Министерство науки и высшего образования Российской Федерации
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ ИТМО»
(Университет ИТМО)

Факультет безопасности информационных технологий

Дисциплина:

«Статистические методы в инженерных исследованиях»

ОТЧЕТ
ПО ЛАБОРАТОРНОЙ РАБОТЕ №1

«Первичная обработка экспериментальных данных.
Статистический анализ экспериментальных данных»

Вариант 5

Студент группы N3346 ______________ Бардышев А. А.
(подпись)

Студент группы N3346 ______________ Волощук А. Н.
(подпись)

Студент группы N3346 ______________ Замахов Е. В.
(подпись)

Студент группы N3346 ______________ Пинус И. В.
(подпись)

Студент группы N3346 ______________ Суханкулиев М.
(подпись)

Студент группы N3346 ______________ Шегай С. Д.
(подпись)

Проверил:

Преподаватель, д.т.н., профессор ______________ Федоров А. В.
(подпись)

«___» ________________ 2025 г.

Санкт-Петербург

2025 г.

СОДЕРЖАНИЕ

ВВЕДЕНИЕ

Цель работы – получить умения и отработать навыки оценивания параметров эмпирических распределений по выборке экспериментальных данных.

Для достижения поставленной цели необходимо решить следующие задачи:

- провести серию экспериментов (измерений);
- обработать результаты измерений для идентификации закона распределения;
- оформить отчет по лабораторной работе.

Объектом исследования являются результаты испытаний на восстановление исправного состояния изделия. Обработка данных будет производиться с использованием языка программирования Python в среде Google Colaboratory с применением библиотек NumPy, Pandas, Matplotlib и SciPy.

1 СТАТИСТИЧЕСКИЙ АНАЛИЗ ДАННЫХ О ПРОДОЛЖИТЕЛЬНОСТИ РЕМОНТА

1.1 Постановка задачи и исходные данные

Обработать результаты испытаний на восстановление исправного состояния изделия, а именно:

1. Определить значения характеристик эмпирического распределения продолжительности ремонта изделия (час.) по результатам испытаний.
2. Осуществить выбор теоретического закона распределения продолжительности ремонта изделия.
3. Провести проверку гипотезы о соответствии эмпирического распределения продолжительности ремонта изделия выбранному теоретическому закону распределения.

Исходные данные, представляющие собой выборку из $n = 50$ результатов измерений продолжительности ремонта изделия, час, представлены в таблице 1.

Таблица 1 – Исходные экспериментальные данные

№	x_i	№	x_i	№	x_i	№	x_i	№	x_i
1	31	**11**	1	**21**	4	**31**	2	**41**	3
2	7	**12**	2	**22**	2	**32**	7	**42**	5
3	2	**13**	6	**23**	6	**33**	3	**43**	4
4	11	**14**	2	**24**	9	**34**	11	**44**	21
5	13	**15**	7	**25**	3	**35**	19	**45**	4
6	8	**16**	16	**26**	9	**36**	6	**46**	7
7	11	**17**	42	**27**	13	**37**	4	**47**	1
8	3	**18**	16	**28**	35	**38**	30	**48**	6
9	2	**19**	12	**29**	16	**39**	2	**49**	2
10	26	**20**	32	**30**	5	**40**	2	**50**	39

1.2 Построение вариационного и статистического рядов

Для первичного анализа данных необходимо упорядочить исходную выборку по возрастанию, сформировав вариационный ряд. Результат представлен таблице 2.

