Аппроксимация функций (метод наименьших квадратов)

Задача:

статистически обработать данные, и составить эмпирические формулы для нахождения зависимости одной величины от другой, когда известна таблица их значений, полученных в результате некоторой серии экспериментов.

Важнейшее **отличие** постановки данной задачи от задачи интерполирования состоит в том, что не требуется обязательное совпадение данных, полученных в результате измерений со значениями искомой функции в выделенных точках.

Анализ задачи:

- результаты измерений не могут быть точными,
- выделенные точки (узлы), как правило, ничем не отличаются от всех остальных и непонятно, почему именно в них мы должны требовать точного совпадения данных.

Меры приближения:

- Максимальное по модулю отклонение искомой функции в узлах от данных значений.
- Сумма модулей отклонений искомой функции в узлах от данных значений.
- Сумма квадратов отклонений искомой функции в узлах от данных значений.

ПОСТАНОВКА ЗАДАЧИ.

Дана таблица зависимости функции Y от аргумента X:

X	X1	X2	………	Xn
У	У1	У2	………	Уn

Надо среди функций основных видов определить такую (найти значения соответствующих параметров), чтобы сумма квадратов разностей значений этой функции в узлах и величин Yi была минимальна.

Обычно ограничиваются функциями одного из следующих видов:

- $Y=ax+b$
- $Y=ax^2+bx+c$
- $Y=cx^n$
- $Y=a\,e^x$
- $Y=1/(ax+b)$
- $Y=a\ln(x)+b$
- $Y=a/(x+b)$

Нахождение наилучшей линейной приближающей функции.

Разберем решение задачи, когда решение ищется в виде линейной функции: Y=ax+b.

Цель - определить коэффициенты a и b таким образом, чтобы величина

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

приняла наименьшее значение

Функция F(a,b) представляет из себя многочлен второй степени относительно величин a и b с неотрицательными значениями, поэтому решение всегда существует.

$$\begin{cases} F'_b(a,b) = \sum\limits_{i=1}^{n} 2(ax_i + b - y_i) = 0 \\ F'_a(a,b) = \sum\limits_{i=1}^{n} 2(ax_i + b - y_i) * x_i = 0 \end{cases}$$

$$\begin{cases} a\sum_{i=1}^{n} x_i^{\ 2} + b\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i x_i \\ a\sum_{i=1}^{n} x_i + bn = \sum_{i=1}^{n} y_i \end{cases}$$

Пусть зависимость задана таблицей

X	**-3**	**-1**	**1**	**3**	**5**
Y	**3**	**4**	**6**	**8**	**10**

Для вычисления искомых моментов построим таблицу:

	X	**Y**	**X^2**	**XY**
	-3	**3**	**9**	**-9**
	-1	**4**	**1**	**-4**
	1	**6**	**1**	**6**
	3	**8**	**9**	**24**
	5	**10**	**25**	**50**
Сумма	**5**	**31**	**45**	**67**
Среднее значение (М)	**1**	**6.2**	**9**	**13.4**

Отсюда получаем систему

$9a+b=13.4$ $\qquad\qquad$ $a=0.9$

$a+b=6.2$ $\qquad$ или $\qquad$ $b=5.3$

Проделайте аналогичные выкладки и получите систему уравнений для поиска коэффициентов a, b, c при подборе эмпирической квадратичной зависимости

X	**-3**	**-1**	**1**	**3**	**5**
Y	**3**	**4**	**6**	**8**	**10**

Сведение поиска функций другого вида к поиску линейной функции

При поиске функций другого вида задача сводится к рассмотренной задаче нахождения наилучшей линейной функции. Для этого производится некоторая замена переменных, которая подбирается таким образом, чтобы вновь полученная задача свелась к нахождению линейной зависимости, а после применения описанной конструкции происходит обратная замена.

Функция вида $y=1/(ax+b)$

При поиске такой функции, для сведения задачи к линейной мы произведем замену **$t=1/y$**, после которой задача сводится к нахождению наилучшей линейной функции **$t=ax+b$**. А коэффициенты, найденные при ее решении и будут искомыми в первоначальной задаче.

Алгоритм вычислений:

- заменяем в исходной таблице переменную **Y** на **t**, а все числа, записанные в нижней строке - на обратные
- для получившейся таблицы находим линейную зависимость
- получившиеся значения **a** и **b** берем без изменения.

Функция вида $Y=a \ln(x)+b$

Аналогичные действия производятся при поиске наилучшей приближающей функции вида $\mathbf{Y=a\ ln(x)+b}$. Но замена, которую необходимо произвести для сведения к линейной задаче, в этом случае имеет вид $\mathbf{u=ln(x)}$.

Алгоритм вычислений:

- заменяем в исходной таблице переменную X на u, а все числа, записанные в верхней строке - на их логарифмы
- для получившейся таблицы находим линейную зависимость
- получившиеся значения a и b берем без изменения.