

Олександр Павленко

старший викладач

кафедри системного аналізу

Мелітопольського інституту

державного та муніципального

управління «Класичного

приватного університету»

ЗГЛАДЖУВАННЯ ЯК ОСНОВНИЙ МЕТОД АНАЛІЗУ ЧАСОВИХ РЯДІВ

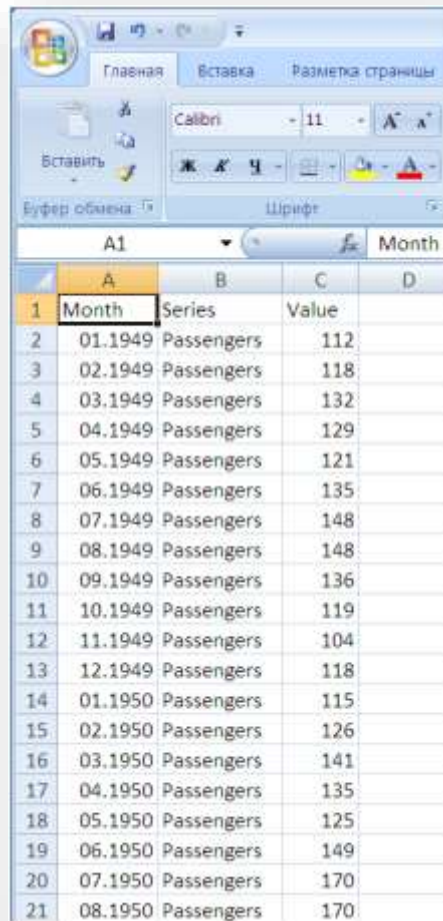
Часовим рядом називається послідовність значень, що змінюються у часі [4, с. 18]. Про деякі прості, але ефективних підходи до роботи з подібними послідовностями розповідається у даній статті. Прикладів таких даних можна зустріти дуже багато - котирування валют, обсяги продаж, звернення клієнтів, дані в різних прикладних науках (соціологія, метеорологія, геологія, спостереження у фізиці) і багато іншого.

Ряди є поширеною і важливою формою опису даних, оскільки дозволяють спостерігати всю історію зміни даного нас значення. Це дає нам можливість судити про «типовому» поведінці величини і про відхилення від такої поведінки [3, с. 34].

Переді мною постало завдання вибрати набір даних, на якому можна було б наочно продемонструвати особливості часових рядів. Я вирішив скористатися статистикою пасажиропотоку на міжнародних авіалініях, оскільки цей набір даних вельми наочний і став свого роду стандартним. Ряд описує кількість пасажирів міжнародних авіаліній на місяць (у тисячах) за період з 1949 по 1960 року.

Для роботи з рядами існує програма Prognoz Platform, в якій є відповідний інструмент «Аналіз часових рядів». Перед імпортом даних у файл потрібно додати стовпець з датою, щоб була прив'язка значень до часу, і

стовпець з ім'ям ряду для кожного спостереження. На рисунку 1 видно, як виглядає вихідний файл, який імпортований в Prognoz Platform за допомогою майстра імпорту безпосередньо з інструмента аналізу часових рядів.



The image shows a screenshot of a spreadsheet application window. The window title is 'Главная' (Home) and it has tabs for 'Главная', 'Вставка' (Insert), and 'Разметка страницы' (Page Layout). The font is 'Calibri' size 11. The active cell is A1, which contains the text 'Month'. The spreadsheet data is as follows:

	A	B	C	D
1	Month	Series	Value	
2	01.1949	Passengers	112	
3	02.1949	Passengers	118	
4	03.1949	Passengers	132	
5	04.1949	Passengers	129	
6	05.1949	Passengers	121	
7	06.1949	Passengers	135	
8	07.1949	Passengers	148	
9	08.1949	Passengers	148	
10	09.1949	Passengers	136	
11	10.1949	Passengers	119	
12	11.1949	Passengers	104	
13	12.1949	Passengers	118	
14	01.1950	Passengers	115	
15	02.1950	Passengers	126	
16	03.1950	Passengers	141	
17	04.1950	Passengers	135	
18	05.1950	Passengers	125	
19	06.1950	Passengers	149	
20	07.1950	Passengers	170	
21	08.1950	Passengers	170	

Рис. 1. Часовий ряд

Перше, що ми зазвичай робимо із часовим рядом, це відображаємо його на графіку. Prognoz Platform дозволяє побудувати графік, просто «перетягнувши» ряд в робочу книгу.

