**Структура модуля:**

Оглавление

[Юнит 1. Введение 2](#_Toc148629444)

[Юнит 2. Временные ряды. Понятие 4](#_Toc148629445)

[Юнит 3. Анализ временных рядов 9](#_Toc148629446)

[Юнит 4. Модели временных рядов. Тренд, сезонность, шум 13](#_Toc148629449)

[Юнит 5. Модели тренда временных рядов 16](#_Toc148629450)

[Юнит 6. Статистический подход к анализу временных рядов. Основные статистические характеристики 24](#_Toc148629452)

[Юнит 7. Анализ стационарности/нестационарности 27](#_Toc148629453)

[Юнит 8. **Преобразования нестационарных временных рядов.** 34](#_Toc148629454)

[Юнит 9. Итоги Полу-модуля 38](#_Toc148629456)

# Юнит 1. Введение

Представьте, что вы устроились дата саентистом в транспортную компанию. У вас есть данные об объеме грузоперевозок на определенном транспортном маршруте в течении нескольких лет, и ваш начальник просит вас посмотреть, как вы можете улучшить работу компании.

Чтобы решить эту задачу, вам понадобится знание временных рядов.

**Временной ряд (ВР)** — это измерения одной и той же случайной величины в разные моменты времени.

Анализируя временные ряды, вы сможете выявить сезонные колебания спроса на перевозки (например, увеличение объемов грузоперевозок летом из-за туристического сезона), а также предсказать будущий спрос на грузоперевозки. Это поможет вашей компании оптимизировать свои маршруты, планировать количество и типы транспортных средств, сократить издержки и повысить эффективность.

В бизнесе временные ряды ценятся за возможность прогнозировать значение целевой переменной на будущее. Самые очевидные примеры использования временных рядов в бизнесе — финансовое моделирование и прогнозирование спроса или выручки.

Модуль даст вам представление о ВР и базовых понятиях, необходимых для его осмысления.

**В этом модуле вы изучите следующие вопросы:**

* базовые методы предварительного анализа ВР;
* типы решаемых задач анализа ВР;
* модели ВР;
* основные статистические свойства ВР;
* особые свойства ВР, например: АКФ, гетероскедастичность, белый гауссов шум;
* особенности стационарного вида ВР;
* виды простых предсказаний ВР, в т.ч. наивные методы предсказаний и методы на основе скользящего среднего и экспоненциального сглаживания;
* особенности использования методов линейной регрессии в анализе ВР;
* особенности использования методов нелинейной регрессии в анализе ВР;
* особенности базовых методов разложения ВР.

Также вы познакомитесь с **методами работы с ВР в рамках языка программирования Python**, включая особенности:

* визуализации ВР,
* определения базовых свойств ВР,
* моделирования ВР,
* работы с фреймворком SKTime,
* выбора методов предсказания ВР.

**В результате прохождения модуля вы будете:**

* знать основные понятия анализ временных рядов, основные свойства ВР;
* уметь предложить подход для решения той или иной задачи анализа временных рядов;
* владеть методами предварительного анализа временных рядов.

Умение работать с временными рядами является отличным навыком, который откроет перед вами множество дверей: финансовая отрасль, маркетинг и продажи, анализ производственных показателей и другие сферы.

# Юнит 2. Временные ряды. Понятие

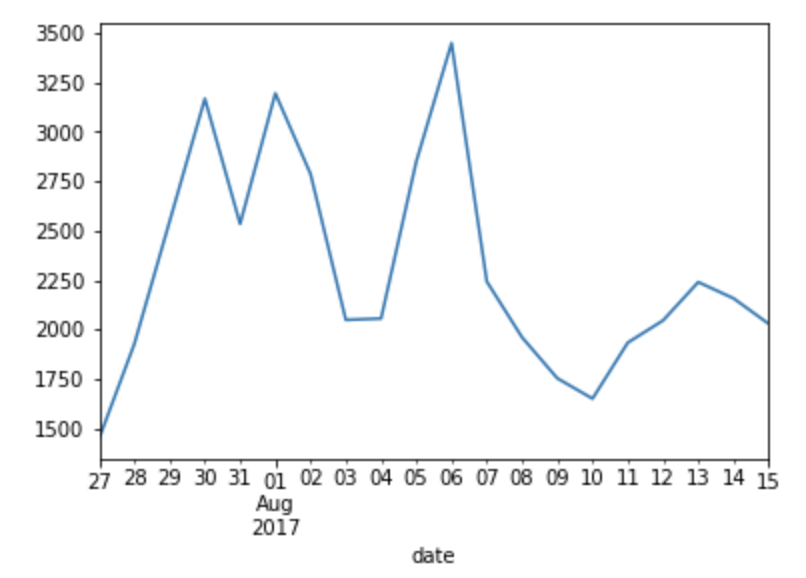
В этом юните мы подробнее рассмотрим основополагающее понятие данного модуля — временной ряд, а также поговорим о его видах и свойствах.

Временные ряды — это набор оцифрованных проиндексированных значений, в которых каждый отсчет связан с другим некоторой зависимостью.

В примере с транспортной компанией в качестве временного ряда может выступать объем грузоперевозок (одна и та же величина) в течении нескольких лет (в разные моменты времени).

Часто временной ряд — это набор наблюдений (выборок), проиндексированных по времени — но также может использоваться другая индексированная переменная.

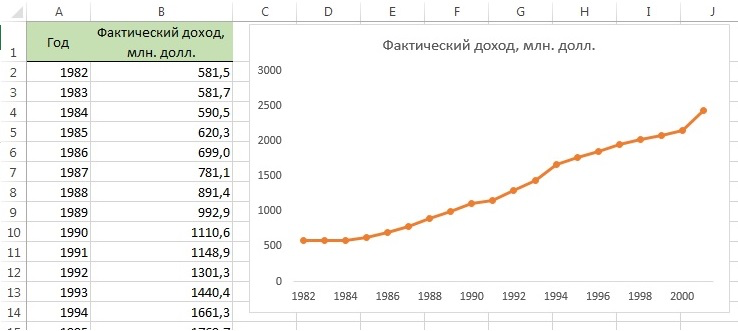
На графике ниже визуализирован временной ряд с получением данных за равные промежутки времени (дни).

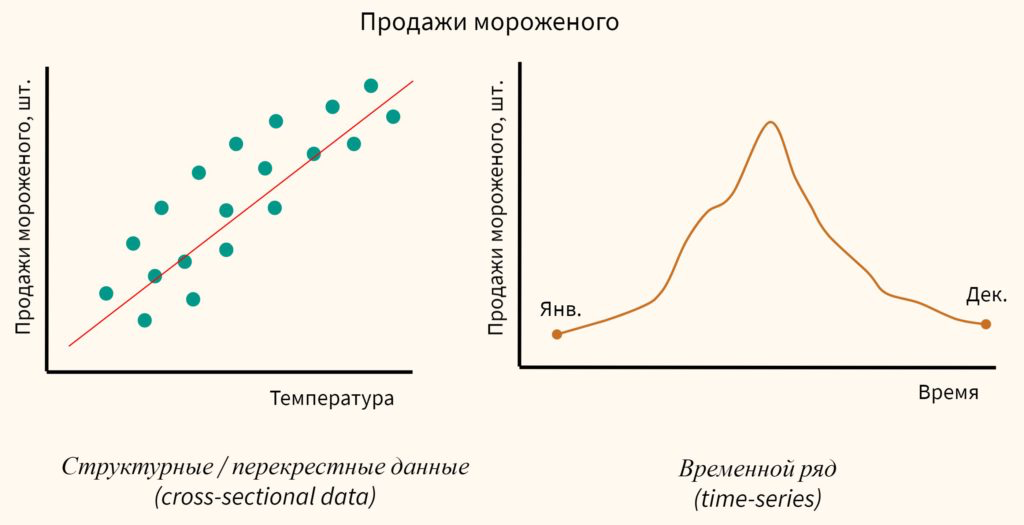


Примеры временных рядов:

* всевозможные экономические индексы (RTS, DJIA и др.);
* стоимость активов на бирже (курсы нефти, ценных бумаг, валют и криптовалют);
* физические показатели спортсмена во время тренировок или специальных нагрузочных тестов;
* средняя дневная скорость интернета;
* регулярные измерения роста, веса, давления;
* температура воды в Красном море по дням в июле.

