Федеральное государственное образовательное бюджетное учреждение высшего образования «Финансовый университет при Правительстве Российской Федерации»

РАСЧЕТНО АНАЛИТИЧЕСКАЯ РАБОТА

По дисциплине: Анализ данных

Выполнил: Ковалев Алексей Игоревич,

студент группы ПИ19-3

Преподаватель: Пяткина Дарья Анатольевна,

к.ф.-м.н., доцент

Москва, 2021

ОГЛАВЛЕНИЕ

[1. ИСХОДНЫЕ ДАННЫЕ 3](#_Toc73285010)

[2. РАСЧЕТ ДОПОЛНИТЕЛЬНЫХ ПРИЗНАКОВ ОЦЕНКИ 4](#_Toc73285011)

[3. ПЕРВИЧНЫЕ ИССЛЕДОВАНИЯ (ГРЯЗНЫЕ ДАННЫЕ) 7](#_Toc73285012)

[4. ОЧИСТКА ДАННЫХ 18](#_Toc73285013)

[5. ИССЛЕДОВАНИЕ ЛОГАРИФМИЧЕСКИХ ДОХОДНОСТЕЙ (ДАННЫЕ БЕЗ ВЫБРОСОВ И ПРОПУСКОВ) 20](#_Toc73285014)

[6. ПРОВЕРКА ГИПОТЕЗ 23](#_Toc73285015)

[6.1 Проверка гипотезы о нормальности логарифмической доходности 23](#_Toc73285016)

[6.2 Проверка гипотезы о равенстве нулю математического ожидания недельной логарифмической доходности 31](#_Toc73285017)

[6.3 Проверка гипотезы о равенстве дисперсий логарифмической доходности за последний и предпоследний годы 32](#_Toc73285018)

[6.4 Проверка гипотезы о равенстве средних значений логарифмической доходности за последний и предпоследний годы 33](#_Toc73285019)

[7. ДИСПЕРСИОННЫЙ АНАЛИЗ 35](#_Toc73285020)

[7.1 Проверка гипотезы о значимости зависимости значений логарифмической доходности акций компаний от номера года 35](#_Toc73285021)

[ЗАКЛЮЧЕНИЕ 37](#_Toc73285022)

1. ИСХОДНЫЕ ДАННЫЕ

Согласно номеру моего варианта – 209, и номеру моей группы – ПИ19-3 я выбрал следующие данные за период 01.01.2016-01.01.2021: 69 – ФСК ЕЭС ао, 39 – МРСКЮга ао, 30 – ММК. Загруженные в Excel исходные данные приведены на рисунке 1.

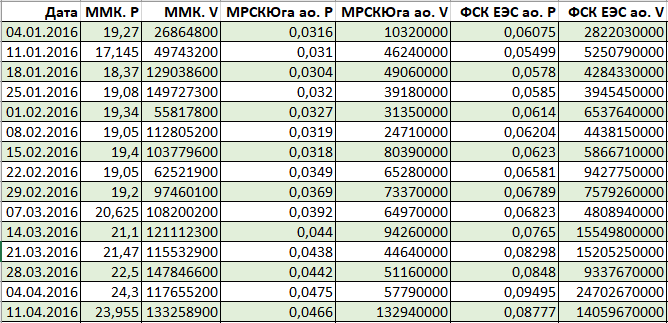


Рисунок 1 - исходные данные (лист "Исходные данные")

Для сокращения наименований оценочных параметров я ввел следующие обозначения: P – цена (price), V – объем (volume), I – доходность (income), LN – натуральный логарифм параметра, R – относительный параметр (relative).

1. РАСЧЕТ ДОПОЛНИТЕЛЬНЫХ ПРИЗНАКОВ ОЦЕНКИ

После загрузки данных я выполнил расчет дополнительных параметров для дальнейшего исследования: натуральный логарифм цены (рисунок 2), доходность (рисунок 3), натуральный логарифм доходности (рисунок 4), натуральный логарифм объема (рисунок 5).

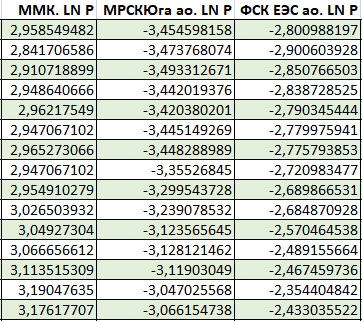


Рисунок 2 - натуральный логарифм цены (лист "Исходные данные")

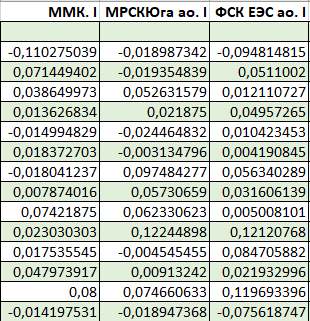


Рисунок 3 - доходность (лист "Исходные данные")

