Министерство образования и науки Российской Федерации
Волгоградский государственный архитектурно-строительный университет

Статистическая обработка данных в системе MathCAD

Методические указания к лабораторной работе

Составители Н. Н. Потапова, О. М. Забродина, О. А. Богомолова

Волгоград
ВолгГАСУ 2014

УДК 004.451(076.5)
ББК 32.973.26-018.2 я73
С781

С781 **Статистическая** обработка данных в MathCAD [Электронный ресурс] : методические указания к лабораторной работе / М-во образования и науки Рос. Федерации, Волгогр. гос. архит.-строит. ун-т ; сост. Н. Н. Потапова, О. М. Забродина, О. А. Богомолова. — Электронные текстовые и графические данные (1,2 Мбайт). — Волгоград : ВолгГАСУ, 2014. — Учебное электронное издание сетевого распространения — Систем. требования: PC 486 DX-33; Microsoft Windows XP; 2-скоростной дисковод CD-ROM; Adobe Reader 6.0; Internet Explorer 6.0. — Официальный сайт Волгоградского государственного архитектурно-строительного университета. Режим доступа: http://www.vgasu.ru/publishing/on-line/ — Загл. с титул. экрана.

В данных методических указаниях рассматриваются вопросы, связанные со статистической обработкой данных в системе Mathcad, которая включает в себя не только типовые статистические вычисления, но и проведение регрессий различного типа, сглаживание данных и предсказание.

В методических указаниях представлена теоретическая и практическая части по использованию статистических функций. Разработаны индивидуальные задания, направленные на приобретение навыков по обработке данных, полученных, например, из эксперимента. Сформулированы контрольные вопросы для проверки знаний. Приведены примеры использования методов статистической обработки данных средствами Mathcad.

Для студентов всех специальностей по дисциплинам «Использование пакетов прикладных программ в инженерных расчетах» и «Информатика».

УДК 004.451(076.5)
ББК 32.973.26-018.2 я73

Содержание

Введение

Обработка данных является важной сферой инженерной деятельности.

Выборка экспериментальных данных, чаще всего, представляется в виде массива, состоящего из пар чисел (x_i, y_i). Поэтому возникает задача аппроксимации дискретной зависимости y(x) непрерывной функцией f(x). Функция f(x), в зависимости от специфики задачи, может отвечать различным требованиям:

- f(x) должна проходить через точки (x_i, y_i), т.е . f (x_i)=y_i, i=1…n. В этом случае говорят об *интерполяции* данных функцией f (x) во внутренних точках между x_i, или экстраполяции за пределами интервала, содержащего все x_i;
- f(x) должна некоторым образом (например в виде определенной аналитической зависимости) приближать y(x_i), не обязательно проходя через точки (x_i, y_i). Такова постановка задачи *регрессии*, которую во многих случаях также можно назвать сглаживанием данных;
- f(x) должна приближать экспериментальную зависимость y(x_i), учитывая к тому же, что данные (x_i, y_i) получены с некоторой погрешностью, выражающей шумовую компоненту измерений. При этом функция f(x), с помощью того или иного алгоритма, уменьшает погрешность, присутствующую в данных (x_i, y_i). Такого типа задачи называют задачами *фильтрации*. Сглаживание – частный случай фильтрации.

Различные виды построения аппроксимирующей зависимости f(x) иллюстрирует рис.1. На нем исходные данные обозначены кружками, интерполяция отрезками прямых линий – пунктиром, линейная регрессия – наклонной прямой линией, а фильтрация – жирной гладкой кривой. Эти зависимости приведены в качестве примера и отражают лишь малую часть возможностей Mathcad по обработке данных. Вообще говоря, в Mathcad имеется целый арсенал встроенных функций, позволяющий осуществлять самую различную регрессию, интерполяцию-экстраполяцию и сглаживание данных.

В интегрированной системе Mathcad существуют огромное количество встроенных функций, позволяющих осуществлять регрессию, интерполяцию, экстраполяцию и сглаживание данных.

Обычные средства обработки данных включают в себя линейную и сплайновую аппроксимации. Статистическая обработка включает в себя не только статистические вычисления, но и проведение линейной и нелинейной регрессий различного типа сглаживание данных и предсказание.

С вопросами линейной интерполяции, сплайн - интерполяции и полиномиальной мы знакомились ранее в курсе изучения информатики. В рамках данных методических указаний рассмотрим возможности системы Mathcad, используя статистические методы.

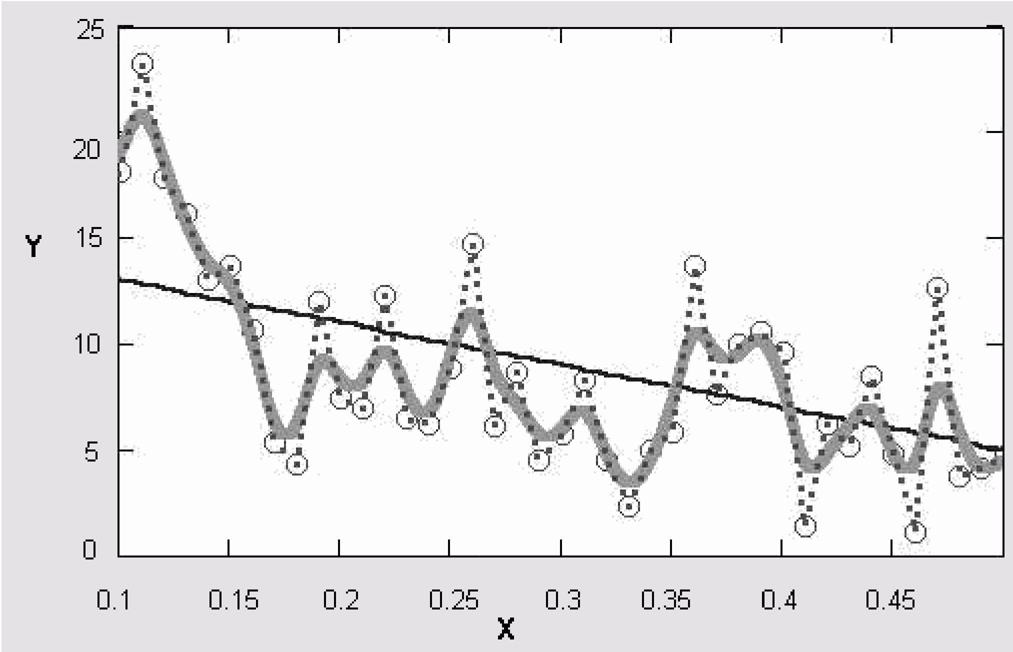


Рис. 1. Разные задачи аппроксимации данных

Статистическая обработка данных

При выполнении физических экспериментов их данные обычно представляются с той или иной случайной погрешностью. Поэтому их обработка нуждается в соответствующих статистических методах.

С помощью Mathcad можно проводить наиболее распространенные статистические расчеты с данными, представленными векторами их значений. Существует также ряд статистических функций для скалярного аргумента.

Типовые статистические функции в MathCAD

Существуют следующие встроенные статистические функции скалярного аргумента x:

erf(x) - функция ошибок;

rnd(x) - функция генерации случайных чисел;

corr(VX,VY) - коэффициент корреляции двух векторов - VX и VY, возвращает скаляр – коэффициент корреляции векторов VX и VY (Пирсона);

cvar(X,Y) - коэффициент ковариации X и Y;

cnorm(x) - функция кумулятивного стандартного нормального распределения.

Через функцию erf(x) легко вычисляется дополнительная функция ошибок:

erfc(x):= 1- erf(x)

Это одна из дополнительных и хорошо известных статистических функций, включенных в состав MathCAD.

Функция rnd(x) при каждом обращении к ней возвращает случайное число с равномерным распределением на отрезке [0, 1]. Эта функция широко применяется при статистическом моделировании различных физических процессов. Числа являются не строго случайными - в действительности это повторяющиеся последовательности из большого количества чисел, распределение которых близко к равномерному.

Функции, устанавливающие связь между парами двух случайных векторов, называются ковариацией и корреляцией (или, по-другому, коэффициентом корреляции). Они различаются нормировкой, как следует из их определения. Коэффициент корреляции – это числовая характеристика совместного распределения двух случайных величин, выражающая их взаимосвязь. Возможное значение от -1 до 1. Если значение 0, то нет зависимости одной величины от другой. Если значение -1, 1, то имеется линейная зависимость одной величины от другой.

Статистические функции для векторов и матриц

Следующая группа функций относится к вычислению основных статистических параметров одномерного массива данных - вектора:

mean(V) - возвращает среднее значение элементов вектора V;

gmean(V) — геометрическое среднее значение элементов вектора V;

median(V) - возвращает медиану элементов вектора V;

var(V) - возвращает дисперсию (вариацию) для элементов вектора V; дисперсия – это характеристика случайной величины, определяемая как математическое ожидание квадрата отклонения случайной величины от её математического ожидания;

stdev(V) - задает стандартное отклонение элементов вектора V; отклонение – это характеристика случайной величины, показывающая степень ее разброса, равная корню квадратному из дисперсии;

hist(int,V) - возвращает вектор частот попадания данных V в заданные интервалы int (служит для построения гистограмм).

В функции hist(int,V) вектор int должен содержать значения границ, в которых подсчитывается число попаданий данных из вектора V. Если строится гистограмма из N элементов, то вектор int должен содержать N + 1 элемент. Функция возвращает вектор из N элементов, числовые значения которых можно использовать для графического построения гистограмм;

max(x) , min(x) — максимальное и минимальное значения выборки;

mode(x) — наиболее часто встречающееся значение выборки;

median (x) — выборочная медиана (median) — значение аргумента, которое делит гистограмму плотности вероятностей на две равные части.

Пример расчета корреляции и ковариации

$N := 100 \qquad \sigma := 1$

$x1 := \mathrm{rnorm}(N, 0, \sigma) \qquad m2 := \mathrm{mean}(x2)$

$x2 := \mathrm{rnorm}(N, 0, \sigma) \qquad \sigma2 := \mathrm{stdev}(x2)$

$m1 := \mathrm{mean}(x1)$

$\sigma1 := \mathrm{stdev}(x1)$

$$\frac{1}{N} \cdot \sum_{i=0}^{N-1} \left[\left(x1_i - m1\right) \cdot \left(x2_i - m2\right)\right] = -0.197$$

$\mathrm{cvar}(x1, x2) = -0.197$

$$\frac{\mathrm{cvar}(x1, x2)}{\sigma1 \cdot \sigma2} = -0.203$$

$\mathrm{corr}(x1, x2) = -0.203$

Как видно, полученное значение корреляции близко к нулю, т. е. псевдослучайные числа генерируются практически независимо.

Здесь векторы x1 и x2, состоящие из 100 значений, формируются случайным образом по нормальному закону распределения rnorm(N, μ, σ) со средним значением μ=0 и стандартным отклонением σ=1.

