итоговая работа

№ 1.1

итоговая работа Зименкова → 11212

№ 1.1

Вероятность одной ошибки составляет 30%, вероятность того, что произойдет более одной ошибки, составляет 10%. Вероятность перезаключения контракта в случае наличия ровно одной ошибки составляет 30%, в случае более одной ошибки составляет 50%, а в случае отсутствия — 90%. Какова вероятность перезаключения?

A – одна ошибка $P(A) = 0{,}3$

B – >1 ошибки $P(B) = 0{,}1$

C – нет ошибок $P(C) = 1 - P(A) - P(B) = 0{,}6$

H – контракт перезаключен

$P(H/A) = 0{,}3$

$P(H/B) = 0{,}5$

$P(H/C) = 0{,}9$

по формуле полной вероятности

$P(H) = P(H/A)\cdot P(A) + P(H/B)\cdot P(B) + P(H/C)\cdot P(C) = 0{,}3\cdot 0{,}3 + 0{,}5\cdot 0{,}1 + 0{,}9\cdot 0{,}6 =$

$= 0{,}09 + 0{,}05 + 0{,}54 = 0{,}68$

Ответ: $P(H) = 0{,}68$

№ 2.1

№ 2.1

В результате каждой покупки вы получаете первый купон с вероятностью 0,8, а второй с вероятностью 0,2. Сколько в среднем нужно сделать покупок, чтобы собрать оба купона?

$P_1 = 0{,}8$ — получен 1й купон

$P_2 = 0{,}2$ – получен 2й купон

M – среднее кол-во покупок, которое нужно сделать, чтобы собрать оба купона

M_1 – кол-во ... чтобы один купон $M_1 = 1$

M_2 – кол-во ... чтобы 2й купон, если уже есть 1й, или 1й, если уже есть 2й.

$M_2 = P_1 \cdot M(2й\ к.) + P_2 \cdot M(1й\ к.)$

т.к. сбор оставшихся купонов распр. геометрически, то:

$M(1й\ к.) = \frac{1}{P_1} = \frac{10}{8}$ $M(2й\ к.) = \frac{1}{P_2} = \frac{10}{2}$

$M_2 = \frac{8}{10}\cdot\frac{10}{2} + \frac{2}{10}\cdot\frac{10}{8} = \frac{8}{2} + \frac{2}{8} = \frac{16}{4} + \frac{1}{4} = \frac{17}{4}$

