

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

*Департамент «Анализ данных и машинного обучения»*  
*Факультет «Информационных технологий и анализа больших данных»*

**Курсовая работа**

*По дисциплине «Теория вероятностей и математическая статистика»*

*на тему:*

**«Проверка гипотезы о нормальном распределении дневной  
логарифмической доходности при условии определенного объема торгов  
накануне»**

*Вид исследуемых данных:*

**NYSE U.S. 100 INDEX и NYSE HEALTH CARE SECTOR INDEX**

Выполнил:

студент гр. ПМ19-1

**Шигаров Д.А.**

Научный руководитель:

к.э.н., доцент

*Иванюк В.А.* **Иванюк В. А.**

Москва – 2021

## План работы

1.	Введение.....	3
1.1	Подробное разъяснение темы .....	3
1.2	Описание выборок.....	3
2.	Предварительный анализ данных .....	3
3.	Теоритическая справка по проверке гипотез.....	8
3.1	Статистическая проверка гипотез.....	8
3.2	Критерий Шапиро-Уилка .....	9
3.3	Критерий Колмогорова .....	9
4.	Проверка гипотезы для модельных данных.....	10
5.	Проверка гипотезы для реальных данных .....	11
6.	Заключение .....	19
7.	Список литературы .....	20
8.	Приложения .....	21
8.1	Приложение .....	21
8.2	Приложение .....	21

## **1. Введение**

### **1.1 Подробное разьяснение темы**

Целью данной курсовой работы является проверка гипотезы о нормальном распределении дневной логарифмической доходности при условии определенного объема торгов накануне.

Для этого будет использоваться закон о нормальном распределении логарифмической доходности. Логарифмы хорошо подходят для изучения доходности и поэтому часто используются. Логарифмическая доходность проста и наглядна, имеет существенные преимущества по сравнению с процентной доходностью. Для исследования выборки будет использоваться нормальный закон распределения. Этот закон удобен и часто используется для анализа различных моделей.

### **1.2 Описание выборок**

В данной курсовой работе по математической статистике используется объединение индексов NYSE U.S. 100 INDEX (\$NY.ID) и NYSE HEALTH CARE SECTOR INDEX (\$NYPTR). Список компаний, входящих в данные индексы, наглядно показывает лидеров в сфере здравоохранения. Тикеры этих компаний: PFE, BMY, MRK, JNJ, WBA, ABT, LLY, DHR, MMM, UNH.

В данной работе будут рассмотрены данные за период в 4 года, с 01.01.2017 по 31.12.2020. Гипотеза о нормальном распределении дневной логарифмической доходности будет проверяться 100 раз.

## **2. Предварительный анализ данных**

Для исследования будут использоваться котировки компаний, составляющих индексы индексов NYSE U.S. 100 INDEX (\$NY.ID) и NYSE HEALTH CARE SECTOR INDEX (\$NYPTR).

NYSE (New York Stock Exchange) является крупнейшей фондовой биржей в мире. Капитализация акций торгуемых на NYSE составляет около 30 трлн долларов, что является практически третью мировых активов. Данные об индексах и котировках были взяты с сайта <https://ru.investing.com/>

Ниже приведен список компаний с соответствующими тикерами, используемых в исследовании:

**Таблица 1. Список компаний**

<b>Тикер</b>	<b>Название компании</b>
PFE	Pfizer
BMJ	Bristol-Myers Squibb
MRK	Merck&Co
JNJ	Johnson & Johnson
WBA	Walgreens Boots Alliance
ABT	Abbott Laboratories
LLY	Eli Lilly & Company
DHR	Danaher Corporation
MMM	3M
UNH	UnitedHealth Group

Для того чтобы выбранные акции было целесообразно проверять на нормальность распределения необходимо проверить количество данных. Так как исследуются котировки, посчитаем количество дней в году, когда проводятся торги. Проверим, что для всех акций торги шли больше двух третей года.

**Таблица 1. Число торговых дней**

Тикер/Год	2017	2018	2019	2020
<b>PFE</b>	251	251	252	253
<b>BMJ</b>	251	251	252	253
<b>MRK</b>	251	251	252	253
<b>JNJ</b>	251	251	252	254
<b>WBA</b>	251	251	252	253
<b>ABT</b>	251	251	252	254
<b>LLY</b>	251	251	252	254
<b>DHR</b>	251	251	252	254
<b>MMM</b>	251	251	252	253
<b>UNH</b>	251	251	252	253

Исходя из данных, представленных в таблице ясно, что все акции удовлетворят условию, имеют достаточное количество изменений. В 2017, 2018 и 2019 все акции учувствуют в торгах одинаковое количество дней. Только в 2020 году **JNJ ABT LLY** и **DHR** имеют на один день больше. Далее важно исследовать отклонения цен за каждый из рассматриваемых годов. Для этого посчитаем максимальные процентные скачки вверх и вниз.

**Таблица 3. Максимальные скачки вверх**

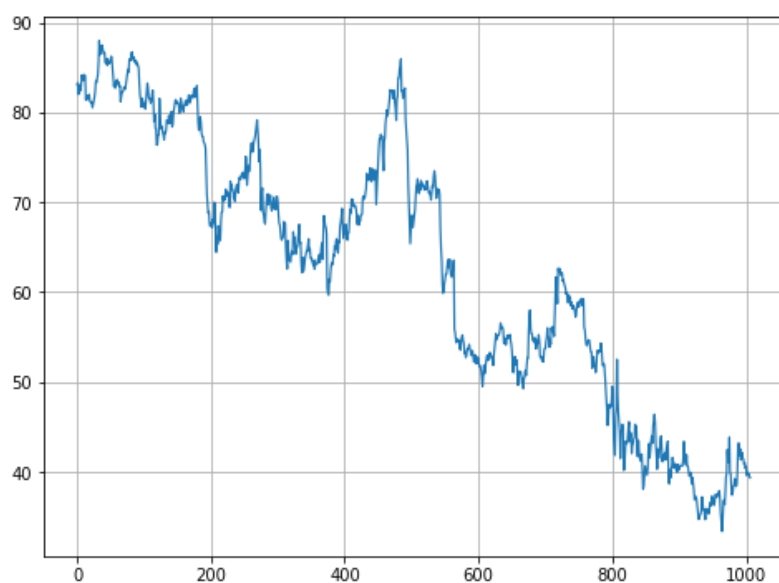
Тикер/Год	2017	2018	2019	2020	Max
<b>PFE</b>	0.030179	0.045845	0.032152	0.059746	0.059746
<b>BMJ</b>	0.038678	0.038873	0.058275	0.043604	0.058275
<b>MRK</b>	0.024585	0.045115	0.023599	0.077994	0.077994
<b>JNJ</b>	0.022663	0.027660	0.028114	0.062544	0.062544
<b>WBA</b>	0.040472	0.046486	0.055840	0.121470	0.121470
<b>ABT</b>	0.027204	0.056449	0.030113	0.073640	0.073640
<b>LLY</b>	0.037441	0.040558	0.027917	0.094908	0.094908
<b>DHR</b>	0.018570	0.050110	0.027911	0.067360	0.067360
<b>MMM</b>	0.024672	0.040116	0.025857	0.085467	0.085467
<b>UNH</b>	0.036026	0.061169	0.054039	0.093532	0.093532

