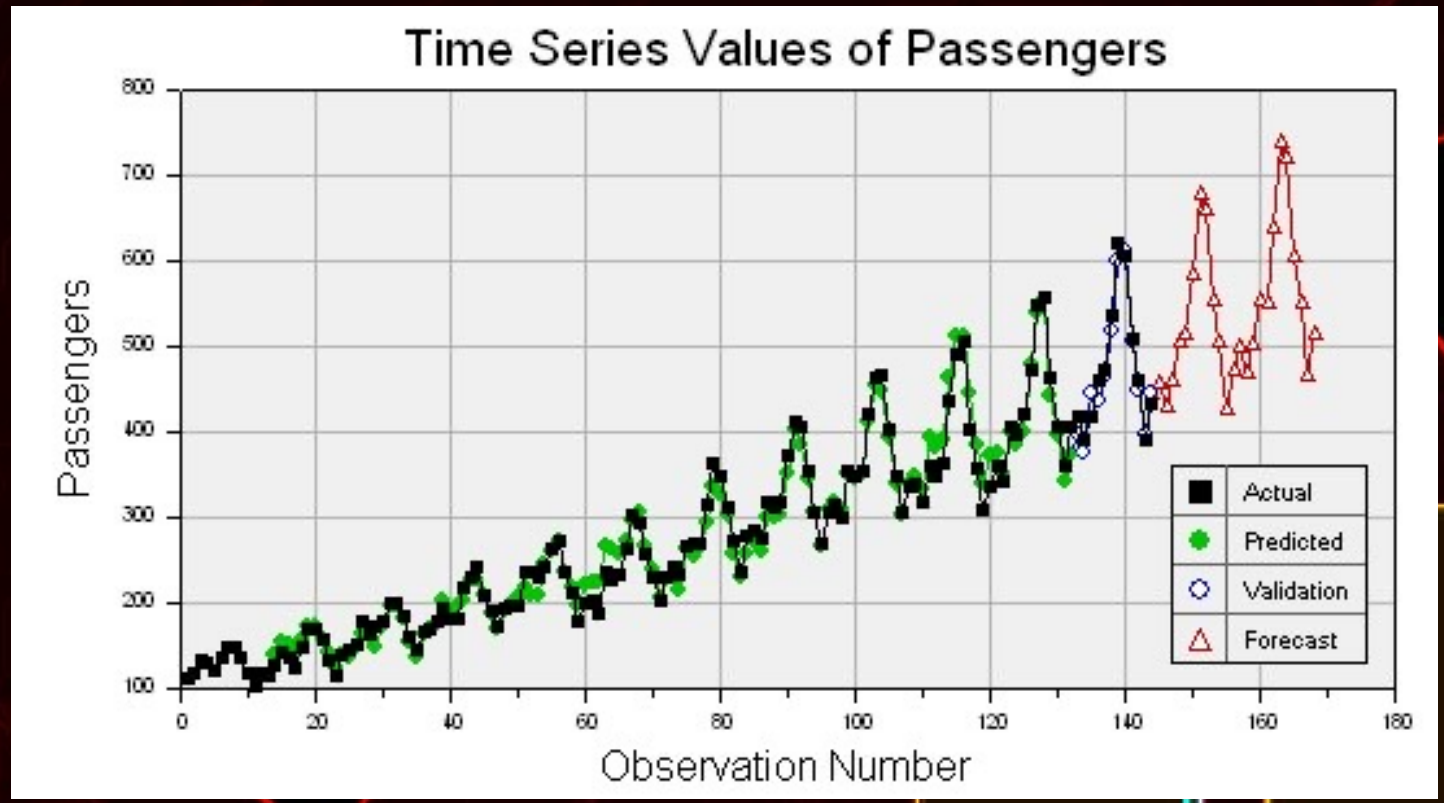
Idősor elemzés

Jónás Dániel
Data scientist

Idősor

- Egymást követő állandó intervallumokkal regisztrált adatok
- Az időbeliség fontos
- Például:
  - Időjárás
  - Energiafogyasztás
  - Forgalmi adatok



