

Autocorrelation and Stationarity



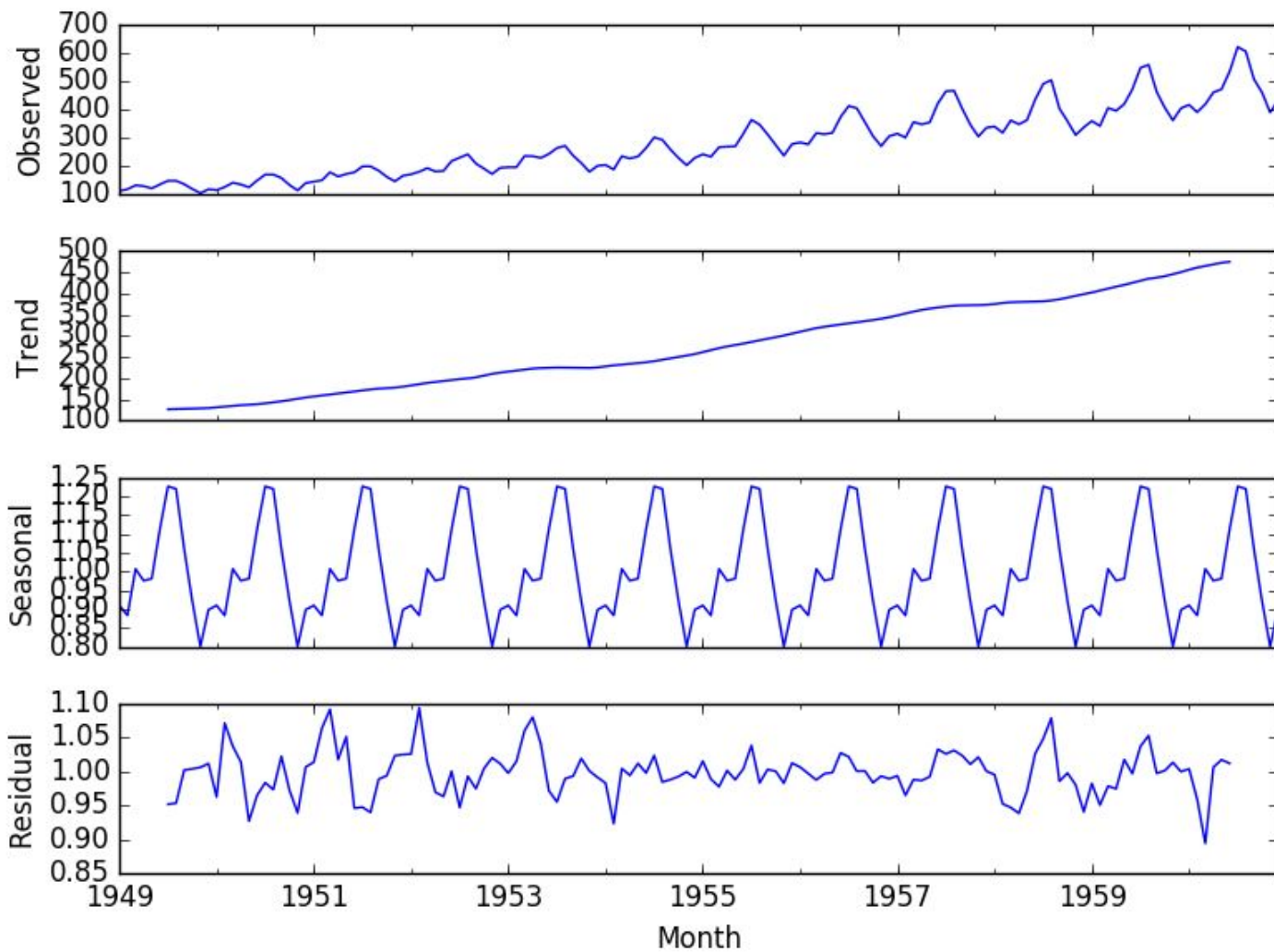
Week 10 - Day 03

Summary

Time series = time + values

Same time interval

(weeks, days, seconds)



!!!

UNIVARIATE TIME SERIES

!!!

