Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Московский государственный технический университет имени Н.Э. Баумана
(национальный исследовательский университет)»
(МГТУ им. Н.Э. Баумана)

ФАКУЛЬТЕТ *Робототехники и комплексной автоматизации*

КАФЕДРА *Системы автоматизированного проектирования (РК-6)*

ОТЧЕТ О ВЫПОЛНЕНИИ ЛАБОРАТОРНОЙ РАБОТЫ

по дисциплине: «Вычислительная математика»

Студент	Абдулаев Баттал Халилович
Группа	РК6-61Б
Тип задания	лабораторная работа
Тема лабораторной работы	Нелинейная регрессия

Студент ____________________ **Абдулаев Б.Х.**
подпись, дата *фамилия, и.о.*

Преподаватель ____________________ **Соколов А.П.**
подпись, дата *фамилия, и.о.*

Оценка ______________________________

Москва, 2020 г.

Оглавление

Задание на лабораторную работу

Дана модель экспоненциального роста:

$$\frac{d}{dt}I = (\beta - \gamma)I, \tag{1}$$

где I – количество инфицированных людей, β– среднее число контактов, приходящихся на человека в единицу времени, и γ – это среднее число выздоровевших, приходящихся на человека в единицу времени (т.е., средняя скорость иммунизации).

Также дана SIS модель:

$$\frac{d}{dt}I = (\beta - \gamma)I - \frac{\beta}{N}I^2 \tag{2}$$

Где N – число людей в популяции (например, число жителей страны). Рассматриваемая страна – Австрия.

Требуется:

1. Найти решения экспоненциального роста и SIS-модели при условии, что $I(0) = I_0$ и продемонстрировать детальный вывод этого решения. После привести решение модели экспоненциального роста к форме:

   $$I(t) = I_0 e^{\chi t}, \tag{3}$$

   где $\chi = \beta - \gamma$, и решение SIS-модели к форме:

   $$I(t) = \frac{I_\infty}{1 + \left(\frac{I_\infty}{I_0} - 1\right) e^{-\chi t}}, \tag{4}$$

   где I_∞ обозначает предельное значение $I(t)$ при $t \to \infty$, т.е.

   $I_\infty = lim_{t \to \infty} I(t)$.

2. Сформировать выборку данных зависимости числа, инфицированных от времени в данной стране.
3. Предполагая, что в первые недели распространения вируса в выбранной стране рост числа зараженных описывается экспоненциальной моделью, требуется оценить значение коэффициента $\chi = \beta - \gamma$ с помощью нелинейной регрессии и нормального уравнения, взяв в качестве аппроксимирующей функции решение экспоненциальной модели и начального числа инфицированных I_0 = 20...30 и вывести на экран в логарифмической шкале полученное решение вместе с исходными дискретными данными о числе инфицированных. Требуется подробно описать формулировку задачи регрессии и явно указать выражения для отдельных векторов и матрицы, входящих в нормальное уравнение.
4. Подставив найденное значение коэффициента $\chi = \beta - \gamma$ в решение SIS-модели, оценить значение коэффициента I_∞ с помощью нелинейной регрессии и нормального уравнения, взяв в качестве аппроксимирующей функции решение SIS- модели и вывести на экран полученное решение вместе с исходными дискретными данными о числе инфицированных. Требуется подробно описать формулировку задачи регрессии и явно указать выражения для отдельных векторов и матрицы, входящих в нормальное уравнение.
5. Ответить на вопросы:
   (a) Какова погрешность полученной аппроксимации (решения SIS-модели) относительно нормы L_∞? Относительно нормы L_2?
   (b) Какой критерий использовался бы для определения максимального числа дней, в течение которых использование модели экспоненциального роста оправдано? Каково такое максимальное число дней в соответствии с критерием?

(c) Предсказать, сколько человек в соответствии с полученной аппроксимацией (решение SIS-модели) будут инфицированы в выбранной стране при $t \to \infty$ и через сколько дней после начала эпидемии наступит её окончание.

