Методы прогнозирования

Анализ временных рядов

Последовательность наблюдений одного показателя, упорядоченная в зависимости от последовательно возрастающих или убывающих значений другого показателя, называется ***динамическим рядом***, или ***рядом динамики***.

Если в качестве признака, в зависимости от которого происходит упорядочивание, берется время, то такой динамический ряд называется ***временным рядом***.

Временной ряд *(экономика)* - последовательность наблюдений некоторого признака (случайной величины) в последовательные моменты времени.

Отдельные наблюдения - ***уровни ряда***, y_t $(t = 1, 2, 3, \ldots, n)$, n число уровней.

Моментные временные ряды содержат значения социально-экономического показателя, относящиеся к определенным моментам времени.

Длина временного ряда - время, прошедшее от начального момента наблюдения до конечного, или число уровней ряда.

Каждый уровень временного ряда формируется под воздействием большого числа факторов:

- факторы, формирующие тенденцию ряда;
- факторы, формирующие сезонные колебания ряда;
- факторы, формирующие циклические колебания ряда;
- случайные факторы.

Тренд- когда во временном ряду проявляется длительная закономерность изменения уровней. Соответствующая математическая модель называется **трендовой моделью**.

Аддитивная модель временного ряда **-**модель, в которой временной ряд представлен как сумма перечисленных компонент.

Общий вид аддитивной модели:

$$y_t = u_t + v_t + c_t + \varepsilon_t \quad (t = 1, 2, \ldots, n),$$

где u_t - ***тренд,*** плавно меняющаяся компонента, описывающая чистое влияние долговременных факторов, то есть длительную тенденцию изменения признака (например, рост населения, экономическое развитие, изменение структуры потребления);

v_t - ***сезонная компонента***, отражающаяся повторяемость экономических процессов в течение не очень длительного периода (года, иногда месяца, например, объем продаж товаров или перевозок пассажиров в различные времена года, уровень безработицы в курортных городах в зимний период выше по сравнению с летним);

c_t - ***циклическая компонента***, отражающаяся повторяемость экономических процессов в течение длительных периодов (например, влияние волн экономической активности Кондратьева, демографических ям, циклов солнечной активности);

ε_t - ***случайная компонента***, отражающая влияние не поддающихся учету и регистрации случайных факторов.

Этапы анализа временных рядов:

- графическое представление и описание поведения временного ряда;
- сглаживание и фильтрация (удаление низко- или высокочастотных составляющих временного ряда);
- выделение и удаление закономерных (неслучайных) составляющих временного ряда (тренда, сезонных и циклических составляющих);
- исследование случайной составляющей временного ряда, построение и проверка адекватности математической модели ее описания;
- прогнозирование развития изучаемого процесса на основе имеющегося временного ряда;
- исследование взаимосвязи между различными временными рядами.

Предварительная обработка временных рядов состоит в выявлении аномальных значений ряда и сглаживании ряда.

Ошибки 1 рода	Ошибки 2 рода
Ошибки при сборе, обработке и передаче информации. Их можно устранить, принять меры к их допущению.	Ошибки, возникающие из-за воздействия факторов, имеющих объективный характер, но действующих эпизодически. Их невозможно устранить, но можно исключить из рассмотрения, заменив значение на среднеарифметическое соседних уровней.

Критерий Ирвина

Аномальной считается точка y_t, отстоящая от предыдущей точки y_{t-1} на величину, большую среднеквадратичного отклонения

$\lambda_t = \dfrac{|y_t - y_{t-1}|}{\sigma}$, где λ_t − критерий Ирвина, а σ−среднеквадратичное отклонение

$$\sigma = \sqrt{\frac{\sum_{t=1}^{n}(y_t - \bar{y})^2}{n-1}}.$$

Точка считается аномальной, если $\lambda_t > \lambda_{таб}$.

Табличные значения $\lambda_{таб}$ уменьшаются с ростом длины ряда, их значения приведены в следующей таблице

n	10	20	30	50	100
$\lambda_{таб}$	1,5	1,3	1,2	1,1	1,0

Метод скользящей средней

Скользящие средние:

$$\tilde{y}_t = \frac{\sum_{i=t-p}^{t+p} y_i}{m},$$ **сглаживание по 3 точкам**:

когда $m=2p+1$ - нечетное число. При $m=3 \quad p=1$.

