**Федеральное агентство связи**

**Ордена Трудового Красного Знамени федеральное государственное бюджетное**

**образовательное учреждение высшего образования**

**«Московский технический университет связи и информатики»**

**Заочный Факультет**

Практическое Задание

на тему

«**Прикладная статистика в анализе данных**»

**Преподаватель**: проф. Саксонов Е.А.

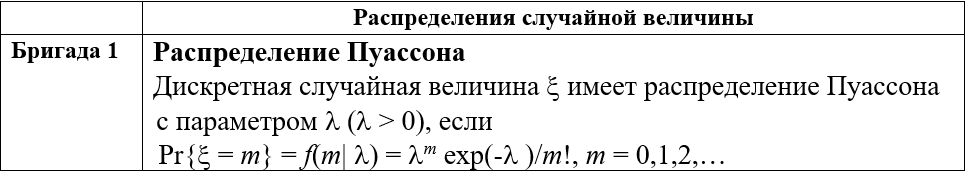
Выполнили студенты **«Бригады-1»**, групп **ЗМБД-2031** и **ЗМБД-2032**:

Бергман Елена  
Комаров Михаил

Черевко Владимир

**Москва 2022 г.**

# Задание 1 – Получение случайных значение с заданным законом распределения

1. Получить выборки случайных чисел с непрерывными и дискретными распределениями. Объем выборки - 50. Для получения выборки использовать генератор случайных чисел, равномерно распределенных на интервале [0,1].  
   
2. Для полученных выборок построить гистограммы.
3. Для полученных выборок провести первичную обработку статистики.
4. Для каждого распределения, полученного с применение генератора случайных чисел (первая практическая работа) подобрать максимально правдоподобные оценки его параметров по результатам полученных выборок.

**Решение.**

Распределение Пуассона является дискретным.

Для генерации случайных чисел из распределения Пуассона будем использовать язык Python и библиотеки numpy и scipy. Функция позволяющая сгенерировать случайное число из данного распределения с заданным значением лямбда называется poisson(lambda=int\_num).

Получим (Рис. 1.1) набор случайных чисел из распределения Пуассона для λ равного 8.

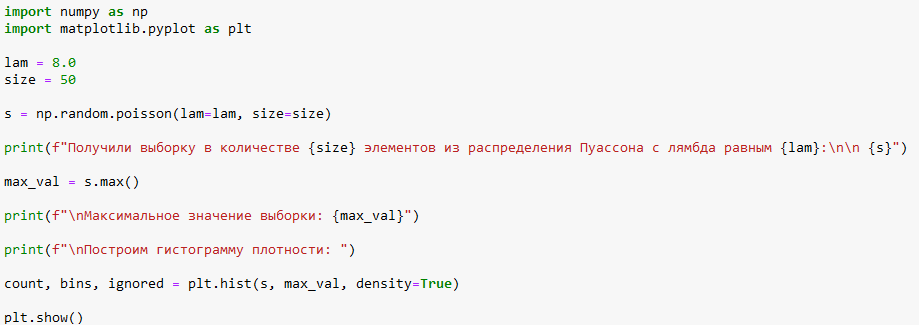


Рис. 1.1 – Получение выборки из распределения Пуассона посредствам языка python.

В результате, мы получили выборку, которая представлена в таблице 1.1.

Таблица 1.1 – Выборка из распределения Пуассона

|  |  |
| --- | --- |
| **Выборка из распределения Пуассона, λзаданное = 8, объём выб. = 50** | |
| **Значения** | **Частота** |
| 0 | 0 |
| 1 | 0 |
| 2 | 0 |
| 3 | 0 |
| 4 | 3 |
| 5 | 6 |
| 6 | 7 |
| 7 | 2 |
| 8 | 4 |
| 9 | 6 |
| 10 | 12 |
| 11 | 4 |
| 12 | 2 |
| 13 | 1 |
| 14 | 1 |
| 15 | 0 |
| 16 | 1 |
| 17 | 0 |
| 18 | 1 |
| **Объём:** | 50 |

Построим гистограмму частот для полученной выборки (Рис. 1.2).

Рис. 1.2 – Гистограмма частот выборки.

На Рис. 1.3 представлен код первичной обработки статистики для полученной выборки.

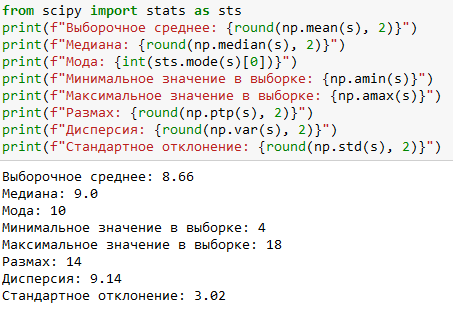


Рис. 1.3 – Первичная обработка статистики для выборки.