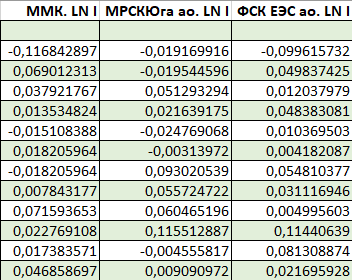


Рисунок 4 - натуральный логарифм доходности (лист "Исходные данные")

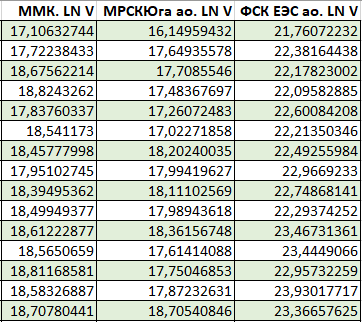


Рисунок 5 - натуральный логарифм объема (лист "Исходные данные")

Итоговая таблица с исходными данными и рассчитанными параметрами представлена на рисунке 6.



Рисунок 6 - итоговая таблица (лист "Исходные данные")

1. ПЕРВИЧНЫЕ ИССЛЕДОВАНИЯ (ГРЯЗНЫЕ ДАННЫЕ)

Для исследования изменения цен с течением времени я визуализировал цены акций 3-ех компаний на одном графике (рисунок 8), предварительно переведя их в относительные величины (рисунок 7).

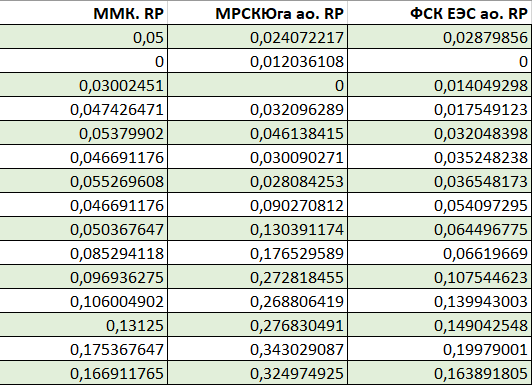


Рисунок 7 - относительные значения цен акций (лист "График цен")

Рисунок 8 - график изменения цен акций компаний с течением времени (лист "График цен")

Так же для визуализации имеющихся данных по логарифмической цене акций были построены ящики с усами (рисунок 9, 10, 11). На графиках наглядно видны выбросы и обозначены значения медиан выборок.

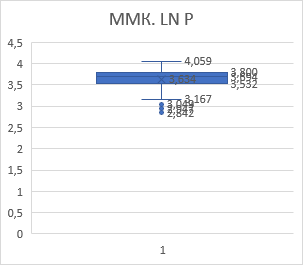


Рисунок 9 - ящик с усами по логарифму цен акций компании ММК (лист "Ящик с усами")

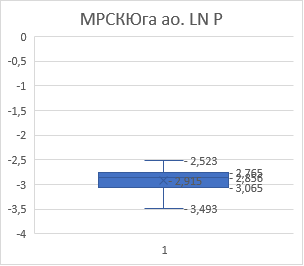


Рисунок 10 - ящик с усами по логарифму цен акций компании МРСКЮга ао (лист "Ящик с усами")

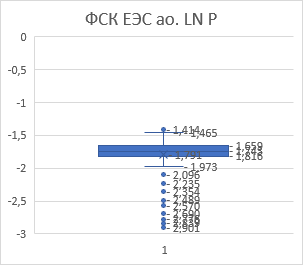


Рисунок 11 - ящик с усами по логарифму цен акций компании ФСК ЕЭС ао (лист "Ящик с усами")

Для получения основных значений статистики по представленным выборкам я воспользовался инструментом «Описательная статистика» из пакета «Анализ данных». На рисунках 12-15 представлена «Описательная статистика» для цен акций, логарифмов цены, логарифмов доходностей и объем продажи акций соответственно.



Рисунок 12 - описательная статистика. Цены акций (лист "Описательная статистика")



Рисунок 13 - описательная статистика. Логарифмы цен (лист "Описательная статистика")



Рисунок 14 - описательная статистика. Логарифмы доходностей (лист "Описательная статистика")



Рисунок 15 - описательная статистика. Логарифмы объемов продаж (лист "Описательная статистика")

Затем, для исследования связей между данными (логарифмов цен и логарифмов объемов) разных компаний и выявления зависимостей и закономерностей я построил диаграммы рассеяния для каждой пары компаний (рисунки 16-18 и 20-22) и корреляционные матрицы признаков (рисунки 19, 23).