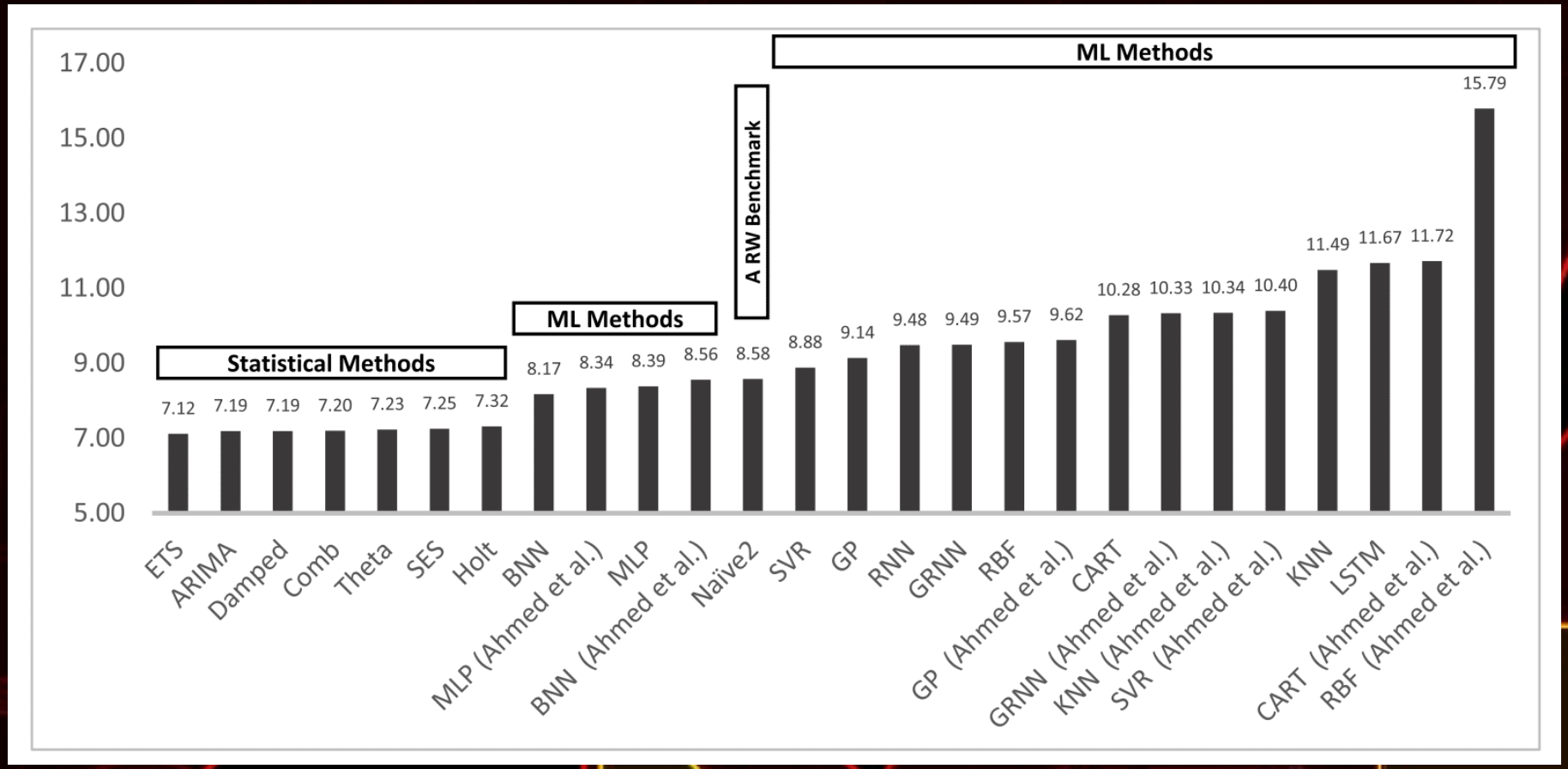
Modell típusok

Hagyományos statisztikai modellek	Gépi tanulási algoritmusok	Neurális hálók
- Exponenciális simítás - Holt-winters-módszer - ARIMA és társai	- Lineáris modellek - Fa alapú modellek - SVM	- RNN - LSTM - GRU - Transformer

Makridakis competition

- M4 (2018)
  - Csak statisztikai módszerek kombinációi ↑
  - Statisztikai módszerek + ML vagy NN ↑
  - Tisztán ML módszerek rosszul teljesítettek ↓
- M5 (2020)
  - Tisztán ML módszerek teljesítettek legjobban
  - ML + NN módszerek is megjelentek