Цель выполнения лабораторной работы

Цель выполнения лабораторной работы: с помощью модели экспоненциального роста и SIS-модели охарактеризовать и, по возможности, предсказать распространение инфекции COVID-19 на территории Австрии.

Выполненные задачи

1. Нахождение решений модели экспоненциального роста и SIS модели.
2. Загружена и отфильтрована выборка данных из csv файла по распространению CODIV-19.
3. Оценка коэффициента χ. Вывод решения экспоненциальной модели и дискретных данных.
4. Оценка коэффициентов I_∞ и I_0. Вывод графика решения SIS-модели и дискретных данных.
5. Рассчитана погрешность полученной из нормального уравнения аппроксимации.
6. Определён и выявлен критерий, определяющий максимальное число дней применения экспоненциальной модели.
7. Анализ эпидемии.

1. Решения моделей экспоненциального роста и SIS – модели.

Для того чтобы решить дифференциальное уравнение модели экспоненциального роста (1), используем метод разделения переменных. Вместо $I(t)$ будет использована I, подразумевая, что численность I есть функция от времени.

$$\frac{dI}{(\beta-\gamma)I} = dt,$$

$$\frac{dI}{I} = (\beta-\gamma)dt$$

Возьмем интегралы от обеих частей дифференциального уравнения:

$$\int \frac{dI}{I} = \int (\beta-\gamma)dt$$

$$ln|I| = ((\beta-\gamma)t + C',$$

$$I = Ce^{(\beta-\gamma)t},$$

где $C = e^{C\prime}$

Константу определим из начального условия I (0) = I_0. Следовательно, получаем, что C = I_0. Подставляем С в зависимость $I(t)$ и производим замену χ = β – γ, получаем решение, совпадающее с формулой (3):

$$I(t) = I_0 e^{\chi t}$$

Для решения дифференциального уравнения SIS- модели (2), воспользуемся заменами $I = \frac{1}{x}$.

$$\frac{d\left(\frac{1}{x}\right)}{dt} = \frac{(\beta - \gamma)}{x} - \frac{\beta}{N}\left(\frac{1}{x}\right)^2,$$

Разделим переменные:

$$\frac{d\left(\frac{1}{x}\right)}{\frac{(\beta - \gamma)}{x} - \frac{\beta}{N}\left(\frac{1}{x}\right)^2} = dt$$

$$\frac{-\left(\frac{1}{x^2}\right)dx}{\frac{(\beta - \gamma)}{x} - \frac{\beta}{N}\left(\frac{1}{x}\right)^2} = dt$$

$$\frac{-dx}{(\beta - \gamma)x - \frac{\beta}{N}} = dt$$

$$\frac{dx}{\frac{\beta}{N} - (\beta - \gamma)x} = dt$$

Возьмем интеграл от обоих частей:

$$\int \frac{dx}{\frac{\beta}{N} - (\beta - \gamma)x} = \int dt$$

$$-\frac{1}{\beta - \gamma} ln\left|\frac{\beta}{N} - (\beta - \gamma)x\right| = t + C$$

$$ln\left|\frac{\beta}{N} - (\beta - \gamma)x\right| = -(\beta - \gamma)t + C$$

Совершаем обратную замену на переменную I:

$$ln\left|\frac{\beta}{N} - \frac{(\beta - \gamma)}{I}\right| = -(\gamma - \beta)t + C$$

$$\frac{\beta}{N} - \frac{(\beta-\gamma)}{I} = e^{-(\beta-\gamma)(t+C)}$$

Получим зависимость $I(t)$:

$$I(t) = \frac{\beta-\gamma}{\frac{\beta}{N} - e^{-(\beta-\gamma)(t+C)}}$$

Находим значение константы C исходя из условия $I(0) = I_0$:

$$C = -\frac{\ln\left(\frac{\beta}{N} - \frac{\beta-\gamma}{I_0}\right)}{\beta-\gamma}$$

