

Федеральное государственное образовательное бюджетное учреждение высшего образования

**Финансовый университет при Правительстве Российской Федерации**

Расчетно аналитическая работа

По дисциплине: **Анализ данных**

Выполнил:

Студент группы ПИ18-1

Факультета прикладной математики

и информационных технологий

**Гриднев Дмитрий Владимирович**,

Преподаватель:

к.э.н., доцент

**Пяткина Дарья Анатольевна**

## Введение

В данной работе производится анализ акций Магнит ао, Таттел ао, НЛМК ао.

Данные использовались недельные за период с 01.01.2015 по 31.12.2019

Данные о котировках акций на Московской бирже можно были взяты с сайта информационного агентства «МФД-ИнфоЦентр» (выбрана торговую площадку (например, «Мосбиржа Акции и ПИФы»)

Для анализа были использованы значения шести признаков:

· **TICKER** - название акции;

· **PER** - период (в нашем случае во всех строках это будет «W» - неделя);

· **DАТЕ** - дата;

· ТIМЕ - время;

· **CLOSE** - цена закрытия (т. е. цена, по которой была совершена последняя сделка в данном периоде);

· **VOL** - объем торговли.

Рассчитаем для каждой акций недельную доходность, недельную лог доходность, логарифм цены и логарифм объема

## Предварительный анализ данных о торгах по акциям трех эмитентов

По выбранным компаниям были составлены таблицы динамики цен, логарифма цен, логарифма доходности и логарифма объема.

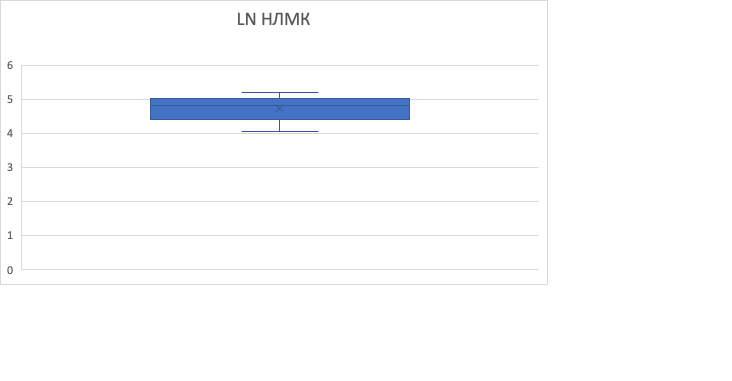
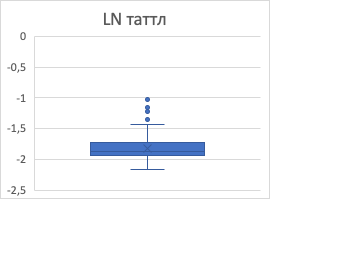
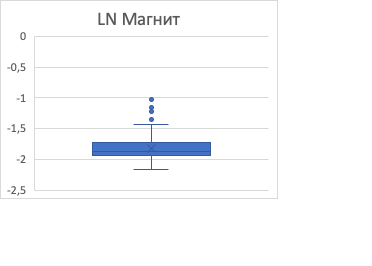
Были сформированы столбцы цен закрытия и объемов торгов, так чтобы для каждого тиккера и каждого параметра был персональный столбец, так же были сформированы столбцы новых параметров: логарифмов цен, доходностей, логарифмических доходностей (логдоходностей) и логарифмов объемов торгов.

