# АНАЛИЗ ВРЕМЕННЫХ РЯДОВ

## 1. Общие сведения

 Очень часто данные об объектах представлены в виде последовательностей значений свойств объектов в различные моменты времени. ***Временным рядом***  называется набор значений какого-либо показателя за несколько последовательных периодов времени.

 Основная задача обработки временных рядов заключается в выявлении особенностей ряда с целью возможного прогноза свойств объекта в последующие моменты времени.

 Анализ поведения объектов во времени показывает, что они обладают рядом общих закономерностей.

1. В поведении объекта может присутствовать регулярная составляющая. Она обычно отражает общую тенденцию развития объекта во времени. Например, спрос на нефть имеет устойчивую тенденцию к возрастанию – рис.1.

Рис. 1

 Представленная на рис.1 регулярная составляющая представляет собой прямую линию. Но не менее часто встречаются и другие ее виды – рис.2.

Рис.1. Пример линейной

составляющей

Рис.2. Примеры нелинейных составляющих

1. На регулярную составляющую, как правило, накладываются периодические колебания свойств. Эти колебания могут быть связаны с влиянием времени года на свойства объекта – так называемые сезонные колебания. Например, потребление нефти в зимнее время меньше, чем в летнее. В результате этого вместо рис.1. будет иметь место рис.3.



#### Рис. 3. Пример временного ряда с сезонной составляющей

1. К сожалению, реальные данные практически никогда не выстраиваются в идеально гладкие кривые, чаще всего они имеют вид, представленный на рис.4.



Рис.4. Пример наложения случайных возмущений на временной ряд.

 Вызвано это тем, что на объект влияет слишком большое количество других факторов, часто действующих в противоположном направлении. Эти факторы не столь существенны, как указанные первые два, но их, как правило, во-первых, очень много и, во-вторых, практически невозможно учесть. Поэтому вклад этих факторов рассматривается как случайная составляющая.

 Применительно к ценам на нефть такими факторами могут быть и стихийные бедствия в районах нефтедобычи, и отдельные заявления руководителей ведущих экономических стран, и попытки скрыть истинные величины потребления, и даже элементарные опечатки при представлении данных.

 Основным методом анализа временных рядов является метод декомпозиции ряда на указанные составляющие.

## 2. Пример

Пусть имеющиеся данные (50 точек) размещены следующим образом:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | A | B | C | D |
| 5 |  |  |  |  |
| 6 |  |  | Yreal |  |
| 7 |  | 1 | 315,3962 |  |
| 8 |  | 2 | 313,9911 |  |
| 9 |  | 3 | 276,0792 |  |
| 10 |  | **…** | **…** |  |
|  |  | **…** | **…** |  |
| 55 |  | 49 | 13,01288 |  |
| 56 |  | 50 | 35,34142 |  |
| 57 |  |  |  |  |
| 58 |  |  |  |  |

Сначала подбирается наиболее подходящая регулярная составляющая ряда. Для рассматриваемого примера использованы следующие варианты этой составляющей:

линейная y=a0 + a1\*x; (1)

логарифмическая y=a0 + a1\* Ln(x); (2)

экспоненциальная y=a0 + a1\*Exp(a2\*x) (3)

степенная y=a0 + a1\*x^a2; (4)

Здесь а0, а1, а2 – некоторые числовые коэффициенты.

Для расчета коэффициентов этих уравнений используется средство Excel «Поиск решения».

// Отметим, что при использовании такого средства Excel, как наложение

//трендов на диаграмму, можно вычислить коэффициенты только //первых двух типов уравнений. Поэтому показывается более общий //метод вычисления коэффициентов для уравнений произвольных типов.

Для этого выполняются следующие шаги:

1. На листе размещаются примерные значения коэффициентов (например, как в следующей таблице)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | M | N | O | P |
| 1 |  |  |  |  |
| 2 |  |  |  |  |
| 3 | 1 | y=a0 + a1\*x; |  |
| 4 |  | a0= | 1 |  |
| 5 |  | a1= | 1 |  |
| 6 |  |  |  |  |
| 7 | 2 | y=a0 + a1\* Ln(x); |  |
| 8 |  | a0= | 1 |  |
| 9 |  | a1= | 1 |  |
| 10 |  |  |  |  |
| 11 | 3 | y=a0 + a1\*Exp(a2\*x) |  |
| 12 |  | a0= | 100 |  |
| 13 |  | a1= | 200 |  |
| 14 |  | a2= | -0,04 |  |
| 15 |  |  |  |  |
| 16 | 4 | y=a0 + a1\*x^a2 |  |
| 17 |  | a0= | 300 |  |
| 18 |  | a1= | -50 |  |
| 19 |  | a2= | -0,01 |  |
| 20 |  |  |  |  |
| 21 |  | y=a0+a1\*sin(a2+a3\*x) |
| 22 |  | a0= | 300 |  |
| 23 |  | a1= | -30 |  |
| 24 |  | a2= | 0 |  |
| 25 |  | a3= | 0,5 |  |
| 26 |  |  |  |  |
|  |  |  |  |  |