Подставим С в зависимость $I(t)$:

$$I(t) = \frac{\beta-\gamma}{\frac{\beta}{N} - \left(\frac{\beta}{N} - \frac{\beta-\gamma}{I_0}\right) e^{(\gamma-\beta)t}}$$

Выражение для I_∞:

$$I_\infty = \frac{N|\chi|}{\beta}$$

Умножим числитель и знаменатель в зависимости $I(t)$ на $\frac{N}{\beta}$ и выполним замены χ и I_∞:

$$I(t) = \frac{I_\infty}{1 - \left(1 - \frac{I_\infty}{I_0}\right) e^{-\chi t}}$$

Для того чтобы привести полученную зависимость к виду (4), вынесем минус за скобки в знаменателе:

$$I(t) = \frac{I_\infty}{1 + \left(\frac{I_\infty}{I_0} - 1\right) e^{-\chi t}}$$

2. Формирование выборки данных из файла в формате csv

Были получены данные об эпидемиологической ситуации в Австрии. Первый случай заражения COVID-19 был выявлен 26 февраля 2020 года. Более подробные данные были представлены в таблице 1[1]

Таблица 1: Статистика инфицированных в Австрии.

Дата	Число заболевших	Дата	Число заболевших	Дата	Число заболевших
2020-02-26	2	2020-03-17	1016	2020-04-06	11983
2020-02-27	2	2020-03-18	1332	2020-04-07	12297
2020-02-28	5	2020-03-19	1646	2020-04-08	12640
2020-02-29	7	2020-03-20	2196	2020-04-09	12969
2020-03-01	10	2020-03-21	2649	2020-04-10	13248
2020-03-02	14	2020-03-22	3024	2020-04-11	13560
2020-03-03	18	2020-03-23	3631	2020-04-12	13807
2020-03-04	24	2020-03-24	4486	2020-04-13	13937
2020-03-05	29	2020-03-25	5282	2020-04-14	14043
2020-03-06	41	2020-03-26	5888	2020-04-15	14234
2020-03-07	74	2020-03-27	7029	2020-04-16	14370
2020-03-08	99	2020-03-28	7697	2020-04-17	14448
2020-03-09	102	2020-03-29	8291	2020-04-18	14603
2020-03-10	131	2020-03-30	8813	2020-04-19	14662
2020-03-11	182	2020-03-31	9618	2020-04-20	14710
2020-03-12	246	2020-04-01	10182	2020-04-21	14783
2020-03-13	361	2020-04-02	10711	2020-04-22	14833
2020-03-14	504	2020-04-03	11129	2020-04-23	14924
2020-03-15	655	2020-04-04	11525	2020-04-24	14985
2020-03-16	860	2020-04-05	11766	2020-04-25	15068

Данные таблицы выведены на рисунке 1.

[1] Таблица числа зараженных взята из электронного источника: https://ourworldindata.org/coronavirus- source-data