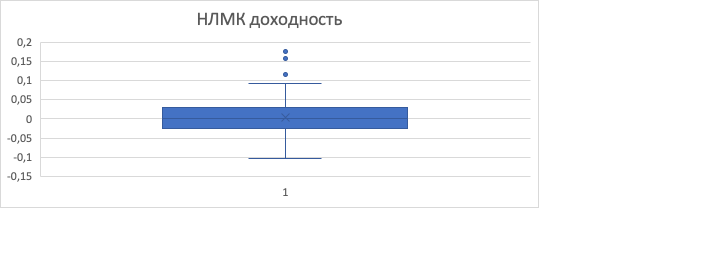
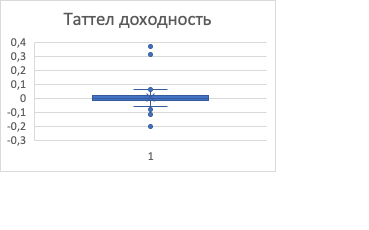
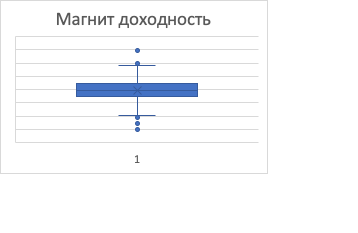
Поскольку значения параметра «Объем торгов» для различных дат и тиккеров могут отличаться на несколько порядков, целесообразно вычислить новый признак «Логарифм объёма». Этот новый признак пропорционален исходному, однако принимает значения одного порядка.

Теперь построим гистограммы для всех характеристик акций. Для единообразия каждую из гистограмм построим с 15 интервалами группирования

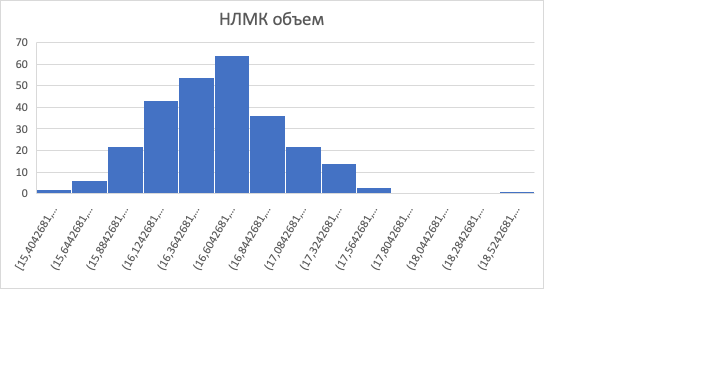
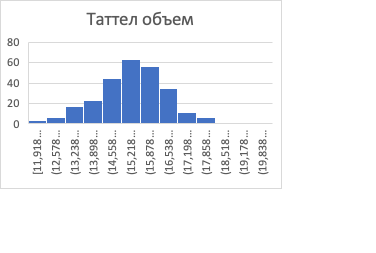
Проанализируем распределения характеристик по 3 акциям.

По акциям БСП ао распределение доходности и логарифма доходности очень близко к нормальному распределению (распределение имеет колоколообразную форму). Распределение логарифма цен и логарифма объемов распределено по закону распределения отличного от нормального. На диаграмме «ящик с усами» можно увидеть наличие сильных выбросов по всем основным показателям.





Точками на диаграмме «ящик с усами» обозначаются выбросы: значения доходности, которые выходят за интервал +-0,3 интерквартильных размах от 3 и 1 квартиля соответственно.



Предварительной обработки данных показала наличие умеренного числа выбросов по всем акциям и почти по всем показателям. Из графиков разброса логдоходностей видно, что по каждой акции наблюдается умеренное количество выбросов как в большую сторону от медианы, так и в меньшую сторону. Гистограммы показывают их умеренные колебания.