1. В ячейку D7 вводится формула (1): =$O$4+$O$5\*B7;
2. В ячейку E7 вводится формула (2): =$O$8+$O$9\*LOG(B7);
3. В ячейку F7 вводится формула (3): =$O$12+$O$13\*EXP($O$14\*B7;
4. В ячейку G7 вводится формула (4) =$O$17+$O$18\*B7^$O$19.
5. Все указанные формулы копируются до 56 строки.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | A | B | C | D | E | F | G | H |
| 5 |  |  |  |  |  |  |  |  |
| 6 |  |  | Yreal | Y1 | Y2 | Y3 | Y4 |  |
| 7 |  | 1 | 315,3962 | 185,4513 | 316,3298 | 329,6659 | 359,1789 |  |
| 8 |  | 2 | 313,9911 | 181,6173 | 263,8741 | 293,1459 | 277,8989 |  |
| 9 |  | 3 | 276,0792 | 177,7833 | 233,1895 | 261,3638 | 235,9401 |  |
| 10 |  | **…** | **…** | **…** | **…** | **…** | **…** |  |
|  |  | **…** | **…** | **…** | **…** | **…** | **…** |  |
| 55 |  | 49 | 13,01288 | 1,417789 | 21,8063 | 48,51841 | 31,38272 |  |
| 56 |  | 50 | 35,34142 | -2,41624 | 20,2774 | 48,47208 | 30,32436 |  |
| 57 |  |  |  |  |  |  |  |  |
| 58 |  |  |  |  |  |  |  |  |

1. Очевидно, что из-за случайности начальных значений коэффициентов, полученные в столбцах D-G расчетные Y мало соответствуют реальным значениям Y, расположенным в столбце C. Для их «подгонки» используется метод наименьших квадратов, согласно которому коэффициенты уравнений подбираются таким образом, чтобы сумма квадратов отклонений расчетных данных от реальных была минимальной.
2. Для этого в ячейку D57 вводится формула:

=СУММКВРАЗН($C7:$C56;D7:D56),

которая затем копируется до столбца G.

1. Для подбора коэффициентов первого уравнения выполняется последовательность команд:

***Курсор устанавливается в ячейку D57 > Вызывается средство «Поиск решения» > в нем устанавливаются следующие параметры: «Целевая ячейка» - $D$57; «Минимум»; Изменяя ячейки $O$4:$O$5 > Найти решение***

1. Аналогично для подбора коэффициентов второго уравнения:

***Курсор устанавливается в ячейку E57 > Вызывается средство «Поиск решения» > в нем устанавливаются следующие параметры: «Целевая ячейка» - $E$57; «Минимум»; Изменяя ячейки $O$8:$O$9 > Найти решение***

Результаты подбора приведены в таблице:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | M | N | O | P |
| 2 |  |  |  |  |
| 3 | 1 | y=a0 + a1\*x; |  |
| 4 |  | a0= | 189,2853 |  |
| 5 |  | a1= | -3,83403 |  |
| 6 |  |  |  |  |
| 7 | 2 | y=a0 + a1\* Ln(x); |  |
| 8 |  | a0= | 316,3087 |  |
| 9 |  | a1= | -174,302 |  |
| 10 |  |  |  |  |

1. Третье уравнение имеет более сложное (хотя бы потому, что имеет три параметра). Заложенные в «Поиск решения» методы поиска в общем случае для подобных уравнений не срабатывают. Они хорошо работают только в том случае, когда начальные значения коэффициентов достаточно близки к «правильным». Для того, чтобы указать «правильные» начальные значения коэффициентов необходимо понимание особенностей обрабатываемого уравнения. У студентов подобный навык отсутствует. Поэтому можно поступить следующим образом:

- по данным столбцов C и F строится диаграмма в виде графика;

- на ряд Yreal накладывается экспоненциальный тренд (серая линия) с показом уравнения на диаграмме;



- выведенное Excel уравнение отличается от уравнения (3) отсутствием параметра a0. Но это не столь важно, поскольку этот параметр отвечает только за сдвиг графика относительно оси Y. А вот остальные параметры определяют форму кривой.

- вводим значения 200 и -0,04 в ячейки O13 и O14 и вручную подгоняем значение a1, так чтобы графики совпадали наилучшим образом. Линию тренда при этом можно удалить.

1. Для точного подбора коэффициентов выполняется последовательность команд:

***Курсор устанавливается в ячейку F57 > Вызывается средство «Поиск решения» > в нем устанавливаются следующие параметры: «Целевая ячейка» - $F$57; «Минимум»; Изменяя ячейки $O$12:$O$14 > Найти решение***

