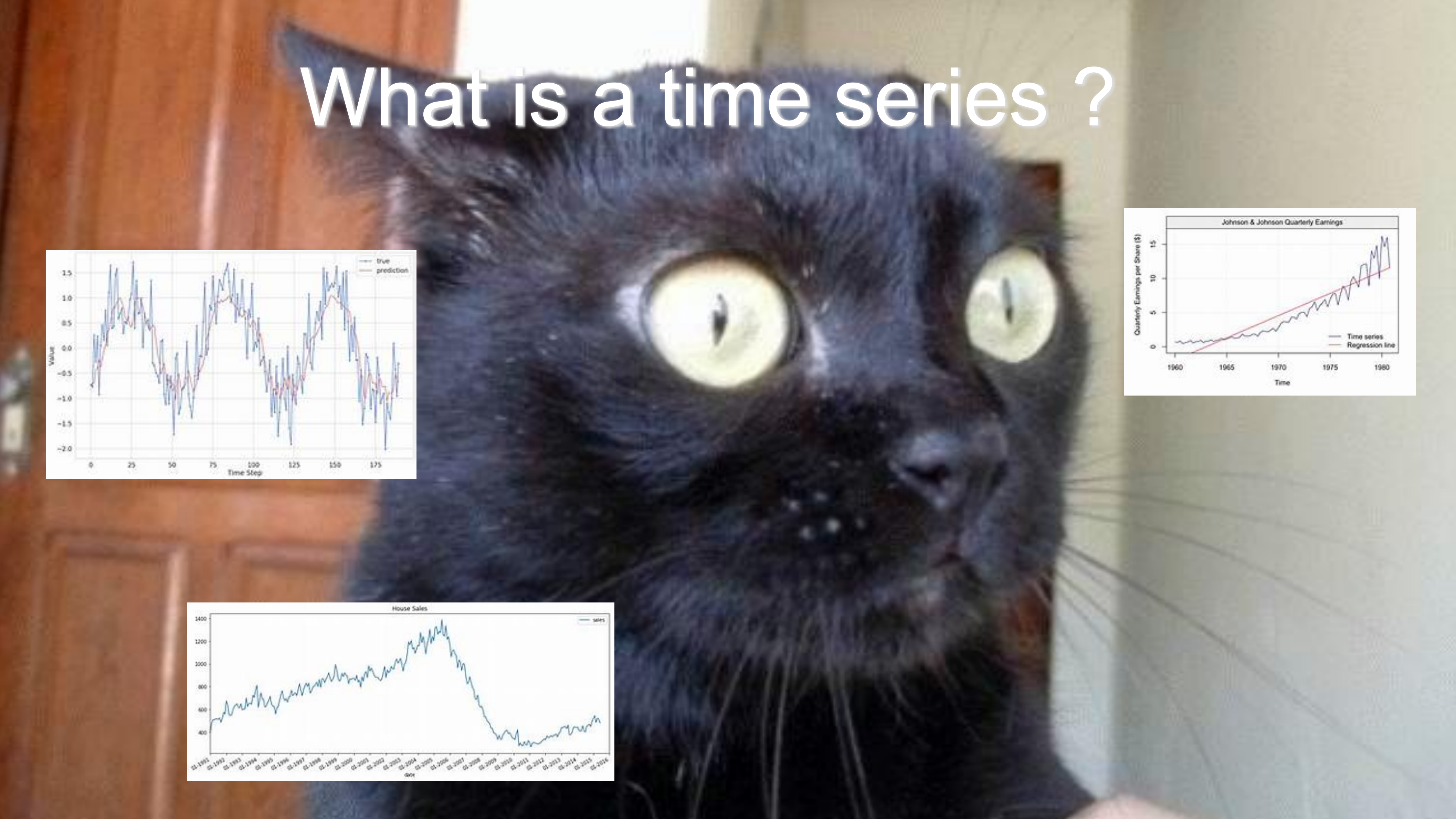
UNIVERSITÉ DE
NEUCHÂTEL

Time series analysis

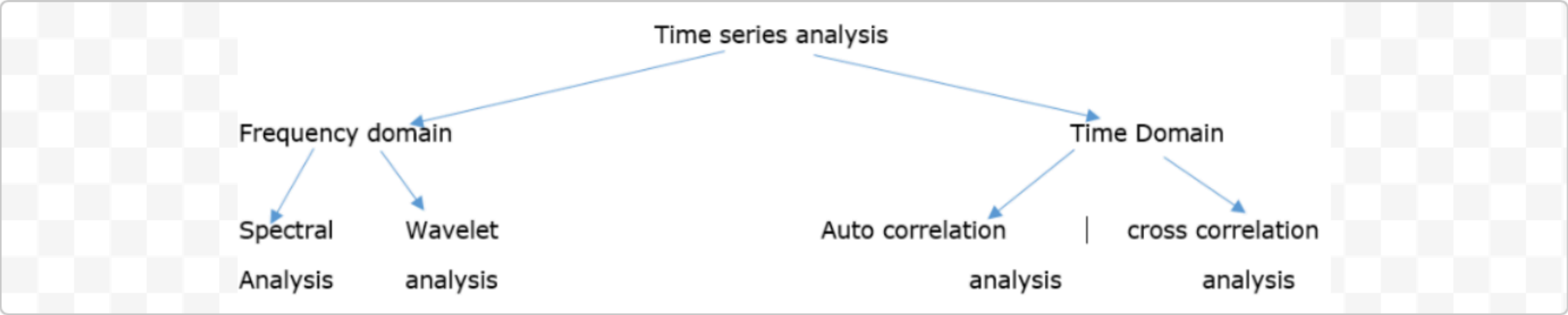
Master of Cognitive Science

Steven Moran & Marco Maiolini

Nov 23, 2022


What is a time series ?
true
prediction
Value
Time Step
Johnson & Johnson Quarterly Earnings
Quarterly Earnings per Share ($)
Time series
Regression line
Time
House Sales
sales
date

Methods for time series analysis



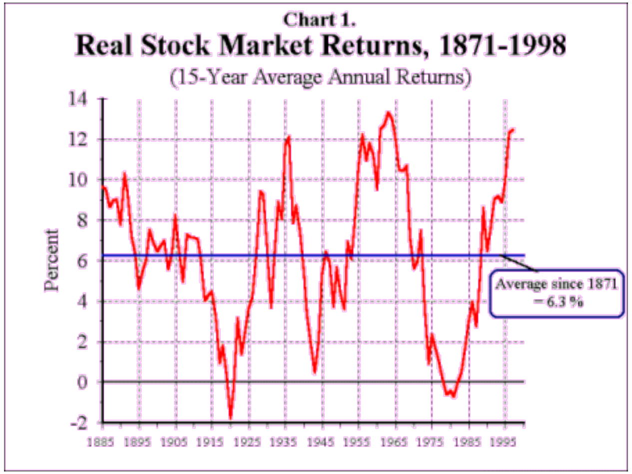
https://datascienceplus.com/time-series-analysis-using-arima-model-in-r/

Time series

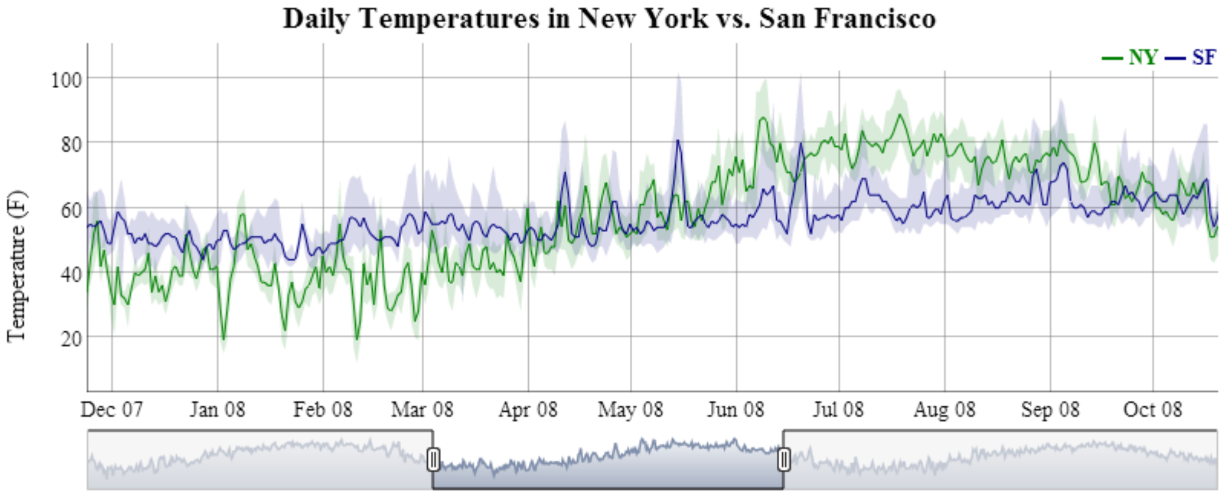
- “Collection of observation made sequentially through time”
- Mathematical model to explain serial correlation in a time series data set
- Time series analysis provides a robust statistical framework for analysis

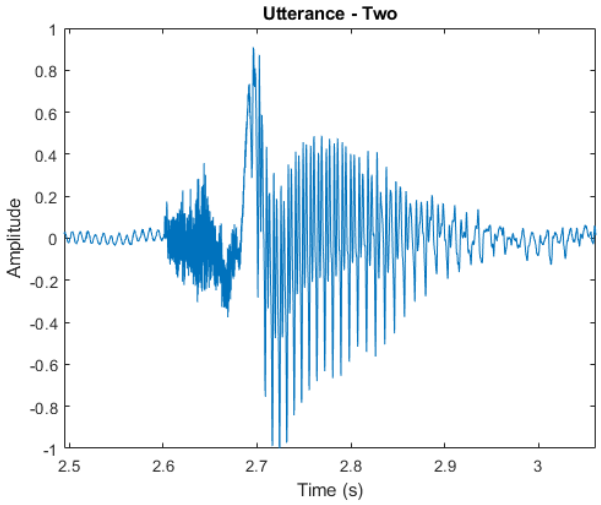
Chris Chatfield. 2019. *The analysis of time series and introduction with R.* Chapman / Hall.

Economics & Finance

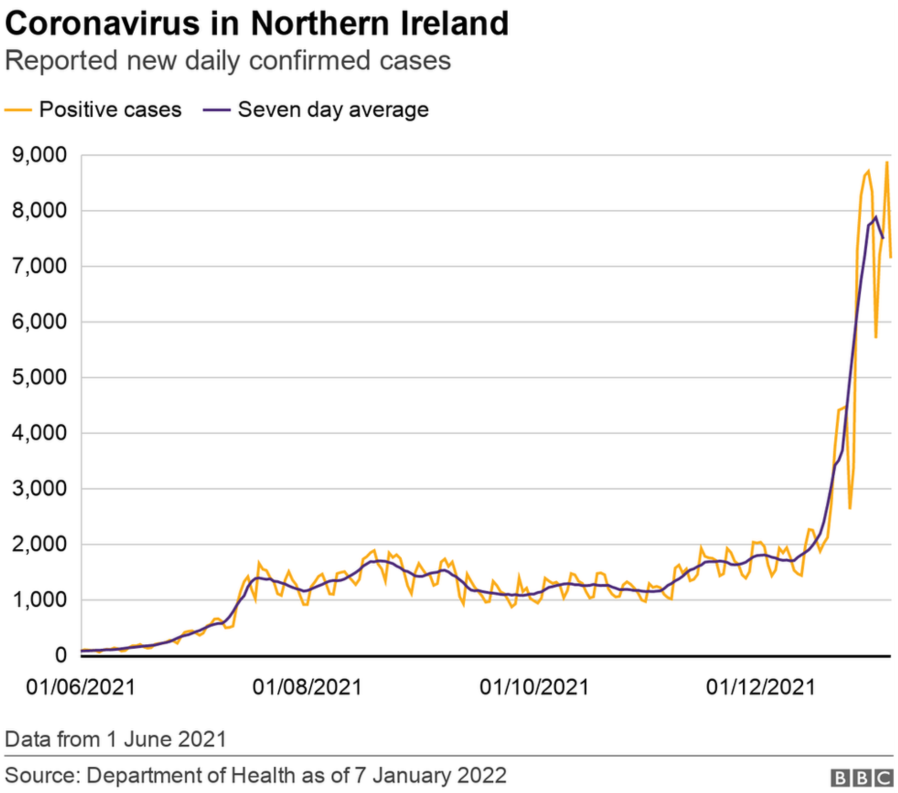


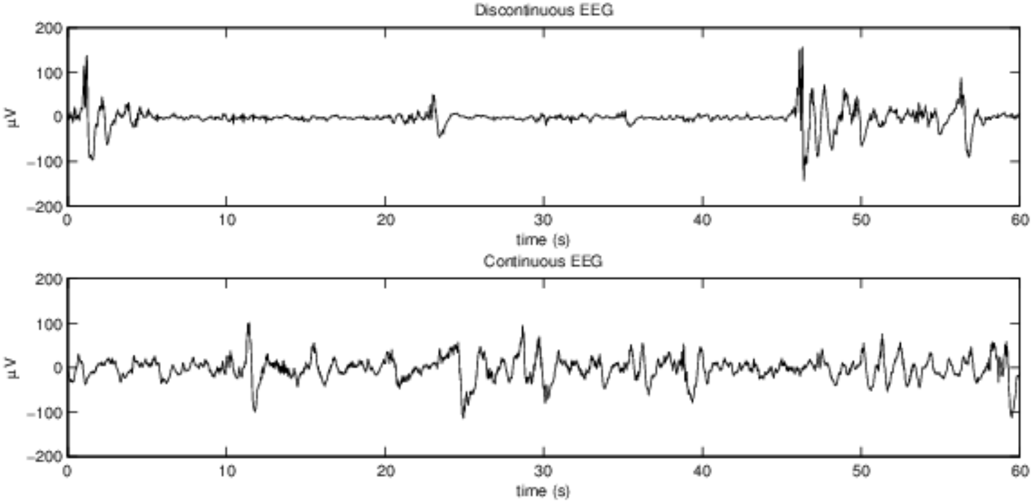
Physical observations





Biological observations





Types of time series

- **Discrete**: Data taken only in specific fixed time
- **Continuous**: Data taken continuously through time
- **Deterministic**: Predicted exactly by previous values
- **Stochastic**: Only partially predicted by past values

Why use time series analysis?

- **Identify** systematic patterns or causes of trends through time
- Time series forecasting to **predict** future events
- **Clean** your data and impute missing values (fill the gaps)

Example of time series

- **Weather** data, e.g. rainfall, temperature

- **Health** data, e.g., heart rate (EKG) or brain (EEG) monitoring

- **Forecasting**, e.g., stocks, business sales, interest rates, trading, gambling

So what can we do?

Describe

So what can we do?

Describe

Explain

So what can we do?

Describe

Explain

Predict

So what can we do?

