

Отчет о проверке на заимствования №1



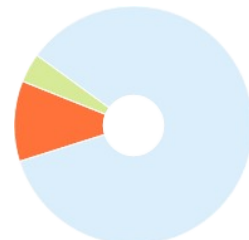
Автор: Башмакова А.
Проверяющий: Пользователь для API (galkinala@fa.ru / ID: 2588)
Организация: Финансовый университет при Правительстве РФ
Отчет предоставлен сервисом «Антиплагиат» - <http://fa.antiplagiat.ru>

ИНФОРМАЦИЯ О ДОКУМЕНТЕ

№ документа: 878356
Начало загрузки: 18.05.2021 12:53:03
Длительность загрузки: 00:00:15
Имя исходного файла: Башмакова АА (191805) [ПМ19-1].docx
Название документа: Башмакова АА (191805) [ПМ19-1].docx
Размер текста: 1 кБ
Символов в тексте: 39594
Слов в тексте: 5085
Число предложений: 295

ИНФОРМАЦИЯ ОБ ОТЧЕТЕ

Последний готовый отчет (ред.)
Начало проверки: 18.05.2021 12:53:20
Длительность проверки: 00:00:49
Комментарии: не указано
Поиск с учетом редактирования: да
Модули поиска: ИПС Адилет, Библиография, Сводная коллекция ЭБС, Интернет Плюс, Сводная коллекция РГБ, Цитирование, Переводные заимствования (RuEn), Переводные заимствования по eLIBRARY.RU (EnRu), Переводные заимствования по Интернету (EnRu), Переводные заимствования издательства Wiley (RuEn), eLIBRARY.RU, Модуль поиска "ФУ", СПС ГАРАНТ, Медицина, Диссертации НББ, Перефразирования по eLIBRARY.RU, Перефразирования по Интернету, Патенты СССР, РФ, СНГ, Шаблонные фразы, Кольцо вузов, Издательство Wiley



ЗАИМСТВОВАНИЯ

10,87%

САМОЦИТИРОВАНИЯ

0%

ЦИТИРОВАНИЯ

3,55%

ОРИГИНАЛЬНОСТЬ

85,58%

Заимствования — доля всех найденных текстовых пересечений, за исключением тех, которые система отнесла к цитированиям, по отношению к общему объему документа.
Самоцитирования — доля фрагментов текста проверяемого документа, совпадающий или почти совпадающий с фрагментом текста источника, автором или соавтором которого является автор проверяемого документа, по отношению к общему объему документа.
Цитирования — доля текстовых пересечений, которые не являются авторскими, но система посчитала их использование корректным, по отношению к общему объему документа. Сюда относятся оформленные по ГОСТу цитаты; общеупотребительные выражения; фрагменты текста, найденные в источниках из коллекций нормативно-правовой документации.
Текстовое пересечение — фрагмент текста проверяемого документа, совпадающий или почти совпадающий с фрагментом текста источника.
Источник — документ, проиндексированный в системе и содержащийся в модуле поиска, по которому проводится проверка.
Оригинальность — доля фрагментов текста проверяемого документа, не обнаруженных ни в одном источнике, по которым шла проверка, по отношению к общему объему документа.
Заимствования, самоцитирования, цитирования и оригинальность являются отдельными показателями и в сумме дают 100%, что соответствует всему тексту проверяемого документа.
Обращаем Ваше внимание, что система находит текстовые пересечения проверяемого документа с проиндексированными в системе текстовыми источниками. При этом система является вспомогательным инструментом, определение корректности и правомерности заимствований или цитирований, а также авторства текстовых фрагментов проверяемого документа остается в компетенции проверяющего.

№	Доля в отчете	Доля в тексте	Источник	Актуален на	Модуль поиска	Блоков в отчете	Блоков в тексте	Комментарии
[01]	3,01%	3,51%	Математика в экономике: учебник: В 3-х ч. Ч. 3. Теория вероятностей и математическая статистика http://studentlibrary.ru	27 Ноя 2017	Сводная коллекция ЭБС	4	5	
[02]	0%	3,51%	Математика в экономике: учебник: В 3-х ч. Ч. 3. Теория вероятностей и математическая статистика http://studentlibrary.ru	19 Дек 2016	Медицина	0	5	
[03]	2,72%	2,72%	не указано	раньше 2011	Библиография	1	1	
[04]	0,9%	1,99%	225779 http://biblioclub.ru	раньше 2011	Сводная коллекция ЭБС	2	3	
[05]	1,82%	1,82%	Обучение прыжку способом "перешагивания" детей 9-10 лет посредством применения нетрадиционного оборудования https://otherreferats.allbest.ru	14 Янв 2021	Интернет Плюс	1	1	
[06]	0%	1,82%	Обучение прыжку способом "перешагивания" детей 9-10 лет посредством применения нетрадиционного оборудования https://otherreferats.allbest.ru	23 Дек 2020	Интернет Плюс	0	1	
[07]	0%	1,82%	Обучение прыжку способом "перешагивания" детей 9-10 лет посредством применения нетрадиционного оборудования https://otherreferats.allbest.ru	17 Ноя 2020	Интернет Плюс	0	1	
[08]	1,22%	1,22%	не указано http://fa.ru	05 Янв 2017	Перефразирования по Интернету	3	3	
[09]	0,79%	1,03%	не указано http://docs.gsu.by	05 Янв 2017	Перефразирования по Интернету	3	3	
[10]	0,83%	0,95%	не указано	раньше 2011	Шаблонные фразы	8	9	

[11]	0,06%	0,93%	на тему: «Проверка гипотезы о нормальном распределении логарифмической доходности... http://kurs.znate.ru	05 Янв 2017	ПЕРЕФРАЗИРОВАНИЯ ПО ИНТЕРНЕТУ	1	2
[12]	0%	0,84%	Запишите общую формулу разложения в ряд по центральным эмпирическим моментам среднего арифметического функции от признака. Рассмотрите функцию - страница 5 http://userdocs.ru	10 Дек 2017	Интернет Плюс	0	5
[13]	0,55%	0,83%	Современный взгляд на проблемы экономики и менеджмента. http://elibrary.ru	раньше 2011	eLIBRARY.RU	6	8
[14]	0%	0,76%	Проверка гипотезы о независимости логарифмической доходности за различные интервалы времени при большом, среднем и малом объеме торгов - Готовые работы https://rectorate.ru	24 Апр 2021	Интернет Плюс	0	3
[15]	0%	0,76%	Проверка гипотезы о независимости логарифмической доходности за различные интервалы времени при большом, среднем и малом объеме торгов https://knowledge.allbest.ru	08 Мар 2021	Интернет Плюс	0	3
[16]	0,4%	0,75%	МАТЕМАТИЧЕСКАЯ СТАТИСТИКА. http://elibrary.ru	раньше 2011	ПЕРЕФРАЗИРОВАНИЯ ПО eLIBRARY.RU	1	2
[17]	0,28%	0,7%	Проверка гипотезы о независимости логарифмической доходности https://revolution.allbest.ru	05 Мая 2021	Интернет Плюс	2	3
[18]	0%	0,59%	Мирзоев М.М.	14 Окт 2019	Модуль поиска "ФУ"	0	1
[19]	0%	0,59%	михайлова полина - дтз	14 Окт 2019	Модуль поиска "ФУ"	0	1
[20]	0%	0,59%	резник анастасия - дтз	14 Окт 2019	Модуль поиска "ФУ"	0	1
[21]	0%	0,59%	петренко полина - дтз	14 Окт 2019	Модуль поиска "ФУ"	0	1
[22]	0%	0,59%	алейник владислава - дтз	14 Окт 2019	Модуль поиска "ФУ"	0	1
[23]	0%	0,59%	Bekbuzarova_U18-1_DTZ_vrednye_privychki	31 Окт 2019	Модуль поиска "ФУ"	0	1
[24]	0%	0,59%	Ненахова	02 Ноя 2019	Модуль поиска "ФУ"	0	1
[25]	0%	0,59%	Терехова Ксения. ДТЗ	06 Ноя 2019	Модуль поиска "ФУ"	0	1
[26]	0%	0,59%	РАР_Резник А.Е._У18-6	07 Июн 2020	Модуль поиска "ФУ"	0	1
[27]	0,58%	0,58%	Современный взгляд на проблемы экономики и менеджмента. http://elibrary.ru	раньше 2011	ПЕРЕФРАЗИРОВАНИЯ ПО eLIBRARY.RU	1	1
[28]	0%	0,45%	Сборник рабочих программ дисциплин по направлениям: 38.04.02 Менеджмент магистерские программы: «Корпоративное управление и ответственность бизнеса», «Образовательный менеджмент» и 38.04.04 Государственное и муниципальное управление магистерская программа https://book.ru	21 Янв 2020	Сводная коллекция ЭБС	0	1
[29]	0,44%	0,44%	chibirev_a_a_obrabotka-i-analiz-kratkovremennyh-bolshih-vybrosov-v-dlitelnyh-vremennyh-ryadah-v-prilozhenii-k-voln.docx	14 Окт 2019	Кольцо вузов	1	1
[30]	0%	0,42%	Бухгалтерский учет:современные вызовы, приоритеты и пути развития. Том 4 https://book.ru	03 Июл 2017	Сводная коллекция ЭБС	0	1
[31]	0%	0,42%	История филиала http://old.fa.ru	01 Дек 2020	Интернет Плюс	0	1
[32]	0%	0,42%	под ред. А. В. Касьянова, Е. В. Шестаковой Постатейный комментарий к Трудовому кодексу Российской Федерации Москва 2015 http://dlib.rsl.ru	05 Авг 2019	Сводная коллекция РГБ	0	1
[33]	0%	0,42%	ВКР	03 Июн 2013	Модуль поиска "ФУ"	0	1
[34]	0%	0,42%	Управление прибылью в акционерных обществах региона: теория и практика. Книга 1 http://studentlibrary.ru	19 Дек 2016	Медицина	0	1
[35]	0%	0,42%	Ресурс: Новое прочтение и геоэкономическое измерение экспортного потенциала http://studentlibrary.ru	19 Дек 2016	Медицина	0	1