VERY LIMITED
(developed in the 70')

(RF = 1995)

**What's the lesson
about?**

Autocorrelation + Stationarity

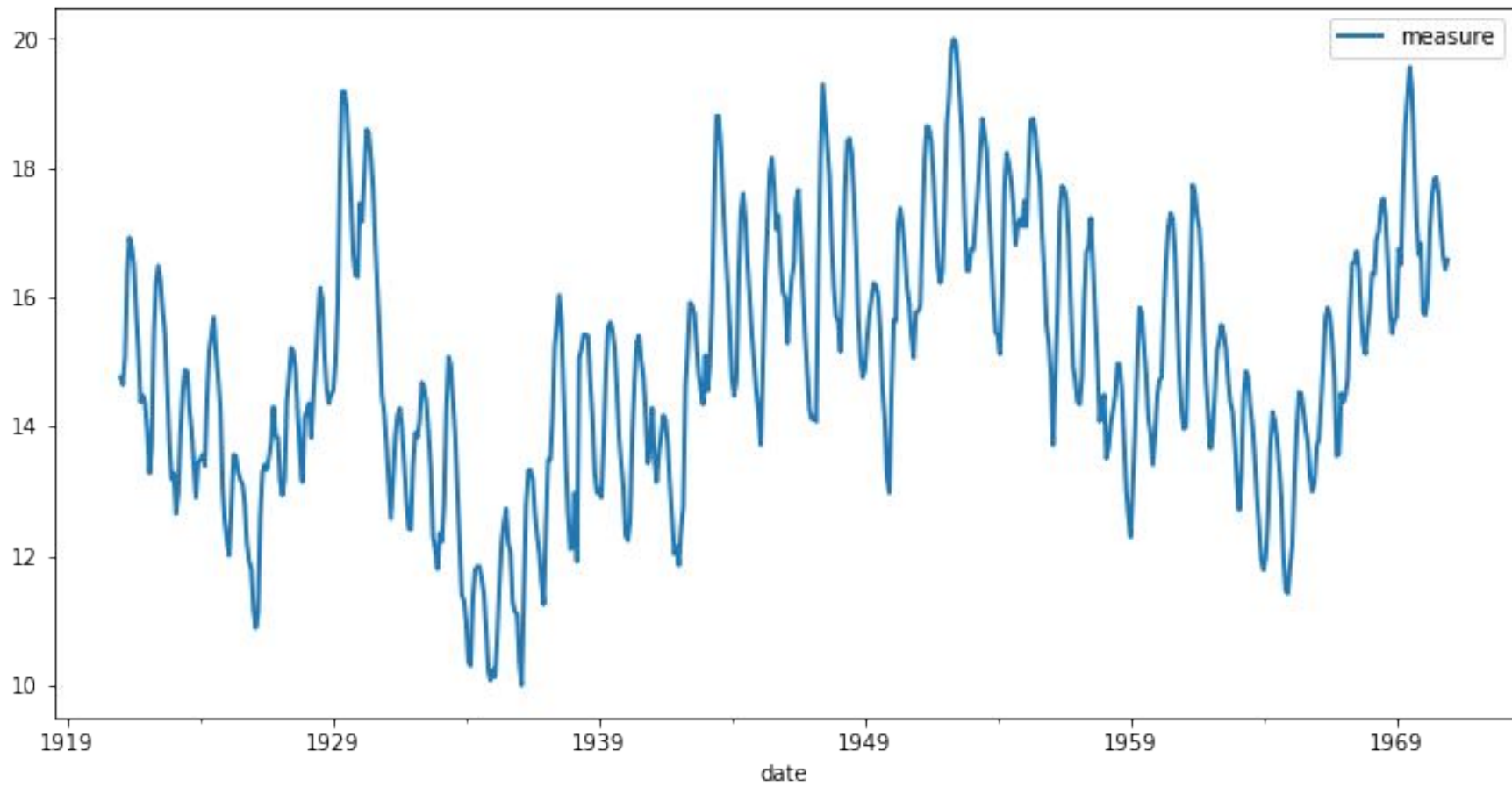
Properties of a TS