Подберём максимально правдоподобную оценку параметра λ (обозначим её как λ\*) для распределения Пуассона по результатам полученной из него выборки X.

Для этого, возьмём натуральный логарифм от функции Пуассона для набора из n элементов выборки.

Затем, получим от этого логарифма частную производную по λ и приравняем её к 0 (наши условия экстремума). Теперь, разрешим уравнение относительно λ\*. В результате, полученное в правой части выражение представляет собой выборочное среднее по X. Описанные преобразования приведены на Рис. 1.4.

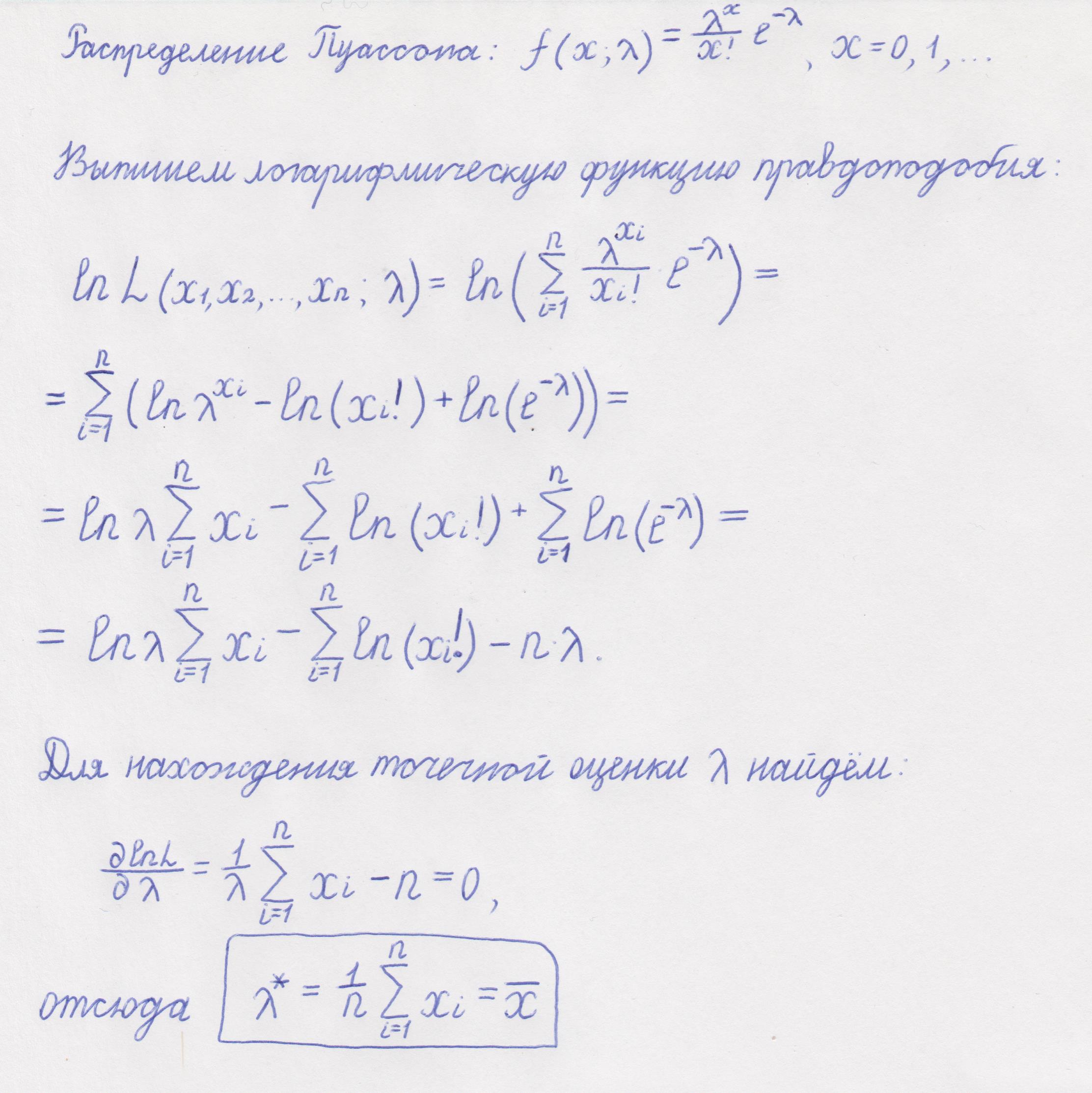


Рис. 1.4 – Вывод формулы для оценки макс. правдоподобия [1].

