

Анализ временных рядов. Лекция 1

Костромина Алина
13.11.2025

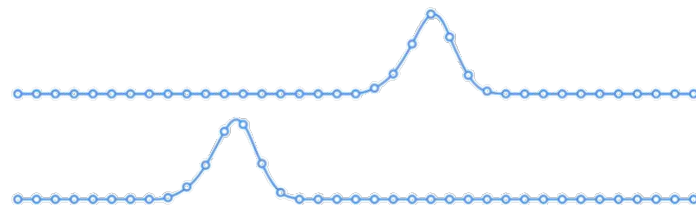
О чем сегодня будем разговаривать?

- Как устроены временные ряды?
- Как подойти к задаче прогноза?
 - Какие модели можно использовать?
 - Как оценить качество прогноза?
 - Как правильно выделить валидационную выборку?

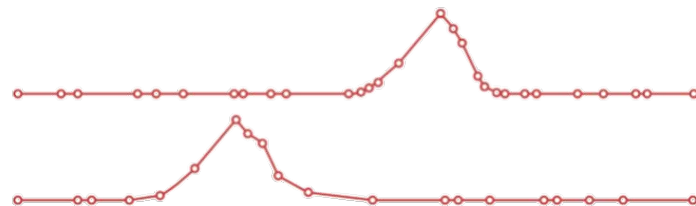
Прогнозирование временных рядов

Зафиксируем терминологию

- **Временной ряд** – последовательность некоторых значений по времени.
- Если рассматриваются одновременно несколько одиночных временных рядов, то такая структура называется **многомерный временной ряд**.
- **Регулярный временной ряд** – временной ряд с равномерными временными интервалами между точками.
- Дополнительные признаки (экзогенные переменные, ковариаты) — это внешние факторы, которые не генерируются самой системой, но оказывают влияние на целевые переменные.