**Таблица 4. Максимальные скачки вниз**

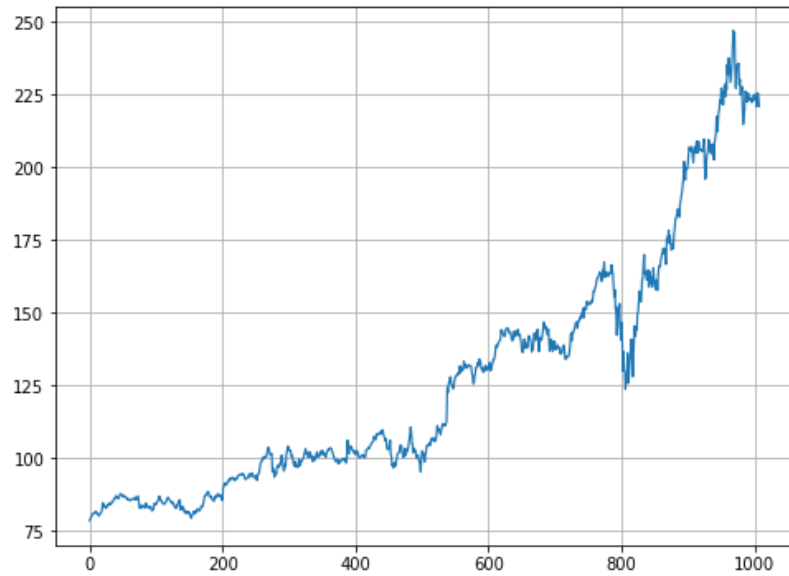
Тикер/ Год	2017	2018	2019	2020	Min
<b>PFE</b>	-0.016182	-0.049616	-0.030791	-0.063028	-0.063028
<b>BMJ</b>	-0.036427	-0.074131	-0.038363	-0.053097	-0.038363
<b>MRK</b>	-0.038150	-0.032092	-0.047545	-0.055073	-0.055073
<b>JNJ</b>	-0.015668	-0.078565	-0.042082	-0.054719	-0.054719
<b>WBA</b>	-0.038994	-0.046807	-0.059845	-0.075590	-0.075590
<b>ABT</b>	-0.025472	-0.046591	-0.043318	-0.084317	-0.084317
<b>LLY</b>	-0.032338	-0.058382	-0.039788	-0.073850	-0.073850
<b>DHR</b>	-0.030526	-0.046821	-0.032894	-0.107098	-0.107098
<b>MMM</b>	-0.022329	-0.049528	-0.047484	-0.095459	-0.095459
<b>UNH</b>	-0.034197	-0.046666	-0.071636	-0.077402	-0.077402

По представленным таблицам видно, что наибольший однодневный скачок вверх в цене произошел у **WBA** в 2020 году. Наибольший однодневный скачок вниз в цене в свою очередь произошел у **DHR** в 2020 году. Ниже представлены графики данных акций на всем исследуемом промежутке.

**Рисунок 5. График цен тикера WBA**



**Рисунок 6. График цен тикера DHR**



По графикам видно, что акции ведут себя нормально, без каких-либо ярких отклонений. Поэтому все данные подходят для дальнейшего исследования.

### 3. Теоретическая справка по проверке гипотез

#### 3.1 Статистическая проверка гипотез

Проверка статистических гипотез – это процедура, которая дает понимание об обоснованности определенной гипотезы.

Статистическая гипотеза – некоторое предположение о свойствах случайной величины.  $H_0$  — это нулевая гипотеза, которая принимается за основную, первоначальную, пока не будет доказано обратное. В этом случае далее рассматривается конкурирующая гипотеза  $H_1$ , которая является взаимоисключающей  $H_0$ .

Ошибка первого рода — это ситуация, когда гипотеза  $H_0$  отвергается, что является ошибкой, так как гипотеза  $H_0$  на самом деле верна. Вероятность ошибки первого рода называется значимостью критерия и обозначается  $\alpha$  (альфа).

Ошибка второго рода — это ситуация, когда гипотеза  $H_0$  принимается, что является ошибкой, так как гипотеза  $H_0$  на самом деле ошибочна. Вероятность ошибки второго рода обозначается  $\beta$  (бета).

P-значение или p-value – это значение статистического критерия, которая помогает принят решение об отклонении или неотклонении нулевой гипотезы. Принято, что нулевая гипотеза не отклоняется, если полученное p-value больше 0.05, в противном случае нулевая гипотеза отвергается.



### 3.2 Критерий Шапиро-Уилка

Критерий Шапиро-Уилка используется для проверки гипотезы  $H_0$  и является одним из наиболее эффективных критериев проверки нормальности.

Критерий Шапиро-Уилка использует отношение оптимальной линейной несмещенной оценки дисперсии к обычной оценке методом максимального правдоподобия. Статистика критерия имеет вид:

$$W = \frac{(\sum_{i=1}^n a_i x_{(i)})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

$x_{(i)}$  –  $i$ -я порядковая статистика,  $\bar{x}$  – среднее значение,  $a_i$  – табличное значение.

### 3.3 Критерий Колмогорова

Критерий Колмогорова помогает определить, принадлежит ли интересующая выборка определенному закону распределения. Статистикой критерия Колмогорова является максимум модуля разности значений интересующей функции распределения  $F_n(x)$  и теоретической функции распределения  $F(x)$ :

$$d = \max_{-\infty \leq x \leq \infty} |F_n(x) - F(x)|$$

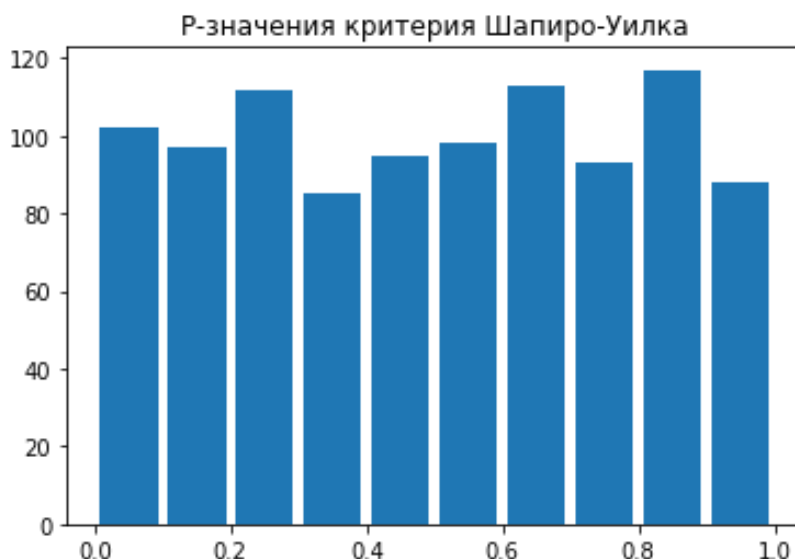
Нулевая гипотеза  $H_0$  принимается на уровне значимости  $\alpha$ , если  $\lambda = d\sqrt{n}$  при условии, что  $\lambda \leq \lambda_\alpha$ .

#### 4. Проверка гипотезы для модельных данных

Прежде всего, чтобы удостовериться в корректности действий и в правильной работе программ, проанализируем модельные данные. Создадим выборку, соответствующую нормальному закону распределения. Посчитаем статистику основного критерия, используя метод Монте-Карло, получим Р-значения и квантили.

**Таблица 7. Квантили модельных данных**

<b>Квантиль 0.1</b>	0.9904
<b>Квантиль 0.2</b>	0.9921
<b>Квантиль 0.3</b>	0.9930
<b>Квантиль 0.4</b>	0.9939
<b>Квантиль 0.5</b>	0.9946
<b>Квантиль 0.6</b>	0.9951
<b>Квантиль 0.7</b>	0.9956
<b>Квантиль 0.8</b>	0.9962
<b>Квантиль 0.9</b>	0.9967



**Рисунок 8. Р-значение при нормальном распределении**

Р-значение критерия Колмогорова: 0.8674

Исходя из вида гистограммы и полученного высокого значения критерия Колмогорова можно сделать вывод о нормальном распределении проверяемой выборки. Это подтверждает правильность действий и работоспособность кода.

## 5. Проверка гипотезы для реальных данных

Убедившись в правильности действий, приступаем к проверке гипотезы о нормальном распределении дневной логарифмической доходности при условии определенного объема торгов накануне. Используя актуальные данные о котировках акций в промежутке с 01.01.17 по 31.12.20, можно проверить гипотезу.