[36]	<div>0%</div>	0,42%	Сборник официальных документов и материалов 4/2016 http://studentlibrary.ru	20 Дек 2016	Медицина	0	1	
[37]	<div>0%</div>	0,41%	Влияние смысловой установки на распознавание эмоций	29 Сен 2020	Кольцо вузов	0	1	
[38]	<div>0%</div>	0,4%	http://vital.lib.tsu.ru/vital/access/services/Download/vtls:000631745/SOURCE1 http://vital.lib.tsu.ru	24 Янв 2020	Интернет Плюс	0	2	
[39]	<div>0%</div>	0,37%	Андрухова, Ольга Витальевна Экономическое развитие нефтесервисных компаний в России : диссертация ... кандидата экономических наук : 08.00.05 Апатиты 2020 http://dlib.rsl.ru	12 Янв 2021	Сводная коллекция РГБ	0	1	
[40]	<div>0%</div>	0,34%	Теоретическая часть, Статистическая проверка гипотез, Статистический критерий Фроцини - Проверка гипотезы о нормальном распределении логарифмической доходности по критерию Фроцини https://studbooks.net	13 Апр 2021	Интернет Плюс	0	2	
[41]	<div>0,18%</div>	0,33%	Коллоидный журнал. 2017. Т. 79, № 4 http://biblioclub.ru	21 Янв 2020	Сводная коллекция ЭБС	1	1	
[42]	<div>0,29%</div>	0,29%	PREFACE http://ben.klemens.org	06 Янв 2018	Переводные заимствования (RuEn)	1	1	
[43]	<div>0,28%</div>	0,28%	200148 http://biblioclub.ru	18 Апр 2016	Сводная коллекция ЭБС	1	1	
[44]	<div>0%</div>	0,27%	Теория вероятностей и математическая статистика http://fa.ru	раньше 2011	Интернет Плюс	0	1	
[45]	<div>0%</div>	0,27%	Проверка гипотезы о равенстве дисперсий логарифмической доходности фондового индекса и входящих в его состав акций., Программирование - Курсовая работа https://referatbooks.ru	05 Мая 2021	Интернет Плюс	0	1	
[46]	<div>0%</div>	0,24%	Формирование креативной компетентности будущих бакалавров-учителей в процессе обучения математике на основе специального комплекса заданий http://research.sfu-kras.ru	05 Мая 2021	Интернет Плюс	0	1	
[47]	<div>0,07%</div>	0,22%	klykov_v_a_procedurey-identifikacii-setevyh-struktur-osnovannye-na-koefficientah-korrelyacii-kendalla-i-spirmena.docx	30 Мая 2020	Кольцо вузов	1	1	
[48]	<div>0%</div>	0,2%	Запишите общую формулу разложения в ряд по центральным эмпирическим моментам среднего арифметического функции от признака. Рассмотрите функцию - страница 4 http://userdocs.ru	10 Дек 2017	Интернет Плюс	0	1	
[49]	<div>0%</div>	0,2%	Для проверки каких гипотез применяется критерий Колмогорова? Каким образом находится значение статистики данного критерия? https://studfile.net	23 Апр 2021	Интернет Плюс	0	1	
[50]	<div>0%</div>	0,17%	Мельников, Сергей Владимирович Повышение точности позиционирования беспилотных летательных аппаратов в условиях искажения или подавления навигационного поля GPS/Глонасс : диссертация ... кандидата технических наук : 05.13.01 Ставрополь 2020 http://dlib.rsl.ru	22 Окт 2020	Сводная коллекция РГБ	0	1	Источник исключен. Причина: Маленький процент пересечения.
[51]	<div>0%</div>	0,15%	Method and tools for fault detection in smart high-performance buildings;Méthode et outils pour l'identification de défauts des bâtiments connectés performants https://tel.archives-ouvertes.fr	04 Мая 2020	Интернет Плюс	0	1	Источник исключен. Причина: Маленький процент пересечения.
[52]	<div>0%</div>	0,15%	Фильтр Калмана для минимизации энтропийного значения случайной погрешности с не Гауссовым распределением / Хабр https://habr.com	10 Апр 2019	Интернет Плюс	0	1	Источник исключен. Причина: Маленький процент пересечения.
[53]	<div>0%</div>	0,15%	Фильтр Калмана. Первый взгляд / СоХабр https://sohabr.net	23 Апр 2021	Интернет Плюс	0	1	Источник исключен. Причина: Маленький процент пересечения.
[54]	<div>0%</div>	0,15%	Фильтр Калмана. Первый взгляд / СоХабр https://sohabr.net	22 Апр 2021	Интернет Плюс	0	1	Источник исключен. Причина: Маленький процент пересечения.
[55]	<div>0%</div>	0,15%	Фильтр Калмана для минимизации энтропийного значения случайной погрешности с не Гауссовым	22 Апр 2021	Интернет Плюс	0	1	Источник исключен. Причина: Маленький процент

			распределением / Хабр https://habr.com					пересечения.
[56]	0%	0,14%	Традиционное АВ-тестирование. Продолжение и резюме https://craftappmobile.com	07 Мая 2021	Интернет Плюс	0	1	Источник исключен. Причина: Маленький процент пересечения.
[57]	0%	0,14%	Байесовский вывод — интуиция и примеры - NOP::Nuances of Programming - Medium https://medium.com	28 Апр 2021	Интернет Плюс	0	1	Источник исключен. Причина: Маленький процент пересечения.
[58]	0%	0,11%	pdf http://vixra.org	29 Ноя 2018	Интернет Плюс	0	1	Источник исключен. Причина: Маленький процент пересечения.
[59]	0%	0,11%	Visualizing the distribution of a dataset — seaborn 0.10.0 documentation http://seaborn.pydata.org	23 Фев 2020	Интернет Плюс	0	1	Источник исключен. Причина: Маленький процент пересечения.
[60]	0%	0,11%	Visualizing the distribution of a dataset — seaborn 0.10.1 documentation https://seaborn.pydata.org	25 Апр 2021	Интернет Плюс	0	1	Источник исключен. Причина: Маленький процент пересечения.
[61]	0%	0,11%	Фильтр Калмана — ProfitTraders.com http://profittraders.com	04 Мая 2021	Интернет Плюс	0	1	Источник исключен. Причина: Маленький процент пересечения.

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

10

Департамент анализа данных, принятия решений и финансовых технологий

41

Курсовая работа

на тему

**«Проверка гипотезы о равенстве дисперсий логарифмической доходности
индекса фондового рынка и входящих в его состав акций»**

Вид данных для исследования:

8

**«Котировки акций компаний, входящих в индекс ММВБ потребительского
сектора»**

Выполнила:

студентка группы ПМ19-1,

Башмакова А. А.

Научный руководитель:

доцент, к.э.н.

Гринева Н. В.

