2 СЛАЙД

Временной ряд - собранный в разные моменты времени статистический материал о значении каких-либо параметров исследуемого процесса.

Временной ряд существенно отличается от простой выборки данных, так как при анализе учитывается взаимосвязь измерений со временем, а не только статистическое разнообразие и статистические характеристики выборки.

3 СЛАЙД

Задачи, возникаемые при работе с временными рядами делятся на 2 больших класса.

Это анализ - когда мы пытаемся по имеющимся данным получить как можно больше информации.

И прогнозирование - когда по имеющимся данным делается прогноз значений временного ряда на какие-то будущие периоды.

Анализ рядов состоит из основных задач: выявление трендов данных - т.е. определение, есть ли какая-то тенденция во временном ряде, выявление сезонных компонент и определение, есть ли в данных какие-то аномалии или выбросы.

Также бывает полезно разложить временной ряд на составляющие, например на тренд, на сезонность и на некоторую ошибку.

Задачи прогнозирования временных рядов осуществляются с помощью различных моделей. Например, самые простые - это простая авторегрессионная модель, скользящее среднее, и их обобщение — это ARMA и ARIMA модели.

Также существует большой класс адаптивных моделей прогнозирования.

4 СЛАЙД

Временные ряды бывают стационарные и нестационарные.

Стационарный ряд подразумевает, что в данных у нас нет ярко выраженного тренда, то есть мы можем сказать, что мат. ожидание наших случайных величин, которые мы собираем в виде ряда, равно некоторой константе.

По нестационарному ряду можно сказать, что у него в данных есть какая-то тенденция к развитию. Например, это может быть возрастающий тренд или наоборот убывающий, либо какие-то цикличные явления.

5 СЛАЙД

Также ряды могут содержать сезонные компоненты. Сезонная компонента — это некоторое периодическое явление, которое возникает в наших данных, например, какие-то периодические подъемы или, наоборот, падения в данных. Не следует забывать, что ряды могут содержать как тренды, так и сезонности.

6 СЛАЙД

Существует несколько подходов к анализу структуры временных рядов, содержащих сезонные или циклические колебания.

Простейший подход - расчет значений сезонной компоненты методом скользящей средней и построение аддитивной или мультипликативной модели временного ряда.

Аддитивная модель предполагает, что каждый уровень временного ряда может быть представлен как сумма трендовой, сезонной и случайной компонент.

Мультипликативная модель предполагает произведение трендовой, сезонной и случайной компонент.

Выбор одной из двух моделей осуществляется на основе анализа структуры сезонных колебаний.

Если амплитуда колебаний приблизительно постоянна, строят аддитивную модель временного ряда, в которой значения сезонной компоненты предполагаются постоянными для различных циклов.
Если амплитуда сезонных колебаний возрастает или уменьшается, строят мультипликативную модель временного ряда, которая ставит уровни ряда в зависимость от значений сезонной компоненты.

Построение аддитивной и мультипликативной моделей сводится к расчету значений трендовой, циклической и случайной компонент для каждого уровня ряда.

Процесс построения модели включает в себя следующие шаги:

1. Выравнивание исходного ряда методом скользящей средней.

2. Расчет значений сезонной компоненты.

3. Устранение сезонной компоненты из исходных уровней ряда и получение выровненных данных в аддитивной или мультипликативной модели.

4. Аналитическое выравнивание уровней и расчет значений тренда с использованием полученного уравнения тренда.

5. Расчет полученных по модели значений или Расчет абсолютных и относительных ошибок.

На практике отличить аддитивную модель от мультипликативной можно по величине сезонной вариации. Аддитивной модели присуща практически постоянная сезонная вариация, тогда как у мультипликативной она возрастает или убывает, графически это выражается в изменении амплитуды колебания сезонного фактора, как это показано на рисунке слайда.

7 СЛАЙД - **МЕТОД STL**

STL расшифровывается как A Seasonal-Trend Decomposition Procedure Based on Loess, это процедура декомпозиции временного ряда на сезонную, трендовую составляющую и на остатки, использующая метод локальных регрессий (LOESS), с помощью которой происходит сглаживание исходного ряда данных.

Метод применяется для аддитивных моделей.

8 СЛАЙД

LOESS методика была предложена Кливлендом в 1979 году для моделирования и сглаживания двумерных данных. Эта техника представляет общий подход для приближения двумерных данных (т.е данных с параметрами времени и различных значений).

9 СЛАЙД

STL позволяет получить из первого ряда на слайде разложение на три ряда. Здесь можно увидеть сезонную составляющую (- также видно, что она довольно постоянна), трендовую и ошибки. То есть их можно анализировать и на основе этого делать выводы.

10 СЛАЙД

STL это делает с помощью двух основных циклов: внутренний и внешний. Во внутреннем, на первом шаге мы задаем какие-то начальные приближения весов которые будут использоваться в LOESS-процедуре, и какое-то начальное приближение тренда, которое обычно нулевое.

Затем вычитаем трендовую составляющую исходного ряда. Потом каждый сезонный компонент сортируем по периодам, после чего сглаживаем LOESS.

Это сглаживание можно привести на таком примере: представим, что у нас есть недельная сезонность: понедельник, вторник, среда и так далее. Мы берем и последовательно сглаживаем значения каждого дня недели и получаем какой-то новый ряд. Это второй шаг.