Describe

Explain

Predict

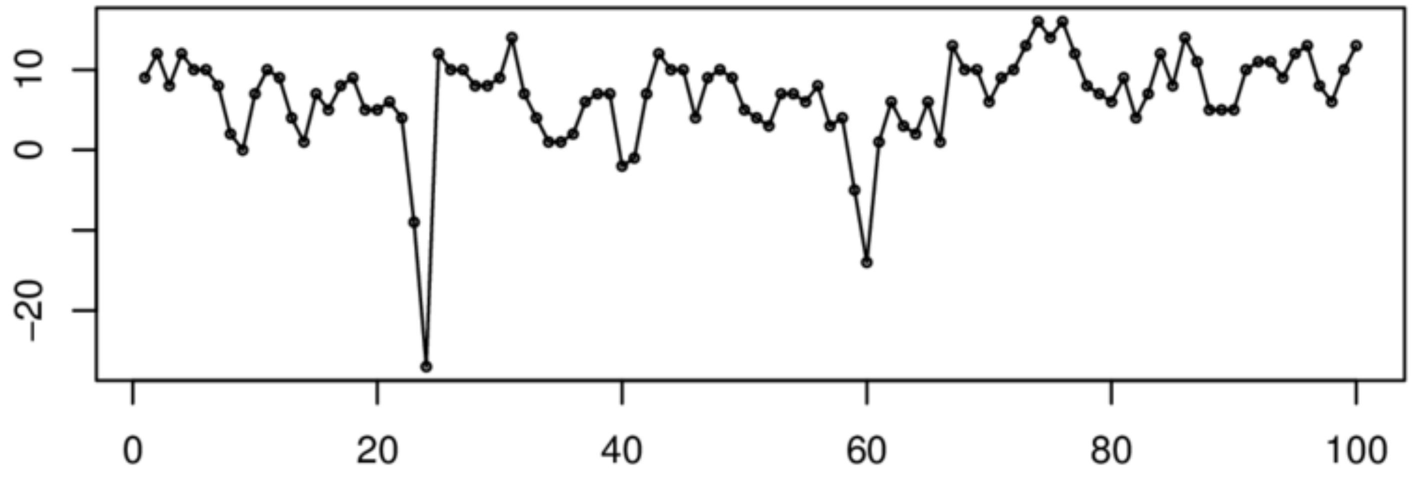
Control

Describe a time series

- **Mean**: The central value of a finite set of numbers
- **Frequency**: Sampling frequency
- **Variance**: Expectation of the squared deviation of a random variable from its mean
- **Autocorrelation**: Correlation among neighboring observations

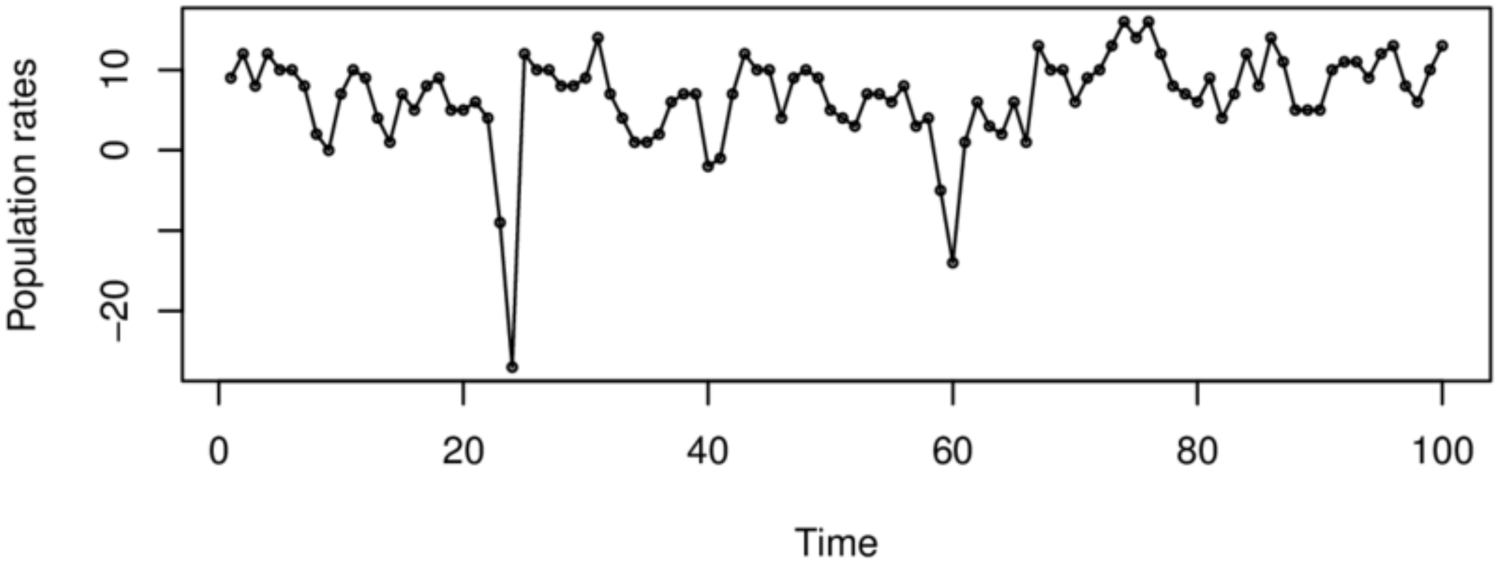
Describe a time series

Population growth rate:



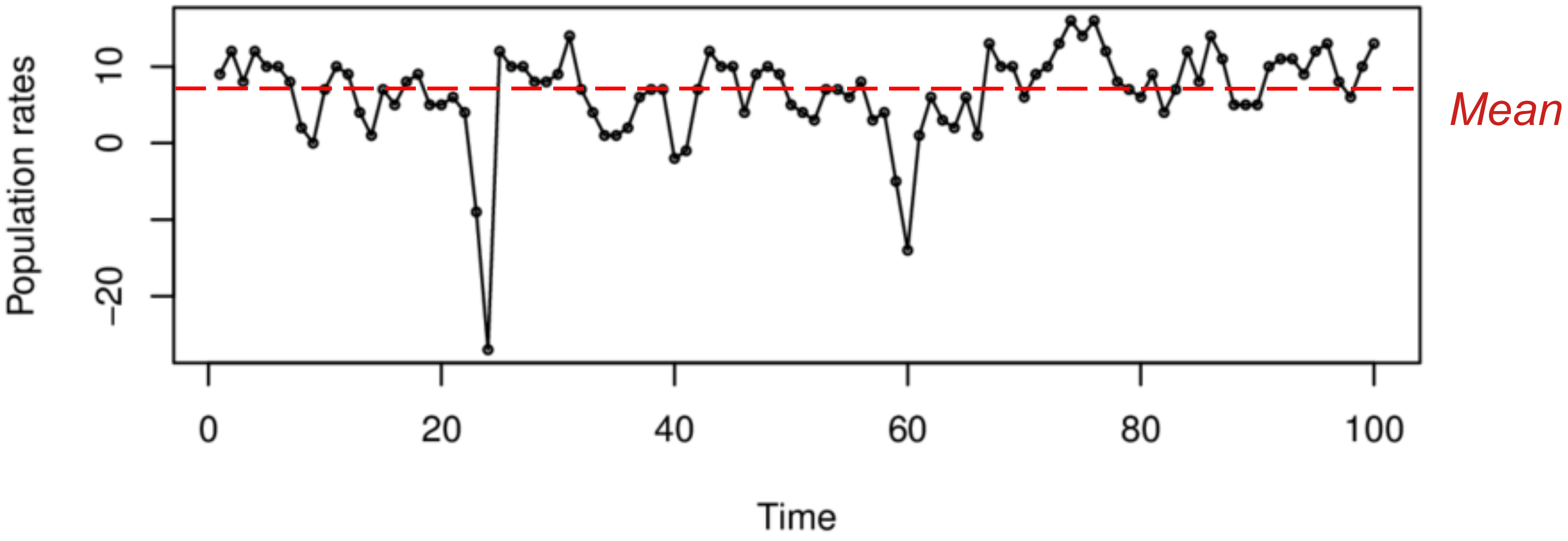
Describe a time series

Population growth rate:



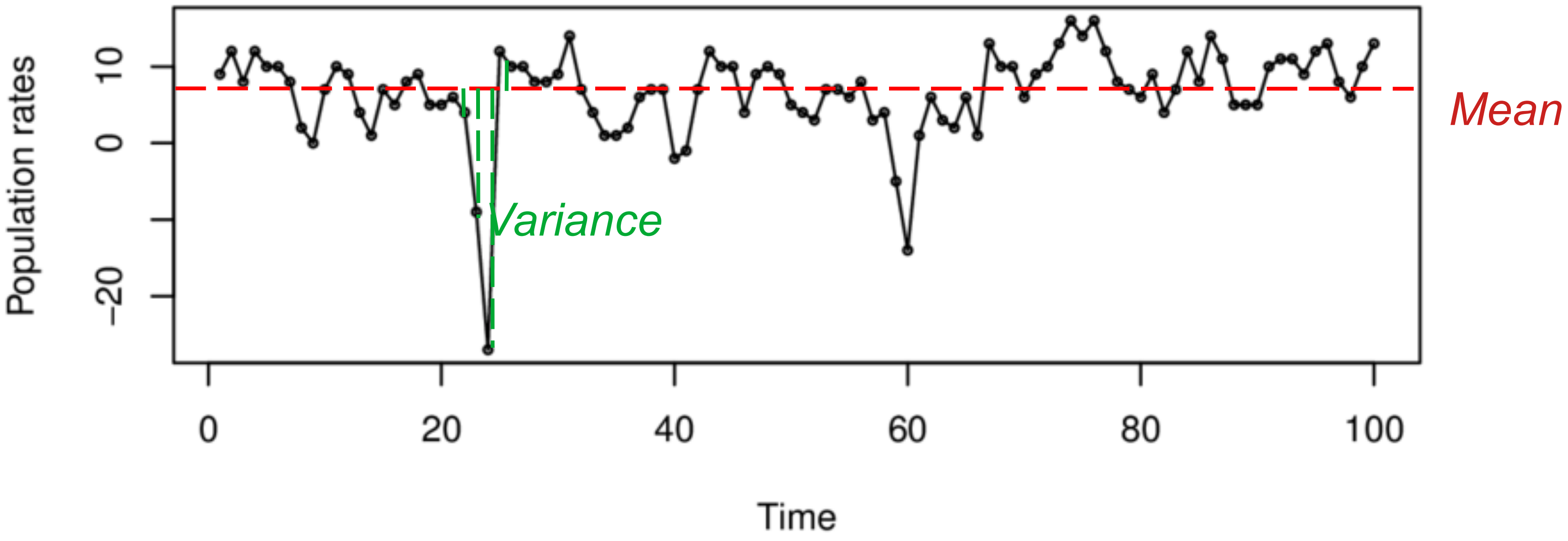
Describe a time series

Population growth rate:



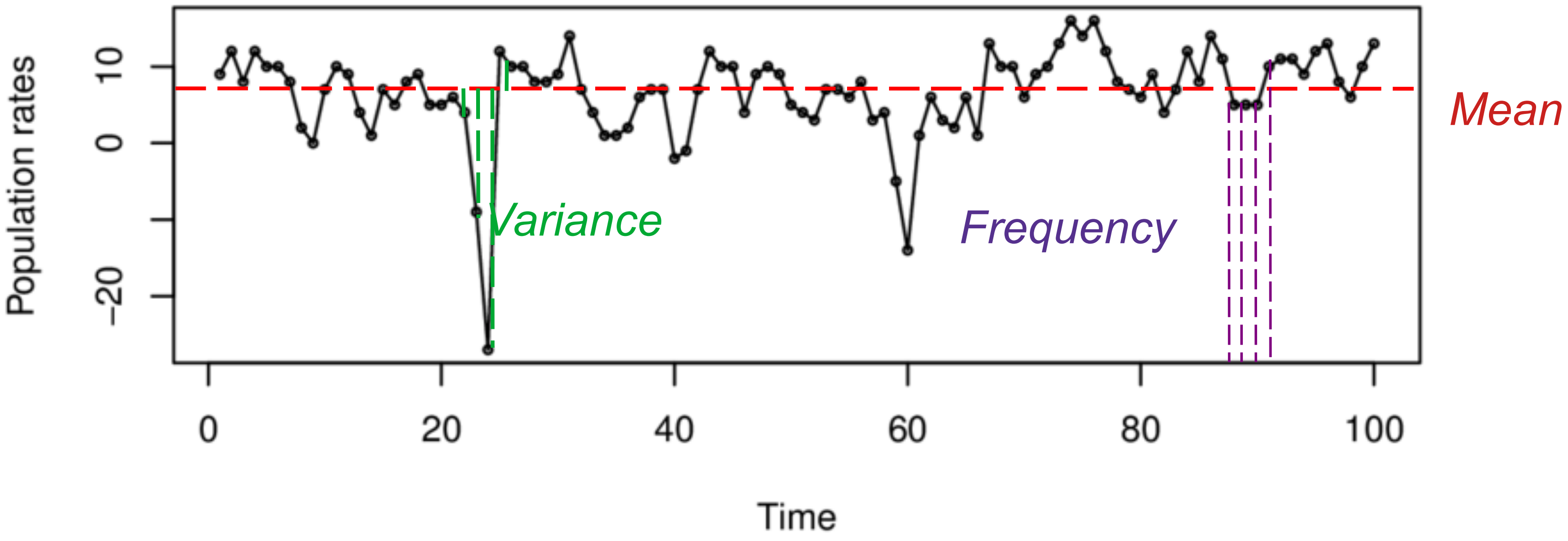
Describe a time series

Population growth rate:



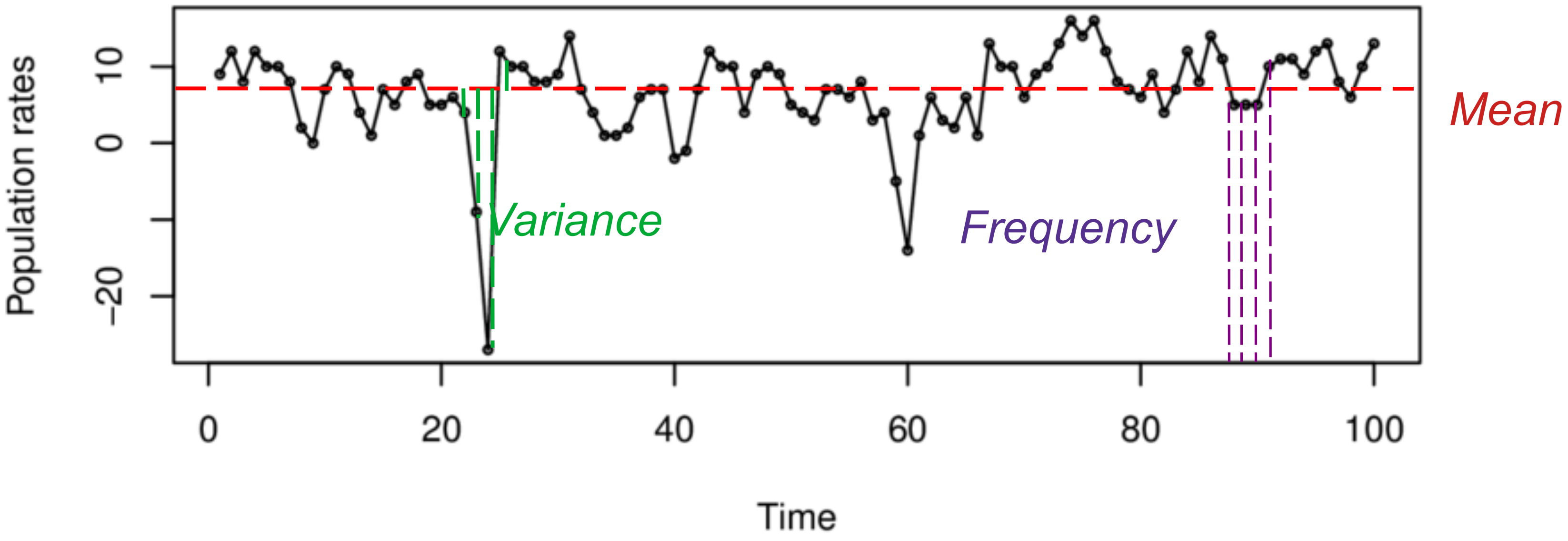
Describe a time series

Population growth rate:



Describe a time series

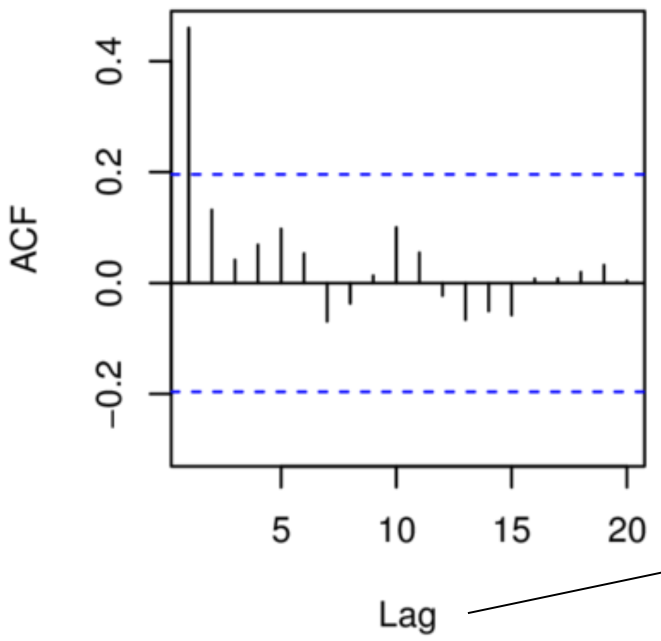
Population growth rate:

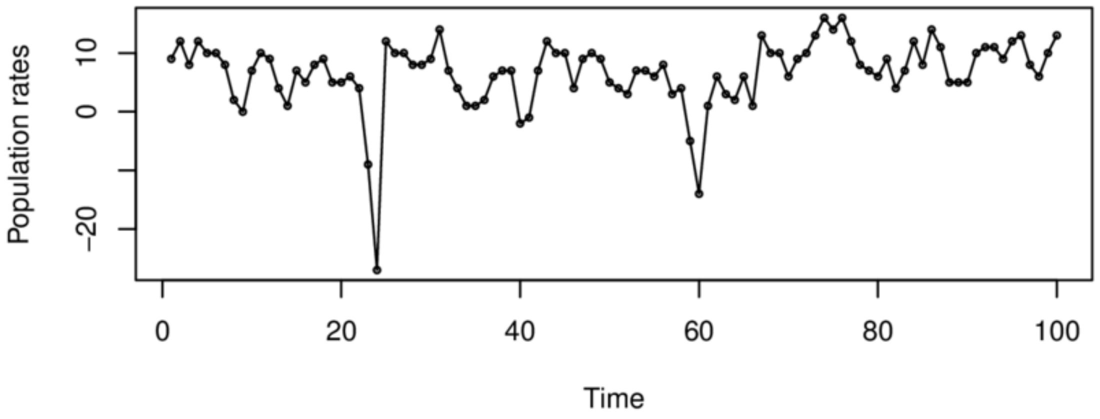


How to plot a time series in R: *plot.ts(x, xlab, ylab, main)*

Describe a time series

Autocorrelation plot



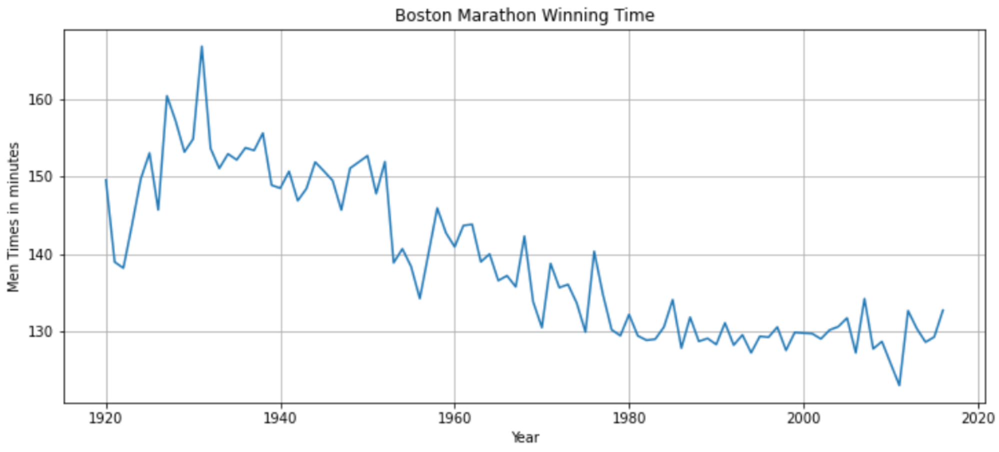


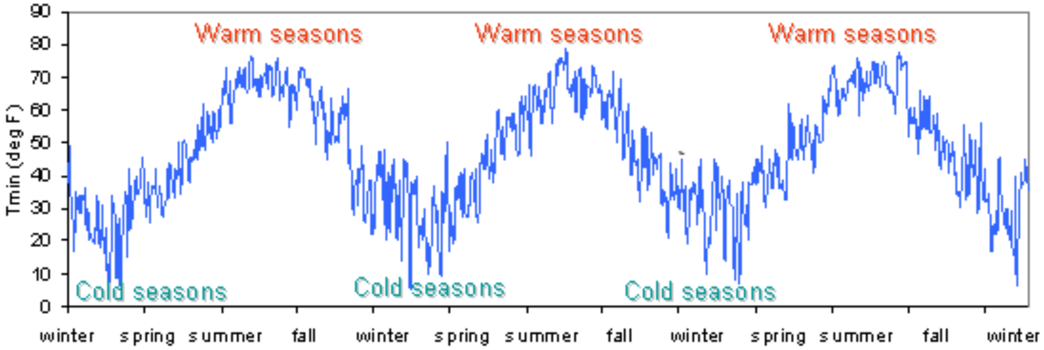
Lag*:* Delay of the observation is taken

How to do an autocorrelation plot for a specific time series in R: acf*(x, main)*

Describe a time series

Trend: General development of a feature over time

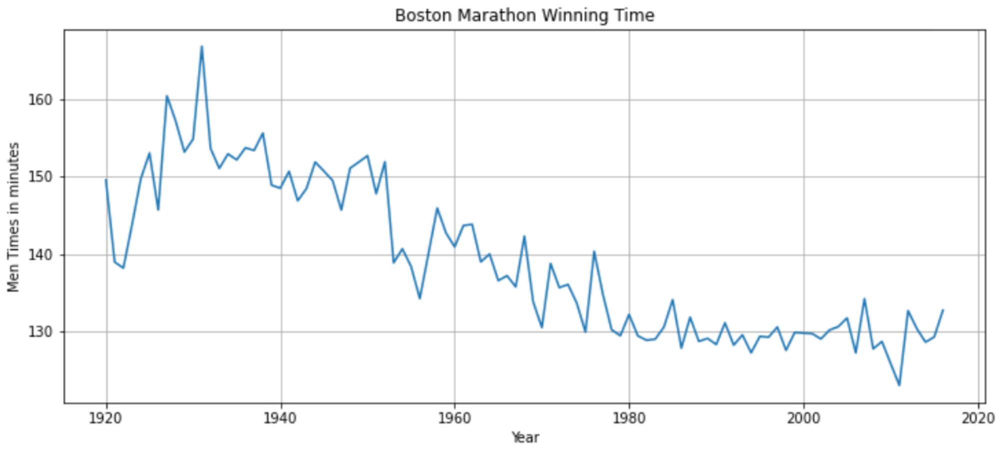


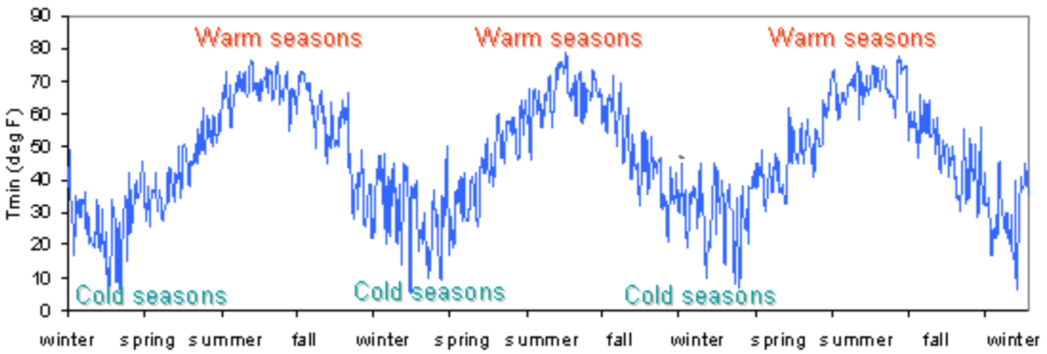


Describe a time series

Trend: General development of a feature over time

Relation between your data and time

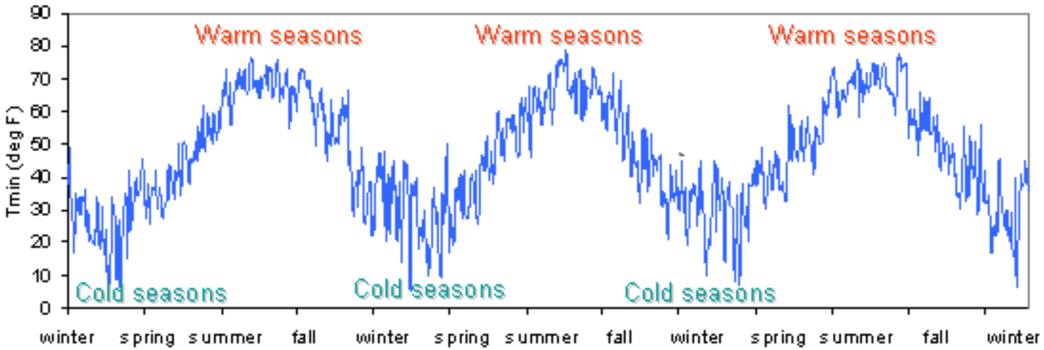




Describe a time series

Trend: General development of a feature over time

Relation between your data and time



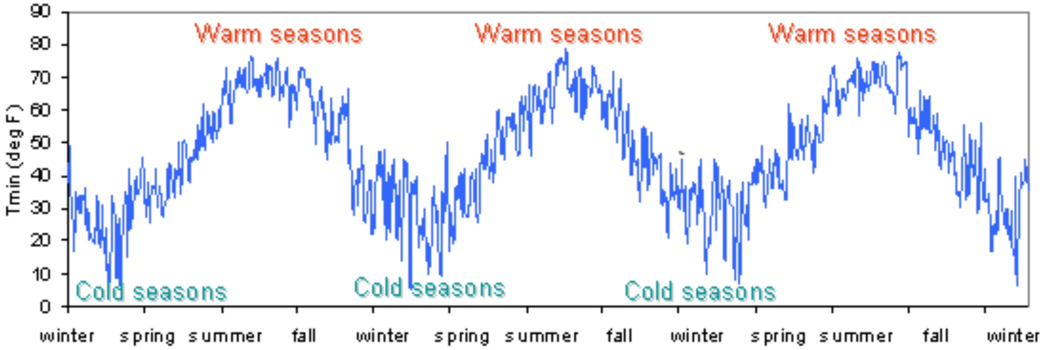
Linear trends

Periodic trends

Describe a time series

Trend: General development of a feature over time

Relation between your data and time



Linear trends

Periodic trends

In R you can remove the linear trend with: *diff()*

Basic models of time series

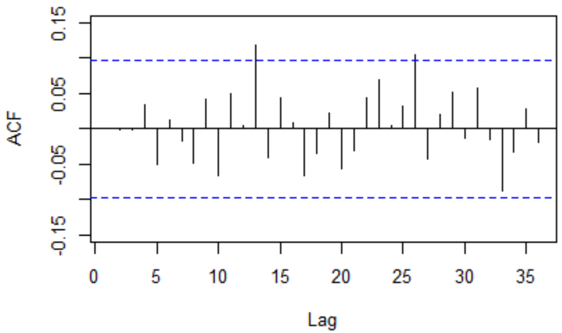
- White Noise model (WN)
- Random walk model (RW)
- Simple moving average model (MA)
- Autoregressive model (AR)

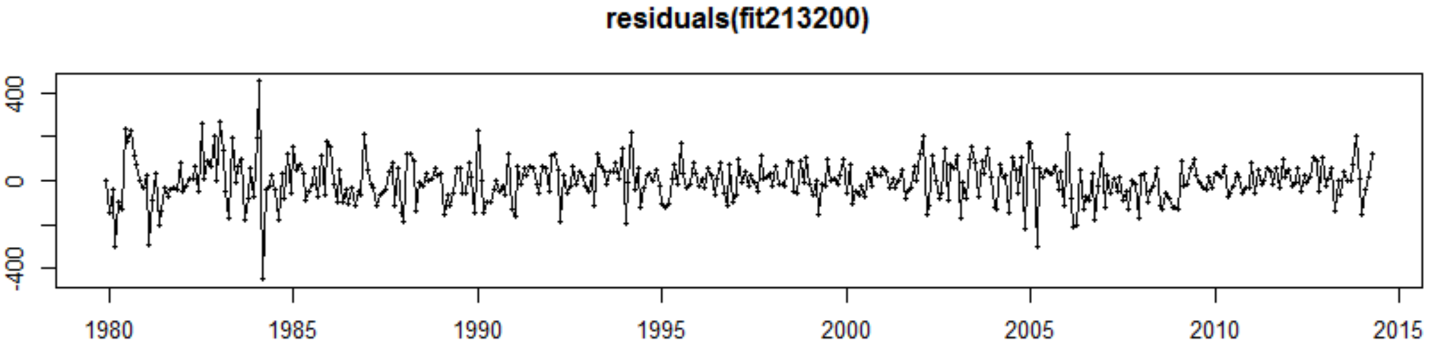
Basic models of time series

- Does our observed time series follow one of these models?
    - White noise
    - Random walk model (RW)
    - Simple moving average model (MA)
    - Autoregressive model (AR)

White Noise model (WN)