Москва
2021

Содержание

<u>ВВЕДЕНИЕ</u>	3
<u>I. ТЕОРЕТИЧЕСКАЯ СПРАВКА</u>	5
1. <u>Математическая статистика</u>	5
2. <u>Статистическая гипотеза</u>	5
3. <u>Ошибки первого и второго рода</u>	6
4. <u>P-значения</u>	7
5. <u>Логарифмическая доходность</u>	8
6. <u>Критерий Колмогорова</u>	9
7. <u>Критерий Фишера</u>	11
<u>II. ПРАКТИЧЕСКАЯ ЧАСТЬ</u>	13
1. <u>Предварительный анализ данных</u>	13
2. <u>Проверка гипотезы на модельных данных</u>	17
3. <u>Проверка гипотезы на реальных данных</u>	18
4. <u>Альтернативные гипотезы и оценка мощности критерия</u>	21
<u>ЗАКЛЮЧЕНИЕ</u>	22
<u>СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ</u>	23
<u>ПРИЛОЖЕНИЯ</u>	24
<u>Приложение 1</u>	24
<u>Приложение 2</u>	25
<u>Приложение 3</u>	25

ВВЕДЕНИЕ

В этой работе я проверю гипотезу о равенстве дисперсий логарифмической доходности фондового рынка и входящих в его состав акций. Для этого буду использовать критерий Фишера как относительно простой, но находящий широкое распространение в дисперсионном анализе.

Для исследования и проверки гипотезы я выбрала данные котировок акций тех компаний, которые входят в индекс ММВБ Московской биржи и относятся к потребительскому сектору – МОЕХСН. Рассматривала я временной промежуток за последние 10 лет, то есть с 1 января 2011 года по 30 декабря 2021 года [9].

Индекс ММВБ представляет собой ценовый композитный фондовый индекс. Он считается одним из основных показателей экономики поскольку взвешен по рыночной капитализации. Он включает в себя акции крупнейших российских эмитентов, которые являются наиболее ликвидными и активно развивающимися и относятся к различным сферам жизни: сельское хозяйство, финансы, транспорт, промышленность, торговля, медицина.

Индекс МОЕХСН (ММВБ потребительский сектор) на Московской бирже включает в себя компании, которые предоставляют различные продукты жизнедеятельности потребителям, например, обувь, продукты питания, игрушки, технику, алкоголь, а также ведут сельскохозяйственную деятельность. На апрель 2020 года в индекс входят 12 компаний: X5 Retail Group (компания, которая управляет продуктовыми магазинами «Пятёрочка», «Перекрёсток», «Чижик», «Карусель»), ПАО «Детский мир» (сеть магазинов товаров для детей), ПАО «Магнит» (сеть магазинов под маркой «Магнит»: магазин «у дома», Супермаркет «Магнит Семейный», суперстор «Магнит Экстра», «Магнит Косметик», «Магнит Аптека»), ПАО «РУСАГРО» (самая крупная аграрная компания в России), ПАО "ЛЕНТА" (сеть продовольственных гипермаркетов первая по величине в России), МГП

«Мать и Дитя» (сеть частных медицинских центров в области акушерства, гинекологии и педиатрии), ПАО «М.Видео» (сеть магазинов торговли бытовой техникой и электроникой), ПАО «Белуга Групп» (компания по производству алкогольных напитков), ПАО «Аптечная сеть 36,6» (сеть магазинов по продаже лекарственных средств), ПАО "Русская Аквакультура" (компания, которая занимается производством и продажей красной рыбы), ПАО «среллоС» (компания, занимающаяся автомобилестроением), OR GROUP (торговая платформа, которая объединяет обувную сеть под брендом Westfalika.

Планируемая новизна заключается в проверке данной гипотезы с использованием языка программирования Python версии 3.0 и интерактивной оболочки для него Jupiter Notebook, благодаря которой возможно хранить и обрабатывать большой объём данных с использованием изображений, графиков, гистограмм, формул.

Актуальность работы заключается в использовании современного инструмента обработки данных, а именно языка программирования Python. Кроме того, сфера потребительских услуг никогда не потеряет свою актуальность что в 2021 году, что в 2031.

По результатам работы можно будет сделать вывод о взаимосвязи колебания логарифмической доходности индекса фондового рынка и колебания логарифмической доходности акций, входящих в его состав, значит, и их биржевых котировок.

I. ТЕОРЕТИЧЕСКАЯ СПРАВКА

1. Математическая статистика

Часть единой дисциплины прикладной математики - теорией вероятностей и математической статистикой представляющая из себя совокупность или же семейство дисциплин, таких как экономическая статистика, финансовая статистика, социальная статистика и так далее, основной задачей которой является обеспечение конкретных статистических дисциплин теоретическим фундаментом, называется математической статистикой.

Целью мат. статистики является получение прикладных и фундаментальных заключений путем создания единой методики по сбору и обработке [7].

2. Статистическая гипотеза

Статистической гипотезой можно назвать всякое утверждение, содержащее информацию о виде или параметрах внутреннего распределения, об отношениях между случайными величинами и тому подобное. При этом её можно считать параметрической, если она базируется на предположении, заключающемся в том, что нам известно генеральное распределение, причём с точностью до конечного числа параметров.

Пусть H_0 и H_1 – это две статистические гипотезы, которые взаимоисключают друг друга, при этом назовём гипотезу H_0 основной, а H_1 – дополнительной. В дальнейшем принимаем в качестве базисного предположения утверждение о том, что одна из гипотез является справедливой.

Так как H_0 и H_1 – взаимоисключающие гипотезы, то принятие H_0 автоматически означает отступление от гипотезы и наоборот отступление от гипотезы H_0 означает принятие гипотезы H_1 .

Правило, по которому, если гипотеза H_0 не принимается, а выборка x_1, \dots, x_n принадлежит некоторой области K , и наоборот если гипотеза H_0 принимается, а выборка x_1, \dots, x_n не принадлежит некоторой области K , называется статистическим критерием или статистикой критерия. Область значений критерия, на которой принимается гипотеза H_0 , называют областью допустимых значений или же областью принятия гипотезы. Соответственно, область значений критерия, на которой отклоняется гипотеза H_0 , называют критической областью.

Как правило критическую область можно задать, используя неравенства:

$$K = \{x_1, \dots, x_n \in R_n : t > c\} \quad (2.1)$$

или

$$K = \{x_1, \dots, x_n \in R_n : t < c\} \quad (2.2)$$

или

$$K = \{x_1, \dots, x_n \in R_n : t < c_1 \cup x_1, \dots, x_n \in R_n : t > c_2\}, \quad (2.3)$$

где c, c_1, c_2 ($c_2 > c_1$) = const, $t = t(x_1, \dots, x_n)$ - статистика критерия [9].

3. Ошибки первого и второго рода

Во время применения статистики критерия возможно возникновение ошибок двух различных типов. Случай, когда отвергается гипотеза H_0 , являющаяся верной, называют ошибкой первого рода. Напротив, ошибка второго рода – ситуация, при которой отвергается гипотеза H_1 , являющаяся верной.

Уровень значимости критерия α - вероятность ошибки первого рода или значимость критерия. За мощность критерия обозначают $1-\beta$, где β – вероятность ошибки второго рода.

Статистическая гипотеза проверяется путем сравнения наблюдаемого значения критерия с критическим значением, связанным с данным уровнем значимости, что позволяет отклонить или принять основную гипотезу. При этом в тех случаях, когда уровень значимости будет другим, то придётся вновь вычислять соответствующее критическое значение [1].

4. Р-значения

P_value позволяет принять или отвергнуть нулевую гипотезу сразу для всех уровней значимости без необходимости вычислять критические значения.

Р-значением статистического критерия для фиксированной реализации x случайной выборки $x = (x_1, \dots, x_n)$ называется такое число $PV(x)$, что $PV(x) \geq \alpha$ для любого уровня значимости α , при котором гипотеза H_0 принимается, и $PV(x) \leq \alpha$, для любого уровня значимости α , при котором гипотеза H_0 отвергается.

Если представить, что P-value найдено или заранее известно, то принятие (или отклонение) H_0 для данного α реализуется так: в случае, гипотеза H_0 отклоняется при $PV(x) < \alpha$, и принимается при $PV(x) \geq \alpha$.

