

Временные ряды 2

...

Построение признаков

2019

План

1. ДЗ1
2. Теоретическая часть
 - a. Тренд , цикл , Сезон
 - b. Понятие стационарного ряда
 - c. Интегральный ряд
 - d. Модели временного ряда
 - e. Сглаживание
 - f. тест ADF
3. Практическая часть - построение признаков от значений ряда
 - a. как запустить анализ стационарности
 - b. как запустить сглаживание
 - c. как собрать модель
 - d. ДЗ 2

Тренд, сезонность, цикл, шум

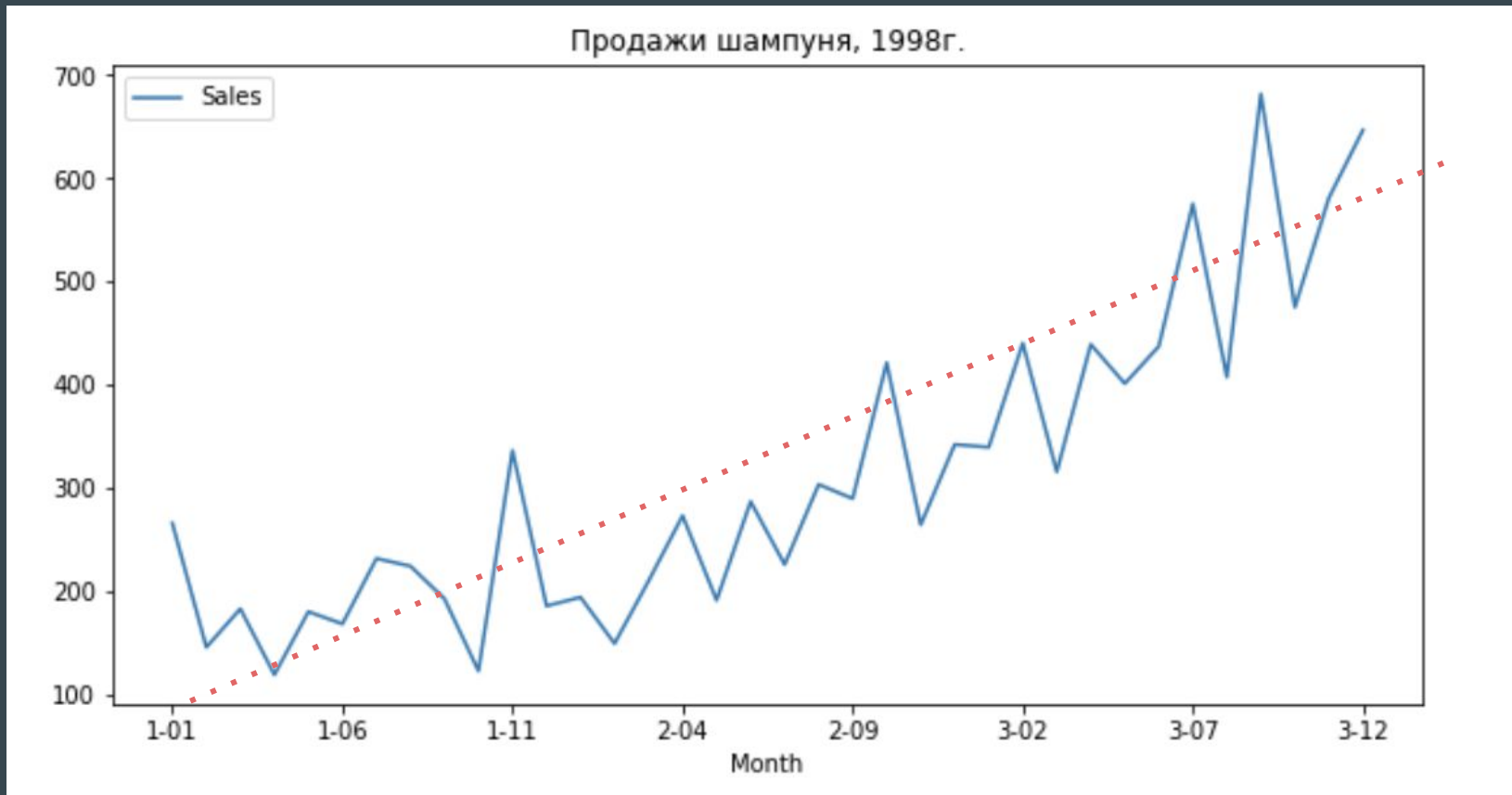
Тренд - плавное долгосрочное изменение ряда