Прежде всего необходимо посчитать логарифмическую доходность каждого дня в течение интересующих 4 лет для всех данных.

Далее используя квантили 0.333 и 0.6666 для каждого тикера в отдельности, разделим данные по объему торгов на три примерно равные группы: маленький объем, средний объем, большой объем.

Рассчитаем Р-значения логарифмических доходностей каждой компании за годовой период при каждом объеме торгов. Строим таблицу по полученным данным.

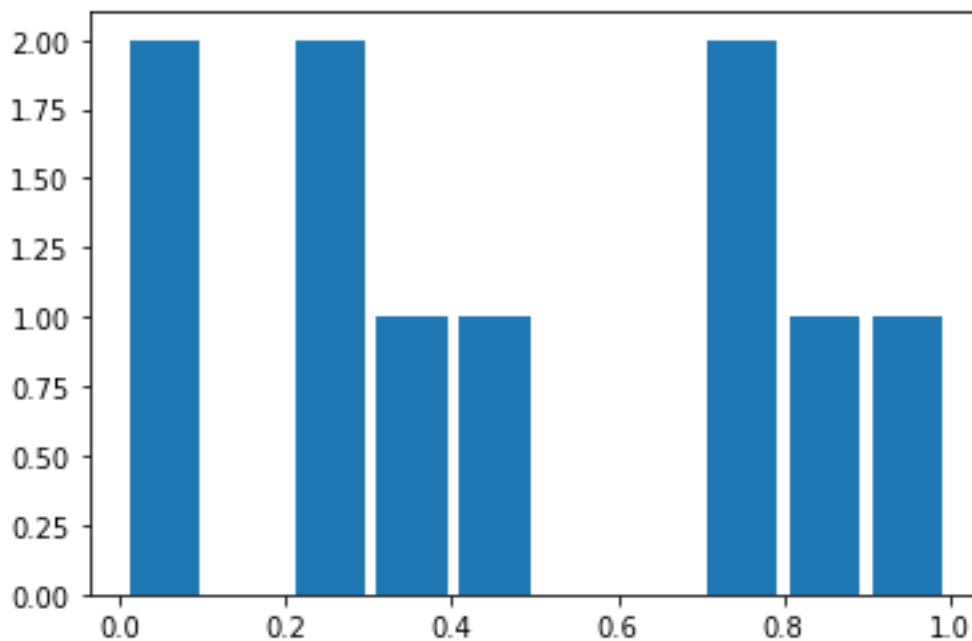
**Рисунок 9. Р-значения реальных данных по годам и объему.**

Тикеры	2017	2018	2019	2020
Маленький объем				
PFE	0.220291	0.060315	0.777501	0.265175
BMJ	0.039484	0.392865	0.560417	0.998637
MRK	0.218394	0.889395	0.386312	0.660998
JNJ	0.380365	0.885054	0.735833	0.403952
WBA	0.710576	0.944062	0.262954	0.582184
ABT	0.003217	0.506536	0.220215	0.499485
LLY	0.833966	0.283243	0.675506	0.760555
DHR	0.999844	0.244059	0.036233	0.378735
MMM	0.411028	0.131883	0.962223	0.620085
UNH	0.749912	0.976982	0.207067	0.204079
Средний объем				
PFE	0.848544	0.208191	0.148532	0.114967
BMJ	0.824017	0.726712	0.326529	0.747581
MRK	0.120153	0.614088	0.137565	0.003769
JNJ	0.184224	0.610466	0.688996	0.593590
WBA	0.047309	0.496514	0.179191	0.416618
ABT	0.403463	0.130624	0.178562	0.385584
LLY	0.090346	0.190592	0.320223	0.507565

DHR	0.495485	0.919026	0.125758	0.109490
MMM	0.114002	0.325854	0.162234	0.270596
UNH	0.683101	0.031266	0.232110	0.132650
Большой объем				
PFE	0.002983	0.591070	0.815606	0.000786
BMJ	0.957618	0.204273	0.316981	0.788126
MRK	0.027340	0.257828	0.127764	0.154076
JNJ	0.831479	0.000013	0.000084	0.000125
WBA	0.846967	0.616717	0.648918	0.012189
ABT	0.956170	0.204077	0.019934	0.288558
LLY	0.432282	0.015088	0.004245	0.013288
DHR	0.058515	0.095198	0.909015	0.001722
MMM	0.768818	0.197205	0.066123	0.000318
UNH	0.886279	0.887532	0.002379	0.221926

Изучим данные из полученной таблицы.

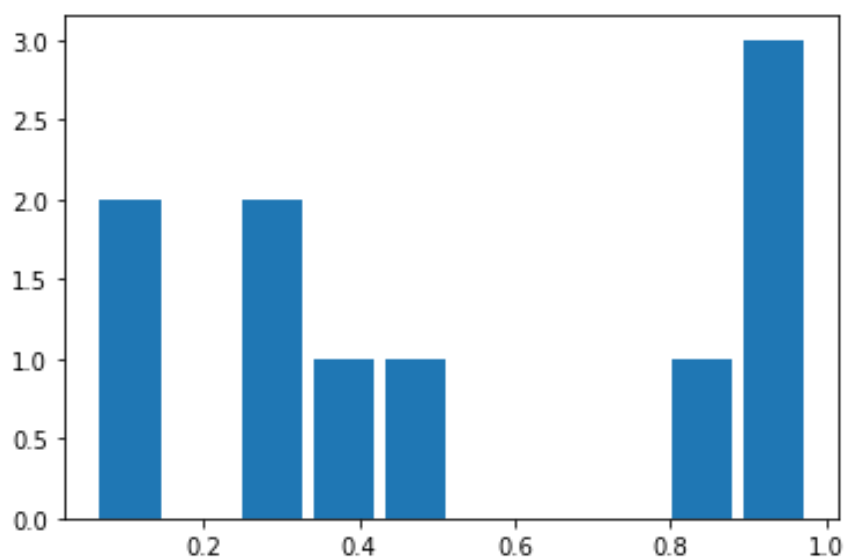
**Рисунок 10. Гистограмма Р-значений маленького объема 2017 г.**



Р-значение критерия Колмогорова: 0.8674663097670999

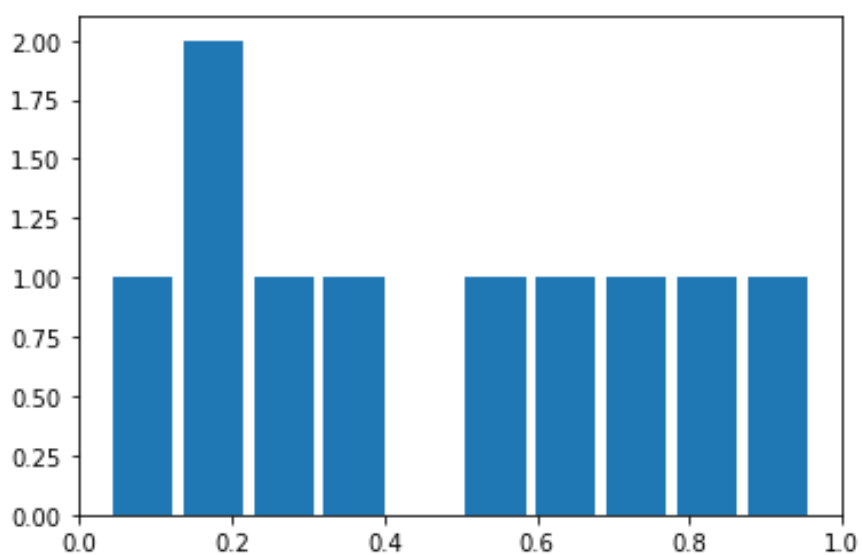
Исходя из гистограммы и критерия Колмогорова ясно, что Р-значения распределены равномерно. Значит нулевая гипотеза не опровергнута для малых объемов торгов в 2017 году.