Согласно данным посчитанным на Рис. 1.3, выборочное среднее равно λ\* и равно 8,66. Результат достаточно близкий к λ заданному при генерации выборки, которая равна 8.

# Задание 2 – Метод наименьших квадратов

**Бригада 1.** Получены следующие результаты экспериментов:

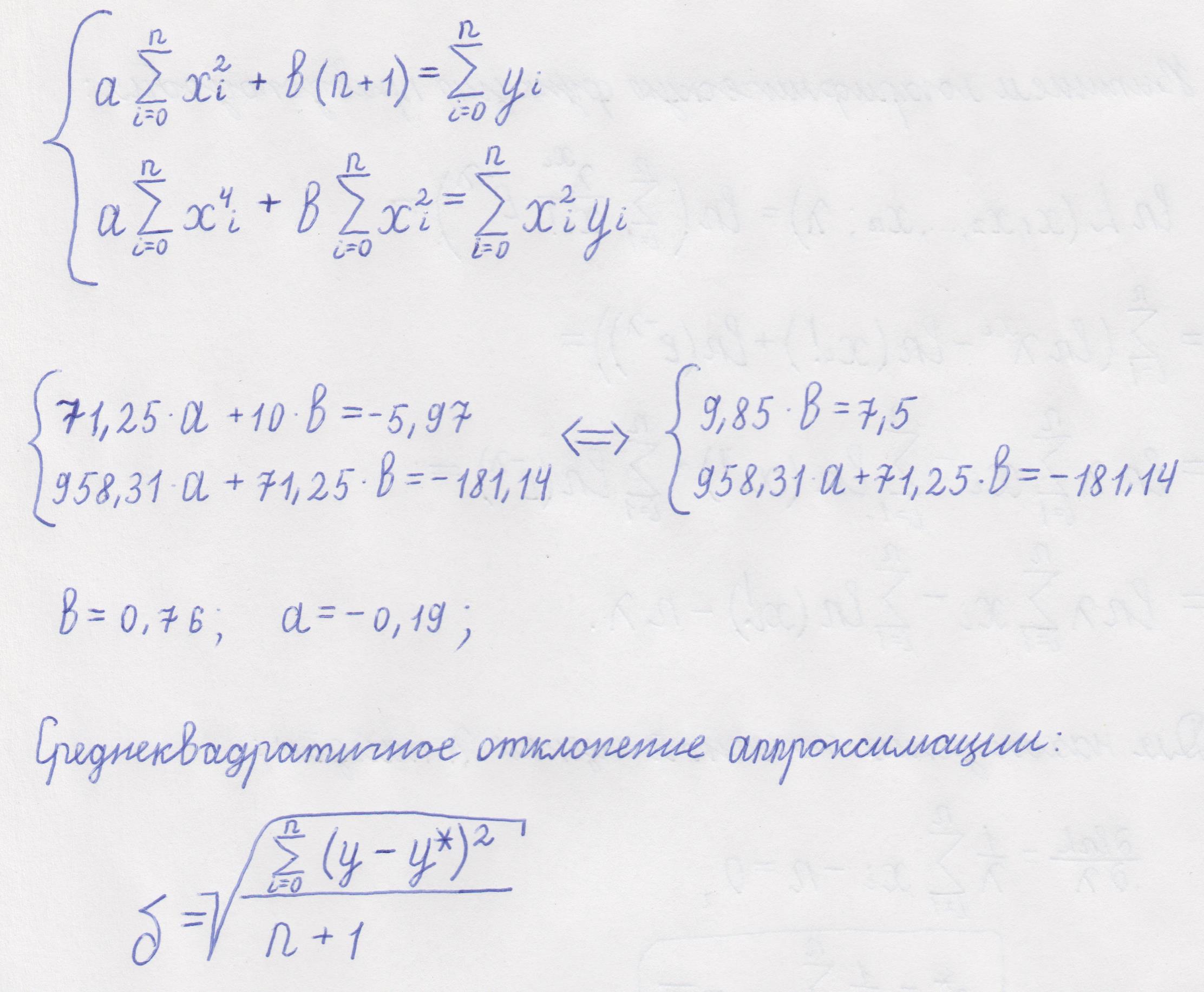
|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| xi | 0,0 | 0,5 | 1,0 | 1,5 | 2,0 | 2,5 | 3,0 | 3,5 | 4,0 | 4,5 |
| yi | 1,67 | 1,32 | 1,10 | 0,81 | 0,5 | 0,18 | -0,10 | -0,45 | -9,80 | -1,20 |

Найти коэффициенты аппроксимирующей функции: y = *a*x2 + b.

Вычислить суммарную квадратичную ошибку аппроксимации. Построить графики.