Теперь выделим несколько формальных свойств временного ряда:

* 1. Данные временного ряда **структурированы**, а атрибуты (так иногда называют признаки) **зависимы от времени**.  
     На графике ниже визуализирован временной ряд фактического дохода компании. Данные структурированы (упорядочены и находятся в таблице), а фактический доход — это атрибут, зависящий от времени.  
     [Источник изображения](https://baguzin.ru/wp/analiz-vremennyh-ryadov/)

2. Данные временного ряда, в отличие от любых других данных, имеют **определенную последовательность**.  
На графиках ниже отображены продажи мороженого с января по декабрь. В структурных данных мы видим зависимость продаж от температуры (выше температура — больше продаж) в рамках временного ряда, а также — динамику количества продаж в рамках заданного временного диапазона: начиная с января и заканчивая декабрем этого года. Как мы видим, летом график продаж возрастает, что соответствует нашему графику слева (зависимость от температуры).  
[Источник изображения](https://www.dmitrymakarov.ru/intro/time-series-20/)

**3. Временные ряды могут быть дискретными и непрерывными (Discrete and Continuous):**

В общем случае временная переменная дискретна и задана с равномерными временными шагами, и в этом случае мы имеем **дискретный временной ряд.**

Примером дискретного временного ряда может быть ряд ежедневных объемов продаж в магазине за последний месяц. Например, можно записать ряд как:

1 день: 100 продаж

2 день: 85 продаж

3 день: 92 продажи

...

30 день: 120 продаж

В данном примере каждое наблюдение представлено в виде дискретной точки, соответствующей определенному дню. Такой ряд может быть использован для анализа динамики продаж в магазине и выявления тенденций или сезонности в объеме продаж.

Однако есть и **непрерывные временные ряды.**

Примером непрерывного временного ряда может быть график изменения акционной цены на фондовой бирже в течение дня. Данные о цене акций обновляются мгновенно и непрерывно в течение рабочего дня. Изменения цены могут быть непрерывными, и график может показывать постоянные флуктуации в течение времени.

# Юнит 3. Анализ временных рядов

В прошлом юните вы познакомились с видами и свойствами временных рядов, теперь настала пора перейти к их анализу.

Анализ представляет собой широкий спектр методов и подходов для изучения временной динамики данных. Правильный анализ может помочь получить полезные инсайты и прогнозы для бизнеса.

**Анализ временных рядов** — это один или последовательный набор методов/алгоритмов для получения (или оценки/прогнозирования) некоторых параметров (или характеристик) ряда.

Такими характеристиками могут быть дальнейшие (предыдущие) значения выборки, некоторые параметры ряда или вообще оценка соответствия ряда некоторой модели.

Как правило, ряды обрабатываются статистическими методами анализа.

**Примеры задач на анализ временных рядов:**

* экономическое прогнозирование (эконометрический анализ, анализ фондовой биржи, потребление ресурсов);
* диагностика промышленных установок (например, диагностика неисправностей станков или других устройств с помощью анализа вибрации);
* прогнозирование поведения физико-химических процессов (например, прогноз солнечных пятен, прогноз погоды);
* медицинская диагностика (ЭЭГ, ЭМГ и т. д.);
* обработка сигналов (например, измерительные сигналы от некоторых промышленных датчиков);
* контроль и диагностика производства продукции (например, прогноз качества продукции по результатам измерений);
* **база данных временных рядов (TSDB)** — это база данных, оптимизированная для значений с временными отметками или база временных рядов. База данных временных рядов создана специально для обработки каких-либо значений и событий с отметками времени. В некоторых случаях базы данных имеют значительные преимущества по сравнению с классическими подходами.

**Задачи анализа временных рядов**

Основные задачи анализа временных рядов:

* прогнозирование будущих значения ряда (может быть прогнозирование значений вне выборки или внутри нее);
* восстановление неполного ряда (при отсутствии некоторых данных внутри ряда);
* оценка параметров ряда — например, период сезонности (или цикличности), наклон тренда;
* классификация участков временных рядов;
* генерация временных рядов;
* обнаружение аномалий;
* поиск закономерностей во временных рядах;
* сегментация временных рядов (поиск значимых участков);
* кластеризация временных рядов.

Кроме основных задач, в анализе временных рядов существует несколько вспомогательных, которые помогают обработать данные и сделать их более удобными для дальнейшего анализа.

**Вспомогательные задачи анализа временных рядов:**

* Шумоподавление - это процесс удаления или снижения уровня шума во временных рядах, чтобы выявить более четкие и значимые структуры в данных.
* Преобразование (трансформация) временного ряда- представляет собой процесс изменения формы или шкалы временного ряда для получения лучшего представления данных.. Нелинейное преобразование для лучшего представления временных рядов, простейшими из них являются производные и логарифмические преобразования.
* Разложение временного ряда (декомпозиция) предназначено для извлечения различных компонентов из временного ряда, таких как тренд, сезонность или цикличность. Это позволяет получить более детальное представление о структуре ряда и выделить основные особенности данных.
* Повторная дискретизация и повторное квантование рядов, а также интерполяция — сжатие, увеличение размерности, компенсация эффектов неравномерного квантования или неравномерной дискретизации, заполнение пропусков.

Для того, чтобы корректно осуществлять анализ, нам необходимо, корректно настраивать параметры ВР, например, шаг. В контексте задачи анализа временных рядов, нам приходит на помощь теорема Котельникова которая, говорит нам, что для эффективного анализа и интерпретации рядов, мы должны соблюдать определенную частоту дискретизации.

Рассмотрим ее подробнее.

**Теорема Котельникова в анализе временных рядов**

* Чем чаще процесс происходит во времени (то есть чем выше частота процесса), тем больше отсчетов дискретного временного ряда необходимо брать на единицу времени (то есть тем выше должна быть частота дискретизации).
* Теорема Котельникова утверждает, что минимум необходимо иметь по два отсчета на период процесса (то есть частота дискретизации должна быть минимум в два раза выше максимально ожидаемой частоты процесса).
* Отклонение от теоремы может привести к отображению данных во временном ряду как аномальных данных (например эффект алиасинга).
* Теорема Котельникова в первую очередь используется в случае равномерного шага дискретизации.

# Юнит 4. Модели временных рядов. Тренд, сезонность, шум

В этом юните мы сфокусируемся на аспектах временных рядов, которые позволяют нам понять тенденции в данных, выявить закономерности и таким образом подготовить основу для прогнозирования будущих значений.

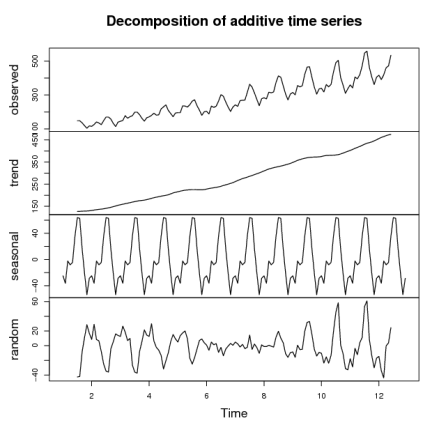
Для того, чтобы корректно описать временной ряд, нам нужно свести его к некой математической зависимости. Для этого нам понадобится термин — модель временного ряда.

**Модели временных рядов**

Модель временного ряда **у** — это специфическая (выбранная, формализованная) зависимость совместного распределения вероятностей значений модели **у** по временным шагам **n**.

В модели временных рядов часто выделяют следующие части:

* Тренд — медленно меняющаяся часть зависимости временного ряда.
* Сезонность — некоторые «относительно быстро меняющиеся» периодические составляющие, как правило тут речь идет о регулярном периоде.
* Цикличность — это некоторые периодические компоненты с «относительно медленным изменением» с нерегулярным периодом и относительно высокой интенсивностью.
* Шум — это некоторое случайное (стохастическое) искажение выходных значений: нерегулярные или случайные колебания (вариации).



Пример разложение наглядного ВР – числа пассажиров авиалиний на составляющие. ВР имеет нарастающий тренд, почти линейной формы. Также есть сезонность, в которой, если присмотреться на регулярные искажения видна и вторая сезонная компонента. Остаток от такого разложения – показан на нижнем графике. Мы всегда хотели бы чтобы этот остаток был бы белым шумом – случайным процессом с нормальным распределением. Однако, тут мы видим, что в остатке разложения есть некоторая информация – какой-то периодический процесс. Вероятно, это неточность выделения сезонной части – она видимо возникла в силу некоторой нерегулярности сезонности. Другими словами, остаток разложения – это не учтенная информация. Чем такой информации больше, тем хуже. Но важно также понимать, что все зависит от задачи. Если предполагается, что требуемая точность позволяет не учитывать какую-то информацию (влияние информации, например, ее амплитуда незначительна), то использованное разложение ВР можно считать достаточным. Другими словами, все зависит от постановки задачи.

Следует сделать несколько заметок касательно моделей временного ряда:

* При проведении тренд-сезонной декомпозиции компонент шум может рассматриваться как вся остаточная часть.
* Мы можем поделить все компоненты на детерминированные и случайные.
* Часто все компоненты , кроме шумов рассматриваются как детерминированный (заданные определенно) характер.

Однако в некоторых случаях компоненты также могут рассматриваться как стохастические (случайные), но только формально. Например, в анализе ВР мы не можем решить уравнение прямой по двум точкам. Уравнение прямой, то есть нахождение коэффициентов наклона и смещения нужно статистически. Например, это можно сделать методом наименьших квадратов. Дело в том, что такой подход учитывает неточности, которые мы ожидаем будут в значения ВР.

