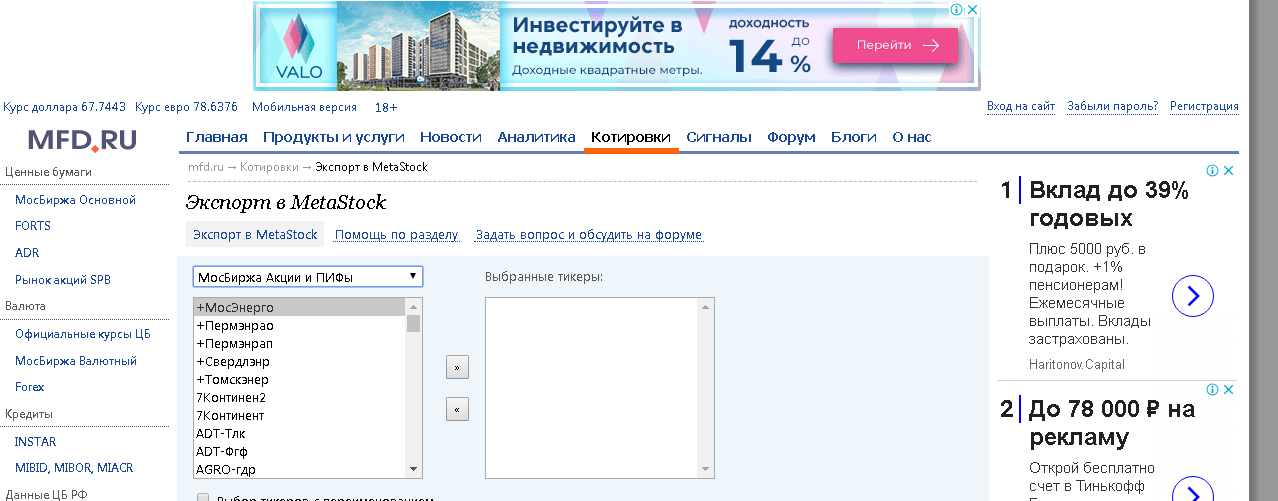


Рисунок 2. Выбор акций нужных компаний.

Выбор акций нужных компаний осуществляется при помощи пары окон (см. Рисунок 1).

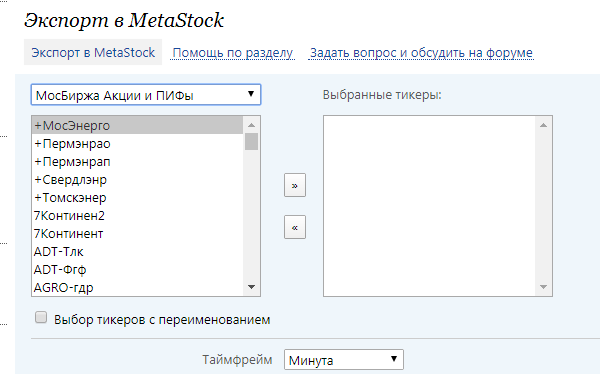


Рисунок 3. Выбор акций нужных компаний (продолжение).

Далее следует задать формат записей формируемого файла данных, выбирая необходимые параметры, как показано на следующем рисунке

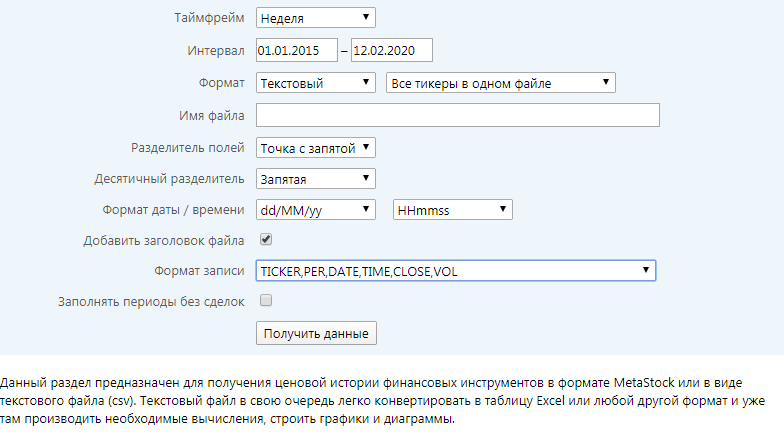
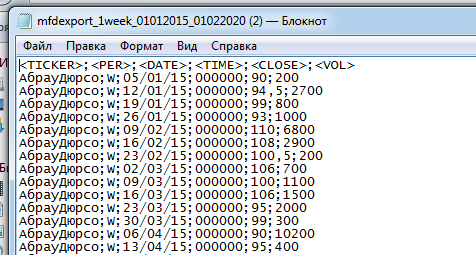


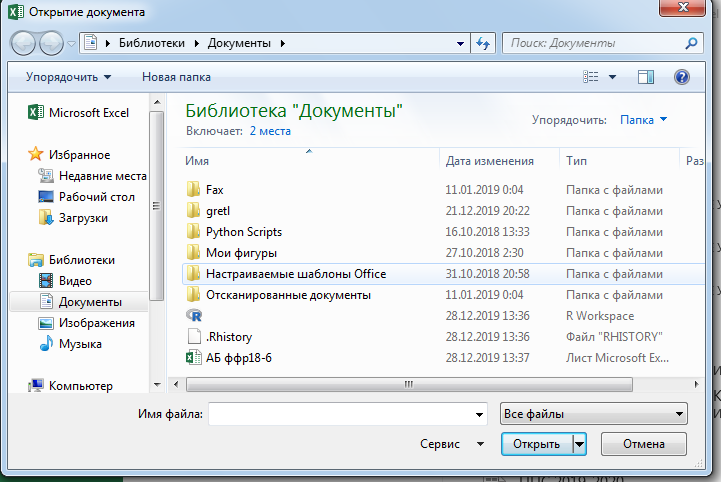
Рисунок 4. Формирование формата записей.

После нажатия «кнопки» «Получить данные» будет создан текстовый файл, фрагмент которого показан ниже



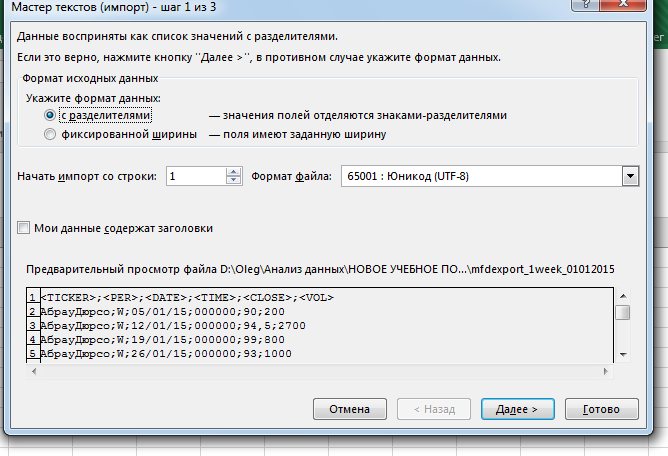
**Рисунок 5.** Фрагмент сформированного файла.

Для дальнейшей обработки этот файл необходимо загрузить на лист ***Excel***. При чтении текстового файла в ***Excel***следует указать тип файла «Все файлы» как показано на **Рисунке 6.**

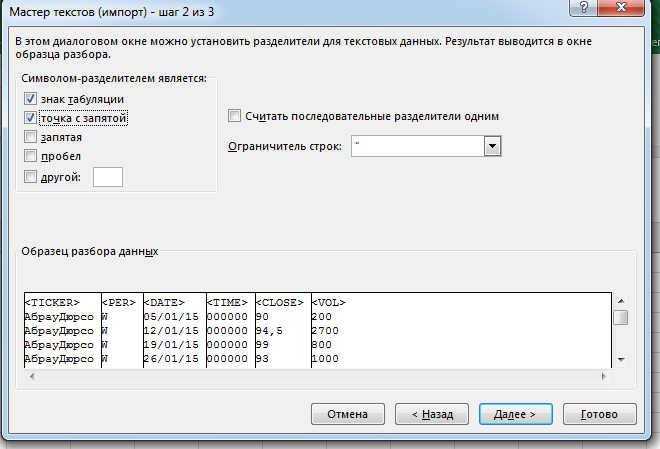


**Рисунок 6.** Чтение текстового файла в ***Excel***следует указать тип файла «Все файлы»

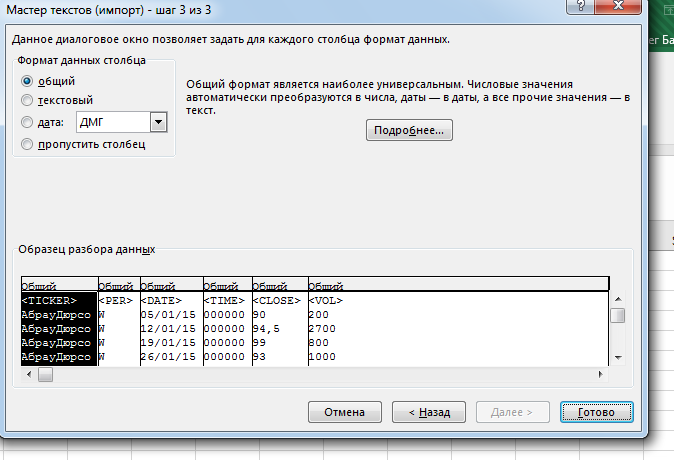
Далее следует указывать опции, как показано на **Рисунках 7-9.**