**Решение**.

Рассмотрим аппроксимирующую функцию y = ax2 + b. Слагаемое a1 при x1 у неё равно 0, следовательно, для подбора коэффициентов a и b методом МНК, из классической матрицы Грама для квадратного уравнения мы исключаем одну строку (частную производную по a1 приравненную к 0) и один столбец (для слагаемого a1). Итоговая матрица у нас будет вида (1), где n = 9.

 (1)

Составляющие системы уравнений вычислим посредствам таблицы excel (Рис. 2.1).

Таблица 2.1 – Вычисление элементов матрицы Грама.



Подставляем рассчитанные значения в матрицу вида (1) и находим её решение методом Гаусса (Рис. 2.1).

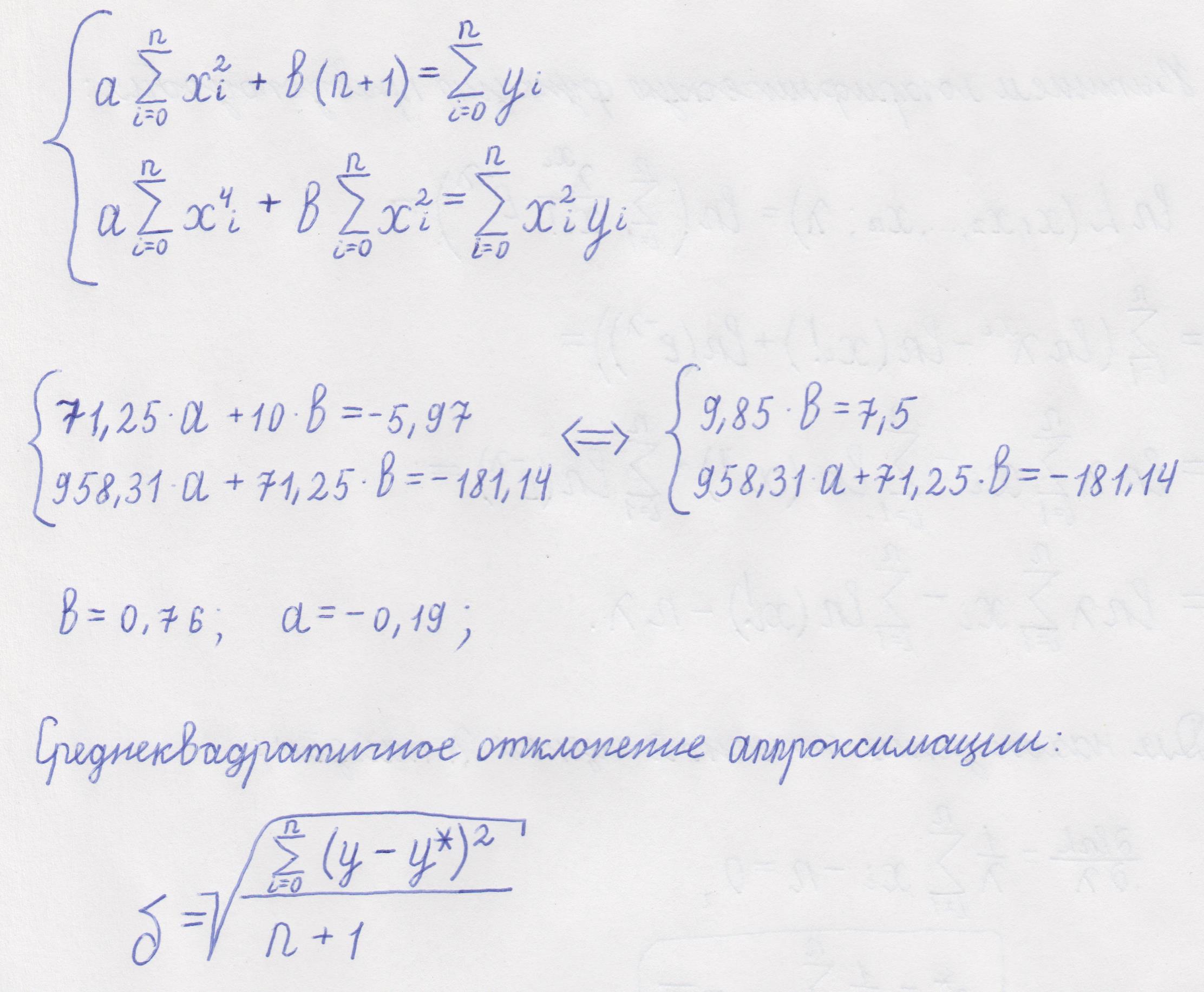


Рис. 2.1 – Решение матрицы методом Гаусса [3].