Относительно модели ряда предполагается, что временные ряды генерируются через регулярные интервалы времени (например, дневная температура), и поэтому называются регулярными временными рядами. Однако данные временного ряда не обязательно должны поступать через определенные промежутки времени. В таком случае это называется нерегулярным временным рядом. Пополнение счета или снятие средств в банкомате являются примерами нерегулярных временных рядов.

# Юнит 5. Модели тренда временных рядов

Простейшим случаем временного ряда является одномерная последовательная зависимость значений некоторого процесса от времени. Зависимость может быть представлена в следующей форме:

* где — это временной ряд;
* — это некоторый начальный уровень (смещение);
* как правило, — это временная метка; метка может быть Также может быть дискрет (номер выборки). В этом случае можно пересчитать выборку во время как где — период дискретизации ;

Часто цикличность и включаются в тренд, в этом случае модель может быть задана как

Иногда модель включает редкие, но регулярные события по типу праздничных дней в амодель

Также иногда следует включать в модель аномалии

В обоих формулах: – это число сезонностей, например, годовая сезонность, сезонность в квартал, сезонность в неделю – это три сезонности, это число редких но регулярных событий, например, новый год, 8 марта и тд.

**Модели временных рядов. Заметки**

В некоторых случаях также можно встретить мультипликативную модель временных рядов.

В противоположность мультипликативной, приведенная выше модель может назваться аддитивной.

В общем случае любая модель временного ряда может быть представлена как комбинация нескольких мультипликативных моделей и аддитивных моделей, например

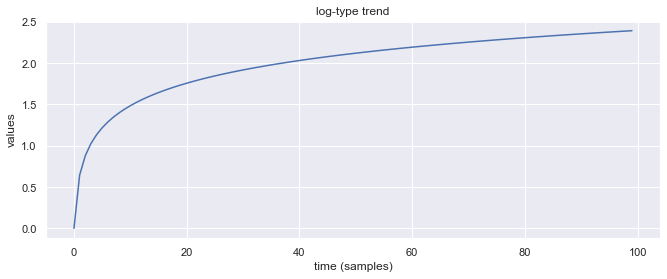
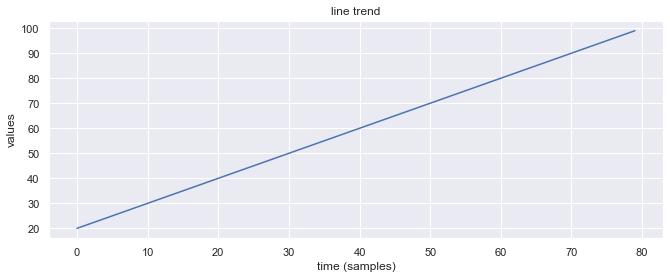
В более общих и сложных случаях модель может иметь несколько составляющих тренда или других компонентов, находящихся в некоторых (необязательно линейных) отношениях одна к другой.

**Популярные виды тренда**

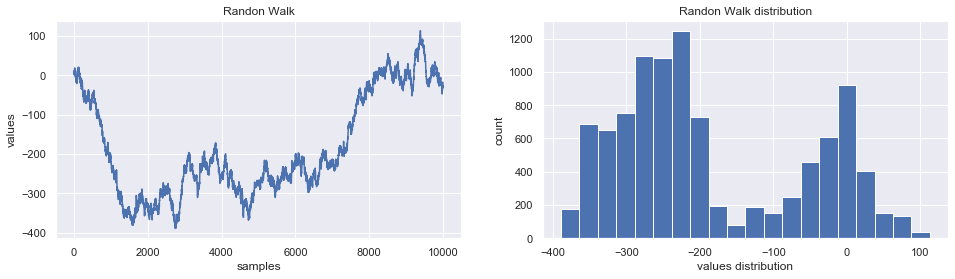
* отсутствие тренда;
* линейный тренд ;
* параболический тренд ;
* полиномиальный тренд ;
* экспоненциальный тренд ;
* насыщение(логистический) тренд ;
* циклический тренд (огибающая) ;

А также многие другие функции, способные оставаться монотонными достаточного долго.

В выражениях выше отсчеты временного ряда от времени ; – параметры тренда. Отметим, также что тренд может быть случайным или стохастический тренд, например, случайное блуждание) типа , где – это предыдущее значение ВР, а – это некоторая случайная величина.

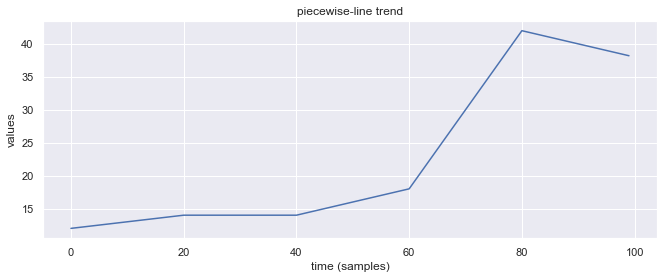
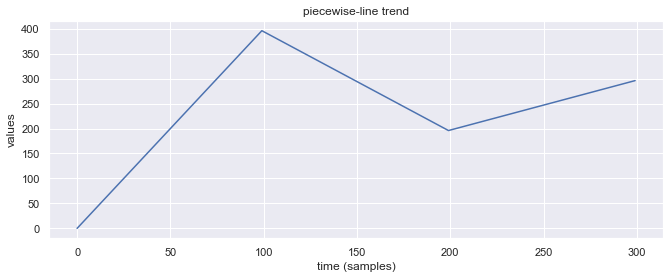


Примеры линейного тренда и тренда с насыщением ( логистический). Первый пример означает постоянный рост, например, рост загруженности колл-центра в первую неделю после открытия. Второй пример означает, что в результате быстрого роста в начале, ВР постепенно перестает расти. В примере с тем же колл-центром это бы значило, что число звонков доходит до максимума (насыщение). То есть все, кто хотел узнали про колл-центр и новых клиентов почти не появляется.



Пример стохастического тренда – тренд типа «случайное блуждание». Такой тренд появляется каждое следующее значение опирается на предыдущие, но меняет его на случайную величину. Например, на рынке, когда вы видите цены остальных участников, но можете или немного увеличить свою цену или поторговаться, или уменьшить ее. То есть изменение цены можно ожидать случайным, но связанным с ее предыдущим значением.

Отметим, что тренд не всегда монотонный. Тренд может иметь точки перегиба.



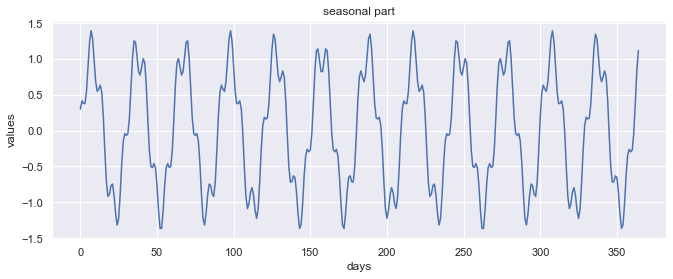
Примеры линейного тренда с точками перегиба. Такой тренд также можно назвать кусочно-линейным. Точки перегиба означают появление новизны в данных. В каком-то смысле такого рода тренд можно было разделить на две части: тренд на рост и цикличность. То есть некоторая медленноменяющаяся зависимость. Примерами точек перегиба могут быть, появление нового фактора. В примере с колл-центром, допустим, что на первой неделе его работы оказалось, что половину оборудование привезли бракованную, и загрузку пришлось уменьшить. Или спустя неделю рядом открылся конкурент, он оттянул на себя часть звонков.

Основные части временных рядов — это тренд и сезонность. Когда у нас есть данные, первое, что нам нужно, это определить часть тренда. В этом юните вы узнаете, что определяет тренд и каковы его виды.

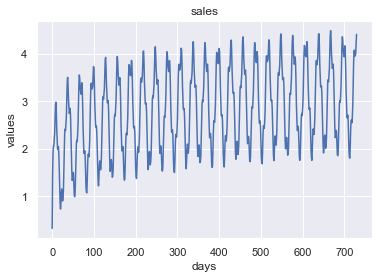
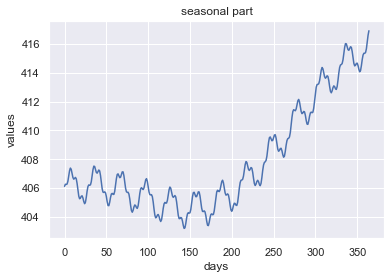
**В** **отношении** **тренда** **важны** **следующие** **моменты:**

* Часто мы не знаем, как аппроксимировать тренд наилучшим образом, если это не линейный тренд.
  + Погрешность аппроксимации тренда имеет основное влияние на погрешность прогноза временного ряда или оценки его параметров.
  + Для большинства реальных временных рядов тренд — это функция немонотонного роста (или падения). В этом случае можно сказать, что мы говорим и о тренде, и о цикличности вместе.
* Тренд может иметь точки перегиба и/или участки насыщения.
  + Часто такие точки необходимо спрогнозировать с максимальной точностью, а их интерпретация — это отдельная задача, которая может дать много подсказок для исследователя.
* В некоторых случая можно неверно предположить отсутствие тренда, если, например, сезонность слишком интенсивная, так происходит, например, в потребление электроэнергии (увидим в практической части).