Рисунок 16 - диаграмма рассеяния значений признаков логарифмов объемов торгов акций компаний "МРСКЮга ао" и "ММК" (лист "LN V. Диаграммы рассеяния")

Рисунок 17 - диаграмма рассеяния значений признаков логарифмов объемов торгов акций компаний "ФСК ЕЭС ао" и "ММК" (лист "LN V. Диаграммы рассеяния")

Рисунок 18 - диаграмма рассеяния значений признаков логарифмов объемов торгов акций компаний "ФСК ЕЭС ао" и "МРСКЮга ао" (лист "LN V. Диаграммы рассеяния")

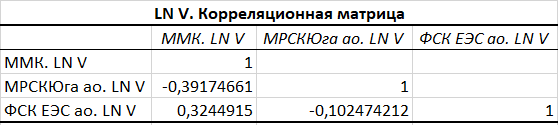


Рисунок 19 - корреляционная матрица логарифмов объемов торгов акций компаний (лист "LN V. Диаграммы рассеяния")

Рисунок 20 - диаграмма рассеяния значений логарифмов цен акций компаний "МРСКЮга ао" и "ММК" (лист "LN P. Диаграммы рассеяния")

Рисунок 21 - диаграмма рассеяния значений логарифмов цен акций компаний "ФСК ЕЭС ао" и "МРСКЮга ао" (лист "LN P. Диаграммы рассеяния")

Рисунок 22 - диаграмма рассеяния значений логарифмов цен акций компаний "ФСК ЕЭС ао" и "ММК" (лист "LN P. Диаграммы рассеяния")

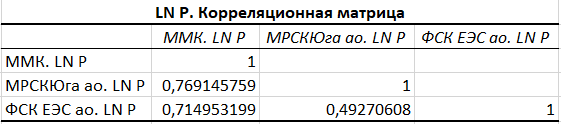


Рисунок 23 - корреляционная матрица логарифмов цен акций компаний (лист "LN P. Диаграммы рассеяния")

Последним пунктом исследования неочищенных данных была визуализация частот значений с помощью гистограмм интервальных частот. Гистограммы были построены с помощью инструмента «Гистограмма» из пакета «Анализ данных». Для использования инструмента сначала необходимо было рассчитать левые и правые границы интервалов (карманы). Полученные диаграммы представлены на рисунках 24-26 – для цен акций, 27-29 – для логарифмов цен.

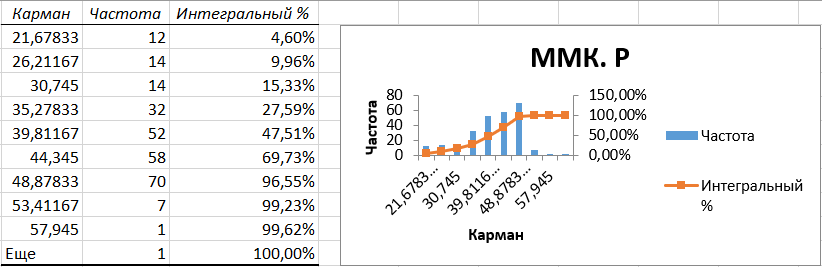


Рисунок 24 - гистограмма интервальных частот для цены акций компании "ММК" (лист "P. Гистограммы частот")

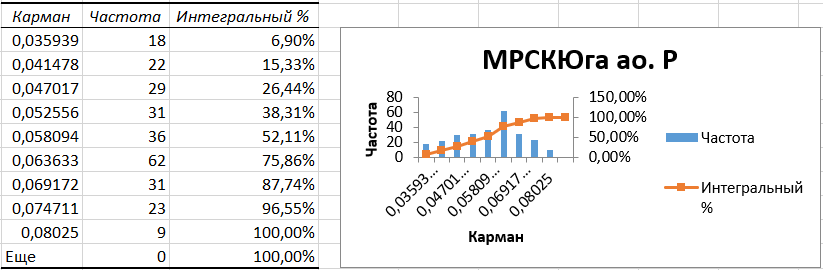


Рисунок 25 - гистограмма интервальных частот для цены акций компании "МРСКЮга" (лист "P. Гистограммы частот")

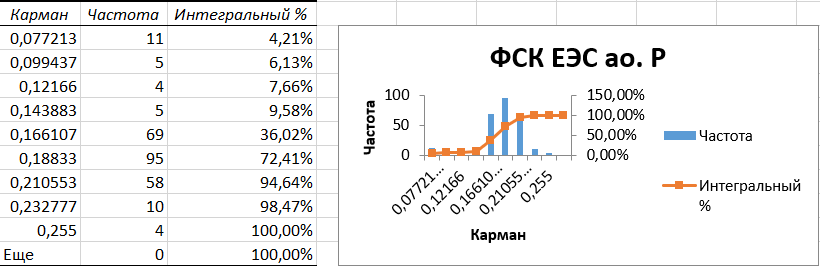


Рисунок 26 - гистограмма интервальных частот для цены акций компании "ФСК ЕЭС ао" (лист "P. Гистограммы частот")

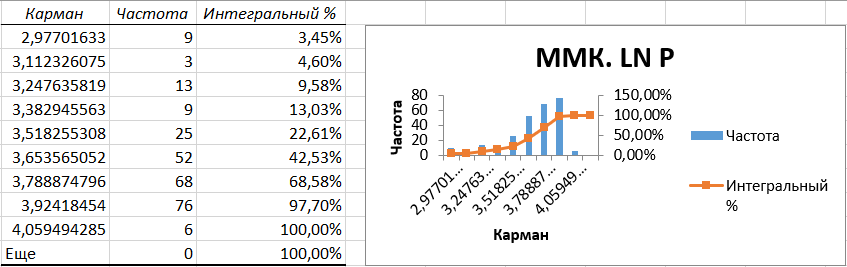


Рисунок 27 - гистограмма интервальных частот для логарифма цены акций компании "ММК" (лист "LN P. Гистограммы частот")