**Рисунок 11. Гистограмма Р-значений маленького объема 2018 г.**



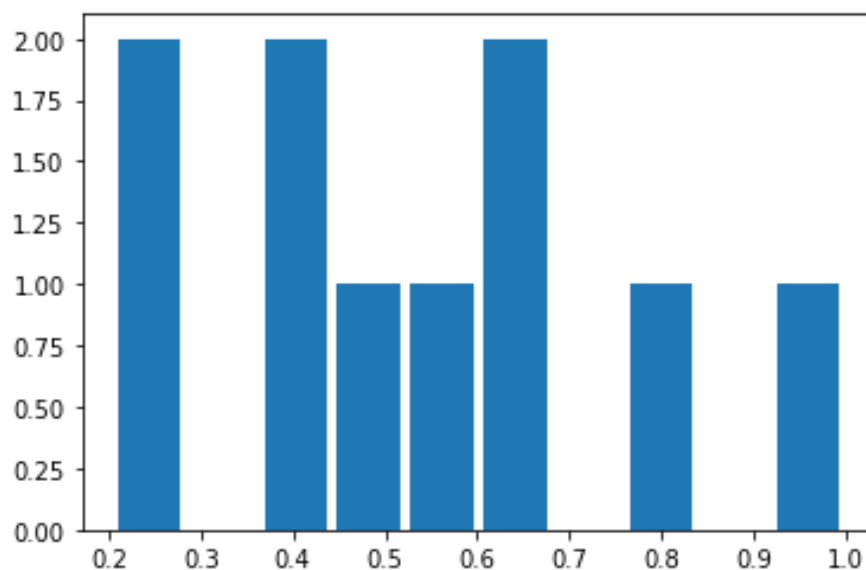
Р-значение критерия Колмогорова: 0.3267646738208149

**Рисунок 12. Гистограмма Р-значений маленького объема 2019 г.**



Р-значение критерия Колмогорова: 0.9918810764056362

**Рисунок 13. Гистограмма Р-значений маленького объема 2020 г.**

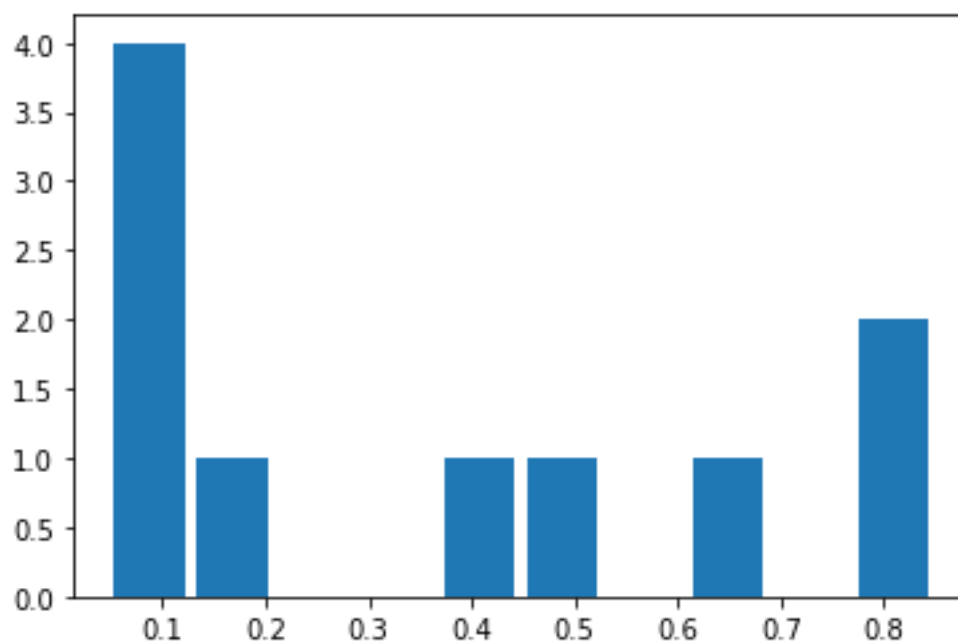


Р-значение критерия Колмогорова: 0.7991692944047911

По гистограммам и значениям критерия Колмогорова, представленным выше, видно, что для малых объемов торгов для всех 4 лет нулевая гипотеза о нормальном распределении дневной логарифмической доходности отвергаться не будет.

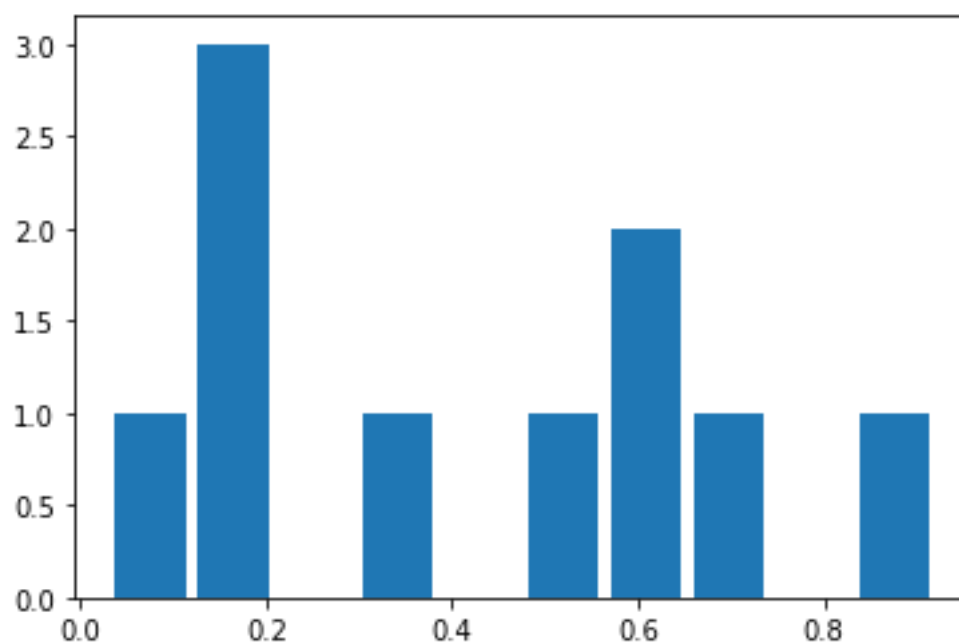
Рассмотрим остальные объемы торгов.

**Рисунок 14. Гистограмма Р-значений среднего объема 2017 г.**



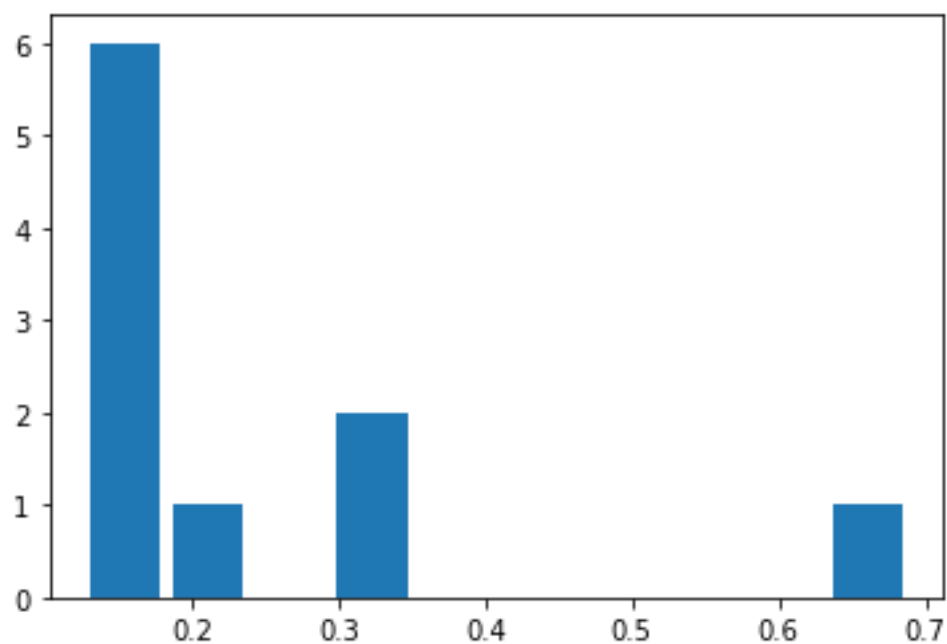
Р-значение критерия Колмогорова: 0.2197802555160243