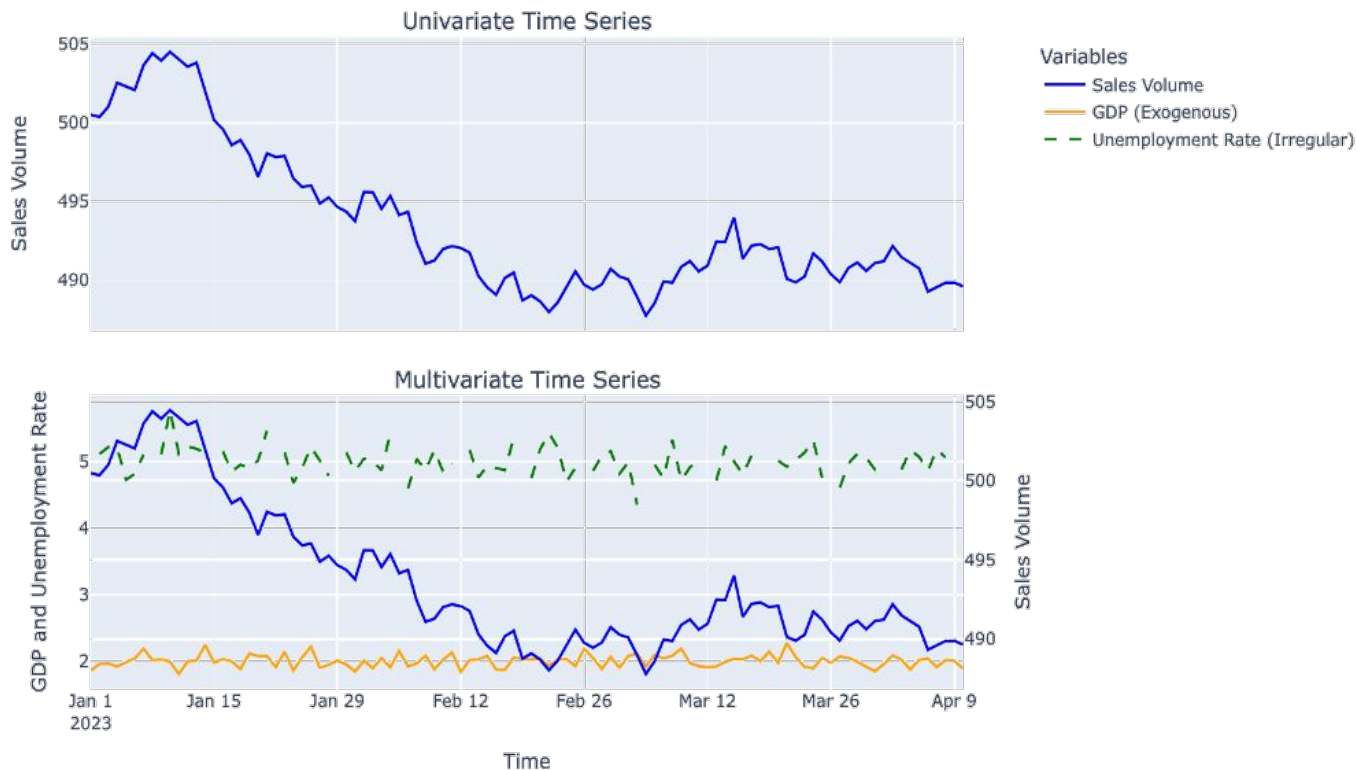
Пример регулярных событий



*Пример **нерегулярных** событий*

Прогнозирование временных рядов

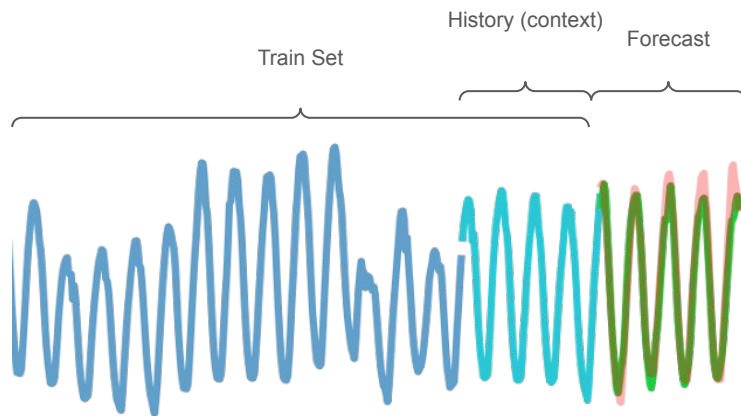
Зафиксируем терминологию



Прогнозирование временных рядов

Зафиксируем терминологию

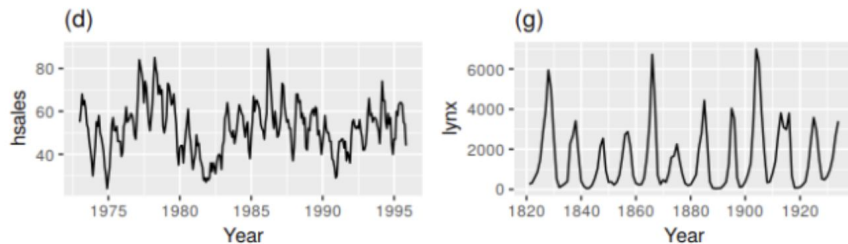
- **Задача прогнозирования** — по известной последовательности (контексту) предсказывать будущие значения на заданном горизонте прогнозирования.
- **Train_set \neq History (context)**



Компоненты временного ряда

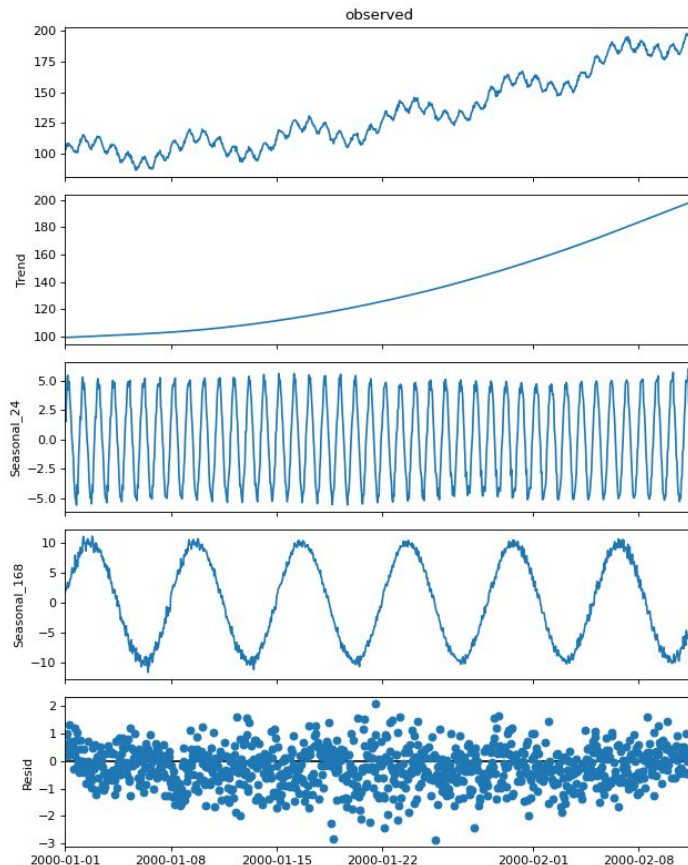
Как разделить сигнал на понятные составляющие?