The simplest model of time series

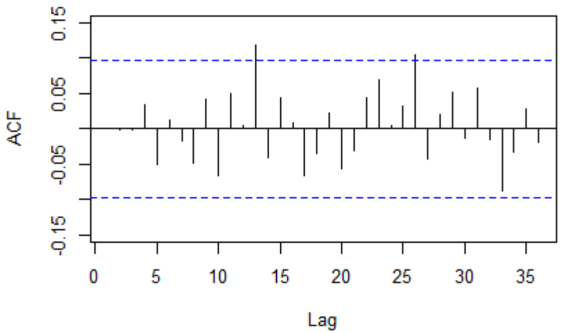


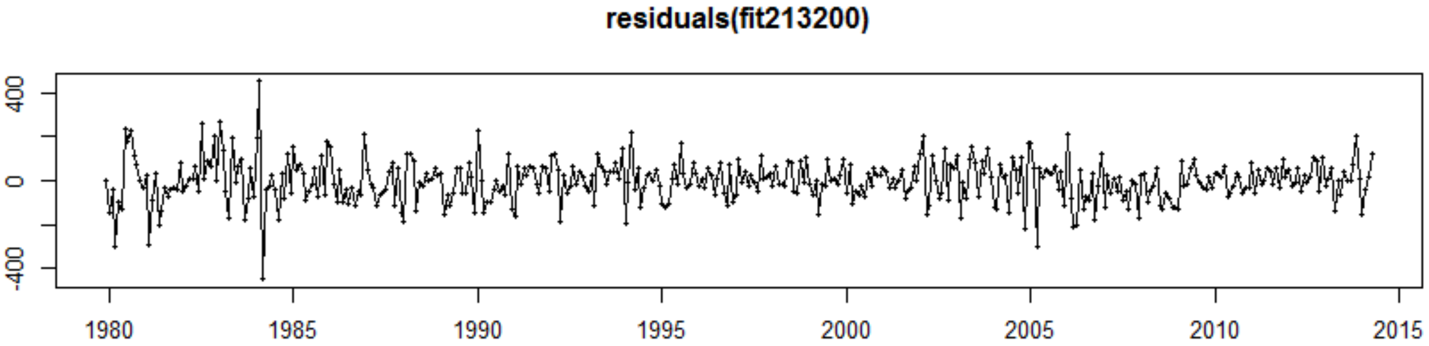


White Noise model (WN)

The simplest model of time series

- A fixed, constant mean

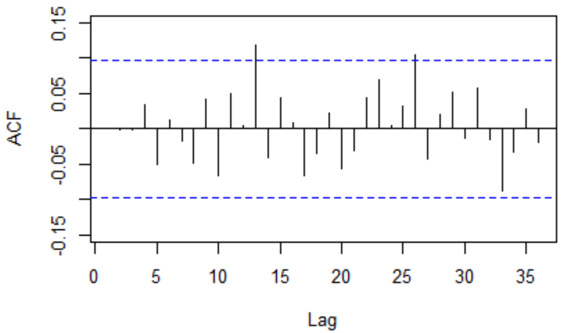


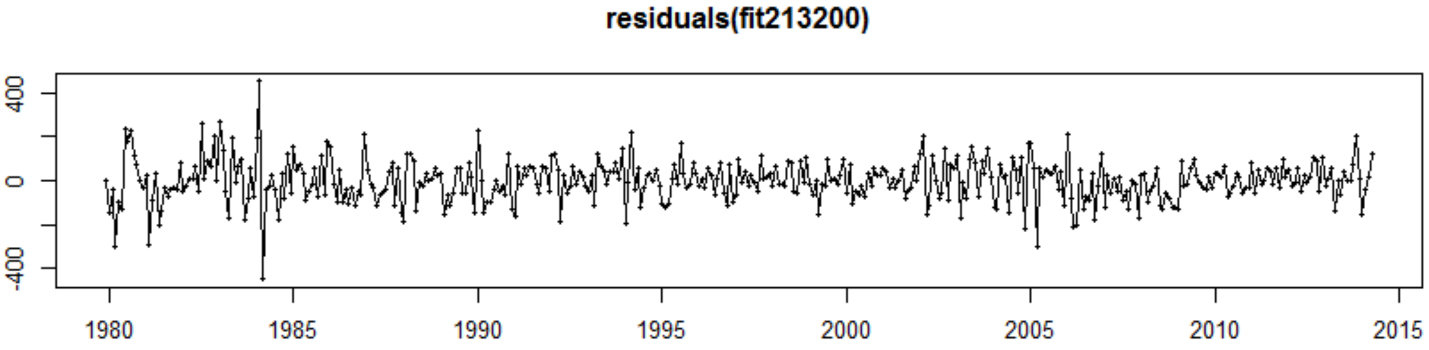


White Noise model (WN)

The simplest model of time series

- A fixed, constant mean
- A fixed, constant variance

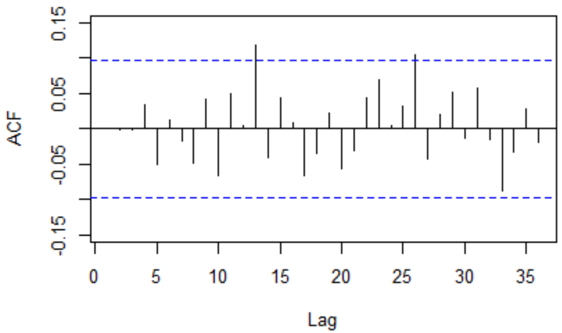


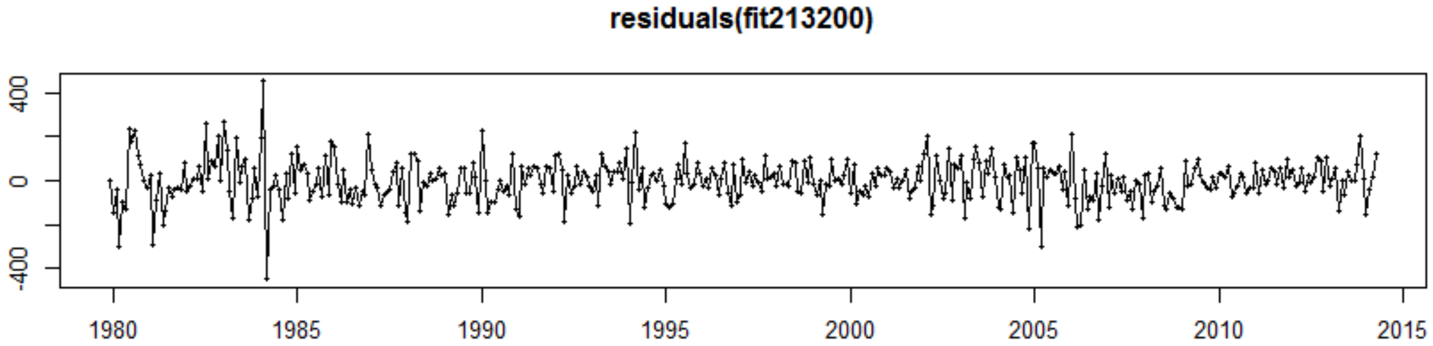


White Noise model (WN)

The simplest model of time series

- A fixed, constant mean
- A fixed, constant variance
- No correlation over time





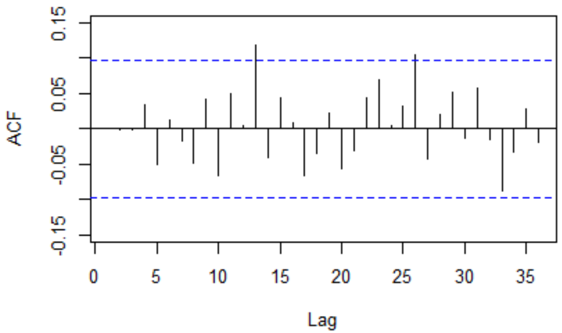
White Noise model (WN)

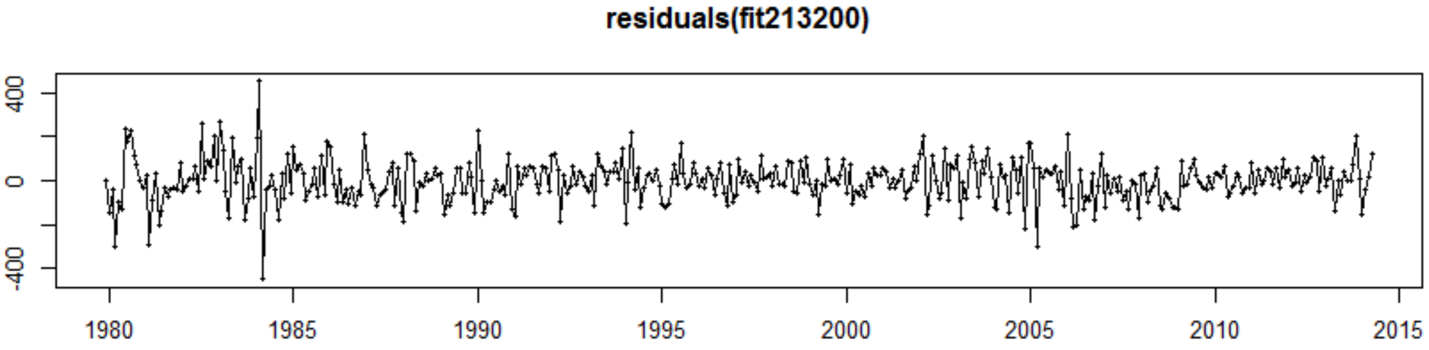
The simplest model of time series

- A fixed, constant mean
- A fixed, constant variance
- No correlation over time

In R you can fit your data in a WN model using this code:

arima(x, order=c(0,0,0))





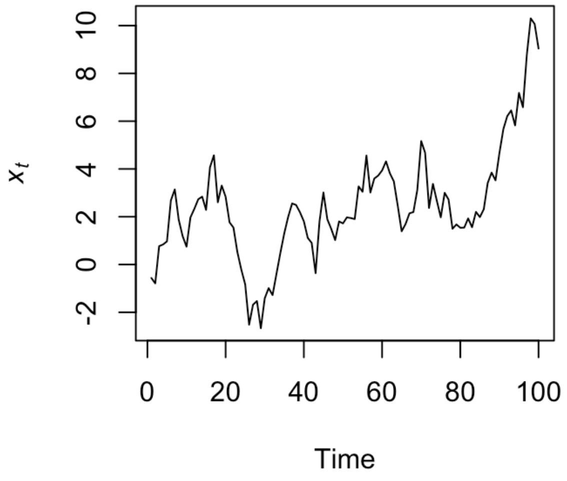
Random walk model (RW)

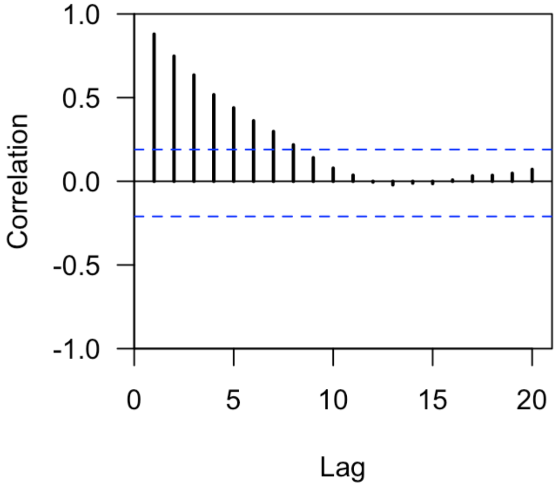
Another simple model of time series

- For each period (x-axis), the variable takes a random step up or down (y-axis)

Random walk model (RW)

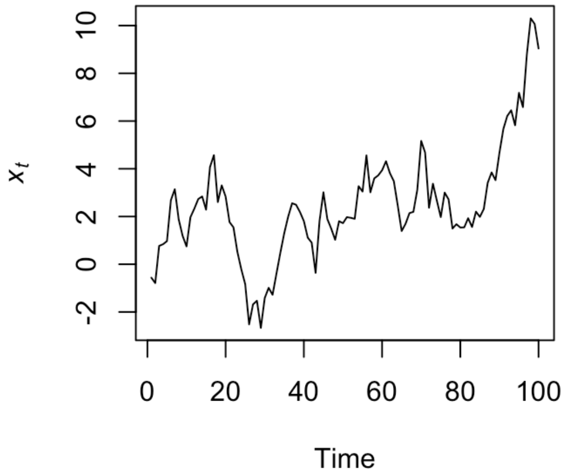
Defined as: Today = Yesterday + Noise

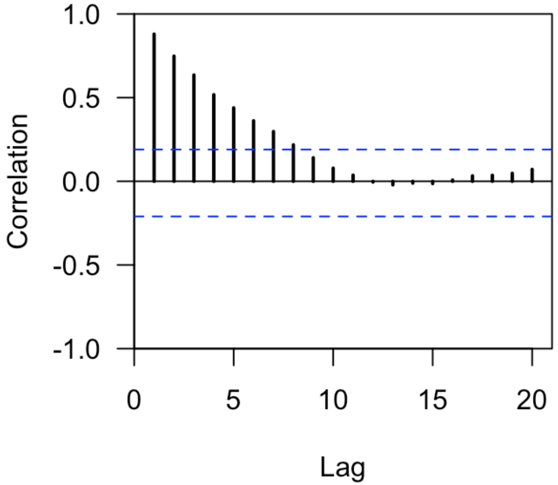




Random walk model (RW)

Defined as: Today = Yesterday + Noise ⟶ Today - Yesterday = Noise



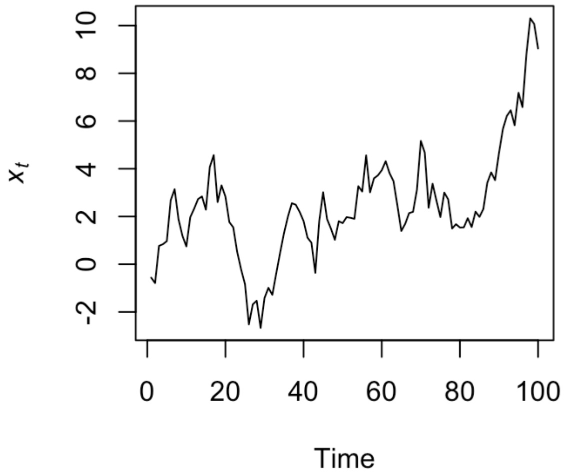


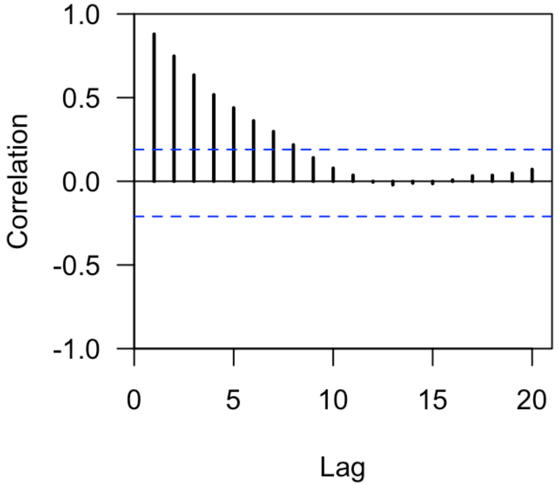
Random walk model (RW)

Defined as: Today = Yesterday + Noise $\longrightarrow$ Today - Yesterday = Noise

White noise

diff(RW) → WN model





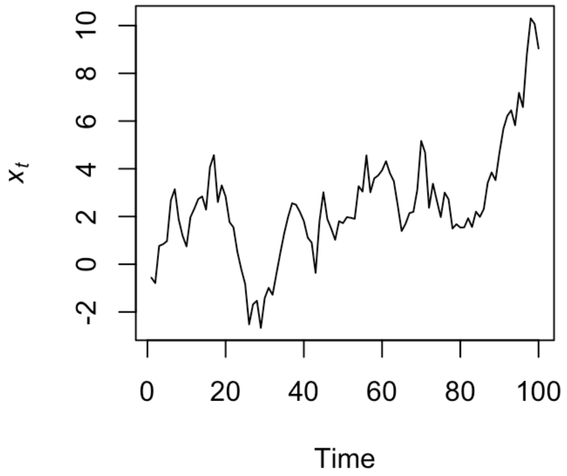
Random walk model (RW)

