**КУРСОВАЯ РАБОТА**

Аппроксимация функции методом наименьших квадратов

**Введение**

эмпирический mathcad аппроксимация

Целью курсовой работы является углубление знаний по информатике, развитие и закрепление навыков работы с табличным процессором Microsoft Excel и MathCAD. Применение их для решения задач с помощью ЭВМ из предметной области, связанной с исследованиями.

В каждом задании формулируются условия задачи, исходные данные, форма выдачи результатов, указываются основные математические зависимости для решения задачи Контрольный расчет позволяет убедиться в правильности работы программы.

Понятие аппроксимация представляет собой приближенное выражение каких-либо математических объектов (например, чисел или функций) через другие более простые, более удобные в использовании или просто более известные. В научных исследованиях аппроксимация применяется для описания, анализа, обобщения и дальнейшего использования эмпирических результатов.

Как известно, между величинами может существовать точная (функциональная) связь, когда одному значению аргумента соответствует одно определенное значение, и менее точная (корреляционная) связь, когда одному конкретному значению аргумента соответствует приближенное значение или некоторое множество значений функции, в той или иной степени близких друг к другу. При ведении научных исследований, обработке результатов наблюдения или эксперимента обычно приходиться сталкиваться со вторым вариантом. При изучении количественных зависимостей различных показателей, значения которых определяются эмпирически, как правило, имеется некоторая их вариабельность. Частично она задается неоднородностью самих изучаемых объектов неживой и, особенно, живой природы, частично обуславливается погрешностью наблюдения и количественной обработке материалов. Последнюю составляющую не всегда удается исключить полностью, можно лишь минимизировать ее тщательным выбором адекватного метода исследования и аккуратностью работы.

Специалисты в области автоматизации технологических процессов и производств имеют дело с большим объёмом экспериментальных данных, для обработки которых используется компьютер. Исходные данные и полученные результаты вычислений могут быть представлены в табличной форме, используя табличные процессоры (электронные таблицы) и, в частности, Excel. Курсовая работа по информатике позволяет студенту закрепить и развить навыки работы с помощью базовых компьютерных технологий при решении задач в сфере профессиональной деятельности.- система компьютерной алгебры из класса систем автоматизированного проектирования, ориентированная на подготовку интерактивных документов с вычислениями и визуальным сопровождением, отличается легкостью использования и применения для коллективной работы.

**1. Общие сведения**

Очень часто, особенно при анализе эмпирических данных возникает необходимость найти в явном виде функциональную зависимость между величинами *x* и *у*, которые получены в результате измерений.

При аналитическом исследовании взаимосвязи между двумя величинами x и y производят ряд наблюдений и в результате получается таблица значений:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| *x* | *x1* | *x1* |  | *xi* |  | *Xn* |
| *у* | *y1* | *y1* |  | *yi* |  | *Yn* |

Эта таблица обычно получается как итог каких-либо экспериментов, в которых *x,* (независимая величина) задается экспериментатором, а *у,* получается в результате опыта. Поэтому эти значения *у,* будем называть эмпирическими или опытными значениями.

Между величинами x и y существует функциональная зависимость, но ее аналитический вид обычно неизвестен, поэтому возникает практически важная задача - найти эмпирическую формулу

*y =* f(x; a1, a2,…, am), *(1)*

(где *a1, a2,…, am* - параметры), значения которой при *x = x,* возможно мало отличались бы от опытных значений *у, (i =* 1,2,…, *п)*.

Обычно указывают класс функций (например, множество линейных, степенных, показательных и т.п.) из которого выбирается функция *f (x)*, и далее определяются наилучшие значения параметров.

Если в эмпирическую формулу (1) подставить исходные *x,* то получим теоретические значения

*YTi = f* (*xi*; a1, a2…… *am)*, где *i =* 1,2,…, *n*.

Разности *yiT - уi,* называются отклонениями и представляют собой расстояния по вертикали от точек *Mi* до графика эмпирической функции.

Согласно методу наименьших квадратов наилучшими коэффициентами *a1, a2,…, am* считаются те, для которых сумма квадратов отклонений найденной эмпирической функции от заданных значений функции

 (2)

будет минимальной.

Поясним геометрический смысл метода наименьших квадратов.