Таблица 2 – Вариационный ряд выборки

```
--- Вариационный ряд ---
[ 1  1  2  2  2  2  2  2  2  2  2  3  3  3  3  4  4  4  4  5  5  6  6  6
  6  7  7  7  7  8  9  9 11 11 11 12 13 13 16 16 16 19 21 26 30 31 32 35
 39 42]
```

Для дальнейшего анализа сгруппируем данные. Определим основные параметры для группировки:

1. Минимальное и максимальное значения: $x_{min} = 1$; $x_{max} = 42$.
2. Размах варьирования: $R = x_{max} - x_{min} = 41$.
3. Количество интервалов по формуле Стерджесса: $k = 1 + 3.322 \cdot \lg 50 \approx 6.64$. Принимаем $k = 7$.
4. Ширина интервала: $\Delta x == \frac{R}{k} = \frac{41}{7} \approx 5.86$. Для удобства принимаем $\Delta x = 6$.

На основе этих параметров построим статистический ряд распределения, который включает в себя частоты, относительные и накопленные частоты, а также плотность относительной частоты.

Формулы для расчёта:

$$\text{Середины интервалов} - x_j$$

$$p_j = \frac{n_j}{n}$$

$$\text{Накопленные частоты} - n_j^{\Sigma}$$

$$F_j = \frac{n_j^{\Sigma}}{n}$$

$$\omega_j = \frac{n_j}{\Delta x}$$

$$f_j = \frac{p_j}{\Delta x}$$

Результаты расчетов сведены в таблице 3.

Таблица 3 – Статистический ряд распределения

--- Статистический ряд распределения ---

	Границы интервалов	Середины интервалов	Частоты (n_j)	Отн. частоты (p_j)	Накопл. частоты	Накопл. отн. частоты (F_j)	Плотность отн. частоты (f_j)
0	[1; 7)	4.0	25	0.50	25	0.50	0.083333
1	[7; 13)	10.0	11	0.22	36	0.72	0.036667
2	[13; 19)	16.0	5	0.10	41	0.82	0.016667
3	[19; 25)	22.0	2	0.04	43	0.86	0.006667
4	[25; 31)	28.0	2	0.04	45	0.90	0.006667
5	[31; 37)	34.0	3	0.06	48	0.96	0.010000
6	[37; 43)	40.0	2	0.04	50	1.00	0.006667

1.3 Визуальный анализ распределения

Для наглядного представления характера распределения построим гистограмму относительных частот и график эмпирической функции распределения (кумуляту).

На рисунке 1 представлена гистограмма плотности относительных частот.

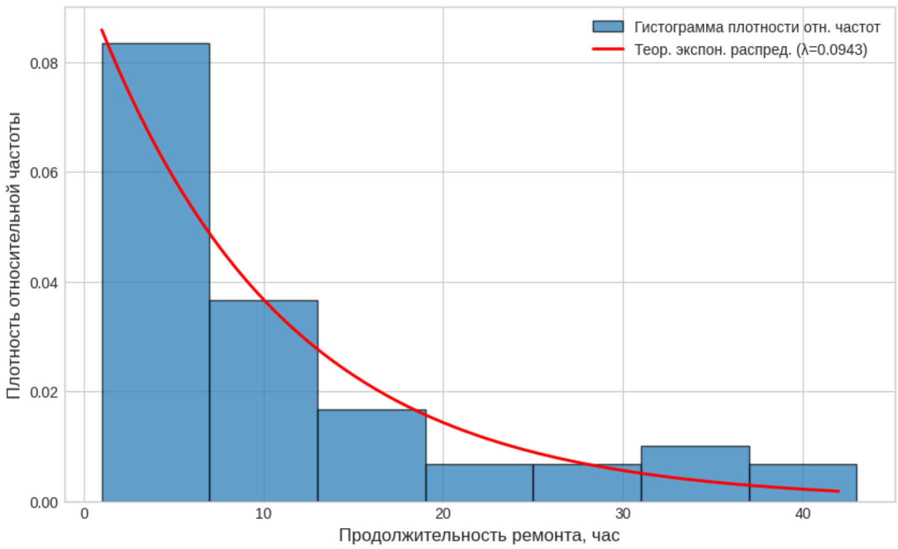

Рисунок 1 – Гистограмма плотности относительных частот

Визуальный анализ гистограммы показывает, что распределение является существенно асимметричным. Наибольшая частота приходится на первый интервал (малые значения продолжительности ремонта), после чего частоты быстро убывают. Такая форма характерна для экспоненциального распределения.