**Рисунок 15. Гистограмма Р-значений среднего объема 2018 г.**



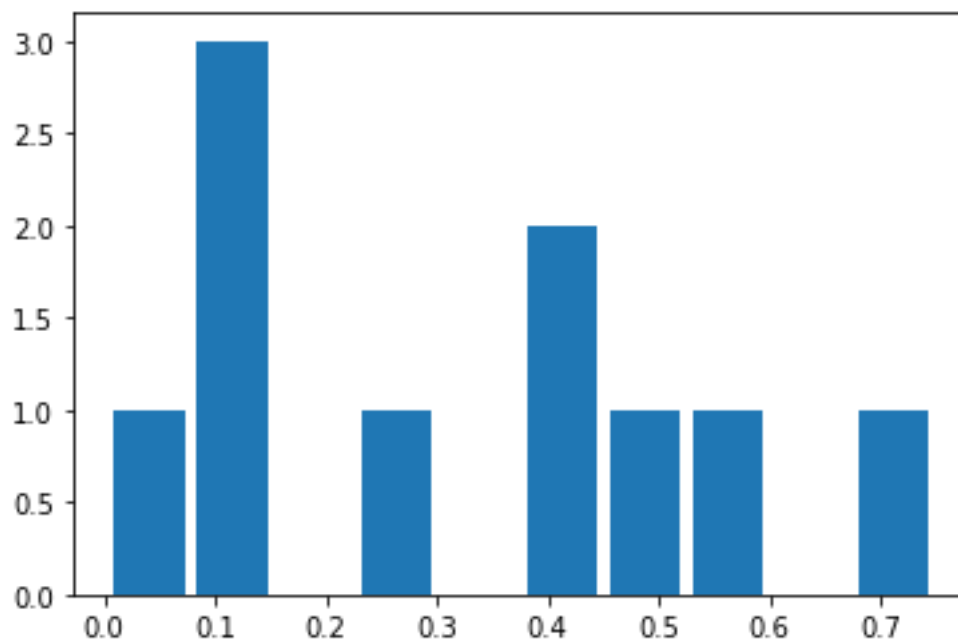
Р-значение критерия Колмогорова: 0.8554884201478801

**Рисунок 16. Гистограмма Р-значений среднего объема 2019 г.**



Р-значение критерия Колмогорова: 0.0012139729997274456

**Рисунок 17. Гистограмма Р-значений среднего объема 2020 г.**



Р-значение критерия Колмогорова: 0.24921592147367755

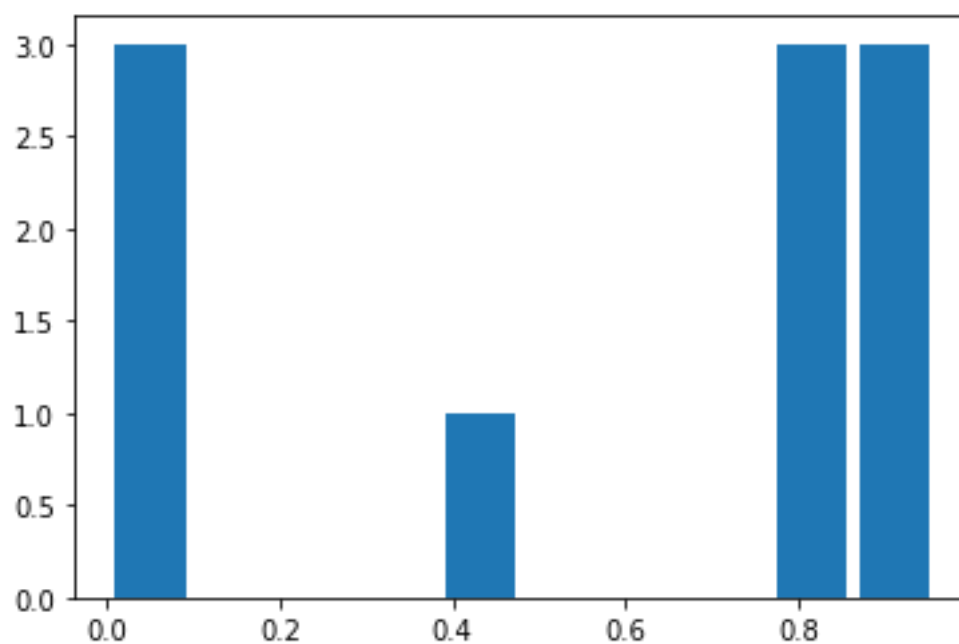
Сверху представлены данные о распределении Р-значения для средних объемов торгов. По ним видно, что для 2017, 2018 и 2020 годов значения распределены равномерно, и значит гипотеза о нормальном распределении логарифмической доходности не отвергается.

Для 2019 года исходя из гистограммы и значения критерия Колмогорова видно, что Р-значения распределены неравномерно. Это означает, что нулевая гипотеза для средних объемов торгов в 2019 году отвергается.

Разберемся с большими объемами торгов.