Каждая пара чисел (*xi*, *yi*) из исходной таблицы определяет точку *Mi* на плоскости *XOY.* Используя формулу (1) при различных значениях коэффициентов *a1, a2,…, am* можно построить ряд кривых, которые являются графиками функции (1). Задача состоит в определении коэффициентов *a1, a2,…, am* таким образом, чтобы сумма квадратов расстояний по вертикали от точек *Mi* (*xi*, *yi*) до графика функции (1) была наименьшей (рис. 1).



Рис. 1

Построение эмпирической формулы состоит из двух этапов: выяснение общего вида этой формулы и определение ее наилучших параметров.

Если неизвестен характер зависимости между данными величинами x и *y*, то вид эмпирической зависимости является произвольным. Предпочтение отдается простым формулам, обладающим хорошей точностью. Удачный выбор эмпирической формулы в значительной мере зависит от знаний исследователя в предметной области, используя которые он может указать класс функций из теоретических соображений. Большое значение имеет изображение полученных данных в декартовых или в специальных системах координат (полулогарифмической, логарифмической и т.д.). По положению точек можно примерно угадать общий вид зависимости путем установления сходства между построенным графиком и образцами известных кривых.

Определение наилучших коэффициентов *a1, a*2,…, *am* входящих в эмпирическую формулу производят хорошо известным аналитическими методами.

Для того, чтобы найти набор коэффициентовa *a1, a2…..am,* которые доставляют минимум функции S, определяемой формулой (2), используем необходимое условие экстремума функции нескольких переменных - равенство нулю частных производных.

В результате получим нормальную систему для определения коэффициентов *ai (i =* 1,2,…, *m)*:

………… (3)

Таким образом, нахождение коэффициентов *ai* сводится к решению системы (3). Эта система упрощается, если эмпирическая формула (1) линейна относительно параметров *ai*, тогда система (3) - будет линейной.

**1.1 Линейная зависимость**

Конкретный вид системы (3) зависит от того, из какого класса эмпирических формул мы ищем зависимость (1). В случае линейной зависимости *y = a1 + a2x* система (3) примет вид:

  (4)

Эта линейная система может быть решена любым известным методом (методом Гаусса, простых итераций, формулами Крамера).

**1.2 Квадратичная зависимость**

В случае квадратичной зависимости *y = a1 + a2x + a*3x2 система (3) примет вид:

(5)

**1.3 Экспоненциальная зависимость**

В ряде случаев в качестве эмпирической формулы берут функцию в которую неопределенные коэффициенты входят нелинейно. При этом иногда задачу удается линеаризовать т.е. свести к линейной. К числу таких зависимостей относится экспоненциальная зависимость

*y = a1\* ea2x (6)*

где a1 иa2, неопределенные коффициенты.

Линеаризация достигается путем логарифмирования равенства (6), после чего получаем соотношение

ln y = ln a1 + a2x (7)

Обозначим ln *у* и ln *ax* соответственно через *t* и *c*, тогда зависимость (6) может быть записана в виде *t = a1 + a2х*, что позволяет применить формулы (4) с заменой *a1* на *c* и *уi* на *ti*

**1.4 Элементы теории корреляции**

График восстановленной функциональной зависимости *у(х)* по результатам измерений (х*i*, *уi*), *i = 1,2, K*, *n* называется кривой регрессии. Для проверки согласия построенной кривой регрессии с результатами эксперимента обычно вводят следующие числовые характеристики: коэффициент корреляции (линейная зависимость), корреляционное отношение и коэффициент детерминированности. При этом результаты обычно группируют и представляют в форме корреляционной таблицы. В каждой клетке этой таблицы приводятся численности *niJ -* тех пар (х, *у)*, компоненты которых попадают в соответствующие интервалы группировки по каждой переменной. Предполагая длины интервалов группировки (по каждой переменной) равными между собой, выбирают центры х*i* (соответственно *уi*) этих интервалов и числа *niJ-* в качестве основы для расчетов.

Коэффициент корреляции является мерой линейной связи между зависимыми случайными величинами: он показывает, насколько хорошо в среднем может быть представлена одна из величин в виде линейной функции от другой.

Коэффициент корреляции вычисляется по формуле:

 (8)

где , , и  - среднее арифметическое значение соответственно *х* и *у*.

Коэффициент корреляции между случайными величинами по абсолютной величине не превосходит 1. Чем ближе |р| к 1, тем теснее линейная связь между х и *у.*

В случае нелинейной корреляционной связи условные средние значения располагаются около кривой линии. В этом случае в качестве характеристики силы связи рекомендуется использовать корреляционное отношение, интерпретация которого не зависит от вида исследуемой зависимости.