На рисунке 2 представлен график эмпирической функции распределения.

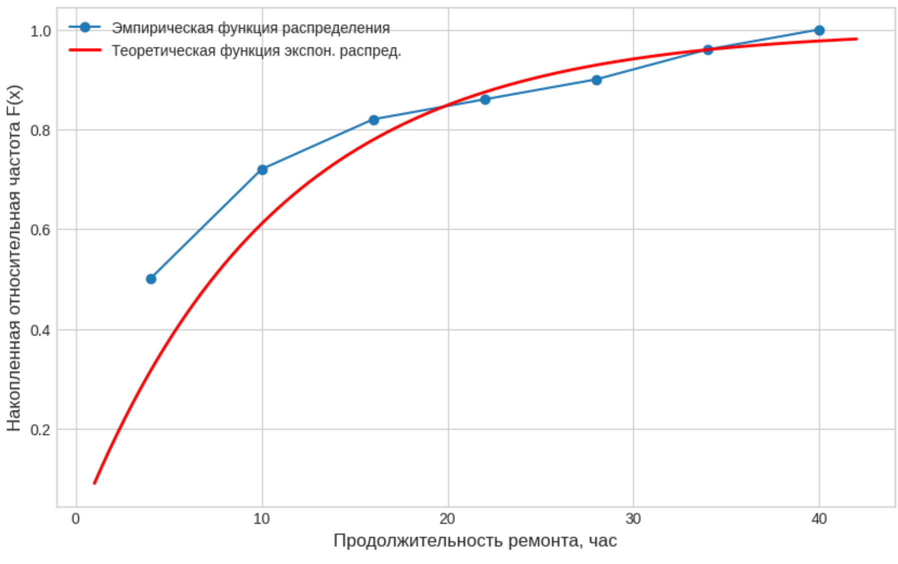

Рисунок 2 – График эмпирической функции распределения (кумулята)

График кумуляты также подтверждает асимметричный характер распределения: быстрый рост вначале и постепенное приближение к единице.

1.4 Расчет числовых характеристик эмпирического распределения

Для количественной оценки свойств распределения рассчитаем его основные числовые характеристики:

1. Математическое ожидание: $M_x = \overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}$ ≈10.60
2. Дисперсия (несмещенная): $D_x = S_x^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}$ ≈114.73
3. Среднеквадратическое отклонение: $S_x = \sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}}$ ≈10.71
4. Коэффициент асимметрии: $As_x = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^3}{nS_x^3}$ ≈1.48
5. Эксцесс: $Ex_x = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^4}{nS_x^4} - 3$ ≈1.22
6. Коэффициент вариации: $\nu = \frac{S_x}{\overline{x}}$ ≈1.01
7. Мода (групп.): $Me_x = min(x_j^{Me}) + \frac{\Delta x\left(0{,}5n - n_{j-1}^{\Sigma}\right)}{n_j^{Me}}$ ≈4.85
8. Медиана (групп.): $Mo_x = min(x_j^{Mo}) + \frac{\Delta x\left(n_j^{Mo} - n_{j-1}^{Mo}\right)}{\left(n_j^{Mo} - n_{j-1}^{Mo}\right) + \left(n_j^{Mo} - n_{j+1}^{Mo}\right)}$ ≈7.00

Анализ числовых характеристик:

– Коэффициент асимметрии $As = 1.48 > 0$, что подтверждает наличие значительной положительной асимметрии, выявленной при визуальном анализе.

– Коэффициент вариации $\nu = 1.01$. Это ключевой показатель. Значение, очень близкое к 1, является сильным аргументом в пользу гипотезы об экспоненциальном законе распределения, для которого теоретически $\nu = 1$.