Рассмотрим отдельно случай, когда $PV(x) = \alpha$. Как правило, критическую область можно представить в виде

$$K_\alpha = \{x : t_x > c_\alpha\}, \quad (4.1)$$

где c_α – непрерывная убывающая функция. Как нетрудно видеть, в этом случае

$$PV(x) = c(tx-1) \quad (4.2)$$

и для $\alpha = PV(x)$ имеет место равенство $t_x = c(\alpha)$ означающее, что H_0 принимается. Отсюда уже легко получить широко применяемую формулу:

$$PV(x) = PNO \quad (t_x > t_x) \quad (4.3)$$

Действительно, при любом уровне значимости α

$$PNO(t_x > c(\alpha) = \alpha. \quad (4.4)$$

$$\text{Аналогично} \quad K\alpha = tx < c\alpha, \quad (4.5)$$

где $c(\alpha)$ – непрерывная возрастающая функция, Р-значение удовлетворяет отношению

$$PV(x) = PNO \quad (t_x < t_x) \quad (4.6) [1]$$

5. Логарифмическая доходность

Логарифмическую доходность или данные об изменении цен довольно часто используют для анализа данных. Данный показатель используется вместо процентной доходности и имеет вид:

$$\ln P_t P_{t-1} \quad (5.1)$$

Или же, применив свойство логарифма от частного, получим:

$$\ln P_t - \ln P_{t-1} \quad (5.2)$$

Где t – рассматриваемый период, P_t – цена акции в данный период, P_{t-1} – цена акции за прошлый период.

Стоит отметить ряд достоинств логарифмической доходности. Во-первых, она позволяет без особых проблем объединять доходы при более низких частотах выплат, что возможно только за счёт обобщения доходов при более высоких частотах выплат, а значит, ежемесячная доходность будет равна сумме ежедневных логарифмических доходностей, что делает расчёты проще, нежели с использованием процентной доходности. Также логарифмическая доходность позволяет найти начальную цену, если произошло её повышение на $x\%$, а затем снижение на такую же величину $x\%$ [3].

6. Критерий Колмогорова

Также в данной работе будет использоваться вспомогательный критерий Колмогорова. Этот критерий позволяет подтвердить или опровергнуть гипотезу о распределении величины по какому-либо закону.

За статистику критерия Колмогорова принято считать меру расхождения между значением эмпирической функции распределения $F_n(x)$ и предполагаемой теоретической функцией распределения $F(x)$. Среди всех полученных значений выбирают максимальное по модулю

Обозначим как $m_{x,x}$ то количество составляющих вектора $x = x_1, \dots, x_n$, которые меньше x для любого $x \in \mathbb{R}_n$. Если мы имеем дело со случайным вектором $x = (x_1, \dots, x_n)$, то обозначение $m_{x,x}$ аналогично, хотя стоит отметить, что оно будет являться дискретной случайной величиной, которая может принимать целые значения от 0 до n . Обозначим за x реализацию случайной выборки X , имеющую объём n , из некоторого распределения с функцией $F(x)$, и тогда эмпирическая функция распределения, которая соответствует данной выборке будет иметь вид:

$$F = F(x, x) = m(x, x)/n. \quad (6.1)$$

Для оценки функции $F(x)$ по случайной выборке X формула аналогична и выглядит следующим образом:

$$F = F(x, X) = m(x, X)/n. \quad (6.2)$$

Отметим, что $F(x, x)$ является числовой функцией, в то время как $F(x, X)$ является случайным процессом, поскольку в каждой точке x принимает случайное значение.

Расстояние между функциями $F(x)$ и $F(x)$ можем определить, используя формулу:

$$d = \sup_x |F(x) - F(x)|. \quad (6.3)$$

Аналогично обратим внимание на то, что в случае с функцией функции $F = F(x, x)$ расстояние $d = d(x)$ будет являться числом, но для $F = F(x, X)$ расстояние $d = d(x)$ уже случайная величина а значит принимает значения от 0 до 1 включительно.

Опираясь на теоремы Колмогорова в случае непрерывной функции $F(x)$ при любом неотрицательном $u \geq 0$ существует предел

$$\lim_{n \rightarrow \infty} P(n d(X) < u) = K(u), \quad (6.4)$$

где

$$K(u) = \sum_{k=-\infty}^{\infty} (-1)^k e^{-2k^2 u^2} \quad (6.5)$$

А также на основе данной теоремы критерий согласия с критической областью $n d(x) > u_\alpha$, где u_α - корень уравнения $K(u) = 1 - \alpha$, где уровень значимости стремится к α при $n \rightarrow \infty$, то есть α является асимптотическим

уровнем значимости. Этот критерий и является критерием Колмогорова, и применяется при $n > 20$, так как в противном случае фактический уровень значимости будет сильно отличаться от номинального значения.

Чтобы определить максимальное абсолютное отклонение гипотетической функции $F(x)$ от эмпирической функции $F(x)$ применяется следующая формула:

$$d(x) = \max_{1 \leq i \leq n} |F(x_i) - F(x)|, \quad (6.6)$$

где x_i - i -й член вариационного ряда

$$x_1 \leq x_2 \leq \dots \leq x_{n-1} \leq x_n \quad [1].$$

Таблица. Критические значения статистики критерия Колмогорова-Смирнова

α	0,15	0,10	0,05	0,03	0,01
$D_{n\alpha}$	0,775	0,819	0,895	0,995	1,035

7. Критерий Фишера

Когда нужно сравнить две дисперсии нормальных распределений используют критерий Фишера. Принято считать, что все доходности финансовых активов подчиняются нормальному распределению по центральной предельной теореме: если количество дней, за которое рассчитывается логарифм кумулятивной доходности, является большим, то распределение суммы логарифмов стремится к нормальному. Но мы убедимся в этом по критерию Колмогорова.

Пусть имеются две независимые выборки из нормальных распределений:

$$x_1, \dots, x_m \sim N(\mu_x, \sigma_x^2),$$

$$y_1, \dots, y_n \sim N(\mu_y, \sigma_y^2).$$

Будем считать, что параметры μ_x , σ_x^2 , μ_y , σ_y^2 известны. В качестве основной гипотезы примем $H_0: \sigma_x^2 = \sigma_y^2$, а в качестве дополнительной одну из трёх гипотез:

- 1) $H_1: \sigma_x^2 > \sigma_y^2$;
- 2) $H_1: \sigma_x^2 < \sigma_y^2$;
- 3) $H_1: \sigma_x^2 \neq \sigma_y^2$.

Теорема о построении критериев для проверки гипотезы с известным уравнением значимости α :

Если верна H_0 , то

$$s_x^2/s_y^2 \sim F_{m-1, n-1} ,$$

где $s_x^2 = \frac{1}{m} \sum_{i=1}^m (X_i - \bar{X})^2$, $s_y^2 = \frac{1}{n} \sum_{i=1}^n (Y_i - \bar{Y})^2$, а $F(m-1, n-1)$ – распределение Фишера с $m-1$ и $n-1$ степенями свободы [2].

II. ПРАКТИЧЕСКАЯ ЧАСТЬ

1. Предварительный анализ данных

Начнём работу с предварительного анализа данных. Котировки компаний, используемые в работе, были взяты с сайта Межбанковского финансового дома [11]. Информацию об индексе и входящих в его состав акциях [10], а распределении акций по секторам [9]. Сопоставим котировки акций и название компаний в таблице 1.

Таблица 1 - Список компаний и тикеров котировок акций

Тикер	Название компании
AGRO	ПАО «РУСАГРО»
PJPq	X5 Retail Group
AQUA	ПАО «Русская аквакультура»
DSKY	ПАО «Детский мир»
MGNT	ПАО «Магнит»
LNTAq	ПАО "ЛЕНТА"
MDMGDR	МГП «Мать и Дитя»
MVID	ПАО «М.Видео»
BELU	ПАО «Белуга Групп»
APTK	ПАО «Аптечная сеть 36,6»
SVAV	ПАО «среллоС»
ORUP	OR GROUP

Для расчёта количества торговых дней использую столбец <DATE>, в котором указана информация о датах котировок. Она представлена в формате ууММдд. Для вычисления торговых дней используется группировка данных по полю <DATE> с использованием информации о годе, после чего осуществляется подсчёт торговых дней в получившихся группах. Информация выводится в таблицу, где строки – это названия компаний, а столбцы – года, и сохраняется в файл формата csv с разделителем «;».

Полученные результаты количества торговых дней для всех компаний можем наблюдать в таблице 2.

Таблица 2 - Вычисленное количество торговых дней для всех компаний

	A	B	C	D	E	F	G	H
1		Ticker	2015	2016	2017	2018	2019	2020
2	0	AGRO	229	252	252	254	252	250
3	1	PJPq	0	0	0	234	252	250
4	2	AQUA	151	231	228	223	251	250
5	3	DSKY	0	0	224	254	252	250
6	4	MGNT	250	252	252	254	252	250
7	5	LNTAq	250	252	252	254	252	250
8	6	MDMGDR	0	0	0	0	0	38
9	7	MVID	250	252	252	254	252	250
10	8	BELU	0	0	104	222	252	250
11	9	APTK	250	252	252	254	252	250
12	10	SVAV	250	252	252	254	252	250
13	11	ORUP	0	0	50	252	252	250

Отметим, что у котировок PJPq данные с 2018 года, MDMGDR с 2020, у DSKY и BELU с 2017, у ORUP с конца 2017, а у AQUA с середины 2015, поэтому их анализировать дальше не будем. Составим новую таблицу дней, где будут только те компании, в которых есть торговые дни в период с 201 по 2020 года. Полученные результаты представлены в таблице 3.