- (M)STL разложение
 - Тренд
 - Сезонность (может быть несколько)
 - Остаток
- Еще отдельно выделяют **циклы** — аperiodические колебания



Сезонность vs цикл

Источник: <https://otexts.com/fpp2/stationarity.html>

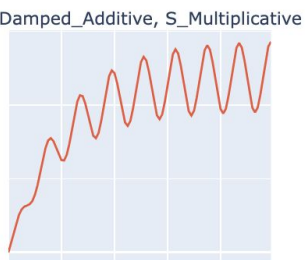
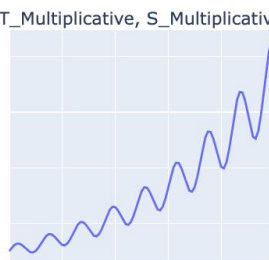
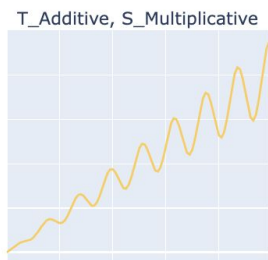
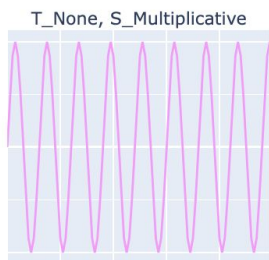
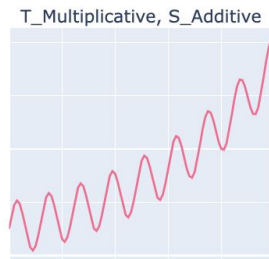
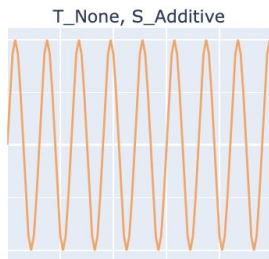
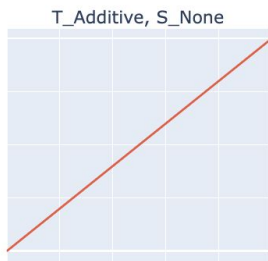
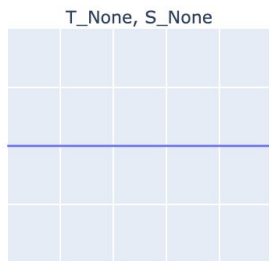


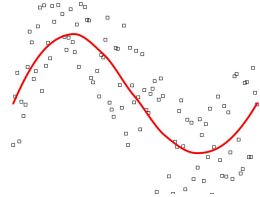
Пример *mstl* разложения

Источник:

<https://www.statmodels.org/dev/generated/statmodels.tsa.seasonal.MSTL.html>

Компоненты временного ряда: виды тренда и сезонности





Компоненты временного ряда

LOESS Smoothing

Каждую точку приближаем с помощью k соседей

$$Y_i = \mu(x_i) + \epsilon_i$$

Вес каждой точки внутри окна k задаем ядром h — дистанция до самого дальнего соседа

$$w_i(x) = W\left(\frac{x_i - x}{h}\right)$$

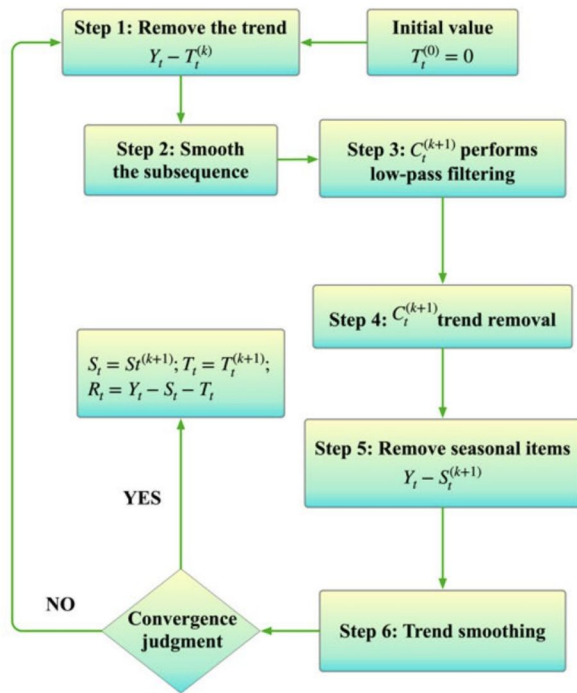
$$W(u) = (1 - |u|^3)^3 \text{ for } |u| < 1,$$

Решаем задачу локальной линейной регрессии

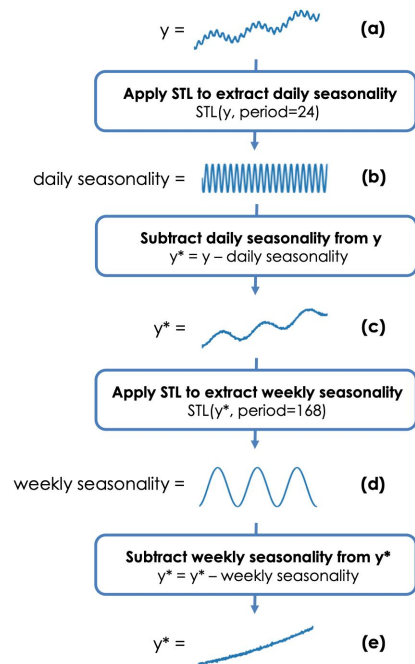
$$\mu(x_i) \approx \beta_0 + \beta_1(x_i - x) + \dots + \beta_p(x_i - x)^p$$

$$\sum_{i=1}^r w_i(x)(Y_i - \beta_0 - \beta_1(x_i - x) - \dots - \beta_p(x_i - x)^p)^2$$