1.5 Выбор теоретического закона и проверка гипотезы о виде распределения

На основании проведенного анализа выдвигается нулевая гипотеза H_0: эмпирическое распределение продолжительности ремонта изделия согласуется с экспоненциальным теоретическим законом распределения.

Обоснование выбора:

1. Визуальный анализ: гистограмма (рисунок 1) имеет характерную для экспоненциального закона форму.
2. Количественный анализ: коэффициент вариации $\nu \approx 1$.
3. Физический смысл: продолжительность ремонта, как и время безотказной работы, часто описывается экспоненциальным законом, который моделирует время до наступления случайного события в простейшем потоке.

Для проверки гипотезы H_0 воспользуемся критерием согласия Пирсона (χ^2). Уровень значимости примем $\alpha = 0.05$.

Плотность вероятности экспоненциального распределения: $f(x) = \lambda \cdot e^{-\lambda x}$, где λ – интенсивность. Оценка параметра λ вычисляется как $\lambda = \frac{1}{M(x)} = \frac{1}{10.6} \approx 0.0943$.

Вычислим теоретические вероятности P_j' попадания в каждый интервал и соответствующие теоретические частоты $n_j' = n \cdot P_j'$. Для предотвращения некорректного применения критерия χ^2, объединим интервалы, для которых теоретическая частота $n_j' < 5$.

Расчетные значения для критерия Пирсона приведены в таблице 4.

Таблица 4 – Расчет наблюдаемого значения критерия χ²

```
--- Расчет критерия Пирсона (χ²) ---
       Границы  n_j  n_j_theor  (n-n')^2/n'
0       [1; 7)   25  19.665866     1.446821
1      [7; 13)   11  11.165702     0.002459
2     [13; 19)    5   6.339558     0.283050
3  [19;    43)    9   7.462172     0.316920
```

*Сумма теоретических частот не равна 50, так как мы рассматриваем интервалы, а не всю числовую ось.

Наблюдаемое значение статистики критерия: $\chi^2_{\text{набл}} = 2.05$.

Определим критическое значение $\chi^2_{\text{кр}}$ по таблице распределения хи-квадрат. Число степеней свободы $r = k' - m - 1$, где:

– k' – число интервалов после объединения ($k' = 4$).

– m – число параметров распределения, оцененных по выборке ($m = 1$, это параметр λ).

Таким образом, $r = 4 - 1 - 1 = 2$.

Для уровня значимости $\alpha = 0.05$ и $r = 2$ критическое значение составляет $\chi^2_{\text{кр}} = 5.99$.

Вывод по проверке гипотезы:

Так как $\chi^2_{\text{набл}} = 2.05 < \chi^2_{\text{кр}} = 5.99$, то нет оснований отвергать нулевую гипотезу H_0. Следовательно, эмпирическое распределение продолжительности ремонта изделия с доверительной вероятностью 95% согласуется с экспоненциальным законом с параметром $\lambda \approx 0.0943$.

ЗАКЛЮЧЕНИЕ

В ходе выполнения лабораторной работы были получены умения и отработаны навыки оценивания параметров эмпирических распределений.

Была проанализирована выборка из 50 измерений продолжительности ремонта изделия. На основе построенных гистограммы, кумуляты и рассчитанных числовых характеристик (в частности, коэффициента вариации $\nu = 1.01$) была выдвинута гипотеза об экспоненциальном характере распределения исследуемой случайной величины.

Проверка гипотезы с помощью критерия согласия Пирсона (χ^2) подтвердила ее состоятельность на уровне значимости $\alpha = 0.05$. Наблюдаемое значение критерия ($\chi^2_{\text{набл}} = 2.05$) оказалось значительно меньше критического ($\chi^2_{\text{кр}} = 5.99$), что позволило принять гипотезу.