Проанализируем теперь данные о котировках акций российских компаний. Воспользуемся «Описательной статистикой» в Excel. Результаты вычисления статистических характеристик акций компактно представлены в таблице.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Среднее | Стандартная ошибка | Медиана | Мода | Стандартное отклонение | Дисперсия выборки |
| Доходность | |  |  |  |  |
| *Магнит ао доходность* | -0,00301 | 0,0025 | -0,00245 | 0 | 0,0416 | 0,0017 |
| *Таттел. Ао доходность* | 0,003627 | 0,0027 | 0 | 0 | 0,0439 | 0,0019 |
| *НЛМК ао доходность* | 0,003487 | 0,0024 | 0,00217 | #Н/Д | 0,04 | 0,0016 |
|  | Логдоходность | |  |  |  |  |
| *Магнит ао логдоходность* | -0,00389 | 0,0026 | -0,00245 | 0 | 0,0418 | 0,0017 |
| *Таттел. Ао логдоходность* | 0,002723 | 0,0026 | 0 | 0 | 0,042 | 0,0018 |
| *НЛМК Ао логдоходность* | 0,002697 | 0,0024 | 0,00217 | #Н/Д | 0,0396 | 0,0016 |
|  | Логарифм цеены | |  |  |  |  |
| *Магнит ао Логарифм цены* | 8,832965 | 0,0296 | 9,11273 | 9,297 | 0,484 | 0,2342 |
| *Таттел. Ао Логарифм цены* | -1,82272 | 0,0111 | -1,87405 | -1,966 | 0,1819 | 0,0331 |
| *НЛМК Логарифм цены* | 4,737062 | 0,0192 | 4,82044 | 4,29 | 0,314 | 0,0986 |
|  | Логарифм объема | |  |  |  |  |
| *Магнит ао Логарифм объема* | 13,69893 | 0,0394 | 13,6263 | #Н/Д | 0,6438 | 0,4145 |
| *Таттел. Ао Логарифм объема* | 15,57927 | 0,0743 | 15,6749 | 14,57 | 1,2137 | 1,473 |
| *НЛМК Ао Логарифм объема* | 16,63254 | 0,0267 | 16,6181 | #Н/Д | 0,4362 | 0,1902 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Эксцесс | Асимметричность | Интервал | Минимум | Максимум |
|  |  |  |  |  |
| 1,59751397 | 0,03729455 | 0,29752682 | -0,1484099 | 0,14911692 |
| 29,8622456 | 2,8477343 | 0,572529 | -0,2027972 | 0,3697318 |
| 1,47211548 | 0,41097184 | 0,27930528 | -0,1022742 | 0,17703109 |
|  |  |  |  |  |
| 1,60381157 | -0,1843776 | 0,29964372 | -0,16065 | 0,13899376 |
| 22,6196454 | 1,42260152 | 0,54126114 | -0,2266462 | 0,31461495 |
| 1,07309921 | 0,22700612 | 0,27088583 | -0,1078906 | 0,16299525 |
|  |  |  |  |  |
| -1,6541444 | -0,3331377 | 1,35986391 | 8,06526521 | 9,42512912 |
| 2,47844543 | 1,12739865 | 1,13420323 | -2,1628232 | -1,0286199 |
| -1,2648889 | -0,363643 | 1,1587935 | 4,05178495 | 5,21057845 |
|  |  |  |  |  |
| 0,04538504 | 0,53521876 | 3,48177963 | 12,2160131 | 15,6977927 |
| 0,6157079 | -0,1282192 | 8,09762916 | 11,9183906 | 20,0160197 |
| 1,15084613 | 0,42178191 | 3,2725011 | 15,4042681 | 18,6767692 |

У большинства признаков не определена мода. Это связано с тем, что значения доходностей очень редко повторяются.

Построим диаграммы рассеяния для анализа возможной взаимосвязи между показателями. Так как различных комбинаций пар признаков очень велико, построим диаграммы рассеяния по основным, где вероятно ожидается какая-то взаимосвязь. Построим диаграммы рассеяния по логарифму объема и логарифму цены по каждой паре акций (всего 3 диаграмм).