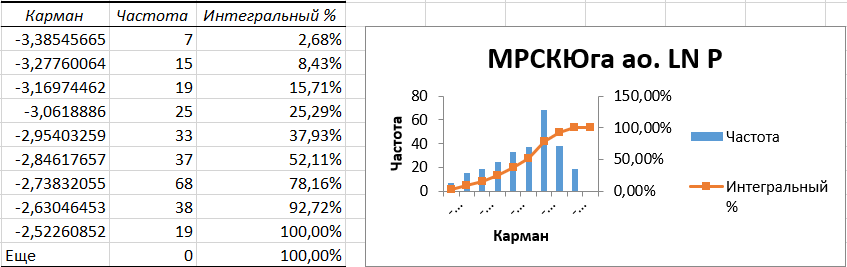


Рисунок 28 - гистограмма интервальных частот для логарифма цены акций компании "МРСКЮга оа" (лист "LN P. Гистограммы частот")

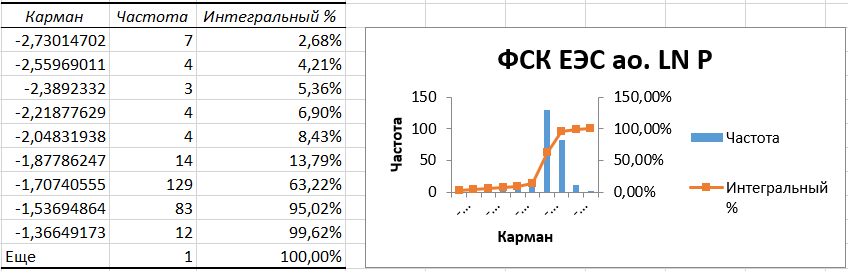


Рисунок 29 - гистограмма интервальных частот для логарифма цены акций компании "ФСК ЕЭС ао" (лист "LN P. Гистограммы частот")

1. ОЧИСТКА ДАННЫХ

Для дальнейшего анализа будут использоваться только очищенные данные по логарифмической доходности акций компаний. Для удаления выбросов были вычислены первый и третий квартили, межквартальные расстояния и пороговые нижние верхние значения для удаления выбросов (рисунок 30).

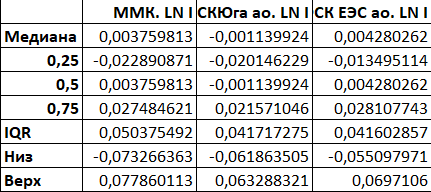


Рисунок 30 - дополнительные расчеты для очистки данных от выбросов (лист "Очищенные данные")

Затем с помощью инструмента «Настраиваемая сортировка» я удалил все строки содержащие выбросы и получил очищенную выборку (рисунок 31) логарифмических доходностей акций.

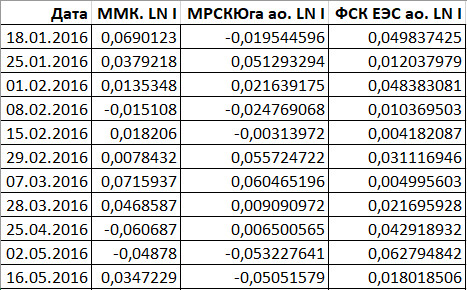


Рисунок 31 - очищенные от выбросов логарифмические доходности (лист "Очищенные данные")

1. ИССЛЕДОВАНИЕ ЛОГАРИФМИЧЕСКИХ ДОХОДНОСТЕЙ (ДАННЫЕ БЕЗ ВЫБРОСОВ И ПРОПУСКОВ)

Для визуализации изменения логарифмических доходностей акций на одном графике я привел их значения к относительным и построил график (рисунок 32).

Рисунок 32 - график изменения логарифмических доходностей трех компаний (лист "C. График LN I")

С помощью пакета «Анализ данных» я получил «Описательную статистику» для очищенных данных (рисунок 33).



Рисунок 33 - описательная статистика. Логарифмы доходностей (лист "C. Описательная статистика")

Затем для дисперсии (рисунок 34) и для стандартного отклонения (рисунок 35) выборки были вычислены и визуализированы доверительные интервалы.