Корреляционное отношение вычисляется по формуле:

(9)

где *ni* = , *nf* = , а числитель характеризует рассеяние условных средних *у,* около безусловного среднего *y*.

Всегда . Равенство*=* 0 соответствует некоррелированным случайным величинам;*= 1* тогда и только тогда, когда имеется точная функциональная связь между*y* и x. В случае линейной зависимости *y* от x корреляционное отношение совпадает с квадратом коэффициента корреляции. Величина*- ρ* 2 используется в качестве индикатора отклонения регрессии от линейной.

Корреляционное отношение является мерой корреляционной связи *y* с *x* в какой угодно форме, но не может дать представления о степени приближенности эмпирических данных к специальной форме. Чтобы выяснить насколько точно построенная кривая отражает эмпирические данные вводится еще одна характеристика - коэффициент детерминированности.

Для его описания рассмотрим следующие величины.- полная сумма квадратов, где  среднее значение .

Можно доказать следующее равенство



Первое слагаемое равно Sост =  и называется остаточной суммой квадратов. Оно характеризует отклонение экспериментальных от теоритических.

Второе слагаемое равно Sрегр =  2 и называется регрессионной суммой квадратов и оно характеризует разброс данных.

Очевидно, что справедливо следующее равенство Sполн = Sост + Sрегр.

Коэффициент детерминированности определяется по формуле:

 (10)

Чем меньше остаточная сумма квадратов по сравнению с общей суммой квадратов, тем больше значение коэффициента детерминированности *r2*, который показывает, насколько хорошо уравнение, полученное с помощью регрессионного анализа, объясняет взаимосвязи между переменными. Если он равен 1, то имеет место полная корреляция с моделью, т.е. нет различия между фактическим и оценочным значениями y. В противоположном случае, если коэффициент детерминированности равен 0, то уравнение регрессии неудачно для предсказания значений y

Коэффициент детерминированности всегда не превосходит корреляционное отношение. В случае когда выполняется равенство *r*2 =  то можно считать, что построенная эмпирическая формула наиболее точно отражает эмпирические данные.

##### **2. Постановка задачи**

1. Используя метод наименьших квадратов функцию , заданную таблично, аппроксимировать

а) многочленом первой степени ;

б) многочленом второй степени ;

в) экспоненциальной зависимостью .

. Для каждой зависимости вычислить коэффициент детерминированности.

. Вычислить коэффициент корреляции (только в случае а).

. Для каждой зависимости построить линию тренда.

. Используя функцию ЛИНЕЙН вычислить числовые характеристики зависимости  от .

. Сравнить свои вычисления с результатами, полученными при помощи функции ЛИНЕЙН.

. Сделать вывод, какая из полученных формул наилучшим образом аппроксимирует функцию .

. Написать программу на одном из языков программирования и сравнить результаты счета с полученными выше.

##### **3. Исходные данные**

Функция  задана рисунком 1.



Рис. 1.

##### **4. Расчет аппроксимаций в табличном процессоре Excel**

Для проведения расчетов целесообразно воспользоваться табличным процессором Microsoft Excel. И данные расположить как показано на рисунке 2.



Рис. 2

Для этого заносим:

· в ячейки A6:A30 заносим значения xi.

· в ячейки B6:B30 заносим значения уi.

· в ячейку C6 вводим формулу =А6^2.

· в ячейки C7:C30 эта формула копируется.

· в ячейку D6 вводим формулу =А6\*В6.

· в ячейки D7:D30 эта формула копируется.

· в ячейку F6 вводим формулу =А6^4.

· в ячейки F7:F30 эта формула копируется.

· в ячейку G6 вводим формулу =А6^2\*В6.

· в ячейки G7:G30 эта формула копируется.

· в ячейку H6 вводим формулу =LN(B6).

· в ячейки H7:H30 эта формула копируется.

· в ячейку I6 вводим формулу =A6\*LN(B6).

· в ячейки I7:I30 эта формула копируется. Последующие шаги делаем с помощью автосуммирования

· в ячейку А33 вводим формулу =СУММ (А6:А30).

· в ячейку B33 вводим формулу =СУММ (В6:В30).

· в ячейку C33 вводим формулу =СУММ (С6:С30).

· в ячейку D33 вводим формулу =СУММ (D6:D30).

· в ячейку E33 вводим формулу =СУММ (E6:E30).

· в ячейку F33 вводим формулу =СУММ (F6:F30).