$M = M_1 + M_2 = 1 + \frac{17}{4} = \frac{21}{4} = 5{,}25$

Ответ: 5,25 покупок

№ 3.2

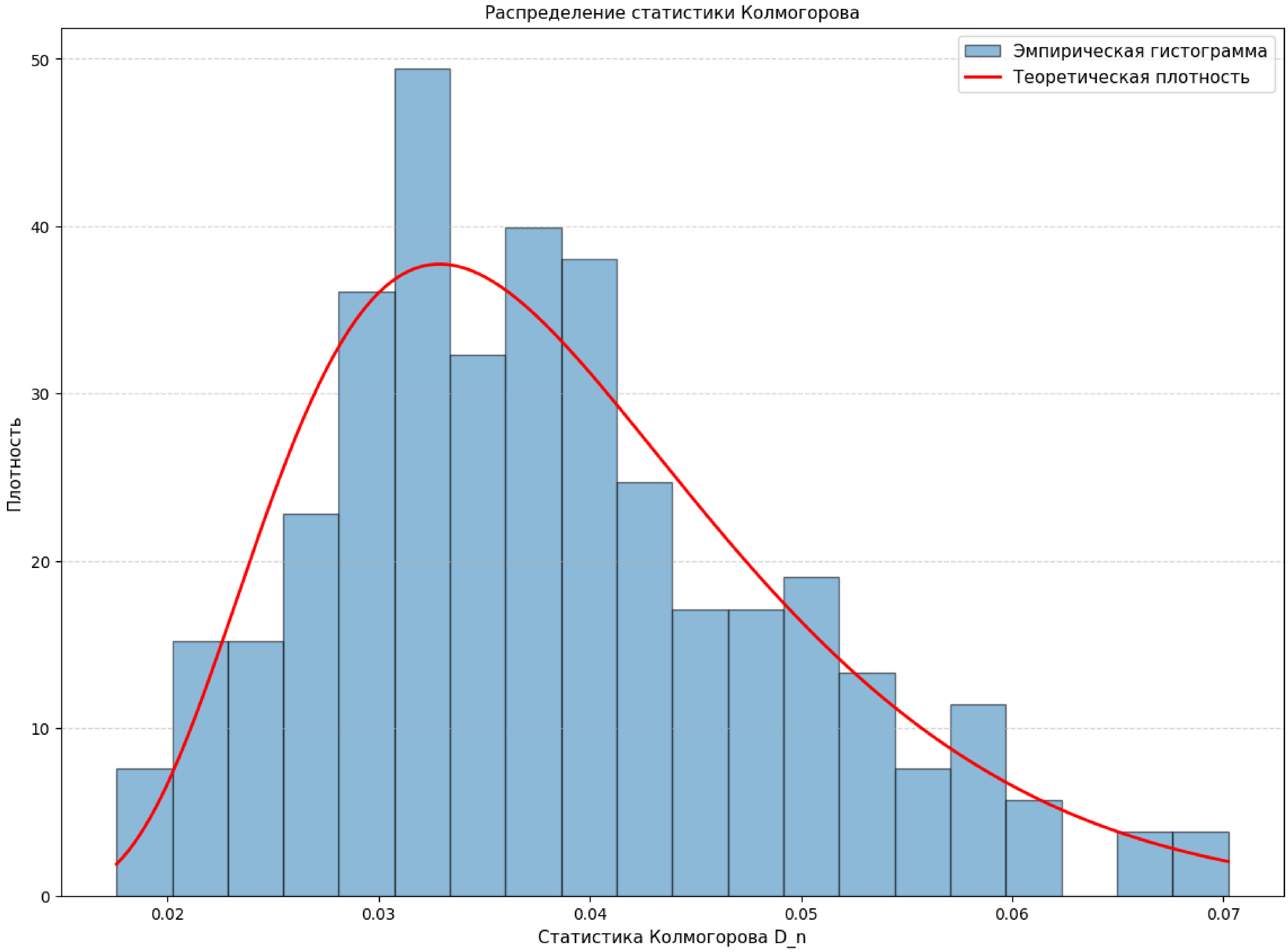
ноутбук colab: [https://colab.research.google.com/drive/1zNVAJ6RLzIYrsBqABYtzh7pYUqy_v3E4?usp=sharing (https://colab.research.google.com/drive/1zNVAJ6RLzIYrsBqABYtzh7pYUqy_v3E4?usp=sharing)](https://colab.research.google.com/drive/1zNVAJ6RLzIYrsBqABYtzh7pYUqy_v3E4?usp=sharing)

Сгенерируйте 200 выборок размера 500 из Exp(1). Посчитайте для каждой выборки статистику Колмогорова. Нарисуйте гистограмму по полученному набору статистик. На нем же изобразите график теоретической плотности распределения Колмогорова. Согласуется ли результат с теорией?

```python
import numpy as np
from scipy.stats import kstest, expon
import matplotlib.pyplot as plt

# параметры
n = 200 # число выборок
sample_size = 500 # размер каждой выборки

# массив с выборками
samples = np.random.exponential(scale=1, size=(n, sample_size))

# массив со статистиками Колмогорова
ks_stats = []

# считаем статистики
for sample in samples:
    stat, p_value = kstest(sample, 'expon', args=(0,1))
    ks_stats.append(stat)

ks_stats = np.array(ks_stats)

# одновыборочный ks-тест для проверки согласия с теорией

from scipy.stats import kstest, kstwobign # распределение Колмогорова

print("KS-тест для проверки согласия с теоретическим распределением:")

ks_stat, ks_pvalue = kstest(
    ks_stats,
    lambda x: kstwobign.cdf(x * np.sqrt(sample_size))
)

print("KS-статистика: ", ks_stat)
print("p-value: ", ks_pvalue)

if ks_pvalue > 0.05:
  print("Гипотезу можно принять - распределение согласуется с теоретическим")
else:
  print("Гипотезу следует отвергнуть - распределение не согласуется с теоретическим")

plt.figure(figsize=(14, 10))

# график - гистограмма с плотностью
x_values = np.linspace(min(ks_stats), max(ks_stats), 200)
kstwobign_pdf = kstwobign.pdf(x_values * np.sqrt(sample_size)) * np.sqrt(sample_size)

plt.hist(ks_stats, bins=20, density=True, edgecolor='black', alpha=0.5, label='Эмпирическая гистограмма')
plt.plot(x_values, kstwobign_pdf, 'r-', lw=2, label='Теоретическая плотность')
plt.xlabel('Статистика Колмогорова D_n', fontsize=11)
plt.ylabel('Плотность', fontsize=11)
plt.title('Распределение статистики Колмогорова', fontsize=11)
plt.legend(fontsize=11)
plt.grid(axis='y', linestyle='--', alpha=0.5)

plt.show()
```