Сглаживание по 5 точкам:

$$\tilde{y}_t = \frac{y_{t-2} + y_{t-1} + y_t + y_{t+1} + y_{t+2}}{5}$$

и взвешенная (средневзвешенная) скользящая средняя

$$\tilde{y}_t = \frac{\sum_{i=t-p}^{t+p} \rho_i y_i}{\sum_{i=t-p}^{t+p} \rho_i}.$$

Для **сглаживания по 5* точкам** используют весовые коэффициенты (-3, 12, 17, 12, -3).

При сглаживании временного ряда по *2p+1* соседним точкам *p* в начале и в конце ряда остаются не сглаженными. Эти точки либо исключают из рассмотрения, либо использовать для них специальные формулы сглаживания для крайних точек. Например, для сглаживания по трем точкам

$$\tilde{y}_1 = \frac{5y_1 + 2y_2 - y_3}{6}, \quad \tilde{y}_n = \frac{5y_n + 2y_{n-1} - y_{n-2}}{6}.$$

Аналитическое выравнивание временного ряда

Этапы аналитического выравнивания временного ряда:

- с помощью графика или методом перебора определяется вид зависимости, которой подчинено развитие явления;
- через систему уравнений из метода наименьших квадратов рассчитываются параметры уравнения;
- рассчитываются значения временной функции, которые определяют тенденцию развития;
- при необходимости проводится экстраполяция или интерполяция тенденции развития.

Метод проверки разностей средних уровней

(наличие тренда)

Исходный ряд из n точек делится на два с примерно одинаковым числом точек n_1 и n_2 ($n = n_1 + n_2$). Для каждой из частей вычисляются средние значения и дисперсия.

$$\tilde{y}_1 = \frac{\sum_{i=1}^{n_1} y_i}{n_1}, \quad \sigma_1^2 = \frac{\sum_{i=1}^{n_1} (y_i - \tilde{y}_1)^2}{n_1 - 1}$$

$$\tilde{y}_2 = \frac{\sum_{i=1}^{n_2} y_i}{n_2}, \quad \sigma_2^2 = \frac{\sum_{i=1}^{n_2} (y_i - \tilde{y}_2)^2}{n_2 - 1}$$

Гипотеза об однородности дисперсий частей ряда проверяется с помощью критерия Фишера:

$$F_{наб} = \begin{cases} \sigma_1^2 / \sigma_2^2, если\ \sigma_1^2 > \sigma_2^2; \\ \sigma_2^2 / \sigma_1^2, если\ \sigma_2^2 > \sigma_1^2. \end{cases}$$

$F_{таб} = F(1 - \alpha / 2, n_1 - 1, n_2 - 1)$ – квантиль распределения Фишера

Если $F_{наб} < F_{таб}$, то гипотеза об однородности дисперсии принимается и переходим к следующему этапу проверки, в противном случае гипотеза об однородности дисперсии отклоняется.

Окончательная проверка гипотезы об отсутствии тренда проводится с использованием критерия Стьюдента: $t_{наб} = \dfrac{|\bar{y}_1 - \bar{y}_2|}{\sigma\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$, где $\sigma = \sqrt{\dfrac{(n_1 - 1)\sigma_1^2 + (n_2 - 1)\sigma_2^2}{n_1 + n_2 - 2}}$.

$t_{таб} = t(1 - \alpha / 2, n - 2)$ –квантиль распределения Стьюдента, где α –уровень значимости (обычно α=0,05).

Если $t_{наб} < t_{таб}$, то гипотеза принимается и тренда нет, в противном случае тренд есть.

Линейная модель тренда

ряд динамики имеет постоянные или почти постоянные ежегодные (ежеквартальные, ежемесячные и т.п.) абсолютные приросты

$$\hat{y}_t = a_0 + a_1 t$$

где $\hat{y}_i$ - уровень тренда для периода или момента с номером t_i;

a - свободный член уравнения, равный среднему уровню тренда для периода (момента) с нулевым номером t_i;

b - главный параметр линейного тренда – его константа – среднее абсолютное изменение за принятую в ряду единицу времени.

Величина параметров a и b определяется по методу наименьших квадратов путем приравнивания частных первых производных функции

$$f(a,b) = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (y_i - a - bt_i)^2$$ к нулю.