На рис.2 представлен фрагмент документа MathCAD, в котором организована генерация вектора X из 1000 случайных чисел, дано их распределение и вычислены основные статистические параметры массива случайных чисел - вектора X. Этот фрагмент иллюстрирует также применение функции hist.

При достаточно большом числе случайных чисел вид гистограммы приближенно говорит о законе их распределения. Так, если высоты столбцов примерно равны, то распределение будет равномерным

Рис.2. Пример генерации случайных чисел и их проверка

.

Указанные функции могут использоваться и для обработки данных, представленных элементами (действительными и комплексными) матриц А размера m x n.

Все рассмотренные примеры работы статистических функций относились к векторам, элементы которых были случайными числами. Но точно так же все эти функции применяются и по отношению к выборкам случайных данных, сгруппированных в матрицы. При этом статистические характеристики рассчитываются для совокупности всех элементов матрицы, без разделения ее на строки и столбцы. Например, если матрица имеет размерность MxN, то и объем выборки будет равен MxN.

Соответствующий пример вычисления среднего значения приведен в листинге ниже. В его первой строке определяется матрица данных х размера 4х2. Действие встроенной функции mean матричного аргумента (последняя строка листинга) иллюстрируется явным суммированием элементов матрицы х (предпоследняя строка). Действие прочих встроенных функций на матрицы совершенно аналогично действию их на векторы.

Пример вычисления среднего значения элементов матрицы:

$$x := \begin{pmatrix} 1.0 & 4 \\ 1.5 & 7 \\ 0.9 & 1.2 \\ 1.2 & 12 \end{pmatrix}$$

$$M := rows(x) \qquad M = 4$$

$$N := cols(x) \qquad N = 2$$

$$\frac{1}{M \cdot N} \cdot \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} x_{i,j} = 3.6$$

$$mean(x) = 3.6$$

Пример листинга в системе MathCAD к определению статических характеристик для вектора x, заданного значениями 5 2 14 3 2:

$$x := (5 \quad 2 \quad 14 \quad 3 \quad 2)^T$$

$$N := length(x) \qquad N = 5$$

$$\frac{1}{N} \cdot \sum_{i=0}^{N-1} x_i = 5.2$$

$$m := mean(x) \qquad m = 5.2$$

$$median(x) = 3 \qquad max(x) = 14$$

$$mode(x) = 2 \qquad min(x) = 2$$

$$\frac{1}{N} \cdot \sum_{i=0}^{N-1} \left(x_i - m\right)^2 = 20.56$$

$$var(x) = 20.56 \qquad Var(x) \cdot \frac{N-1}{N} = 20.56$$

$$\sqrt{var(x)} = 4.534 \qquad stdev(x) = 4.534$$

$$Stdev(x) \cdot \sqrt{\frac{N-1}{N}} = 4.534$$

$$\frac{1}{N-1} \cdot \sum_{i=0}^{N-1} \left(x_i - m\right)^2 = 25.7$$

$$Var(x) = 25.7 \qquad var(x) \cdot \frac{N}{N-1} = 25.7$$

$$\sqrt{Var(x)} = 5.07 \qquad Stdev(x) = 5.07$$

$$stdev(x) \cdot \sqrt{\frac{N}{N-1}} = 5.07$$

Функции создания векторов с различными законами распределения

Последняя группа статистических функций служит для создания векторов с определенными законами распределения значений их элементов:

·rbeta(m,s1,s2) - бета-распределение;
·rbinom(m,n,p) - биномиальное распределение;
·rcauchy(m,l,s) - распределение Коши;
·rchisq(m,d) - хи-квадрат-распределение;
·rexp(m,r) - экспоненциальное распределение,
·rF(m,d1,d2) - распределение Фишера;
·rgamma(m,s) - гамма-распределение;
·rgeom(m,p) - геометрическое распределение;
·rlnorm(m,m,s) - логарифмическое нормальное распределение;
·rlogis(m,l,s) - логистическое распределение;
·rnbinom(m,n,p) - отрицательное биномиальное распределение;
·rnorm(m,m,s) - нормальное распределение;
·rpois(m,l) - распределение Пуассона;
·rt(m,d) - распределение Стьюдента;
·runif(m,a,b) - равномерное распределение;
·rweibull(m,s) - распределение Вейбулла.

На рис. 3. показан фрагмент документа MathCAD с примерами построения графиков различных статистических функций и задания наборов чисел с различным распределением.

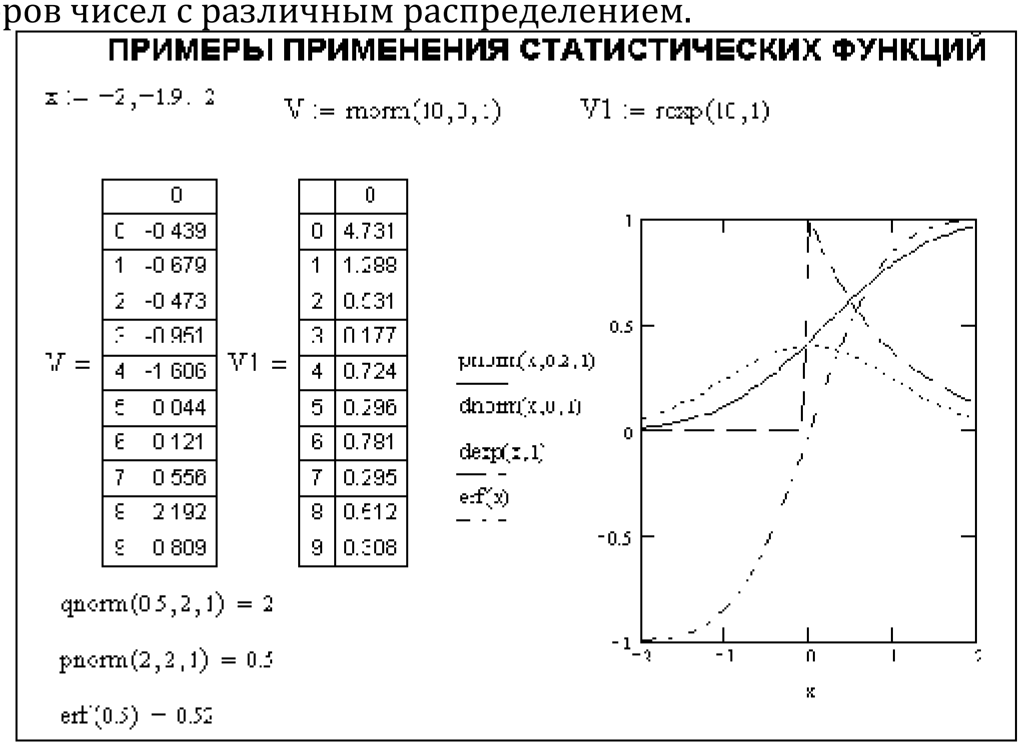


Рис. 3. Примеры применения статистических функций

В Приложении 1 дается перечень других статистических функций: функции вычисления плотности распределения вероятности, функции распределения, квантили распределения.

Обилие статистических функций, включенных в систему MathCAD, позволяет с ее помощью выполнять достаточно сложные статистические расчеты.

Регрессия

Другой широко распространенной задачей обработки данных является представление их совокупности некоторой функцией f(x). Задача регрессии сводится к подбору неизвестных коэффициентов этой функции такими, чтобы функция приближала «облако» исходных точек, заданных векторами x и y, с наименьшей среднеквадратической погрешностью.

Другими словами, задачи математической регрессии имеют смысл приближения выборки данных (x_i,y_i) некоторой функцией f(x), определенным образом минимизирующей совокупность ошибок $|f(x_i)-y_i|$.

Геометрически задача построения эмпирической формулы состоит в проведении кривой возможно ближе примыкающей к системе точек

$$M_i(x_i,y_i) \qquad (i = 0,1, \dots , n).$$

В силу производимого действия большинство задач регрессии являются частным случаем более общей проблемы сглаживания данных.

Линейная регрессия

Самый простой и наиболее часто используемый вид регрессии – линейная регрессия. Приближение данных (x_i,y_i) осуществляется линейной функцией y(x) = b+ax. На координатной плоскости (x,y) линейная функция, как известно, представляется прямой линией. Еще линейную регрессию часто называют методом наименьших квадратов, поскольку коэффициенты a и b вычисляются из условия минимизации суммы квадратов ошибок $|b+ax_i-y_i|$.

Для расчета линейной регрессии в Mathcad имеются два дублирующих друг друга способа. Правила их применения представлены в листингах 1 и 2.

Результат обоих листингов получается одинаковым:

- line (x, y) – вектор из двух элементов (b,a) коэффициентов линейной регрессии b+ax;
- intercept (x, y) – коэффициент b линейной регрессии;
- slope (x,y) – коэффициент a линейной регрессии:

  x – вектор действительных данных аргумента;

  y – вектор действительных данных значений того же размера.

Листинг1. Линейная регрессия:

$$x := (0\ \ 1\ \ 2\ \ 3\ \ 4\ \ 5\ \ 6)^T$$

$$y := (4.1\ \ 2.4\ \ 3\ \ 4.3\ \ 3.6\ \ 5.2\ \ 5.9)^T$$

$$\mathrm{line}(x,y) = \begin{pmatrix} 2.829 \\ 0.414 \end{pmatrix}$$

$$f(t) := \mathrm{line}(x,y)_0 + \mathrm{line}(x,y)_1 \cdot t$$

Листинг 2. Другая форма записи линейной регрессии:

$$x := (0\ \ 1\ \ 2\ \ 3\ \ 4\ \ 5\ \ 6)^T$$

$$y := (4.1\ \ 2.4\ \ 3\ \ 4.3\ \ 3.6\ \ 5.2\ \ 5.9)^T$$

$$\mathrm{intercept}(x,y) = 2.829$$

$$\mathrm{slope}(x,y) = 0.414$$

$$f(t) := \mathrm{intercept}(x,y) + \mathrm{slope}(x,y) \cdot t$$

В Mathcad имеется альтернативный алгоритм, реализующий не минимизацию суммы квадратов ошибок, а медиан-медианную линейную регрессию для расчета коэффициентов a и b (Листинг 3) :

- medfit**(x,y)** – вектор из двух элементов **(**b,a**)** коэффициентов линейной медиан-медианной регрессии b+ax:

  x, y – векторы действительных данных одинакового размера.