STL



MSTL



Источники:

<https://www.nature.com/articles/s41598-023-31569-w>

<https://towardsdatascience.com/multi-seasonal-time-series-decomposition-using-mstl-in-python-136630e67530/>

<https://www.wessa.net/download/stl.pdf>

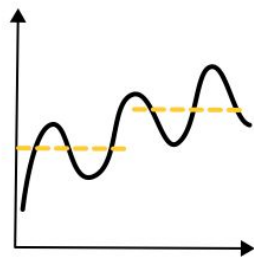
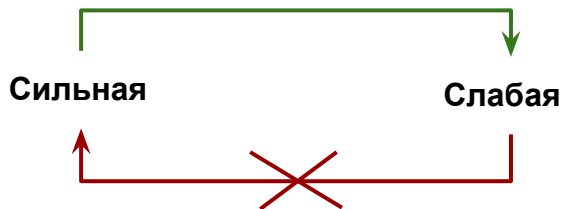
Стационарность

Постоянство свойств во времени

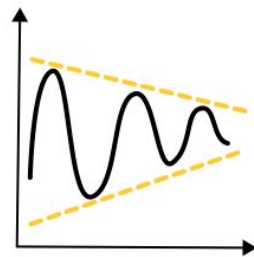
В узком смысле (сильная стационарность) — временной ряд $y_1 \dots, y_t$ стационарен, если для любого s совместное распределение $y_1 \dots, y_{t+s}$ не зависит от t .

В широком смысле (слабая стационарность) — временной ряд $y_1 \dots, y_t$ стационарен, если:

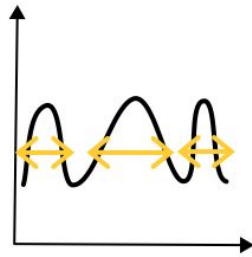
1. Существуют конечное матожидание $m(t)$ и дисперсия $v(t)$ в каждой точке $t \in T$.
2. $m(t) = \text{const.}$ $v(t) = \text{const.}$
3. Автоковариационная функция зависит только от разности аргументов — $\text{Cov}(y_t, y_{t+\tau}) = \gamma(\tau)$.



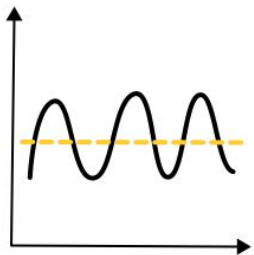
Mean dependent on time



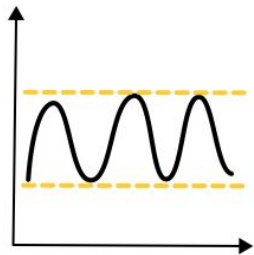
Variance dependent on time



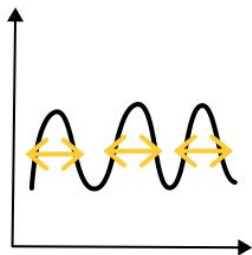
Covariance dependent on time



Mean independent on time



Variance independent on time



Covariance independent on time

Источник:

<https://medium.com/data-science/stationarity-in-time-series-a-comprehensive-guide-8beabe20d68>

Стационарность

Два наиболее популярных теста на стационарность.

Augmented Dickey-Fuller

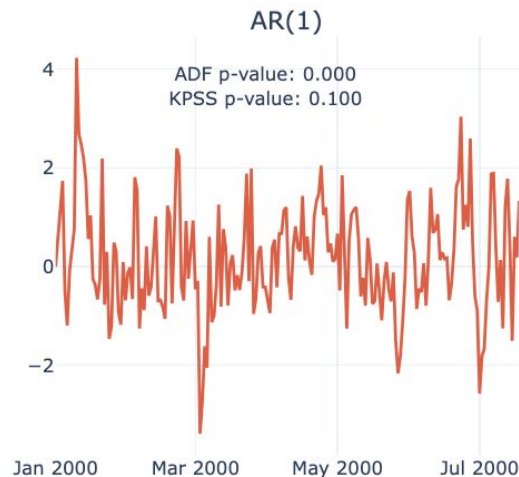
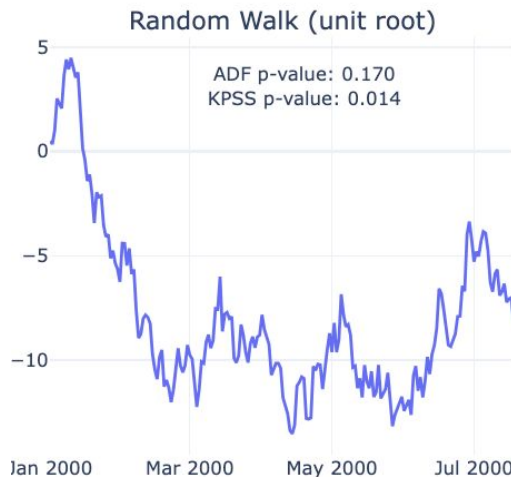
Рассмотрим модель временного ряда (есть расширенный вариант с учетом лаговых разностей):

$$Z_t = \theta_0 + \theta_1 Z_{t-1} + \varepsilon_t$$

$H_0: \theta_1 = 1$ (ряд имеет единичный корень).