После алгебраических преобразований получаем два «нормальных уравнения» метода наименьших квадратов (МНК) для прямой:

$$na + b\sum_{i=1}^{n} t_i = \sum_{i=1}^{n} y_i$$

$$a\sum_{i=1}^{n} t_i + b\sum_{i=1}^{n} t_i^2 = \sum_{i=1}^{n} (y_i t_i).$$

Показательная модель тренда

ряд динамики описывает процесс развития объекта и имеет ежегодно возрастающий абсолютный прирост

$$\hat{y}_t = a_0(1+a_1)^t$$

Экспоненциальный тренд $\hat{y}_i = a \cdot k^{t_i}$.

Для нахождения параметров a и k уравнение логарифмируем:

$$\ln \hat{y}_i = \ln a + t_i \ln k .$$

Итоговый вид систему уравнений:

$$n \ln a + \ln k \sum_{i=1}^{n} t_i = \sum_{i=1}^{n} \ln y_i$$

$$\ln a \sum_{i=1}^{n} t_i + \ln k \sum_{i=1}^{n} t_i^2 = \sum_{i=1}^{n} t_i \ln y_i$$

Особенность этого типа тренда - логарифмировать необходимо номера периодов (моментов) времени: $\hat{y} = a + b \ln t$. Следовательно, все номера должны быть положительными числами.

Полиномиальный (параболический) тренд

$$\hat{y}_i = a + bt_i + ct_i^2$$

Для вычисления параметров a, b, c по методу наименьших квадратов

три частные производные функции: $f(a,b,c) = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$

приравниваются к нулю, и после преобразований получаем систему трех уравнений с тремя неизвестными:

$$na + b\sum_{i=1}^{n} t_i + c\sum_{i=1}^{n} t_i^2 = \sum_{i=1}^{n} y_i$$

$$a\sum_{i=1}^{n} t_i + b\sum_{i=1}^{n} t_i^2 + c\sum_{i=1}^{n} t_i^3 = \sum_{i=1}^{n} y_i t_i$$

$$a\sum_{i=1}^{n} t_i^2 + b\sum_{i=1}^{n} t_i^3 + c\sum_{i=1}^{n} t_i^4 = \sum_{i=1}^{n} y_i t_i^2$$

Гиперболический тренд

$$\hat{y}_i = a + \frac{b}{t_i}$$

Соответственно нормальные уравнения метода наименьших квадратов получат вид:

$$na + b\sum_{i=1}^{n}\frac{1}{t_i} = \sum_{i=1}^{n} y_i$$

$$a\sum_{i=1}^{n}\frac{1}{t_i} + b\sum_{i=1}^{n}\frac{1}{t_i^2} = \sum_{i=1}^{n}\frac{y_i}{t_i}.$$

Сезонные *колебания- колебания, которые связанны со сменой времен года и поэтому ежегодно повторяются.*

Под сезонными колебаниями понимаются устойчивые изменения какого-либо явления внутри отдельно взятого периода. Н/р, объемы производства и потребления. Для изучения сезонных колебаний используется метод одногодичных (когда нет резких перепадов в уровнях ряда) или многолетних средних.

Расчет индексов (коэффициентов) сезонности

для каждого месяца (квартала) внутри года

$$J_t = \frac{y_t}{\overline{y}} 100\%$$

(темп роста по отношению к у, за базу взято среднее значение),

где $\overline{y} = \frac{1}{n}\sum y_t$, n - число временных интервалов в году

(если $n = 12$, то рассматриваются месячные колебания, при $n = 4$ -квартальные).

Индексы сезонности обычно выражаются в процентах.

Виды сезонных колебаний

По форме

- ***Выпуклые*** – явление возрастает к середине года и снижается к началу или концу (строительство, добыча полезных ископаемых открытым способом, торговля и т.д.).
- ***Вогнутые*** – явление снижается к середине года и возрастает к концу (торговля товарами зимнего ассортимента, мясная промышленность, сахарная, электроника, потребление тепла и т.д.).

По промежутку времени

- колебание внутри года, месяца, квартала, рабочего дня.

Ряд Фурье

$$y_t = a_0 + \sum_{k=1}^{m} (a_k \cos kt + b_k \sin kt),$$

где: y_t – выравненные уровни ряда;

a_0, a_k, b_k - параметры уравнения;

cos t, sin t – тригонометрические функции, соответствующие каждому месяцу года;

k – номер гармоники; m – взятое количество гармоник.