Листинг 3. Построение медиан-медианной линейной регрессии:

$$\mathrm{medfit}(x,y) = \begin{pmatrix} 2.517 \\ 0.55 \end{pmatrix}$$

$$g(t) := \mathrm{medfit}(x,y)_0 + \mathrm{medfit}(x,y)_1 \cdot t$$

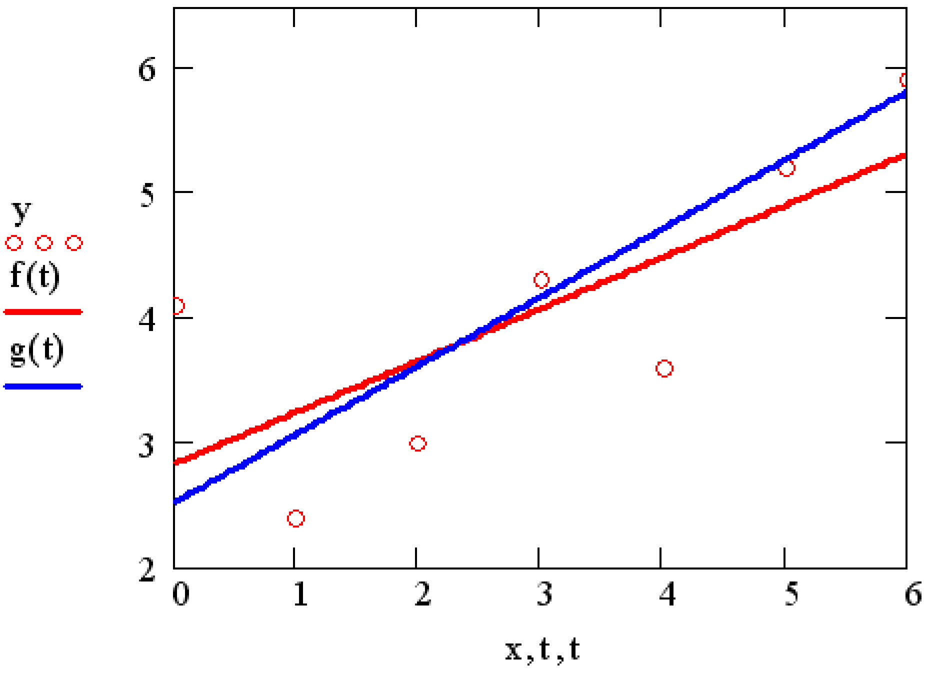


Рис.4. Линейная регрессия по методу наименьших квадратов и методу медиан

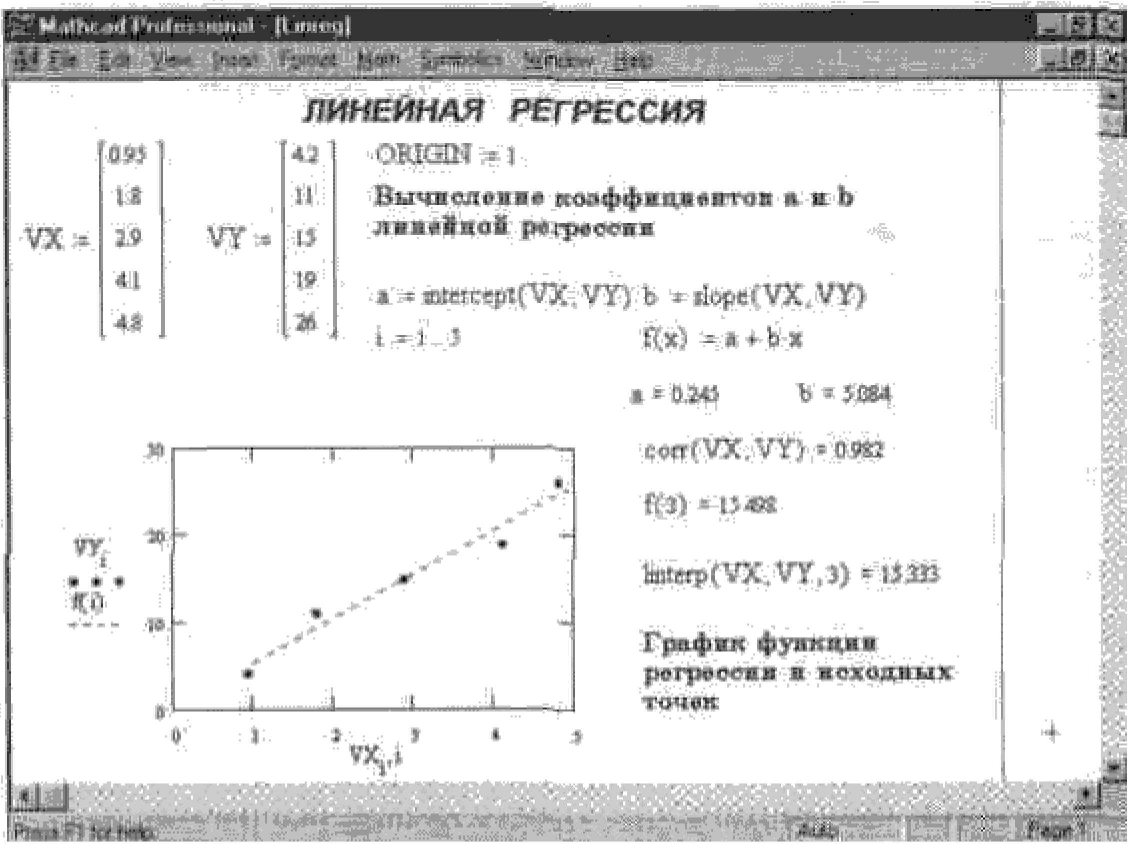


Рис.5. Пример линейной регрессии

Как видно на рис.5 прямая регрессии проходит в «облаке» исходных точек с максимальным среднеквадратичным приближением к ним. Чем ближе коэффициент корреляции к 1, тем точнее представленная исходными точками зависимость приближается к линейной.

Линеаризация некоторых зависимостей

К линейной регрессии можно свести многие виды нелинейной регрессии при двухпараметрических зависимостях y(x).

Наиболее часто берется класс, состоящий из функций следующего вида

1) $y = a_1+b_1x$	5) $y = e^{a_5+b_5x}$
2) $y = a_2 \cdot e^{b_2 x}$	6) $y = a_6 + b_6/x$
3) $y = a_3 \cdot b_3^{x}$	7) $y = a_7 + b_7 \cdot e^{x}$
4) $y = a_4 \cdot x^{b_4}$	8) $y = a_8 + b_8 \cdot \ln x$

Для определения параметров зависимостей используется метод наименьших квадратов. Если рассматривается зависимость первого типа, то составляется сумма

$$S = \sum_{i=1}^{n} (y_i - a_1 - b_1 \cdot x_i)^2$$

и наилучшими считаются те значения параметров a_1 и b_1, для которых S принимает наименьшие значения.

Случаи 2) – 8) сводятся к случаю 1).

Для случая 2) логарифмируя получим

$$\ln y = \ln a_2 + b_2 x$$

Если обозначить ln y через z и $\ln a_2$ через a, то получим зависимость

$$z=a + b_2x$$

Найдя отсюда a и b_2, находим затем a_2 по формуле $a_2=\exp(a)$. $y=\exp(z)$.

Для случая 3) логарифмируя получим

$$\ln y = \ln a_3 + x \ln b_3$$

Вводя обозначения

$$z = \ln y, \quad a = \ln a_3, \quad b = \ln b_3,$$

получим зависимость

$$a_3 = \exp(a), \; b_3 = \exp(b).$$

Для случая 4) логарифмируя получим

$$\ln y = \ln a_4 + b_4 \ln x$$

Вводя обозначения

$$z = \ln y, \; x1=\ln x, \; a=\ln a_4,$$

получим зависимость

$$z = a + b_4\, x1.$$

Найдя отсюда a и b_4, полагаем $a_4=\exp(a)$.

Для случая 5) логарифмируя получим

$$\ln y = a_5 + b_5 x$$

Вводим обозначение $z = \ln y$. Получим зависимость

$$z = a_5 + b_5 x$$

Находим значения параметров a_5 и b_5.

Для случая 6) вводим обозначение $x1 = 1/x$, получаем зависимость

$$y = a_6 + b_6 x1,$$

затем находим a_6 и b_6.

Для случая 7) вводим обозначение $x1 = \exp(x)$, получаем зависимость

$$y = a_7 + b_7 x1,$$

затем находим a_7 и b_7.

Для случая 8) вводим обозначение $x1 = \ln(x)$, получаем зависимость

$$y = a_8 + b_8 x1,$$

находим значения a_8 и b_8.

Выбор наилучшей функции производится следующим образом: для каждой функции вычисляется средняя квадратичная ошибка аппроксимации

$$S_1 = \frac{1}{n-2}\sum_{i=0}^{n}\left(y_i - y1_i\right)^2 \qquad (1)$$

и относительная ошибка аппроксимации

$$S_2 = \frac{1}{n+1}\sum_{i=0}^{n}\left|\frac{y_i - y1_i}{y_i}\right|, \qquad (2)$$

где y_i – эмпирические значения, $y1_i$ – расчетные значения.

Затем функции 1)–8) ранжируем в порядке возрастания значений S1 и S2. Функция, для которой сумма рангов S_1 и S_2 наименьшая, считается лучшей.

Реализация линейной регрессии общего вида

В Mathcad можно осуществить регрессию в виде линейной комбинации $C_1 * f_1(x) + C2 * f_2(x) + \ldots$, где $f_i(x)$ - любые функции пользователя, а C_i - подлежащие определению коэффициенты. Кроме того, имеется путь проведения регрессии более общего вида, когда комбинацию функций и искомых коэффициентов задает сам пользователь.

Приведем примеры встроенных функций для регрессии общего вида и примеры их использования (листинги 1 и 2).

linfit (x, y, F) - вектор параметров линейной комбинации функций пользователя, осуществляющей регрессию данных.

genfit (x, y, g, G) - вектор параметров, реализующих регрессию данных с помощью функций пользователя общего вида.

x - вектор действительных данных аргумента, элементы которого расположены в порядке возрастания.

y - вектор действительных значений того же размера.

F(x) - пользовательская векторная функция скалярного аргумента.

g - вектор начальных значений параметров регрессии размерности N.

G(x,C) - векторная функция размерности N+1, составленная из функции пользователя и ее N частных производных по каждому из параметров C.

Примечание. Число данных (количество элементов в векторах x и y) должно быть не меньше, чем N.