KPSS

H_0 : Ряд стационарен (по уровню или тренду).



Примеры временных рядов

Какие ряды стационарны? Какие компоненты в них можно выделить?

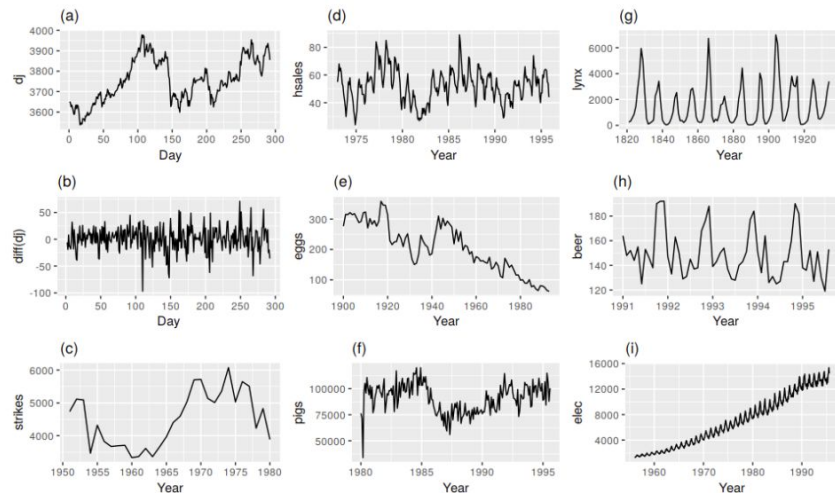


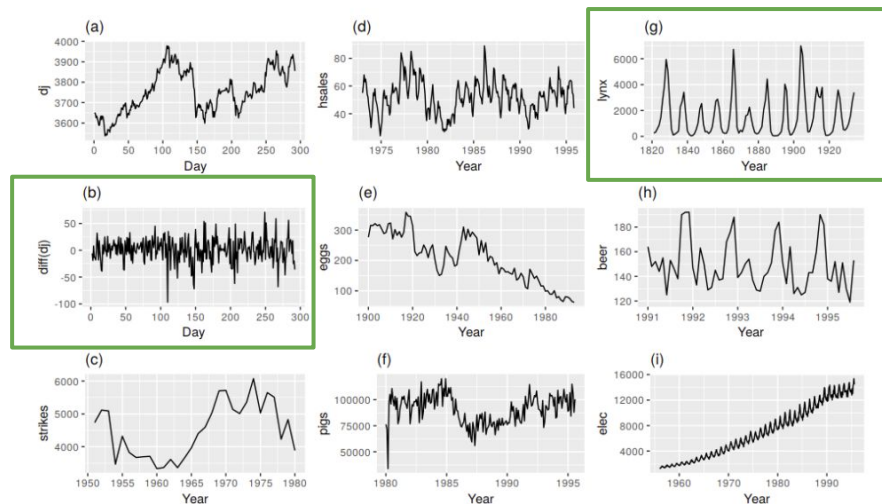
Figure 8.1: Which of these series are stationary? (a) Dow Jones index on 292 consecutive days; (b) Daily change in the Dow Jones index on 292 consecutive days; (c) Annual number of strikes in the US; (d) Monthly sales of new one-family houses sold in the US; (e) Annual price of a dozen eggs in the US (constant dollars); (f) Monthly total of pigs slaughtered in Victoria, Australia; (g) Annual total of lynx trapped in the McKenzie River district of north-west Canada; (h) Monthly Australian beer production; (i) Monthly Australian electricity production.

Источник:

<https://otexts.com/fpp2/stationarity.html>

Примеры временных рядов

Какие ряды стационарны? Какие компоненты в них можно выделить?



- **Тренд:** (a), (c), (e), (f) и (i)
- **Сезонность:** (d), (h) и (i)
- **Изменение дисперсии:** (i)

(b) и (g) стационарны!

Циклы аperiodичны — их наступление непредсказуемо, поэтому их наличие не делает ряд нестационарным.