Результат

```
KS-тест для проверки согласия с теоретическим распределением:
KS-статистика: 0.04529923507533962
p-value: 0.7889880982071726
Гипотезу можно принять - распределение согласуется с теоретическим
```



Вывод

Получившаяся гистограмма наглядно показывает, что сгенерированный результат хорошо совпадает с теоретическим распределением. Это также подтверждается KS-тестом: получилось довольно большое значение p-value, а значит, эмпирические данные действительно хорошо совпадают с теоретическим распределением. Небольшие отклонения (выходы столбцов и выбросы в правом хвосте) обоснованы случайными вариациями и небольшим объемом выборки.

Конечно, если позапускать случайную генерацию много раз, результаты могут получиться разными - иногда появляются заметные выбросы, которые визуально сильно выбиваются из графика, а KS-тест показывает, что распределение не совпало, но это в целом ожидаемо из-за случайной генерации и небольшого объема выборки.

№ 4.1

Из загруженного набора акций с 2014 года по конец 2022 года возьмите нужную по счету акцию в файле, соответствующую номеру этой задачи, и для нее по close values посчитайте 5-дневный 5-% VaR логарифмической доходности (по выборке). На тестовой части выборки (2023 год) проведите бэктестинг оцененного VaR одним из частотных тестов.

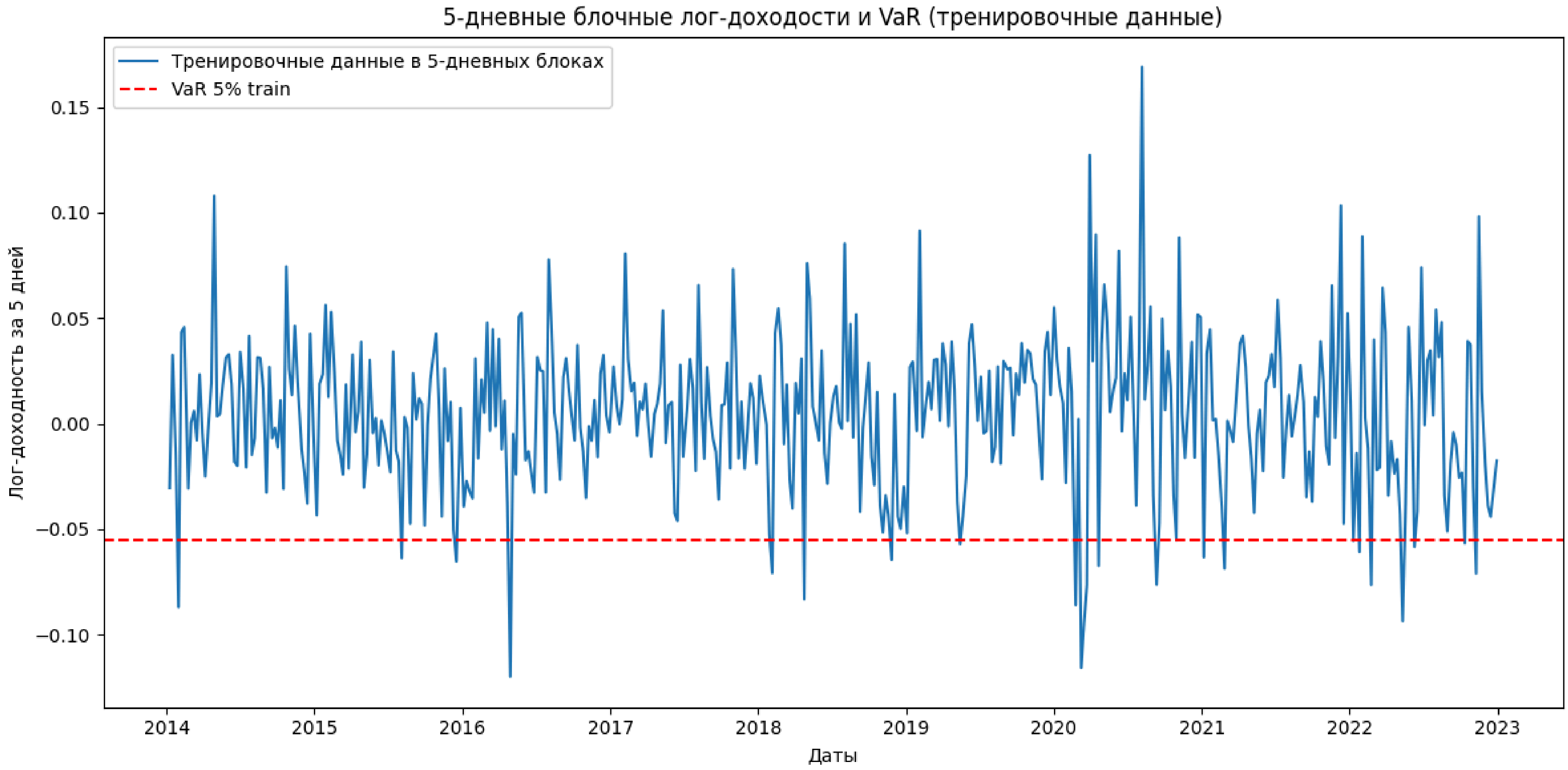
1-й вариант - AAPL

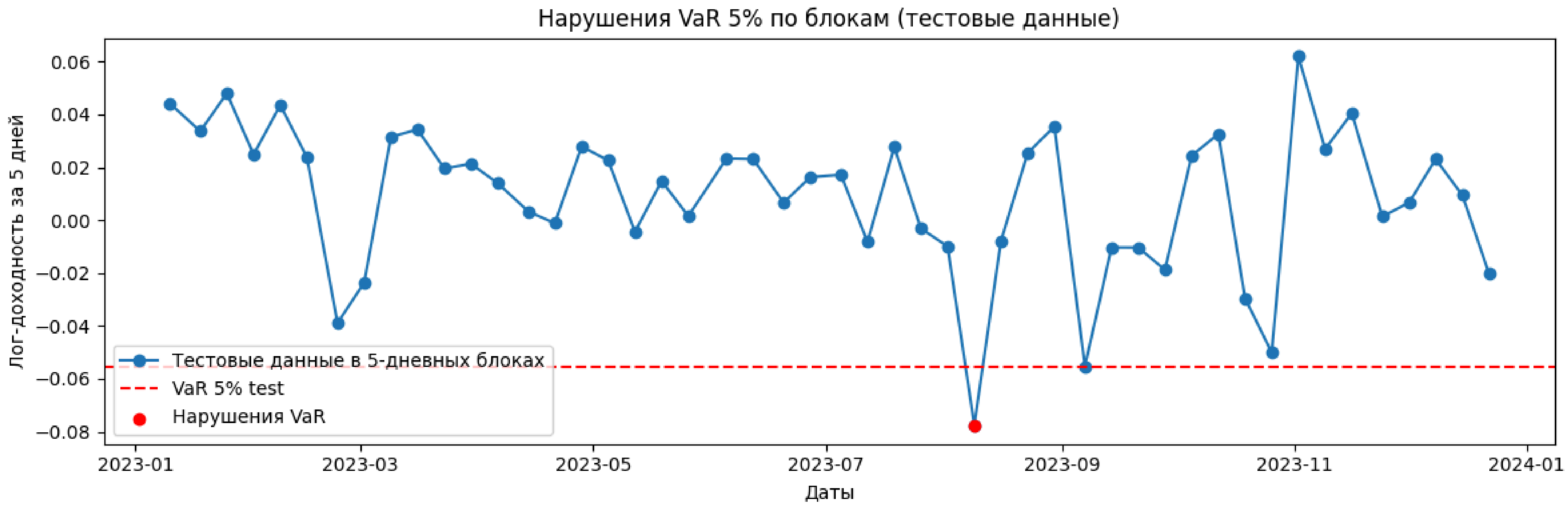
```python
import pandas as pd
import numpy as np
from scipy.stats import chi2, chi2_contingency
import matplotlib.pyplot as plt

# загружаем данные
df  = pd.read_csv('data.csv')
df['Date'] = pd.to_datetime(df['Date'])
df = df.sort_values('Date')