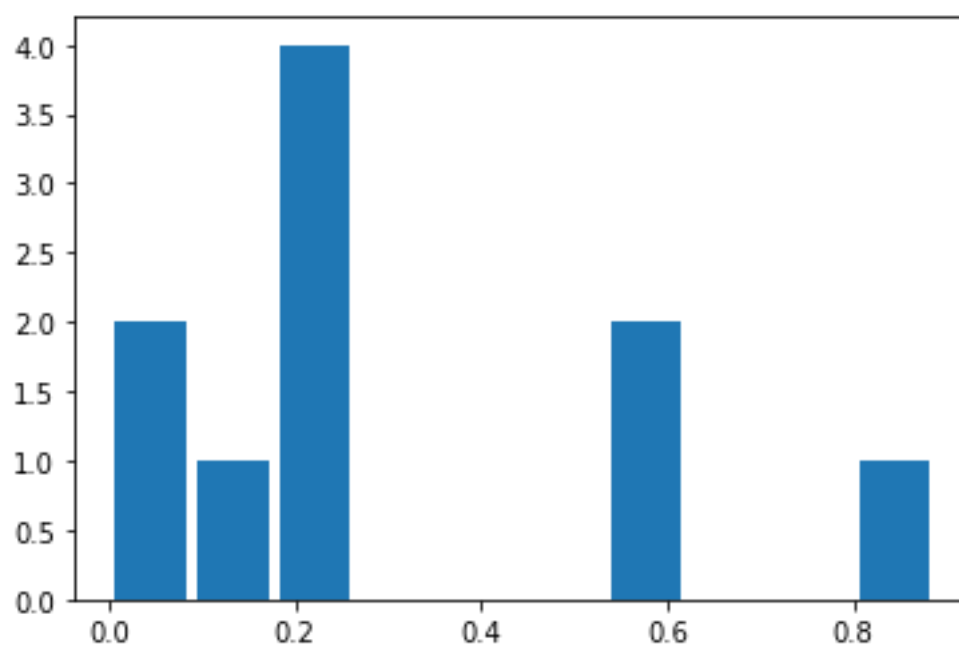


**Рисунок 18. Гистограмма Р-значений большого объема 2017 г.**



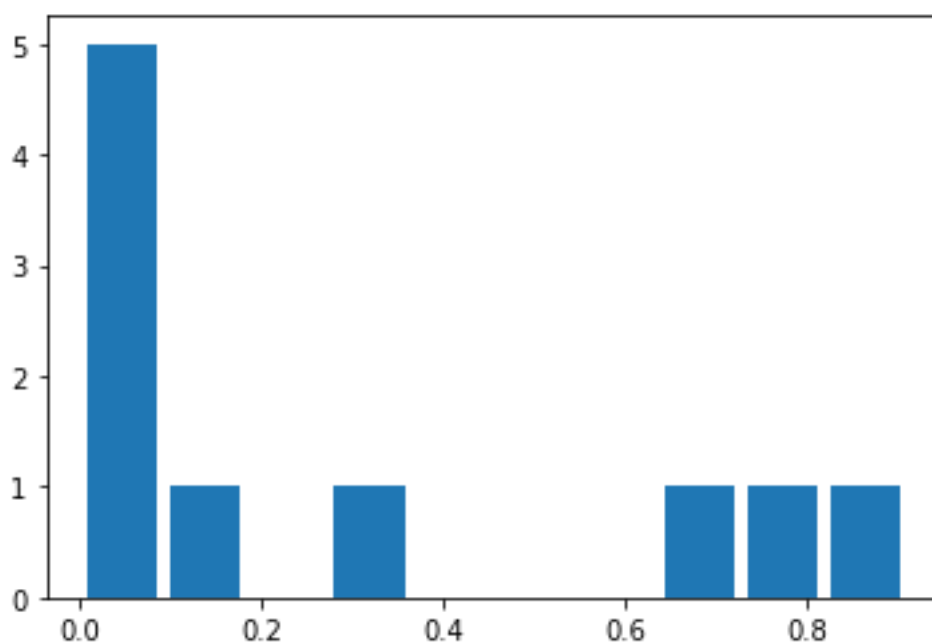
Р-значение критерия Колмогорова: 0.09975096198564863

**Рисунок 19. Гистограмма Р-значений большого объема 2018 г.**



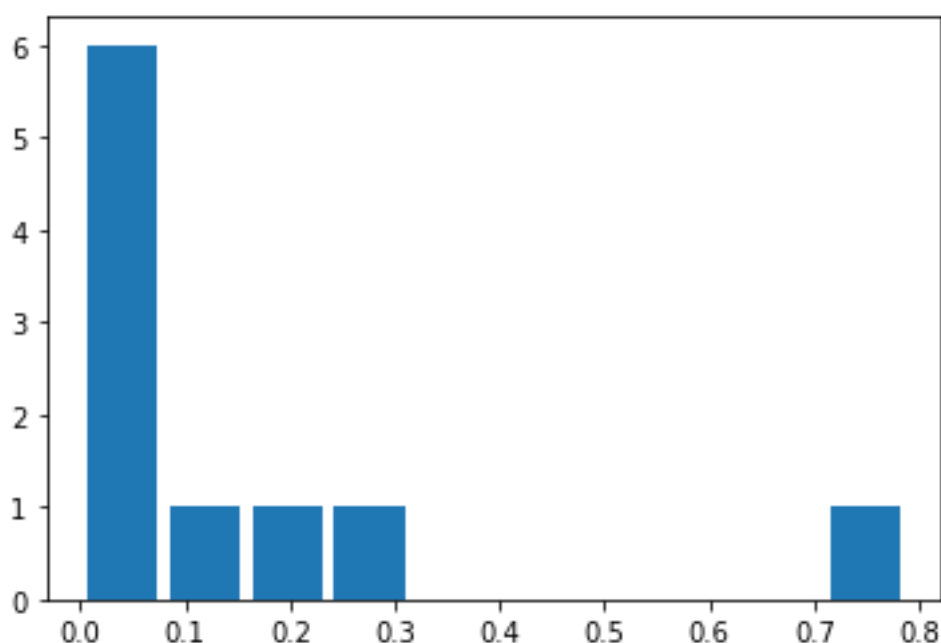
Р-значение критерия Колмогорова: 0.02677419342644842

**Рисунок 20. Гистограмма Р-значений большого объема 2019 г.**



Р-значение критерия Колмогорова: 0.01441300691238881

**Рисунок 21. Гистограмма Р-значений большого объема 2019 г.**



Р-значение критерия Колмогорова: 0.0004030691014627627

Выше представлены данные о распределении Р-значений при больших объемах торгов. По ним видно, что в 2018, 2019 и 2020 годах значения распределены неравномерно, а значит нулевая гипотеза отклоняется. В 2017 году данные показывают, что Р-значения распределены равномерно. Следовательно нулевая гипотеза не

отвергается. Однако значения критерия Колмогорова для 2017 года достаточно малы, что показывает ненадежность утверждения о равномерности распределения.

## **6. Заключение**

Целью работы было проверка гипотезы о нормальном распределении дневной логарифмической доходности при условии определенного объема торгов накануне. Для этого использовались актуальные данные о котировках акций, одновременно входящих в индексы NYSE U.S. 100 INDEX (\$NY.ID) и NYSE HEALTH CARE SECTOR INDEX (\$NYPTR) в промежутке с 01.01.17-31.12.20.

По результатам проведенного исследования с использованием критерия Шапиро-Уилка и критерия Колмогорова видно, что для малых объемов торгов гипотеза о нормальном распределении дневной логарифмической доходности за все 4 года подтверждается. Для средних объемов нулевая гипотеза подтверждается в трех и четырех годах. Для больших объемов данных нулевая гипотеза отклоняется.

Правильность вычислений была проверена на модельных данных, где гипотеза была подтверждена, как и ожидалось.

Можно сделать вывод, что объем торгов значительно влияет на нормальность распределения дневной логарифмической доходности. Чем больше объем, тем больше вероятность отклонения выборки от нормального распределения.

## 7. Список литературы

1. Браилов А. В. Лекции по математической статистике. М.: Финакадемий, 2007
2. Гусеева Е. Н. Теория вероятностей и математическая статистика.
3. Ивченко Г. И., Медведев Ю. И. Введение в математическую статистику: Учебник. М.: Издательство ЛКИ, 2010
4. Кобзарь А.И. Прикладная математическая статистика. Для инженеров и научных работников. – М.: Физматлит, 2006.
5. Лемешко Б. Ю., Лемешко С. Б. Сравнительный анализ критериев проверки отклонения распределения от нормального закона/Метрология. 2005. № 2. С. 3–24.
6. Курсы по анализу данных Вадима Леонардовича Аббакумова. - URL: <https://habr.com/ru/company/JetBrains-education/blog/438058/>
7. Финансовая платформа и новостной сайт. - URL: <https://ru.investing.com/>
8. Нью-Йоркская фондовая биржа. - URL: <https://www.nyse.com/index>