Таким образом, задача идентификации закона распределения для заданной выборки была успешно решена. Продолжительность ремонта изделия может быть адекватно смоделирована экспоненциальным распределением с параметром $\lambda \approx 0.0943$.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Сидняев, Н. И. Теория планирования эксперимента и анализ статистических данных : учебное пособие для вузов / Н. И. Сидняев. – 2-е изд., перераб. и доп. – Москва : Издательство Юрайт, 2011. – 399 с. – Текст : непосредственный.
2. Рыков, В. В. Математическая статистика и планирование эксперимента : учебное пособие / В. В. Рыков, В. Ю. Иткин. – Москва : РГУ нефти и газа им. И. М. Губкина, 2008. – 210 с. – Текст : непосредственный.
3. ГОСТ 24026-80. Исследовательские испытания. Планирование эксперимента. Термины и определения. – Введ. 1981-07-01. – Москва : Издательство стандартов, 1980. – 15 с. – Текст : непосредственный.
4. NumPy: the fundamental package for scientific computing with Python : официальный сайт. – Текст : электронный. – 2025. – URL: https://numpy.org/doc/stable/ (дата обращения: 15.10.2025).
5. Pandas: powerful Python data analysis toolkit : официальный сайт. – Текст : электронный. – 2025. – URL: https://pandas.pydata.org/docs/ (дата обращения: 15.10.2025).
6. Varoquaux, G. SciPy: high-level scientific computing / G. Varoquaux, A. Chauve, A. Espaze [et al.]. – Текст : электронный // Scientific Python Lecture Notes. – 2025. – URL: https://main--lectures-scientific-python-org.netlify.app/intro/scipy/ (дата обращения: 15.10.2025).
7. Matplotlib: Visualization with Python : официальный сайт. – Текст : электронный. – 2025. – URL: https://matplotlib.org/stable/contents.html (дата обращения: 15.10.2025).

ПРИЛОЖЕНИЕ А

(обязательное)

Листинг А.1 – Код для Google Colaboratory

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from scipy import stats

# Улучшение визуального стиля графиков
plt.style.use('seaborn-v0_8-whitegrid')

data = np.array([
    31, 7, 2, 11, 13, 8, 11, 3, 2, 26,
    1, 2, 6, 2, 7, 16, 42, 16, 12, 32,
    4, 2, 6, 9, 3, 9, 13, 35, 16, 5,
    2, 7, 3, 11, 19, 6, 4, 30, 2, 2,
    3, 5, 4, 21, 4, 7, 1, 6, 2, 39
])

n = len(data)
print("--- Исходные данные ---")
print(data)
print(f"Объем выборки n = {n}")

x_min = variational_series.min()
x_max = variational_series.max()
R = x_max - x_min

# Количество интервалов по формуле Стерджесса
k = round(1 + 3.322 * np.log10(n))

# Ширина интервала
dx = R / k
# Округление
dx_rounded = np.ceil(dx)

print(f"Минимальное значение x_min = {x_min}")
print(f"Максимальное значение x_max = {x_max}")
print(f"Размах варьирования R = {R}")
print(f"Число интервалов (по Стерджессу) k ≈ {1 + 3.322 * np.log10(n):.2f},
      принимаем k = {k}")
print(f"Ширина интервала dx = {R}/{k} ≈ {dx:.2f}, принимаем dx =
      {dx_rounded}")

# Интервалы
bins = np.arange(x_min, x_max + dx_rounded, dx_rounded)
# Частоты
frequencies, bin_edges = np.histogram(data, bins=bins)

# DataFrame для красивого вывода
df_stats = pd.DataFrame()
df_stats['Границы интервалов'] = [f"[{int(bin_edges[i])};
      {int(bin_edges[i+1])})" for i in range(len(bin_edges)-1)]
df_stats['Середины интервалов'] = (bin_edges[:-1] + bin_edges[1:]) / 2
df_stats['Частоты (n_j)'] = frequencies
df_stats['Отн. частоты (p_j)'] = frequencies / n
df_stats['Накопл. частоты'] = frequencies.cumsum()
```