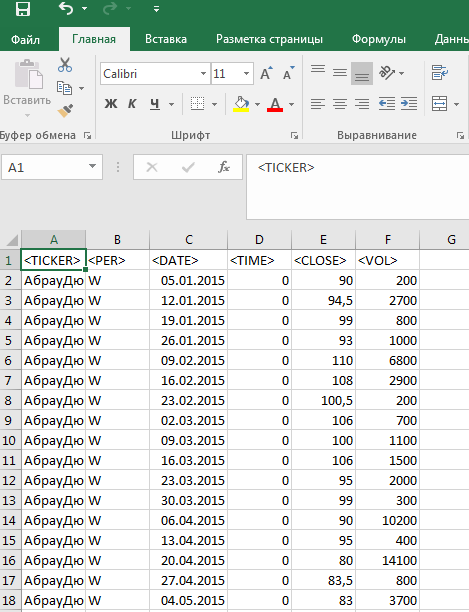
**Рисунок 7. Этапы форматирования файла.**



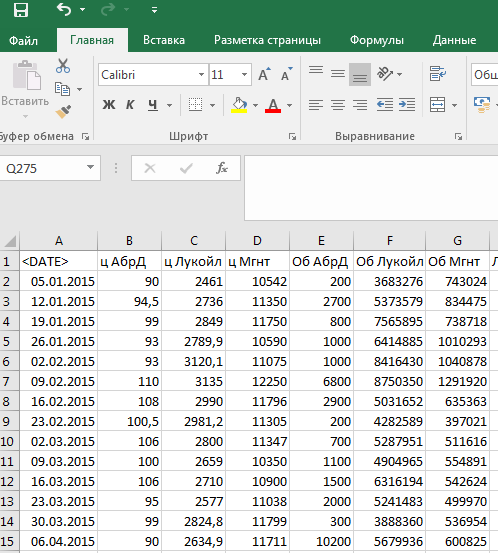
**Рисунок 8. Этапы форматирования файла (продолжение).**



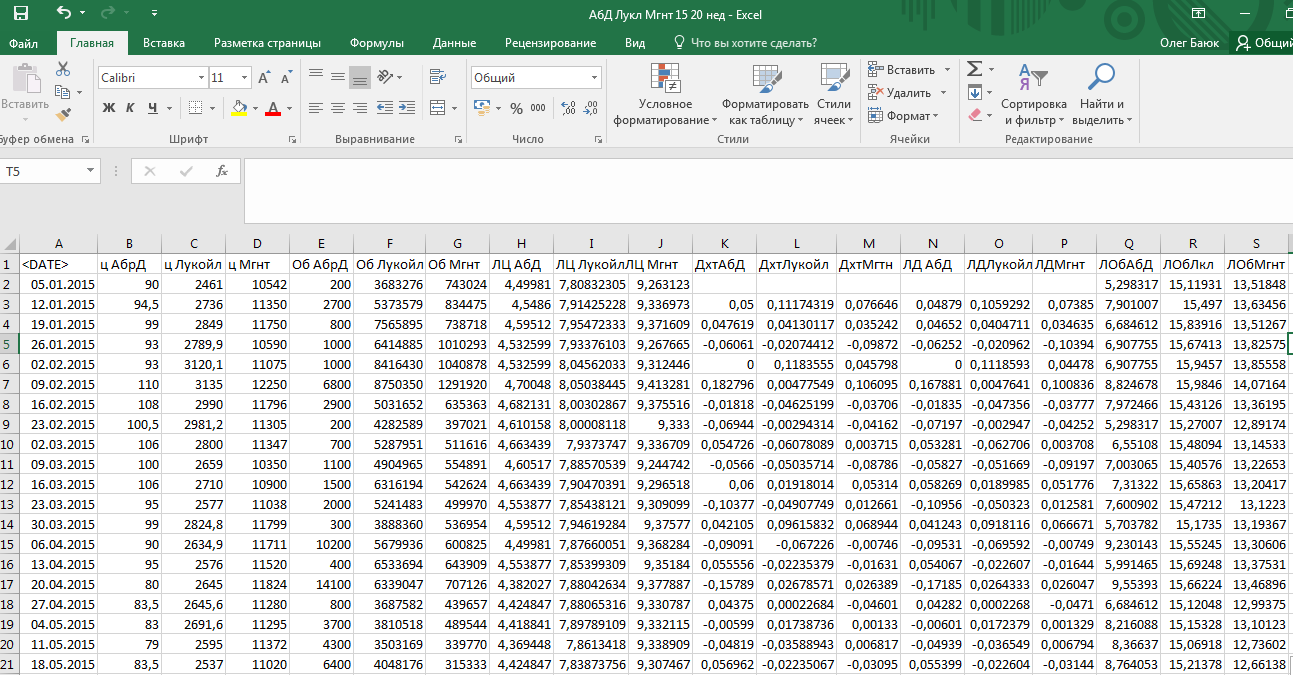
**Рисунок 9. Этапы форматирования файла (завершение).**



**Рисунок 10. Образец исходного листа.**



**Рисунок 11. Размещение новых признаков.**



**Рисунок 12. Образец завершенного листа исходных данных.**

# **2. Вычисление дополнительных признаков для каждой компании**

На первом листе следует:

- удалить столбцы периода и времени;

- сформировать столбцы цен закрытия и объемов торгов, так чтобы для каждого тиккера и каждого параметра был персональный столбец;

- сформировать столбцы новых параметров: логарифмов цен, доходностей, логарифмических доходностей (логдоходностей) и логарифмов объемов торгов.

После этого следует удалить столбец **NAME**и неиспользуемые значения дат.

Доходностьи логдоходность вычисляются следующим образом.

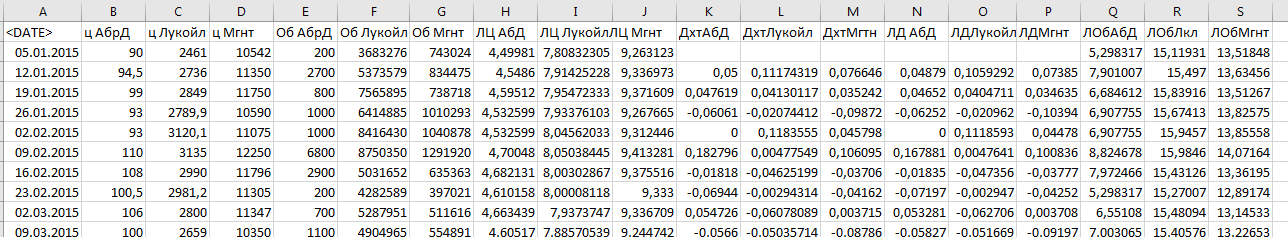
Пусть и – цены актива соответственно для ***i–***й и ***i-1*** –й, то есть предшествующей недель, тогда доходность () и логдоходность () для ***i–***й недели вычисляются по формулам

Анализируя последние формулы можно заметить, что

оэтому, как ожидается коэффициент корреляции этих величин незначительно отличается от 1.

Поскольку значения параметра «Объем торгов» для различных дат и тиккеров могут отличаться на несколько порядков, целесообразно вычислить новый признак «Логарифм объёма». Этот новый признак пропорционален исходному, однако принимает значения одного порядка.

После вычисления перечисленных новых признаков лист исходных данных будет иметь вид (см. **Рисунок 13.**).



**Рисунок 13. Лист исходных данных.**

В рассматриваемом примере в столбце **A1:A266**размещены значения дат, в диапазоне**B1:D266** – значения цен акций компаний «Абрау-Дюрсо», «Лукойл», «Магнит»; в диапазоне**E1:G266** – соответствующие значения объемов торгов акций указанных компаний, в диапазоне **H1:J266** - соответствующие значения логарифмов цен, в диапазоне **K1:M266** - соответствующие значения доходностей, в диапазоне**N1:P266**соответствующие значения логдоходностей, в диапазоне**Q1:S266**- значения логарифмов объемов торгов.

