

## ПЗ №3: Статистическая обработка данных в системе Mathcad

**Аннотация:** В лекции рассматриваются функции статической обработки встроенные в систему Mathcad. Особенности и недостатки данного программного пакета. Рассматривается возможность аппроксимации, представленные в системе Mathcad.

**Ключевые слова:** [Доверительная](#)

[вероятность](#), [Add](#), [create](#), [Data](#), [editor](#), [input](#), [logic](#), [math](#), [net](#), [OPEN](#), [programmable logic device](#), [rules](#), [surface](#), [target](#), [XML](#), [агрегация среднего значения](#), [в диапазоне от x до y](#), [генератор](#), [двойной щелчок](#), [жизненный цикл](#), [зависимость между признаками](#), [имя переменной](#), [ключевое слово](#), [логический](#), [маршрутам](#), [надежность](#), [область определения](#), [память](#), [работам](#), [связь](#), [таблица](#), [универсального множества](#), [фашификация фактических данных](#), [частотность](#), [шкала](#), [элементарные приемы](#), [язык программирования](#)

**Цель лекции:** Ознакомиться с основными статистическими функциями в системе Mathcad. Уяснить предназначение этих функций. Понять для каких целей применяются данные функции.

Mathcad — 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.

Широкий интерес к системе Mathcad привел к тому, что в России наконец-то появились книги по отдельным версиям. Система Mathcad была создана в 80-х годах в университете Станфорда (США).

Современные версии для ПК готовит *фирма MathSoft Application*.

Это универсальный

- Математический интерфейс.
- Есть мощная поддержка графики.
- Возможен импорт графики из других программ.
- Большое количество встроенных математических функций (сотни).
- Встроенные справочники по предметным областям.
- Возможна анимация.
- Символьная математика.

Достоинство – *программирование* на языке математики.

**Недостатки:**

- Это интерпретатор.
- Возможности программирования ограничены.

Системы Mathcad пользуются огромной популярностью во всем мире благодаря удобным средствам подготовки документов, имеющих вид обычных статей или книг. В то же время оказывается, что эти средства вполне достаточны для решения подавляющего большинства задач по математике, физике и других направлений науки и техники.

Mathcad является математически ориентированными универсальными системами. Помимо собственно вычислений они позволяют решать оформительские задачи. Они позволяют готовить статьи, книги, диссертации, научные отчеты, дипломные и курсовые проекты с доступным набором самых сложных математических формул и изысканным графическим представлением результатов.

С самого первого своего появления системы класса Mathcad имели удобный пользовательский *интерфейс* – совокупность средств общения с пользователем в виде масштабируемых и перемещаемых окон, кнопок и иных элементов. У этой системы есть эффективные средства типовой научной графики, они просты в применении и

интуитивно понятны. Словом, системы Mathcad ориентированы на массового пользователя – от ученика начальных классов до академика.

Несмотря на то, что эта *программа* в основном ориентирована на пользователей-непрограммистов, Mathcad также используется в сложных проектах, чтобы визуализировать результаты математического моделирования, путем использования распределённых вычислений и традиционных языков программирования. Также Mathcad часто используется в крупных инженерных проектах, где большое *значение* имеет трассируемость и соответствие стандартам.

Mathcad достаточно удобно использовать для обучения, вычислений и инженерных расчетов . Открытая *архитектура* приложения в сочетании с поддержкой технологий *NET* и *XML* позволяют легко интегрировать.

Мы рассмотрим наиболее часто применяемые функции для статистических расчетов, которые встроены в программу.

### Функция RND(X).

В статистических расчетах при моделировании различных физических процессов широко применяется *встроенная функция*  $RND(X)$ , инициализирующая *генератор* случайных чисел. Здесь  $X$  задаётся как ранжированная *переменная* и определяет число случайных чисел. Например, для получения десяти случайных чисел следует задать *интервал*  $X := 0..9$ .

### Функции аппроксимации.

Для представления физических закономерностей, а также при проведении научно-технических расчетов используются зависимости вида  $y(x)$ , но число точек этих зависимостей ограничено. При этом возникает задача приближенного вычисления значений функций в промежутках между узловыми точками и за их пределами. Такая задача решается аппроксимацией исходной зависимости, то есть ее подменой какой-либо достаточно простой функцией. В системе Mathcad предоставляется возможность аппроксимации двумя типами функций: кусочно-линейной и сплайновой.