Таблица 3 - Вычисленное количество торговых дней для сокращенного списка компаний (скорректированной таблицы)

	A	B	C	D	E	F	G	H
1		Ticker	2015	2016	2017	2018	2019	2020
2	0	AGRO	229	252	252	254	252	250
3	1	MGNT	250	252	252	254	252	250
4	2	LNTAq	250	252	252	254	252	250
5	3	MVID	250	252	252	254	252	250
6	4	APTK	250	252	252	254	252	250
7	5	SVAV	250	252	252	254	252	250

Оставшиеся 6 компании имеют более 229 торговых дней в каждом из рассматриваемых годов, что обеспечивает достаточное количество информации для дальнейшего анализа и обработки.

Далее рассмотрим максимальные относительные скачки цен вверх и вниз. Для этого рассмотрим столбец <CLOSE>, в котором содержится цена, сформированная на момент закрытия акции. В таблице 4 мы можем наблюдать полученный результат для максимальных скачков цен вниз, а в таблице 5 – для максимальных скачков вверх. Для наглядности применим условное форматирование в MS Excel по цветам (зеленый – максимальное значение, красный – минимальное).

Таблица 4 - Максимальные относительные скачки цен вниз

	A	B	C	D	E	F	G	H
1		Ticker	2015	2016	2017	2018	2019	2020
2	0	AGRO	-0,112	-0,051	-0,047	-0,055	-0,032	-0,066
3	1	MGNT	-0,095	-0,073	-0,112	-0,102	-0,035	-0,136
4	2	LNTAq	-0,075	-0,055	-0,043	-0,077	-0,035	-0,134
5	3	MVID	-0,118	-0,077	-0,057	-0,055	-0,069	-0,064
6	4	APTK	-0,193	-0,104	-0,088	-0,065	-0,071	-0,109
7	5	SVAV	-0,064	-0,059	-0,056	-0,064	-0,047	-0,074

Таблица 5 - Максимальные относительные скачки цен вверх

	A	B	C	D	E	F	G	H
1		Ticker	2015	2016	2017	2018	2019	2020
2	0	AGRO	0,209	0,055	0,051	0,058	0,043	0,104
3	1	MGNT	0,082	0,078	0,059	0,079	0,047	0,07
4	2	LNTAq	0,1	0,042	0,051	0,081	0,043	0,088
5	3	MVID	0,095	0,077	0,111	0,054	0,065	0,122
6	4	APTK	0,556	0,189	0,129	0,124	0,455	0,162
7	5	SVAV	0,167	0,124	0,077	0,049	0,04	0,092

Максимальный скачок вниз у АРТК (19, 3%) и у этой же компании самый сильный скачок вверх (55,6%), из чего можно сделать вывод, что ПАО «Аптечная сеть 36,6» обладает максимальными относительными изменениями цен. Построим график изменения цен для этой компании. Результаты отображены на рисунке 1.

Рисунок 1. График изменения цен для компании АРТК

Так как скачок цены больше, чем 50%, значит данная компания не дает нам репрезентативную выборку, и мы исключаем её из дальнейшего исследования.

По итогам предварительного анализа данных мы будем рассматривать 5 компаний, которые имеют более 229 торговых дней в году и не имеют большого скачка цен, что предоставляет нам репрезентативную выборку, которую мы и будем исследовать.

2. Проверка гипотезы на модельных данных

Проверим гипотезу о равенстве дисперсий логарифмической доходности фондового индекса и входящих в его состав акций для модельных данных, сгенерированных случайным образом. Используя метод Монте-Карло, сгенерируем две выборки из 252 элементов с помощью критерия Колмогорова, чтобы проверить гипотезу на модельных данных, что позволит

убедиться в корректной работе программы, вычисляющей значение критерия и р-значения.

Также построим гистограмму Р-значений для смоделированных. Она представлена на рисунке 2 и 3.



Рисунок 2. Гистограмма Р-значения для модельных данных 1

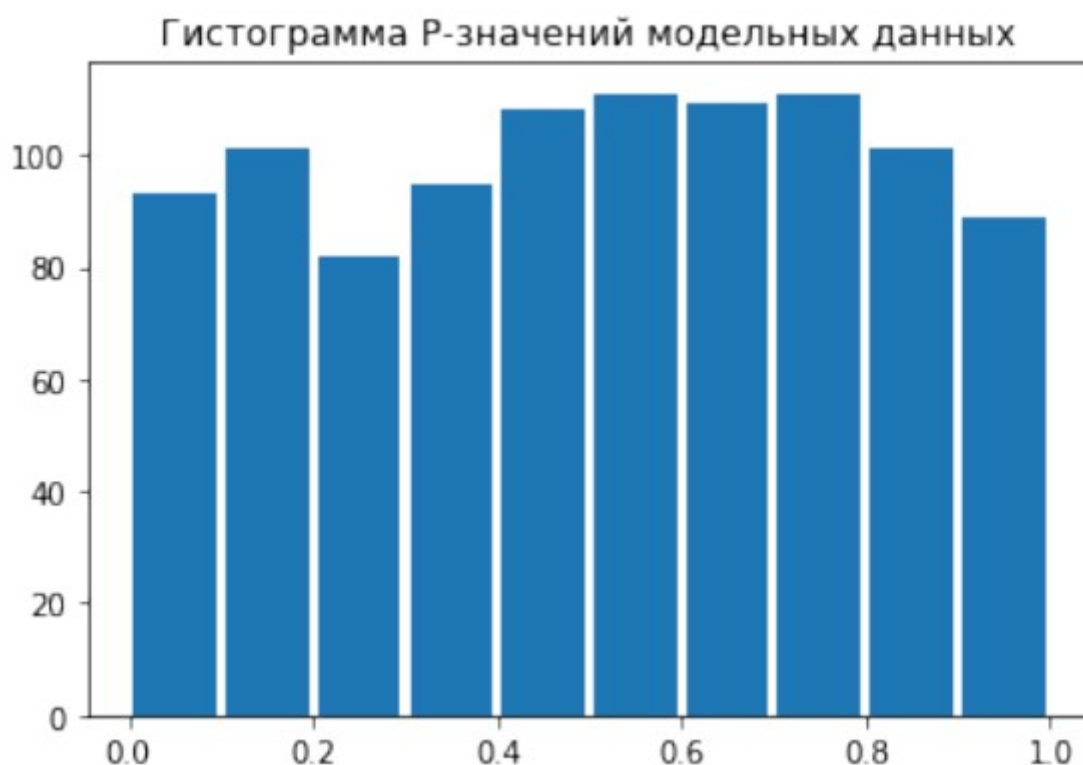


Рисунок 3. Гистограмма Р-значения для модельных данных 2

Получили Р-значение критерия Колмогорова равным 0.7810093429954852 для первой выборки и 0.9155674384900221 для второй выборки, что говорит о равномерности распределений Р-значений для модельных данных [8].

Проверим Критерий Фишера на модельных данных. Критерий Фишера равен 1.2487322294814869, значит дисперсии равны и можно сделать вывод, что гипотеза о равенстве дисперсий логарифмической доходности фондового рынка и входящих в его состав акций принимается.

3. Проверка гипотезы на реальных данных

Убедившись в том, что все необходимые нам программы работают корректно и обработка данных происходит корректно, проверим выполнение гипотезы о равенстве дисперсий логарифмической доходности фондового рынка и входящих в его состав акций по критерию Фишера на реальных данных, а именно на котировках акций компаний, которых входят в индекс

ММВБ потребительский сектор. Как и на этапе предварительной обработке данных воспользуемся полями «<DATE>», показывающее дату информации о котировках, и «<CLOSE>», показывающее цену последней сделки, которая была совершена. Отметим, что критерий работает с логарифмической доходностью, которой в исходных данных нет, соответственно вычислим её и сохраним значение в списке. Для этого задана специальная функция.

После чего вычислим Р-значения уже для реальных данных, что тоже делается через заранее прописанную функцию, для выборки за все 5 лет по каждому году, с помощью Критерия Колмогорова. После чего выведем полученный результат, которая представлен в таблице 6.

