

УДК 621.391

ОСОБЕННОСТИ ИСПОЛЬЗОВАНИЯ МЕТОДА НАИМЕНЬШИХ КВАДРАТОВ В СТАТИСТИЧЕСКИХ ЗАДАЧАХ

Легайда Е.П. *ар.263101*

Белорусский государственный университет информатики и радиоэлектроники

г. Минск, Республика Беларусь

Фильченкова Т.М. – ст. преподаватель каф. ИКТ

Аннотация. В данной статье рассмотрена тема метода наименьших квадратов в статистических задачах обработки результатов измерений. Представлена методика обработки результатов измерений с использованием метода наименьших квадратов. Приведены примеры применения метода наименьших квадратов.

Ключевые слова: метод наименьших квадратов, статистика, обработка результатов измерений, оценка погрешности измерений.

При проведении измерений в физике, инженерии, экономике и других науках важно не только получить числовые результаты, но и оценить их точность. Один из основных вопросов, которые возникают при обработке экспериментальных данных, – это определение наиболее вероятного значения измеряемой величины. В этом докладе рассматривается метод наименьших квадратов, который широко используется для решения этой проблемы.

Первый шаг в определении вероятнейшего значения измеренной величины – это оценка погрешности измерений. Эта погрешность может быть вызвана различными факторами, включая неточность измерительного прибора, нестабильность условий эксперимента, ошибки оператора и другие. Чтобы учесть эти погрешности, обычно проводят несколько измерений и вычисляют среднее значение.

Однако даже при использовании многократных измерений остается неопределенность относительно того, какое значение является наиболее вероятным. Для решения этой проблемы используется метод наименьших квадратов.

Метод наименьших квадратов (МНК) – это статистический метод, который используется для нахождения наилучшей линейной аппроксимации экспериментальных данных. Он основан на минимизации суммы квадратов отклонений между исходными данными и прямой линией, наиболее близкой к этим данным.

Предположим, что у нас есть набор данных, представляющий собой набор значений x и соответствующих им значений y . Мы хотим найти уравнение линейной функции, которая наилучшим образом описывает эти данные. Уравнение линейной функции имеет следующий вид:

$$y = ax + b$$

Чтобы найти наилучшую линейную аппроксимацию, мы ищем значения a и b , которые минимизируют сумму квадратов отклонений между исходными данными и линией. Эта сумма квадратов отклонений выглядит следующим образом:

$$\sum_{i=1}^n (y_i - ax_i - b)^2$$

где n – количество точек данных.

Для минимизации этой суммы квадратов можно применить метод дифференциального исчисления и найти значения a и b , приравняв производные по a и b к нулю. Это приводит к следующим уравнениям:

$$\sum_{i=1}^n y_i = a \sum_{i=1}^n x_i + nb$$

$$\sum_{i=1}^n x_i y_i = a \sum_{i=1}^n x_i^2 + \sum_{i=1}^n x_i$$

Решая эти уравнения относительно а и b, мы получаем следующие значения:

$$a = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2}$$

$$b = \frac{\sum_{i=1}^n y_i - b \sum_{i=1}^n x_i}{n}$$

Небольшой пример:

x	1	2	3	4	5
y	1.1	3.8	6.5	10.2	13.1

$$y = a + bx$$

$$\sum_{i=1}^5 x_i y_i = 134.5$$

$$\sum_{i=1}^5 x_i = 15$$

$$\sum_{i=1}^5 y_i = 34.7$$

$$\sum_{i=1}^5 x_i^2 = 55$$

$$a = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} = 3.04$$

$$b = \frac{\sum_{i=1}^n y_i - b \sum_{i=1}^n x_i}{n} = -2.18$$

$$y = 3.14x - 2.18$$

Таким образом, мы получаем уравнение линейной функции, которое наилучшим образом аппроксимирует исходные данные.

Одним из преимуществ метода наименьших квадратов является то, что он позволяет оценить точность вероятнейшего значения. Для этого используется среднеквадратическая ошибка (Mean Squared Error (MSE)), которая определяется как среднее значение квадратов отклонений между исходными данными и линейной аппроксимацией:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - ax_i - b)^2$$

Стандартная ошибка оценки (SE) может быть вычислена как квадратный корень из MSE:

$$SE = \sqrt{\frac{1}{n-2} \sum_{i=1}^n (y_i - ax_i - b)^2}$$

Эта ошибка оценки показывает, насколько точно линейная аппроксимация соответствует исходным данным. Чем меньше SE, тем более точна линейная аппроксимация.

Заключение. Метод наименьших квадратов является мощным инструментом для определения вероятнейшего значения измеренной величины и оценки его точности. Этот метод широко используется в физике, инженерии, экономике и других областях, где необходимо аппроксимировать данные линейной функцией. Кроме того, метод наименьших квадратов может быть обобщен на другие типы функций, такие как полиномы и экспоненциальные функции. В этих случаях процесс минимизации суммы квадратов отклонений аналогичен, но используются соответствующие функциональные формы.

Несмотря на все преимущества метода наименьших квадратов, следует помнить, что он предполагает линейную зависимость между переменными. Если зависимость не является линейной, результаты метода могут быть неточными или непригодными для использования.