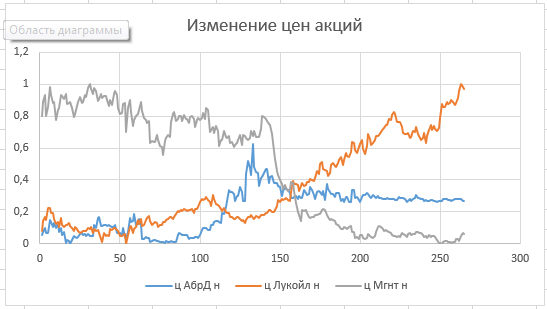
# **3. Исследование изменения цен и объемов торгов акций**

Для получения первой начального представления о поведении исследуемых величин, как правило используются средства визуализации, например, построение графиков. Поскольку цены акций различных компаний могут отличаться на несколько порядков, изображение их на одном графике невозможно. Для решения отмеченной проблемыиспользуется масштабирование. Например, вместо цен (*x*) можно использовать относительные цены (*t*), определяемые формулой

где

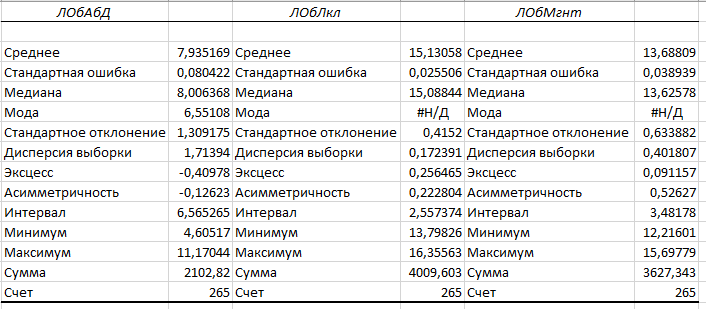
- соответственно минимальное и максимальное значения цены.

На **рисунке 14** представлены графики изменения цен акций.

****

**Рисунок 14. Графики изменения цен акций.**

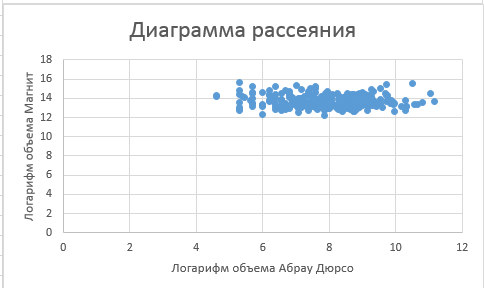
Как было указано выше, для исследования изменения значении объемов торгов удобней исследовать соответствующие изменения логарифмов объемов торгов.



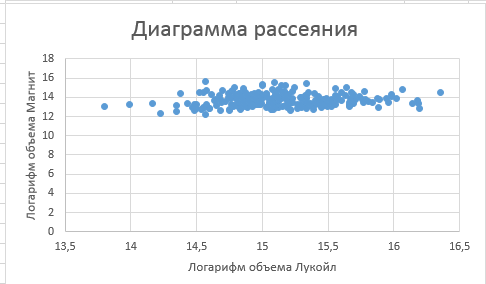
**Рисунок 15. Описательная статистика признаков «логарифмов объемов торгов» акций компаний «Абрау-Дюрсо», «Лукойл», «Магнит».**



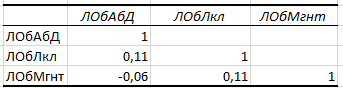
**Рисунок 16. Диаграмма рассеяния значений признаков «логарифмов объемов торгов» акций компаний «Абрау-Дюрсо» и «Лукойл».**



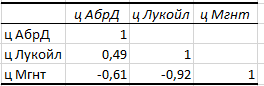
**Рисунок 17. Диаграмма рассеяния значений признаков «логарифмов объемов торгов» акций компаний «Абрау-Дюрсо» и «Магнит».**



**Рисунок 18. Диаграмма рассеяния значений признаков «логарифмов объемов торгов» акций компаний «Лукойл» и «Магнит».**



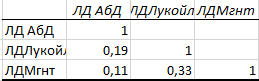
**Рисунок 19. Корреляционная матрица признаков «логарифмов объемов торгов» акций компаний «Абрау-Дюрсо», «Лукойл», «Магнит».**



**Рисунок 20. Корреляционная матрица признаков «цена» акций компаний «Абрау-Дюрсо», «Лукойл», «Магнит».**

# **4. Исследование логарифмических доходностей (логдоходностей) акций**

## **4.1. Исследование до удаления выбросов**



**Рисунок 21. Корреляционная матрица признаков «логдоходность» акций компаний «Абрау-Дюрсо», «Лукойл», «Магнит».**

## **4.2.Удалениевыбросов**

Рассмотрим некоторые числовые характеристики распределения значений числового признака *X,* определенного для всех элементов выборки объема *n*[1, 2].

Пусть наименьшее и наибольшее значения этого признака обозначены соответственно Запись значений признака *X* для всех элементов выборки, упорядоченная по возрастанию (*неубыванию*) называется вариационным рядом. Медианой называется такое число, которое делит вариационный ряд на две равные по объему (числу элементов части). Если *n*– нечетное число, то

,

если *n*–четное число, то

*.*

Аналогично определяются квартили, числа делящие вариационный ряд на четыре равные по численности (объему) части [1, 2]:

- 25% всех элементов вариационного ряда не больше первой (или нижней) квартили , а оставшиеся 75% значений – не меньше ;

- 50% всех элементов вариационного ряда не больше первой (или средней) квартили , а оставшиеся 50% значений – не меньше (очевидно ;

- 75% всех элементов вариационного ряда не больше первой (или верхней) квартили , а оставшиеся 25% значений – не меньше .

Разность

называется *межквартильным* размахом и характеризует разброс значений признака *X*.

Выбросами называются элементы вариационного ряда, не принадлежащие отрезку

**Пример.** Найдем для вариационного ряда

1, 2, 3, 4, 5, 6, 7, 7, 7, 8, 9, 10, 11, 11, 15.

Поскольку вариационный ряд содержит нечетное число элементов

(*n* =15), то медиана, которая также является средней квартилью равна восьмому элементу вариационного ряда

Теперь рассмотрим часть вариационного ряда, содержащую элементы, расположенные не правее медианы:

1, 2, 3, 4, 5, 6, 7, 7,

и найдем их медиану

.

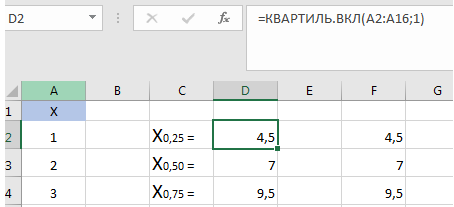
Аналогично найдем . Рассмотрим часть вариационного ряда, содержащую элементы, расположенные не левее медианы:

7, 7, 8, 9, 10, 11, 11, 15,

и найдем их медиану

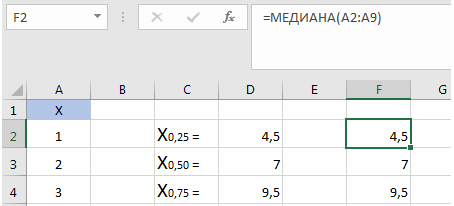
Для определения квартилей в **MSExcel** имеется статистическая функция КВАРТИЛЬ.ВКЛ(<массив>, <часть>).

На **Рисунке 22** представлен результат определения квартилей для данного примера



**Рисунок 22. Вычисление квартилей.**

На **Рисунке 22** представлена возможность вычислить как левой части вариационного ряда.



**Рисунок 23. Вычисление квартилей с помощью статистической функции МЕДИАНА.**

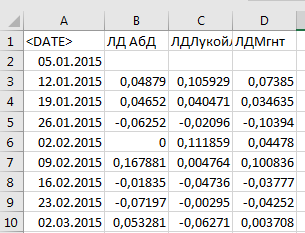
Теперь найдем и проверим имеются ли выбросы в данном вариационном ряде:

значит

Поскольку все элемент выборки менее , то выбросы ниже нормы в данном вариационном ряде отсутствуют, выброс выше нормы – один:

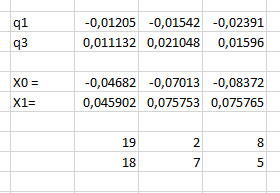
15 >12.

Для исследования выбросов в выборках логдоходностей скопируем столбцы соответствующих значений на новый лист с опцией «значения» в диапазон **B1:D266**, в диапазоне **A1:A266** ранее должны быть размещены соответствующие даты



**Рисунок 24. Лист логдоходностей.**

Вычислим квартили (*q1*) и (*q3*):



**Рисунок 25. Значения квартилей и границ нормы.**

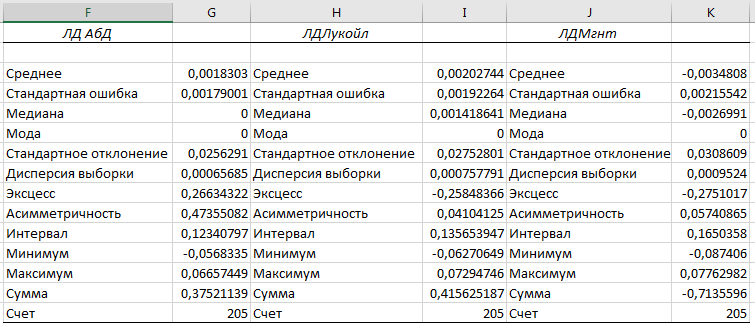
Далее вычислим значения

.