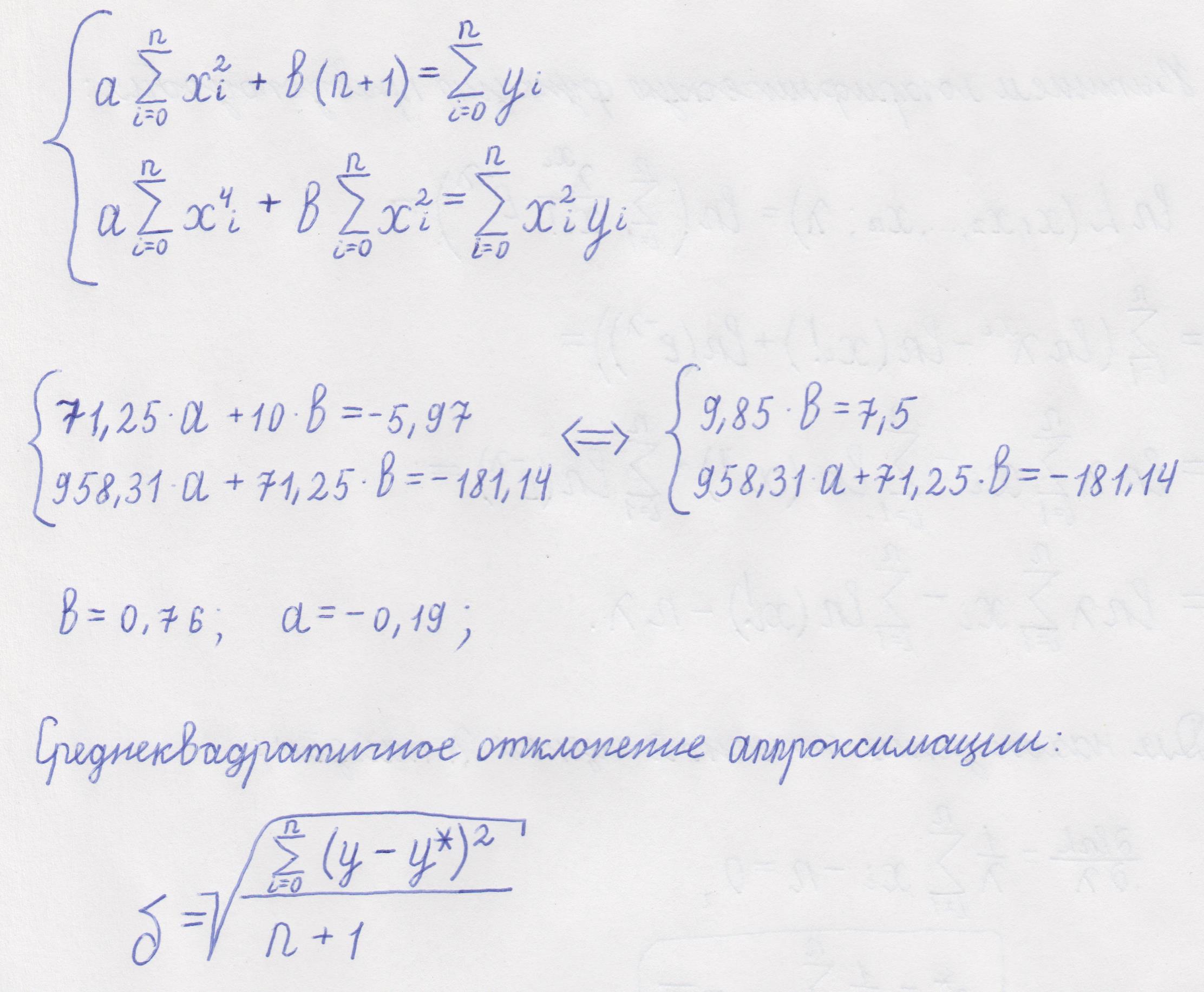
В результате, аппроксимирующая функция будет выглядеть следующим образом.

y = -0,19\*x2 + 0,76

Построим график аппроксимирующей функции и покажем на нём точками исходные значения.

Рис. 2.2 – Графики экспериментальной и аппроксимирующей функций.

Вычислим суммарное квадратичное отклонение аппроксимации по формуле (2). Воспользовавшись суммой квадратов разностей экспериментальной и аппроксимирующей функций, сведенных  
в таблицу 2.1.

 (2)

Таким образом, мы можем полагать, что наша аппроксимирующая функция отклоняется от исходной функции на (64,41/10)^(1/2) = 2,54.

# Задание 3 – Проверка гипотез

Каждая бригада при выполнении задания 1 получала по 50 случайных чисел с заданными законами распределения.