Used during EDA and modelling

Autocorrelation

Example

lake levels (meters) x 50 years

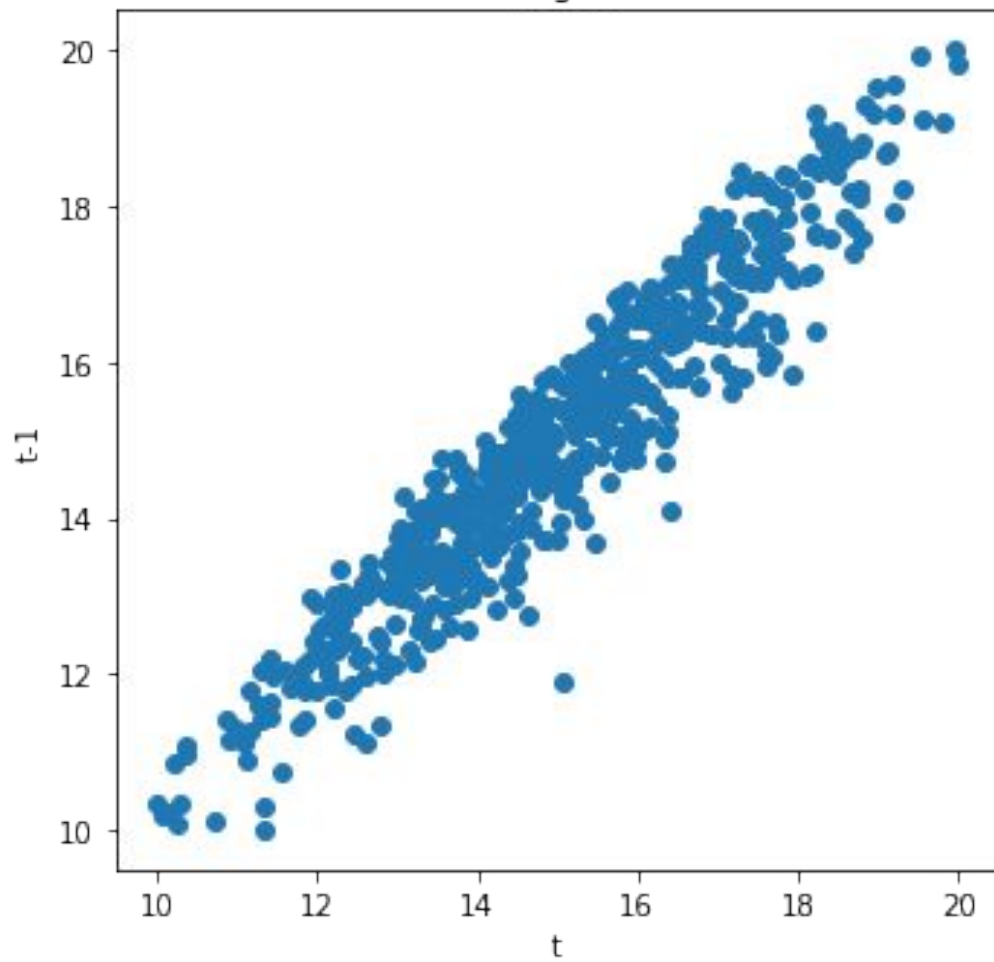


How can you easily
predict the next value?