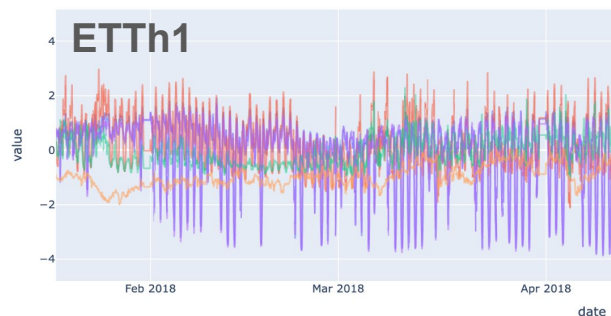
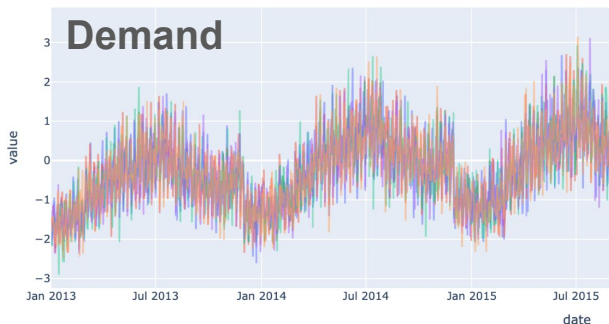
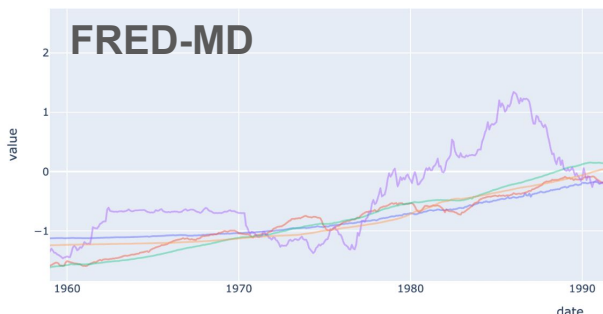
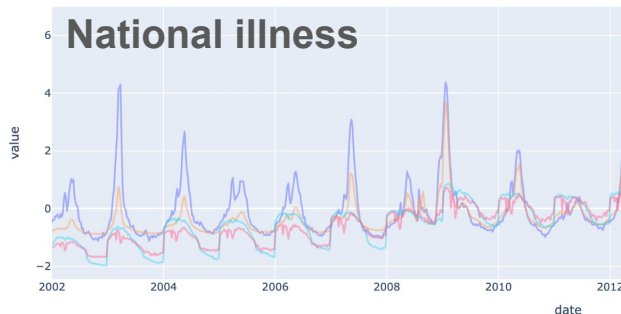
Figure 8.1: Which of these series are stationary? (a) Dow Jones index on 292 consecutive days; (b) Daily change in the Dow Jones index on 292 consecutive days; (c) Annual number of strikes in the US; (d) Monthly sales of new one-family houses sold in the US; (e) Annual price of a dozen eggs in the US (constant dollars); (f) Monthly total of pigs slaughtered in Victoria, Australia; (g) Annual total of lynx trapped in the McKenzie River district of north-west Canada; (h) Monthly Australian beer production; (i) Monthly Australian electricity production.

Источник:

<https://f0nzie.github.io/hyndman-bookdown-rsuite/stationarity-and-differencing.html>

Примеры временных рядов

Какие ряды стационарны? Какие компоненты в них можно выделить?



Нужна ли предобработка?

Универсально:

- **Логарифмирование / Вох-Сох / Yeo-Johnson** — стабилизируем дисперсию и мультипликативные эффекты (не всегда удастся подобрать хорошие параметры).
- **Сезонное дифференцирование** — удаляем сезонность.
- **Дифференцирование** — удаляем тренд и переходим к относительным изменениям.