Символ «М» в кінці імені ряду означає, що ряд має місячну динаміку (інтервал між спостереженнями дорівнює одному місяцю) [1, с. 112]. Вже з графіку ми бачимо, що ряд демонструє дві особливості:

- тренд - на нашому графіку це довгострокове зростання спостережуваних значень. Видно, що тренд практично лінійний.
- сезонність - на графіку це періодичні коливання величини.

Наш ряд досить «акуратний», проте часто зустрічаються ряди, які крім двох описаних вище характеристик демонструють ще одну - наявність «шуму»,

тобто випадкових варіацій в тій чи іншій формі. Приклад такого ряду можна побачити на рисунку 2. Це синусоїдальний сигнал, змішаний з випадковою величиною.

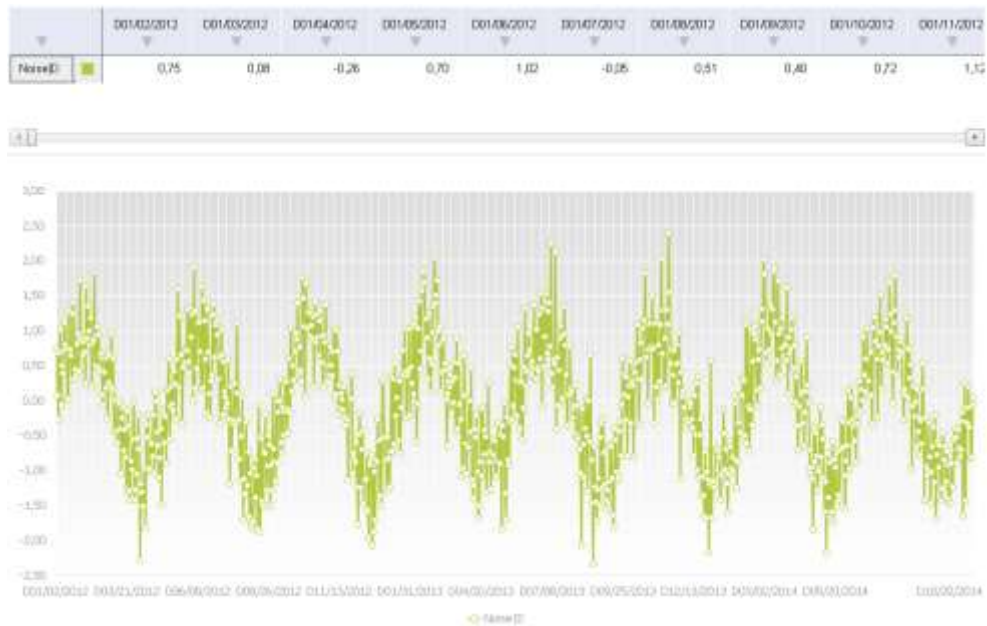


Рис. 2. Часовий ряд з шумом

При аналізі рядів нас цікавить виявлення їх структури та оцінка всіх основних компонентів - тренду, сезонності, шуму та інших особливостей, а також можливість будувати прогнози зміни величини в майбутніх періодах.

При роботі з рядами наявність шуму часто ускладнює аналіз структури ряду. Щоб виключити його вплив і краще побачити структуру ряду, можна використовувати методи згладжування рядів [2, с. 87].

Ковзаюче середнє. Найпростіший метод згладжування рядів - ковзаюче середнє. Ідея полягає в тому, що для будь-якої непарної кількості точок послідовності ряду замінювати центральну точку на середнє арифметичне інших точок:

$$s_i = \frac{1}{2k+1} \sum_{j=-k}^k x_{i+j}, \text{ де } x_i - \text{вихідний ряд, } s_i - \text{згладжений ряд.}$$

Нижче можна побачити результат застосування даного алгоритму до двох наших рядів. Prognoz Platform за замовчуванням пропонує використовувати згладжування з розміром вікна в 5 точок (k в нашій формулі вище дорівнюватиме 2). Слід зазначити, що згладжений сигнал вже не так схильний до впливу шуму, однак разом з шумом, природно, пропадає і частину корисної

інформації про динаміку ряду. Також видно, що у згладженому ряді відсутні перші (і також останні) k точки. Це пов'язано з тим, що згладжування виконується для центральної точки вікна (у нашому випадку для третьої точки), після чого вікно зсувається на одну точку, і обчислення повторюються. Для другого, випадкового ряду, використовується згладжування з вікном рівним 30, щоб краще виявити структуру ряду. Оскільки ряд «високочастотний», точок дуже багато.