```
df_stats['Накопл. отн. частоты (F_j)'] = df_stats['Отн. частоты
      (p_j)'].cumsum()
df_stats['Плотность отн. частоты (f_j)'] = df_stats['Отн. частоты (p_j)'] /
      dx_rounded

print("--- Статистический ряд распределения ---")
print(df_stats.to_string())

# Гистограмма
plt.figure(figsize=(10, 6))
weights = np.ones_like(data) / (len(data) * dx_rounded) # Нормировка для
      плотности
plt.hist(data, bins=bins, weights=weights, edgecolor='black', alpha=0.7,
      label='Гистограмма плотности отн. частот')

# Теоретическая кривая для наглядности
lambda_param = 1 / data.mean()
x_theoretical = np.linspace(x_min, x_max, 200)
y_theoretical = lambda_param * np.exp(-lambda_param * x_theoretical)
plt.plot(x_theoretical, y_theoretical, 'r-', linewidth=2, label=f'Теор.
      экспон. распред. (λ={lambda_param:.4f})')

plt.title('Гистограмма плотности относительных частот', fontsize=14)
plt.xlabel('Продолжительность ремонта, час', fontsize=12)
plt.ylabel('Плотность относительной частоты', fontsize=12)
plt.legend()
plt.grid(True)
plt.show()

# Кумулята
plt.figure(figsize=(10, 6))
# Эмпирическая функция
plt.plot(df_stats['Середины интервалов'], df_stats['Накопл. отн. частоты
      (F_j)'], 'o-', label='Эмпирическая функция распределения')
# Теоретическая функция
cdf_theoretical = 1 - np.exp(-lambda_param * x_theoretical)
plt.plot(x_theoretical, cdf_theoretical, 'r-', linewidth=2,
      label='Теоретическая функция экспон. распред.')

plt.title('График эмпирической функции распределения (кумулята)',
      fontsize=14)
plt.xlabel('Продолжительность ремонта, час', fontsize=12)
plt.ylabel('Накопленная относительная частота F(x)', fontsize=12)
plt.legend()
plt.grid(True)
plt.show()

# --- Расчет числовых характеристик ---
mean_val = np.mean(data)
variance_val = np.var(data, ddof=1) # ddof=1 для несмещенной оценки
std_dev_val = np.std(data, ddof=1)
skewness_val = stats.skew(data)
kurtosis_val = stats.kurtosis(data)
variation_coeff_val = std_dev_val / mean_val

# Мода
modal_interval_index = df_stats['Частоты (n_j)'].idxmax()
n_m = df_stats.loc[modal_interval_index, 'Частоты (n_j)']
n_m_minus_1 = df_stats.loc[modal_interval_index - 1, 'Частоты (n_j)'] if
      modal_interval_index > 0 else 0
n_m_plus_1 = df_stats.loc[modal_interval_index + 1, 'Частоты (n_j)'] if
      modal_interval_index < len(df_stats) - 1 else 0
```

```
x0_m = bin_edges[modal_interval_index]
mode_val_grouped = x0_m + dx_rounded * (n_m - n_m_minus_1) / (2 * n_m -
      n_m_minus_1 - n_m_plus_1)