Рисунок 34 - доверительные интервалы для дисперсии выборки логарифмических доходностей (лист "C. Доверительные интервалы")

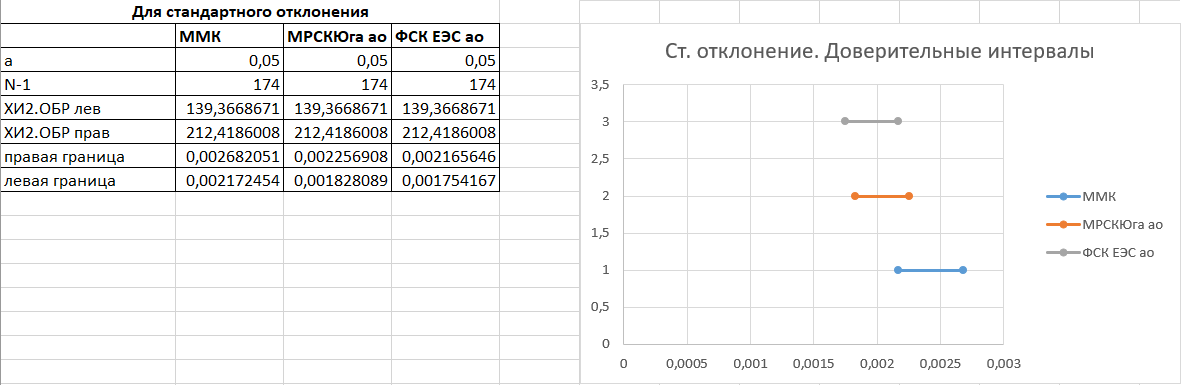


Рисунок 35 - доверительные интервалы для стандартного отклонения выборки логарифмических доходностей (лист "C. Доверительные интервалы")

1. ПРОВЕРКА ГИПОТЕЗ
   1. Проверка гипотезы о нормальности логарифмической доходности

Проверяемые гипотезы: H0 – логарифмические доходности распределены по закону нормального распределения, H1 – распределение не является нормальным.

Для проверки гипотез воспользуемся двумя способами: проверка по критерию Пирсона(хи-квадрат), проверка по критерию Колмогорова-Смирнова.

Начнем с критерия Пирсона. Для начала я разбил размах варьирования очищенной выборки логарифмических доходностей на интервалы и определил частоты попадания значений (рисунки 36-38) в каждый интервал с помощью инструмента «Гистограмма» из пакета «Анализ данных».

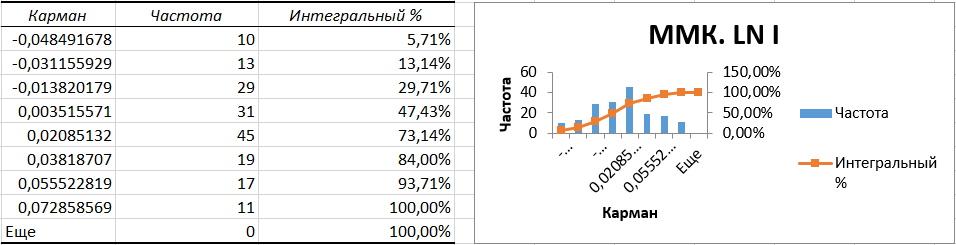


Рисунок 36 - гистограмма интервальных частот для логарифмов цены акций компании "ММК"

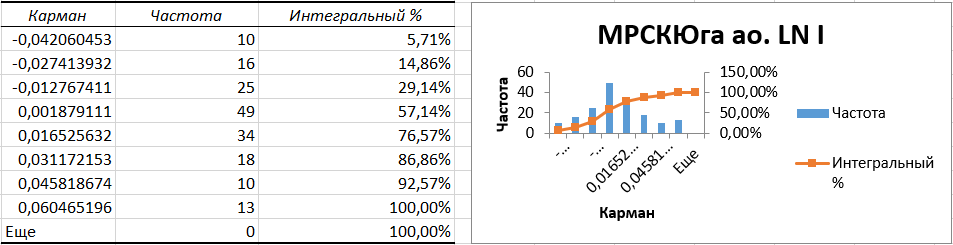


Рисунок 37 - гистограмма интервальных частот для логарифмов цены акций компании "МРСКЮга ао"

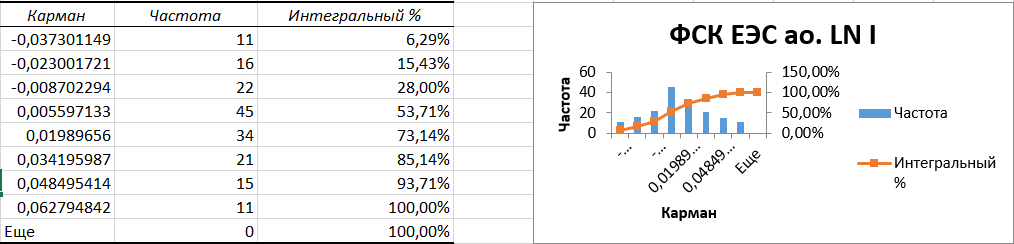


Рисунок 38 - гистограмма интервальных частот для логарифмов цены акций компании "ФСК ЕЭС ао"

Затем я вычислил теоретические значения нормального распределения (плотности и функции) (рисунки 39, 42, 45). Для сравнения их с фактическими данные были визуализированы с помощью графиков (рисунки 40, 41, 43, 44, 46, 47).