Обратим внимание также на то, что такое сезонность в данных. **Сезонность** — это периодический процесс, имеющий достаточно высокий период для нашего ВР. То есть, если мы видим периодически повторяющийся процесс на фоне тренда, то называем его сезонностью. Самую простую сезонность можно представить, как синусоиду. Может быть и несколько синусоид вместе, например, одна — это годовое изменение спроса, а другая это еженедельное измените спроса. Тогда периоду сезонностей будут год и неделя соответственно. Периоду могут быть другими, например, 1 месяц. Сезонности складываются, образуя общую периодичность. Такая зависимость может казаться искаженной, но на самом деле она имеет четкую интерпретацию. Отметим также, что иногда сезонность может быть не регулярной, то есть или ее интенсивность (разброс) или ее период могут «гулять». Например, в зависимости от выходных дней или, например, внешних факторов, таких как температура, если речь идет о потребление электроэнергии. Иногда в качестве сезонности могут рассматриваться также редкие но регулярные события (например праздники для интервала в 10 лет).



Примеры одного сезона (слева) и нескольких сезонностей вместе (справа).



Пример сезонности на фоне циклического тренда и на фоне тренда с насыщением.

# Юнит 6. Статистический подход к анализу временных рядов. Основные статистические характеристики

В анализе временных рядов мы рассматриваем анализ временного ряда как статистическую задачу. В этом юните мы рассмотрим несколько важных статистических характеристик временных рядов, а также методы оценки их значимости. Статистические характеристики и методы имеют важное значение при анализе временных рядов и позволяют нам лучше понять и использовать информацию, заключенную в данных.

Благодаря такому подходу мы можем ввести следующие характеристики ряда:

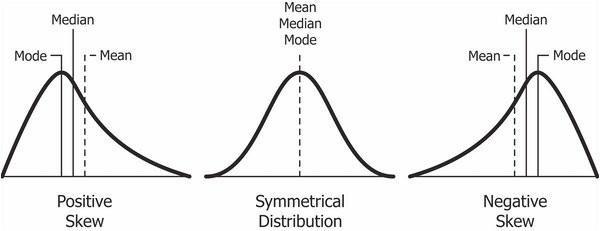
* **среднее значение** (или ожидаемое значение): напомним, что указанное выражение является лишь некоторой эмпирикой, где – это среднее значение.
* **стандартное отклонение** (или корень из дисперсии): где – это дисперсия, а – стандартное отклонение.

Отметим, что кроме среднего значения мы можем ввести еще ряд характеристик ВР, например, Медиану и Моду.

**Медиана** — это средний балл для набора данных, упорядоченных по порядку величины.

Например, учитывая упорядоченный список оценок учащихся, [14 35 45 55 55 56 58 65 87 89 92], медиана составляет 56, потому что это средняя оценка, так как перед ним 5 пунктов, а после него 5 пунктов.

**Мода** — это наиболее частая оценка в нашем гонаборе данных.



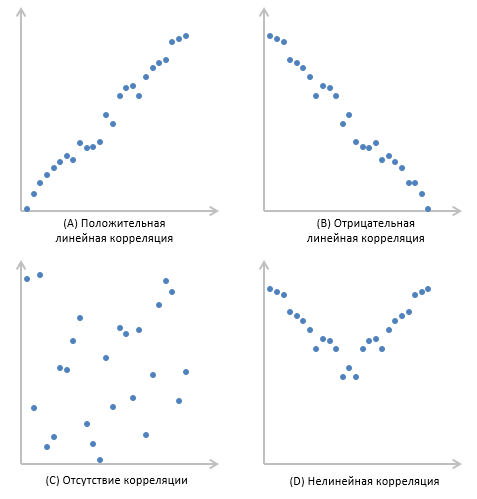
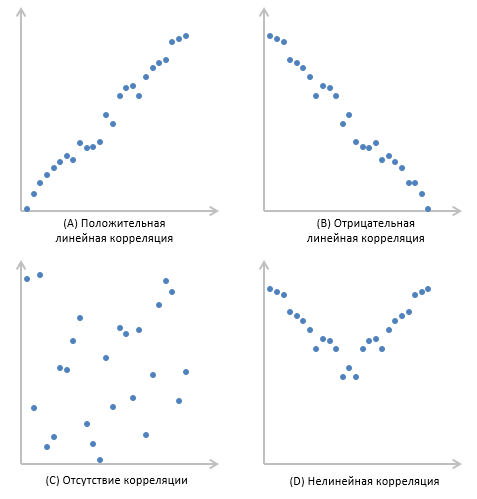
**Автокорреляционная функция**

Автокорреляционная функция (ACF) — это степень линейности временного ряда, связанная с отстающей (отложенной) версией самого себя.

В целом Корреляция (или взаимная корреляция) — это степень линейности двух временных рядов, связанная с одним из них с запаздывающей (отложенной) версией другого.

Обратите внимание: называется -м лагом. То есть временной задержкой на отсчетов. Другими словами, мы берем два временных ряда, один из которых запаздывает или опережает другой на отсчетов и считаем между ними корреляцию. Если оба ряда одинаковы. То есть второй ряд – это сдвинутая копия первого, то мы будем иметь автокорреляцию.

Напомним, что корреляционной коэффициент (корреляция двух выборок) , - это степень взаимосвязанности двух выборок ( и ). Каждая выборка имеет длину , коэффициент их корреляции, он имеет диапазон от до . То есть коэффициент, близкий к говорит о высокой степени схожести двух выборок, тогда как 0 коэффициент корреляции говорит об абсолютной независимости двух выборок.

**

-1

1

отсутствие корреляций – нелинейная связь

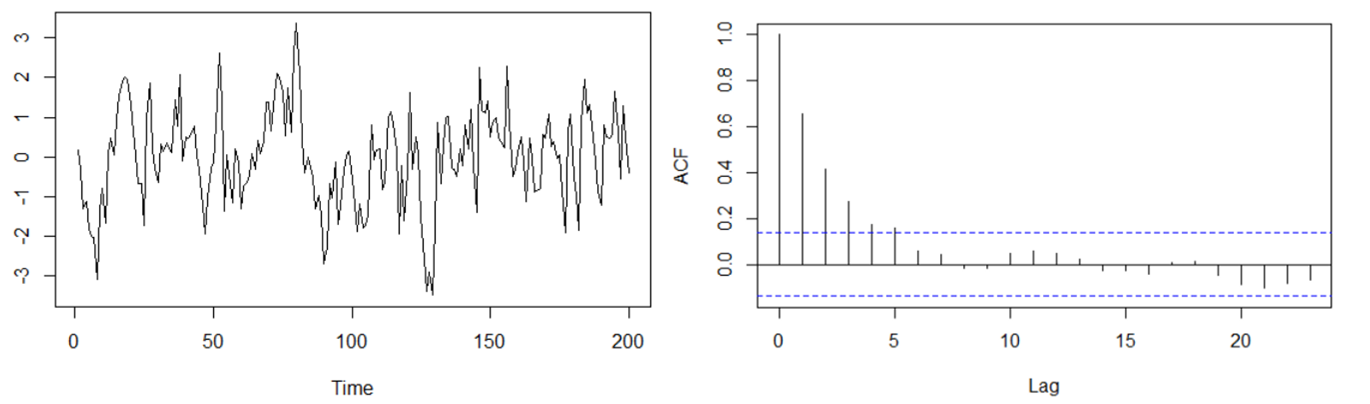
отсутствие   
корреляций

Отрицательная   
линейная корреляция

Положительная   
линейная корреляция

Пример высокой взаимосвязи –это температура на улице и в доме. Особенно если нет отопления. Если отопление в доме создало бы климат-контроль (то есть постоянную температуру в течение года), то корреляция внешней температуры и в доме была бы нулевой.

На самом деле функция автокорреляции обычно отображается на графике в виде значений для заданных лагов так, как это показано ниже. Мы ожидаем что ВР такая корреляция достаточно быстро пойдет на спад. То есть рано или поздно окажется, что сдвинутая копия ВР перестанет быть похожей на себя. Но это происходит не всегда. Более того если во ВР есть шумы, то они будут влиять и на значения корреляции. Причем если эти значения будут небольшими, то сложно будет сказать, что больше влияет шумы или детерминированная часть ВР. Поэтому на практике мы считаем, что есть некоторый уровень корреляции, ниже которого уверенности в достоверности значений нет. Это уровень уверенности. Его мы также отображаем на графиках.