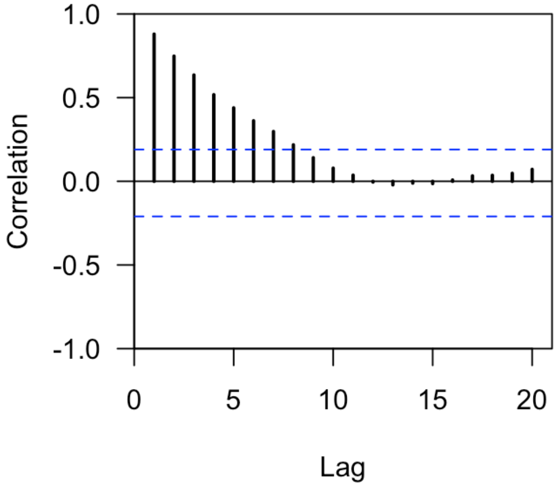
Defined as: Today = Yesterday + Noise $\longrightarrow$ Today - Yesterday = Noise

White noise

- No specific mean
- No specific variance
- Strong dependence over time

$\boxed{diff(RW) \rightarrow WN\ model}$





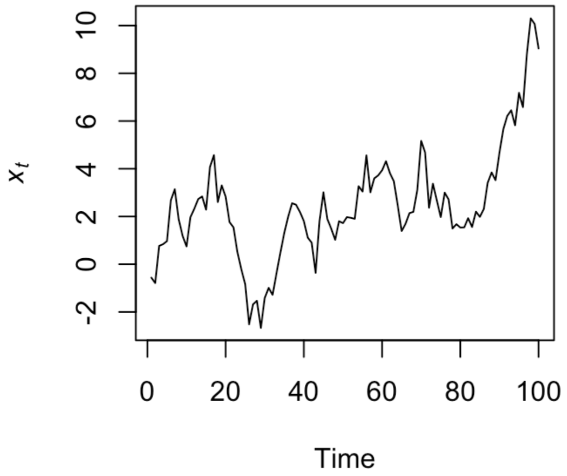
Random walk model (RW)

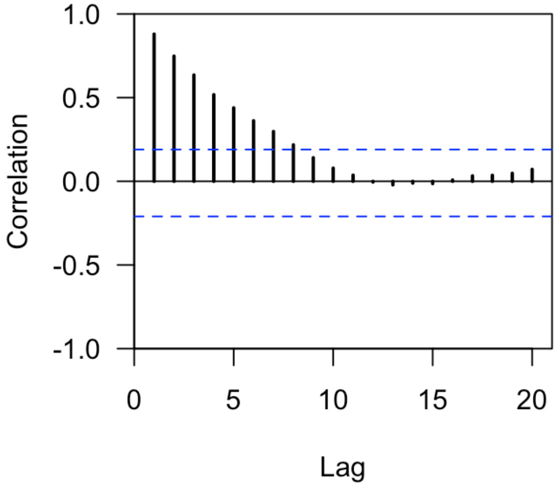
Defined as: Today = Yesterday + Noise ⟶ Today - Yesterday = Noise

White noise

- No specific mean
- No specific variance
- Strong dependence over time

diff(RW) → WN model





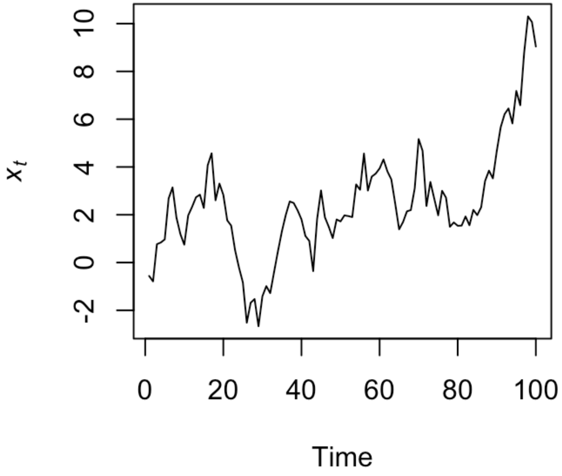
In R you can fit your data in a RW model using this code:

arima(x, order=c(0,1,0))

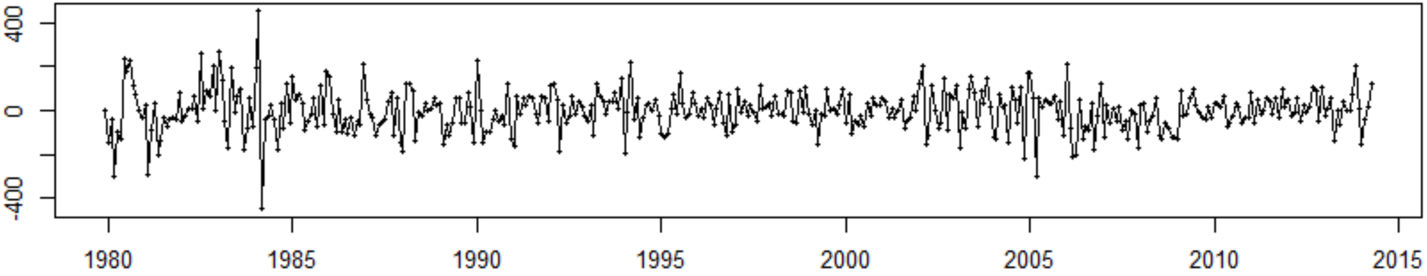
Stationarity

Parsimonious models with distributional stability over time (values are not a function of time)

Random walk model



White Noise model



<u>Stationary models can be modeled with fewer values, however few time series are stationary</u>

Autocorrelation: correlation with past observations

- Study how each time series observation is related to its recent past

- Greater autocorrelation is more predictable

- Lag 1 autocorrelation
  - E.g. compare today versus yesterday for each day
  - cor()
  - Previous stock price is high, likely to be high

- Lag 2 autocorrelation
  - Defined by similarity
  - Compare today versus yesterday AND the day before yesterday

- Autocorrelation function plot: acf()

Autoregressive models (AR)

Linear trend where each observation is regressed on the previous observation

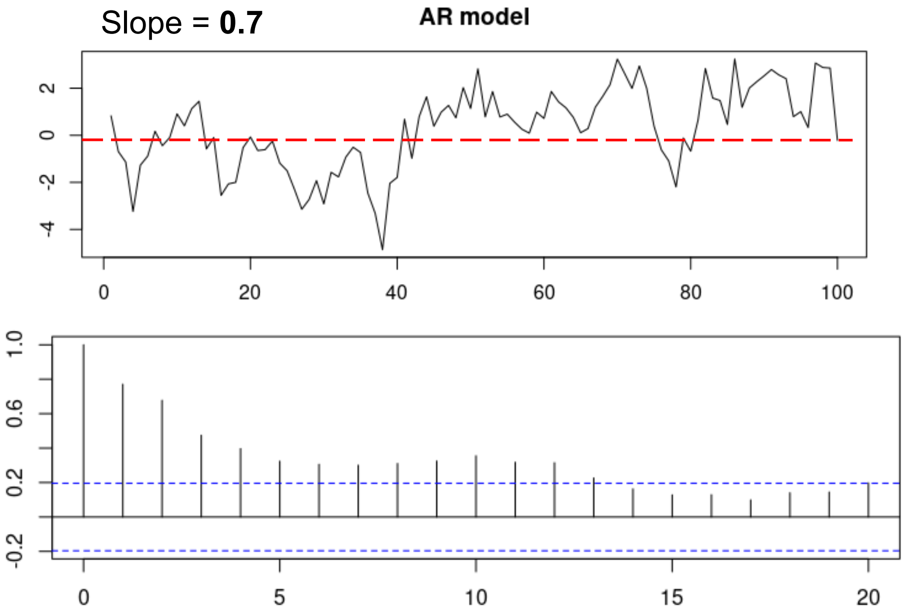
(Today - Mean) = Slope * (Yesterday - Mean) + Noise

Autoregressive models (AR)

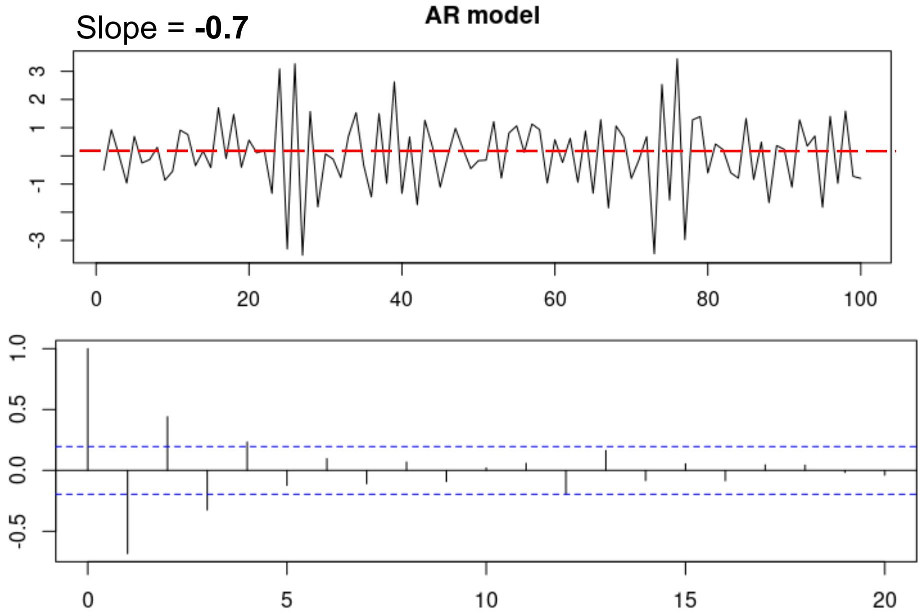
Linear trend where each observation is regressed on the previous observation

(Today - Mean) = Slope * (Yesterday - Mean) + Noise

Slope = **0.7**



Slope = **-0.7**



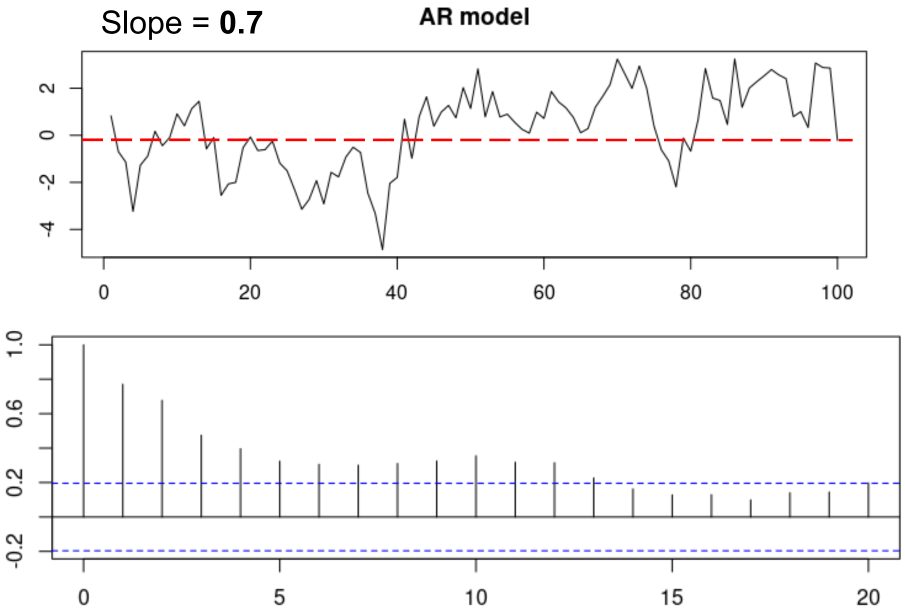
Autoregressive models (AR)

Linear trend where each observation is regressed on the previous observation

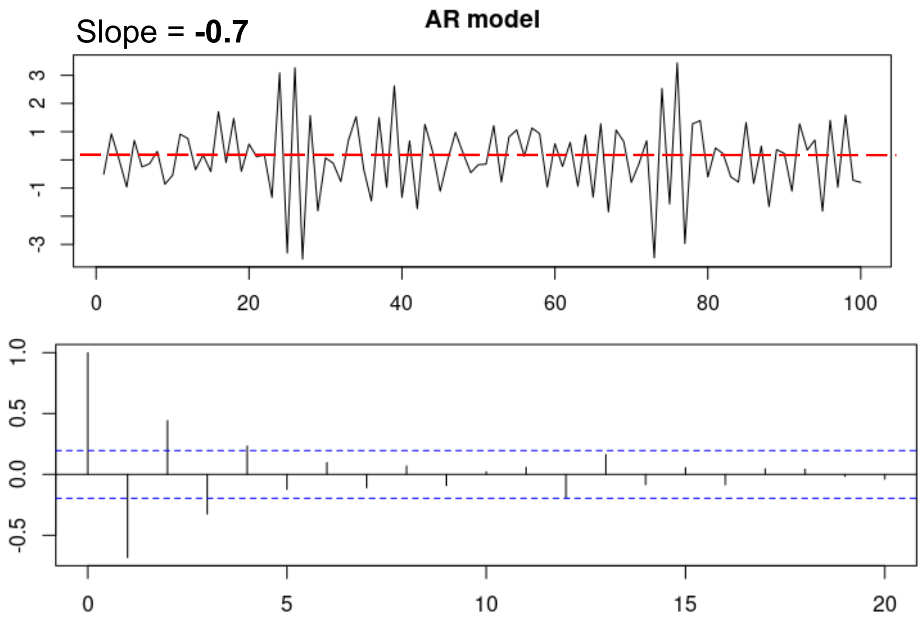
Slope = 1 & Mean = 0

(Today - Mean) = Slope * (Yesterday - Mean) + Noise

Slope = **0.7**



Slope = **-0.7**



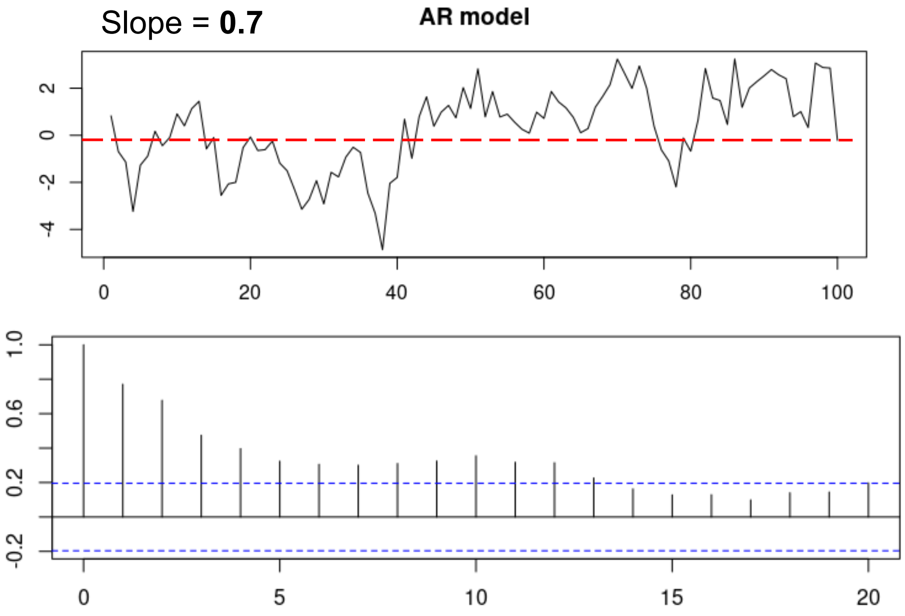
Autoregressive models (AR)