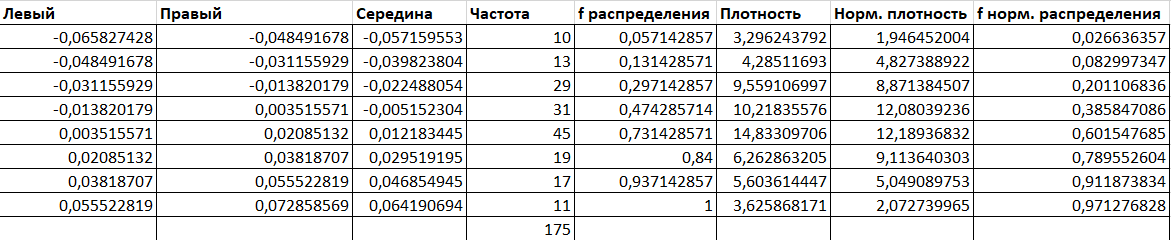


Рисунок 39 - фактические значения и значения плотности и функции нормального распределения для компании "ММК"

Рисунок 40 - график теоретической плотности и гистограмма для логарифмической доходности акций компании "ММК"

Рисунок 41 - график эмпирической функции распределения и график теоретической функции распределения для логарифмической доходности акций компании "ММК"

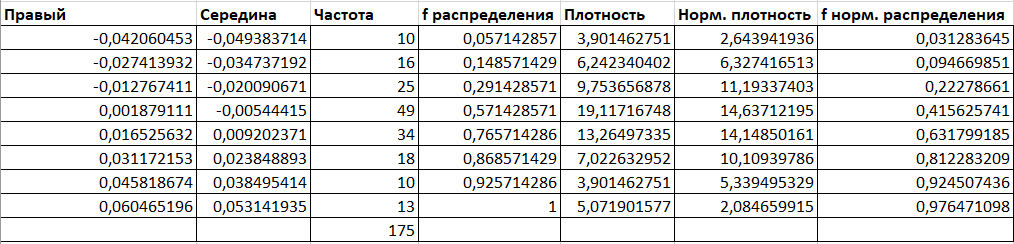


Рисунок 42 - фактические значения и значения плотности и функции нормального распределения для компании "МРСКЮга ао"

Рисунок 43 - график теоретической плотности и гистограмма для логарифмической доходности акций компании "МРСКЮга ао"

Рисунок 44 - график эмпирической функции распределения и график теоретической функции распределения для логарифмической доходности акций компании "МРСКЮга ао"

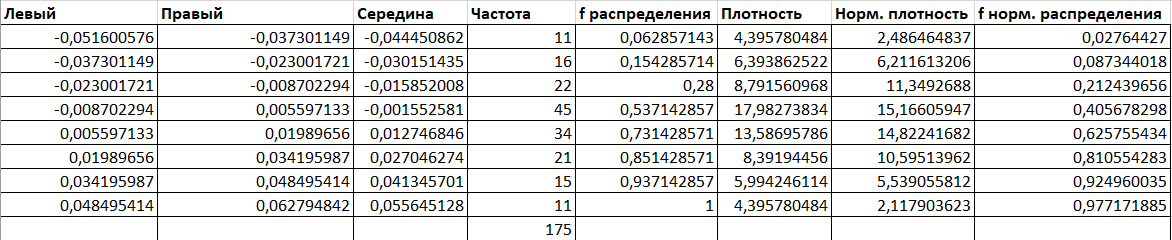


Рисунок 45 - фактические значения и значения плотности и функции нормального распределения для компании "ФСК ЕЭС ао"

Рисунок 46 - график теоретической плотности и гистограмма для логарифмической доходности акций компании "ФСК ЕЭС ао"

Рисунок 47 - график эмпирической функции распределения и график теоретической функции распределения для логарифмической доходности акций компании "ФСК ЕЭС ао"

Визуально для всех трех компаний фактические значения плотности и функции близки к значениям нормального распределения. Для более точной оценки найдем значение критерия хи-квадрат (рисунки 48-50).



Рисунок 48 - проверка гипотезы с помощью критерия Пирсона. ММК

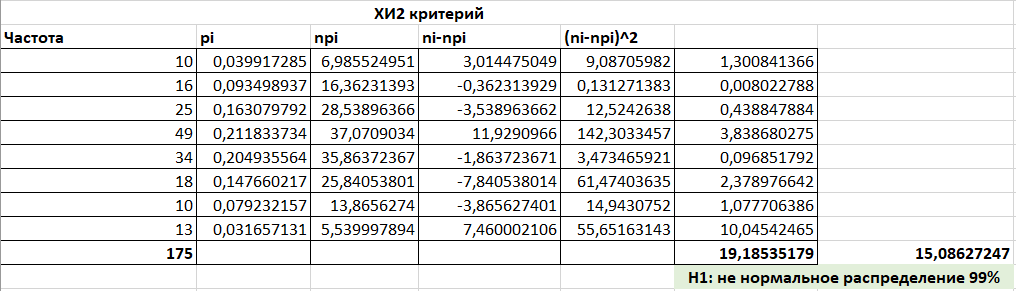


Рисунок 49 - проверка гипотезы с помощью критерия Пирсона. МРСКЮга ао



Рисунок 50 - проверка гипотезы с помощью критерия Пирсона. ФСК ЕЭС ао

Для компании «ММК» рассчитанное значение (рисунок 48) меньше критического (не попадает в критическую область), следовательно, нет оснований отвергать гипотезу H0 – распределение является нормальным.