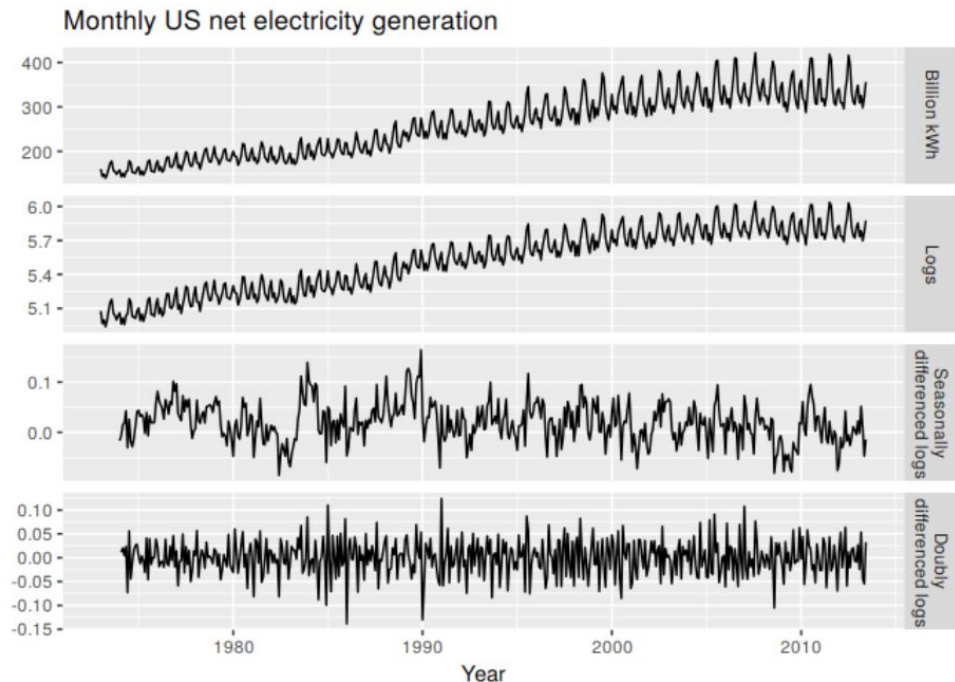
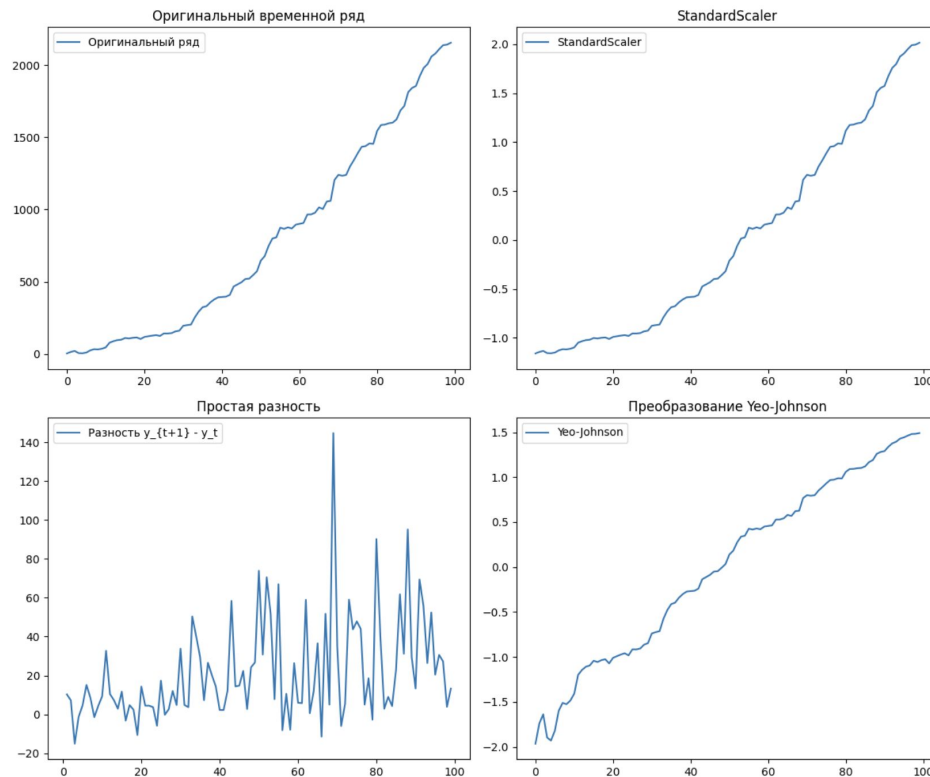


Figure 8.4: Top panel: US net electricity generation (billion kWh). Other panels show the same data after transforming and differencing.

Нужна ли предобработка?

Больше актуально для ML и DL:

- **StandardScaler / MinMaxScaler / RobustScaler** — контролируем диапазон значений.
- **WindowNormalizing** — боремся с артефактами на уровне окон.



Как получить прогнозы?

Группы методов, которые используются на практике



Naive methods

- (Seasonal) Naive
- Mean, Median



Statistical methods

- ETS
- Theta
- ARIMA



ML methods

- LinearRegression
- GradientBoosting



DL methods

- DLinear
- NBEATS
- PatchTST
- GPT4TS

Как не запутаться в `pd.Series.shift(...)`?

Временной ряд можно представить в виде таблицы признаков и использовать знакомые ML и DL методы вместо специализированных моделей

«Длинная» таблица с
временным рядом

STEP [
 L
 H]

time	value	Additional features
0	4	...
1	4	...
2	5	...
3	6	...
4	3	...
5	2	...
6	8	...

H = 2
L = 3
STEP = 1



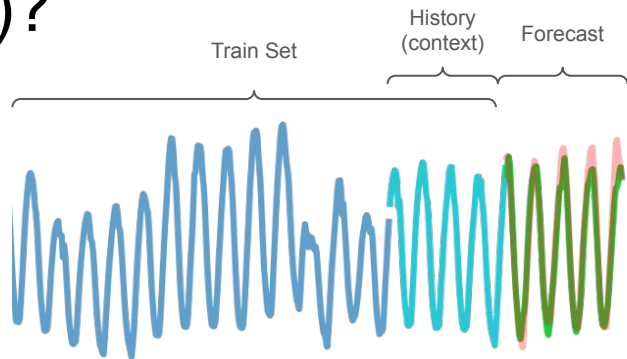
Составляем индексы

id	long ids features	long ids targets
0	[0, 1, 2]	[3, 4]
1	[1, 2, 3]	[4, 5]
2	[2, 3, 4]	[5, 6]



«Широкая» таблица со
сгенерированными признаками

Base ts	Current Value	Max Value	target1	target2
2	5	5	6	3
3	6	6	3	2
4	3	6	2	8



Как оценить качество прогноза?