## 8. Приложения

### 8.1 Приложение

Intel(R) Core (TM) i7-9700 CPU @ 3.00GHz 3.00 GHz

Кэш L3: 12 Мб

Частота шины: 8GT/s

Тактовая частота: 3,00 GHz

### 8.2 Приложение

#### 2. Предварительный анализ данных

```
import numpy as np
import pandas as pd
from google.colab import files
import datetime
import matplotlib.pyplot as plt
import scipy.stats as sp
```

```
uploaded = files.upload()
```

```
WBA=pd.read_csv('WBA.csv', delimiter=',')
WBA=WBA.stack().str.replace(',','').unstack()
UNH=pd.read_csv('UNH.csv', delimiter=',')
UNH=UNH.stack().str.replace(',','').unstack()
PFE=pd.read_csv('PFE.csv', delimiter=',')
PFE=PFE.stack().str.replace(',','').unstack()
MRK=pd.read_csv('MRK.csv', delimiter=',')
MRK=MRK.stack().str.replace(',','').unstack()
MMM=pd.read_csv('MMM.csv', delimiter=',')
MMM=MMM.stack().str.replace(',','').unstack()
LLY=pd.read_csv('LLY.csv', delimiter=',')
LLY=LLY.stack().str.replace(',','').unstack()
JNJ=pd.read_csv('JNJ.csv', delimiter=',')
JNJ=JNJ.stack().str.replace(',','').unstack()
DHR=pd.read_csv('DHR.csv', delimiter=',')
DHR=DHR.stack().str.replace(',','').unstack()
BMV=pd.read_csv('BMV.csv', delimiter=',')
BMV=BMV.stack().str.replace(',','').unstack()
ABT=pd.read_csv('ABT.csv', delimiter=',')
ABT=ABT.stack().str.replace(',','').unstack()
```

```
for i in range(len(JNJ['Дата'])):
    if JNJ['Объём'][i][len(JNJ['Объём'])-1]=='K':
        JNJ['Объём'][i]=str(float(JNJ['Объём'][i][0:len(JNJ['Объём'])-1])/1000)[0:4]+'M'
    if JNJ['Объём'][i]=='-':
        JNJ['Объём'][i]='0.00M'
for i in range(len(ABT['Дата'])):
    if ABT['Объём'][i][len(ABT['Объём'])-1]=='K':
        ABT['Объём'][i]=str(float(ABT['Объём'][i][0:len(ABT['Объём'])-1])/1000)[0:4]+'M'
    if ABT['Объём'][i]=='-':
        ABT['Объём'][i]='0.00M'
for i in range(len(LLY['Дата'])):
    if LLY['Объём'][i][len(LLY['Объём'])-1]=='K':
        LLY['Объём'][i]=str(float(LLY['Объём'][i][0:len(LLY['Объём'])-1])/1000)[0:4]+'M'
    if LLY['Объём'][i]=='-':
        LLY['Объём'][i]='0.00M'
for i in range(len(DHR['Дата'])):
    if DHR['Объём'][i][len(DHR['Объём'])-1]=='K':
        DHR['Объём'][i]=str(float(DHR['Объём'][i][0:len(DHR['Объём'])-1])/1000)[0:4]+'M'
    if DHR['Объём'][i]=='-':
        DHR['Объём'][i]='0.00M'
```

```
def numdays(Tiker):
    Tikern7,Tikern8,Tikern9,Tikern20=0,0,0,0
    for i in Tiker['Дата']:
        if int(i[6:10])==2017: Tikern7+=1
        if int(i[6:10])==2018: Tikern8+=1
        if int(i[6:10])==2019: Tikern9+=1
        if int(i[6:10])==2020: Tikern20+=1
    return(Tikern7,Tikern8,Tikern9,Tikern20)

PFEd=numdays(PFE)
BMYd=numdays(BMY)
MRKd=numdays(MRK)
JNJd=numdays(JNJ)
WBAd=numdays(WBA)
ABTd=numdays(ABT)
LLYd=numdays(LLY)
DHRd=numdays(DHR)
MMMd=numdays(MMM)
UNHd=numdays(UNH)
```

Таблица 2. Число торговых дней

```
d = {"2017": pd.Series([PFEd[0],BMYd[0],MRKd[0],JNJd[0],WBAd[0],ABTd[0],LLYd[0],DHRd[0],MMMd[0],UNHd[0]],
    index=['PFE', 'BMY', 'MRK', 'JNJ', 'WBA', 'ABT', 'LLY', 'DHR', 'MMM', 'UNH']),
    "2018": pd.Series([PFEd[1],BMYd[1],MRKd[1],JNJd[1],WBAd[1],ABTd[1],LLYd[1],DHRd[1],MMMd[1],UNHd[1]],
    index=['PFE', 'BMY', 'MRK', 'JNJ', 'WBA', 'ABT', 'LLY', 'DHR', 'MMM', 'UNH']),
    "2019": pd.Series([PFEd[2],BMYd[2],MRKd[2],JNJd[2],WBAd[2],ABTd[2],LLYd[2],DHRd[2],MMMd[2],UNHd[2]],
    index=['PFE', 'BMY', 'MRK', 'JNJ', 'WBA', 'ABT', 'LLY', 'DHR', 'MMM', 'UNH']),
    "2020": pd.Series([PFEd[3],BMYd[3],MRKd[3],JNJd[3],WBAd[3],ABTd[3],LLYd[3],DHRd[3],MMMd[3],UNHd[3]],
    index=['PFE', 'BMY', 'MRK', 'JNJ', 'WBA', 'ABT', 'LLY', 'DHR', 'MMM', 'UNH'])}

pd.DataFrame(d)
```

```
def deviations(Tiker):
    dev2017,dev2018,dev2019,dev2020=[],[],[],[]
    for i in range(len(Tiker['Откр.'])):
        if int(Tiker['Дата'][i][6:10])==2017: dev2017.append(float(Tiker['Цена'][i])/float(Tiker['Откр.'][i])-1)
        if int(Tiker['Дата'][i][6:10])==2018: dev2018.append(float(Tiker['Цена'][i])/float(Tiker['Откр.'][i])-1)
        if int(Tiker['Дата'][i][6:10])==2019: dev2019.append(float(Tiker['Цена'][i])/float(Tiker['Откр.'][i])-1)
        if int(Tiker['Дата'][i][6:10])==2020: dev2020.append(float(Tiker['Цена'][i])/float(Tiker['Откр.'][i])-1)
    return([max(dev2017),min(dev2017)], [max(dev2018),min(dev2018)], [max(dev2019),min(dev2019)], [max(dev2020),min(dev2020)])
```

Таблица 3. Максимальные скачки вверх

```
d = {"2017": pd.Series([deviations(PFE)[0][0],deviations(BMY)[0][0],deviations(MRK)[0][0],deviations(JNJ)[0][0],deviations(WBA)[0][0],
    index=['PFE', 'BMY', 'MRK', 'JNJ', 'WBA', 'ABT', 'LLY', 'DHR', 'MMM', 'UNH']),
    "2018": pd.Series([deviations(PFE)[1][0],deviations(BMY)[1][0],deviations(MRK)[1][0],deviations(JNJ)[1][0],deviations(WBA)[1][0],
    index=['PFE', 'BMY', 'MRK', 'JNJ', 'WBA', 'ABT', 'LLY', 'DHR', 'MMM', 'UNH']),
    "2019": pd.Series([deviations(PFE)[2][0],deviations(BMY)[2][0],deviations(MRK)[2][0],deviations(JNJ)[2][0],deviations(WBA)[2][0],
    index=['PFE', 'BMY', 'MRK', 'JNJ', 'WBA', 'ABT', 'LLY', 'DHR', 'MMM', 'UNH']),
    "2020": pd.Series([deviations(PFE)[3][0],deviations(BMY)[3][0],deviations(MRK)[3][0],deviations(JNJ)[3][0],deviations(WBA)[3][0],
    index=['PFE', 'BMY', 'MRK', 'JNJ', 'WBA', 'ABT', 'LLY', 'DHR', 'MMM', 'UNH']),
    "Max": pd.Series([max(deviations(PFE))[0],max(deviations(BMY))[0],max(deviations(MRK))[0],max(deviations(JNJ))[0],max(deviations(WBA))[0],
    index=['PFE', 'BMY', 'MRK', 'JNJ', 'WBA', 'ABT', 'LLY', 'DHR', 'MMM', 'UNH'])}

pd.DataFrame(d)
```