****

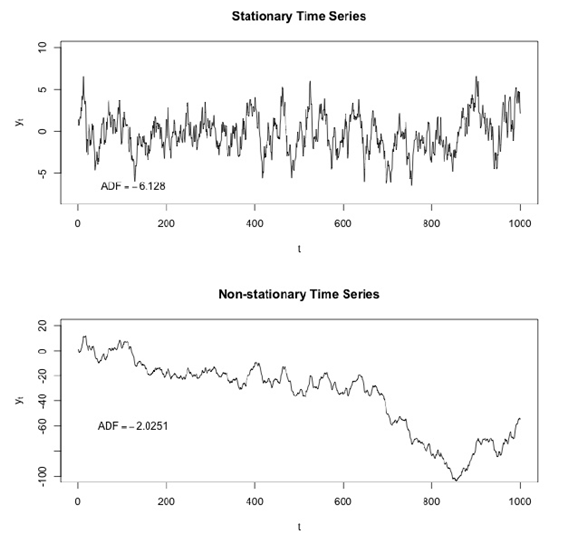
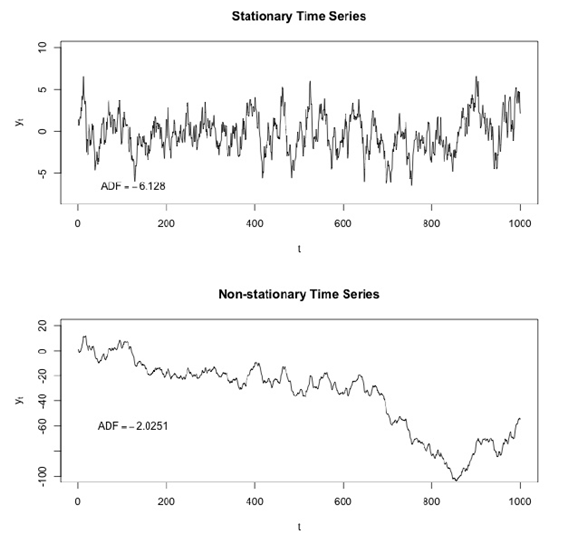
Пример АКФ. Синяя линия показывает уверенность (зависит от дисперсии).

# Юнит 7. Анализ стационарности/нестационарности

Понятие стационарности является одним из основных понятий для анализа временных рядов. Если ряд стационарен, это означает, что его свойства повторяются, и это позволяет нам использовать методы анализа, которые основаны на стационарности. Другими словами все шаблоны поведения ВР должны быть регулярными. Если мы говорим о шуме во ВР, то шум должен всегда иметь одни и те же параметры распределения.

Без стационарности мы не можем предположить, что среднее значение и дисперсия процесса постоянны во времени, и, следовательно, не можем использовать статистические методы, которые основаны на этом предположении.

В данном блоке мы разберемся с понятием стационарности и нестационарности, а дальше поговорим о методах преобразования временных рядов, чтобы мы могли работать с ними как со стационарными.



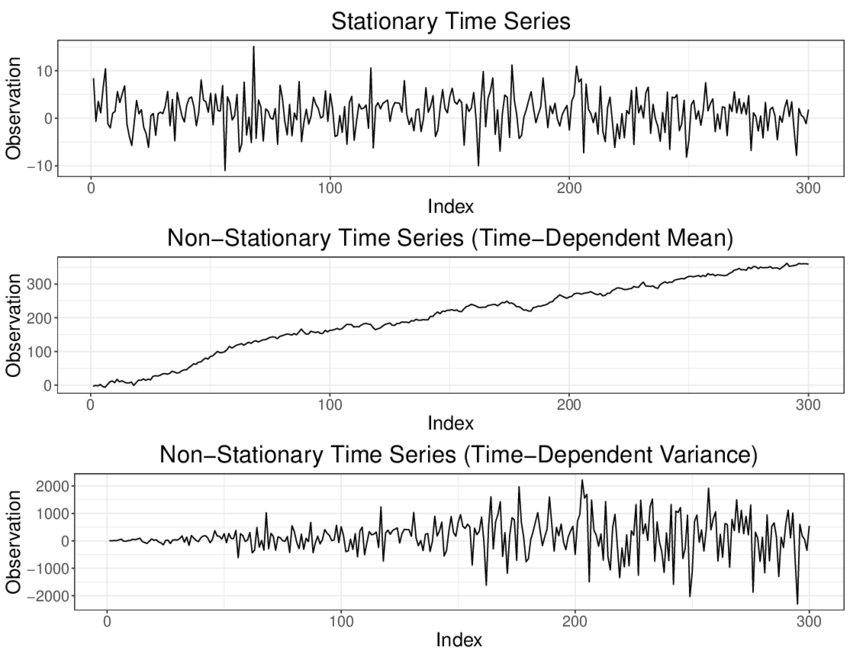
Пример стационарного и нестационарного ВР. Обратите внимание, что у стационарного ВР не видны четкие изменения со временем. Вероятно, распределение будет всегда одним и тем же. Для нестационарного ВР шаблоны явно не повторяются. То есть если мы возьмем окно в один промежуток времени, оно не будет эквивалентно окну в другой промежуток времени.

Можно выделить два определения стационарности: сильное и слабое.

Временной ряд называется **стационарным (в слабом смысле)**, если его статистические свойства (моменты) не меняются с течением времени. Другими словами, ряд имеет постоянное среднее значение и дисперсию, а ковариация зависит только от разницы индексов лагов. Указанные статистические характеристики не должны иметь бесконечные значения.

**Сильная стационарность** означает, что ряд (и его распределение со статистической точки зрения) не меняются с течением времени.

Временной ряд считается **нестационарным**, если некоторые из его статистических свойств меняются с течением времени. Другими словами, он имеет переменную среднего времени, дисперсию или другие статистические характеристики.

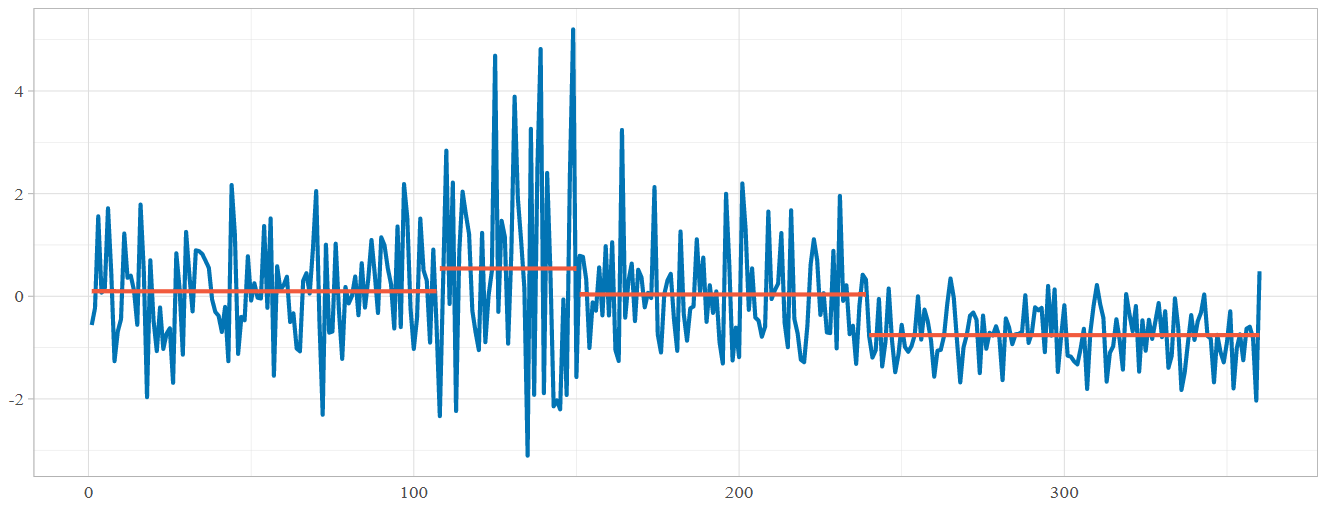


Еще примеры нестационарных ВР. Не стационарность может быть отдельно по среднему значению, по дисперсии или одновременно по обоим параметрам.

В идеале мы хотим иметь стационарный временной ряд. То есть такой, в котором найдя решения для одной его части мы всегда будем уверены, что решение останется тем же для любой другой его части. Для этого нам достаточно слабого определения стационарности. На самом деле нам даже можно тут **неформально!** Провести замену определения стационарности на определение регулярности. То **есть** мы ожидаем именно регулярность шаблонов поведения ВР, а не строгое соответствие определению. Но такая замена справедлива не всегда. Есть ряд алгоритмов, формально допускающих указанную не стационарность ВР. Далее мы увидим, как приводить ряд к стационарному.