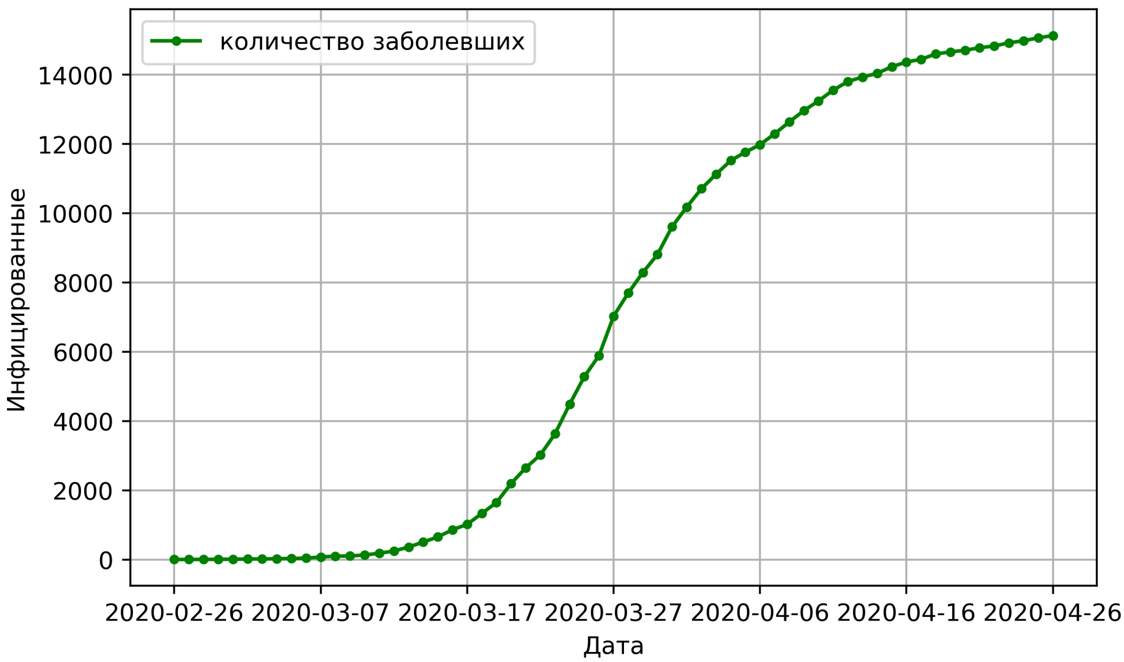

Рисунок 1 – Зависимость числа заражённых в Австрии.

3. Оценка коэффициента χ. Вывод решения экспоненциальной модели и дискретных данных.

Для того чтобы отсечь начальный период пандемии, выберем начальное число инфицированных равным 10. Период, который описывает экспоненциальную зависимость, выберем 25 дней.

Для оценки коэффициента χ воспользуемся нормальным уравнением.

$$\chi = (X^T X)^{-1} X^T \tilde{y}, \tag{5}$$

где

$$X = \begin{Bmatrix} t_1 \\ t_2 \\ \dots \\ t_n \end{Bmatrix} \tag{6}$$

В транспонированном виде:

$$X^T = \{t_1, t_2, \dots, t_n\} \tag{7}$$

Получим более удобную форму параметра χ с учетом (6), (7):

$$\chi = \left(\sum_{i=1}^{n} t_i^2\right)^{-1} \left(\sum_{i=1}^{n} t_i \tilde{y}_i\right) = \frac{\sum_{i=1}^{n} t_i \tilde{y}_i}{\sum_{i=1}^{n} t_i^2} \tag{8}$$

Так как параметр χ входит в формулу (3) линейно, поэтому изменим вид уравнения взял логарифм от обеих частей выражения:

$$\ln\big(I(t)\big) = \ln(I_0) + \chi\mathrm{t} \tag{9}$$

$$\ln\left(\frac{I(t)}{I_0}\right) = \chi\mathrm{t} \tag{10}$$

Аппроксимирующая функция будет иметь вид:

$$\hat{I} = \chi\mathrm{t} \tag{11}$$

Выборка данных о числе инфицированных имеет следующий вид:

$$Y = \{(y_i)\}_{i=1}^{n} \tag{12}$$

Чтобы оценить квадрат отклонения, необходимо преобразовать Y, к такому же виду, что и I:

$$\tilde{y} = \ln(Y) - \ln(I_0) \tag{13}$$

Таким образом метод наименьших квадратов будет иметь следующий вид:

$$\min_{\chi} \sum_{i=1}^{n} \left(\tilde{y}_i - \hat{I}_i(\chi, \mathrm{t})\right)^2 \tag{14}$$

По формуле (8) было получено значение χ равное 0.2939

Подставив коэффициент χ в решение для экспоненциальной модели (3), получим график зависимости аппроксимирующей функции, представленный на рисунке 2.