Листинг 1. Регрессия линейной комбинацией функций пользователя

$$x := (0\ \ 1\ \ 2\ \ 3\ \ 4\ \ 5\ \ 6)^T$$

$$y := (4.1\ \ 2.4\ \ 3\ \ 4.3\ \ 3.6\ \ 5.2\ \ 5.9)^T$$

$$F(x) := \begin{pmatrix} \dfrac{1}{x+1} \\ x \\ e^x \end{pmatrix} \qquad C := \mathrm{linfit}(x, y, F)$$

$$C = \begin{pmatrix} 3.957 \\ 0.854 \\ 5.605 \times 10^{-4} \end{pmatrix}$$

$$f(x) := \frac{C_0}{x+1} + C_1 \cdot x + C_2 \cdot e^x$$

Листинг 2. Регрессия общего вида

$$x := (0\ \ 1\ \ 2\ \ 3\ \ 4\ \ 5\ \ 6)^T$$

$$y := (4.1\ \ 2.4\ \ 3\ \ 4.3\ \ 3.6\ \ 5.2\ \ 5.9)^T$$

$$G(x,C) := \begin{pmatrix} C_0 \cdot x^{C_1 x} \\ x^{C_1 x} \\ C_0 \cdot x^{C_1 x + 1} \end{pmatrix}$$

$$g := \begin{pmatrix} 0.1 \\ 1 \end{pmatrix}$$

$$C := \mathrm{genfit}(x, y, g, G)$$

$$C = \begin{pmatrix} 3.151 \\ 0.056 \end{pmatrix}$$

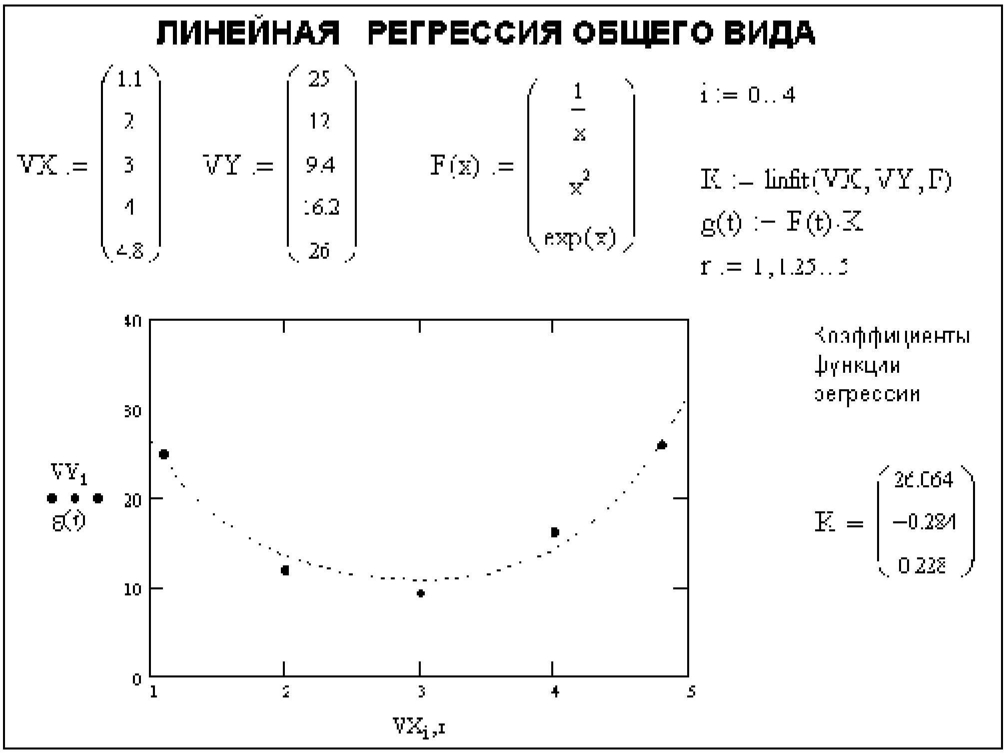


Рис.6. Пример построения линейной регрессии общего вида

Полиномиальная регрессия

Полиномиальная регрессия означает приближение данных (x_i,y_i) полиномом k-й степени $A(x)=a+bx+cx^2+dx^3+..\ .+hx^k$ (рис. 6). При k=1 полином является прямой линией, при k=2 - параболой, при k=3 - кубической параболой и т. д. Как правило, на практике применяются k<5.

Примечание. Для построения регрессии полиномом k-й степени необходимо наличие, по крайней мере, (k+1) точек данных.

В Mathcad полиномиальная регрессия осуществляется комбинацией встроенной функции regress и полиномиальной интерполяции:

regress(x,y,k) - вектор коэффициентов для построения полиномиальной регрессии данных;

interp(s,x,y, t) - результат полиномиальной регрессии:

s=regress(x,y,k);

x - вектор действительных данных аргумента, элементы которого расположены в порядке возрастания;

y - вектор действительных данных значений того же размера;

k - степень полинома регрессии (целое положительное число);

t - значение аргумента полинома регрессии.

Для построения полиномиальной регрессии после функции regress вы обязаны использовать функцию interp

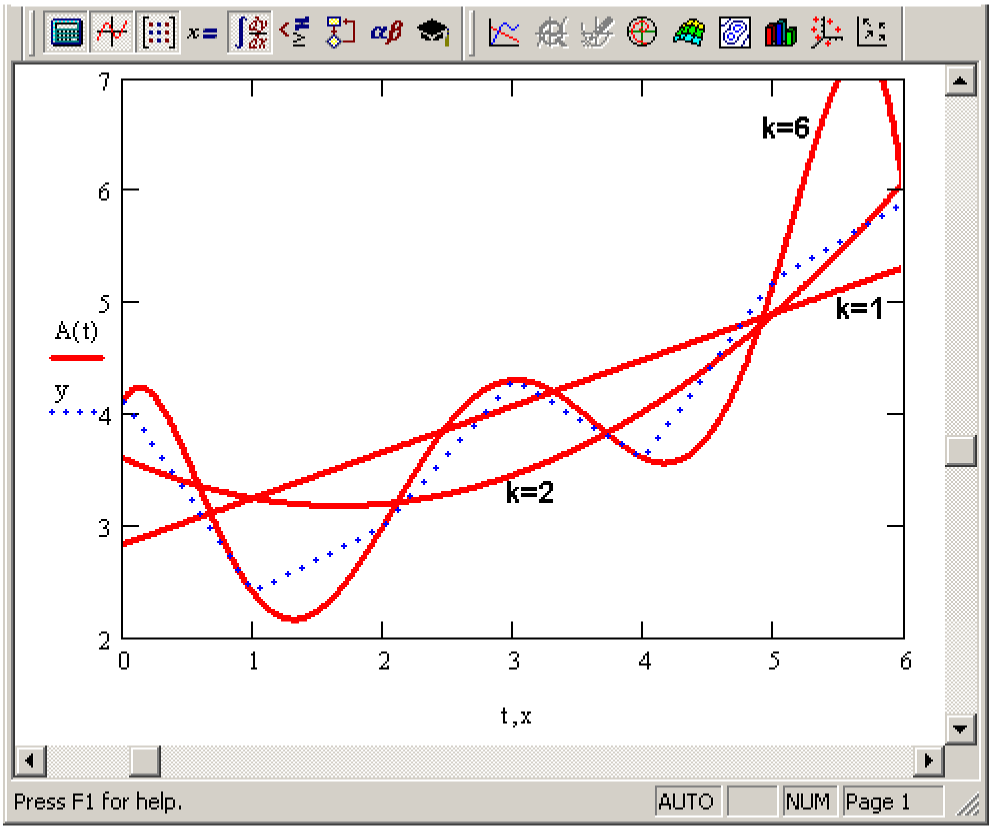


Рис. 7. Регрессия полиномами разной степени (коллаж результатов листинга 1 для разных k)

Листинг 1. Полиномиальная регрессия квадратичной параболой:

```
x := (0  1  2  3  4  5  6)^T

y := (4.1  2.4  3  4.3  3.6  5.2  5.9)^T

  k := 2

 s := regress(x,y,k)

A(t) := interp(s,x,y,t)
```

Регрессия отрезками полиномов

Помимо приближения массива данных одним полиномом имеется возможность осуществить регрессию сшивкой отрезков (точнее говоря, участков, т. к. они имеют криволинейную форму) нескольких полиномов.

Для этого имеется встроенная функция loess, применение которой аналогично функции regress (листинг 1 и рис. 8):

- loess (x, y, span) – вектор коэффициентов для построения регрессии данных отрезками полиномов;
- interp(s,x,y,t) – результат полиномиальной регрессии:

s=loess(x,y,span);

x – вектор действительных данных аргумента, элементы которого расположены в порядке возрастания;

y – вектор действительных данных значений того же размера;

span – параметр, определяющий размер отрезков полиномов (положительное число, хорошие результаты дает значение порядка span=0.75).

Параметр span задает степень сглаженности данных. При больших значениях span регрессия практически не отличается от регрессии одним полиномом (например, span=2 дает почти тот же результат, что и приближение точек параболой).

Листинг 1. Регрессия отрезками полиномов:

$$x := (0\ \ 1\ \ 2\ \ 3\ \ 4\ \ 5\ \ 6)^{T}$$

$$y := (4.1\ \ 2.4\ \ 3\ \ 4.3\ \ 3.6\ \ 5.2\ \ 5.9)^{T}$$

$$s := \mathrm{loess}(x, y, 0.9)$$

$$A(t) := \mathrm{interp}(s, x, y, t)$$

Примечание. Регрессия одним полиномом эффективна, когда множество точек выглядит как полином, а регрессия отрезками полиномов оказывается полезной в противоположном случае.

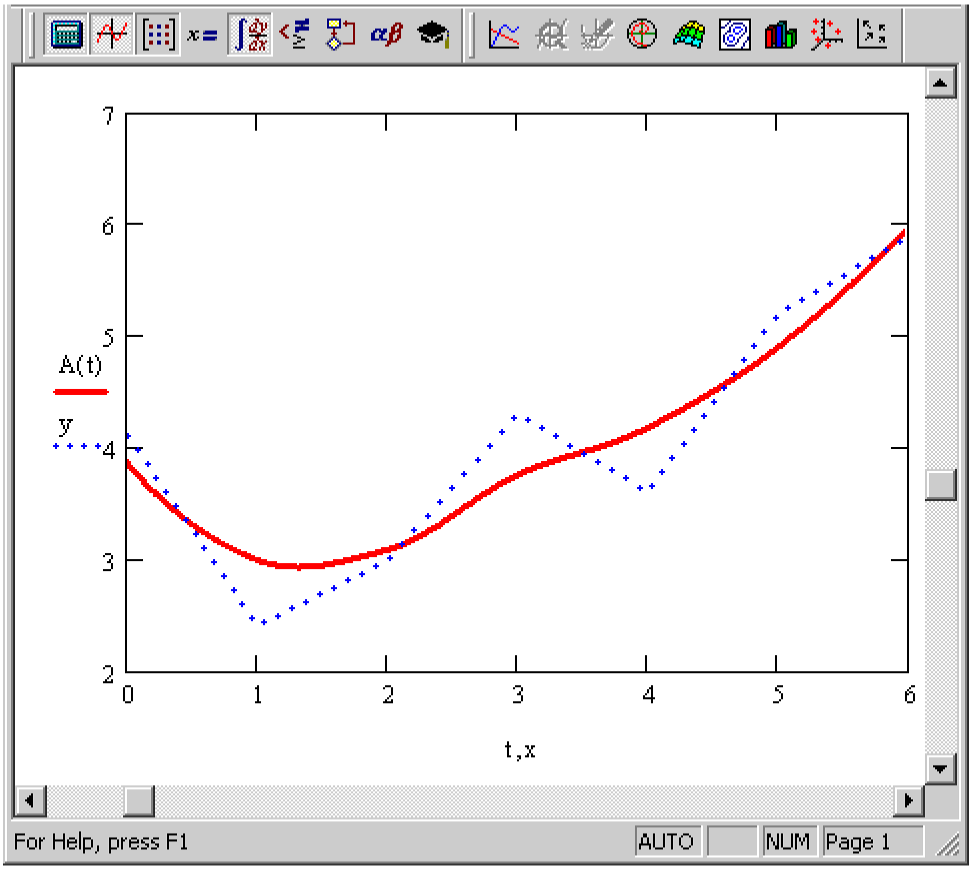


Рис. 8. Регрессия отрезками полиномов

Регрессия специального вида

Кроме рассмотренных, в MathCAD встроено еще несколько видов трехпараметрической регрессии. Их реализация несколько отличается от приведенных выше вариантов регрессии тем, что для них, помимо массива данных, требуется задать некоторые начальные значения коэффициентов a, b, c. Используйте соответствующий вид регрессии, если хорошо представляете себе, какой зависимостью описывается ваш массив данных. Когда тип регрессии плохо отражает последовательность данных, то ее результат часто бывает неудовлетворительным и даже сильно различающимся в зависимости от выбора начальных значений. Каждая из функций выдает вектор уточненных параметров a, b, c.