Linear trend where each observation is regressed on the previous observation

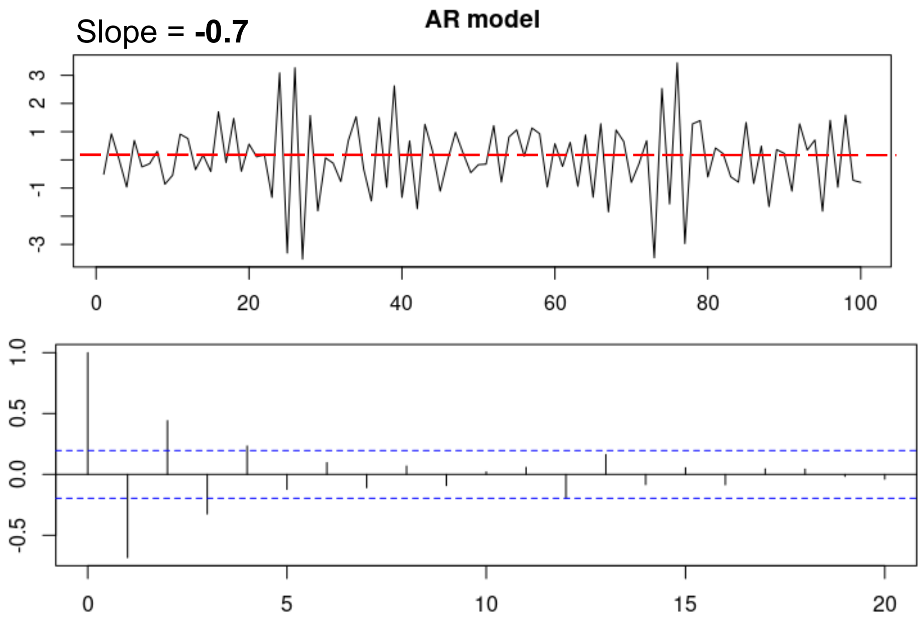
Slope = 1 & Mean = 0 (Today - 0) = 1 * (Yesterday - 0) + Noise

(Today - Mean) = Slope * (Yesterday - Mean) + Noise

Slope = **0.7**



Slope = **-0.7**



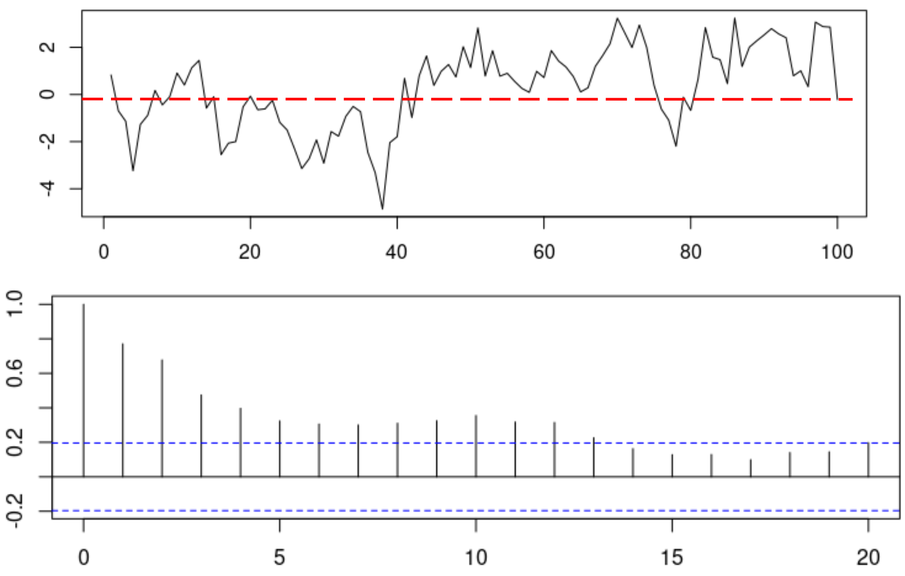
Autoregressive models (AR)

Linear trend where each observation is regressed on the previous observation

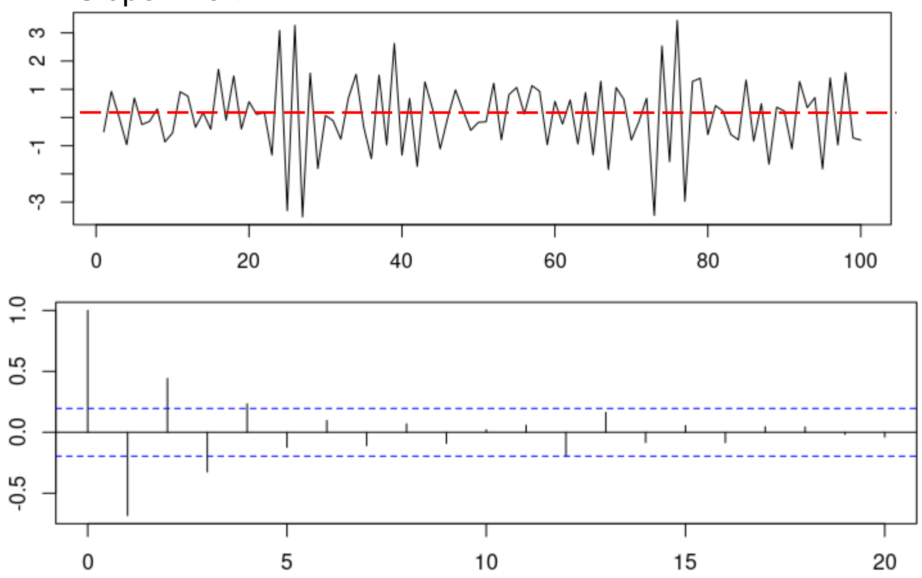
Slope = 1 & Mean = 0 Today = Yesterday + Noise

(Today - Mean) = Slope * (Yesterday - Mean) + Noise

Slope = **0.7**



Slope = **-0.7**



Autoregressive models (AR)

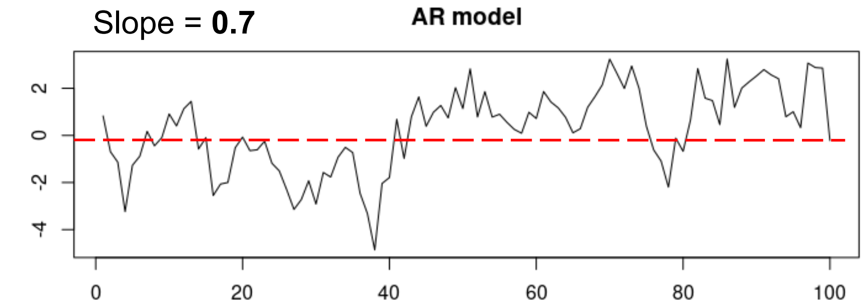
Linear trend where each observation is regressed on the previous observation

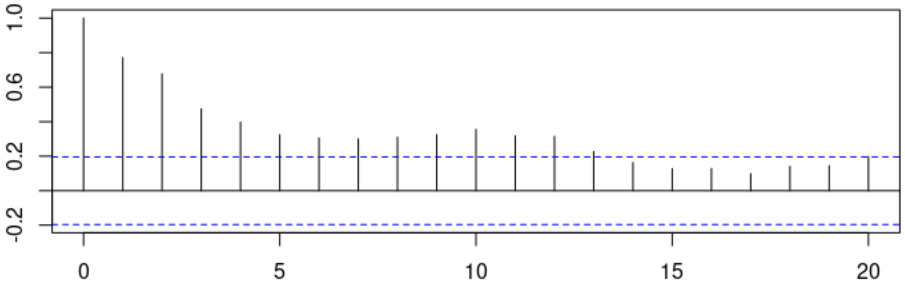
Slope = 1 & Mean = 0 — Today = Yesterday + Noise

(Today - Mean) = Slope * (Yesterday - Mean) + Noise

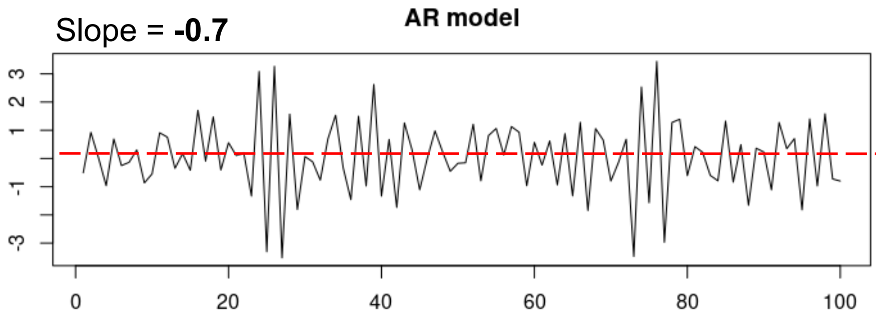
Slope = 0

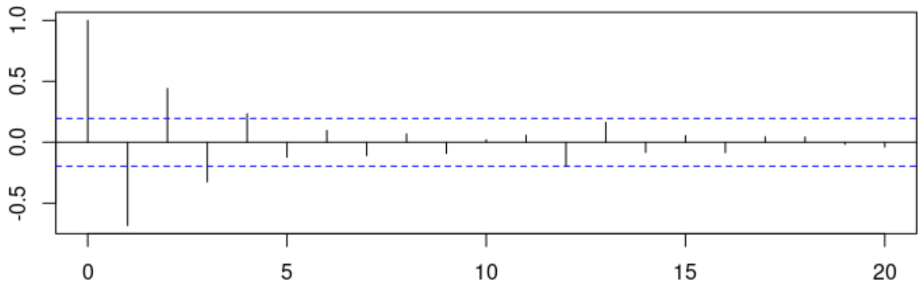
Slope = **0.7**





Slope = **-0.7**





Autoregressive models (AR)

Linear trend where each observation is regressed on the previous observation

Slope = 1 & Mean = 0

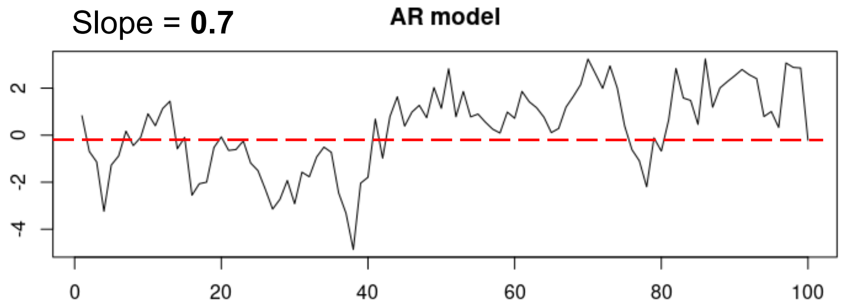
Today = Yesterday + Noise

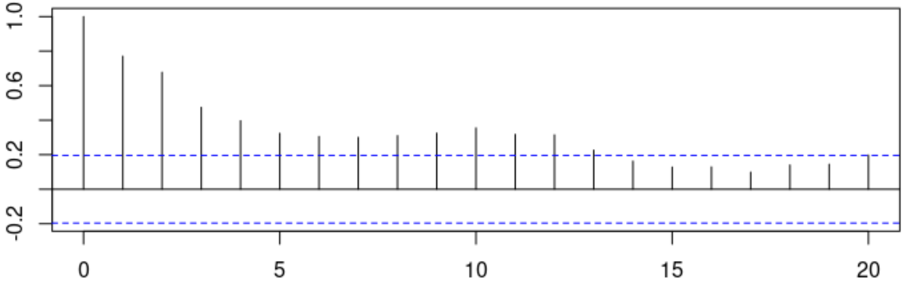
(Today - Mean) = Slope * (Yesterday - Mean) + Noise

(Today - Mean) = 0 * (Yesterday - Mean) + Noise

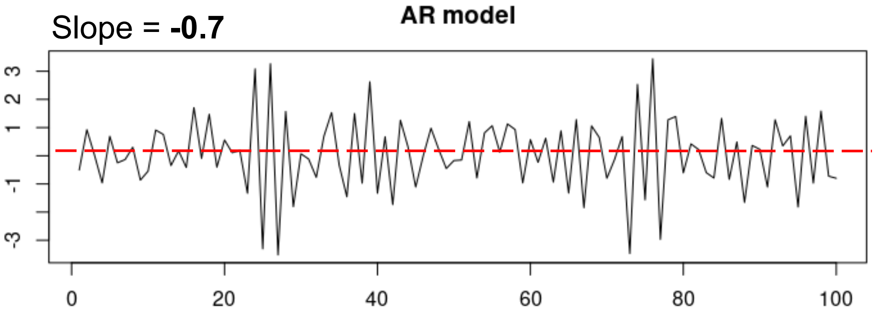
Slope = 0

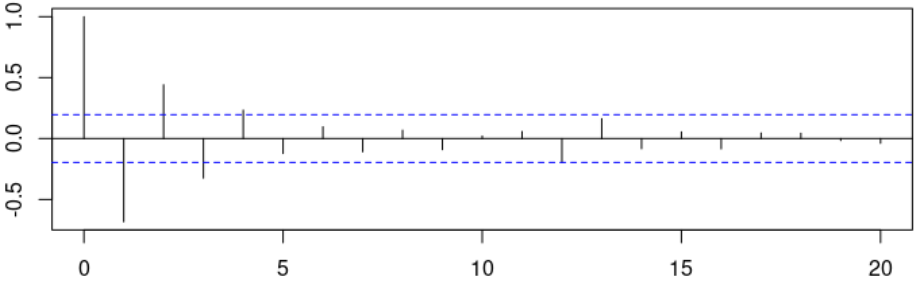
Slope = **0.7**





Slope = **-0.7**





Autoregressive models (AR)

Linear trend where each observation is regressed on the previous observation

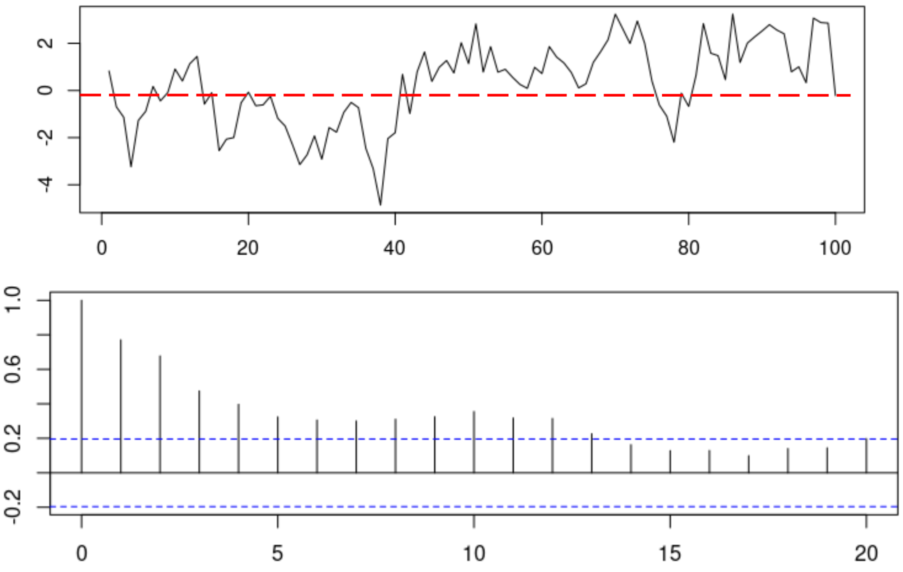
(Today - Mean) = Slope * (Yesterday - Mean) + Noise