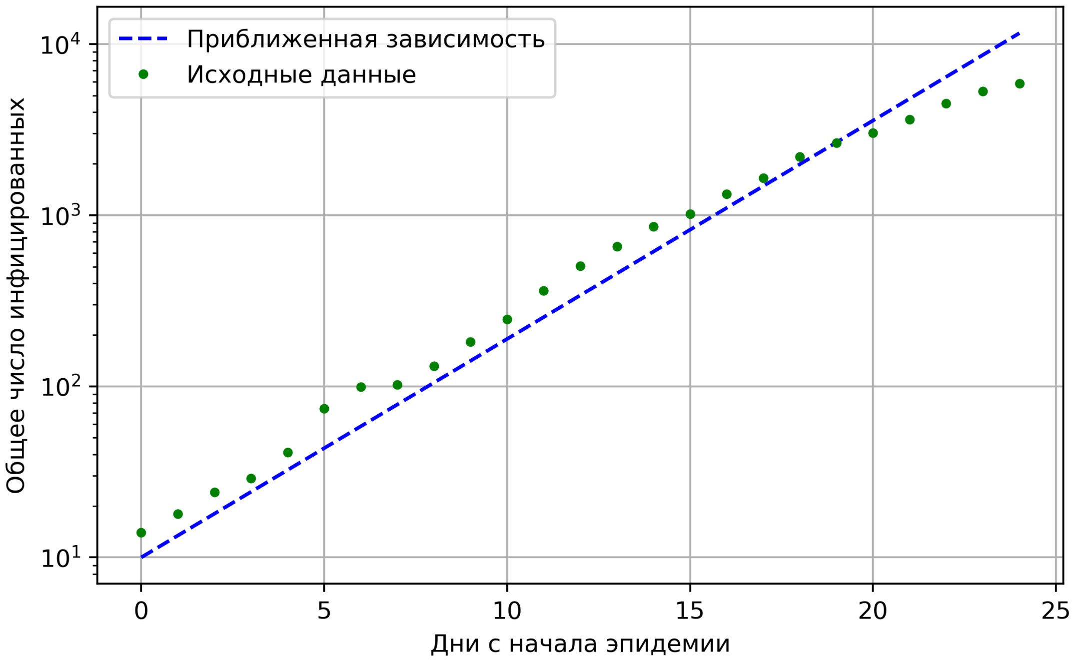

Рисунок 2 – Экспоненциальная модель эпидемии и исходные данные.

Как видно из рисунка 2, на рассматриваемом временном промежутке распространение вируса описывается экспоненциальной моделью.

4. Оценка коэффициентов I_∞ и I_0. Вывод графика решения SIS-модели и дискретных данных.

Для того чтобы получить решения SIS – модели (4), подставим найденное значение $\chi = 0.2939$. Чтобы найти оптимальные I_∞ и I_0 используем библиотеку scipy. С помощью функции scipy.optimize.basinhopping( ) найдем такие коэффициенты, при которых SIS-модель наиболее точно аппроксимирует исходные данные.

Получены следующие значения:

$$I_0 = 9$$

$$I_\infty = 13226$$

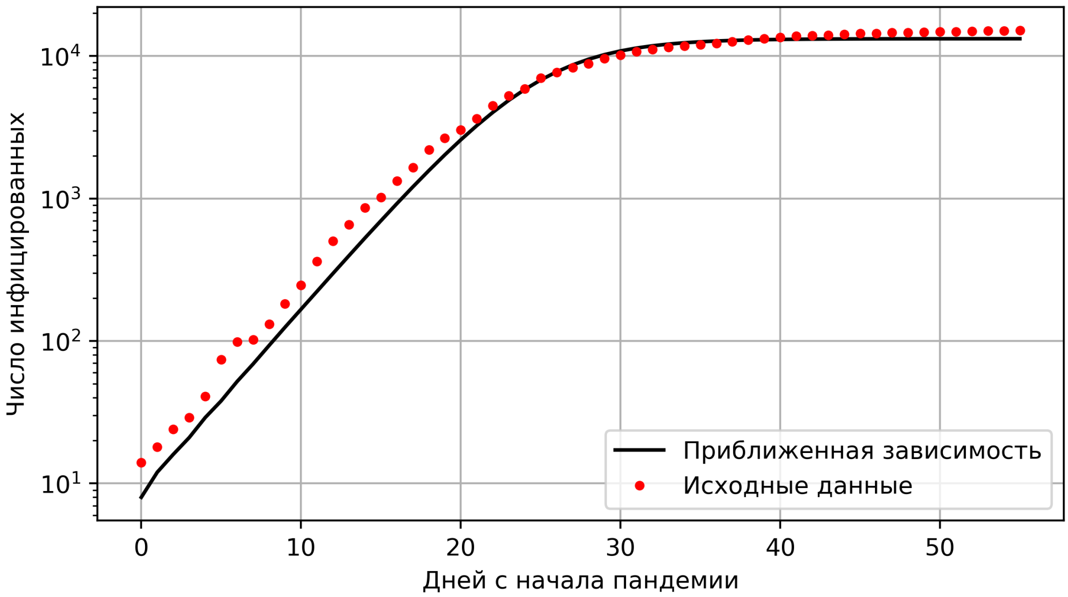

Рисунок 3 – Решение SIS-модели с исходными данными.

По рис. 3 видно, что данная зависимость достаточно точно аппроксимирует исходную.

5. Погрешность полученной из нормального уравнения аппроксимации.

Для оценки погрешности аппроксимации найдем расстояния относительно нормы L_2 и L_∞. Расстояние относительно нормы L_2 можно найти по следующей формуле:

$$\|I(t) - y(t)\|_2 = \left[\frac{1}{n}\sum_t (I(t) - y(t))^2\right]^{\frac{1}{2}} \quad (15)$$

Значение погрешности относительно нормы $L_2 = 809$

Погрешность решения SIS – модели относительно нормы L_∞

$$\|I(t) - y(t)\|_\infty = \max_t |I(t) - y(t)| \qquad (16)$$

Значение погрешности относительно нормы L_∞= 1911

6. Критерий использования модели экспоненциального роста

Для того чтобы сформулировать критерий, построим график зависимости прироста числа инфицированных от количества дней, прошедших с начала пандемии.