Сезонность - циклические изменения уровня с постоянным периодом

Цикл - изменения уровня ряда с переменным периодом

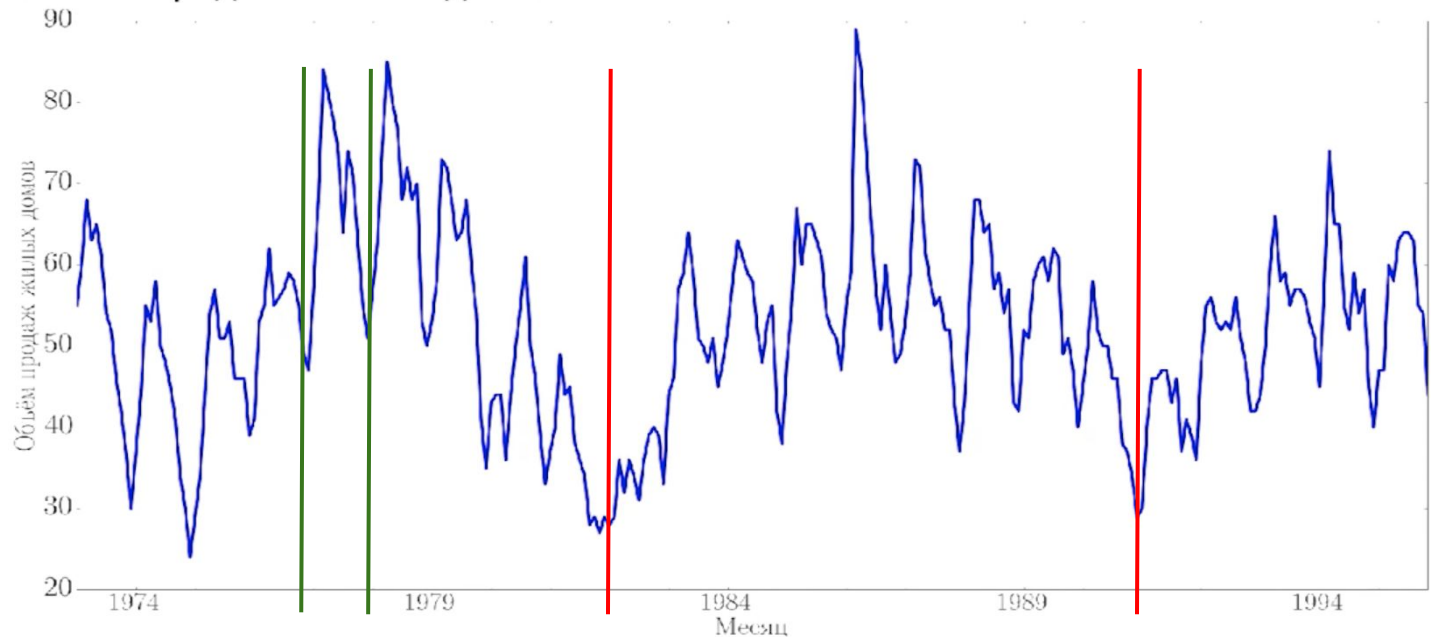
Шум - ошибка, случайные события, которые мы не можем описать