Параметры уравнения ряда Фурье рассчитывается по формулам:

$$a_0 = \frac{\sum_{t=1}^{n} y_t}{n}; a_1 = \frac{\sum_{t=1}^{n} y_t \cos t}{6}; a_2 = \frac{\sum_{t=1}^{n} y_t \cos 2t}{6};$$

$$b_1 = \frac{\sum_{t=1}^{n} y_t sit}{6}; b_2 = \frac{\sum_{t=1}^{n} y_t \sin 2t}{6}.$$

При исследовании ***случайной компоненты*** ε_i на нормальность

используют R/S критерий.

$$R/S = \frac{\max_i \varepsilon_i - \min_i \varepsilon_i}{S_e}, \quad \text{где } S_e = \sqrt{\frac{\sum_{i=1}^{n_1} \varepsilon_i^{\,2}}{n-1}} - \text{стандартная ошибка.}$$

Значения критерия сравниваются с табличными нижней
и верхней границами отношения. Если вычисленное значение критерия
попадает в табличный интервал, то гипотеза о нормальности
распределения принимается, в противном случае гипотеза отвергается.
Критические значения границ критерия приведены в
таблице 1 приложения (УП).

Построенная модель временного ряда используется
для прогнозирования будущих значений **(экстраполяция)**
либо для восстановления отсутствующих в
исходных данных внутренних значений **(интерполяция)**.

Качественные методы прогнозирования (метод экспертных оценок)

Проведение экспертами интуитивно-логического анализа проблемы с количественной оценкой суждений и формальной обработкой результатов. Метод экспертных оценок используется для решения различных сложных неформализуемых проблем.

Первый класс

- Проблемы, в отношении которых имеется достаточный информационный потенциал

Второй класс

- Проблемы, в отношении которых информационный потенциал знаний недостаточен

Типовые задачи

- составление перечня возможных событий в различных областях за определенный промежуток времени;
- определение наиболее вероятных интервалов времени наступления совокупности событий;
- определение целей и задач управления, упорядочение их по степени важности;
- определение альтернативных вариантов решения задачи, оценка их предпочтения;
- альтернативное распределение ресурсов для решения задач, оценка их предпочтительности;
- выбор альтернативных вариантов принятия решений в определенной ситуации, оценка их предпочтительности.

Основные этапы проведения экспертизы

формулирование цели экспертизы;

формирование группы специалистов-аналитиков;

отбор и формирование группы экспертов;

проведение опроса;

анализ и обработка информации экспертов; *синтез объективной (статистической) информации и информации, полученной в результате экспертизы, с целью приведения их в форму, удобную для принятия решения.*

Коэффициент ***конкордации*** - числовой критерий согласованности мнений экспертов в рассматриваемой группе.

$$W = \frac{12\sum_{i=1}^{n} S_i^2}{m^2(n^3 - n)},$$ где n — число объектов; m — число экспертов; S_i — отклонение суммы рангов i-го объекта от средней их суммы для всех объектов.

Если оценки всех экспертов совпадают, то $W = 1$.

Если же оценки экспертов полностью не совпадают, то коэффициент $W < 1$.

Наименьшее возможное его значение равно нулю.

Если $W < 0{,}4$ - слабая согласованность экспертов,

Большие величины $W > 0{,}7$ свидетельствуют о сильной согласованности экспертов.

Причины слабой согласованности:

— в рассматриваемой группе экспертов действительно отсутствует общность мнений;

— внутри группы существуют коалиции с высокой согласованностью мнений, однако обобщенные мнения коалиций противоположны.

Метод средних арифметических рангов.

- Подсчитывается сумма рангов, затем эту сумму делим на число экспертов, получаем средний арифметический ранг. По средним рангам строится итоговая ранжировка по принципу: чем меньше средний ранг, чем лучше проект.

Метод медиан.

- Все ранги по каждой оцениваемой позиции располагаются в порядке неубывания. Вычисляется медиана рангов. Итоговую совокупность медиан снова ранжируем. Лучшей считается позиция с наименьшим итоговым рангом.

Сравнение ранжировок по методу средних арифметических и методу медиан.