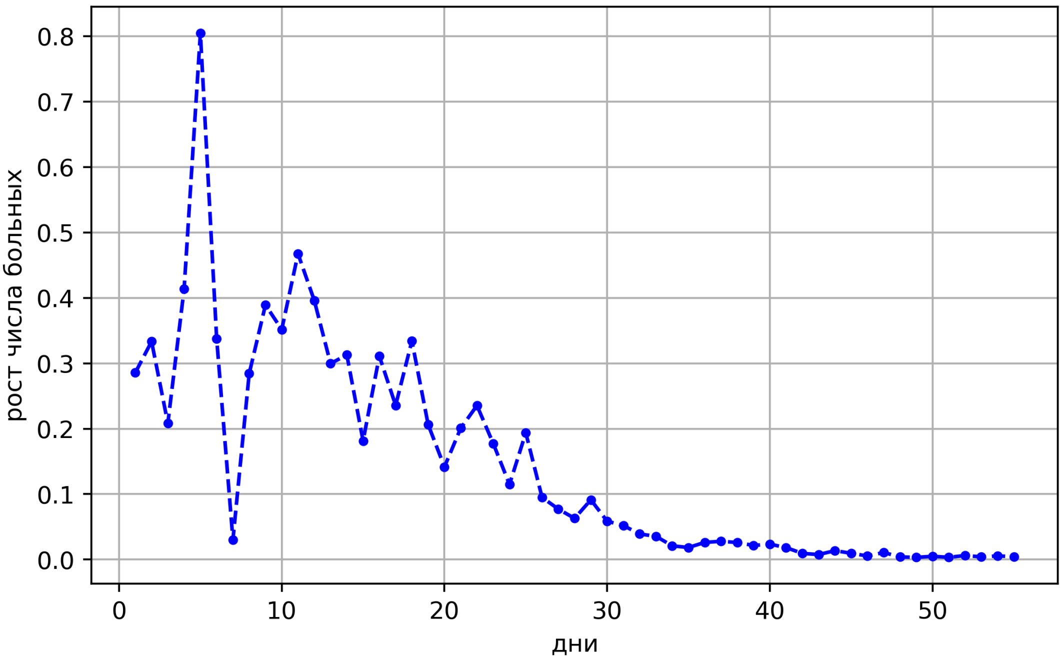

Рисунок 4 – Зависимость прироста заражённых от дней.

Из графика на рис. 4 видно, что в первые дни рост числа больных большой. В это период экспоненциальная модель может хорошо описать данные. Но дальше прирост падает, и данная модель уже не подходит, т.к. экспоненциальная модель хорошо описывает пандемию в период ее роста.

Это связано с тем, что данная модель не учитывает, что большинство инфицированных выздоравливают. Максимальное число дней применения экспоненциальной модели было выявлено на основании построенных графиков (рис. 2, рис. 3). В течение первых 25 дней экспоненциальная модель довольно хорошо аппроксимировала данные, чего нельзя сказать о следующем интервале времени. SIS-модель, напротив, начиная с 25 дня, очень точно аппроксимирует исходные данные (рис.3). Функция этой модели с этого дня практически лежит на функции числа заболевших, что говорит о больших возможностях прогнозирования. Таким образом, критерий был выбран из графического представления предложенных аппроксимаций.

7. Анализ эпидемии

В примечании было сказано считать, что эпидемия заканчивается, когда прирост общего числа заболевших за один день становится меньше 1% от общего числа заболевших за предыдущий день:

Листинг 1. Реализация поиска дня, когда эпидемия завершится.

```
def prognoz(I_0, I_inf):
    i = np.arange(1, 100)
    I_sis_inf = func_SIS(I_0,I_inf, chi,i)
    for i in range(len(I_sis_inf) - 1):
        if I_sis_inf[i] * 1.01 < I_sis_inf[i + 1]:
            end = i
    return end
```

Предельное количество человек, которые будут инфицированы в Австрии, определяется предельным значением $I(t)$, которое достигает полученная аппроксимация при $t \to \infty$. Это значение равно оптимальному коэффициенту I_∞, величина которого была вычислена с помощью метода нелинейной регрессии: I_∞=13225.

Заключение

В процессе выполнения лабораторной работы на примере данных о распространении вируса COVID-19 в Австрии были исследованы модели распространения эпидемий: экспоненциальная модель и SIS-модель.

Было выяснено, что экспоненциальная модель, как правило, применяется для исследования эпидемии в самом её начале и позволяет получить довольно точную аппроксимацию. А SIS-модель применяется для изучения динамики эпидемии на любом её этапе, однако стоит отметить, что аппроксимация с помощью данной модели дает значительные погрешности. Объясняется это тем, что SIS- модель предназначена для оценки «естественного» развития эпидемии, в то время как карантин оказывает значительное влияние на ее ход.

Исследование полученных аппроксимаций показало, что пандемия в Австрии закончится через 34 дня после начала, это означает, что в настоящее время количество новых случаев заражения уменьшается с каждым днем. Также, решение SIS-модели предсказало, что максимальное число случаев заболевания будет равно 13225, что меньше, чем действительное число заболевших (к 25.04.2020) примерно на 1800.

Список использованных источников

1. **ПершинА.Ю.** *Лекции по вычислительной математике.* [Электронный ресурс] // МГТУ им. Н. Э. Баумана, 2020. – 145с.

2. **ПершинА.Ю.** *Семинар 4 по курсу «Вычислительная математика».* [Электронный ресурс] // МГТУ им. Н. Э. Баумана, 2020. – 43с.