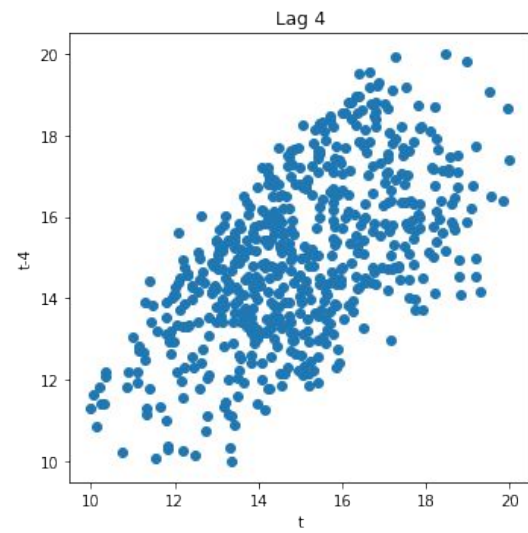
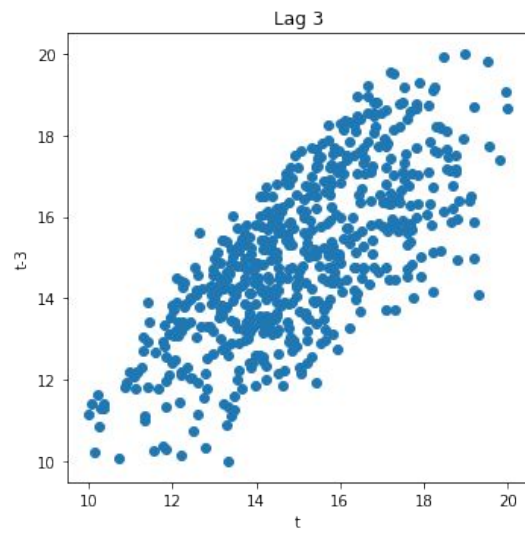
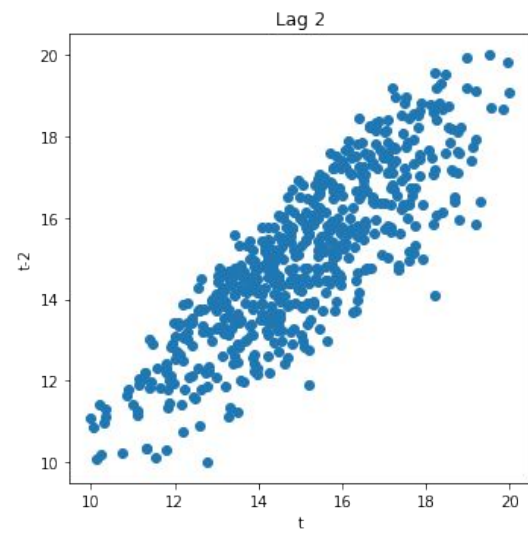
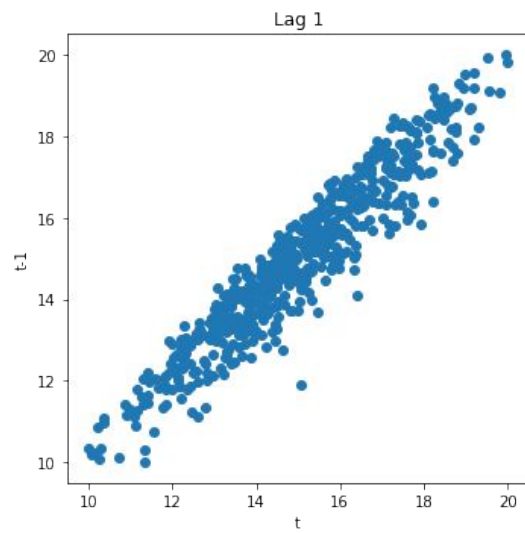
Next value = previous value!

Is the current measure related to
previous measure?