ticker = 'AAPL'

# разделяем train/test
train = df[df['Date'] < '2023-01-01'].copy()
test = df[(df['Date'] >= '2023-01-01') & (df['Date'] < '2024-01-01')].copy()

# лог-доходности будем разбивать на блоки по 5 дней
size = 5

# тренировочные данные
train_prices = train[ticker].values
train_dates = train['Date'].values
train_logreturns = np.log(train_prices[1:] / train_prices[:-1])
n_train_blocks = len(train_logreturns) // size

train_block_logreturn = [
    train_logreturns[i * size: (i+1) * size].sum() for i in range(n_train_blocks)
]
train_block_ends = [
    train_dates[(i + 1) * size] for i in range(n_train_blocks)
]

# тестовые данные
test_prices = test[ticker].values
test_dates = test['Date'].values
test_logreturns = np.log(test_prices[1:] / test_prices[:-1])
n_test_blocks = len(test_logreturns) // size

test_block_logreturn = [
    test_logreturns[i * size: (i+1) * size].sum() for i in range(n_test_blocks)
]
test_block_ends = [
    test_dates[(i + 1) * size] for i in range(n_test_blocks)
]

# расчет VaR 5% (квантиль по train)
var_5p_5d = np.quantile(train_block_logreturn, 0.05)
print("5-дневный VaR 5% по блочным лог-доходностям (тренировочные данные): ", var_5p_5d)

# проверяем нарушения VaR на тестовых данных

# exs = np.array([float(y) for y in test_block_logreturn])
# print(exs)

exc = np.array([int(x < var_5p_5d) for x in test_block_logreturn])
n = len(exc)
x = exc.sum() # исключения
alpha = 0.05 # целевая частота исключений
```

```python
# бэктест методом Купика (пропорциональный тест, POF)
if 0 < x < n:
  pi_hat = x / n
  # LR-статистика Купика
  lr = -2 * ((x*np.log(alpha) + (n - x) * np.log(1 - alpha)) -
