**ПРАКТИЧЕСКАЯ РАБОТА №12 РЕАЛИЗАЦИЯ АЛГОРИТМОВ РЕГРЕССИОННОГО АНАЛИЗА ДАННЫХ НА ЯЗЫКЕ C#**

***Цель практической работы:*** Ознакомиться с методом наименьших квадратов.

Разработать программу нахождения полинома наилучшего приближения.

***Задание:*** Экспериментально получены n − значений величины y при различных значениях величины x. Построить полиномы первой и второй степени, аппроксимирующие результаты эксперимента, с применением метода наименьших квадратов. Результаты выводятся в виде значений и графиков, полученных полиномов. Программа должна быть реализована на языке C#.

***12.1 Регрессионный анализ***

Регрессия – односторонняя случайная зависимость, устанавливающая соответствие между случайными переменными, то есть математическое выражение, отражающее связь между зависимой переменной y и независимыми переменными x при условии, что это выражение будет иметь статистическую значимость, то есть иметь малую вероятность случайного возникновения той или иной величины.

Регрессионный анализ — набор статистических методов исследования влияния одной или нескольких независимых переменных на зависимую y. Суть регрессионного анализа заключается в нахождении наиболее важных факторов, которые влияют на зависимую переменную. При его помощи можно получить конкретные сведения о том, какую форму и характер имеет зависимость между исследуемыми переменными. Связь между переменными может быть положительной, отрицательной, или неявной.

Создание регрессионной модели представляет собой итерационный процесс, направленный на поиск эффективных независимых переменных, чтобы объяснить зависимые переменные, которые мы пытаемся смоделировать или понять, запуская инструмент регрессии, чтобы определить, какие величины являются эффективными предсказателями.

Регрессии бывают линейные и нелинейные.

*.*

Нелинейные регрессии делятся на два класса: регрессии, нелинейные относительно включённых в анализ объясняющих переменных, но линейные по оцениваемым параметрам, и регрессии, нелинейные по оцениваемым параметрам.

Регрессии, нелинейные по объясняющим переменным:

Полином разных степеней .

Равносторонняя гипербола *.*

Регрессии, нелинейные по оцениваемым параметрам:

Степенная .

Показательная .

*12.2. Линейный регрессионный анализ*

При решении ряда прикладных задач нужно не только оценивать уже существующие статистические данные, но и составлять функциональную зависимость, позволяющую производить моделирование характеристик и значений, предсказывать дальнейшее поведение оцениваемых величин. Для построения функций наиболее популярными методами являются интерполяция и аппроксимация исходных значений.

Интерполяция – способ нахождения промежуточных значений величины по имеющемуся дискретному набору известных значений.

Её применяют в случае, когда требуется найти значение функции y(x) при значении аргумента , принадлежащего интервалу , но не совпадающего по значению ни с одним известным значением.

Интерполяция часто встречается при ограниченности возможностей при проведении эксперимента. Если проведение эксперимента является дорогим или трудоёмким, размер выборки может быть достаточно мал. При этом во многих случаях аналитическое выражение функции y(x) не известно и получить его по известным значениям в большинстве случаем невозможно.

Полином, полученный в результате интерполяции, требует совпадения полученной функции со статистическими значения в заданных точках и, как правило, высокой степени, на каждом отрезке полином с уникальными коэффициентами, что на практике вызывает трудности.

Поэтому часто для проведения анализа решают задачу аппроксимации данных. Аппроксимация – замена одних математических объектов другими, в том или ином смысле близкими к исходным.

При этом методе функция проходит не через узлы интерполяции, а между ними. Одним из методов аппроксимации является метод наименьших квадратов.

Смысл метода наименьших квадратов заключается в нахождении таких значений , при которых сумма квадратов отклонений (ошибок) будет стремиться к минимуму

Т.к. каждое значение в общем случае «сопровождается» соответствующим коэффициентом , то задача сводится к нахождению данных коэффициентов. Введём обозначение функции

*.*

Тогда, на основе обращения в точке минимума функции F в нуль её производных, для определения вышеупомянутых коэффициентов составляется нормальная система:

.

Существенным недостатком метода является громоздкость вычислений, вследствие чего к нему прибегают при достаточно точных экспериментальных данных при необходимости получения очень точных значений функции. Этот метод помогает добиться эффекта сглаживания ошибок и получить более точного прогнозируемого результата.

*12.3 Метод наименьших квадратов*

Введём функцию определяющую близость полинома и статистических данных:

Из условия минимума найдём

.

Запишем систему уравнений:

,

k = 0, 1… m,

*.*

Если все точки различны, то система имеет единственное решение. Решая полученную систему относительно находим явное выражение для аппроксимирующего полинома .

F (=

Запишем систему уравнений:

F (=.

Запишем систему уравнений:

Для того чтобы построить полиномы первой и второй степени, нужно найти соответственные значения для уравнений:

Для этого нужно решить системы уравнений для каждого полинома.

Для полинома первой степени система уравнений выглядит следующим образом:

.

Коэффициенты для полинома первой степени находятся через следующую систему уравнений:

.