Затем с использованием инструмента «Сортировка и фильтр», а также инструмента «Удалить» исключаем строки содержащие выбросы для каждого тиккера.

**Описание алгоритма исключения выбросов**

На новый лист «Удаление выбросов» в диапазоне **А1:А266** размещаются даты наблюдений, в диапазоне **B1:D266**размещаютсясоответствующие значения логарифмических доходностей (акций компаний «Абрау-Дюрсо», «Лукойл», «Магнит»). Далее, находим значения числовых характеристик выборки с помощью надстройки «Описательная статистика»



**Рисунок 26. Описательная статистика логдоходностей до удаления выбросов.**

Вычисляем первые (*q1*) и третьи (*q3*) квартили с использованием статистической функции КВАРТИЛЬ.ВКЛ



**Рисунок 27. Значения первых и третьих квартилей.**

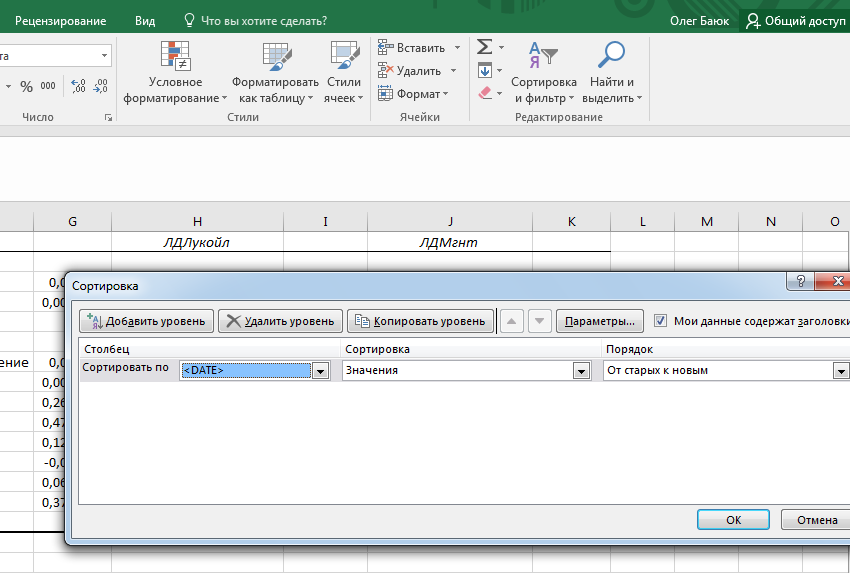
Затем вычисляются *межквартильный* размах

и границы нормы

.

**Рисунок 28. Значения нижних и верхних границ нормы**

После этого диапазон **А1:D266**упорядочивается по возрастанию значений диапазона**B2:B266.** Для этого следует выделить весь диапазон **А1:D266**и воспользоваться сервисом«Сортировка и фильтр-Настраиваемая сортировка» при этом следует указать опцию «Мои данные содержат заголовки» и указать в окне «Сортировать по» заголовок «ЛД Абд». Затем удалить все строки диапазона **А1:D266,** в которых значение столбца **B**меньше или больше. После этого описанную процедуру повторить для столбцов **С** и**D**.



**Рисунок 29. Удаление выбросов.**

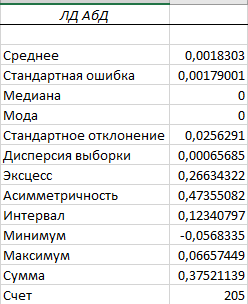
После реализации описанной процедуры, все выбросы будут исключены и остается 205 наблюдений (в рассматриваемом примере).Дальнейшие исследования проводятся с данными, не содержащими выбросов.

# **5. Детальный анализ логдоходностей для каждой компании**

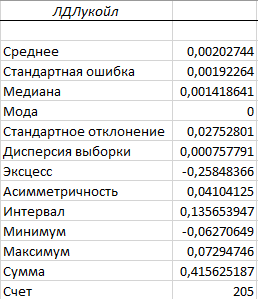
# **(с данными без выбросов)**

## **5.1. Точечные оценки параметров нормального распределения**

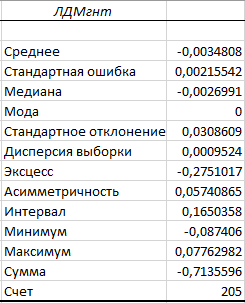
Точечные оценки числовых характеристик параметров целесообразно находить, используя сервис «Описательная статистика»[3]. Результаты работы указанной процедуры представлены на **рисунках 30-32**.



**Рисунок 30.«Описательная статистика» для логдоходностей акций компании «Абрау-Дюрсо».**



**Рисунок 31. «Описательная статистика» для логдоходностей акций компании «Лукойл».**



**Рисунок 32.«Описательная статистика» для логдоходностей акций компании «Магнит».**

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

***Доверительный интервал для оценки математического ожидания нормально распределенной случайной величины при неизвестной дисперсии.***

Пусть имеется выборка из генеральной совокупности распределенной по нормальному закону, тогда имеет место интервальная оценка математического ожидания

,

где

выборочное среднее,

–исправленное стандартное отклонение,

,

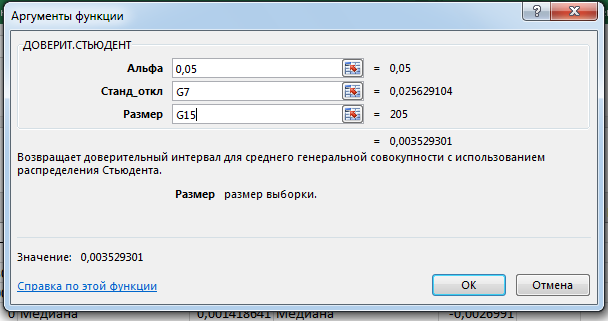
процентная точка распределения Стьюдента с

степенями свободы.

Для вычисления точности интервальной оценки

в **MSExcel**применяется статистическая функция

ДОВЕРИТ.СТЬЮДЕНТ(<>,<>, <>),

при этом вероятность задается пользователем, исправленное стандартное отклонение*s*иобъем выборки *n*вычисляются по выборке, например с помощью сервиса «Описательная статистика»[3]. В рассматриваемом примере ; Значения *s*и вычисляются надстройкой «Описательная статистика» для каждого диапазона. Пример обращения к этой функции представлен на **рисунке33**. В ячейке G7 находится значение исправленного стандартного отклонения, в ячейкеG15 – значение объема выборки

**Рисунок 33.**Пример обращения к этой функции ДОВЕРИТ.СТЬЮДЕНТ

***Доверительный интервал для дисперсии*** определяется следующим образом

,

то есть

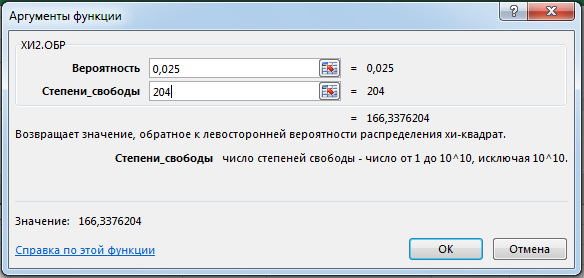
,

здесь

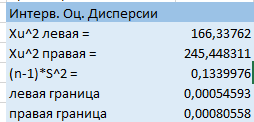
- исправленная выборочная дисперсия (вычисляются по выборке с помощью сервиса «Описательная статистика»);

процентнаяипроцентная точки распределения хи-квадрат с степенями свободы ( – уровень значимости).

Значение вычисляется в **MSExcel**статистической функцией ХИ2.ОБР. При обращении к это функции следует указать значения и (см. **Рисунок 34**. )



**Рисунок 34.**Пример обращения к функции ХИ2.ОБР (, )



**Рисунок 35. Вычисление границ доверительного интервала для дисперсии.**

***Доверительные интервалы для стандартного отклонения***определяется следующим образом

то есть



# **5.3. Проверка гипотезы о нормальности логдоходности**

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

Критерий согласия Пирсона (χ2) применяют для проверки гипотезы о соответствии эмпирического [распределения](http://www.termist.com/gloss/6221/p_n/p_n_1.htm) предполагаемому теоретическому распределению *F(x)* при большом объеме выборки (*n ≥ 100*). Критерий применим для любых видов функции *F(x)*, даже при неизвестных значениях их параметров, что обычно имеет место при анализе результатов механических испытаний. В этом заключается его универсальность.