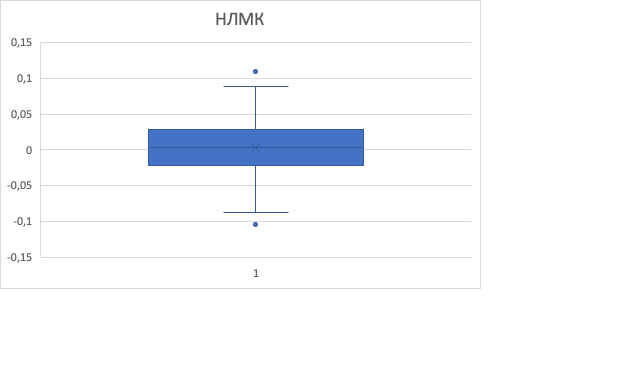
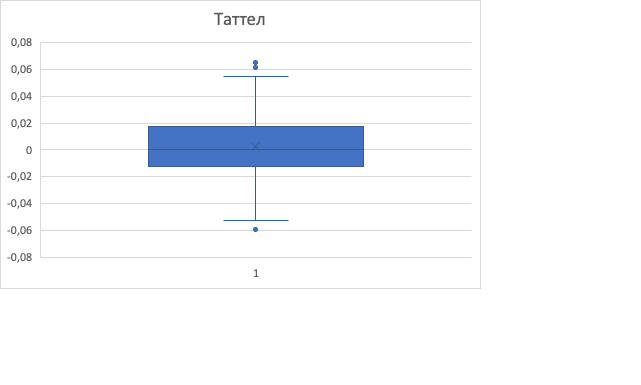
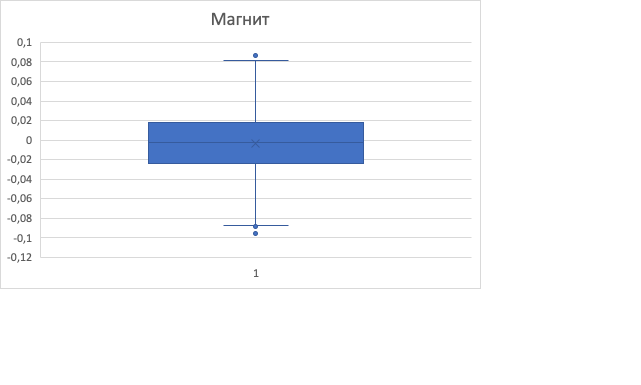
Так же рассчитаем коэффициенты корреляции

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Корреляционная матрица | | | | | | |
|  | *ц Магнит ао LN(ц)* | *ц Таттел. Ао LN(ц).* | *ц НЛМК LN(ц)* | *об Магнит ао LN(об)* | *об Таттел. Ао LN(об)* | *об НЛМК Ао LN(об)* |
| ц Магнит ао LN(ц) | 1 |  |  |  |  |  |
| ц Таттел. Ао LN(ц). | -0,7948914 | 1 |  |  |  |  |
| ц НЛМК LN(ц) | -0,8373646 | 0,6719325 | 1 |  |  |  |
| об Магнит ао LN(об) | -0,6729355 | 0,56825338 | 0,566718 | 1 |  |  |
| об Таттел. Ао LN(об) | -0,4586438 | 0,55522419 | 0,29569637 | 0,31694648 | 1 |  |
| об НЛМК Ао LN(об) | -0,2132344 | 0,30041638 | 0,15652211 | 0,26475918 | 0,17773219 | 1 |

Вычисление коэффициентов корреляции (вычисление через «Анализ данных» в Excel) и диаграммы рассеяния показывают умеренную положительную статистическую связь между логарифмами цен и логарифмами объема компаний у Магнита и Таттел.

В связи с тем, что в данных присутствуют не типичные данные (выбросы). Далее удалим их удалим их.

После удаления нетипичных, амплитуда колебаний уменьшилась незначительно, связно это с тем, что выбросов в исходной выборке было немного. С помощью формулы удалось удалить только сильно влияющие выбросы.



Изменились и корреляционные характеристики распределений, изменение не значительное и отражает такую же силу и направление связи как это было до удаления выбросов.

Поправленная гистограмма относительных частот и соответствующая ей эмпирическая функция распределения показывает незначительное изменение корреляций

Детальный анализ распределений доходностей показал, что утверждение об их логнормальности обладает достаточной надежностью.

С помощью «Анализа данных» в Excel были вычислены математические ожидания и дисперсии логдоходностей компаний, которые в дальнейшем будут использоваться для проверки теста на нормальность распределения

## Анализ логдоходностей эмитентов без выбросов

Пересчитаем статистические показатели трем акциям.