При кусочно-линейной интерполяции вычисления дополнительных точек выполняются по линейной зависимости. То есть, при линейной аппроксимации узловые точки соединяются отрезками прямых линий, для чего используется *функция*  $Linterp(VX, VY, x)$  – линейная *интерполяция*, где  $VX, VY$  – векторы координат узловых точек, а  $x$  – заданный *аргумент*. При небольшом числе узловых точек линейная *интерполяция* оказывается довольно грубой. Даже первая производная *функция* аппроксимации получается с резкими скачками в узловых точках. Поэтому применяют функцию сплайн-аппроксимации.

При сплайн-аппроксимации исходная *функция* заменяется отрезками кубических полиномов, проходящих через три смежные узловые точки. *Коэффициенты* этой функции рассчитываются таким образом, чтобы первая и вторая *производные* были непрерывными. Линия, описывающая сплайн-функцию, напоминает по форме гибкую линейку, закрепленную в узловых точках.

Для осуществления сплайновой аппроксимации в Mathcad существуют четыре специальные встроенные функции:

- $VY, pspline(VX, VY)$  – возвращает вектор  $VS$  вторых производных при приближении в опорных точках к параболической кривой
- $lspline(VX, VY)$  – возвращает вектор  $VS$  вторых производных при приближении в опорных точках к прямой;

- $interp(VS, VX, VY, x)$  – возвращает значение  $y(x)$  для заданных  $VS, VX, VY$  и заданного значения  $x$

Таким образом, сплайн-аппроксимация производится в два этапа:

1. Для выбранного способа приближения к узловым точкам вычисляется вектор вторых производных функции  $y(x)$ , заданной векторами  $VX, VY$  ее значений с помощью одной из функций `cspline`, `pspline` или `lspline`.
2. С помощью функции  $Interp(VS, VX, VY, x)$  вычисляются значения  $y(x)$  для каждой искомой точки.

### Функции регрессии.

Широко распространенной задачей обработки данных является *представление* их совокупности некоторой функцией  $y(x)$ . Такое *представление* называется регрессией. Задача регрессии в том чтобы получить параметры функции такими, при которых функция приближает "облако" исходных точек с наименьшей квадратичной погрешностью.

Известны:

1. Линейная регрессия (прямая линия),
2. Полиномиальная регрессия (полином),
3. Линейная регрессия общего вида (линейная сумма произвольных функций),
4. Нелинейная регрессия общего вида (произвольная функция)

### Линейная регрессия

При линейной регрессии функция  $y(x)$  имеет вид  $y(x) = a + bx$  и описывает отрезок прямой.

Для проведения линейной регрессии в Mathcad имеются ряд встроенных функций:

- $corr(VX, VY)$  – функция корреляции. Возвращает скаляр – коэффициент корреляции векторов  $VX$  и  $VY$  (Пирсона);
- $intercept(VX, VY)$  – возвращает параметр  $a$  – смещение линии регрессии векторов  $VX$  и  $VY$  по вертикали (свободный член прямой регрессии);
- $slope(VX, VY)$  – возвращает значение параметра  $b$  – наклон линии регрессии векторов  $VX$  и  $VY$  (угловой коэффициент линии регрессии),

где  $VX$  и  $VY$  – векторы координат заданных точек. Координаты отдельной точки занимают в векторах  $VX$  и  $VY$  одинаковые позиции.

### Полиномиальная регрессия

При полиномиальной регрессии функция  $y(x)$  имеет вид:

$$y(x) = a_0 + a_1x + a_2x^2 + a_3x^3 + \dots + a_nx^n \quad (3.1)$$

Для реализации полиномиальной регрессии имеется *встроенная функция*  $regress(VX, VY, n)$ , которая возвращает *вектор*  $VS$ , содержащий *коэффициенты* полинома  $n$ -й степени. Эти *коэффициенты* вычисляются по всей совокупности заданных точек, т. е. глобально. Полученный *полином* наилучшим образом приближается к "облаку" точек с координатами, хранящимися в векторах  $VX$  и  $VY$ .

