

Анализ временных рядов и прогнозирование в Excel

Бокуть Л.В.

Белорусский национальный технический университет

Одной из наиболее быстро развивающихся областей математической статистики является анализ временных рядов.

Под временным рядом понимается последовательность наблюдений некоторой случайной величины X в последовательные равноотстоящие моменты времени. Отдельные наблюдения называются уровнями x_t , ($t = \overline{1, n}$) ряда, где n -число уровней. В общем случае временной ряд состоит из нескольких составляющих: $x_t = u_t + v_t + c_t + \varepsilon_t$ ($t = \overline{1, n}$), где u_t - тренд, то есть плавно изменяющаяся компонента, описывающая чистое влияние долговременных факторов; v_t - сезонная компонента, характеризующая повторяемость экономических процессов в течение кратковременных периодов; c_t - циклическая компонента, отражающая повторяемость экономических процессов в течение длительных периодов, ε_t - случайная компонента, характеризующая влияние случайных факторов.

К основным этапам анализа временных рядов относятся:

- графическое представление и описание поведения временного ряда;
- выделение и удаление закономерных составляющих временного ряда;
- сглаживание и фильтрация;
- исследование случайной составляющей временного ряда;
- прогнозирование изучаемого процесса на основе данного временного ряда.

Заметим, что отделить тренд и сезонность в общем случае невозможно, так как они взаимно проникают в друг друга.

Для сглаживания временного ряда применяют скользящую среднюю, которая содержит сведения о тенденциях изменения данных. Пусть имеется временной ряд x_1, x_2, \dots, x_n . К первым $2m+1$ точкам ряда подбирается полином

$$Q_r(t) = a_r t^r + a_{r-1} t^{r-1} + \dots + a_1 t + a_0.$$

Затем составляется сумма:

$$\sum_{t=-m}^m (x_t - a_r t^r - a_{r-1} t^{r-1} - \dots - a_1 t - a_0)^2,$$

которая минимизируется по методу наименьших квадратов. Затем подбирают полином порядка r для второго, третьего, ..., $2m+2$ наблюдения. Эта процедура продолжается вдоль всего ряда до последней группы из $2m+1$ точек. На самом деле нет необходимости подбирать полином каждый раз, так как эта процедура соответствует некоторой линейной комбинации наблюдений с постоянными коэффициентами.

Скольльзящие средние являются симметрическими (то есть коэффициенты веса симметричны относительно среднего). Для прогнозирования в Excel используют асимметричные фильтры. Скользящая средняя в Excel заменяет не средний, а последний уровень ряда в промежутке сглаживания и вычисляется по формуле:

$$F_{t+1} = \frac{1}{N} \sum_{h=0}^N x_{t-h+1},$$

где N - число предшествующих периодов, входящих в скользящую среднюю, x_h - значение в момент времени h , F_h - прогнозируемое значение в момент времени h . Итак, данный асимметричный фильтр позволяет рассчитать значения в прогнозируемом периоде на основе среднего значения переменной для указанного числа предшествующих периодов.

Асимметричные скользящие средние могут учитывать устаревание данных, то есть каждое новое наблюдение, будет иметь вес больше предыдущих:

$$\begin{aligned} F_{t+1} &= (1-\alpha)(A_{t+1} + \alpha A_t + \alpha^2 A_{t-1} + \dots) = \\ &= (1-\alpha) \sum_{i=0}^{\infty} \alpha^i A_{t-i+1}, \quad 0 < \alpha < 1, \end{aligned}$$

где A_t - фактическое значение в момент времени t .

Подход, основанный на использовании асимметричных скользящих средних, называется экспоненциальным сглаживанием. Существуют и другие методы сглаживания и экстраполяции

ции: модели Хольта- Уинтерса, Харрисона, Бокса- Дженкинса, фильтры Калмана и Бюсси.

Excel содержит ряд родственных статистических функций для аналитического выравнивания и прогноза: ТЕНДЕНЦИЯ, ЛИНЕЙН, ЛГРФПРИБЛ, ПРЕДСКАЗ.

Функция ТЕНДЕНЦИЯ аппроксимирует прямой линией по методу наименьших квадратов массивы известные_значения_y и известные_значения_x. Возвращает значения y для заданного массива новые_значения_x в соответствии с этой прямой. Синтаксис:

ТЕНДЕНЦИЯ(известные_значения_y;известные_значения_x;новые_значения_x;конст).

Известные_значения_y- множество значений y, которые уже известны для соотношения $y=mx+b$.

Известные_значения_x- необязательное множество значений x, которые уже известны.

Новые_значения_x- это новые значения x, для которых ТЕНДЕНЦИЯ возвращает соответствующие значения y.

Конст- логическое значение, которое указывает, требуется ли, чтобы константа b была равна 0 (при конст=ЛОЖЬ). Значения m в этом случае подбираются таким образом, чтобы выполнялось соотношение $y=mx$. Функция ЛИНЕЙН позволяет получить коэффициенты уравнения регрессии с помощью метода наименьших квадратов, которые можно использовать для выравнивания и прогноза. Функция ЛГРФПРИБЛ может использоваться аналогично ЛИНЕЙН. Функция ПРЕДСКАЗ вычисляет или предсказывает будущее значение по существующим значениям.

Кроме того, осуществлять прогнозирование в Excel можно с помощью диаграмм.

Прогнозирование процессов на основе временных рядов относится к современным технологиям прогнозирования деятельности предприятия. Оно позволяет, в частности автоматизировать процесс среднесрочного и долгосрочного планирования, моделировать управленческие решения, связанные с финансовой деятельностью предприятия.