Найдем доверительные интервалы для средней доходности и стандартного отклонения

В результате получим

|  |  |
| --- | --- |
| *ц Магнит ао логдох.* | |
|  |  |
| Среднее | -0,003315202 |
| Стандартная ошибка | 0,002237568 |
| Медиана | -0,002311462 |
| Мода | 0 |
| Стандартное отклонение | 0,034808392 |
| Дисперсия выборки | 0,001211624 |
| Эксцесс | 0,074050465 |
| Асимметричность | -0,118259607 |
| Интервал | 0,182384756 |
| Минимум | -0,09599646 |
| Максимум | 0,086388296 |
| Сумма | -0,80227894 |
| Счет | 242 |
| Наибольший(1) | 0,086388296 |
| Наименьший(1) | -0,09599646 |
| Уровень надежности(95,0%) | 0,004407688 |

|  |  |
| --- | --- |
| *ц Таттел. Ао логдох.* | |
|  |  |
| Среднее | 0,002477074 |
| Стандартная ошибка | 0,001528552 |
| Медиана | 0 |
| Мода | 0 |
| Стандартное отклонение | 0,023778693 |
| Дисперсия выборки | 0,000565426 |
| Эксцесс | 0,004565802 |
| Асимметричность | 0,130748965 |
| Интервал | 0,12410113 |
| Минимум | -0,05942342 |
| Максимум | 0,06467771 |
| Сумма | 0,599451968 |
| Счет | 242 |
| Наибольший(1) | 0,06467771 |
| Наименьший(1) | -0,05942342 |
| Уровень надежности(95,0%) | 0,003011028 |

|  |  |
| --- | --- |
| *ц НЛМК Ао логдох.* | |
|  |  |
| Среднее | 0,002879477 |
| Стандартная ошибка | 0,002342986 |
| Медиана | 0,003331756 |
| Мода | 0 |
| Стандартное отклонение | 0,036448307 |
| Дисперсия выборки | 0,001328479 |
| Эксцесс | -0,188117009 |
| Асимметричность | -0,073808583 |
| Интервал | 0,212324412 |
| Минимум | -0,103582286 |
| Максимум | 0,108742126 |
| Сумма | 0,696833449 |
| Счет | 242 |
| Наибольший(1) | 0,108742126 |
| Наименьший(1) | -0,103582286 |
| Уровень надежности(95,0%) | 0,004615345 |

|  |  |  |  |
| --- | --- | --- | --- |
|  | ц Магнит ао логдох. | ц Таттел. Ао логдох. | ц НЛМК Ао логдох. |
| Точность инт. Оц. | 0,004407688 | 0,003011028 | 0,004615345 |
| Хи^2 левая | 200,8051831 | 200,8051831 | 200,8051831 |
| Хи^2 правая | 286,9811707 | 286,9811707 | 286,9811707 |
| (n-1)\*S^2 | 0,000353796 | 7,70494E-05 | 0,00042533 |
| Левая граница | 8,55028E-05 | 3,99014E-05 | 9,37491E-05 |
| Правая граница | 7,15223E-05 | 3,33772E-05 | 7,84202E-05 |

Построим доверительные интервал визуально

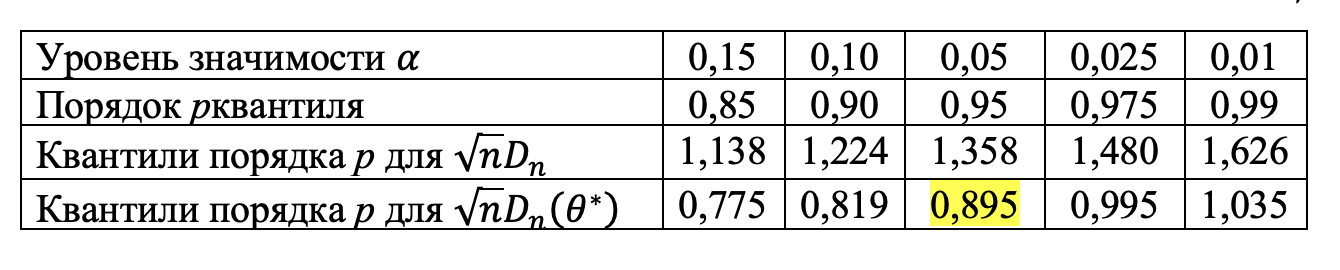
## Проверка гипотез нормального закона распределения логдоходностей

Проанализируем является распределение логдоходностей нормальной величиной

Разобьем доходность каждой акции на интервалы и проверим гипотезу о том, что распределение является нормальным

Кол-во интервалов вычисляется по формуле =ЦЕЛОЕ(1+КОРЕНЬ(2)\*LN(COUNT)), в нашем случае 9 интервалов.