- expfit (x, y, g) – регрессия экспонентой f (x) $=ae^{bx}+c$.
- lgsfit (x,y,g) – регрессия логистической функцией f (x)$=a/(1+be^{-cx})$.
- sinfit (x,y,g) – регрессия синусоидой f (x) =a-sin (x+b) +c.
- pwfit(x,y,g) – регрессия степенной функцией f (x)$=ax^{b}+c$.
- logfit (x, y, g**)** – регрессия логарифмической функцией f (x)=aln(x+b) +c.
- infit (x,y**)** – регрессия двухпараметрической логарифмической функцией f (x)=aln(x)+b:

x – вектор действительных данных аргумента.

y – вектор действительных значений того же размера.

g – вектор из трех элементов, задающий начальные значения a, b, c.

Примечание. Правильность выбора начальных значений можно оценить по результату регрессии – если функция, выданная Mathcad, хорошо приближает зависимость y (x), значит, они были подобраны удачно.

Пример расчета одного из видов трехпараметрической регрессии (экспоненциальной) приведен в листинге 1 и на рис.9. В предпоследней строке листинга выведены в виде вектора вычисленные коэффициенты a, b, c, а в последней строке через эти коэффициенты определена искомая функция f(x).

Листинг 1. Экспоненциальная регрессия:

$$x := (0\ \ 1\ \ 2\ \ 3\ \ 4\ \ 5\ \ 6)^T$$

$$y := (4.1\ \ 2.4\ \ 3\ \ 4.3\ \ 3.6\ \ 5.2\ \ 5.9)^T$$

$$g := \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

$$C := \mathrm{expfit}(x, y, g)$$

$$C = \begin{pmatrix} 0.109 \\ 0.548 \\ 3.104 \end{pmatrix}$$

$$f(t) := C_0 \cdot \exp(C_1 \cdot t) + C_2$$

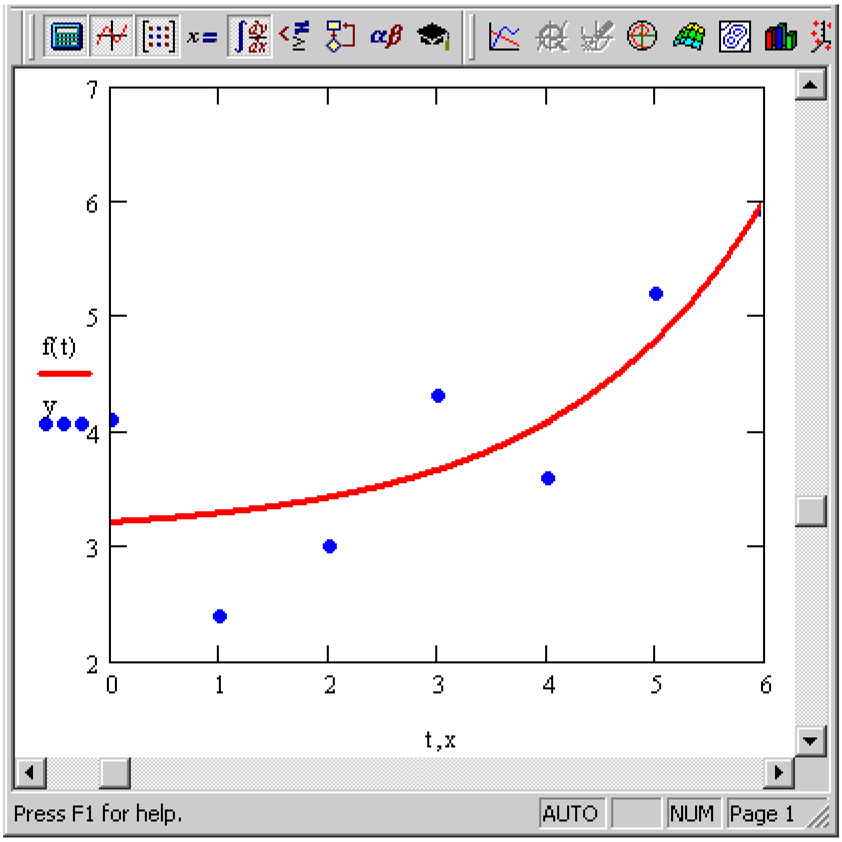


Рис. 9. Экспоненциальная регрессия

Примечание. Многие задачи регрессии данных различными двухпараметрическими зависимостями **y(x)** можно свести к более надежной, с вычислительной точки зрения, линейной регрессии. Делается это с помощью соответствующей замены переменных.

Экстраполяция функцией предсказания

Стандартные функции интерполяции-экстраполяции стоит применять только в непосредственной близости границ интервала данных. В MathCAD имеется более развитый инструмент экстраполяции, который учитывает распределение данных вдоль всего интервала. В функцию predict встроен линейный алгоритм предсказания поведения функции, основанный на анализе, в том числе осцилляции:

predict (y, m, n) - функция предсказания вектора, экстраполирующего выборку данных:

y - вектор действительных значений, взятых через равные промежутки значений аргумента;

m - количество последовательных элементов вектора y, согласно которым строится экстраполяция;

n - количество элементов вектора предсказаний.

Пример использования функции предсказания на примере экстраполяции осциллирующих данных y_j с меняющейся амплитудой приведен в листинге 1. Полученный график экстраполяции, наряду с самой функцией, показан на рис. 13.10. Аргументы и принцип действия функции predict отличаются от рассмотренных выше встроенных функций интерполяции-экстраполяции. Значений аргумента для данных не требуется, поскольку по определению функция действует на данные, идущие друг за другом с равномерным шагом. Обратите внимание, что результат функции predict вставляется "в хвост" исходных данных.

Листинг 1. Экстраполяция при помощи функции предсказания

$$k := 100$$

$$j := 0..k$$

$$y_j := e^{\frac{-j}{100}} \cdot \sin\left(\frac{j}{10}\right)$$

$$m := 50 \qquad n := 150$$

$$A := \mathrm{predict}(y, m, n)$$

Как видно из рис. 10, функция предсказания может быть полезна при экстраполяции данных на небольшие расстояния. Вдали от исходных данных результат часто бывает неудовлетворительным. Кроме того, функция predict хорошо работает в задачах анализа подробных данных с четко прослеживающейся закономерностью (типа рис.10), в основном осциллирующего характера.