# Медиана
median_interval_index = df_stats[df_stats['Накопл. частоты'] >= n/2].index[0]
n_me = df_stats.loc[median_interval_index, 'Частоты (n_j)']
sum_n_me_minus_1 = df_stats.loc[median_interval_index - 1, 'Накопл. частоты']
      if median_interval_index > 0 else 0
x0_me = bin_edges[median_interval_index]
median_val_grouped = x0_me + dx_rounded * (0.5 * n - sum_n_me_minus_1) / n_me

print("--- Числовые характеристики ---")
print(f"Математическое ожидание: {mean_val:.2f}")
print(f"Дисперсия (несмещенная): {variance_val:.2f}")
print(f"Среднеквадратическое отклонение: {std_dev_val:.2f}")
print(f"Коэффициент асимметрии: {skewness_val:.2f}")
print(f"Эксцесс: {kurtosis_val:.2f}")
print(f"Коэффициент вариации: {variation_coeff_val:.2f}")
print(f"Мода (групп.): {mode_val_grouped:.2f}")
print(f"Медиана (групп.): {median_val_grouped:.2f}")

# --- Проверка гипотезы по критерию Пирсона (χ²) ---
alpha = 0.05  # Уровень значимости

# Теоретические вероятности для экспоненциального распределения
# CDF(x) = 1 - exp(-lambda*x)
exp_cdf = lambda x, l: 1 - np.exp(-l * x)
theoretical_probs = []
for i in range(len(bin_edges) - 1):
    prob = exp_cdf(bin_edges[i+1], lambda_param) - exp_cdf(bin_edges[i],
      lambda_param)
    theoretical_probs.append(prob)

theoretical_freqs = np.array(theoretical_probs) * n

# Таблица для расчета
df_chi2 = pd.DataFrame({
    'Границы': df_stats['Границы интервалов'],
    'n_j': df_stats['Частоты (n_j)'],
    'n_j_theor': theoretical_freqs
})

# --- Объединение интервалов с n_j_theor < 5 ---
i = len(df_chi2) - 1
while i > 0:
    if df_chi2.loc[i, 'n_j_theor'] < 5:
        # Объединение текущей строки с предыдущей
        df_chi2.loc[i-1, 'n_j'] += df_chi2.loc[i, 'n_j']
        df_chi2.loc[i-1, 'n_j_theor'] += df_chi2.loc[i, 'n_j_theor']
        # Обновление границы
        new_boundary = f"[{df_chi2.loc[i-1, 'Границы'].split(';')[0][1:]};
      {df_chi2.loc[i, 'Границы'].split(';')[1][:-1]})"
        df_chi2.loc[i-1, 'Границы'] = new_boundary
        # Удаление текущей строки
        df_chi2 = df_chi2.drop(i)
    i -= 1
# Сброс индексов после удаления строк
df_chi2 = df_chi2.reset_index(drop=True)

df_chi2['(n-n\')^2/n\''] = (df_chi2['n_j'] - df_chi2['n_j_theor'])**2 /
      df_chi2['n_j_theor']
```

```
print("--- Расчет критерия Пирсона (χ²) ---")
print(df_chi2.to_string())
print("\n")

# Расчет наблюдаемого и критического значений
chi2_observed = df_chi2['(n-n\')^2/n\''].sum()
k_prime = len(df_chi2) # Число интервалов после объединения
m = 1 # Число оцененных параметров (lambda)
df = k_prime - m - 1 # Число степеней свободы
chi2_critical = stats.chi2.ppf(1 - alpha, df)

print("--- Результаты проверки гипотезы ---")
print(f"Наблюдаемое значение хи-квадрат = {chi2_observed:.2f}")
print(f"Число степеней свободы r = {k_prime} - {m} - 1 = {df}")
print(f"Критическое значение хи-квадрат = {chi2_critical:.2f}")

# Вывод
if chi2_observed < chi2_critical:
    print(f"Вывод: Так как {chi2_observed:.2f} < {chi2_critical:.2f}, нулевая
      гипотеза H0 не отвергается.")
    print("Эмпирическое распределение согласуется с теоретическим
      экспоненциальным законом.")
else:
    print(f"Вывод: Так как {chi2_observed:.2f} >= {chi2_critical:.2f},
      нулевая гипотеза H0 отвергается.")
    print("Эмпирическое распределение НЕ согласуется с теоретическим
      экспоненциальным законом.")
```