### Линейная регрессия общего вида

При линейной регрессии общего вида *функция*  $y(x)$  описывается как:

$$y(x) = F(x, K_1, K_2, \dots, K_n) = K_1 F_1(x) + K_2 F_2(x) + \dots + K_n F_n(x) \quad (3.2)$$

*Функция регрессии* является линейной комбинацией нескольких функций  $F_1(x), F_2(x), \dots, F_n(x)$ , причем сами эти функции могут быть и нелинейными, что резко расширяет возможности такой аппроксимации и распространяет ее на нелинейные функции.

Для реализации линейной регрессии общего вида используется *функция*  $linfit(VX, VY, F)$ , которая возвращает *вектор* коэффициентов  $K$  линейной регрессии общего вида, при котором *коэффициент корреляции* "облака" исходных точек и функции  $y(x)$  максимален. *Вектор*  $F$  функции  $linfit$  должен содержать функции  $F_1(x), F_2(x), \dots, F_n(x)$ , записанные в символьном виде. *Вектор*  $VX$  должен содержать абсциссы точек в возрастающем порядке, а ординаты в векторе  $VY$  должны соответствовать абсциссам в векторе  $VX$ .

### Нелинейная регрессия общего вида

Нелинейная регрессия общего вида это нахождение вектора  $K$  параметров произвольной функции  $F(x, K_1, K_2, \dots, K_n)$  при котором обеспечивается минимальная среднеквадратичная *погрешность* приближения "облака" исходных точек. Для этого используется *функция*

$$genfit(VX, VY, VS, F)$$

Эта *функция* возвращает *вектор*  $K$  параметров функции  $F$ , дающий минимальную среднеквадратичную *погрешность* приближения  $F(x, K_1, K_2, \dots, K_n)$  исходных данных.

*Функция*  $F$  должна быть вектором с символьными выражениями, они должны содержать аналитические выражения для исходной функции и ее производных по всем параметрам.

*Вектор*  $VS$  содержит начальные значения элементов вектора  $K$ .

При решении задачи возникает две проблемы:

1. Надо вычислять производные по переменным, обычно это делается с помощью символьных операций;
2. Необходимо применять функцию  $genfit$  в ее стандартном виде, в связи с этим все параметры заменяются на  $K_1, K_2, \dots, K_n$ .

## Функция предсказания

На практике нередко приходится сталкиваться с задачей расчёта последующих точек по ряду известных точек – задачей предсказания. Для предсказания поведения функциональной зависимости в Mathcad имеется функция: `predict (data, k, N)` – предсказание, где `data` – вектор данных, `k` – число точек предшествующих предсказанию, `N` – число предсказываемых точек данных.

Функция предсказания обеспечивает высокую *точность* при монотонных исходных функциях или функциях, представляемых полиномом невысокой степени.

Функция `predict` применима к предсказуемым событиям, поведение которых описывается реальной математической зависимостью. В странах со стабильной экономикой эта функция вполне применима для описания сезонных и стабильных колебаний курса валют, прибылей фирм, *продажа* товаров и.т.д.

## Краткие итоги

Рассмотрен пакет Codecad — система компьютерной алгебры из класса систем автоматизированного проектирования понятия эксперимента и математической модели.

Рассмотрены основные встроенные функции, применяемые для статистической обработки данных: сплайновой аппроксимации (`pspline`, `lspline`, `interp`), линейной регрессии (`corr`, `intercept`, `slope`), полиномиальной регрессии (`regress`), линейной регрессии общего вида (`linfit`), Нелинейная регрессия общего вида (`genfit`), функция предсказания (`predict`). Предназначение данных функций способы их применения для обработки данных.

Вопросы для самопроверки

1. Перечислите достоинства и недостатки системы Mathcad.
2. Перечислите основные функции, применяемые в статистической обработке.
3. Для чего применяется функция `RND(x)`?
4. В каких случаях применяются функции сплайновой аппроксимации?
5. Чем отличается функция линейной регрессии от функции линейной регрессии общего вида?
6. Дайте определение нелинейной регрессии общего вида?
7. Дайте определение функции предсказания?
8. В каких случаях функция предсказания применяется, а в каких применение ее является не целесообразным?