Критическое значение критерия Колмогорова Смирнова при вероятности 5% будет равно 0,895



С помощью функции анализ данных => гистограмма, используя карманы построим график распределения частот и с помощью него произведем вычисления с помощью критериев Колмагорова Смирнова и Пирсона.

Критическое значение распределения критерия Пирсона при вероятности 5% равно 12,59.

Магнит **5,124075333**

Таттел **2,426831645**

НЛМК **4,815170781**

Значения всех тиккеров **меньше** критического значения критерия Пирсона, что подтверждает гипотезу о том, что логдоходности выбранных компаний распределены по нормальному закону распределения.

По критерию Колмогорова Смирнова

Магнит **7,60098323**

Таттел **3,921228514**

НЛМК **6,556808334**

Значения всех тиккеров **больше** критического значения критерия Колмогорова Смирнова, что подтверждает гипотезу о том, что логдоходности выбранных компаний распределены не по нормальному закону распределения.

Сравнение логдоходности тиккеров по разным критериям показало, что по данным акциям нельзя сделать однозначного вывода, так как Критерий Колмогорова Смирнова дал противоположный ответ Пирсону.

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

Воспользуемся Z-тестом Excel

|  |  |  |  |
| --- | --- | --- | --- |
|  | ц Магнит ао логдох. | ц Таттел. Ао логдох. | ц НЛМК Ао логдох. |
| n | 242 | 242 | 242 |
| X ср | -0,003315202 | 0,002477074 | 0,002879477 |
| S | 0,034808392 | 0,023778693 | 0,036448307 |
| |T| | 1,481609496 | 1,62053614 | 1,228977531 |
| t(кр 95%) | 1,969856213 | 1,969856213 | 1,969856213 |
| t(кр 99%) | 2,596382898 | 2,596382898 | 2,596382898 |
| альфа | 0,05 | 0,05 | 0,05 |
| альфа | 0,01 | 0,01 | 0,01 |
| Z-тест | 0,930777885 | 0,052558579 | 0,109540115 |

В трех случаях среднее значение меньше => принимаем для всех

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

|  |  |
| --- | --- |
| Гипотеза | ***Проверка о равенстве значений логдоходности за 2018 и 2019 год*** |
| H0: E(X) = E(Y) | *X = 2018 год* |
| H1: E(X) != E(Y) | *Y = 2019 год* |

Для проверки гипотезы воспользуемся СТЬЮДЕНТ.ТЕСТ-ом.

|  |  |  |  |
| --- | --- | --- | --- |
|  | ц Магнит ао логдох. | ц Таттел. Ао логдох. | ц НЛМК Ао логдох. |
| Nx | 48 | 48 | 48 |
| Ny | 50 | 50 | 50 |
| Sx | 0,001334803 | 0,000517814 | 0,00129725 |
| Sy | 0,000851526 | 0,000615999 | 0,001046506 |
| T | 1,776019454 | 1,459359645 | 0,52881735 |
| k - степени свободы | 89,86608649 | 95,80181696 | 93,94283328 |
| t (кр 95%) | 1,9869787 | 1,985251004 | 1,985801814 |
| t (кр 99%) | 2,632204191 | 2,628575671 | 2,629732145 |

|T| < t кр во всех случаях => принимаем теорию для трех компаний с вероятностью 99%.

## Сравнительный анализ логдоходностей трех компаний. Однофакторный дисперсионный анализ.

- сумма квадратов отклонений,

*nj* - число единиц в *j*-ой группе,

- среднее по j-ой группе,

среднее по всей совокупности.

- общая дисперсия.