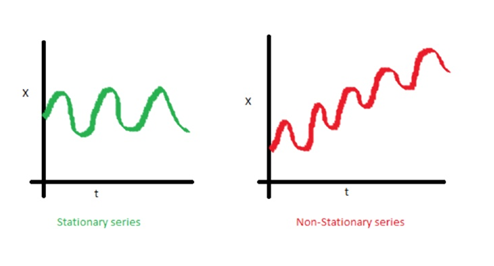
Поскольку не все ВР являются стационарными, мы можем выполнять различные преобразования, чтобы сделать их стационарными. Стационарный временной ряд может быть описан только его средним значением и значениями дисперсии (и ковариации), для нестационарных временных рядов существует множество других статистических характеристик.

Мы можем рассматривать стационарное поведение для каждой части временного ряда: тренд, цикличность, сезонность, шумы.

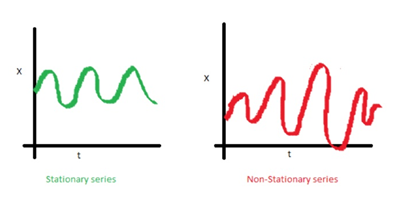


Пример нестационарного ряда с разными средними значениями и значениями дисперсии

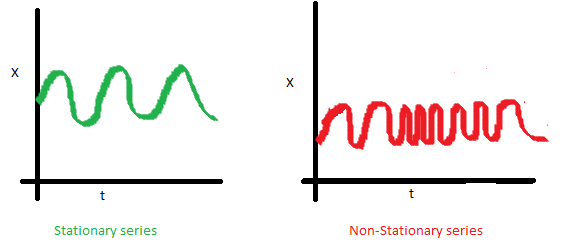
• Пример нестационарности линии тренда



* Пример сезонной нестационарности (в смысле дисперсии)



Пример сезонной нестационарности (в смысле периода сезонности)



Мы упоминали с вами шум, когда говорили о модели временного ряда. Шум является важным фактором, который может влиять на результаты анализа. **Шум —**  это некоторое случайное (стохастическое) искажение выходных значений.

Шум бывает белым и цветным, существует даже шум «соли и перца» (шум характерен для изображений и представляет собой случайно возникающие черные (перец) и белые пиксели (соль)).

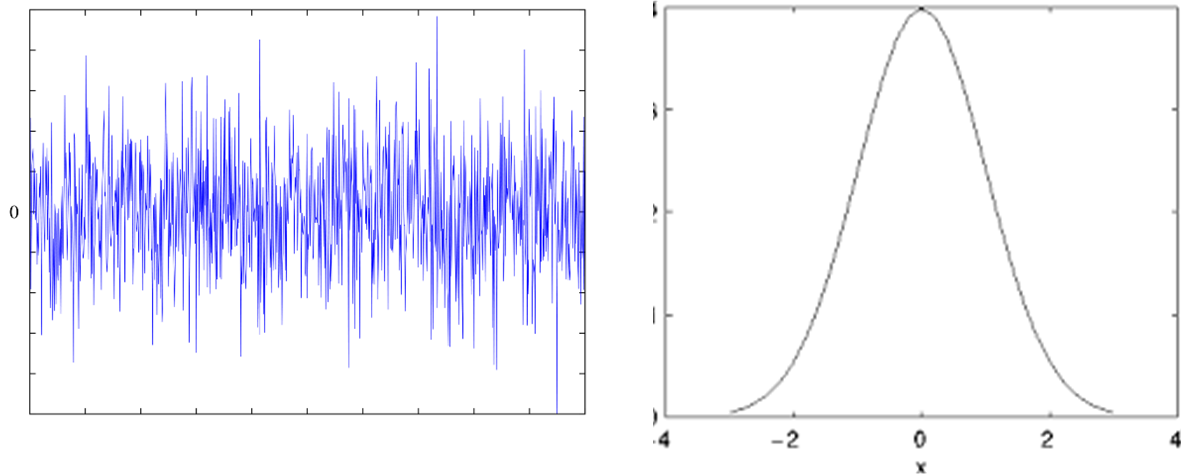
Но так как мы говорим с вами про стационарность, одним из наиболее важных случаев стационарного анализа является стационарность шума.

**Белый гауссов шум (white gaussian noise, WGN)**

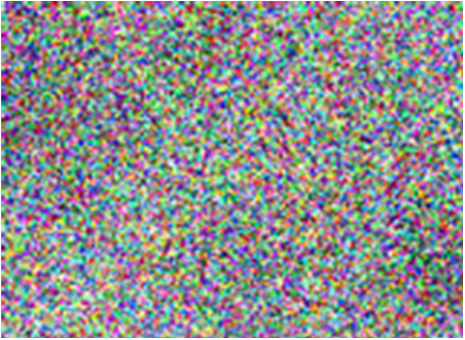
* Шум стационарности с независимыми выборками во времени (с автокорреляцией 0) называется шумом белого гаусса (WGN).
* Таким образом, WGN — это ВР с 𝑒𝑣 = 0; 𝑣𝑎𝑟 <∞; 𝑐𝑜𝑣 (𝑛) = 𝑠𝑡𝑑⋅𝛿 (𝑘 − 𝑛), где 𝛿 (𝑘 − 𝑛) = 1, если 𝑘 = 𝑛, и 0 в остальных случаях.

Обратите внимание, что дисперсия также может быть названа мощностью шума.

* Если WGN имеет неизменное среднее и дисперсию (константы), его также можно назвать независимой и идентично распределенной (i.i.d).
* Многие стационарные процессы (задачи) во временных рядах моделируются как детерминированные на фоне белых гауссовских шумов.
* Обратите внимание, что ВР типа WGN имеет так называемое нормальное распределение с нулевым средним.

****

Пример WGN и его распределения. Такое поведение можно слышать, например, когда неверно настроил радиоприемник. Также WGN возникает, когда вы выключаете антенну с телевизора и видите так называемый «снег».

****

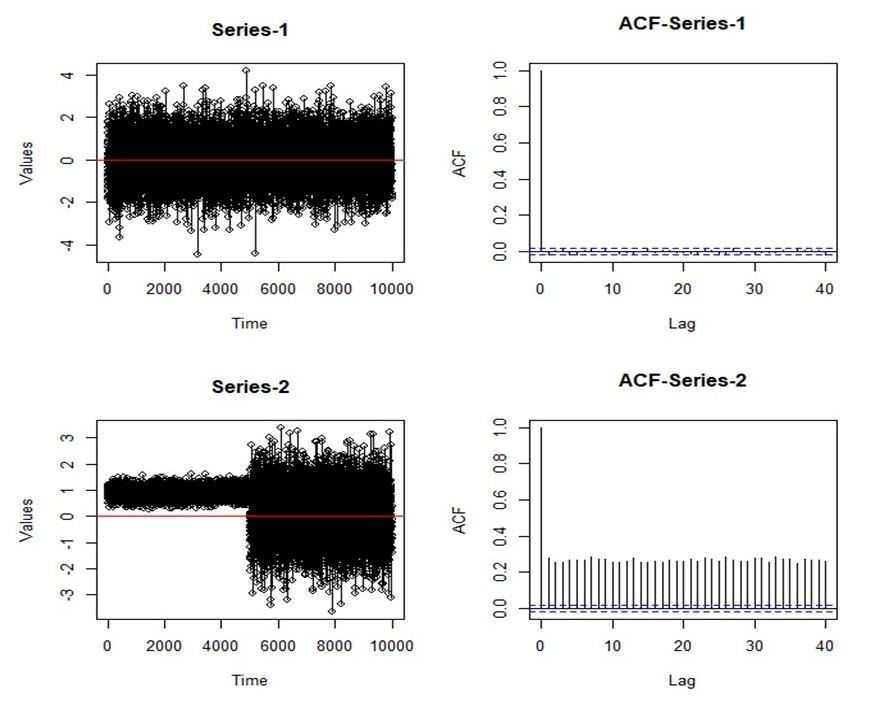
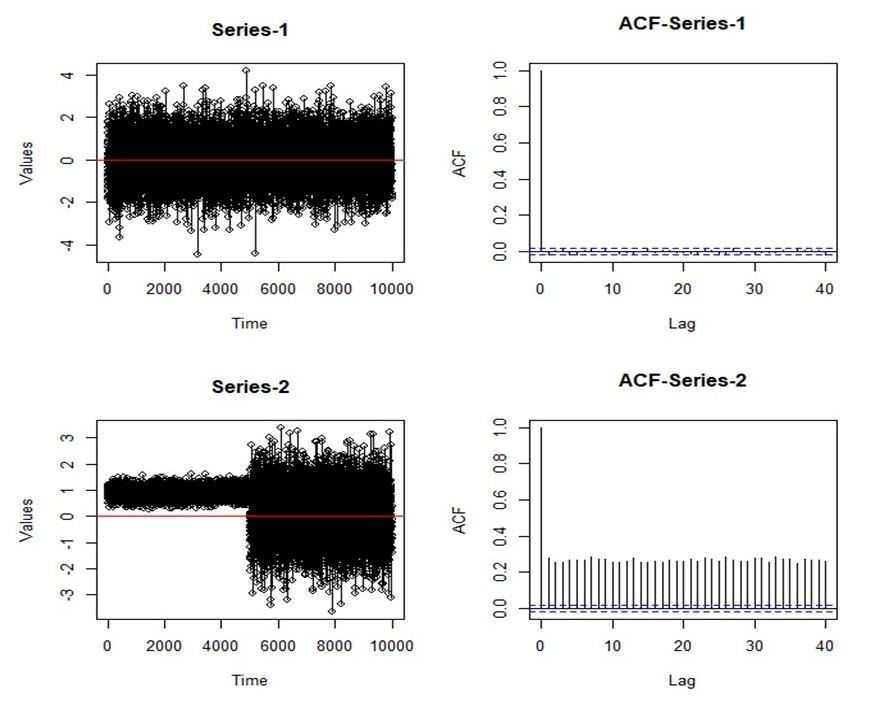
В конце также отметим, что влияние нестационарности, тем сильнее, чем дальше горизонт прогнозирования. То есть на сравнительно небольшом горизонте оказывается, что не слишком значительная нерегулярность дает лишь не слишком значительное увеличение ошибки прогнозирования. Однако, неучтенная каким-либо образом нестационарность как накопительная погрешность может привести к росту ошибки прогноза. То есть виляние нестационарности следует учитывать сообразно длительности прогноза (горизонту прогнозирования) либо некоторых эквивалентнов этого параметра в других задачах.