1. Точно также рассчитываются коэффициенты уравнения (4). Только здесь используется степенной тренд.
2. Результаты расчетов приведены в таблицах.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | M | N | O | P |
| 9 |  |  |  |  |
| 10 |  |  |  |  |
| 11 | 3 | y=a0 + a1\*Exp(a2\*x) |  |
| 12 |  | a0= | 48,16145 |  |
| 13 |  | a1= | 323,4681 |  |
| 14 |  | a2= | -0,13895 |  |
| 15 |  |  |  |  |
| 16 | 4 | y=a0 + a1\*x^a2 |  |
| 17 |  | a0= | -200,684 |  |
| 18 |  | a1= | 559,852 |  |
| 19 |  | a2= | -0,22629 |  |
| 20 |  |  |  |  |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | B | C | D | E | F | G | H |
| 5 |  |  |  |  |  |  |  |
| 6 |  | Yreal | Y1 | Y2 | Y3 | Y4 |  |
| 7 | 1 | 315,3962 | 185,4513 | 316,3087 | 329,6653 | 359,1682 |  |
| 8 | 2 | 313,9911 | 181,6172 | 263,8384 | 293,1452 | 277,8951 |  |
| 9 | 3 | 276,0792 | 177,7832 | 233,1453 | 261,3629 | 235,9387 |  |
|  |  | **…** | **…** | **…** | **…** | **…** |  |
|  |  | **…** | **…** | **…** | **…** | **…** |  |
| 55 | 49 | 13,01288 | 1,417835 | 21,70337 | 48,51858 | 31,3765 |  |
| 56 | 50 | 35,34142 | -2,41619 | 20,17405 | 48,47225 | 30,31802 |  |
| 57 |  |  | 104293,6 | 35366,74 | 24740,76 | 30476,19 |  |
| 58 |  |  |  |  |  |  |  |

1. В результате расчетов в строке 57 вычислены значения сумм квадратов разностей между реальными значениями Y и четырьмя расчетными Y. При этом минимальной оказалась сумма квадратов разностей для третьего уравнения. Следовательно это уравнение и следует выбрать в качестве основного тренда (или регулярной составляющей) для имеющихся данных.
2. Вычислим остатки ряда первого уровня. Для этого в ячейку H7 вводим формулу: =C7-F7, которую копируем до 56 строки.
3. Для столбца H построим диаграмму типа график. Построенный график напоминает сильно искаженную синусоиду.



1. Для определения параметров этой синусоиды используется ее общее уравнение:

y=a0+a1\*Sin(a2+a3\*x) (5)

1. Для определения параметров уравнения используем то же средство «Поиск решения» по методу наименьших квадратов. Для этого на листе задаются начальные значения параметров уравнения:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | M | N | O | P |
| 19 |  |  |  |  |
| 20 |  |  |  |  |
| 21 | 5 | y=a0+a1\*sin(a2+a3\*x) |
| 22 |  | a0= | 0 |  |
| 23 |  | a1= | 40 |  |
| 24 |  | a2= | 0 |  |
| 25 |  | a3= | 0,5 |  |
| 25 |  |  |  |  |

В ячейку I7 вводится формула: =$O$22+$O$23\*SIN($O$24+$O$25\*B7), которая копируется до 56 строки.

В ячейку I57 вводится формула: =СУММКВРАЗН(H7:H56;I7:I56).

1. Вводим примерные значения коэффициентов уравнения (5).
* коэффициент a0 отвечает за сдвиг синусоиды по оси Y. Судя по графику этот сдвиг примерно равен нулю;
* коэффициент a1 отвечает за размах синусоиды по оси Y. Судя по графику этот размах примерно равен 40;
* коэффициент a2 отвечает за сдвиг синусоиды по оси X. Судя по графику этот сдвиг примерно равен нулю;
* коэффициент a3 отвечает за сжатие или растяжение синусоиды по оси X. Судя по графику этот сдвиг примерно равен 0,5.
1. Уточняем значения коэффициентов выполнив команды:

***Курсор устанавливается в ячейку I57 > Вызывается средство «Поиск решения» > в нем устанавливаются следующие параметры: «Целевая ячейка» - $I$57; «Минимум»; Изменяя ячейки $O$22:$O$25 > Найти решение***

1. В результате получим уточненные значения коэффициентов:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | M | N | O | P |
|  |  |  |  |  |
| 20 |  |  |  |  |
| 21 | 5 | y=a0+a1\*sin(a2+a3\*x) |
| 22 |  | a0= | 0,086249 |  |
| 23 |  | a1= | 26,55213 |  |
| 24 |  | a2= | 0,168449 |  |
| 25 |  | a3= | 0,492218 |  |
| 26 |  |  |  |  |

1. Для столбца I построим диаграмму типа график.



1. Находим остатки ряда второго уровня. Для этого в ячейку J7 вводим формулу: =H7-I7 и копируем ее до строки 56. Для столбца J строим график.



1. График представляет собой некоторые случайные данные, в которых отсутствует какая-либо закономерность. Т. е. дальнейшая декомпозиция ряда не имеет смысла. Поэтому итоговое выражение для ряда представляет собой сумму уравнений (3) и (5) с найденными коэффициентами:

Y=48,1 + 323,47\*Exp(-0,14\*x) + 0,09 + 26,55\*Sin(0,17 + 0,49\*x)

1. Используя полученное уравнение можно спрогнозировать значение ряда на последующие периоды времени.

Например, для x=51 значение Y будет равно = 52, 19.