Lag 1



Is the current value related to
previous previous values?



Autocorrelation!

Shift \rightarrow correlation

Shift shift \rightarrow correlation

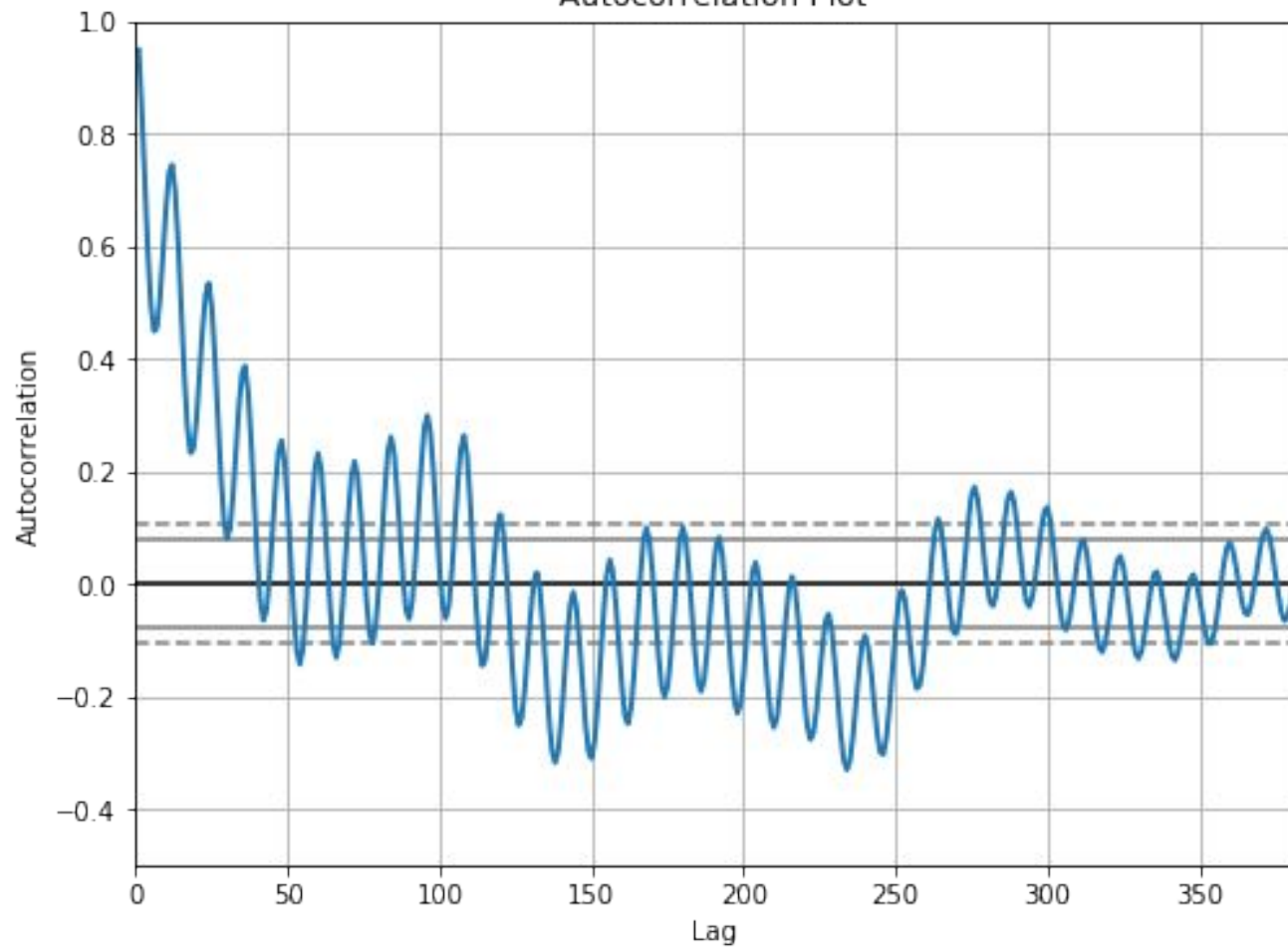
Shift shift shift \rightarrow correlation