Использование критерия χ2 предусматривает разбиение [размаха варьирования](http://www.termist.com/laborat/stat/war_rjad.htm#in_xar)[выборки](http://www.termist.com/bibliot/stud/ma_en_sl/23/126_20.htm) на интервалы и определения числа наблюдений (частоты) *nj* для каждого из *l* интервалов. Для удобства [оценок](http://www.termist.com/bibliot/stud/ma_en_sl/63/559_9.htm) параметров распределения интервалы выбирают одинаковой длины (*h*).

Число интервалов зависит от объема выборки.

При выборе количества интервалов часто пользуются правилом Стерджеса*l≃1+*ln*n*, где *n*– объём выборки.

Обычно принимают: при *n = 100, l = 10 ÷ 15*; при *n = 200,l*= *15 ÷ 20;*

при *n = 400, l = 25 ÷ 30*; при *n = 1000, l = 35 ÷ 40*.

Интервалы, содержащие менее пяти наблюдений, объединяют с соседними. Однако, если число таких интервалов составляет менее 20% от их общего количества, допускаются интервалы с частотой *nj* ≥ 2.

Статистикой критерия Пирсона служит величина

где *pj*- вероятность попадания изучаемой случайной величины в *j*-й

интервал, вычисляемая в соответствии с гипотетическим законом распределением *F(x)*. При вычислении вероятности *pj*нужно иметь в виду, что левая граница первого интервала и правая последнего должны совпадать с границами области возможных значений случайной величины. Например, при нормальном распределении первый интервал простирается от -∞, а последний  ̶ до +∞.

***Теорема К. Пирсона.****Предположим, что гипотеза H0 верна. Тогда при распределение величины сходится к распределению хи-квадрат с l-1степенью свободы, то есть,*

***.***

Практический смысл этой теоремы в том, что при *большом объеме* выборки распределение  можно считать распределением хи-квадрат с *l-1* степенью свободы.

Нулевую гипотезу о соответствии выборочного распределения теоретическому закону *F(x)* проверяют путем сравнения вычисленной по формуле (3.91) величины с критическим значением , найденным по таблице значений распределения χ2 для уровня значимости α и числа степеней свободы *k*= *l*1 *- m - 1*. Здесь *l*1 ̶ число интервалов после объединения; *m*  ̶ число параметров, оцениваемых по рассматриваемой выборке. Если выполняется неравенство , то нулевую гипотезу не отвергают. При несоблюдении указанного неравенства принимают альтернативную гипотезу о принадлежности выборки неизвестному распределению.

Зная уровень значимости и число степеней свободы , можно найти критическое значение статистики и сравнить его с эмпирическим значением критерия , определенным по выборке. Если , то гипотеза о независимости случайных величин принимается, иначе основная гипотеза отвергается.

**Алгоритм решения задачи вMSExcel**

1. Скопируем диапазон значений логдоходности (без выбросов) на новый лист;

2. Вычисляем основные выборочные числовые характеристики с помощью сервиса «Описательная статистика;

3. Вычислим длину интервала *h*;

4. Вычисляем границы интервалов, на которые делится диапазон значений логдоходности для данного актива, при этом в качестве левой границы первого интервала выбирается очень маленькое число (например -100000), в качестве правой границы последнего интервала - очень большое число (например 100000);

5. Вычисляем эмпирические интервальные частоты *nj*с помощью статистической функции ЧАСТОТА;

6. Вычисляем теоретические интервальные частоты , при этом вероятности того, что случайная величина примет значение на соответствующем интервале, вычисляются с помощью статистической функции НОРМ.РАСП;

7. Вычисляем значения

,

после чего находим сумму всех этих значений, то есть вычисленное значение статистики критерия :

8. По заданному значению уровня значимости с помощью статистической функции ХИ2.ОБР.ПХ находим критическое значение критерия ;

9. Если , то гипотеза о нормальном распределении логдоходности принимается, если данное неравенство не выполняется, то гипотеза отклоняется.

Указанные расчеты проводятся для всех трех активов. При этом целесообразно использовать отдельный лист для каждого актива.

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

Пусть имеется выборка из некоторой генеральной совокупности, ставится задача проверки гипотезы о том, что генеральная совокупность подчинена заданной непрерывной функции распределения

Пусть – эмпирическая функции распределения, соответствующая данной выборке, определенная следующим образом

Критерий Колмогорова применяется для проверки гипотезы

Нулевую гипотезу можно записать в виде [4]:

(2.2.1)

при этом альтернативная гипотеза определяется так:

Согласно теореме В. И. Гливенко [5] для любой непрерывной функции при эмпирическая функция равномерно сходится к ней

(2.2.2)

Таким образом, эмпирическая функция распределения является состоятельной оценкой теоретической функции распределения

А. Н. Колмогоровым доказана теорема [5]:

Вероятность неравенства

стремится при равномерно по λ к

для любой непрерывной функции распределения

Таким образом, проверка гипотезы (2.2.1) - (2.2.2) производится по следующему алгоритму.

1. Выбирается уровень значимости .

2. Вычисляется значение статистики

3. Определяется квантиль функции распределения

4. При нулевая гипотеза не отклоняется, при

нулевая гипотеза отклоняется в пользу альтернативной гипотезы.

Значения для некоторых значений приведены в таблице 2.2.1.

Если теоретическая функция распределения известна с точностью до параметров, то в качестве значений параметров используются их оценки, вычисленные по элементам выборки. В этом случае для проверки гипотезы о соответствии выборки теоретическому закону распределения используется статистика типа Колмогорова [5]:

Рассмотрим случай нормального распределения с математическим ожиданием . Если истинные значения указанных параметров неизвестны, то по выборке вычисляются их оценки

,

Введем обозначения, ,

Тогда

Квантили распределений статистик и для указанного случая приведены в ***таблице 1.***

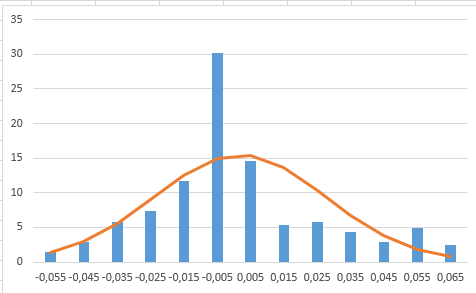
***Таблица 1.***

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Уровень значимости | 0,15 | 0,10 | 0,05 | 0,025 | 0,01 |
| Порядок *p*квантиля | 0,85 | 0,90 | 0,95 | 0,975 | 0,99 |
| Квантили порядка *p* для | 1,138 | 1,224 | 1,358 | 1,480 | 1,626 |
| Квантили порядка *p* для | 0,775 | 0,819 | 0,895 | 0,995 | 1,035 |

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

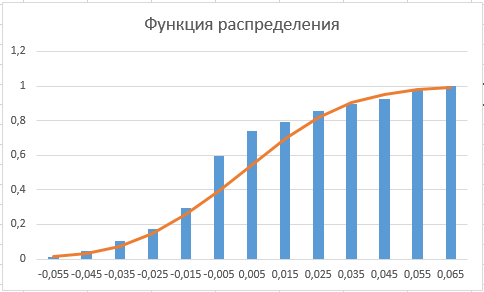
Применение обоих критериев позволяет отклонить гипотезу о принадлежности выборки для компании «Абрау-Дюрсо» к нормальному распределению. Аналогичные расчеты были выполнены для остальных двух компаний. По их результатам можно сделать вывод о принятии гипотезы о нормальном распределении.

# **5.4. Построение диаграммы, содержащей гистограмму эмпирической плотности и график теоретической плотности *(это желательный, но не обязательный пункт).***



**Рисунок 36.График теоретической плотности и гистограмма для логдоходности акций компании «Абрау-Дюрсо».**

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



**Рисунок 37.Графиков эмпирической функции распределения и график теоретической функции распределения для логдоходности акций компании «Абрау-Дюрсо».**

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

**Задача.** Для каждой акции проверьте на 5%-ном и 1%-ном уровне значимости гипотезу о том, что математическое ожидание недельной доходности равно нулю при альтернативной гипотезе о том, что оно больше нуля. В каждом случае вычислите также наблюдаемый уровень значимости.

Для решения поставленной задачи следует использовать статистику

,

то есть

.

Поскольку конкурирующая гипотеза

*Н*1: ,

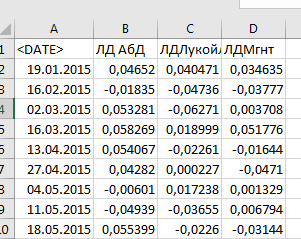
то для принятия решения находится критическая точка распределения Стьюдента, соответствующая уровню значимости и числу степеней свободы . Нулевая гипотеза принимается, если

и отклоняется при

.