*.*

, - однофакторная дисперсия

Для более простых вычислений воспользуемся функцией “Однофакторный дисперсионный анализ” из пакета анализа данных.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ИТОГИ |  |  |  |  |
| *Группы* | *Счет* | *Сумма* | *Среднее* | *Дисперсия* |
| 2015 | 47 | 0,05034634 | 0,0010712 | 0,00132553 |
| 2016 | 46 | 0,07764527 | 0,00168794 | 0,00116126 |
| 2017 | 48 | -0,3043308 | -0,0063402 | 0,0013993 |
| 2018 | 48 | -0,5918574 | -0,0123304 | 0,0013348 |
| 2019 | 50 | -0,0218903 | -0,0004378 | 0,00085153 |
| 2020 | 3 | -0,012192 | -0,004064 | 0,00086562 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Дисперсионный анализ | |  |  |  |  |  |
| *Источник вариации* | *SS* | *df* | *MS* | *F* | *P-Значение* | *F критическое* |
| Между группами | 0,00681175 | 5 | 0,00136235 | 1,12737109 | 0,34647965 | 2,25229073 |
| Внутри групп | 0,28518966 | 236 | 0,00120843 |  |  |  |
|  |  |  |  |  |  |  |
| Итого | 0,29200141 | 241 |  |  |  |  |

p-value = 0,34647965 > 0,05 => , что логдоходности не зависят от года. Исходя из этого можно сделать вывод о том, что логдоходности тиккеров независимы друг от друга и подходят для составления потфеля.

## Графики временных рядов логдоходностей трех компаний

Данные результаты достаточно сложно анализировать, поэтому для анализа динамики логарифма доходностей по всем акциям построим диаграммы.

Анализируя данные на данном графике, можно сделать вывод о схожести представленных зависимостей.

Для того, чтобы это подтвердить, постоим общую диаграмму, для каждого показателя создадим линию тренда и сравним их.

Как видно, данная диаграмма подтверждает верхний вывод.

## Исследование тесноты связи между логдоходностями

Для исследования тесноты связи между значениями логдоходностей акций компаний используем функцию «Корреляция» из пакета анализ данных.

|  |  |  |
| --- | --- | --- |
| **Гипотеза** | | |
| H0: r = 0 | | |
| H1: r != 0 | | |
| *n* | 242 | |
|  | | |  | | 1 | 2 | 3 |
|  | | |  | | *ц Магнит ао логдох.* | *ц Таттел. Ао логдох.* | *ц НЛМК Ао логдох.* |
| 1 | | | ц Магнит ао логдох. | | 1 |  |  |
| 2 | | | ц Таттел. Ао логдох. | | 0,143710681 | 1 |  |
| 3 | | | ц НЛМК Ао логдох. | | 0,208092985 | 0,043031688 | 1 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | r | t | tkp (5%) | tkp (1%) |
| 1 и 2 | 0,143710681 | 2,249708848 | 1,969897635 | 2,596469182 |
| 1 и 3 | 0,208092985 | 3,295913348 | 1,969897635 | 2,596469182 |
| 2 и 3 | 0,043031688 | 0,667262131 | 1,969897635 | 2,596469182 |

|  |  |
| --- | --- |
| **Вывод** | t1 = 2,249708848 > tkp => принимаем теорию H0 (95%) |
| t2 = 3,295913348 < tkp => принимаем теорию H1 |
| t3 = 0,175976303 < tkp => принимаем теорию H0 (99%) |

## Заключение

В ходе работы были проанализированы характеристики акций Магнит ао, Таттел ао, НЛМК ао. В целом можно сказать, что распределение доходностей симметричное.

Также были проведены тесты проверки изменения доходности по времени, и было подтверждено, что логарифм доходности акций во все периоды колеблется около нуля.

В целом анализ корреляционных свойств логдоходностей компаний показал их слабую, но при этом похожую друг на друга статистическую зависимость. Это можно заметить, рассматривая и графики временных рядов логдоходностей компаний

## Исследование тесноты связи между логдоходностями

1. Соловьев В. И. Анализ данных: теория вероятностей и прикладная статистика, обработка и визуализация данных в Microsoft Excel: 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)