Тренд



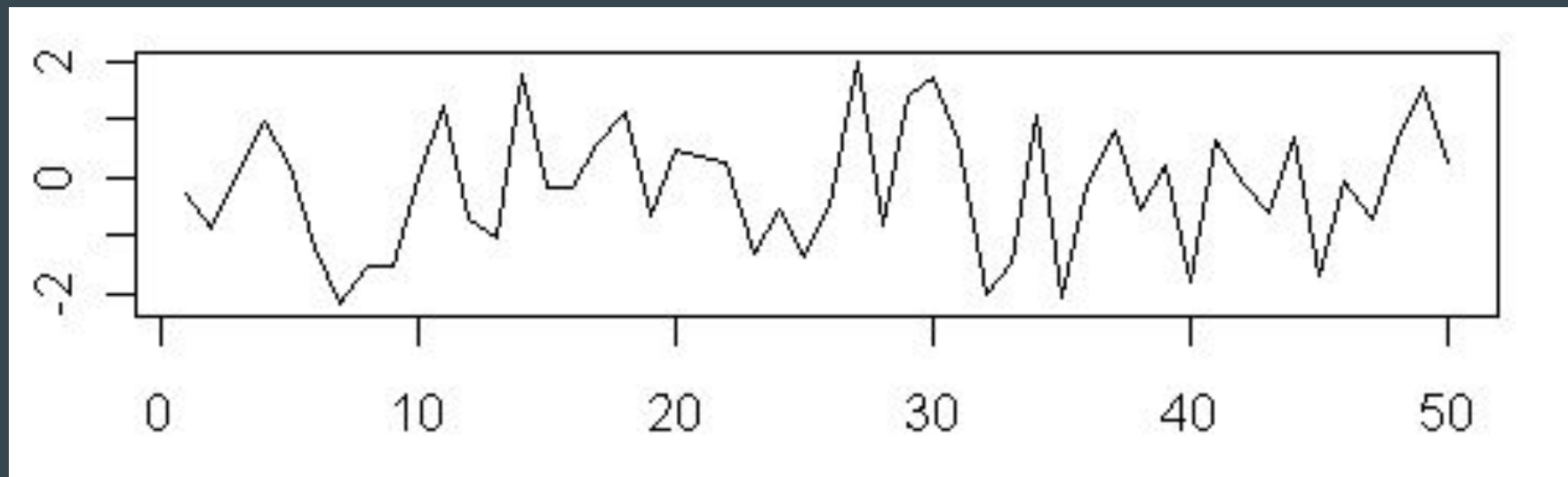
Цикл и Сезонность

Объём продаж жилых домов:

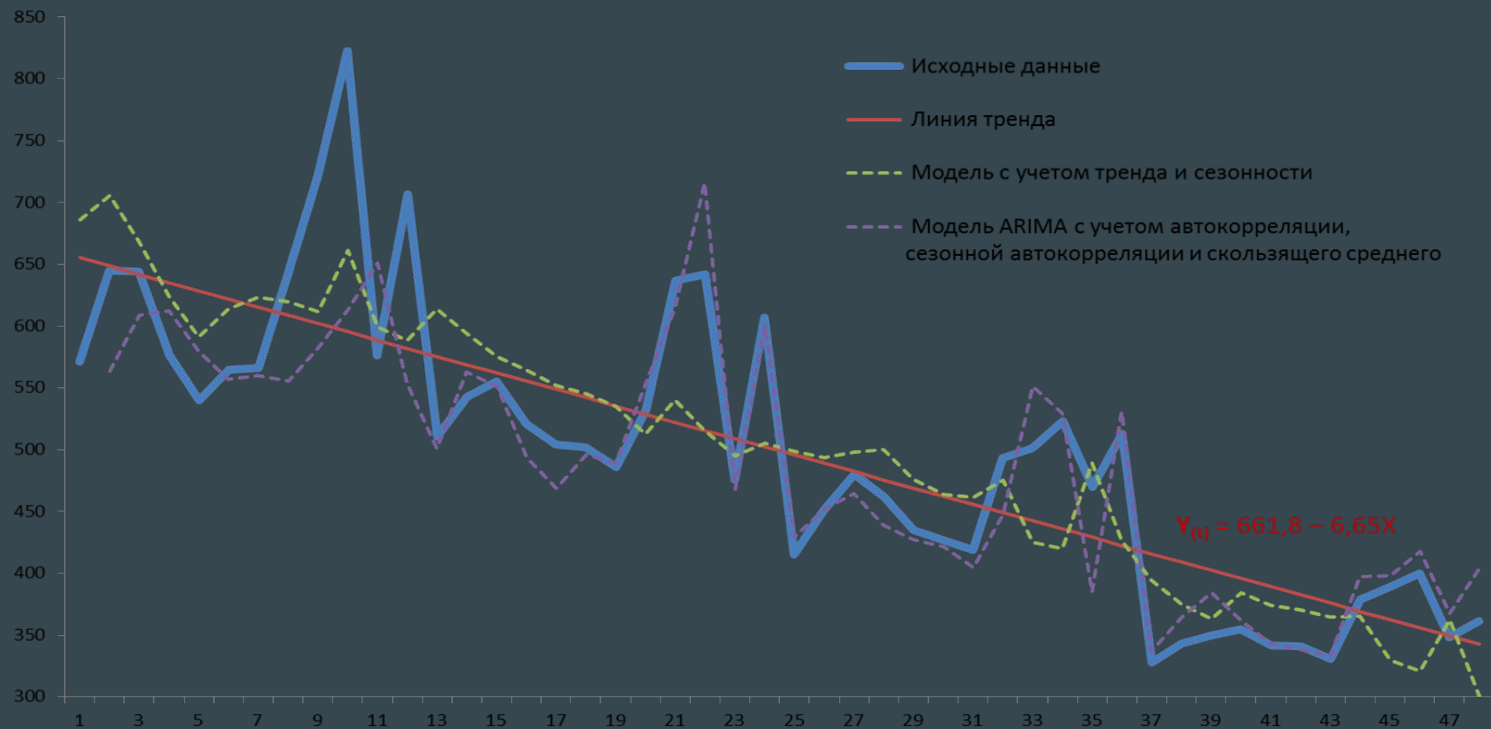


Годовая сезонность, экономические циклы

шум



ВСЕ ВМЕСТЕ



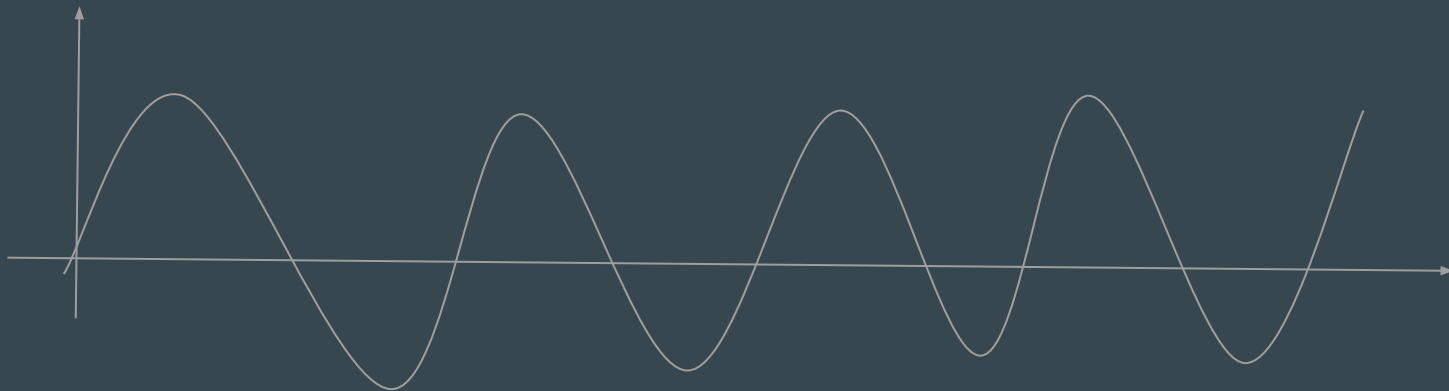
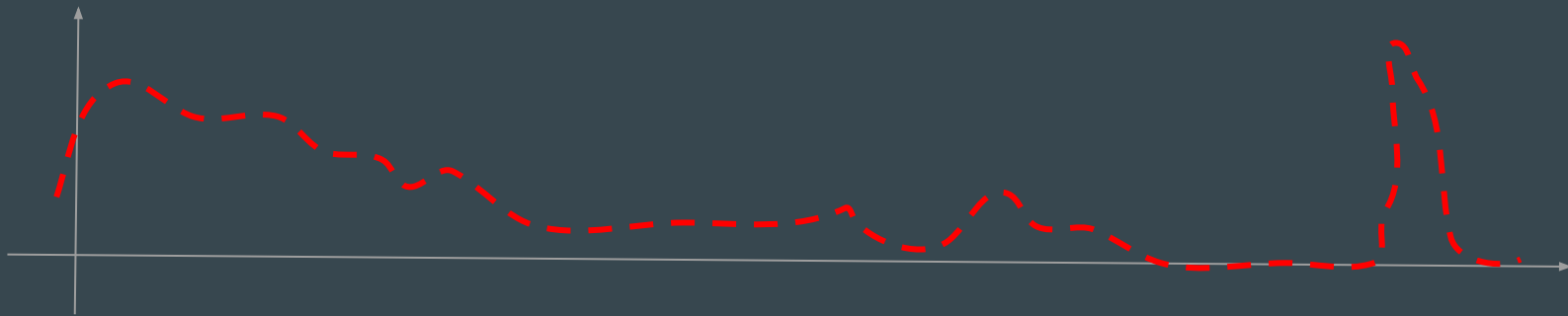
Стационарность