Для полинома второй степени система уравнений выглядит следующим образом:

Коэффициенты для полинома второй степени соответственно находятся через следующую систему уравнений:

Подставив вместо полученное выражение упростим систему из трёх уравнений в систему из двух, получим:

Приведя подобные в полученной системе уравнений получим:

Для упрощения уравнений проведём замены переменных:

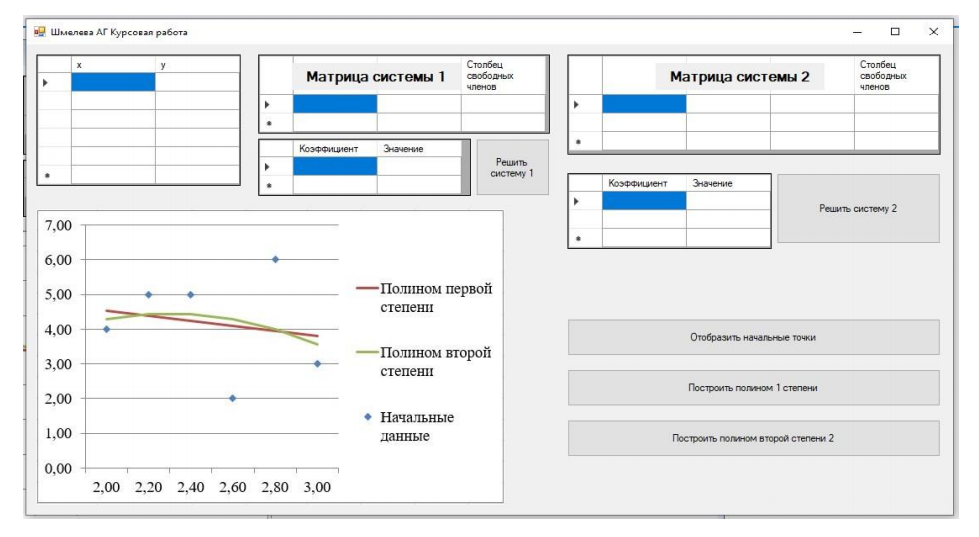
Полученное выражение схоже с первым случаем:

.

Тогда коэффициенты , находятся следующим образом:

А выражение для коэффициент выглядит следующим образом:

.

Рис. 12.1. Пример интерфейса

В результате работы программы на экране должна отображаться следующая информация - таблица исходных данных и график. На графике присутствуют полиномы первой и второй степени, а также исходные данные.

Для построения графика требуется использовать компонент **Chart**. При этом необходимо в свойстве **Series** выбрать тип графика установив **ChartType** в **Spline.** Здесь же настраиваются толщина и цвет графика. Также для каждого графика настраивается его легенда.

***12.4. Выполнение индивидуального задания***

Вариант выбирается в соответствии с номером зачетной книжки (последняя цифра в номере). Например, если номер зачетной книжки заканчивается на 2 – вариант 2, если на 0, то вариант 10.Вариант 1

|  |  |
| --- | --- |
| x | y |
| 0,0 | 3,0 |
| 0,2 | 6,0 |
| 0,4 | 3,0 |
| 0,6 | 6,0 |
| 0,8 | 4,0 |
| 1,0 | 3,0 |

Вариант 2

|  |  |
| --- | --- |
| x | y |
| 0,0 | 5,0 |
| 0,2 | 5,0 |
| 0,4 | 4,0 |
| 0,6 | 4,0 |
| 0,8 | 6,0 |
| 1,0 | 6,0 |

Вариант 3

|  |  |
| --- | --- |
| x | y |
| 3,0 | 2,0 |
| 3,2 | 3,0 |
| 3,4 | 3,0 |
| 3,6 | 3,0 |
| 3,8 | 2,0 |
| 4,0 | 4,0 |

Вариант 4

|  |  |
| --- | --- |
| x | y |
| 3,0 | 6,0 |
| 3,2 | 2,0 |
| 3,4 | 6,0 |
| 3,6 | 4,0 |
| 3,8 | 3,0 |
| 4,0 | 4,0 |

Вариант 5

|  |  |
| --- | --- |
| x | y |
| 5,0 | 2,0 |
| 5,2 | 4,0 |
| 5,4 | 4,0 |
| 5,6 | 3,0 |
| 5,8 | 3,0 |
| 6,0 | 3,0 |

Вариант 6

|  |  |
| --- | --- |
| x | y |
| 4,0 | 4,0 |
| 4,2 | 3,0 |
| 4,4 | 6,0 |
| 4,6 | 6,0 |
| 4,8 | 4,0 |
| 5,0 | 4,0 |

Вариант 7

|  |  |
| --- | --- |
| x | y |
| 1,0 | 2,0 |
| 1,2 | 6,0 |
| 1,4 | 4,0 |
| 1,6 | 4,0 |
| 1,8 | 2,0 |
| 2,0 | 5,0 |

Вариант 8

|  |  |
| --- | --- |
| x | y |
| 5,0 | 3,0 |
| 5,2 | 2,0 |
| 5,4 | 5,0 |
| 5,6 | 2,0 |
| 5,8 | 2,0 |
| 6,0 | 3,0 |

Вариант 9

|  |  |
| --- | --- |
| x | y |
| 2,0 | 4,0 |
| 2,2 | 2,0 |
| 2,4 | 4,0 |
| 2,6 | 2,0 |
| 2,8 | 5,0 |
| 3,0 | 2,0 |

Вариант 10

|  |  |
| --- | --- |
| x | y |
| 0,0 | 6,0 |
| 0,2 | 3,0 |
| 0,4 | 2,0 |
| 0,6 | 6,0 |
| 0,8 | 2,0 |
| 1,0 | 5,0 |