· в ячейку G33 вводим формулу =СУММ (G6:G30).

· в ячейку H33 вводим формулу =СУММ (H6:H30).

· в ячейку I33 вводим формулу =СУММ (I6:I30).

Аппроксимируем функцию *y = f* (x) линейной функцией *y = a1 + a*2x. Для определения коэффициентов a1 и a2 воспользуемся системой (4). Используя итоговые суммы таблицы 2, расположенные в ячейках A33, B33, C33 и D33, запишем систему (4) в виде

(11)

решив которую, получим a1 = -24,7164 и a2 = 11,63183

Таким образом, линейная аппроксимация имеет вид *y= -24,7164 + 11,63183х* (12)

Решение системы (11) проводили, пользуясь средствами Microsoft Excel. Результаты представлены на рисунке 3:



Рис. 3

В таблице в ячейках A38:B39 записана формула {=МОБР (A35:B36)}. В ячейках E38:E39 записана формула {=МУМНОЖ (A38:B39, C35:C36)}.

Далее аппроксимируем функцию *y = f* (x) квадратичной функцией *y = a1 + a2x + a3x*2. Для определения коэффициентов a1, a2 и a3 воспользуемся системой (5). Используя итоговые суммы таблицы 2, расположенные в ячейках A33, B33, C33, D33, E33, F33 и G33 запишем систему (5) в виде:

(13)

Решив которую, получим a1 = 1,580946, a2 = -0,60819 и a3 = 0,954171 (14)

Таким образом, квадратичная аппроксимация имеет вид:

*у = 1,580946 -0,60819х +0,954171 х2*

Решение системы (13) проводили, пользуясь средствами Microsoft Excel. Результаты представлены на рисунке 4.



Рис. 4

В таблице в ячейках A46:C48 записана формула {=МОБР (A41:C43)}. В ячейках F46:F48 записана формула {=МУМНОЖ (A41:C43, D46:D48)}.

Теперь аппроксимируем функцию *y = f* (х) экспоненциальной функцией *y = a1ea2x.* Для определения коэффициентов *a1* и *a2* прологарифмируем значения *yi* и используя итоговые суммы таблицы 2, расположенные в ячейках A26, C26, H26 и I26 получим систему:

 (15)

где *с = ln(a1).*

Решив систему (10) найдем *с =* 0,506435, a2 = 0.409819.

После потенцирования получим a1 = 1,659365.

Таким образом, экспоненциальная аппроксимация имеет вид *y = 1,659365\*e0,4098194x*

Решение системы (15) проводили, пользуясь средствами Microsoft Excel. Результаты представлены на рисунке 5.



Рис. 5

В таблице в ячейках A55:B56 записана формула {=МОБР (A51:B52)}. В ячейках E54:E56 записана формула {=МУМНОЖ (A51:B52, С51:С52)}. В ячейке E56 записана формула =EXP(E54).

Вычислим среднее арифметическое x и у по формулам:

; 

Результаты расчета x и *y* средствами Microsoft Excel представлены на рисунке 6.



Рис. 6

В ячейке B58 записана формула =A33/25. В ячейке B59 записана формула =B33/25.

Для того чтобы рассчитать коэффициент корреляции и коэффициент детерминированности данные целесообразно расположить в виде таблице 2.

Таблица 2



Поясним как таблица на рисунке 7 составляется.

Ячейки A6:A33 и B6:B33 уже заполнены (см. рис. 2).

Далее делаем следующие шаги:

· в ячейку J6 вводим формулу =(A6-$B$58)\*(B6-$B$59).

· в ячейки J7:J30 эта формула копируется.

· в ячейку K6 вводим формулу =(А6-$В$58)^2.

· в ячейки K7:K30 эта формула копируется.

· в ячейку L6 вводим формулу =(В1-$В$59)^2.

· в ячейки L7:L30 эта формула копируется.

· в ячейку M6 вводим формулу =($Е$38+$Е$39\*А6-В6)^2.

· в ячейки M7:M30 эта формула копируется.

· в ячейку N6 вводим формулу =($F$46 +$F$47\*A6 +$F$48\*A6 Л6-В6)^2.

· в ячейки N7:N30 эта формула копируется.

· в ячейку O6 вводим формулу =($Е$56\*ЕХР ($Е$55\*А6) - В6)^2.

· в ячейки O7:O30 эта формула копируется.

Последующие шаги делаем с помощью автосуммирования.

· в ячейку J33 вводим формулу =CYMM (J6:J30).