Для компании «МРСКЮга ао» рассчитанное значение (рисунок 49) больше критического (попадает в критическую область), следовательно, нужно принять гипотезу H1 – распределение не является нормальным с вероятностью 99%.

Для компании «ФСК ЕЭС» рассчитанное значение (рисунок 50) меньше критического (не попадает в критическую область), следовательно, принимаем гипотезу H0 – распределение является нормальным.

Далее я выполнил проверку этих же гипотез с помощью критерия Колмогорова-Смирнова (рисунки 51-53).

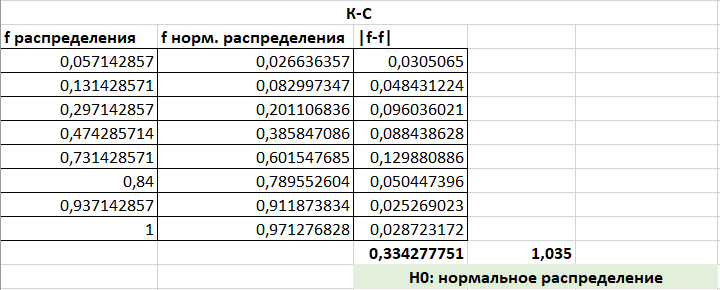


Рисунок 51 - проверка гипотезы с помощью критерия Колмогорова-Смирнова. ММК



Рисунок 52 - проверка гипотезы с помощью критерия Колмогорова-Смирнова. МРСКЮга ао



Рисунок 53 - проверка гипотезы с помощью критерия Колмогорова-Смирнова. ФСК ЕЭС ао

Выводы, которые можно сделать по вычисленным значениям, полученным с помощью критерия Колмогорова-Смирнова, полностью совпадают с выводами, сделанными ранее с помощью критерию Пирсона.

* 1. Проверка гипотезы о равенстве нулю математического ожидания недельной логарифмической доходности

Проверяемые гипотезы: H0 – математическое ожидание недельной логарифмической доходности акций компании равно нулю, H1 – математическое ожидание недельной логарифмической доходности акций компании отлично от нуля.

Для проверки этих гипотез я воспользовался встроенной функцией Z.ТЕСТ, результаты работы функции приведены на рисунке 54.



Рисунок 54 - значения, вычисленные с помощью функции Z.ТЕСТ

Для компании «ММК» полученное значение оказалось равным 0,046, то есть большим , но меньшим , следовательно, можно принять гипотезу H0 - математическое ожидание недельной логарифмической доходности акций «ММК» равно нулю.

Для компании «МРСКЮга ао» полученное значение оказалось равным 0,454, то есть большим и , следовательно, нет оснований отклонить гипотезу H0 - математическое ожидание недельной логарифмической доходности акций «МРСКЮга ао» равно нулю.

Для компании «ФСК ЕЭС ао» полученное значение оказалось равным 0,454, то есть большим и , следовательно, нужно принять гипотезу H1 - математическое ожидание недельной логарифмической доходности акций «ФСК ЕЭС ао» не равняется нулю.

* 1. Проверка гипотезы о равенстве дисперсий логарифмической доходности за последний и предпоследний годы

Проверяемые гипотезы: H0 – дисперсии логарифмической доходности за последний и предпоследний годы равны, H1 – дисперсии логарифмической доходности за последний и предпоследний годы не равны.

Результат работы функции F.ТЕСТ представлен на рисунке 55.

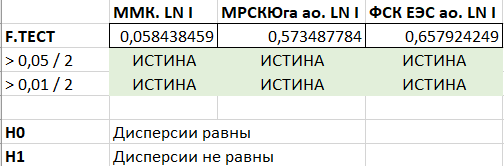


Рисунок 55 - значения, вычисленные с помощью функции F.ТЕСТ

Для всех компаний вычисленные значения значительно превосходят и , следовательно, нет оснований отклонить гипотезу H0 – дисперсии логарифмической доходности за последний и предпоследний годы равны для всех рассматриваемых акций компаний.

* 1. Проверка гипотезы о равенстве средних значений логарифмической доходности за последний и предпоследний годы

Проверяемые гипотезы: H0 – средние значения логарифмической доходности за последний и предпоследний годы равны, H1 – дисперсии логарифмической доходности за последний и предпоследний годы не равны.

Для проверки этих гипотез я воспользовался встроенной функцией Z.ТЕСТ, результаты работы функции приведены на рисунке 56.



Рисунок 56 - значения, вычисленные с помощью функции T.ТЕСТ

Результаты вычислений позволяют сделать вывод о справедливости выдвигаемой основной гипотезы. То есть гипотеза о равенстве средних значений логарифмической доходности за последний и предпоследний годы принимается при уровнях значимости и

1. ДИСПЕРСИОННЫЙ АНАЛИЗ
   1. Проверка гипотезы о значимости зависимости значений логарифмической доходности акций компаний от номера года

Проверяемые гипотезы: H0 – зависит, H1 – не зависит.