   (x * np.log(pi_hat) + (n - x) * np.log(1 - pi_hat)))
  p_value = 1 - chi2.cdf(lr, df=1)
else:
  lr = np.False_p_value = np.nan

print()
print('-------------------------------------------------')
print("Бэктестинг VaR на тестовых данных:")
print()
print("Ожидаемых нарушений VaR (n * alpha): ", n * alpha)
print("Фактических нарушений VaR (x): ", x)
print("p_value: ", p_value)

if p_value > 0.05:
  print("p_value > 0.05, нет оснований отвергать правильность VaR")
else:
  print("p_value < 0.05, VaR отвергается, частота нарушений не совпадает с теоретической")

# график train (тренировочные данные и VaR)
plt.figure(figsize=(12, 6))
plt.plot(train_block_ends, train_block_logreturn, linestyle='-',
         label='Тренировочные данные в 5-дневных блоках')
plt.axhline(var_5p_5d, color='red', linestyle='--', label='VaR 5% train')
plt.title('5-дневные блочные лог-доходости и VaR (тренировочные данные)')
plt.xlabel('Даты')
plt.ylabel('Лог-доходность за 5 дней')
plt.legend()
plt.tight_layout()
plt.show

# график test (тестовые данные и нарушения VaR)
plt.figure(figsize=(12, 4))
plt.plot(test_block_ends, test_block_logreturn, marker='o',
         label='Тестовые данные в 5-дневных блоках')
plt.axhline(var_5p_5d, color='red', linestyle='--', label='VaR 5% test')
plt.scatter([test_block_ends[i] for i in range(n) if exc[i] == 1],
 [test_block_logreturn[i] for i in range (n) if exc[i] == 1],
 color='red', zorder=5, label='Нарушения VaR')
plt.title('Нарушения VaR 5% по блокам (тестовые данные)')
plt.xlabel('Даты')
plt.ylabel('Лог-доходность за 5 дней')
plt.legend()
plt.tight_layout()
plt.show
```