Требуется, применяя критерий согласия χ2, проверить гипотезу о том, что полученная выборка взята из генеральной совокупности с заданным распределением. При проверке гипотезы проводить разбиение на 7 – 9 интервалов.

**Решение.**

**Гипотеза H0** – Дискретная случайная величина X извлечена из генеральной совокупности, распределённой по закону Пуассона.

**Гипотеза H1** - Дискретная случайная величина X извлечена из генеральной совокупности, не подчиняющейся закону Пуассона.

Вычисления будем проводить в пакете MS Excel 2021.

Проверим гипотезу H0 при уровне значимости 0,05 согласно следующему плану [2].

1. Найдём по данному эмпирическому распределению выборочную среднюю, которую примем в качестве оценки параметра λ распределения Пуассона, обозначим её как **λнаблюдаемое**. Вычисление будем посредствам функции SUMPRODUCT(A3:A21;B3:B21)/B22 (см. Рис. 3.1).  
   λнаблюдаемое = 8,66.
2. Найдём посредствам функции POISSON.DIST(A3; $B$23; FALSE), реализующей формулу распределения Пуассона, вероятности **Pi(X)** данного i-го события в серии из n испытаний.
3. Найти теоретические частоты по формуле **fe = f0 \* Pi(X)** для набора из n испытаний.  
     
   Рис 3.2 – Получение теоретических частот для выборки X.
4. Произвести коррекцию интервалов принимаемых значений (аналогичную коррекции Йетса), чтобы теоретические частоты (посредствам их объединения) имели значения больше единицы. Для новых интервалов вычислить набор частот f0 и теоретических частот fe.   
   В результате, получилось 11 интервалов, обозначим их **s**. Результаты отображены на Рис. 3.2.
5. Сравнить эмпирические и теоретические частоты с помощью критерия Пирсона, приняв число степеней свободы:  
   k = s – 2 = 11 – 2 = 9.  
   По таблицам [1] находим значение правостороннего критического χ2 для чиста степеней свободы k = 9 и уровня значимости 0,05.  
   Итого, **χ2критич. правостор.** = 16,919.  
   Вычисляем **χ2наблюдаемое** как сумму значений (f0 - fe)2/fe для каждого из n испытаний.  
   Итого, **χ2наблюдаемое** = 15,0191.  
   Как можем видеть на рисунке ниже  
   **χ2наблюдаемое < χ2критич. правостор.**



Рис. 3.2 – Вычисление χ критериев и проверка гипотезы H0.

**Вывод.**

Нет оснований отвергать гипотезу H0 о распределении случайной величины X по закону Пуассона.

# Задание 4 – Моделирование

1. Найти площадь S заданной фигуры или ограниченную пересечением графиков функций на интервале [a,b]



Построить модель вычислений, провести 10 сеансов моделирования, в каждом сеансе по 100 реализаций случайных величин. Построить графики вычисляемых значений площади по сеансам моделирования: S1 , S2 ,…, S10 и средних значений от числа сеансов моделирования: S1, (S1+ S2)/2, (S1+ S2+ S3)/3, …,(S1+ S2+ S3+…+ S10)/10.

**Решение.**

Для того, чтобы найти площадь фигуры, ограниченной кривыми, воспользуемся методом Монте-Карло. Для этого нам нужно найти границы нашей фигуры и заключить их в прямоугольник с известной нам площадью. Затем, приняв стороны прямоугольника за границы интервала, следует сгенерировать на этом интервале 100 случайных чисел из равномерного распределения. Всего следует провести 10 таких экспериментов [4].

Полученные n = 100 значений для каждой из выборок X следует проанализировать на попадание в границы нашей целевой фигуры и сосчитать количество попаданий m из возможных n. Для нахождения площади целевой фигуры [2], следует умножить площадь прямоугольника на m и разделить на n.

Проанализируем заданные функции на предмет их пересечения на заданном интервале [0;3] (приравняв их друг к другу) и определим условия попадания случайных пар значений (x; y) в границы нашей фигуры. Вычисления представлены на (Рис. 4.1).