Рассмотрим решение данной задачи для рассматриваемого примера с использованием**MSExcel.**

Создадим новый лист «Проверка гипотез». Скопируем на этот лист данные о логдоходностяхакций трех компаний, оставшиеся после удаления выбросов. Таким образом, на листе останутся 206 строк (для данного примера):



**Рисунок 38. Исходные данные для проверки гипотез.**

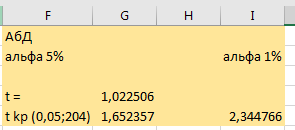
Для вычисления статистики

используем формулу



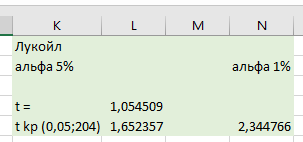
Для вычисления критических значений распределения Стьюдента используем статистическую функцию СТЬЮДЕНТ.ОБР

Ниже приведены результаты вычислений

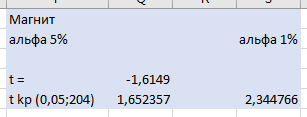


**Рисунок 39. Данные для принятия решения о принятии гипотезы (акция компании «Абрау-Дюрсо»).**

Поскольку вычисленное значение критерия оказалось менее критического значения, как при , так и при (см. **Рисунок** ), то в обоих случаях нет оснований отвергнуть гипотезу о равенстве нулю математического ожидания недельной логдоходности данного актива (акция компании «Абрау-Дюрсо»). Аналогичная ситуация – с компаниями «Лукойл» и «Магнит» (см. **Рисунки40 и 41**).



**Рисунок 40. Данные для принятия решения о принятии гипотезы (акция компании «Лукойл»).**



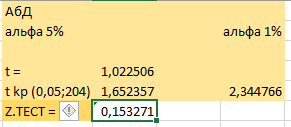
**Рисунок 41. Данные для принятия решения о принятии гипотезы (акция компании «Магнит»).**

Данную задачу в **MSExcel** можно решить другим способом, используя статистическую функцию Z.ТЕСТ[3]. Эта функция вычисляет вероятность того, что значение статистики окажется более значения, вычисленного по данной выборке (здесь- предполагаемое значение математического ожидания). Если вычисленное с помощью функции Z.ТЕСТ значение окажется более заданного уровня значимости, то основная гипотеза не отклоняется, в противном случае основную гипотезу следует отклонить.

Рассмотрим использование этой функции для решения предложенных задач. Пусть требуется проверить гипотезу о том, что математическое ожидание недельной доходности компании «Абрау-Дюрсо» равно нулю при альтернативной гипотезе о том, что оно больше нуляпри уровнях значимости , так и при . Обращаемся к функции Z.ТЕСТ:

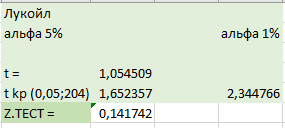
Z.ТЕСТ(**B2:B206**;0).

В диапазоне **B1:B206**размещаются значения логарифмических доходностей акций компании «Абрау-Дюрсо», 2-й аргумент – предполагаемое значение математического ожидания, в данном случае равное нулю, у этой функции возможен третий аргумент – среднее квадратичное отклонение (СКО), однако, если оно не известно, то третий аргумент отсутствует и вместо точного значения СКО используется его выборочная оценка. Полученное значение Z.ТЕСТ(**B2:B206**;0) Z.ТЕСТ(**B2:B206**;0) оказалось равным 0,153, то есть большим как , так и . Поэтому основная гипотеза о равенстве нулю математического ожидания недельной логдоходности данного актива (акция компании «Абрау-Дюрсо») не отклоняется. (см. **Рисунок42**).

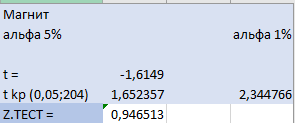


**Рисунок 42. Принятиегипотезы с помощью функции Z.ТЕСТ(акция компании «Абрау-Дюрсо»).**

Аналогичная ситуация – с компаниями «Лукойл» и «Магнит» (см. **Рисунки43** и **44**).



**Рисунок 43. Принятие гипотезы с помощью функции Z.ТЕСТ (акция компании «Лукойл»).**



**Рисунок 44. Принятие гипотезы с помощью функции Z.ТЕСТ (акция компании «Магнит»).**

# **7. Проверкагипотезы о равенстве средних значений логарифмической доходности за последний и предпоследний годы (двусторонний t-тест без каких-либо предположений об однородности дисперсий)**

При решения данной задачи применяется критерий для проверки гипотезы о сравнении двух математических ожиданий при отсутствии предположения о равенстве дисперсий. Рассмотрим решение указанной задачи в общем виде.

**Критерии для проверки гипотезы о сравнении двух математических ожиданий при отсутствии предположения о равенстве дисперсий**

Пусть имеются две нормально распределенные генеральные совокупности *Х* и *Y*. Из них извлечены независимые выборки объемов соответственно *п*x и *п*y, по которым вычислены выборочные средние и .

Пусть при этом дисперсии генеральных совокупностей*х* и *y*не известны и различны.

По извлеченным независимым выборкам объемов соответственно *п*x и *п*y вычисляются исправленные выборочные дисперсии  и .

Рассмотрим нулевую гипотезу *Н*0: .

Альтернативная гипотеза также имеет вид:

1. *Н*1:;
2. *Н*1:;
3. *Н*1:;

Для проверки *Н*0 во всех трех случаях используется статистика

число степеней свободы определяется по формуле

Критическая область определяется в зависимости от вида *Н*1 :

1. *Н*1:: ;
2. *Н*1:: *T*< - ;
3. *Н*1:: .

При решении задачи в **MSExcel** используется функция СТЬЮДЕНТ.ТЕСТ(**массив1; массив2; хвосты; тип)**  с параметрами:

- **массив1** – диапазон, содержащий первую выборку;

- **массив2** – диапазон, содержащий вторую выборку;

- **хвосты** - 1, если конкурирующая гипотеза вида 2 или 3,

2, если конкурирующая гипотеза вида 1;

- **тип** - 3 – если нет предположения о равенстве дисперсий.

Значение вычисляется с помощью функции СТЬЮДЕНТ.ОБР(вероятность; степени\_свободы),

где

вероятность – 1-

степени\_свободы -

Для проверки гипотезы о равенстве средних значений логарифмических доходностей (после удаления выбросов) рассматриваемых акций за последний и предпоследний годы используем описанный выше метод. В диапазоне **А1:А206** размещаются даты наблюдений, в диапазоне **B1:D206**размещаютсясоответствующие значения логарифмических доходностей (акций компаний «Абрау-Дюрсо», «Лукойл», «Магнит»).

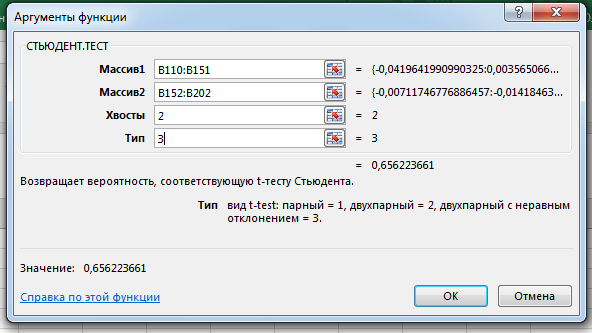
Рассмотрим более подробно решение задачи для акций «Абрау-Дюрсо». В диапазоне **B110:B151**размещены значения логдоходностейакций «Абрау-Дюрсо» для 2018 года, а в диапазоне **B152:B202 –** для2019 года. Обращаемся к статистической функции СТЬЮДЕНТ.ТЕСТ, указывая значения ее параметров (см. **Рисунок** ), как показано ниже:

СТЬЮДЕНТ.ТЕСТ(**B110:B151; B152:B202;2;3).**

Значение параметра «**хвосты»** полагается равным 2, так как выдвигается конкурирующая гипотеза вида

*Н*1:.

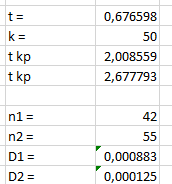
Значение параметра «**тип»** полагается равным 3, поскольку нет предположения о равенстве дисперсий и как будет показано далее дисперсии заметно отличаются.



**Рисунок 45. Ввод данных для функции СТЬЮДЕНТ.ТЕСТ.**

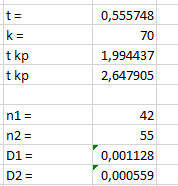
Результатом работы этой функции является вычисленное значение статистики критерия в данном случае оно равно 0,656. Далее вычисляется число степеней свободы

и критические значения критерия для значений с помощью функции СТЬЮДЕНТ.ОБР. Результаты вычислений (см. **Рисунок46**) позволяют сделать вывод о справедливости выдвигаемой основной гипотезы. То есть гипотеза о равенствесредних значений логарифмической доходности за последний и предпоследний годы принимается при уровнях значимости .