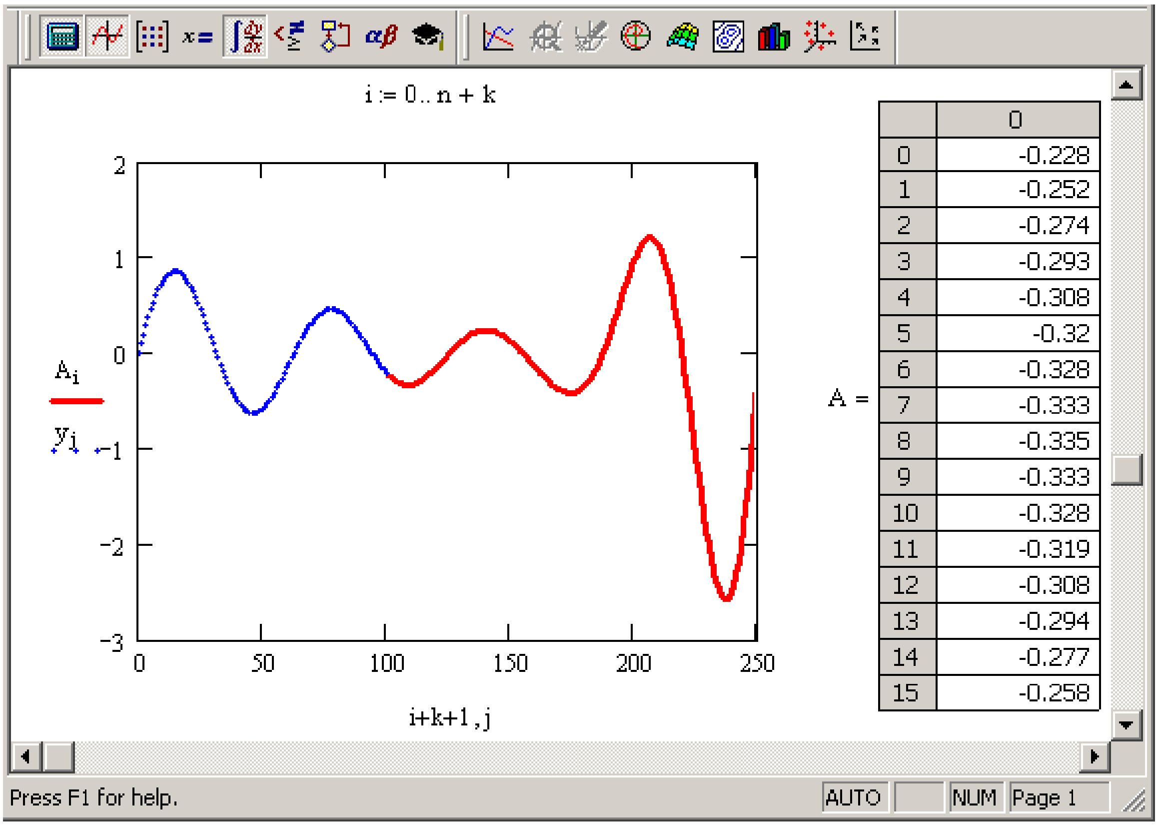


Рис.10. Экстраполяция при помощи функции предсказания

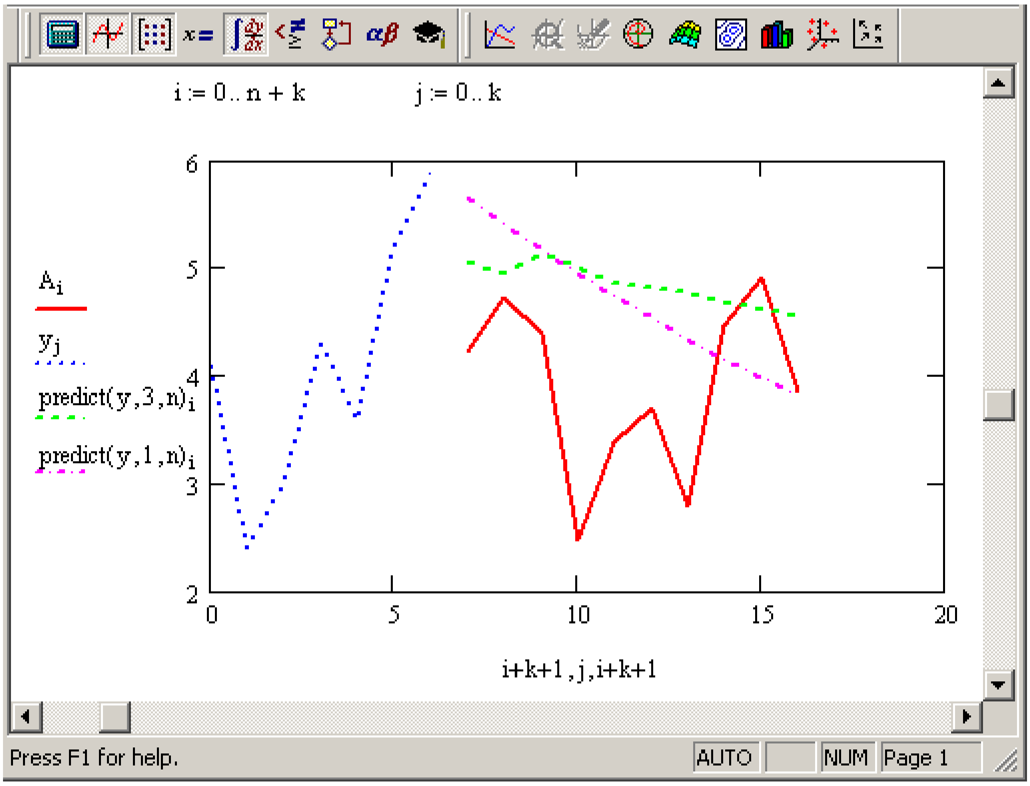


Рис.11. Работа функции предсказания в случае малого количества данных

Если данных мало, то предсказание может оказаться бесполезным. В листинге 2 приведена экстраполяция небольшой выборки данных. Соответствующий результат показан на рис. 11 для различных крайних точек массива исходных данных, для которых строится экстраполяция.

Листинг 2. Экстраполяция при помощи функции предсказания

$$y := (4.1\ \ 2.4\ \ 3\ \ 4.3\ \ 3.6\ \ 5.2\ \ 5.9)^T$$

$$k := \mathrm{rows}(y) - 1$$

$$m := 5 \qquad n := 10$$

$$A := \mathrm{predict}(y, m, n)$$

Лабораторная работа

Тема: «Использование статистических функции в математическом пакете MathCAD при обработке данных»

Цель работы: Научиться основным методам обработки опытных данных, овладеть навыками расчета с помощью ЭВМ в системе Mathcad основных числовых характеристик выборки, научиться выполнять различные виды регрессий оценивать их эффективность.

Задание

1. В соответствии с вариантом для заданных векторов x и y рассчитать следующие характеристики, используя статистические функции Mathcad :

- число элементов в выборке,
- наибольшее и наименьшее значения выборки,
- средние значения,
- среднее геометрическое векторов,
- дисперсию,
- среднее квадратическое отклонение,
- моду,
- медиану,
- кэффициенты корреляции и ковариации.

Результат оформить в виде таблицы.

2. Найти эмпирическую формулу зависимости одного из видов 1- 8 в разделе «Линеаризация некоторых зависимостей», заданных таблицей величин x и y с минимальной среднеквадратической ошибкой аппроксимации. Построить чертеж.

3. Студентам с нечетными номерами вариантов выполнить линейную регрессию с использованием встроенной функции line.

Студентам с четными номерами вариантов выполнить линейную регрессию с использованием встроенных функций intersept и slope.

Построить графики зависимостей.

4. Студентам с нечетными номерами вариантов выполнить полиномиальную регрессию.

Студентам с четными номерами вариантов выполнить регрессию отрезками полиномов.

Построить графики зависимостей.

5. Студентам с нечетными номерами вариантов выполнить регрессию специального вида: экспонентой, логарифмической функцией, синусоидой.

Студентам с четными номерами вариантов выполнить регрессию специального вида: логистической функцией, степенной функцией, двухпараметрической логарифмической функцией.

Построить графики зависимостей.

6. Выполнить регрессию общего вида, если функция представляет собой следующую линейную комбинацию:

$$f(x) = C_0 \cdot \frac{1}{x+2} + C_1 \cdot x^2 + C_2 \cdot e^x \quad .$$

Найти коэффициенты функции C_i и построить график.

Варианты заданий

		Номер варианта									
		1	2	3	4	5	6	7	8	9	10
x		y									
1	0.29	2.37	1.41	2.45	0.22	1.98	8.42	4.21	2.39	1.09	3.35
2	0.35	2.55	1.43	2.48	0.32	2.12	7.49	4.44	2.75	2.53	4.08
3	0.39	2.71	1.45	2.51	0.41	2.25	6.83	4.57	3.05	3.37	5.63
4	0.49	3.12	1.49	2.57	0.64	2.78	6.13	4.62	3.52	3.58	5.69
5	0.57	3.28	1.52	2.62	0.82	3.11	5.76	4.67	3.78	4.87	6.03
6	0.65	3.57	1.55	2.66	1.03	4.92	5.48	4.74	4.02	5.92	7.15

7	0.94	4.86	1.72	2.92	2.27	6.37	4.82	4.96	4.85	6.48	7.28
8	1.35	5.54	1.83	3.11	3.06	9.13	4.65	5.23	5.21	7.32	8.34
9	1.54	6.47	1.97	3.29	4.29	10.53	4.48	5.68	5.52	7.44	9.05
10	2.66	6.93	2.03	3.38	4.83	12.95	4.43	5.87	5.56	9.15	10.11

		Номер варианта									
		11	12	13	14	15	16	17	18	19	20
x		y									
1	0.49	3.37	2.41	2.45	0.52	2.98	9.42	5.41	3.39	2.09	4.35
2	0.54	3.55	2.43	2.78	0.62	3.12	8.49	5.44	3.75	3.53	5.08
3	0.59	3.71	2.45	3.21	0.71	3.25	7.83	5.47	4.05	4.37	6.63
4	0.69	4.12	2.49	3.57	0.94	3.78	7.13	5.52	4.52	4.58	6.69
5	0.76	4.28	2.52	3.62	1.12	4.11	6.76	5.57	4.78	5.87	7.03
6	0.84	4.57	2.55	3.66	1.33	5.92	5.48	5.63	5.02	6.92	8.15
7	1.22	5.86	2.72	3.92	2.57	6.37	5.82	5.96	5.85	7.48	8.28
8	1.43	6.54	2.83	4.11	3.36	10.13	5.65	6.23	6.21	8.32	9.34
9	1.72	7.47	2.97	4.29	4.59	11.53	5.48	6.68	6.52	8.44	10.05
10	1.84	7.93	3.03	4.38	5.13	13.95	5.43	6.87	6.56	10.15	11.11

		Номер варианта									
		21	22	23	24	25	26	27	28	29	30
x		y									
1	1.49	4.37	3.41	4.45	1.52	3.98	10.42	6.41	4.39	3.09	5.35
2	1.54	4.55	3.43	4.48	1.62	4.02	9.49	6.44	4.75	4.53	6.08
3	1.59	4.75	3.45	4.54	1.73	4.25	8.83	6.47	5.05	5.37	7.63
4	1.69	5.22	3.49	4.57	1.93	4.78	8.13	6.52	5.52	5.58	7.69
5	1.76	5.38	3.52	4.62	2.12	5.11	7.76	6.57	5.78	6.87	8.03
6	1.84	5.57	3.66	4.66	2.43	6.92	6.48	6.63	6.02	7.92	9.15
7	2.22	6.86	3.72	4.92	2.57	7.37	6.82	6.96	6.85	8.48	9.28
8	2.43	7.54	3.83	5.11	4.36	11.13	6.65	7.23	7.21	9.32	10.34

9	2.72	8.47	3.97	5.29	5.59	12.53	6.48	7.68	7.52	9.44	11.05
10	2.84	8.93	34.03	5.38	6.13	14.95	6.43	7.87	7.56	11.15	12.11

Порядок выполнения работы

1. Подготовить данные для ввода в ПЭВМ.
2. Получить решение заданий 1 - 6 по лабораторной работе в системе Mathcad.
3. Оценить погрешность расчетов.
4. Оформить результаты расчетов.
5. Составить отчет.

Содержание отчета

1. Название темы. Цель работы.
2. Данные ввода в ПЭВМ.
3. Результаты расчетов на ПЭВМ.
4. Графическая иллюстрация результатов.
5. Выводы.

Контрольные вопросы

1. Типовые статистические функции в MathCAD.
2. Статистические функции для векторов и матриц.
3. Как формулируется задача по подбору эмпирических формул.
4. Отличие интерполяции от регрессии.
5. Геометрическая интерпретация задачи построения эмпирической формулы.
6. Линейная регрессия.
7. Линеаризация некоторых зависимостей.
8. Реализация линейной регрессии общего вида.
9. Полиномиальная регрессия.
10. Регрессия отрезками полиномов.
11. Регрессия специального вида.
12. Экстраполяция функцией предсказания.

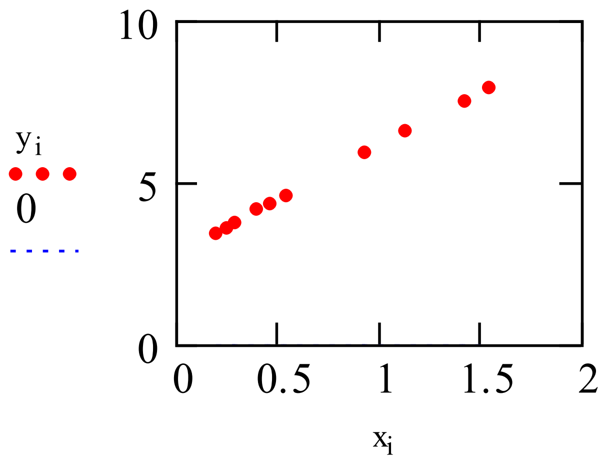
Пример выполнения лабораторной работы

Задание 1.