Таблица 6. Р-значений для реальных данных за 5 лет по каждому году

	Тикер	2015	2016	2017	2018	2019	2020
0	AGRO	0.0	0.0	0.0	0.0	0.0	0.0
1	MGNT	0.0	0.0	0.0	0.0	0.0	0.0
2	LNTAq	0.0	0.0	0.0	0.0	0.0	0.0
3	MVID	0.0	0.0	0.0	0.0	0.0	0.0
4	APTK	0.0	0.0	0.0	0.0	0.0	0.0
5	SVAV	0.0	0.0	0.0	0.0	0.0	0.0
6	MOEXCN	0.0	0.0	0.0	0.0	0.0	0.0

Отразим полученные значения на гистограмме.

Гистограмма Р-значений модельных данных 1, вычисленных с помощью критерия Колмогорова

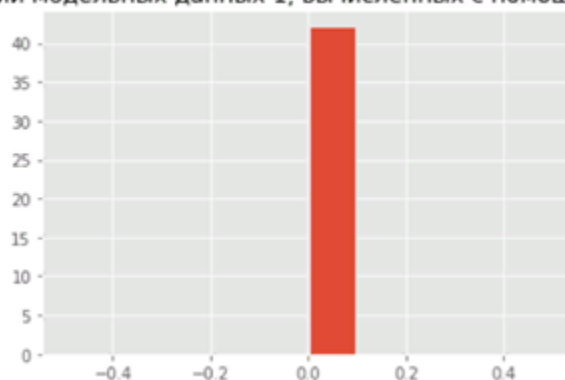


Рисунок 4. Гистограмма Р-значения.

Так как значения Р-значения сосредотачиваются вокруг нуля, то мы не можем сказать, что реальные данные распределены равномерно.

Вычислим значения дисперсий для каждой компании по каждому году и получим следующие значения.

Таблица 7. Дисперсии логарифмической доходности за 5 лет по каждому году

	Тикер	2015	2016	2017	2018	2019	2020
0	AGRO	0.055917	0.055917	0.055917	0.055917	0.055917	0.055917
1	MGNT	0.234778	0.234778	0.234778	0.234778	0.234778	0.234778
2	LNTAq	0.136766	0.136766	0.136766	0.136766	0.136766	0.136766
3	MVID	0.111443	0.111443	0.111443	0.111443	0.111443	0.111443
4	SVAV	0.078697	0.078697	0.078697	0.078697	0.078697	0.078697
5	MOEXCN	0.008455	0.008455	0.008455	0.008455	0.008455	0.008455

В дисперсионном анализе не применяется прямое сравнение дисперсий, поэтому переходим к анализу значений логарифмической доходности, но уже с применением рассматриваемого критерия Фишера. Для этого воспользуемся специально написанной для этого функцией для компаний и индекса MOEXFN и получим значения, которые представлены в таблице 8.

Таблица 8. Проверка критерия Фишера

	A	B	C
1	AGRO	43.73868939867883	
2	MGNT	771.065195439747	
3	LNTAq	261.6566468132922	
4	MVID	173.73253262154708	
5	SVAV	86.6344458183792	

У логарифмической доходности акций MGNT было наибольшее значение дисперсии, для неё мы получили наибольшее значение критерия Фишера. У логарифмической доходности акций AGRO было наименьшее значение дисперсии, для неё мы получили наименьшее значение критерия Фишера.

Для нормально распределённой выборки на модельных данных мы получили значение Фишера. По реальным данным видно, что эти значения сильно отличаются от нормально распределённой.

Значит, можем сделать вывод, что гипотеза о равенстве дисперсий логарифмической доходности индекса фондового рынка и входящих в его состав акций отвергается.

4. Альтернативные гипотезы и оценка мощности критерия

В качестве альтернативных гипотез было предположено, что логарифмическая доходность имеет:

1. распределение по модулю закона Стьюдента с тремя степенями свободы;
2. логнормальное распределение.

Вычислим мощность критерия Фишера, вычисляя 1000 раз Р-значения при уровне значимости $\alpha=0.05$ для каждого альтернативного распределения.

Результаты можно видеть в таблице 15.

Таблица 9. Мощность критерия Фишера для различных распределений.

Вид распределения	20	125	252
Распределение по модулю закона Стьюдента с тремя степенями свободы	0.6406	0.9874	0.9998
Логнормальное распределение со стандартным отклонением равным 1/4	0.6603	0.9893	0.9989

Из таблицы следует, что мощность критерия Фишера для логнормального распределения и для распределения Стьюдента с тремя степенями свободы увеличивается с увеличением объёма выборки. Значит,

чем больше выборка, тем меньше вероятность допустить ошибку второго рода.

ЗАКЛЮЧЕНИЕ

10

В данной курсовой работе была рассмотрена проверка гипотезы о равенстве дисперсий логарифмической доходности индекса фондового рынка и входящих в его состав акций по критерию Фишера на примере котировок акций компаний, входящих в индекс MOEXCN (потребительского сектора), а также рассмотрена правильность выполнения программ как на модельных данных, так и на реальных.

В результате проверки поставленной гипотезы о равенстве дисперсий логарифмической доходности индекса фондового рынка и входящих в его состав акций я получила следующие результаты: гипотеза не подтвердилась для выбранных мной реальных данных, но зато подтвердилась для сгенерированных случайным образом модельных данных.

Из полученных результатов следует лишь одно: выбранная гипотеза не верна для реальных данных, а значит дисперсии логарифмической доходности индекса фондового рынка и входящих в его состав акций не равны. На подобный результат могло повлиять множество факторов. Одним из них можно назвать и то, что у каждого из критериев существуют недостатки. Так, например, критерий Фишера крайне чувствителен к отклонению от нормального распределения и достигает максимальной эффективности для выборок одинакового объёма, что крайне редко можно наблюдать в современных реалиях, а не на моделях.

Но сам факт того, что на модельных данных, сформированных случайным образом из диапазона чисел $[0;1]$, гипотеза подтверждается показывает то, что в вычислениях нет ошибок, а значит и критерии были проверены правильно.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. А.В. Браилов, А.С. Солодовников; под ред. В.А. Бабайцева, В.Б. Гисина Математика в экономике: учебник: В 3-х ч. Ч. 3. Теория вероятностей и математическая статистика – М.: Финансы и статистика, 2013
2. Браилов А.В. Лекции по математической статистике. – М.: Финакадемия, 2008.
3. Браилов А.В. Лекции по теории вероятности. – М.: Финакадемия, 2008.
4. Красс М. С., Б.П. Чупрынов Б.П. Математика в экономике - М.: Финансы и статистика, 2007.
5. С.С. Бондарчук, И.С. Бондарчук Статобработка экспериментальных данных в MS Excel: учебное пособие. Издательство Томского государственного педагогического университета, 2018.
6. Глебов Криволапов Практикум по математической статистике. Проверка гипотез с использованием Excel, MatCale, R и Python. М.: Прометей, 2019.
7. Малугин В.А. Математическая статистика. М.: Юрайт, 2020.
8. В.Я. Крохалев, С.А. Скопинов, В.А. Телешев СТАТИСТИКА, Издательство УГМУ 2018
9. Смирнова З.М., Крейнина М.В. ПРОВЕРКА СТАТИСТИЧЕСКИХ ГИПОТЕЗ.
10. https://smart-lab.ru/q/index_stocks/MOEXCN/
11. <https://ru.investing.com/equities/>
12. <https://mfd.ru/>

ПРИЛОЖЕНИЯ

Приложение 1

Технические характеристики компьютера:

Процессор Intel Core i5-8250U CPU @

Тактовая частота 1.60 GHz 1.80 GHz

Частота системной шины 4 GT/s QPI

Объём кэша второго уровня 1,0 Мб

Время выполнения программы:

--- 26.84362006187439 seconds ---

Список файлов

Имя файла
Таблица 1 – Список компаний и тикеров котировок акций
Таблица 2 - Вычисленное количество торговых дней для всех компаний
Таблица 3 - Вычисленное количество торговых дней для сокращенного списка компаний
Таблица 4 - Максимальные относительные скачки цен вниз 13
Таблица 5 - Максимальные относительные скачки цен вверх
Рисунок 1. График изменения цен для компании АРТК 17
Рисунок 2. Гистограмма Р-значений для модельных данных 1 17
Рисунок 3. Гистограмма Р-значений для модельных данных 2 13
Таблица 6. Р-значений для реальных данных за 5 лет по каждому году
Рисунок 4. Гистограмма Р-значения.
Таблица 7. Дисперсии логарифмической доходности за 5 лет по каждому году
Таблица 8. Проверка критерия Фишера
Таблица 9. Мощность критерия Фишера для различных распределений.

```

import time

start_time = time.time() # время начала

#загружаем необходимые библиотеки

import numpy as np

import pandas as pd

import statistics as st

import scipy.stats as stats

import matplotlib.pyplot as plt
29
from scipy.stats import norm

import csv
47
from math import *

from scipy.stats import chisquare

import matplotlib

import math