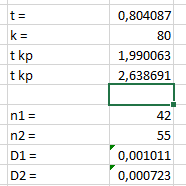


**Рисунок 46. Данные для вывода о справедливости выдвигаемой основной гипотезы. (акция компании «Абрау-Дюрсо»).**

Аналогичные расчеты выполняются для акций двух других компаний. Как можно видеть результаты аналогичны (см. **Рисунки 47 и 48** )



**Рисунок 47. Данные для вывода о справедливости выдвигаемой основной гипотезы. (акция компании «Лукойл»).**



**Рисунок 48. Данные для вывода о справедливости выдвигаемой основной гипотезы.(акция компании «Магнит»).**

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

**Критерий для проверки гипотезы о сравнении двух дисперсий.**

Пусть имеются две нормально распределенные генеральные совокупности *Х* и *Y*. Из них извлечены независимые выборки объемов соответственно *п*x и *п*y, по которым вычислены исправленные выборочные дисперсии  и . Требуется при заданном уровне значимости α проверить нулевую гипотезу

*Н*0: *D*(*X*) = *D*(*Y*)

о равенстве дисперсий рассматриваемых генеральных совокупностей.

В качестве критерия принимается случайная величина

 (6)

- отношение большей выборочной дисперсии к меньшей. Она имеет распределение Фишера-Снедекора со степенями свободы *k*1 = *n*1 – 1 и *k*2 = *n*2 – 1, где *n*1 – объем выборки, по которой вычислена большая исправленная дисперсия, а *n*2 – объем второй выборки.

пусть *Н*1: *D*(*X*) >*D*(*Y*).

При

*Fкр*(*α; п*x -1; *п*y-1)

нулевая гипотеза принимается,

при

*>Fкр*(*α; п*x -1; *п*y-1)

отвергается.

При решении задачи в **Excel** используется функция

**F.ТЕСТ**(**массив1; массив2**) с параметрами:

- **массив1** – диапазон, содержащий первую выборку;

- **массив2** – диапазон, содержащий вторую выборку.

Функция [3]возвращает вероятность сходства дисперсий тестируемых выборок (2*α*). Если вычисленное значение оказывается больше заданного уровня значимости, то гипотеза *Н*0принимается, в противном случае – отклоняется.

Так же, как в предыдущем разделе в диапазоне **А1:А206** размещаются даты наблюдений, в диапазоне **B1:D206**размещаютсясоответствующие значения логарифмических доходностей (акций компаний «Абрау-Дюрсо», «Лукойл», «Магнит»).

Рассмотрим более подробно решение задачи для акций «Абрау-Дюрсо». В диапазоне **B110:B151**размещены значения логдоходностейакций «Абрау-Дюрсо» для 2018 года, а в диапазоне **B152:B202 –** для2019 года. Обращаемся к статистической функции **F.ТЕСТ**



Вычисленное значение 3,64Е-10 менее как для

, что позволяет отклонить гипотезу о равенстве дисперсий.

Для компании «Лукойл» получается иной результат. Вычисленное значение функции**F.ТЕСТ**оказывается равным 0,0249. Легко убедится, что

0,01/2=0,005<0,0249<0,025=0,05/2,

что позволяет отклонить гипотезу о равенстве дисперсийпри уровне значимости отклонить эту гипотезу при .

В случае компании «Магнит» вычисленное значение функции **F.ТЕСТ**оказывается равным 0,22, что больше 0,025=0,05/2 и тем более 0,01/2=0,005. Поэтому в этом случае нет оснований отклонить гипотезу о равенстве дисперсий.

# **9. Сравнительный анализ логдоходностей трех компаний.**

**Однофакторный дисперсионный анализ.**

При обработке результатов наблюдений часто возникает вопрос о том насколько существенное влияние оказывает какой-либо фактор или группа факторов на измеряемую величину[3]. Для ответа на этот вопрос может быть использован дисперсионный анализ.

***Формулировка задачи однофакторного дисперсионного анализа.***

Пусть случайная величина *X* наблюдается при *m* постоянных значениях некоторого фактора *Y*. Результаты наблюдений составляют *m*выборок (групп) объемами *nk*, принадлежащие *m*генеральным совокупностям, имеющим равные и неизвестные дисперсии и математические ожидания *k,(k= 1,…,m)*

Проверяется гипотеза .

Обозначим *Xiki* -е значение в *k* -й выборке (*k = 1,…,l; i=1,…, nk*).

Можно определить 3 показателя вариации признака *X*:

1) **Общая дисперсия**

- сумма квадратов отклонений,

-общая дисперсия.

2) **Межгрупповая дисперсия**

Характеризует вариацию изучаемого признака, возникающую под влиянием признака-фактора, положенного в основание группировки

,

где *m* - число групп,

*nj* - число единиц в *j*-ой группе,

- среднее по j-ой группе,

среднее по всей совокупности.

3) **Внутригрупповая дисперсия**

Внутригрупповая дисперсия отражает часть вариации, происходящей под влиянием неучтенных факторов, которая не зависит от группировочного признака.

- сумма квадратов отклонений,

**Средняя из внутригрупповых дисперсий** по совокупности в целом - вариация значений признака под влиянием прочих факторов.

***Правило сложения дисперсий:*** *общая дисперсия, которая возникает под влиянием всех факторов, равна сумме средней из внутригрупповых и межгрупповой:*

*.*

***Общая дисперсия*** по комплексу на одну степень свободы вариации

,

***Факторная дисперсия*** по комплексу на одну степень свободы вариации

,

***Остаточная дисперсия*** по комплексу на одну степень свободы вариации

,

Для степеней свободы

*,*

или

.

Если

,

то гипотеза принимается.

**Общая схема дисперсионного анализа:**

1) Вычисление общей дисперсии комплекса;

2) Вычисление факторной дисперсии комплекса;

3) Вычисление остаточной дисперсии комплекса;

4) Вычисление корреляционного отношения;

5) Проверка равенства

*,*

6) Проверка равенства

*,*

7) Вычисление

8) Вычисление статистики

.

9) Определение критического значенияпри заданном уровне значимости;

10) Сравнение

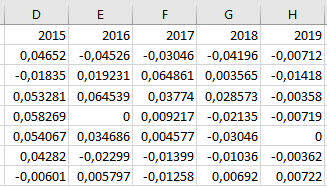
- влияние изучаемого фактора существенно,

- влияние изучаемого фактора недостоверно.

**Решение задачи вMSExcel.**

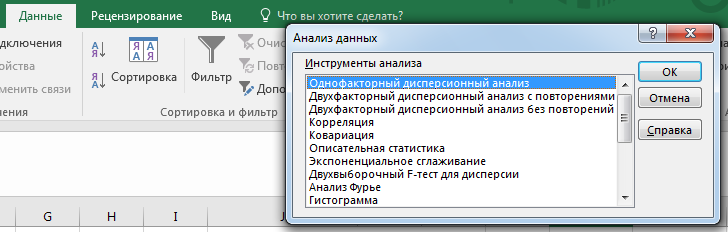
***Задача 1.***Проверить гипотезу о значимости зависимости значений логдоходности акций компании «Абрау-Дюрсо» от номера года.

***Решение.*** На новом листе «Дисп анализ» в диапазоне **А1:А206** размещаются даты наблюдений, в диапазоне **B1:B206**размещаютсясоответствующие значения логарифмических доходностей акций компании «Абрау-Дюрсо». Далее в столбцах D-:-Hразмещаем значения логдоходностей, для 2015, 2016,…, 2019 года (см. **Рисунок49.**)

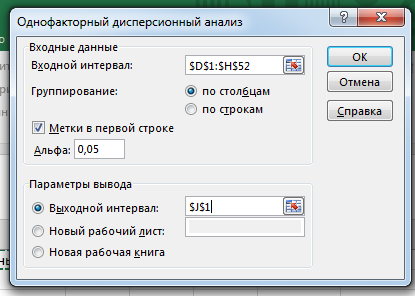


**Рисунок 49.Размещение значений логдоходностей для каждого года.**

Затем обращаемся к процедуре «Однофакторный дисперсионный анализ» (Данные-Анализ данных –Однофакторный дисперсионный анализ) (см. **Рисунок50**). При этом следует указать значения входных параметров (см. **Рисунок51**):

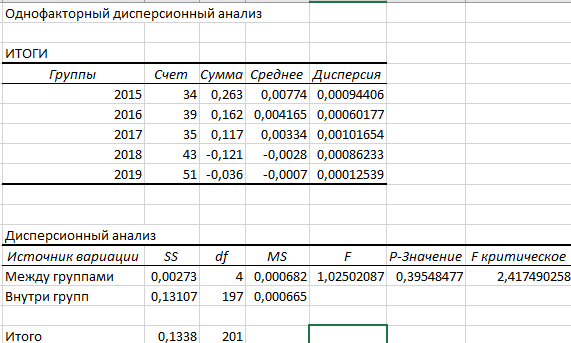


**Рисунок 50.Обращение к процедуре «Однофакторный дисперсионный анализ».**



**Рисунок 51. Ввод входных параметров.**

Результаты работы процедуры представлены с помощью двух таблиц (см. **Рисунок52**).