Експоненціальне згладжування. Більш просунутий метод згладжування, який також можна використовувати для прогнозування - експоненціальне згладжування, також іноді зване методом Хольта-Уінтерса (Holt-Winters) на честь імен його творців [2, с. 58]. Існує наскільки варіантів даного методу:

- одинарне згладжування для рядів, у яких немає тренду і сезонності;
- подвійне згладжування для рядів, у яких є тренд, але немає сезонності;
- потрійне згладжування для рядів, у яких є і тренд, і сезонність.

Метод експоненціального згладжування обчислює значення згладженого ряду шляхом оновлення значень, розрахованих на попередньому кроці, використовуючи інформацію з поточного кроку. Інформація з попереднього і поточного кроків береться з різними вагами, якими можна курувати.

У найпростішому варіанті одинарного згладжування співвідношення таке:

$$s_i = \alpha x_i + (1 - \alpha)s_{i-1}, \text{ де } 0 \leq \alpha \leq 1$$

Параметр α визначає співвідношення між незгладжені значенням на поточному кроці і згладженим значенням з попереднього кроку. При $\alpha = 1$ ми будемо брати тільки точки вихідного ряду, тобто ніякого згладжування не буде. При $\alpha = 0$ ряд ми будемо брати тільки згладжені значення з попередніх кроків, тобто ряд перетвориться на константу. Однак, якщо в даних є тренд, просте згладжування буде «відставати» від нього (або доведеться брати значення α близькими до 1, але тоді згладжування буде недостатнім). Потрібно використовувати подвійне експоненціальне згладжування.

Подвійне згладжування використовує вже два рівняння - одне рівняння оцінює тренд як різницю між поточним і попереднім згладженим значеннями, потім згладжує тренд простим згладжуванням. Друге рівняння виконує згладжування як у випадку простого варіанту, але в другому доданку використовується сума попереднього згладженого значення і тренду.

Потрійне згладжування включає ще один компонент - сезонність, і використовує ще одне рівняння. При цьому розрізняються два варіанти сезонного компонента - адитивний і мультиплікативний. У першому випадку амплітуда сезонного компонента постійна і з часом не залежить від базової амплітуди ряду. У другому випадку амплітуда змінюється разом зі зміною базової амплітуди ряду. Це якраз наш випадок, як видно з графіка. З ростом ряду амплітуда сезонних коливань збільшується.

Так як наш перший ряд має і тренд, і сезонність, тому для нього підбираємо параметри потрійного згладжування. У Prognoz Platform це досить просто зробити, тому що при оновленні значення параметра платформа відразу ж перемальовує графік згладженого ряду, і візуально можна відразу побачити, наскільки добре він описує наш вихідний ряд. Prognoz Platform також дозволяє автоматично підібрати оптимальні значення параметрів, використовуючи систематичний пошук в просторі значень параметрів і мінімізуючи суму квадратів відхилень згладженого ряду від початкового. Описані методи дуже прості, їх легко застосовувати, і вони є чудовою відправною точкою для аналізу структури та прогнозування часових рядів.

Перелік використаних джерел

1. Афанасьев В.Н., Юзбашев М.М. Анализ временных рядов. – М.: Финансы и статистика, 2001.
2. Бриллинджер Д. Временные ряды. Обработка данных и теория. – М.: Мир, 1980.
3. Кендалл М.Дж., Стьюарт А. Многомерный статистический анализ и временные ряды. – М.: Наука, 1976.
4. Кендэл М. Временные ряды. – М.: Финансы и статистика, 1980.

5. Верещага В. М., Конопацький Є. В., Павленко О. М. Визначення площі, обмеженої топографічною замкненою плоскою кривою //Комп'ютерно-інтегровані технології: освіта, наука, виробництво. – 2015. – №. 20. – С. 119-123.

6. Павленко О. М. Геометричне моделювання вертикального планування горизонтальної земельної ділянки засобами точкового БН-числення: дис. – Мелітопольський державний педагогічний університет імені Богдана Хмельницького, 2017.

7. Верещага, В. М. Спосіб згортання (розгортання) чарунок [Текст] / В. М. Верещага, Є. О. Адоньєв, О. М. Павленко // Сучасні проблеми моделювання. – 2016. – №. 7. – С. 32–38.

8. Павленко О.М., Верещага В.М., Кучеренко В.В.Вертикальне планування на місцевості земельної ділянки до ідеально горизонтальної площини // Сучасні проблеми моделювання. – 2014. – №. 5. – С. 32–38.