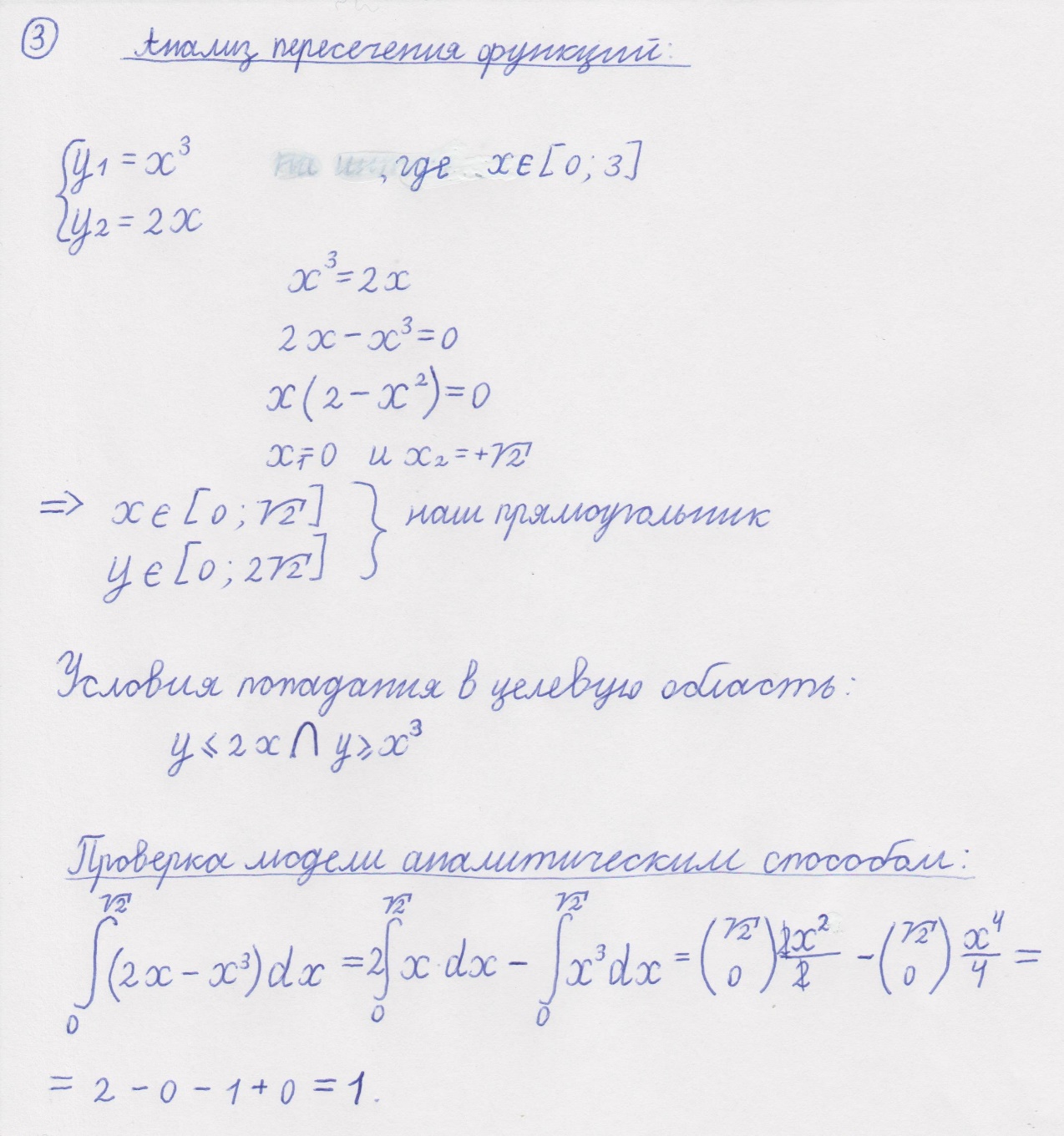


Рис. 4.1 – Анализ пересечения функций.

Для наглядности, ниже построен график, на котором можно видеть пересечение функций на заданном интервале.

Рис. 4.2 – График пересечения функций y1 и y2.

Изложенный выше план был реализован посредствам таблиц MS EXCEL. Ниже представлены 2 таблицы (Таблица 4.1 и 4.2) в которой описаны 10 выборок по 100 элементов в каждой, посчитаны оцениваемые площади для каждой. Также, посчитаны средние площади для каждого набора выборок и построены графики их зависимостей от количества выборок (Рис. 4.3). **График и таблицы можно в более крупном масштабе рассмотреть в одноимённом excel-файле, во вкладке «Задание-4»**.

Таблица 4.1 – Оценочные выборки с 1 по 5.



Таблица 4.2 – Оценочные выборки с 6 по 10.



Рис. 4.3 – Изменение значения площади и её среднего от выборки

Теперь, для проверки того, что мы правильно произвели моделирование, посчитаем площадь нашей фигуры аналитическим способом посредствам вычисления определённого интеграла [3]  
 (Рис. 4.4).