· в ячейку K33 вводим формулу =СУММ (К6:К30).

· в ячейку L33 вводим формулу =CYMM (L6:L30).

· в ячейку M33 вводим формулу =СУММ (М6:М30).

· в ячейку N33 вводим формулу =СУММ (N6:N30).

· в ячейку O33 вводим формулу =СУММ (06:030).

Теперь проведем расчеты коэффициента корреляции по формуле (8) (только для линейной аппроксимации) и коэффициента детерминированности по формуле (10). Результаты расчетов средствами Microsoft Ехcеl представлены на рисунке 7.



Рис. 7

В таблице 8 в ячейке B61 записана формула =J33/(K33\*L33^(1/2). В ячейке B62 записана формула =1 - M33/L33. В ячейке B63 записана формула =1 - N33/L33. В ячейке B64 записана формула =1 - O33/L33.

Анализ результатов расчетов показывает, что квадратичная аппроксимация наилучшим образом описывает экспериментальные данные.

**4.1 Построение графиков в Excel**

Выделим ячейки A1:A25, после этого обратимся к мастеру диаграмм. Выберем точечный график. После того как диаграмма будет построена, щелкнем правой кнопкой мышки на линии графика и выберем добавить линию тренда (соответственно линейную, экспоненциальную, степенную и полиномиальную второй степени).



График линейной аппроксимации



График квадратичной аппроксимации



График экспоненциальной аппроксимации.

##### **5. Аппроксимация функции с помощью MathCAD**

Аппроксимация данных с учетом их статистических параметров относится к задачам регрессии. Они обычно возникают при обработке экспериментальных данных, полученных в результате измерений процессов или физических явлений, статистических по своей природе (как, например, измерения в радиометрии и ядерной геофизике), или на высоком уровне помех (шумов). Задачей регрессионного анализа является подбор математических формул, наилучшим образом описывающих экспериментальные данные.

**.1 Линейная регрессия**

Линейная регрессия в системе Mathcad выполняется по векторам аргумента **Х** и отсчетов **Y** функциями:

**intercept (x, y)** - вычисляет параметр **а1,** смещение линии регрессии по вертикали (см. рис.)



**slope (x, y)** - вычисляет параметр **a2,** угловой коэффициент линии регрессии (см. рис.)



Полученные значения коэффициентов используем в уравнении регрессии

**y(x) = a1+a2\*x**

Функция **corr (у, y(x))** вычисляет *коэффициент корреляции Пирсона.* Чем он ближе к **1,** тем точнее обрабатываемые данные соответствуют линейной зависимости (см. рис.)





**.2 Полиноминальная регрессия**

Одномерная полиномиальная регрессия с произвольной степенью n полинома и с произвольными координатами отсчетов в Mathcad выполняется функциями:

**regress (х, у, n)** - вычисляет вектор **S,** в составе которого находятся коэффициенты **ai** полинома **n**-й степени;

Значения коэффициентов **ai** могут быть извлечены из вектора **S** функцией **submatrix (S, 3, length(S) - 1, 0, 0).**

Полученные значения коэффициентов используем в уравнении регрессии

**y(x) = a1+a2\*x+a3\*x2** (см. рис.)



**.3 Нелинейная регрессия**

Для простых типовых формул аппроксимации предусмотрен ряд функций нелинейной регрессии, в которых параметры функций подбираются программой Mathcad.

К их числу относится функция **expfit (x, y, s),** которая возвращает вектор, содержащий коэффициенты **a1, a2** и **a3** экспоненциальной функции

**y(x) = a1 ^exp (a2 •x) + a3.** В вектор **S** вводятся начальные значения коэффициентов **a1, a2** и **a3** первого приближения.

##### **Заключение**

Анализ результатов расчетов показывает, что линейная аппроксимация наилучшим образом описывает экспериментальные данные.

Результаты полученные с помощью программы MathCAD полностью совпадают со значениями полученными с помощью Excel. Это говорит о верности вычислений.

##### **Список используемой литературы**

1. Информатика: Учебник / Под ред. проф. Н.В. Макаровой. М.: Финансы и статистика 2007
2. Информатика: Практикум по технологии работы на компьютере / Под. Ред. проф. Н.В. Макаровой. М Финансы и статистика, 2011.
3. Н.С. Пискунов. Дифференциальное и интегральное исчисление, 2010.
4. Информатика, Аппроксимация методом наименьших квадратов, методические указания, Санкт-Петербург, 2009.