

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ
ФЕДЕРАЦИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ
ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«БЕЛГОРОДСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНОЛОГИЧЕСКИЙ
УНИВЕРСИТЕТ им. В. Г. ШУХОВА» (БГТУ им. В.Г. Шухова)
Кафедра программного обеспечения
вычислительной техники и автоматизированных
систем

Лабораторная работа №1
по дисциплине: Анализ Данных

Выполнил: студент группы ВТ-231
Масленников Д. А.
Проверили:
Кабальянц П. С.

Лабораторная работа № 1

Возьмите открытые данные о росте числа зараженных COVID-19 в мире за три месяца. Первый месяц из трех определяется сдвигом на номер по списку с февраля 2020 года. Проверьте гипотезу о том, что этот рост описывается экспоненциальной функцией. Полученное р-значение сравните с уровнем значимости 0,05.

Решение:

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

data = pd.read_csv('owid-covid-data.csv')

data['date'] = pd.to_datetime(data['date'])

start_date = '2021-04-01'
end_date = '2021-06-30'

filtered_data = data[(data['date'] >= start_date) & (data['date'] <= end_date)]

global_data = filtered_data.groupby('date')['total_cases'].sum().reset_index()

global_data['log_total_cases'] = np.log(global_data['total_cases'])

global_data = global_data.dropna(subset=['log_total_cases'])

slope, intercept, r_value, p_value, std_err = linregress(global_data.index, global_data['log_total_cases'])

# Вывод результатов
print(f"Коэффициент наклона (slope): {slope}")
print(f"Коэффициент корреляции (r_value): {r_value}")
print(f"p-значение: {p_value}")

alpha = 0.05
if p_value < alpha:
    print("Гипотеза о экспоненциальном росте отвергается (p-значение < 0.05)")
else:
    print("Гипотеза о экспоненциальном росте не отвергается (p-значение >= 0.05)")

x = global_data.index
y_real = global_data['total_cases']
y_exp = np.exp(intercept + slope * x)

plt.figure(figsize=(10, 6))
plt.plot(global_data['date'], y_real, label='Реальные данные', linestyle='-', linewidth=2)
plt.plot(global_data['date'], y_exp, label='Экспоненциальная модель',
```

```

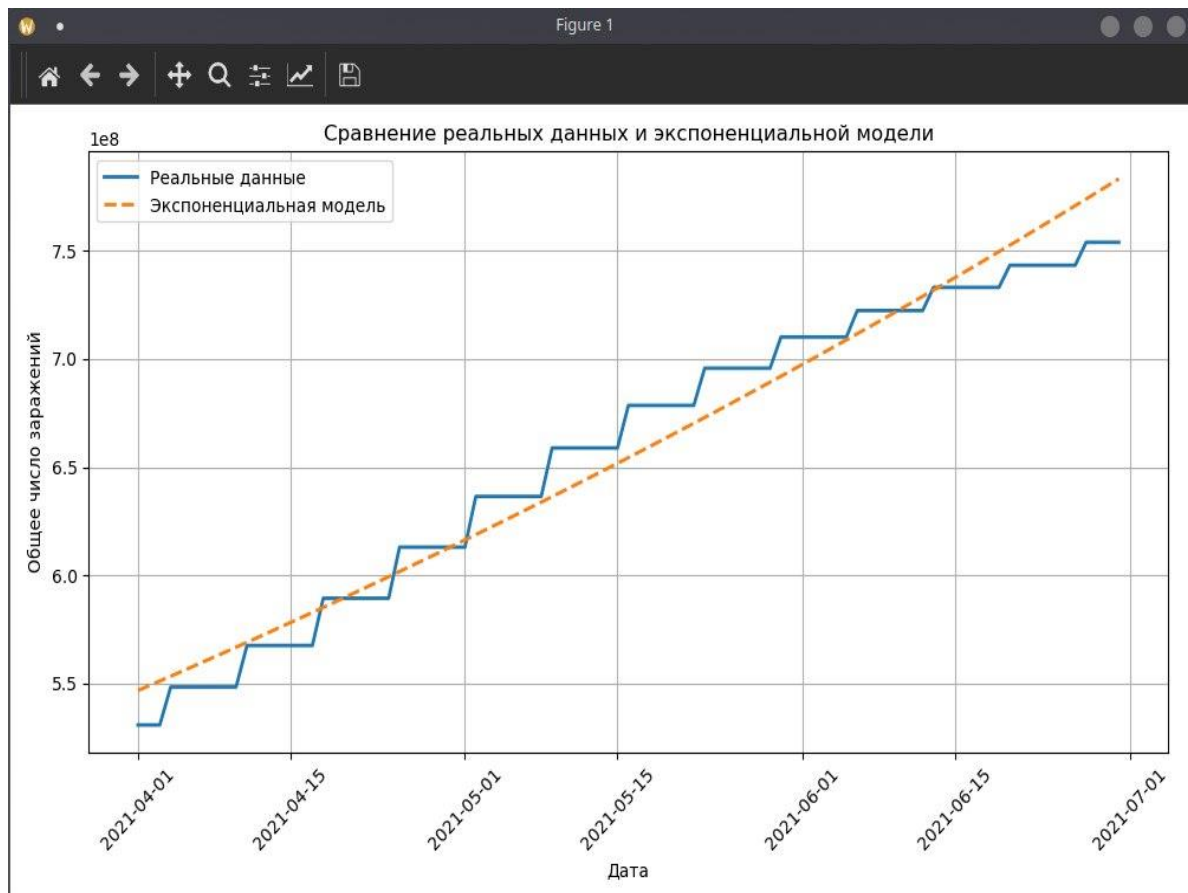
linestyle '--', linewidth 2)
plt.xlabel('Дата')
plt.ylabel('Общее число заражений')
plt.title('Сравнение реальных данных и экспоненциальной модели')
plt.legend()
plt.grid(True)
plt.xticks(rotation 45)
plt.tight_layout()
plt.show()

```

```

> python3 nlr.py
Коэффициент наклона (slope): 0.003995077469683306
Коэффициент корреляции (r_value): 0.9820020277119358
p-значение: 3.253253185816641e-66
Гипотеза о экспоненциальном росте отвергается (p-значение < 0.05)

```



Вывод: На основе анализа данных о росте числа заражений COVID-19 за три месяца (апрель–июнь 2021 года) гипотеза о том, что рост описывается экспоненциальной функцией, была проверена с использованием линейной регрессии. Полученное p-значение меньше уровня значимости 0.05, что позволяет отвергнуть гипотезу об экспоненциальном росте, а визуальное сравнение реальных данных и экспоненциальной модели подтверждает это расхождение.