# Юнит 8. **Преобразования нестационарных временных рядов.**

На самом деле почти все временные ряды нестационарны (в той или иной степени). Однако, как уже обсуждалось, в зависимости от требований к точности, в некоторых случаях мы можем аппроксимировать ряд (сделать предположение), что он является стационарным для нашей задачи.

Один из самых простых способов преобразовать ряд — взять его численную производную, получив ряд:

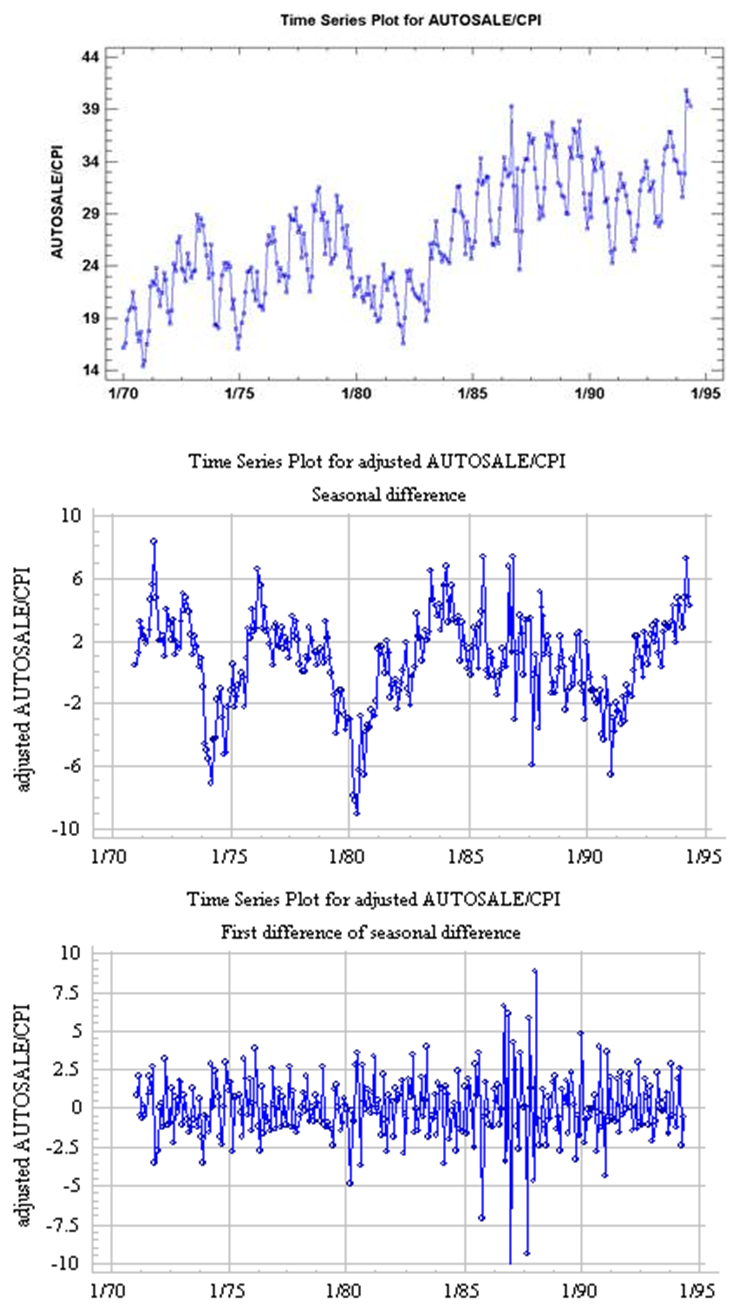
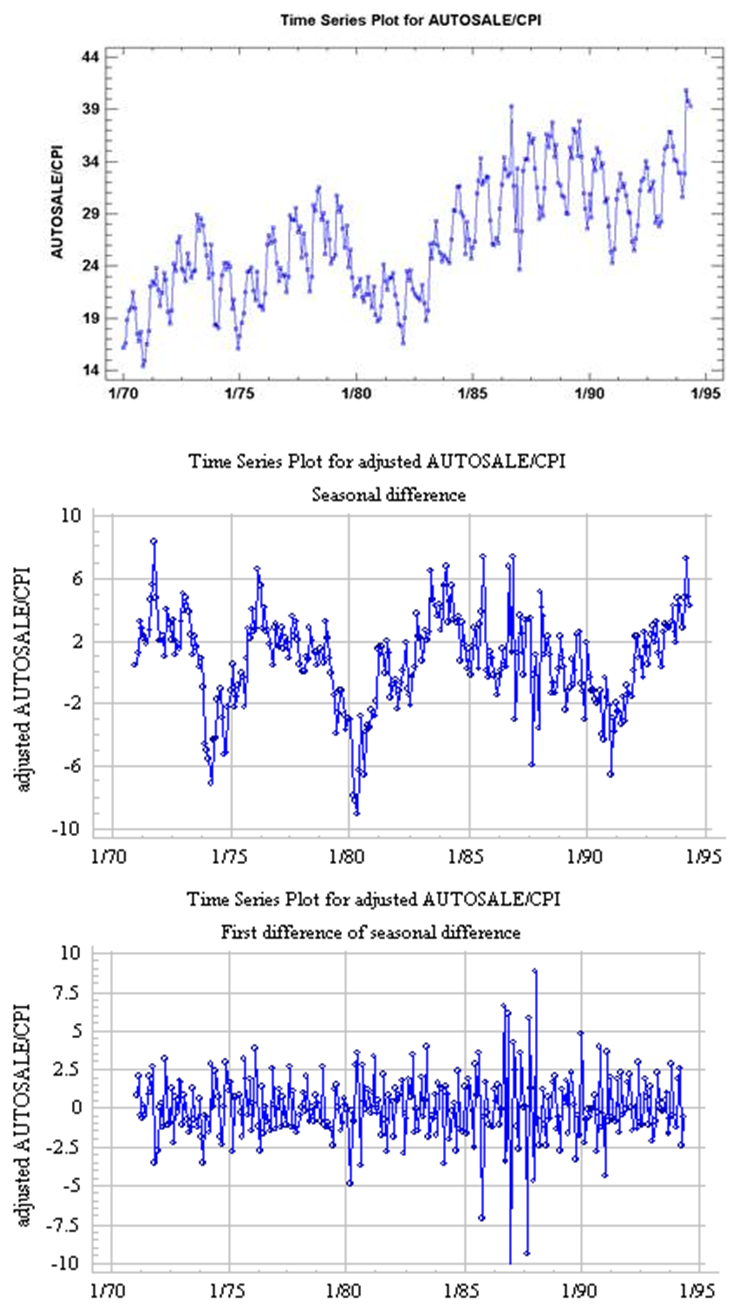
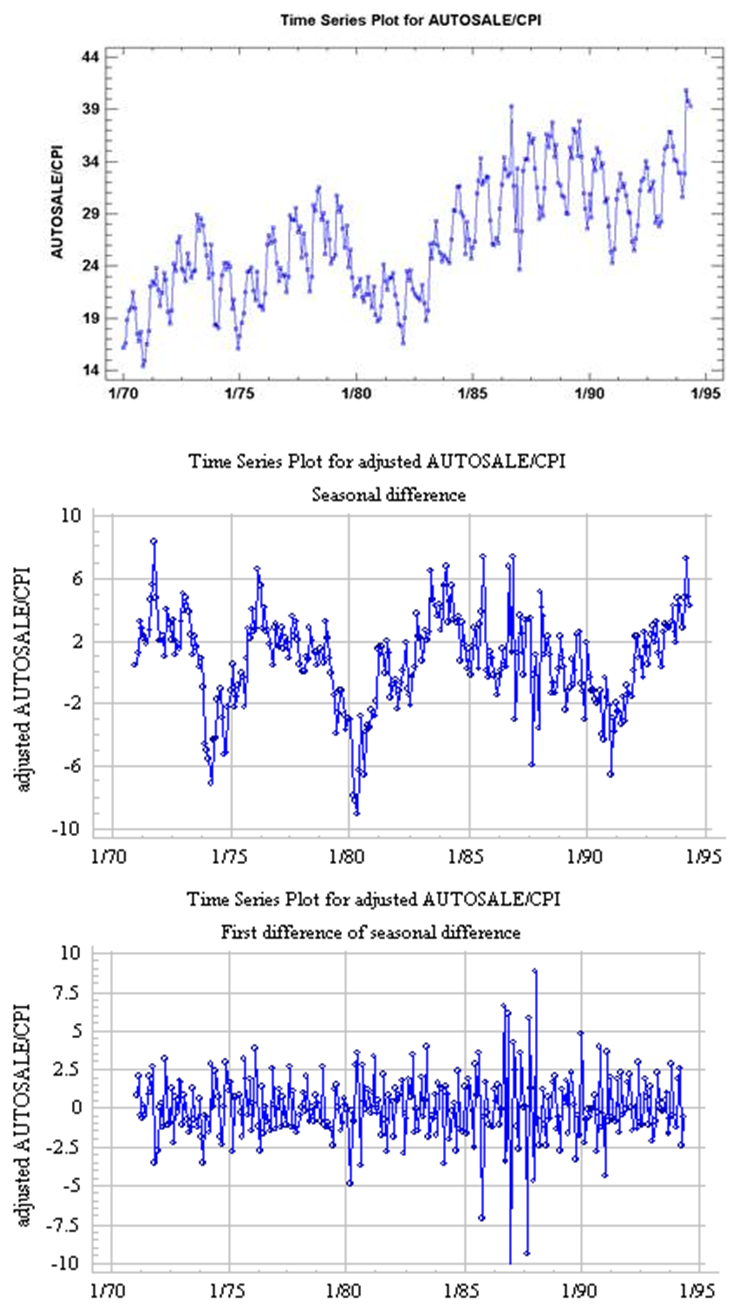
где– ВР в момент отсчета ; - производная ВР; – вторая производная. Степень производной при получении стационарного значения ВР также называется **параметром интегрирования**. Тут мы ожидаем   
(и есть такие процессы) для которых сам ВР нестационарен, однако его изменения (или ускорение изменений и т.д.) стационарны. К таким процессом относятся случайное блуждание или например турбулентность, а также многие другие. Читатель может без труда убедиться, что линейный тренд может быть компенсирован при помощи первой производной, а тренд парабола может быть компенсирован при помощи производной второго порядка.