Для заданных векторов x и y расчет статистических характеристик, используя встроенные функции Mathcad :

$$i := 0..9$$

$$x := \begin{pmatrix} 0.19 \\ 0.24 \\ 0.29 \\ 0.39 \\ 0.46 \\ 0.54 \\ 0.92 \\ 1.13 \\ 1.42 \\ 1.54 \end{pmatrix} \quad y := \begin{pmatrix} 3.35 \\ 3.53 \\ 3.69 \\ 4.10 \\ 4.26 \\ 4.55 \\ 5.84 \\ 6.52 \\ 7.45 \\ 7.91 \end{pmatrix}$$



$length(x) = 10 \quad length(y) = 10$

$min(x) = 0.19 \quad min(y) = 3.35$

$max(x) = 1.54 \quad max(y) = 7.91$

$mean(x) = 0.712 \quad mean(y) = 5.12$

$gmean(x) = 0.558 \quad gmean(y) = 4.888$

$median(x) = 0.5 \quad median(y) = 4.405$

$stdev(x) = 0.477 \quad stdev(y) = 1.597$

$\mathrm{var}(x) = 0.227 \qquad \mathrm{var}(y) = 2.55$

$\mathrm{corr}(x, y) = 1 \qquad \mathrm{cvar}(x, y) = 0.762$

Таблица полученных значений:

Параметры	**Вектор x**	**Вектор y**
Число элементов вектора	10	10
Минимальное значение	0.19	3.35
Максимальное значение	1.54	7.91
Среднее значение	0.712	5.12
Среднее геометрическое	0.588	4.888
Медиана	0.5	4.405
Среднее квадратическое	0.477	1.597
Дисперсия (вариация)	0.227	2.55
Коэффициент корреляции	1	1
Коэффициент ковариации	0.762	0.762

Задание 2.

Найти эмпирическую формулу y1=f(x) зависимости величин x и y, заданных табличной формой.

x	0.19	0.24	0.29	0.39	0.46	0.54	0.92	1.13	1.42	1.54
y	3.35	3.53	3.69	4.1	4.26	4.55	5.84	6.52	7.45	7.91

Решение: Загружаем систему Mathcad. Исходя из заданной таблицы, определяем две матрицы-столбца размерности 10×1.
Задаем функции S_1 и S_2 (см.(5.3) и (5.4))

$$S1(y,y_1) := \frac{1}{7}\sum_i (y_i - y1_i)^2$$

$$S2(y,y_1) := \frac{1}{10}\sum_i \left|\frac{y_i - y1_i}{y_i}\right|$$

Печатаем параметр i изменяющийся от 0 до 9.

Затем для каждого из 8 видов функций с помощью операторов **intercept(x,y)** и **slope(x,y)** определяем коэффициенты a_i и b_i эмпирических зависимостей.

Печатаем a1:=intercept(x,y) a1=

b1:=slope(x,y) b1=.

Затем определяем функцию y1:

$$y1_i := a1 + b1 * x_i$$

и находим значения S1 и S2

S1(y,y1)= S2(y,y1)=

Данную процедуру повторяем.

Затем функции 1) – 8), рассмотренные в разделе «Линеаризация некоторых зависимостей» ранжируем в порядке возрастания значений S1 и S2. Функция, для которой сумма рангов S1 и S2 наименьшая, считается лучшей.

Анализируя полученные значения S1 и S2 (см. рис. 5.7–5.8), заключаем, что функция $y=a_1+b_1\cdot x=2.736+3.348\cdot x$ дает наилучший вид аналитической зависимости заданных значений величин x и y.

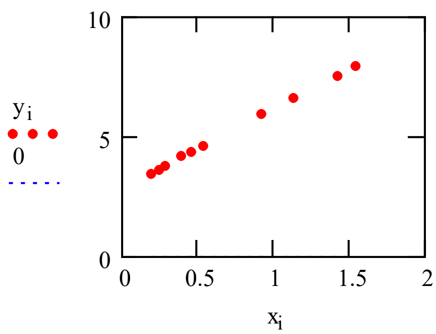
Строим график этой функции и на него наносим экспериментальные точки. Документ Mathcad с решением задачи 5.3 дан на рис. 5.7–5.8.

Листинг проделанной в системе Mathcad работы:

Исходные данные:

$i := 0..9$

$$x := \begin{pmatrix} 0.19 \\ 0.24 \\ 0.29 \\ 0.39 \\ 0.46 \\ 0.54 \\ 0.92 \\ 1.13 \\ 1.42 \\ 1.54 \end{pmatrix} \quad y := \begin{pmatrix} 3.35 \\ 3.53 \\ 3.69 \\ 4.10 \\ 4.26 \\ 4.55 \\ 5.84 \\ 6.52 \\ 7.45 \\ 7.91 \end{pmatrix}$$



Результаты расчетов:

$$S1(y,y1) := \frac{\sum_i \left(y_i - y1_i\right)^2}{7} \qquad S2(y,y1) := \frac{\sum_i \left|\frac{\left(y_i - y1_i\right)^2}{y_i}\right|}{10}$$

1)

$a1 := \text{intercept}(x,y) \quad a1 = 2.736$

$b1 := \text{slope}(x,y) \quad b1 = 3.348$

$y1_i := a1 + b1 \cdot x_i$

$S1(y,y1) = 1.009 \times 10^{-3} \quad S2(y,y1) = 1.504 \times 10^{-4}$

2)

$z_i := \ln\left(y_i\right)$

$a := \text{intercept}(x,z) \quad a2 := \exp(a) \quad a2 = 3.127$

$b2 := \text{slope}(x,z) \quad b2 = 0.627$

$y1_i := a2 \cdot e^{b2 \cdot x_i}$

$S1(y,y1) = 0.044964 \quad S2(y,y1) = 5.605036 \times 10^{-3}$

3)

$z_i := \ln\left(y_i\right)$

$a := \text{intercept}(x,z) \quad a3 := \exp(a) \quad a3 = 3.127$

$b := \text{slope}(x,z) \quad b3 := \exp(b) \quad b3 = 1.873$

$y1_i := a3 \cdot b3^{x_i}$

$S1(y,y1) = 0.044964 \quad S2(y,y1) = 5.605036 \times 10^{-3}$

4)

$z_i := \ln\left(y_i\right) \quad x1_i := \ln\left(x_i\right)$

$a := \text{intercept}(x1,z) \quad a4 := \exp(a) \quad a4 = 6.233$

$b4 := \text{slope}(x1,z) \quad b4 = 0.417$

$y1_i := a4 \cdot \left(x_i\right)^{b4}$

$S1(y,y1) = 0.07169 \quad S2(y,y1) = 9.101149 \times 10^{-3}$

5)

$z_i := \ln\left(y_i\right)$

$a5 := \text{intercept}(x,z) \quad a5 = 1.14$

$b5 := \mathrm{slope}(x, z) \qquad b5 = 0.627$

$y1_i := \exp\left(a5 + b5 \cdot x_i\right)$

$S1(y, y1) = 0.044964 \quad S2(y, y1) = 5.605036 \times 10^{-3}$

6)

$x1_i := \dfrac{1}{x_i}$

$a6 := \mathrm{intercept}(x1, y) \quad a6 = 7.231$

$b6 := \mathrm{slope}(x1, y) \qquad b6 = -0.926$

$y1_i := a6 + \dfrac{b6}{x_i}$

$S1(y, y1) = 0.873214 \quad S2(y, y1) = 0.123034$

7)

$x1_i := \exp\left(x_i\right)$

$a7 := \mathrm{intercept}(x1, y) \quad a7 = 2.095$

$b7 := \mathrm{slope}(x1, y) \qquad b7 = 1.315$

$y1_i := a7 + b7 \cdot \exp\left(x_i\right)$

$S1(y, y1) = 0.096787 \quad S2(y, y1) = 0.013477$

8)

$x1_i := \ln\left(x_i\right)$

$a8 := \mathrm{intercept}(x1, y) \quad a8 = 6.385$

$b8 := \mathrm{slope}(x1, y) \qquad b8 = 2.167$

$y1_i := a8 + b8 \cdot \ln\left(x_i\right)$

$S1(y, y1) = 0.20994 \quad S2(y, y1) = 0.030799$

График функции, дающей наилучшее приближение

$t := 0, 0.01 .. 2$

$y1(t) := 2.736 + 3.348t$

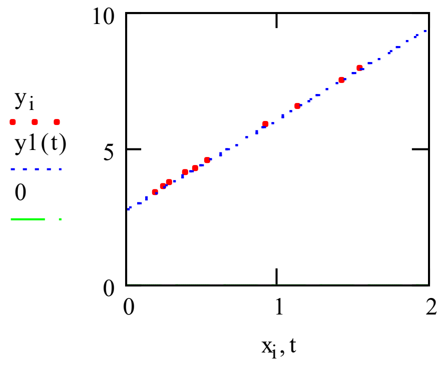


Рис. 5.7.

Вывод: функция $y=a_1+b_1 \cdot x=2.736+3.348 \cdot x$ дает наилучший вид аналитической зависимости заданных значений величин x и y.

Задание 3.

$x := (0.19\ 0.24\ 0.29\ 0.39\ 0.46\ 0.54\ 0.92\ 1.13\ 1.42\ 1.54)^T$

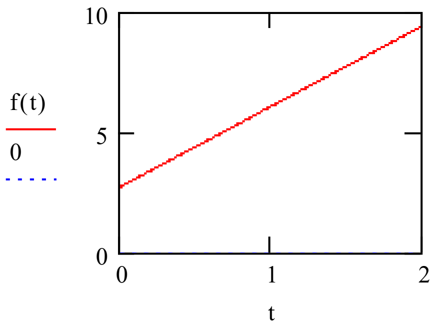
$y := (3.35\ 3.53\ 3.69\ 4.10\ 4.26\ 4.55\ 5.84\ 6.52\ 7.45\ 7.91)^T$

a)

$$\mathrm{line}(x,y) = \begin{pmatrix} 2.736 \\ 3.348 \end{pmatrix}$$

$f(t) := \mathrm{line}(x,y)_0 + \mathrm{line}(x,y)_1 \cdot t$

$t := 0, 0.01 .. 2$



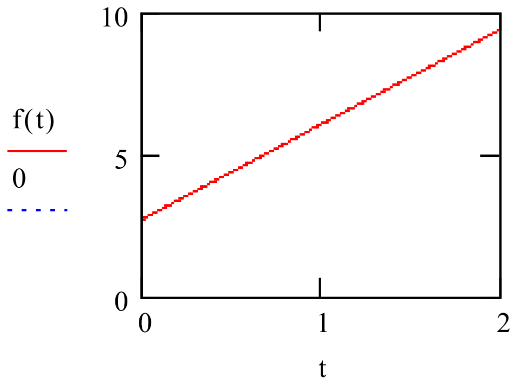
b)

$\mathrm{intercept}(x,y) = 2.736$

$\mathrm{slope}(x,y) = 3.348$

$f(t) := \mathrm{intercept}(x,y) + \mathrm{slope}(x,y)\, t$

$t := 0, 0.01 .. 2$



Задание 4.