Ряд называется стационарным, если его свойства не зависят от времени

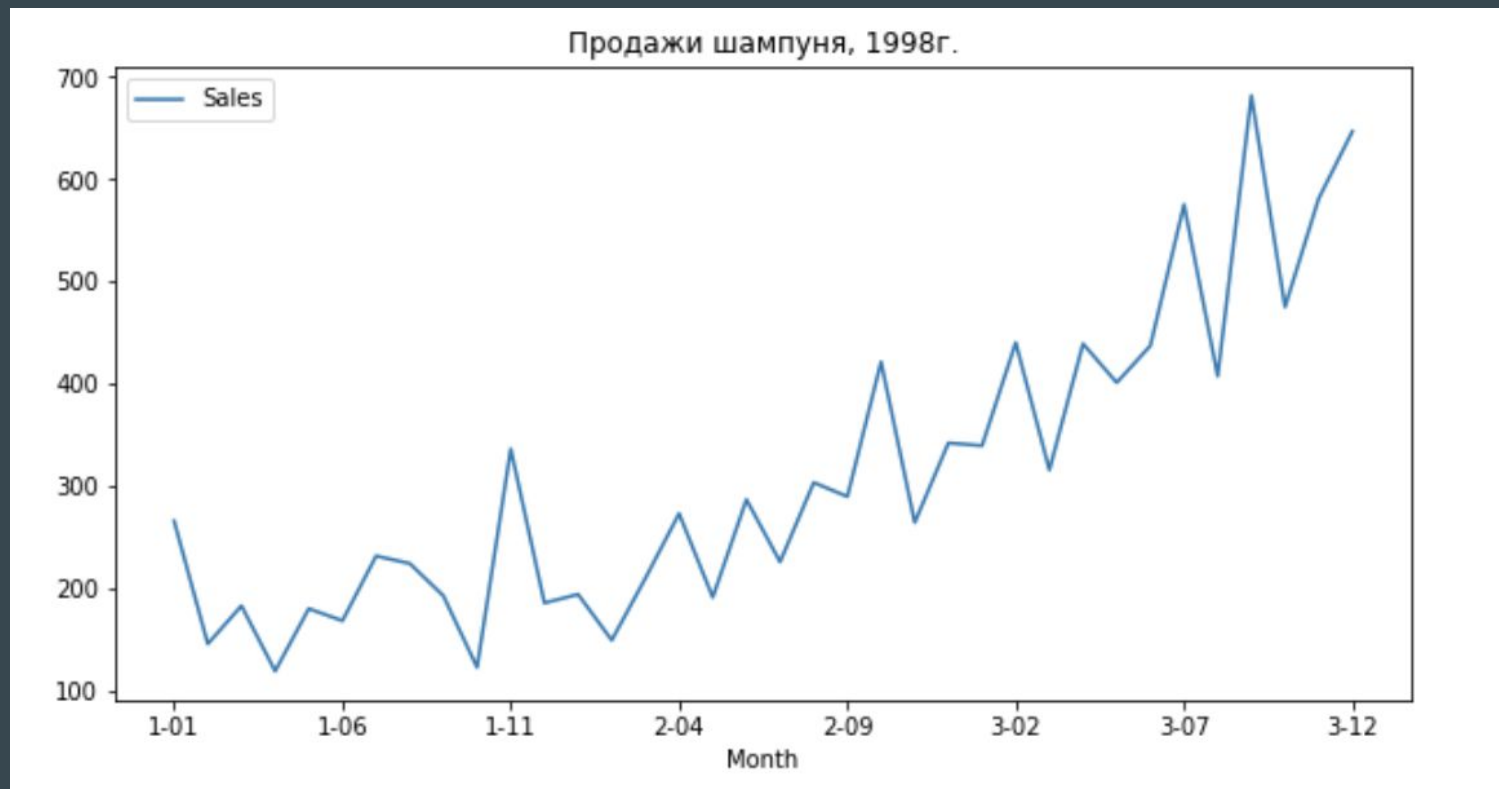
Сезонность и тренд -> нестационарность

Цикличность ~~->~~ нестационарность

Стационарность

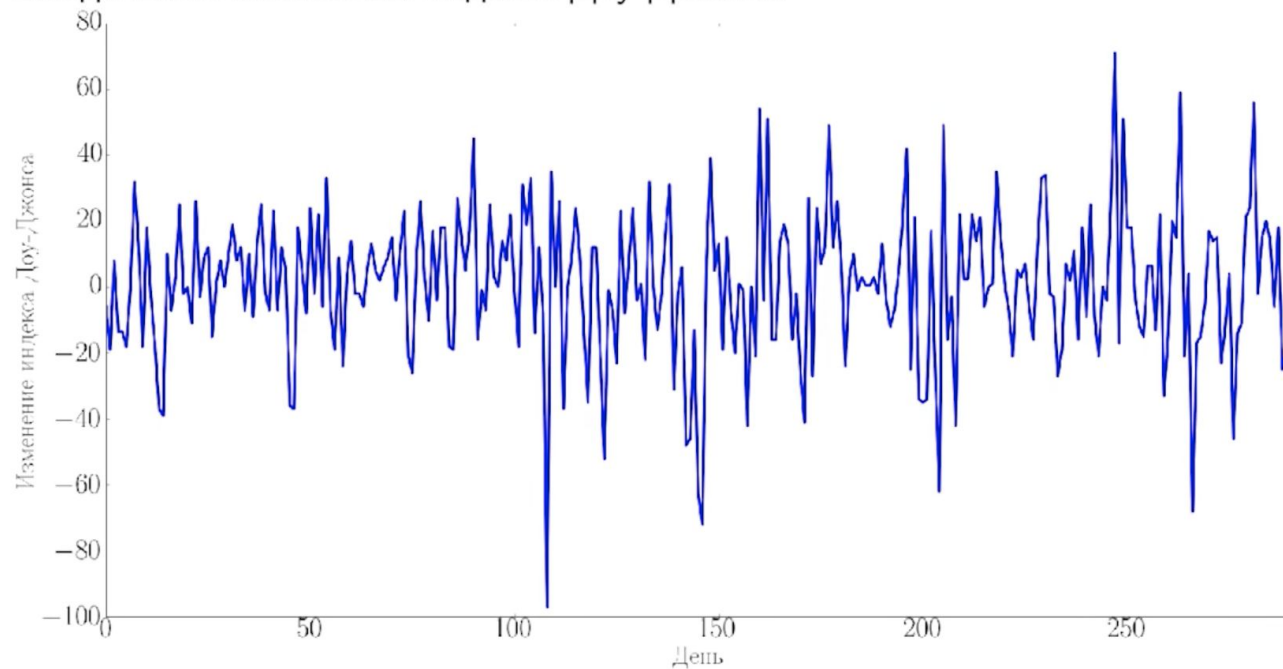


Пример: продажи шампуня



Пример: изменения индекса Доу-Джонса

Ежедневные изменения индекса Доу-Джонса:



Ничего

Стационарность подразумевает:

Постоянство :

- Матожидания
- Дисперсии
- Автоковариация стационарного ряда с лагом L

p-value

Ошибка первого рода (α) - вероятность отклонить нулевую гипотезу, когда она верна

$\alpha = 0.05, 0.1, 0.01$

p-value - вероятность получить такие отклонения или еще больше от нулевой гипотезы

Если $p\text{-value} < \alpha$, то отклоняем нулевую гипотезу в пользу альтернативной

Тест Дики-Фуллера - простая версия

Авторегрессия для 1-х разностей $dY(t) = Y(t) - Y(t-1) = b * Y(t-1) + e(t)$

При $b = 0$ у нас от разностей остается только шум \rightarrow шум стационарен, процесс не зависит от времени

Нулевая гипотеза H_0 : ряд не стационарен ($b \neq 0$)