Пример стационарного ВР и нестационрного, но с двумя участками стационарных приращений. Последний можно будет считать стационарным вплоть то точки перехода одного приращения в другое. Примеры синтетические.

Таким же образом можно выполнить компенсацию сезонности с использованием таковых разностей, когда ожидается, что значения выборки будут иметь одинаковые значения в случае отсутствия тренда.

где — сезонная производная; — период сезона; — период сезонности второго порядка. Отметим, что сезонная и обычная производная часто комбинируются (например, сначала обычная, потом сезонная).



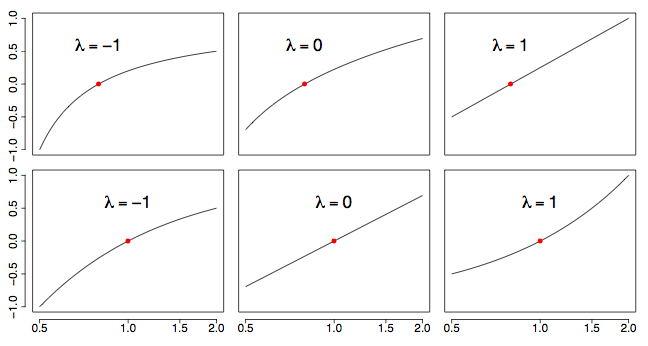
Пример ВР, в котором видно значительное влияние как тренда, так и сезонности. Для ВР сначала произведена сезонная производная, потом для результат обычная производная для устранения тренда.   
Отметим, что на самом деле не всегда требуется такой подход. Возможно было бы достаточно компенсации только тренда. Однако, как показывает практика, чем более стационарен ВР, тем меньше требует параметров для его аппроксимации (например, для предсказания его значений), а значит тем выше обобщающая способность.

В зависимости от типа нестационарности ВР для его преобразования к стационарному виду могут быть использованы различные методы. Примерами простых методов преобразования временных рядов являются:

* логарифмические преобразования,
* преобразования Гетероскедастичности
* преобразования Бокса-Кокса где – результат преобразования, – параметр

Преобразование Бокса-Кокса имеет смысл компенсации нелинейности тренда и/или снижения волатильности – временной неравномерности дисперсии. Отметим, что это параметр также называется **Гетероскедастичность.**

**То есть, напомним, что Гетероскедастичность** — неоднородность наблюдений, выражающуюся в неодинаковой (непостоянной) дисперсии случайной ошибки ВР. Свойство одинаковой дисперсии возмущений называется **гомоскедастичностью.**



Примеры результатов Влияния преобразования Бокса-Кокса на линейный тренд.

# Юнит 9. Итоги Полу-модуля

В этом модуле вы познакомились с основными понятиями, задачами анализа и методами моделирования временных рядов, это знание позволит вам делать более сложные предсказания и станет фундаментом для понимания практических методов работы с временными рядами.

**Благодаря изучению модуля вы познакомились с:**

1. Временным рядом, его особенностями и сферой применения.

2. Анализом временных рядов: основными и вспомогательными задачами анализа временных рядов.

3. Моделями временных рядов: тренд, сезонность и шум как общие компоненты временных рядов.

4. Моделями тренда: моделями, используемыми для выявления особенностей тренда во временных рядах

5. Статистическим подходом к анализу: методами статистического анализа при анализе данных временных рядов

6. Анализом стационарности/нестационарности: свойствами стационарных рядов, различиями между стационарными и нестационарными временными рядами.

7. Преобразованием нестационарных временных рядов: методами, используемыми для превращения нестационарных временных рядов в стационарные

8. Гетероскедастичностью и гомоскедастичностью: неодинаковой и одинаковой дисперсией случайной ошибки временного ряда

Анализ и моделирование временных рядов являются важным инструментом датасаентиста в различных областях: экономике, финансах, бизнесе, инженерии, метеорологии, медицине и др. Специалист, который владеет работой с ВР, может помочь спрогнозировать будущие тенденции и поведение на рынках, дать рекомендации для инвестирования, понять и устранить проблемы в производственном процессе, в медицинских данных и т.д.

Мы продолжим с вами разбираться со временными рядами в следующем модуле!

**Дополнительная литература**

1. A Gentle Introduction to Autocorrelation and Partial Autocorrelation / machinelearningmastery.com — [Электронный ресурс]. URL: https://machinelearningmastery.com/gentle-introduction-autocorrelation-partial-autocorrelation/ (дата обращения: 10.06.2023).
2. A Gentle Introduction to Exponential Smoothing for Time Series Forecasting in Python / machinelearningmastery.com — [Электронный ресурс]. URL: https://machinelearningmastery.com/exponential-smoothing-for-time-series-forecasting-in-python/ (дата обращения: 10.06.2023).
3. Autocorrelation / www.itl.nist.gov — [Электронный ресурс]. URL: https://www.itl.nist.gov/div898/handbook/eda/section3/eda35c.htm (дата обращения: 10.06.2023).
4. Computer Science Center (2016) [лекция] Лекция 9. Экспоненциальное сглаживание. Распознавание образов: метод к-го ближайшего соседа / YouTube. 28 декабря (<https://www.youtube.com/watch?v=M0Hz1u59Ysw>). Просмотрено: 10.06.2023.
5. How to Perform a Durbin-Watson Test in Python / www.statology.org — [Электронный ресурс]. URL: https://www.statology.org/durbin-watson-test-python/ (дата обращения: 10.06.2023).
6. Installing statsmodels / www.statsmodels.org — [Электронный ресурс]. URL: https://www.statsmodels.org/devel/install.html (дата обращения: 10.06.2023).