$x := (0.19\ 0.24\ 0.29\ 0.39\ 0.46\ 0.54\ 0.92\ 1.13\ 1.42\ 1.54)^T$

$y := (3.35\ 3.53\ 3.69\ 4.10\ 4.26\ 4.55\ 5.84\ 6.52\ 7.45\ 7.91)^T$

Полиномиальная регрессия

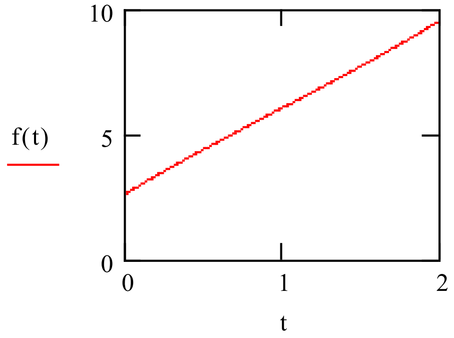
a)

$k := 3$

$s := \text{regress}(x, y, k)$

$f(t) := \text{interp}(s, x, y, t)$

$t := 0, 0.01..2$



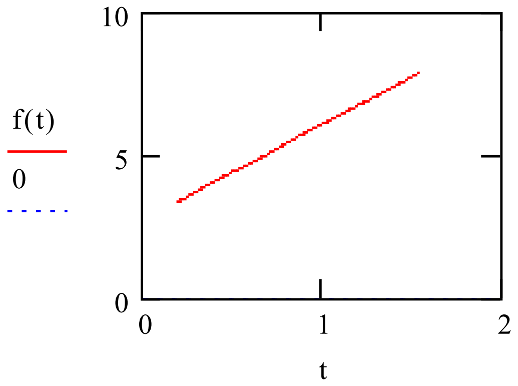
Регрессия отрезками полиномов

b)

$s := \text{loess}(x, y, 0.75)$

$f(t) := \text{interp}(s, x, y, t)$

$t := 0, 0.01..2$



Задание 5.

Примечание. Здесь приведены два примера. При выполнении остальных примеров воспользуйтесь теоретическим материалом раздела «Регрессия специального вида». Для получения соответствующих функций используйте указанные в этом разделе формулы.

$x := (0.19\ 0.24\ 0.29\ 0.39\ 0.46\ 0.54\ 0.92\ 1.13\ 1.42\ 1.54)^T$

$y := (3.35\ 3.53\ 3.69\ 4.10\ 4.26\ 4.55\ 5.84\ 6.52\ 7.45\ 7.91)^T$

$g := \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$

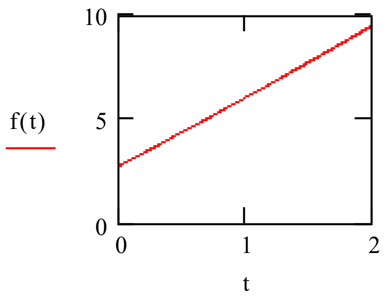
a)

$C := \text{expfit}(x, y, g)$

$$C = \begin{pmatrix} 60.055 \\ 0.053 \\ -57.276 \end{pmatrix}$$

$$f(t) := C_0 \cdot \exp(C_1 \cdot t) + C_2$$

$$t := 0, 0.01 .. 2$$



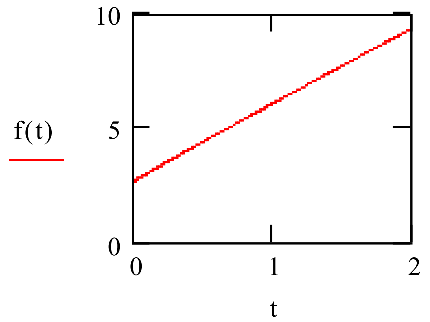
b)

$$C := \mathrm{pwrfit}(x, y, g)$$

$$C = \begin{pmatrix} 3.42 \\ 0.972 \\ 2.676 \end{pmatrix}$$

$$f(t) := C_0 \cdot t^{C_1} + C_2$$

$$t := 0, 0.01 .. 2$$



Задание 6.

$$x := (0.19\ \ 0.24\ \ 0.29\ \ 0.39\ \ 0.46\ \ 0.54\ \ 0.92\ \ 1.13\ \ 1.42\ \ 1.54)^T$$

$$y := (3.35\ \ 3.53\ \ 3.69\ \ 4.10\ \ 4.26\ \ 4.55\ \ 5.84\ \ 6.52\ \ 7.45\ \ 7.91)^T$$

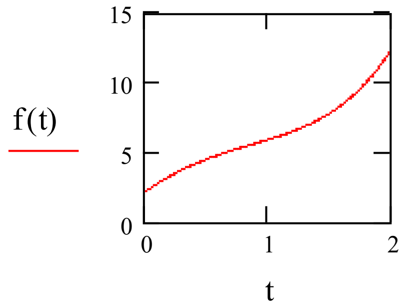
$$F(x) := \begin{pmatrix} \frac{1}{x+2} \\ x^2 \\ \exp(x) \end{pmatrix}$$

$$K := \mathrm{linfit}(x, y, F)$$

$$K = \begin{pmatrix} -5.539 \\ -5.684 \\ 4.954 \end{pmatrix}$$

$$f(t) := F(t) \cdot K$$

$$t := 0, 0.01 .. 2$$



Список литературы

1.<http://www.sistemair.ru/dok/mathcad12/Glava_12/Index03.htm>
2.<http://www.piter.com/attachment.php?barcode=978531800362&at=exc&n=0>
3.<http://ru.wikipedia.org/wiki/MathCad>
4. Самоучитель Mathcad 12. – СПб.: БХВ – Петербург, 2004.- 576 с.

Функции вычисления плотности распределения вероятности

Функции вычисления плотности вероятности распределения представлены следующим набором:

·dbeta(x,s1,s2) - бета-распределение (s1, s2>0 - параметры формы, 0 dbinom(k,n,p) - биномиальное распределение (возвращает значение вероятности P(x = k), где n и k целые числа, причем 0JkJn и 0JpJ1);

·dcauchy(x,l,s) - распределение Коши (l - параметр разложения, s>0 - параметр масштаба);

·dchisq(x,d) - хи-квадрат-распределение (x, d>0, причем d - число степеней свободы);

·dexp(x,r) - экспоненциальное распределение (r,x>0);

·dF(x,d1,d2) - распределение Фишера (d1, d2>0 - числа степеней свободы, x>0);

·dgamma(x,s) - гамма-распределение (s>0 - параметр формы, xi0);

·dgeom(k,p) - геометрическое распределение (0<p≤1 - вероятность успеха в отдельном испытании, k - целое неотрицательное число);

·dlnorm(x,m,s) - логарифмическое нормальное распределение (m - натуральный логарифм среднего значения, s>0 - натуральный логарифм среднеквадратичного отклонения, x>0);

·dlogis(x,l,s) - логистическое распределение (l - параметр разложения, s>0 - параметр масштаба);

·dnbinom(k,n,p) - отрицательное биномиальное распределение (n>0 и k>0 - целые числа, 0<p≤1);

·dnorm(x,m,s) - нормальное распределение (m - среднее значение, s>0 - среднеквадратичное отклонение);

·dpois(k,l) - распределение Пуассона (l>0, k - целое неотрицательное число);

·dt(x,d) - распределение Стьюдента (d>0 - число степеней свободы, x - вещественное число);

·dunif(x,a,b) - равномерное распределение (a и b - граничные точки интервала, причем a<b и a≤x≤b);

·dweibull(x,s) - распределение Вейбулла (s>0 - параметр формы).

Функции распределения

Функции распределения дают вероятность того, что случайная величина будет иметь значения, меньшие или равные определенной величине. Они представлены ниже:

·pbeta(x,s1,s2) - значение в точке x функции бета-распределения;

·pbinom(k,n,p) - значение функции распределения биномиального закона для k успехов в серии из n испытаний;

·pcauchy(x,l,s) - значение в точке x функции распределения Коши со шкалой параметров l и s;

·pchisq(x,d) - значение в точке x кумулятивного хи-квадрат-распределения, в котором d - степень свободы;

·pexp(x,r) - значение в точке x функции экспоненциального распределения;

·pF(x,d1,d2) - значение в точке x функции распределения Фишера;

·pgamma(x, s) - значение в точке x функции гамма-распределения;

·pgeom(k,p) - значение в точке x функции геометрического распределения;

·plnorm(x,m,s) - значение в точке x функции логарифмического нормального распределения;

·plogis(x,l,s) - значение в точке x функции логистического распределения;

·pnorm(x,m,s) - значение в точке x функции нормального распределения;

·pnbinom(k,n,p) - значение в точке x функции отрицательного биномиального распределения;

·ppois(k,l) - значение для k функции распределения Пуассона;

·pt(x,d) - значение в точке x функции распределения Стьюдента;

·punif(x,a,b) - значение в точке x функции равномерного распределения;

·pweibull(x,s) - значение в точке x функции распределения Вейбулла.

Квантили распределения

Следующая группа задает обращения (квантили) функций распределения случайных величин. Они позволяют по заданной вероятности вычислить такое значение x, при котором вероятность равна или меньше заданного значения p:

·qbeta(p,s1,s2) - квантили обратного бета-распределения с параметрами формы s1 и s2;

·qbinom(p,n,q) - количество успешных определений при решении уравнения Бернулли, если число испытаний равно n, вероятность этого количества успешных определений равна p, а q - вероятность успеха при однократном испытании (0≤q≤1 и 0≤p≤1);

·qcauchy(p,l,q) - квантили обратного распределения Коши со шкалой параметров l и s (s>0 и 0<p<1);

·qchisq(p,d) - квантили обратного хи-квадрат-распределения;

·qexp(p,r) - квантили обратного экспоненциального распределения, при котором параметр r>0 определяет частоту (0≤p<1);

·qF(p,d1,d2) - квантили обратного распределения Фишера, в котором d1 и d2 - степени свободы;

·qgamma(p,s) - квантили обратного гамма-распределения;

·qgeom(p,q) - квантили обратного геометрического распределения;

·qlnorm(p,m,s) - квантили обратного логарифмического нормального распределения;

·qlogis(p,l,s) - квантили обратного логистического распределения;

·qnbinom(p,n,q) - квантили обратного отрицательного биномиального распределения с размером n и вероятностью ошибки q;

·qnorm(p,m,s) - квантили обратного нормального распределения со средним значением m и стандартным отклонением s;

·qpois(p,l) - квантили обратного распределения Пуассона;

·qt(p,d) - квантили обратного распределения Стьюдента (d определяет степени свободы, d>0 и 0 qunif(p,a,b) - квантили обратного равномерного распределения;

·qweibull(p,s) - квантили обратного распределения Вейбулла.

План выпуска учеб.-метод. документ. 2014 г., поз. 28

Публикуется в авторской редакции

Подписано в свет 27.06.2014.
Гарнитура «Таймс». Уч.-изд. л. 1,3. Объем данных 1,2 Мбайт.

Федеральное государственное бюджетное образовательное учреждение
высшего профессионального образования
«Волгоградский государственный архитектурно-строительный университет»
400074, Волгоград, ул. Академическая, 1
http://www.vgasu.ru, info@vgasu.ru