Альтернативная H_1 : ряд стационарен ($b = 0$)

Тест есть в python <https://www.statsmodels.org/stable/generated/statsmodels.tsa.stattools.adfuller.html>

Тест Дики-Фуллера (расширенная)

Авторегрессия $Y(t) = a_1 * Y(t-1) + e(t)$

Авторегрессия для 1-х разностей $dY(t) = Y(t) - Y(t-1) = b * Y(t-1) + e(t)$

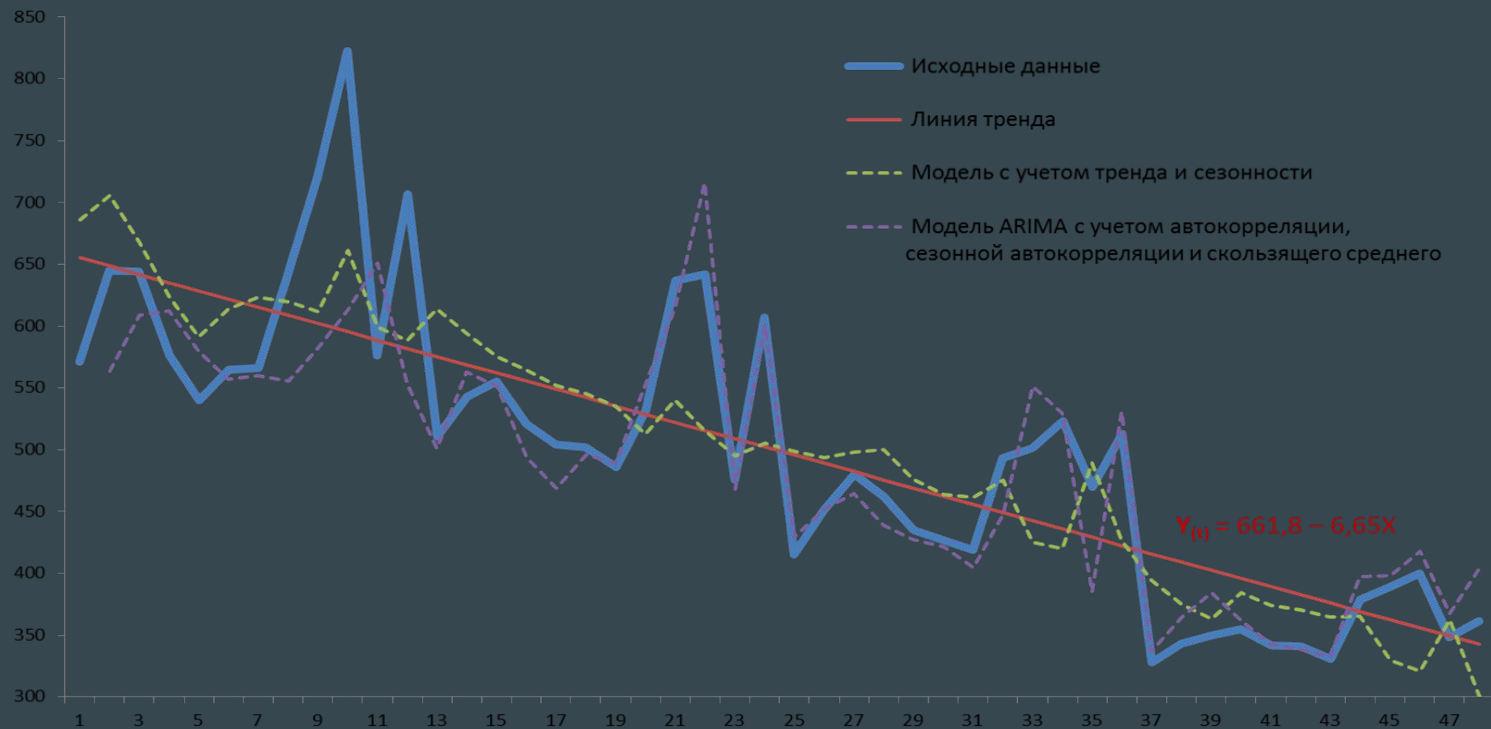
$Y(t) = a_1 Y(t-1) + a_2 Y(t-2) + e_t$

$dY(t) = Y(t) - Y(t-1)$

$dY(t) = (a_1 + a_2 - 1) Y(t-1) - a_2 dY(t-1) + e_t$

проверяем $a_1 + a_2 - 1 = 0$