```

Таблица 2 - Вычисленное количество торговых дней для всех компаний

```

#посчитаем количество торговых дней для всех компаний

tickers=['AGRO','PJPq', 'AQUA', 'DSKY', 'MGNT', 'LNTAq', 'MDMGDR', 'MVID', 'BELU', 'APTK', 'SVAV',
'ORUP']

kolvo_dney_vse=pd.DataFrame()

kolvo_dney_vse['Ticker'] = tickers

years=range(2015,2021)

for i in years:

    kolvo_year=[]

    for j in range(len(tickers)):

        file=pd.read_csv(tickers[j]+''.csv',sep=';')

        diap=(file['<DATE>']>=i*10000)&(file['<DATE>']<(i+1)*10000)

        kolvo=len(file[diap])

        kolvo_year.append(kolvo)

```

```

kolvo_dney_vse[str(i)]=kolvo_year

kolvo_dney_vse.to_csv('Количество торговых дней всех компаний по годам.csv',sep=';')

kolvo_dney_vse

```

Таблица 3 - Вычисленное количество торговых дней для сокращенного списка компаний

```

#посчитаем количество торговых дней компаний для скорректированной таблицы

tickers=['AGRO', 'MGNT', 'LNTAq', 'MVID', 'APTK', 'SVAV']

kolvo_dney_kor=pd.DataFrame()

kolvo_dney_kor['Ticker'] = tickers

years=range(2015,2021)

for i in years:

    kolvo_year=[]

    for j in range(len(tickers)):

        file=pd.read_csv(tickers[j]+'.csv',sep=';')

        diap=(file['<DATE>']>=i*10000)&(file['<DATE>']<(i+1)*10000)

        kolvo=len(file[diap])

        kolvo_year.append(kolvo)

    kolvo_dney_kor[str(i)]=kolvo_year

kolvo_dney_kor.to_csv('Количество торговых дней компаний для скорректированной
таблицы.csv',sep=';')

kolvo_dney_kor

```

Таблица 4 - Максимальные относительные скачки цен вниз

```

#посчитаем относительные скачки цен вниз

tickers=['AGRO', 'MGNT', 'LNTAq', 'MVID', 'APTK', 'SVAV']

down=pd.DataFrame()

down['Ticker'] = tickers

years=range(2015,2021)

for i in years:

    down_znach=[]

```



```

for j in range(len(tickers)):

    file=pd.read_csv(tickers[j]+''.csv',sep=';')

    file['<LEAP>'] = file['<CLOSE>'].pct_change().round(3)

        diap=(file['<DATE>']>=i*10000)&(file['<DATE>']<(i+1)*10000)

        znach_year=file[diap]['<LEAP>'].min()

        down_znach.append(znach_year)

    down[str(i)]=down_znach

down.to_csv('Относительные изменения цен вниз.csv',sep=';', decimal=',')

down

```

Таблица 5 - Максимальные относительные скачки цен вверх

```

#посчитаем относительные скачки цен вверх

tickers=['AGRO', 'MGNT', 'LNTAq', 'MVID', 'APTK', 'SVAV']

up=pd.DataFrame()

up['Ticker'] = tickers

years=range(2015,2021)

for i in years:

    up_znach=[]

    for j in range(len(tickers)):

        file=pd.read_csv(tickers[j]+''.csv',sep=';')

        file['<LEAP>'] = file['<CLOSE>'].pct_change().round(3)

            diap=(file['<DATE>']>=i*10000)&(file['<DATE>']<(i+1)*10000)

            znach_year=file[diap]['<LEAP>'].max()

            up_znach.append(znach_year)

    up[str(i)]=up_znach

up.to_csv('Относительные изменения цен вверх.csv',sep=';', decimal=',')

up

#проверка гипотезы на модельных данных

kvanlil_1 = [kv for kv in np.arange(0.1,1,0.1)]

```

```

znach_1=[]

for i in kvanlil_1:

    znach_1.append(np.quantile(Shapir1,i))

df_1 = pd.DataFrame({'Квантиль':kvanlil_1, 'Значение':znach_1})

df_1.to_csv('Первые 9 квантилей статистики Шапиро-Уилка1.csv',sep=';')

df_1

kvanlil_2 = [kv for kv in np.arange(0.1,1,0.1)]

znach_2=[]

for i in kvanlil_2:

    znach_2.append(np.quantile(Shapir2,i))

df_2 = pd.DataFrame({'Квантиль':kvanlil_2, 'Значение':znach_2})

df_2.to_csv('Первые 9 квантилей статистики Шапиро-Уилка2.csv',sep=';')

df_2

```

Таблица 6 - Таблица квантилей 1

```

from scipy.stats import kstest

kvanlil_9991 = [kv for kv in np.arange(0.001,1,0.001)]

znach_9991 = []

for i in kvanlil_9991:

    znach_9991.append(np.quantile(Shapir1,i))

print(kstest(znach_9991,'norm'))

df_9991 = pd.DataFrame({'Квантиль':kvanlil_9991, 'Значение':znach_9991})

df_9991.to_csv('999 квантилей статистики Шапиро-Уилка1.csv')

df_9991

```

Рисунок 2. Гистограмма Р-значений для модельных данных 1

```

pvalues=[]

for i in range(1000):

    x=np.random.normal(loc=0, scale=1, size=252)

    pvalues.append(stats.shapiro(x)[1])

```

```
plt.hist(pvalues,rwidth=0.9)

plt.title('Гистограмма Р-значений модельных данных')

plt.show()
```

Таблица 7 - Таблица квантилей 2

```
from scipy.stats import kstest

kvanlil_9992 = [kv for kv in np.arange(0.001,1,0.001)]

znach_9992 = []

for i in kvanlil_9992:

    znach_9992.append(np.quantile(Shapir2,i))

print(kstest(znach_9992,'norm'))

df_9992 = pd.DataFrame({'Квантиль':kvanlil_9992, 'Значение':znach_9992})

df_9992.to_csv('999 квантилей статистики Шапиро-Уилка2.csv')

df_9992
```

Рисунок 3. Гистограмма Р-значений для модельных данных 2

```
pvalues1=[]

for i in range(1000):

    x=np.random.normal(loc=0, scale=1, size=252)

    pvalues1.append(stats.shapiro(x)[1])

plt.hist(pvalues1,rwidth=0.9)

plt.title('Гистограмма Р-значений модельных данных')

plt.show()

#выполним проверку с помощью критерия Колмогорова

def p_kolm(n):

    viborka=np.random.normal(loc=0, scale=1, size=n)

    p=stats.kstest(viborka,'norm')

    return round(p[1],3)

#построение гистограммы Р-значений, вычисленных с помощью критерия Колмогорова

from scipy.stats import kstest
```

```

pvalues_kolm1=[]

for i in range(1000):

    pvalues_kolm1.append(p_kolm(n))

print(kstest(pvalues_kolm1,'norm'))

plt.hist(pvalues_kolm1,rwidth=0.9)

plt.title('Гистограмма Р-значений модельных данных 1, вычисленных с помощью критерия
Колмогорова')

plt.show()

#построение гистограммы Р-значений, вычисленных с помощью критерия Колмогорова

from scipy.stats import kstest

pvalues_kolm2=[]

for i in range(1000):

    pvalues_kolm2.append(p_kolm(n))

print(kstest(pvalues_kolm2,'norm'))

plt.hist(pvalues_kolm2,rwidth=0.9)

plt.title('Гистограмма Р-значений модельных данных 1, вычисленных с помощью критерия
Колмогорова')

plt.show()

# проверим, что наши выборки взяты из распределения одного вида

viborka_1=np.random.normal(loc=0, scale=1, size=n)

viborka_2=np.random.normal(loc=0, scale=1, size=n)

pvalue1 = stats.ks_2samp(viborka_1,viborka_2)

print(pvalue1[1])

#проверка критерия Фишера

def fisher_criterion(v1, v2):

    return (abs(np.mean(v1) - np.mean(v2)))/ (np.var(v1) + np.var(v2))

fisher_criterion (znach_9991, znach_9992)

#проверка мощности критерия

m = 10000 #количество экспериментов

```

```

f1=[]

f2=[]

for i in range(m):

    viborka11 = stats.t.rvs(df=3, loc = 0, scale = 1,size = 252)

    viborka12 = stats.t.rvs(df=3, loc = 0, scale = 1,size = 252)

    viborka21 = stats.lognorm.rvs(s=1/4 ,loc = 0, scale = 1,size = 252)

    viborka22 = stats.lognorm.rvs(s=1/4 ,loc = 0, scale = 1,size = 252)

    f1.append(fisher_criterion(viborka11, viborka12))

    f2.append(fisher_criterion(viborka21, viborka22))

pow1 = 0

pow2 = 0

```

Таблица 11. Мощность критерия Фишера для различных распределений.