Таблица 4. Максимальные скачки вниз

```
d = {"2017": pd.Series([deviations(PFE)[0][1],deviations(BMY)[0][1],deviations(MRK)[0][1],deviations(JNJ)[0][1],deviations(WBA)[0][1],
    index=['PFE', 'BMY', 'MRK', 'JNJ', 'WBA', 'ABT', 'LLY', 'DHR', 'MMM', 'UNH']),
    "2018": pd.Series([deviations(PFE)[1][1],deviations(BMY)[1][1],deviations(MRK)[1][1],deviations(JNJ)[1][1],deviations(WBA)[1][1],
    index=['PFE', 'BMY', 'MRK', 'JNJ', 'WBA', 'ABT', 'LLY', 'DHR', 'MMM', 'UNH']),
    "2019": pd.Series([deviations(PFE)[2][1],deviations(BMY)[2][1],deviations(MRK)[2][1],deviations(JNJ)[2][1],deviations(WBA)[2][1],
    index=['PFE', 'BMY', 'MRK', 'JNJ', 'WBA', 'ABT', 'LLY', 'DHR', 'MMM', 'UNH']),
    "2020": pd.Series([deviations(PFE)[3][1],deviations(BMY)[3][1],deviations(MRK)[3][1],deviations(JNJ)[3][1],deviations(WBA)[3][1],
    index=['PFE', 'BMY', 'MRK', 'JNJ', 'WBA', 'ABT', 'LLY', 'DHR', 'MMM', 'UNH']),
    "Min": pd.Series([max(deviations(PFE))[1],max(deviations(BMY))[1],max(deviations(MRK))[1],max(deviations(JNJ))[1],max(deviations(WBA))[1],
    index=['PFE', 'BMY', 'MRK', 'JNJ', 'WBA', 'ABT', 'LLY', 'DHR', 'MMM', 'UNH'])}

pd.DataFrame(d)
```

Рисунок 5. График цен тикера WBA

```
plt.figure(figsize=(8,6))
li=[]
for i in range(len(WBA['Цена'])):
    li.append(float(WBA['Откр.'][i]))
plt.plot(range(len(WBA['Цена'])),li[::-1],linewidth = 1.2)
plt.grid()
```

Рисунок 6. График цен тикера DHR

```
plt.figure(figsize=(8,6))
li=[]
for i in range(len(DHR['Цена'])):
    li.append(float(DHR['Откр.'][i]))
plt.plot(range(len(DHR['Цена'])),li[::-1],linewidth = 1.2)
plt.grid()
```

#### 4. Проверка гипотезы для модельных данных

Рисунок 8. P-значения при нормальном распределении

```
pvalues = []
statistics = []
ks=[]
for i in range(1000):
    distr = sp.shapiro(sp.norm.rvs(size = 250))
    statistics.append(distr[0])
    pvalues.append(distr[1])
kl=(sp.kstest(pvalues, cdf='uniform')[1])
print("P-значение критерия Колмогорова-Смирнова:", round(kl,4))
plt.title('P-значения критерия Шапиро-Уилка')
plt.hist(pvalues,rwidth=0.85)
plt.show()
```

Таблица 7. Квантили модельных данных

```
q9 = np.quantile(statistics, np.arange(0.1, 1, 0.1))
d = {'':pd.Series([round(q9[0],4),round(q9[1],4),round(q9[2],4),round(q9[3],4),round(q9[4],4),round(q9[5],4),round(q9[6],4),round(q9[7],4),round(q9[8],4),round(q9[9],4)],
    index=['Квантиль 0.1', 'Квантиль 0.2', 'Квантиль 0.3', 'Квантиль 0.4', 'Квантиль 0.5', 'Квантиль 0.6', 'Квантиль 0.7', 'Квантиль 0.8', 'Квантиль 0.9', 'Квантиль 1.0'])}
pd.DataFrame(d)
```

#### 5. Проверка гипотезы для реальных данных

```
def profvolume(Tiker):
    ldox=[]
    for i in range(1007):
        ldox.append(np.log(float(Tiker["Цена"][i])/float(Tiker["Откр."][i])))
    d = {"Дата":Tiker["Дата"],"Откр":Tiker["Откр."],"Закр":Tiker["Цена"],"Л.Доход":ldox, "Объем": Tiker["Объем"]}
    df = pd.DataFrame(d, index=range(1007))
    for i in range(len(df['Объем'])):
        df['Объем'][i]=float(df['Объем'][i][0:4])
        df['Дата'][i]=int(df['Дата'][i][6:10])

    q9 = np.quantile(df["Объем"], np.linspace(0, 1,4))
    df1 = df[df['Объем'] <= q9[1]]
    df11 = df[(df['Объем'] > q9[1]) & (df['Объем'] < q9[2])]
    df2 = df11[df11['Объем'] < q9[2]]
    df3 = df[df['Объем'] >= q9[2]]
    return df1,df2,df3
```

```
def pvolumes(df1,df2,df3):
    pm,pc,pb=[],[],[]
    dfd1 = df1[df1['Дата'] == 2020]
    pm.append(sp.shapiro(dfd1['Л.Доход'])[1])
    dfd2 = df1[df1['Дата'] == 2019]
    pm.append(sp.shapiro(dfd2['Л.Доход'])[1])
    dfd3 = df1[df1['Дата'] == 2018]
    pm.append(sp.shapiro(dfd3['Л.Доход'])[1])
    dfd4 = df1[df1['Дата'] == 2017]
    pm.append(sp.shapiro(dfd4['Л.Доход'])[1])

    dfd1 = df2[df2['Дата'] == 2020]
    pc.append(sp.shapiro(dfd1['Л.Доход'])[1])
    dfd2 = df2[df2['Дата'] == 2019]
    pc.append(sp.shapiro(dfd2['Л.Доход'])[1])
    dfd3 = df2[df2['Дата'] == 2018]
    pc.append(sp.shapiro(dfd3['Л.Доход'])[1])
    dfd4 = df2[df2['Дата'] == 2017]
    pc.append(sp.shapiro(dfd4['Л.Доход'])[1])

    dfd1 = df3[df3['Дата'] == 2020]
    pb.append(sp.shapiro(dfd1['Л.Доход'])[1])
    dfd2 = df3[df3['Дата'] == 2019]
    pb.append(sp.shapiro(dfd2['Л.Доход'])[1])
    dfd3 = df3[df3['Дата'] == 2018]
    pb.append(sp.shapiro(dfd3['Л.Доход'])[1])
    dfd4 = df3[df3['Дата'] == 2017]
    pb.append(sp.shapiro(dfd4['Л.Доход'])[1])
    return pm,pc,pb
```

```
def datafPvalues(Tiker):
    df1,df2,df3=profvolume(Tiker)
    Tikervp=pvolumes(df1,df2,df3)
    d = {"2017": pd.Series([Tikervp[0][3],Tikervp[1][3],Tikervp[2][3]], index=['М.Объем', 'С.Объем', 'Б.Объем']),
        "2018": pd.Series([Tikervp[0][2],Tikervp[1][2],Tikervp[2][2]], index=['М.Объем', 'С.Объем', 'Б.Объем']),
        "2019": pd.Series([Tikervp[0][1],Tikervp[1][1],Tikervp[2][1]], index=['М.Объем', 'С.Объем', 'Б.Объем']),
        "2020": pd.Series([Tikervp[0][0],Tikervp[1][0],Tikervp[2][0]], index=['М.Объем', 'С.Объем', 'Б.Объем'])}
    f=pd.DataFrame(d)
    return f
```

Рисунок 9. Р-значения реальных данных по годам и объему.

datafPvalues(WBA)

Гистограммы Р-значений Рисунок 10-21

```
xm2020=[datafPvalues(PFE)['2020'][2],
        datafPvalues(BMY)['2020'][2],
        datafPvalues(MRK)['2020'][2],
        datafPvalues(JNJ)['2020'][2],
        datafPvalues(WBA)['2020'][2],
        datafPvalues(ABT)['2020'][2],
        datafPvalues(LLY)['2020'][2],
        datafPvalues(DHR)['2020'][2],
        datafPvalues(MMM)['2020'][2],
        datafPvalues(UNH)['2020'][2]]
print("Р-значение критерия Колмогорова:", sp.kstest(xm2020, cdf='uniform')[1])
plt.hist(xm2020,rwidth=0.85)
plt.show()
```