Существует множество вариантов метрик, но нет ни одной универсальной

- Лучше использовать сразу несколько.

1 Scale-Dependent

MSE (Mean Squared Error)

$$MSE = \frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2$$

MAE (Mean Absolute Error)

$$MAE = \frac{1}{n} \sum_{t=1}^n |y_t - \hat{y}_t|$$

MdAE (Median Absolute Error)

$$MdAE = \text{median}(|y_t - \hat{y}_t|)$$

2 Scale-Independent

MAPE (Mean Absolute Percentage Error)

$$MAPE = 100 \times \frac{1}{n} \sum_{t=1}^n \left| \frac{y_t - \hat{y}_t}{y_t} \right|$$

SMAPE (Symmetric Mean Absolute Percentage Error)

$$SMAPE = 100 \times \frac{1}{n} \sum_{t=1}^n \frac{|y_t - \hat{y}_t|}{\frac{|y_t| + |\hat{y}_t|}{2}}$$

msMAPE (Modified SMAPE)

$$msMAPE = 100 \times \frac{1}{n} \sum_{t=1}^n \frac{|y_t - \hat{y}_t|}{\frac{\max(|y_t| + |\hat{y}_t| + \epsilon, 0.5 + \epsilon)}{2}}$$

3 Relative

RelRMSE (Relative RMSE)

$$RelRMSE = \sqrt{\frac{MSE}{MSE_{baseline}}}$$

RelMAE (Relative MAE)

$$RelMAE = \frac{MAE}{MAE_{baseline}}$$

MASE (Mean Absolute Scaled Error)

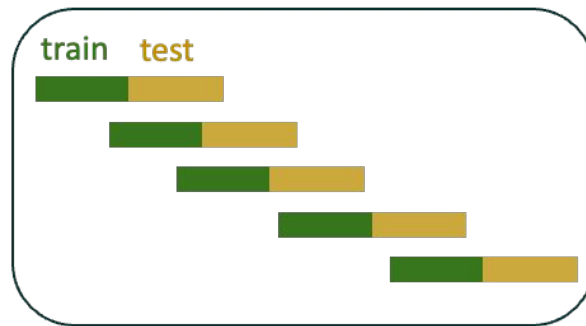
$$MASE = \frac{MAE}{\frac{n}{m-s} \sum_{k=s+1}^m |y_k - y_{k-s}|}$$

Как оценить качество прогноза?

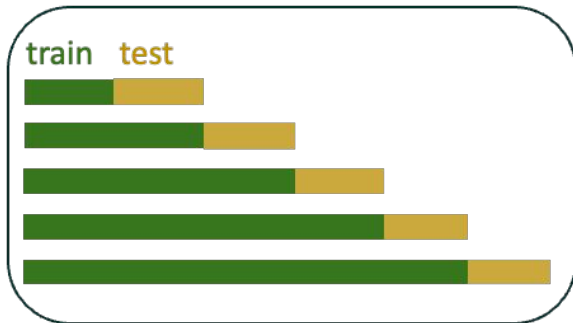
1 Blocked KFold



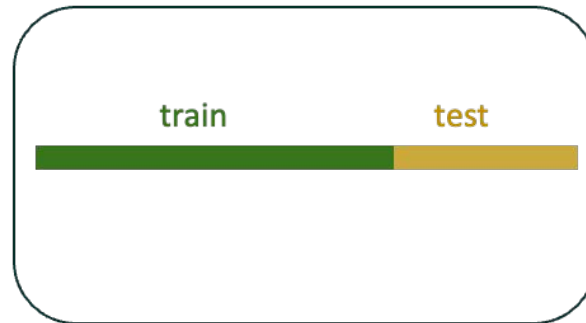
2 Rolling window



3 Expanding window



4 Holdout



Куда обращаться?

- <https://otexts.com/fpp3/some-simple-statistics.html> — ***Forecasting: Principles and Practice (3rd ed) (Rob J Hyndman and George Athanasopoulos)***: несложный учебник, к которому можно обращаться как к входной точке, если возникла проблема
- https://www.youtube.com/watch?v=S_Z8RnTE5dl&t=5463s — 2.5-часовая ***лекция Christoph Bergmeir*** про прогнозирование временных рядов
- <https://arxiv.org/abs/2203.10716> — Forecast Evaluation for Data Scientists: Common Pitfalls and Best Practices