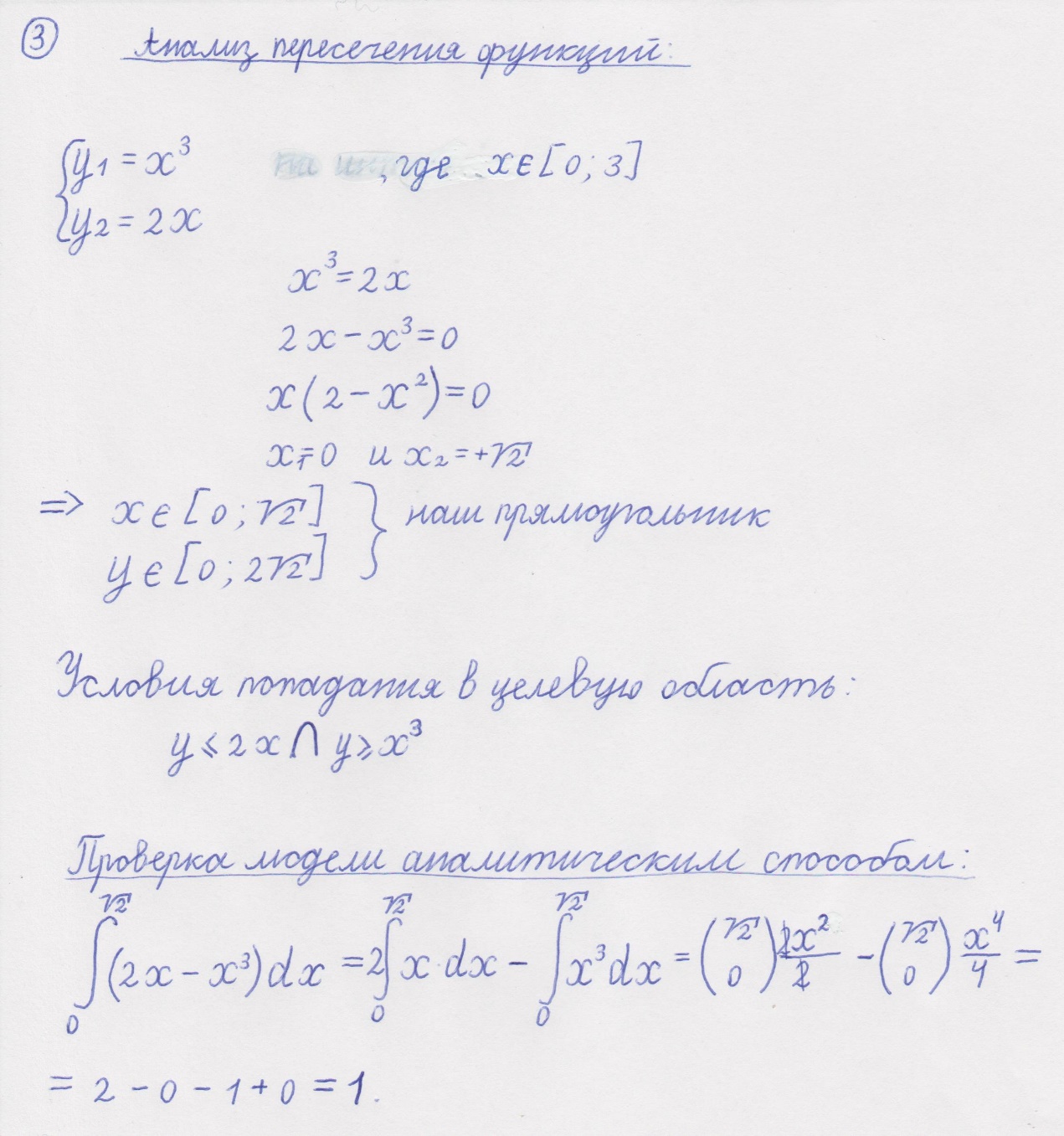


Рис. 4.4 – Вычисление площади аналитическим способом.

Как можно видеть, значение площади, полученное точным аналитическим способом равно единице. Среднее значение, полученное посредствам моделирования равно 0,94. Итого, мы на 94% успешно смоделировали площадь нашей фигуры.

# Источники

1. Теория вероятностей и математическая статистика: учеб. пособие / В. С. Мхитарян, Л. И. Трошин, Е. В. Астафьева, Ю. Н. Миронкина; под ред. В. С. Мхитаряна. – М. : Маркет ДС, 2010 (Университетская серия).
2. В. Е. Гмурман – Руководство к решению задач по теории вероятностей и математической статистике; Издание шестое, дополненное. – М. : «Высшая школа», 2002.
3. Письменный Д.Т. – Конспект лекций по высшей математике. 1 часть. – 4-е изд. – М.: Айрис-пресс, 2004.
4. Ивченко Г.И., Медведев Ю.И. – Введение в математическую статистику: Учебник. М.: Издательство ЛКИ, 2010.