С помощью одноименного инструмента из пакета «Анализ данных» я провел однофакторный дисперсионный анализ, результаты которого представлены на рисунках 57-59.

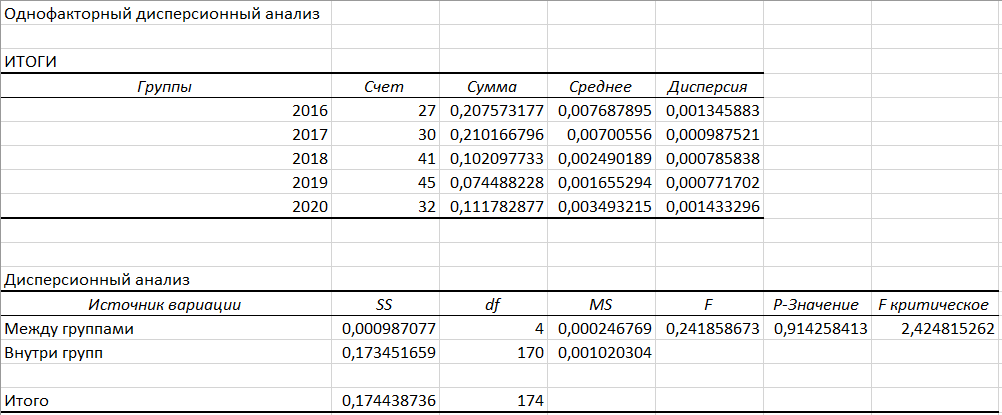


Рисунок 57 - однофакторный дисперсионный анализ. ММК

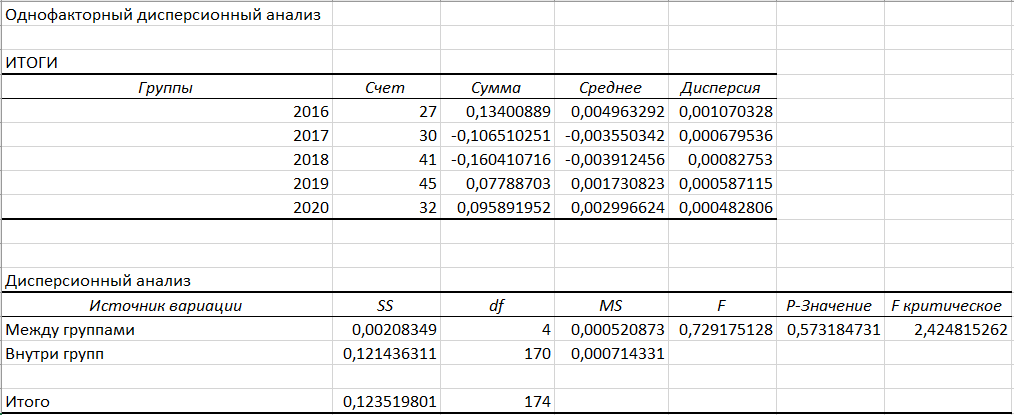


Рисунок 58 - однофакторный дисперсионный анализ. МРСКЮга ао

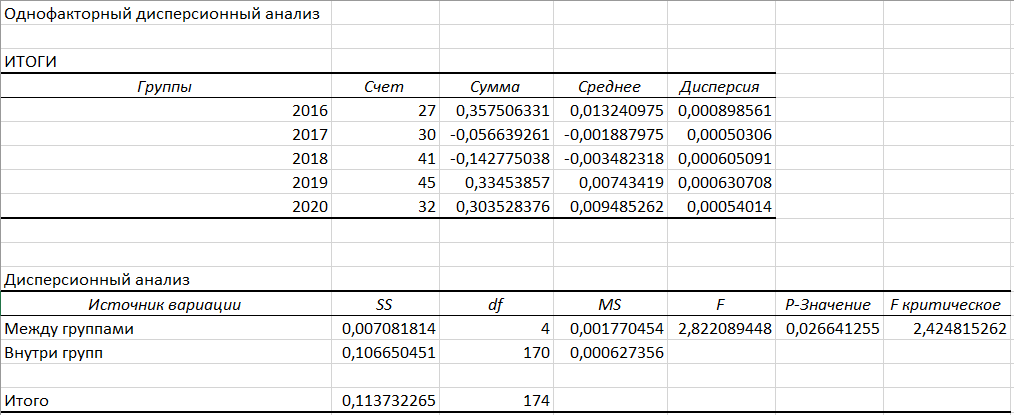


Рисунок 59 - однофакторный дисперсионный анализ. ФСК ЕЭС ао

Полученные неравенства для F и F-критического, а также значения P-значение позволяют сделать выводы о том, что нет оснований отклонять гипотезу H0 при уровне значимости 0,05.

ЗАКЛЮЧЕНИЕ

Ну вот какой-то такой анализ данных получился.