11 СЛАЙД

На третьем шаге уже идёт глубокое сглаживание сезонной компоненты, и здесь мы сглаживаем всё подряд с помощью различных методов - например скользящим средним, и получаем еще один ряд.

На четвертом шаге идёт детрендирование сглаженных сезонных компонент получением нового ряда, т.е. из второго ряда, полученного на втором шаге вычитаем ряд, полученный на третьем шаге. Получаем опять новый ряд.

Пятым шагом идет десезонализация, то есть мы из исходного ряда вычитаем ряд, полученный на четвёртом шаге.

Последний во внутреннем цикле шестой шаг — это сглаживание тренда. То есть теперь получаем тренд как LOESS, то есть применением LOESS к ряду, полученному на шаге 5.

Весь внутренний цикл делаем 2-3 раза в зависимости от задачи.

12 СЛАЙД

На внешнем цикле считаются лишь остатки по формуле (на слайде).

Затем пересчитываются веса P, которые используются в LOESS по новой.

Весь внешний цикл рассчитывается раз 10–15, также в зависимости от данных. Т.е. если в данных много выбросов, лучше запустить больше циклов, если мало, может хватить и одного.

13 СЛАЙД **STL разложение с помощью R**

В R прогнозирование и анализ временных рядов можно обрабатывать с помощью обычных функций из пакета stats, например glm() или большого количества специализированных пакетов.

Также существует специальный класс объектов для работы с данными, представляющими собой временные ряды - ts (от time series - временной ряд). Для создания объектов этого класса служит одноименная функция - ts().

В качестве примера возьмём данные, представляющие собой количество кликов на сайте в определенные моменты времени из CSV файла. И запишем их в переменную data.

Посмотрим, как выглядит наш ряд. Воспользуемся функцией summary.

Воспользуемся функцией plot(ts), отобразив лишь клики и получим график.

Можно сказать, что здесь есть тренд, а также сезонность на фоне случайной составляющей. И можно заметить, что увеличение дисперсии в сезонности не наблюдается, восхождение тренда происходит линейно. Поэтому у нас есть все основания предполагать, что у нас аддитивная сезонность и мы можем применить STL процедуру.

14 СЛАЙД

Поместим во временной ряд результаты STL разложения исходного ряда, применив процедуру STL и указав только второй столбец. Также стоит заметить, что robust=TRUE, означает оценку с помощью робастных методов. Это полезно, если нет уверенности в своих данных. Указываем 2 итерации внутреннего цикла. И выводим разложение на график.

15 СЛАЙД

Здесь видно четыре графика. Первый —исходные данные. Второй — выделенная из них сезонная составляющая. Третий —трендовая составляющая. И четвертый — это остатки.

Видно, что сезонность хорошо выделилась и она очень периодичная. Тренд восходящий.

Остатки около нуля, чистые, симметричные и без особых выбросов. Следовательно мы можем сказать, что STL разложение прошло успешно.

16 СЛАЙД **AR-Модель**

AR-модель, т.е. авторегрессионная модель - это модель временных рядов, в которой значения временного ряда в данный момент линейно зависят от предыдущих значений этого же ряда. Это простая модель, с помощью которой можно также прогнозировать, если мы сможем правильно её определить.

Обозначается как AR(p),

где p — это порядок регрессии, которая представляется в виде, где a1, ..., ap — это параметры модели (- коэффициенты авторегрессии),

c — это некоторая постоянная ( -часто для упрощения применяют равное 0),

и εt — это некоторый белый шум ( -т.е. это такой процесс, матожидание которого равно 0 и он симметричный)

17 СЛАЙД **Построение авторегрессионной модели прогнозирования временных рядов с помощью R.**

Преобразуем наши данные во временной ряд.

Так как мы будем прогнозировать, то соответственно сразу отрежем хвост в 30 последних точек и запишем в переменную ts. Это будут данные, по которым мы будем строить наши прогнозные модели

Также нам необходимы данные, на которых мы будем сравнивать прогноз. Это будут 30 последних наблюдений. Запишем в переменную newts и выведем на графике.

Данные содержат явный недельный тренд и ярко выраженную недельную сезонность.

18 СЛАЙД

Попробуем построить по ним простейшую прогнозную модель — модель скользящего среднего. Переменная order.max максимальный порядок построения модели. Единица означает авторегрессию первого порядка.

Получаем коэффициент для авторегрессии.

Построим прогноз по этой модели на 30 последних точках и получим результат.

Далее так как мы видели, что в данных присутствует явная недельная сезонность, то логично построить авторегрессию более высокого порядка.

19 СЛАЙД

Построим до порядка 10. Здесь модель уже сама подбирает оптимальный порядок.

В результате произошёл подбор авторегрессии с размерностью 7, что логично с учетом того, что в данных недельная сезонность.

20 СЛАЙД

Далее попробуем построить прогноз для 30 последних точек и увидим, как выглядит прогнозы по полученным моделям.

Потом сравним их с реальными данными, которые были на 30 последних точках

В итоге получился прогноз по этим двум моделям — это регрессия первого порядка и более высокого порядка до 7-го.

Так как в данных есть явная недельная сезонность, то модель более высокого порядка авторегрессии даёт более точный результат, чем авторегрессия первого порядка.