Shift ... shift \rightarrow correlation

```
df[ 'measure' ].autocorr(lag=1)
```

```
plot_acf(ts)  
(statsmodels)
```


Autocorrelation Plot



Confidence bars!

Useful to choose
the parameters of our models

Part of the EDA

*“Cool, there is a correlation with the previous
3 weeks and with the 52nd week (seasonality)!”*

Stationarity

Characteristic of a TS

Stationarity

=

characteristics of a TS

don't change with time

Stationarity

=

characteristics of a TS

~~don't change with time~~

are time-invariant

Stationarity

=

same mean + same variance

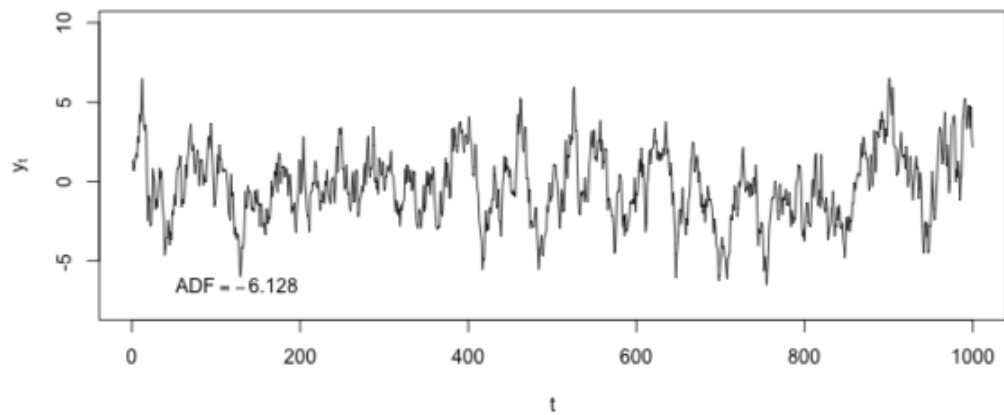
Stationarity

=

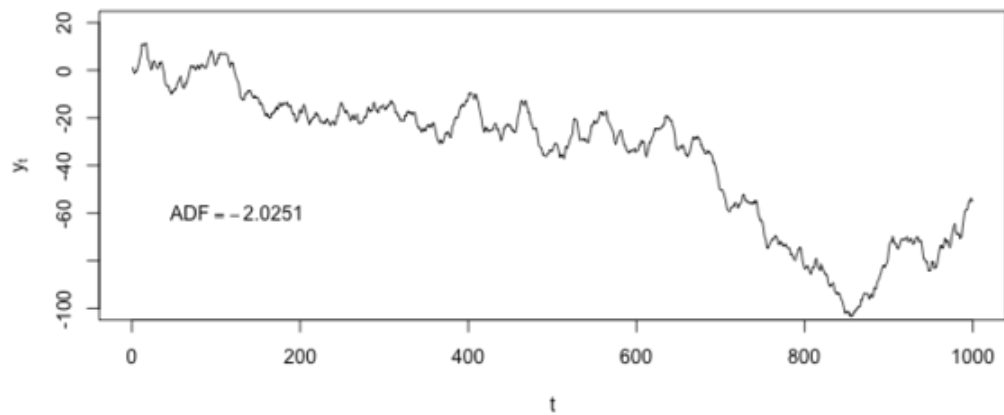
same mean + same variance

Stationarity = “no trends”!

Stationary Time Series



Non-stationary Time Series



Augmented Dickey-Fuller (ADF) test

Really important to check!

(many models requires stationarity)

Stationarity -> no dependence on time ->
nice statistical properties

How to treat non-stationary TSs

We need to get stationarity!

We have a statistical test!

(with p-value)

Differencing

Remove seasonality

Log

etc.

Differencing

Current val = current val - previous val

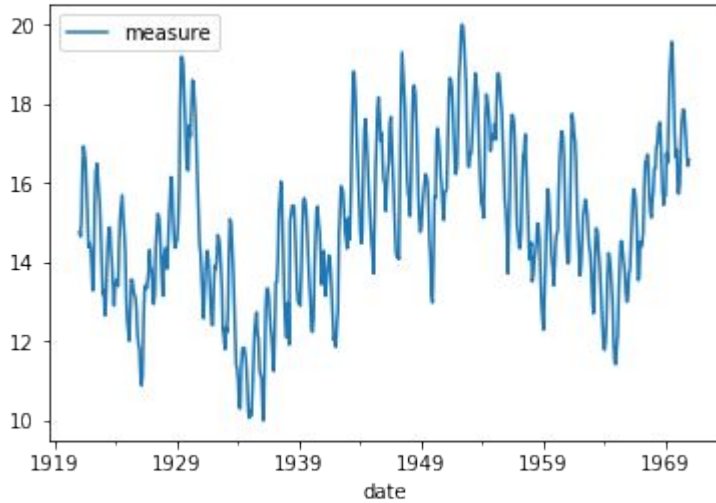
Increase/decrease wrt previous value

Time	Original	Diff
w1	10	NA
w2	12	2
w3	7	-5
w4	9	2
w5	15	6

The diagram illustrates the calculation of differences between consecutive 'Original' values. Arrows point from the 'Original' column to the 'Diff' column:

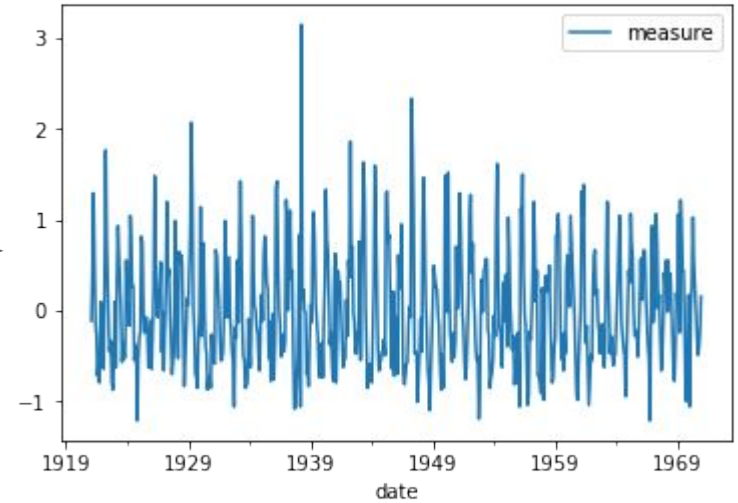
- From 10 (w1) to 2 (w2)
- From 12 (w2) to 2 (w2)
- From 9 (w4) to 6 (w5)
- From 15 (w5) to 6 (w5)

Original



Non stationary

Diff



Stationary

Summary

- Autocorrelation = correlation + lag/shift
- Autocorrelation plot (statmodels)
- Useful for EDA and to choose parameters

- Stationarity = time invariant characteristics
- Stationarity = often needed for modelling
- No Stationarity → transformations (diff)
- Statistical test (p-value)