Вычисление мощности критерия

```

for i in f1:

    if i<0.05:

        pow1+=1

print(pow1/m)

for i in f2:

    if i<0.05:

        pow2+=1

print(pow2/m)

#проверка гипотезы на реальных данных

#импорт таблиц

tickers=['AGRO', 'MGNT', 'LNTAq', 'MVID', 'SVAV', 'MOEXCN']

def tables(ticker):

    company = pd.read_csv(ticker+'.csv', sep=';') #чтение из файла

    company['<DATE>'] = pd.to_datetime(company['<DATE>'], dayfirst=True, infer_datetime_format=True)

#преобразование даты

```

```

return company

# список логарифмов доходности по годам
def logYeld(ticker, a):

    LD = []

    ya=['2015', '2016', '2017', '2018', '2019', '2020']

    company = tables(ticker)

    company['Логарифм доходности'] = np.log(company['<CLOSE>'])

    for year in ya:

        table = pd.read_csv('Количество торговых дней компаний для скорректированной
таблицы.csv', sep=';')

        n = (table[str(year)])[a])

        z = company['Логарифм доходности'][company['<DATE>']<=str(year)+'-12-31'].size-(n+1)

        LD.append((company['Логарифм доходности'][company['<DATE>']<=str(year)+'-12-
31'])[z:].kurtosis())

    print (LD)

    return LD

matplotlib.style.use('ggplot')

PV = []

for ticker in tickers:

    a = tickers.index(ticker)

    log = logYeld(ticker, a)

13
Таблица 8. Р-значений для реальных данных за 5 лет по каждому году

def log_prof_max(year,file): #создаю функцию для подсчета логарифмической доходности

    company = pd.read_csv(file, sep=';') #считываю файл

    #обрезаю дату до четырех цифр года

    company['<DATE>'] = [int(str(company['<DATE>'])[i])[4:] for i in range(len(company['<DATE>']))]

    number = list(np.where(company['<DATE>'] == year)[0])[:-1]

    #ищу логарифмическую доходность

    #округляю до двух знаков после запятой

```

```

    logg_doch_max = np.log(company['<CLOSE>'])

    return stats.kstest(logg_doch_max, 'norm')[1]

#далее записываю в массив названия тикеров компаний для формирования столбца DataFrame и
считывания файлов

tickers=['AGRO', 'MGNT', 'LNTAq', 'MVID', 'APTK', 'SVAV', 'MOEXCN']

#для формирования названий столбцов DataFrame создаю переменную, хранящую массив всех
годов с 2015 по 2020

years = range(2015,2021)

#создаю DataFrame

table_log_max = pd.DataFrame() #добавляю номер тикера с 1 для наглядности

#записываю в столбец названия тикеров

table_log_max['Тикер'] = tickers

#высчитываю максимальное отклонение цен по каждому году

all_val=[]

for year in years:

    log_number_max = [] #создаю пустой массив для заполнения максимального отклонения по
каждой акции

    for ticker in tickers:

        #добавляю в массив значение логарифмических доходностей

        #по каждому году исследуемой акции из файла рассчитываю функции и добавляю в массив

        log_number_max.append(round(log_prof_max(year, ticker+'.csv'), 10))

        all_val.append(round(log_prof_max(year, ticker+'.csv'), 10))

#записываю в столбец по году значения массива log_number_max, который на каждом шаге цикла

#хранит логарифмическую доходность в определенный год по каждой акции

    table_log_max[str(year)] = log_number_max

#текст вывожу жирным

print('\033[1m' + 'Таблица. Р-значения по критерию Колмогорова')

table_log_max

plt.hist(all_val,rwidth=0.9)

```

```
plt.title('Гистограмма P-значений модельных данных 1, вычисленных с помощью критерия Колмогорова')
```

```
plt.show()
```

Таблица 9. Дисперсии логарифмической доходности за 5 лет по каждому году

```
def log_prof_max(year,file): #создаю функцию для подсчета логарифмической доходности
```

```
    company = pd.read_csv(file, sep=';') #считываю файл
```

```
#обрезаю дату до четырех цифр года
```

```
    company['<DATE>'] = [int(str(company['<DATE>'])[i]):4] for i in range(len(company['<DATE>']))]
```

```
    number = list(np.where(company['<DATE>'] == year)[0])[:-1]
```

```
#ищу логарифмическую доходность (т.е логарифм отношения цены закрытия текущей акции к предыдущей)
```

```
#округляю до двух знаков после запятой
```

```
    logg_doch_max = np.log(company['<CLOSE>'])
```

```
    return np.var(logg_doch_max)
```

```
#далее записываю в массив названия тикеров компаний для формирования столбца DataFrame и считывания файлов
```

```
tickers=['AGRO', 'MGNT', 'LNTAq', 'MVID', 'SVAV', 'MOEXCN']
```

```
#для формирования названий столбцов DataFrame создаю переменную, хранящую массив всех годов с 2015 по 2020
```

```
years = range(2015,2021)
```

```
#создаю DataFrame
```

```
table_log_max = pd.DataFrame() #добавляю номер тикера с 1 для наглядности
```

```
#записываю в столбец названия тикеров
```

```
table_log_max['Тикер'] = tickers
```

```
#высчитываю максимальное отклонение цен по каждому году
```

```
for year in years:
```

```
    log_number_max = [] #создаю пустой массив для заполнения максимального отклонения по каждой акции
```

```
    for ticker in tickers:
```

```
#добавляю в массив значение логарифмических доходностей
```



```

#по каждому году исследуемой акции из файла рассчитываю функции и добавляю в массив
    log_number_max.append(log_prof_max(year, ticker+'.csv'))

#записываю в столбец по году значения массива log_number_max, который на каждом шаге цикла
#хранит максимальную логарифмическую доходность в определенный год по каждой акции

    table_log_max[str(year)] = log_number_max

#текст вывожу жирным
print('\033[1m' + 'Таблица. Дисперсии логарифмической доходности')

table_log_max

Таблица 10. Проверка критерия Фишера

def log_prof_max(file): #создаю функцию для подсчета логарифмической доходности

    company = pd.read_csv(file, sep=';') #считываю файл

#обрезаю дату до четырех цифр года

    company['<DATE>'] = [int(str(company['<DATE>'])[i][:4]) for i in range(len(company['<DATE>']))]

#ищу логарифмическую доходность (т.е логарифм отношения цены закрытия текущей акции к
предыдущей)

#округляю до двух знаков после запятой

    logg_doch_max = np.log(company['<CLOSE>'])

    return (logg_doch_max)

#далее записываю в массив названия тикеров компаний для формирования столбца DataFrame и
считывания файлов

tickers=['AGRO', 'MGNT', 'LNTAq', 'MVID', 'SVAV', 'MOEXCN']

#для формирования названий столбцов DataFrame создаю переменную, хранящую массив всех
годов с 2015 по 2020

#создаю DataFrame

table_log_max = pd.DataFrame() #добавляю номер тикера с 1 для наглядности

#записываю в столбец названия тикеров

table_log_max['Тикер'] = tickers

#высчитываю максимальное отклонение цен по каждому году

log_number_max = [] #создаю пустой массив для заполнения максимального отклонения по
каждой акции

```

```

for ticker in tickers:

#добавляю в массив значение логарифмических доходностей

#по каждому году исследуемой акции из файла рассчитываю функции и добавляю в массив

    log_number_max.append(log_prof_max(ticker+'.csv'))

#записываю в столбец по году значения массива log_number_max, который на каждом шаге цикла

#хранит максимальную логарифмическую доходность в определенный год по каждой акции

    table_log_max = log_number_max

#текст вывожу жирным

print("\033[1m' + 'Таблица. Логарифмическая доходность:')

table_log_max

for i in table_log_max[:5]:

    y = fisher_criterion(i, table_log_max[5])

    print(y)

finish_time = time.time()

print("--- %s seconds ---" % (finish_time - start_time)) # Время выполнения программы в секундах

```