Результат

```
5-дневный VaR 5% по блочным лог-доходностям (тренировочные данные): -0.055315535814685804

-------------------------------------------------
Бэктестинг VaR на тестовых данных:

Ожидаемых нарушений VaR (n * alpha): 2.45
Фактических нарушений VaR (x): 1
p_value: 0.283020441561747
p_value > 0.05, нет оснований отвергать правильность VaR
```





Вывод

5-дневный VaR 5% = -0.05531553. За 5 дней лог-доходность акции не упадет ниже -5,53%.
Нарушений: 1 из 49. Ожидаемое кол-во - 49 · 0,05 = 2,45.

Достаточно большое p-value и маленькое количество нарушений VaR (1 < 2,45) показывают, что VaR рассчитан верно. Одно значение было довольно близко к VaR: -0.05526984, когда VaR -0.05531553, и его видно на графике, однако все же лог-доходность в этом блоке больше, чем VaR. В любом случае, 2 нарушения это все еще в пределах ожидаемого количества.

№ 5.2

№ 5.2

$M(X) = \mu_X = 8 \quad M(Y) = \mu_Y = 100$

$Var(X) = \sigma_X^2 = 2^2 \quad Var(Y) = \sigma_Y^2 = 10^2$

$Corr(X,Y) = \rho = -0{,}5$

регрессия Y на X:

$Y = \beta_0 + \beta_1 X + \varepsilon \qquad M(\varepsilon) = 0 \qquad x_0 = 9$

1. найти β_0, β_1

$\beta_0 = M(Y) - \beta_1 M(X) \qquad \beta_1 = \dfrac{Cov(X,Y)}{Var(X)}$

$Cov(X,Y) = Corr \cdot \sigma_X \cdot \sigma_Y \Rightarrow \beta_1 = \dfrac{\rho \cdot \sigma_X \cdot \sigma_Y}{\sigma_X^2} = \dfrac{\rho \sigma_Y}{\sigma_X} = \dfrac{-0{,}5 \cdot 10}{2} = -2{,}5$

$\beta_0 = 100 - (-2{,}5) \cdot 8 = 100 + 20 = 120$

2. найти уравнение $\hat{Y} = \beta_0 + \beta_1 X$

$\hat{Y} = 120 - 2{,}5X$

3. найти $\hat{Y}$ при $x_0 = 9$

$\hat{Y}(x_0) = 120 - 2{,}5 \cdot 9 = 97{,}5$

4. найти R^2

т.к. регрессия простая, то $R^2 = Corr^2(X,Y) = \rho^2 = 0{,}25$

Ответ: 1. $\beta_0 = 120$, $\beta_1 = -2{,}5$ 2. $\hat{Y} = 120 - 2{,}5X$

3. $\hat{Y}(x_0) = 97{,}5$ 4. $R^2 = 0{,}25$

спасибо за курс! это было трудно, но интересно ··

— любительница математических операторов,

Σοφία