Slope = 1 & Mean = 0 — Today = Yesterday + Noise

Slope = 0 — (Today - Mean) = Noise

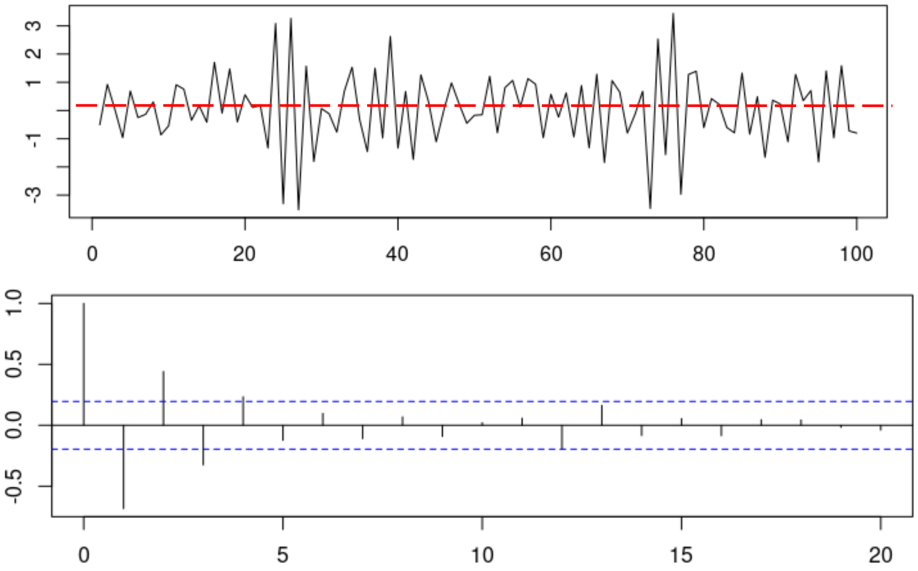
Slope = **0.7**

AR model



Slope = **-0.7**

AR model



Simple Moving Average models (MA)

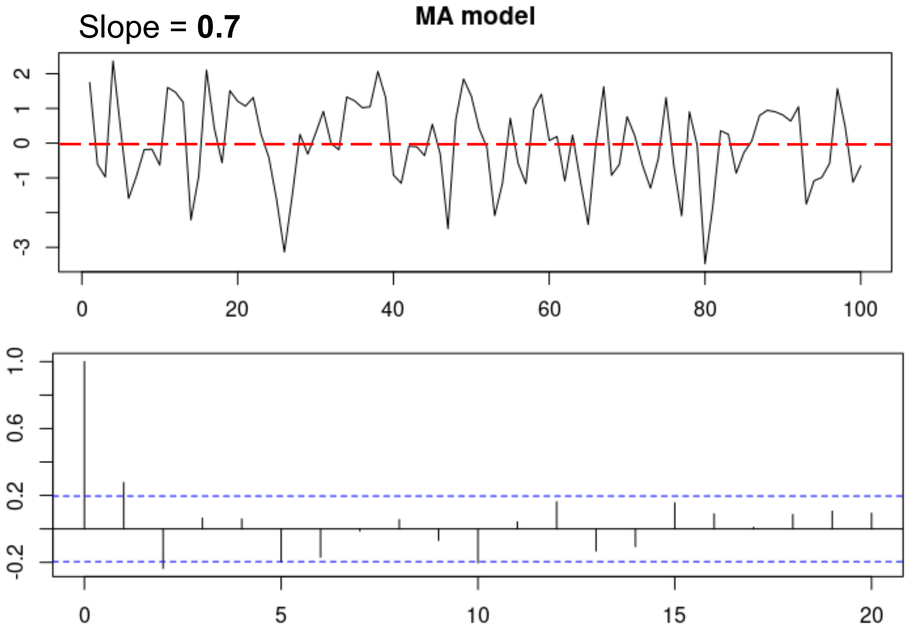
Linear trend where each observation is regressed on the previous innovation, which is not actually observed

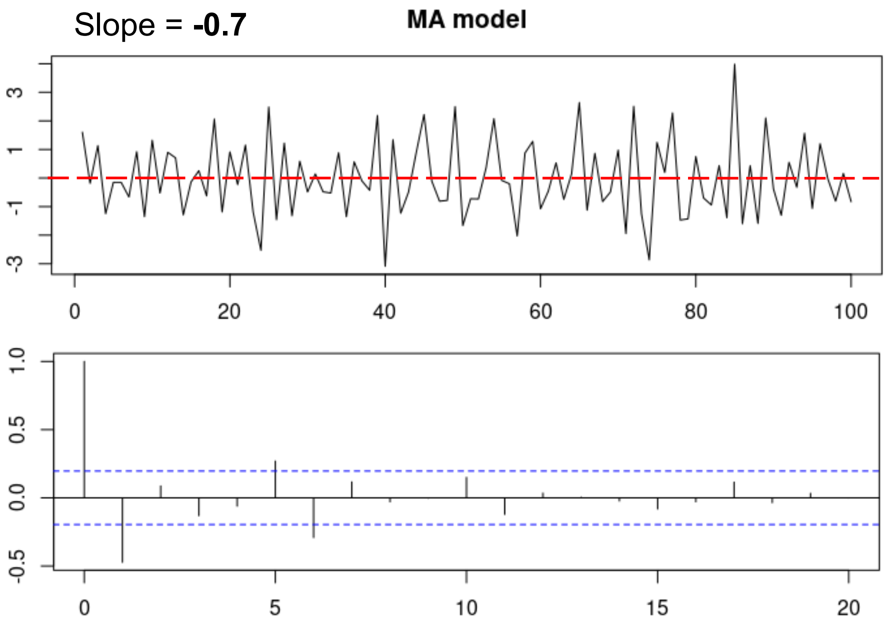
Today = Mean + Noise + Slope * (Yesterday’s Noise)

Simple Moving Average models (MA)

Linear trend where each observation is regressed on the previous innovation, which is not actually observed

Today = Mean + Noise + Slope * (Yesterday's Noise)





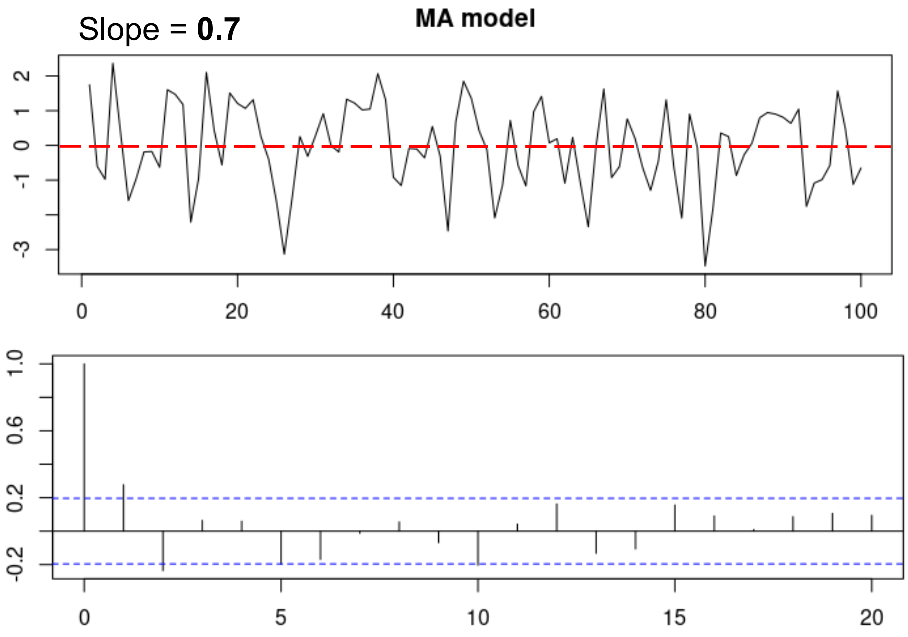
Simple Moving Average models (MA)

Linear trend where each observation is regressed on the previous innovation, which is not actually observed

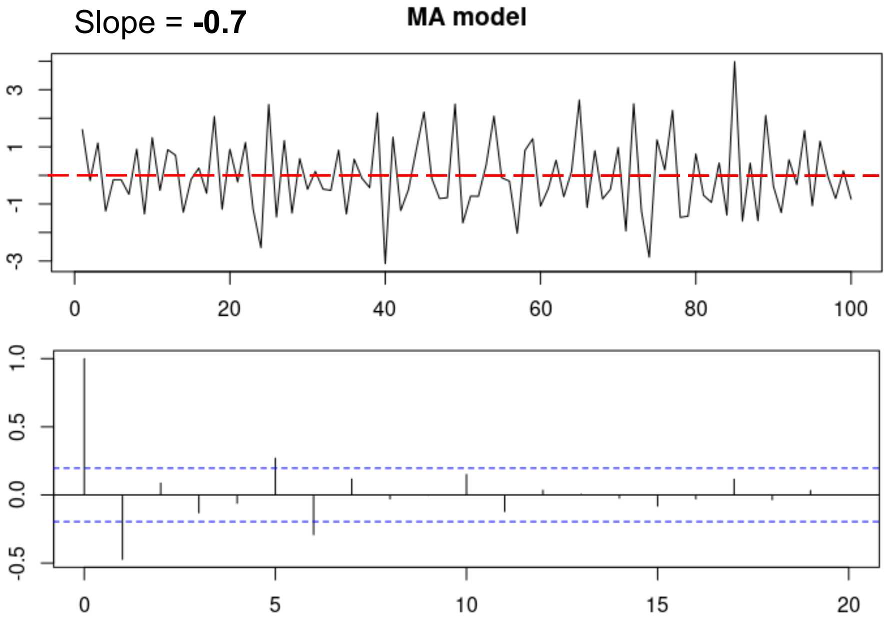
Slope = 0

Today = Mean + Noise + Slope * (Yesterday's Noise)

Slope = **0.7**



Slope = **-0.7**



Simple Moving Average models (MA)

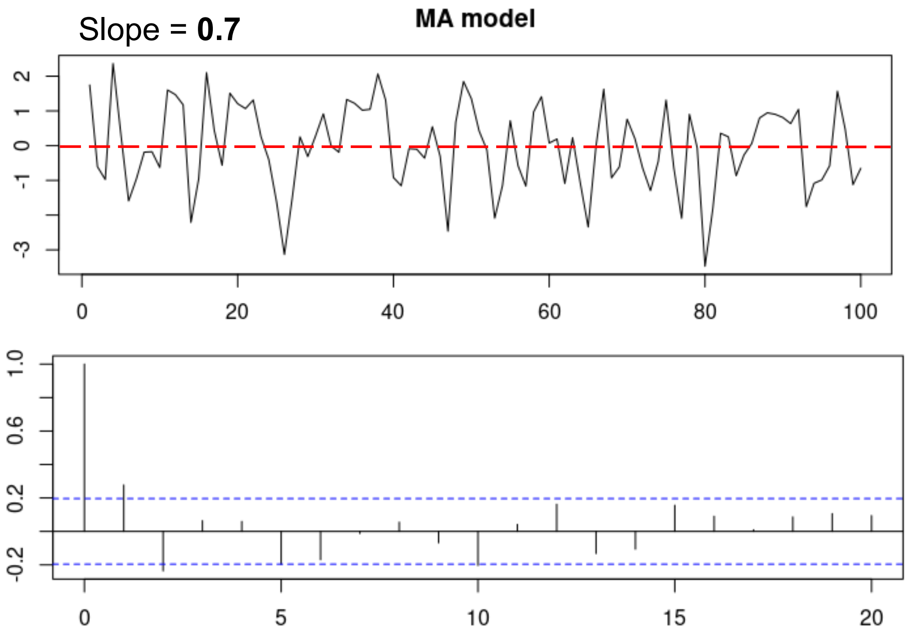
Linear trend where each observation is regressed on the previous innovation, which is not actually observed

Slope = 0

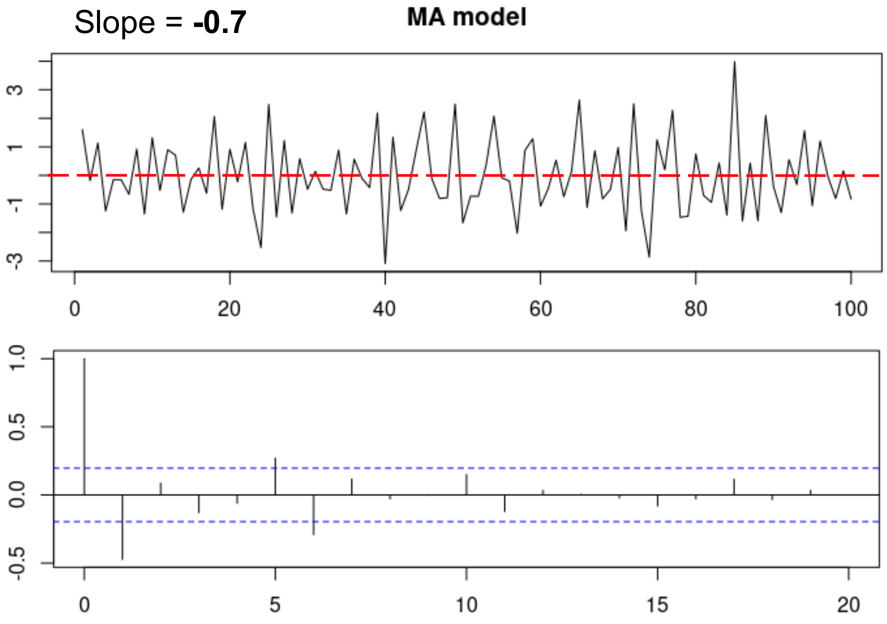
Today = Mean + Noise + 0 * (Yesterday's Noise)

Today = Mean + Noise + Slope * (Yesterday's Noise)

Slope = **0.7**



Slope = **-0.7**



Simple Moving Average models (MA)

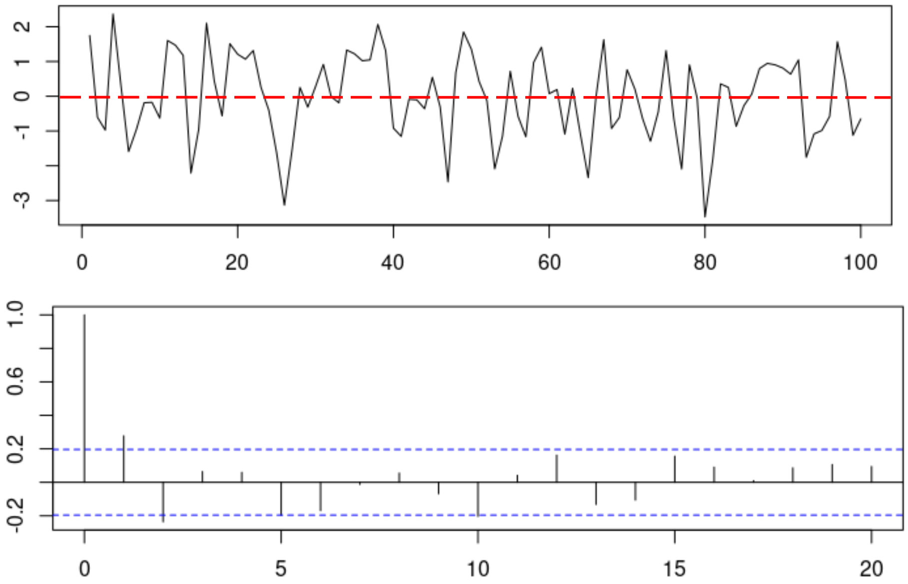
Linear trend where each observation is regressed on the previous innovation, which is not actually observed