**Рисунок 52.Результаты работы процедуры.**

Первая таблица - **ИТОГИ**:

“Счет” – частота – для каждого года объем выборки;

“Сумма” – сумма значений показателя по строкам;

“Среднее” – среднее значение показателя (логдоходности) в каждой группе

(для каждого года);

“Дисперсия” – частная дисперсия показателя (логдоходности)в каждой

группе (для каждого года);

Вторая таблица - Таблица **«Дисперсионный анализ»** представляет результаты дисперсионного анализа однофакторного комплекса, в котором первая колонка “Источник вариации” содержит наименование дисперсий.

Графа “SS” - это сумма квадратов отклонений,

“df” - степень свободы,

графа “MS” - средний квадрат (MS= SS/df),

“F” - критерий фактического F – распределения.

“P - значение” - вероятность того, что дисперсия, воспроизводимая уравнением, равна дисперсии остатков. Определяет вероятность того, что полученная количественная определенность взаимосвязи между факторами и результатом может считаться случайной.

“F - критическое” - это значение F – теоретического, которое впоследствии сравнивается с F – фактическим.

В рассматриваемом примере

.

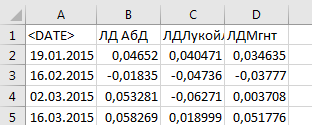
Кроме того,

P – значение=0,395> 0,05=

Каждое из последних двух неравенств позволяет сделать вывод о том, что гипотезу об отсутствии зависимости значений логдоходности акций компании «Абрау-Дюрсо» от номера года нет оснований отклонить при уровне значимости = 0,05.

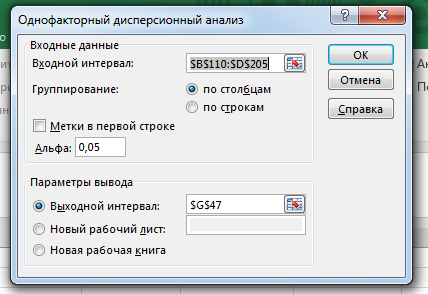
***Задача 2.***Проверить гипотезу о значимости зависимости значений логдоходности акций компаний «Абрау-Дюрсо», «Лукойл», «Магнит» от названия компании на промежутке времени от 1 января 2018 года по 27 января 2020 года при уровне значимости = 0,05.

***Решение.*** На новом листе «Дисперсионный анализ 2» в диапазоне **А1:А206** размещаются даты наблюдений, в диапазоне **B1:D206**размещаютсясоответствующие значения логарифмических доходностей акций компаний «Абрау-Дюрсо», «Лукойл» и «Магнит» (см. **Рисунок53**).



**Рисунок 53.Значения логарифмических доходностей акций компаний «Абрау-Дюрсо», «Лукойл» и «Магнит».**

Затем обращаемся к процедуре «Однофакторный дисперсионный анализ» (Данные-Анализ данных – Однофакторный дисперсионный анализ) (см. **Рисунок54).** При этом следует указать значения входных параметров:

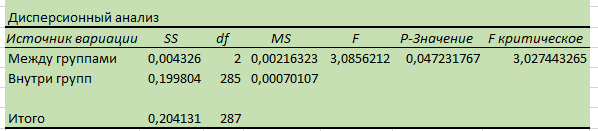


**Рисунок 54. Ввод значений входных параметров.**

Результаты работы процедуры представлены с помощью двух таблиц (см. **Рисунки 55 и56**).



**Рисунок 55.Результаты работы процедуры (итоги).**



**Рисунок 56.Результаты работы процедуры (дисперсионный анализ).**

В этом примере

.

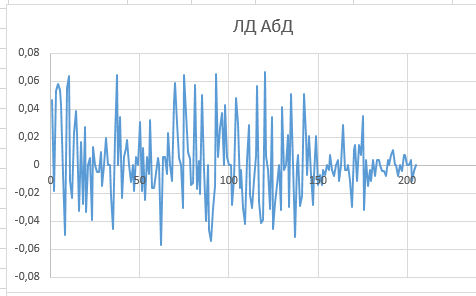
Кроме того,

P – значение =0,0470,05=

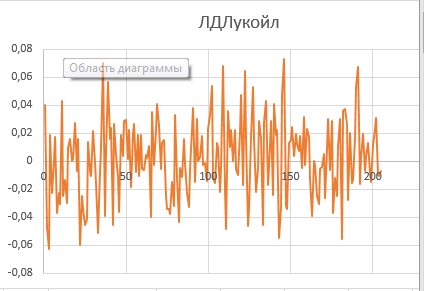
Каждое из последних двух неравенств позволяет сделать вывод о том, что гипотезу об отсутствии зависимости значений логдоходности акций компаний«Абрау-Дюрсо», «Лукойл», «Магнит» от названия компании на промежутке времени от 1 января 2018 года по 27 января 2020 года следует отклонить при уровне значимости = 0,05.

# **10. Графики временных рядов логдоходностей трех компаний**

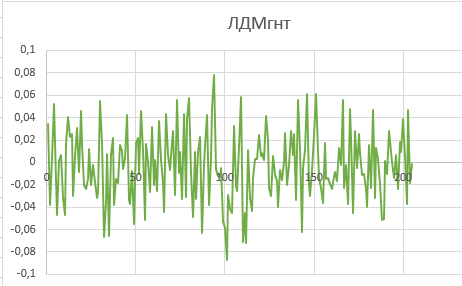
Для построения графиков зависимостей логдоходностей от времени используем сервис **MSExcel**«Вставка- Диаграммы». Для большей наглядности целесообразно каждый график изображать на отдельной системе координат (см. **Рисунки 57, 58 и59**).

****

**Рисунок 57.График зависимостей логдоходностей от времени(«Абрау-Дюрсо»).**

****

**Рисунок 58.График зависимостей логдоходностей от времени(«Лукойл»).**

****

**Рисунок 59.График зависимостей логдоходностей от времени(«Магнит»).**

Анализируя представленные графики можно сделать вывод о схожести представленных зависимостей.

# **11. Исследование тесноты связи между логдоходностями**

Для исследования тесноты связи между значениями логдоходностей акций компаний «Абрау-Дюрсо», «Лукойл», «Магнит» используем процедуру «Корреляция».

Для оценки значимости статистического коэффициента корреляции применяется следующий алгоритм.

1. Выдвигается нулевая гипотеза

при альтернативной гипотезе

2. Вычисляется статистика

,

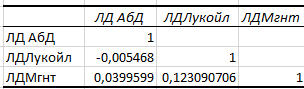
- распределение Стьюдента с степенями свободы.

3. При заданном уровне значимости *α* определяется критическое значение .

Если то гипотеза отклоняется, в противном случае – не отклоняется.

Опишем решение задачи проверки гипотезы о значимости корреляций в **MSExcel**.

На новом листе «Удаление выбросов» в диапазоне **А1:А206** размещены даты наблюдений, в диапазоне **B1:D206**размещенысоответствующие значения логарифмических доходностей акций компаний «Абрау-Дюрсо», «Лукойл» и «Магнит». Вызываем процедуру «Корреляция». Обращение к этой процедуре «Данные-Анализ данных – Корреляция»(см. **Рисунок** ).

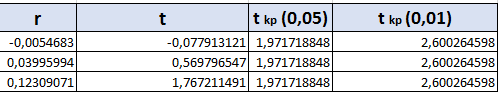


**Рисунок 60. Корреляционная матрица логдоходностей.**

Для каждого значения статистического коэффициента корреляции вычисляем значение

,

далеенаходим критические значения распределения Стьюдента с помощью статистической функции СТЬЮДЕНТ.ОБР.2Х, для которой следует указать значения параметров и . Результаты вычислений приведены в следующей таблице



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

# **Заключение**

В учебном пособии рассматривается решение основных типовых задач визуализации и обработки данных средствами **MSExcel**. Следует отметить, что подобным образом могут быть решены другие похожие задачи. Например, для проверки гипотез о виде распределения логдоходностей могут быть использованы иные критерии и иные законы распределения. Кроме того,следует отметить, что все предложенные задачи могут быть решены средствами языка R.

# **Список литературы**

1. Соловьев В. И. Анализ данных: теория вероятностей и прикладная статистика, обработка и визуализация данных в MicrosoftExcel: учебник/ В.И. Соловьев. – Москва: КНОРУС,2019.-498 с. – (Бакалавриат).

2. Браилов А. В., Глебов В. И., Криволапов С. Я., Рябов П. Е.

Теория вероятностей и математическая статистика : учебник-практикум. – М.-Ижевск : НИЦ «Регулярная и хаотическая динамика» ; Институт компьютерных исследований, 2016. – 414 с.

3. Козлов А.Ю., Мхитарян В.С., Шишов В.Ф. Статистический анализ данных в MSExcel: Учебное пособие. –М.: ИНФРА-М, 2013. – 320 С. –(Высшее образование)

4. Орлов А.И. Непараметрические критерии согласия Колмогорова, Смирнова, омега-квадрат и ошибки при их применении. Научный журнал КубГАУ, №97(03), 2014 года, с. 1-29. <http://ej.kubagro.ru/2014/03/pdf/47.pdf>