ВСЕ ВМЕСТЕ



Модель ряда

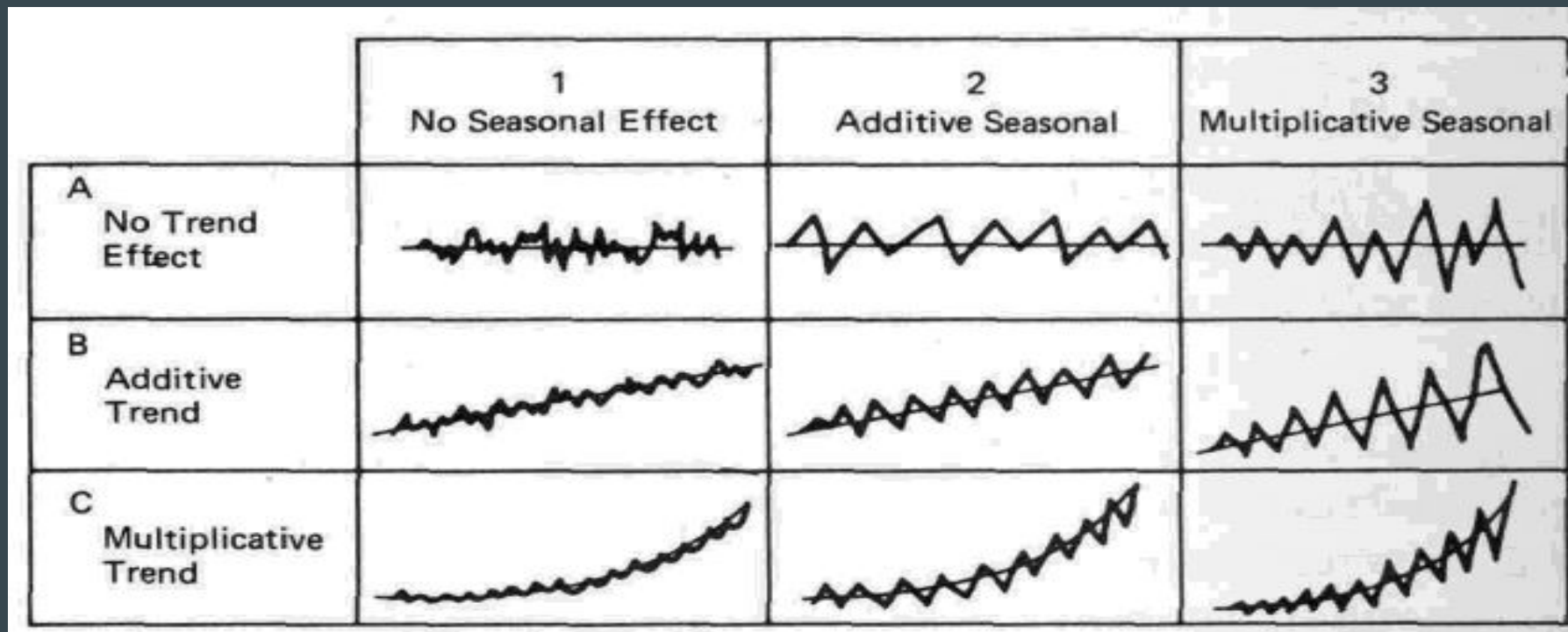
- аддитивная :

$$y(t) = X_{\text{trend}}(t) + X_{\text{cicle}}(t) + e(t)$$

- мультипликативная :

$$y(t) = X_{\text{trend}}(t) X_{\text{cicle}}(t) e(t)$$

Виды рядов



Сглаживание

Скользящие средние

$$x_NEW(i) = (x(i-n) + x(n-i+1) + \dots + x(i)) * 1/(n + 1).$$

n - порядок

Сглаживание

Экспоненциальное

$$x_NEW(i) = (x(i)*a + (1-a)*(x(n-i+1)* a + (1-a)*(...+ (1-a)*(x(i-n))))$$

$$x_NEW(i) = (x_NEW(i-1)*(1-a) + a*x(i))$$

a - порядок