Slope = 0

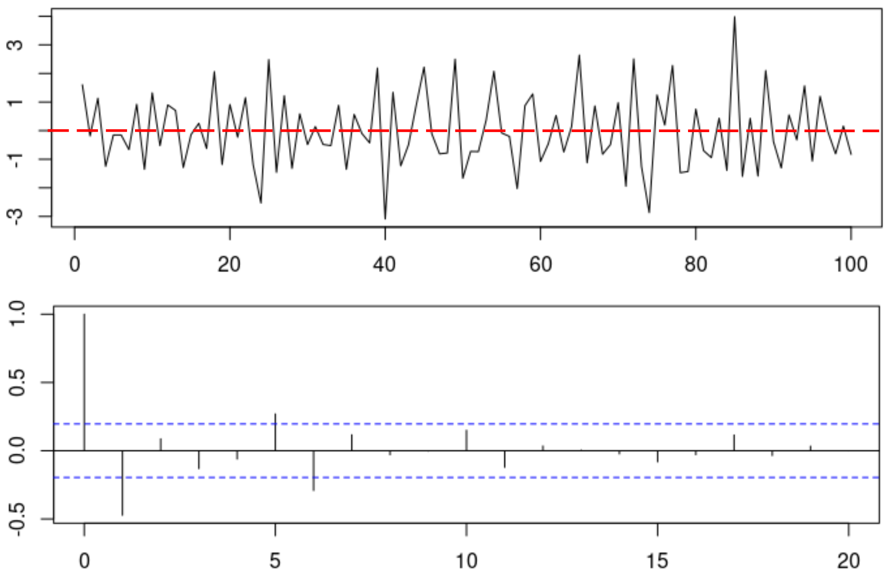
Today = Mean + Noise

Today = Mean + Noise + Slope * (Yesterday's Noise)

Slope = **0.7** **MA model**



Slope = **-0.7** **MA model**



Forecast

We can use the Autoregressive models (AR) and the Simple Moving Average models (MA) models to fit our data and try to forecast our time series

- *Fitted values:* Forecast (estimation) of an observation using all previous ones

Forecast

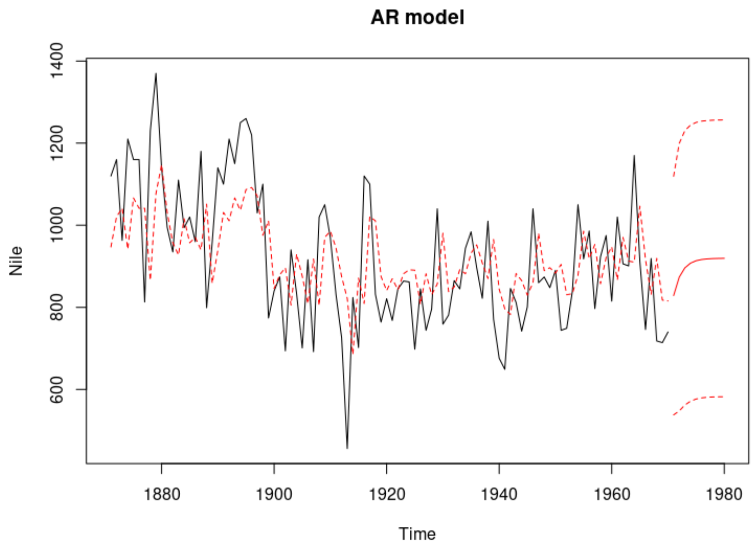
We can use the Autoregressive models (AR) and the Simple Moving Average models (MA) models to fit our data and try to forecast our time series

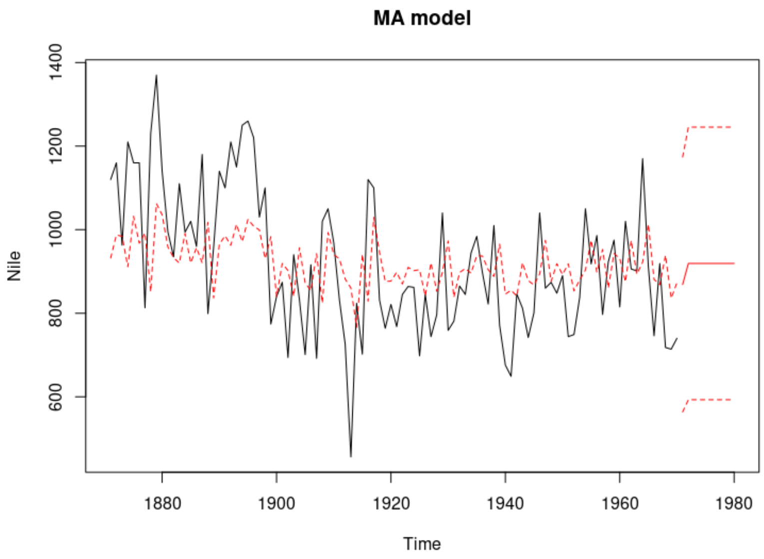
- *Fitted values:* Forecast (estimation) of an observation using all previous ones
- *Residuals:* Difference between the observation and the corresponding fitted values

Forecast

We can use the Autoregressive models (AR) and the Simple Moving Average models (MA) models to fit our data and try to forecast our time series

- *Fitted values:* Forecast (estimation) of an observation using all previous ones
- *Residuals:* Difference between the observation and the corresponding fitted values

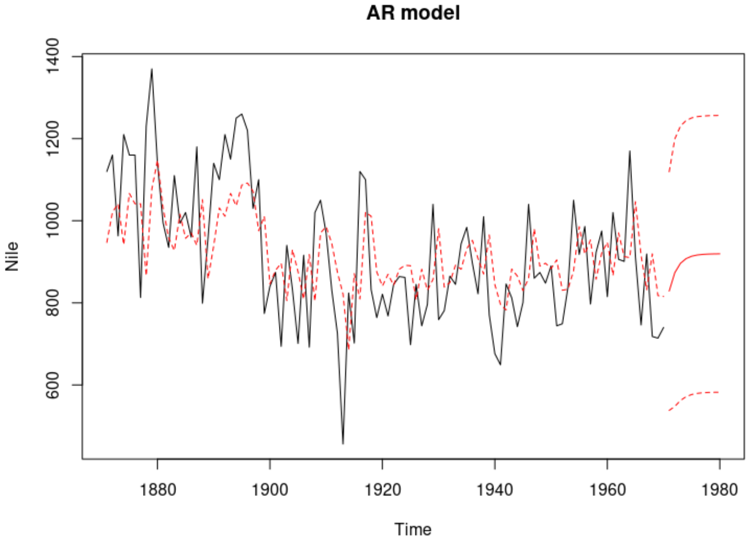


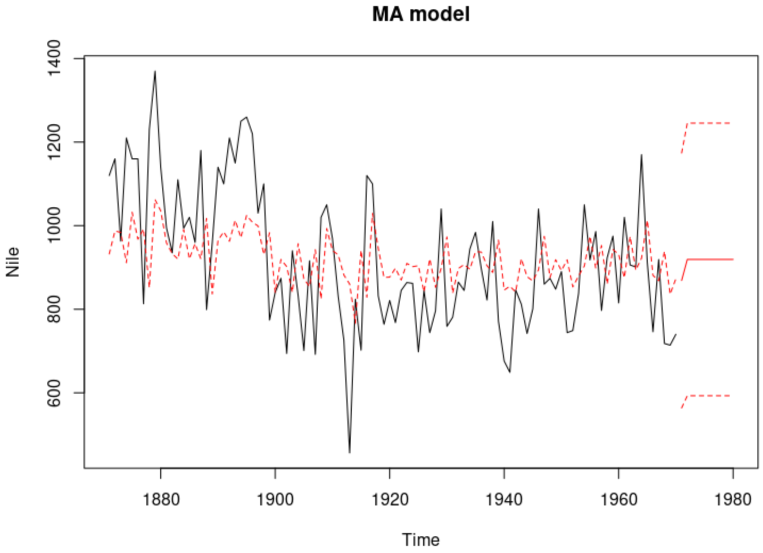


Forecast in R

To fit our time series on AR model:

AR model ← *arima(x, order=c(1,0,0))*





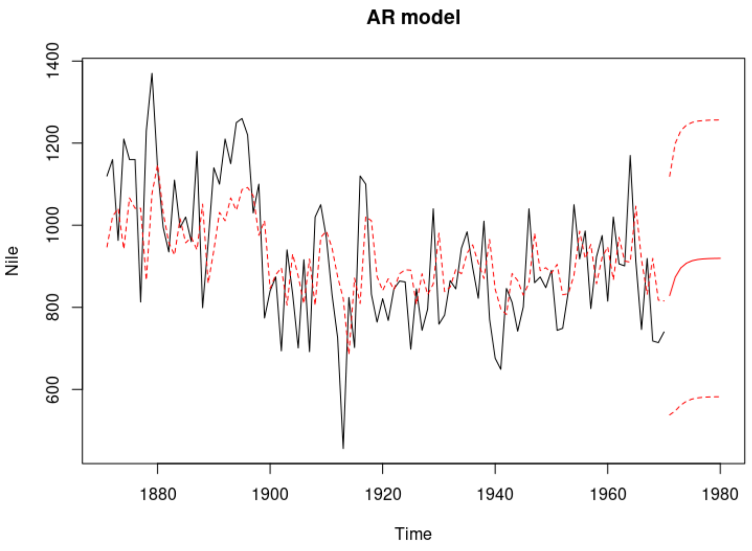
Forecast in R

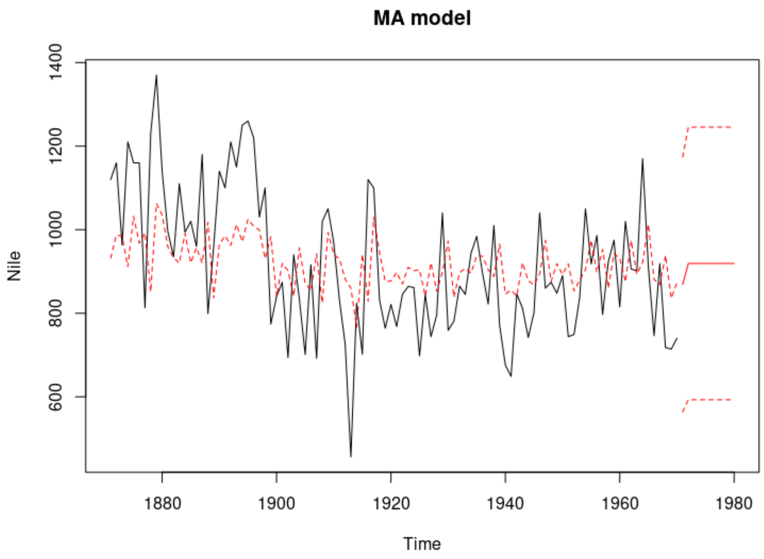
To fit our time series on AR model:

AR model ← *arima(x, order=c(1,0,0))*

To fit our time series on MA model:

MA model ← *arima(x, order=c(0,0,1))*





Forecast in R

To fit our time series on AR model:

AR model ← *arima(x, order=c(1,0,0))*

Fitted values ← *x - residuals(x)*

Forecast ← *predict(x)$pred*

Forecast SD ← *predict(x)$sd*

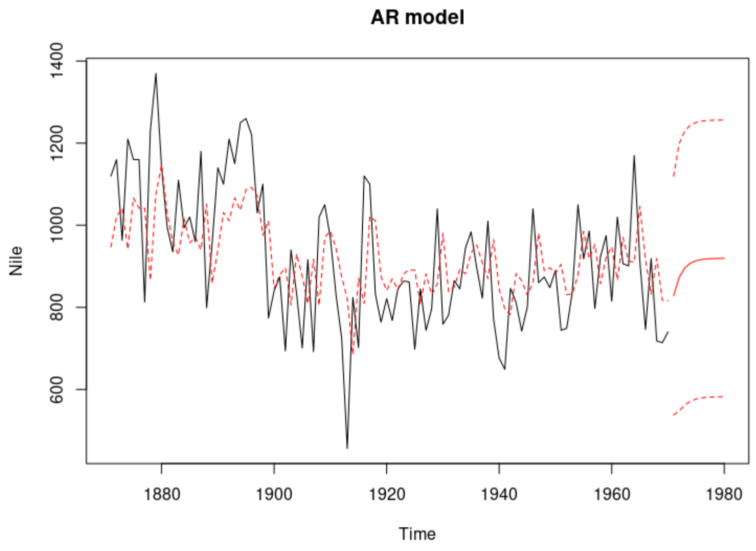
To fit our time series on MA model:

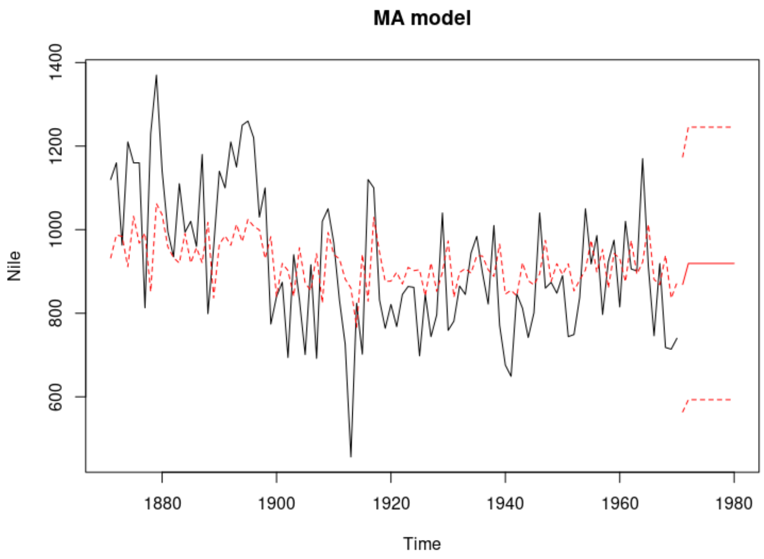
MA model ← *arima(x, order=c(0,0,1))*

Fitted values ← *x - residuals(x)*

Forecast ← *predict(x)$pred*

Forecast SD ← *predict(x)$sd*





Compare different models

When comparing different models we want to find out which model explains better our data regardless of the **individual independent variables** in the model.

It doesn’t mean that there is a right and a wrong model!

Compare different models

When comparing different models we want to find out which model explain better my data regardless of the **individual independent variables** in the model.

! It doesn't mean that there is a right and a wrong model!

Akaike's information criteria (AIC):

AIC estimates model complexity. It works by estimating the expected performance of the model's predictions, for that scope it uses observed data and hypothetical sample generated by the same model.

The best model show the smallest value; a difference within **4 - 7** units indicate less support, a difference over **10** indicate that the worse model can be omitted

Limits of time series analysis

- Limited to previous observations
- Seriously affected by NAs values
- When forecasting doesn’t take into account variables

General workflow

- Outline a hypothesis for a particular time series and its behavior

One liners for time series terminology

- **Trend**: long-term change in the mean of the data;
- **Seasonality**: regular and predictable changes;
- **Residuals**: de-seasonalised and de-trended series;
- **Stationarity**: When time series properties remain constant over time;
- **Autocorrelation**: Correlation with past observations;
- **Heteroskedasticity**: Changes in the variance;
- **Regularity**: Whether the series is captured at regular intervals;
- **Frequency**: Frequency at which the series is observed;
- **Reflexivity**: When the forecast affects the outcome;
- **Outliers**: Rare but possibly interesting observations;
- **Regimes and Change Detection**: When the data distribution changes;
- **Dimensionality**: Number of variables in the time series.

https://towardsdatascience.com/12-things-you-should-know-about-time-series-975a185f4eb2

R