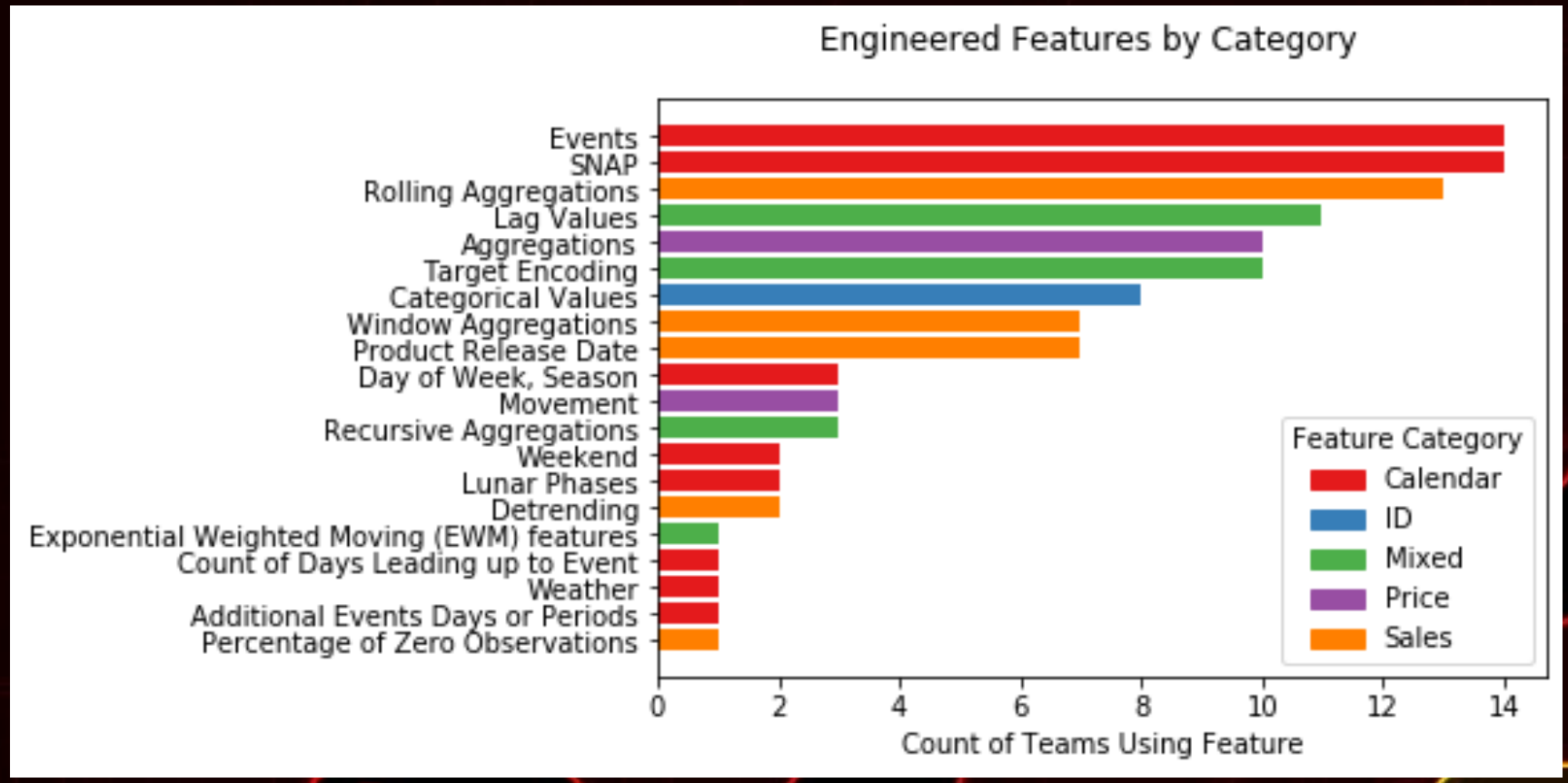
Makridakis competition



Makridakis competition

Team*	Grad. Boosting**	Neural Networks	Proprietary	Statistical Model	Recursive?***
A1	220				Both
A2	10	2 (NBEATS)			R
A4	40				NR
A5	7				R
A18	1				R
U1	126				NR
U2	10		1 (histogram)	1 (SSA)	R
U3	10	3 (Keras - Dense)			R
U4		24 (PyTorch - LSTM)			R
U5	280				NR
U7		2 (Keras - LSTM)			R
U12	9				NR
U18	4	1 (TF Keras - Dense)		1 (Scipy Stats)	Both
U24	10 + 1 (ngboost)	2 (TF Keras - Dense)		3 (Statsmodel QuantReg)	R
U48		1 (PyTorch - Seq2Seq)			R

Makridakis competition



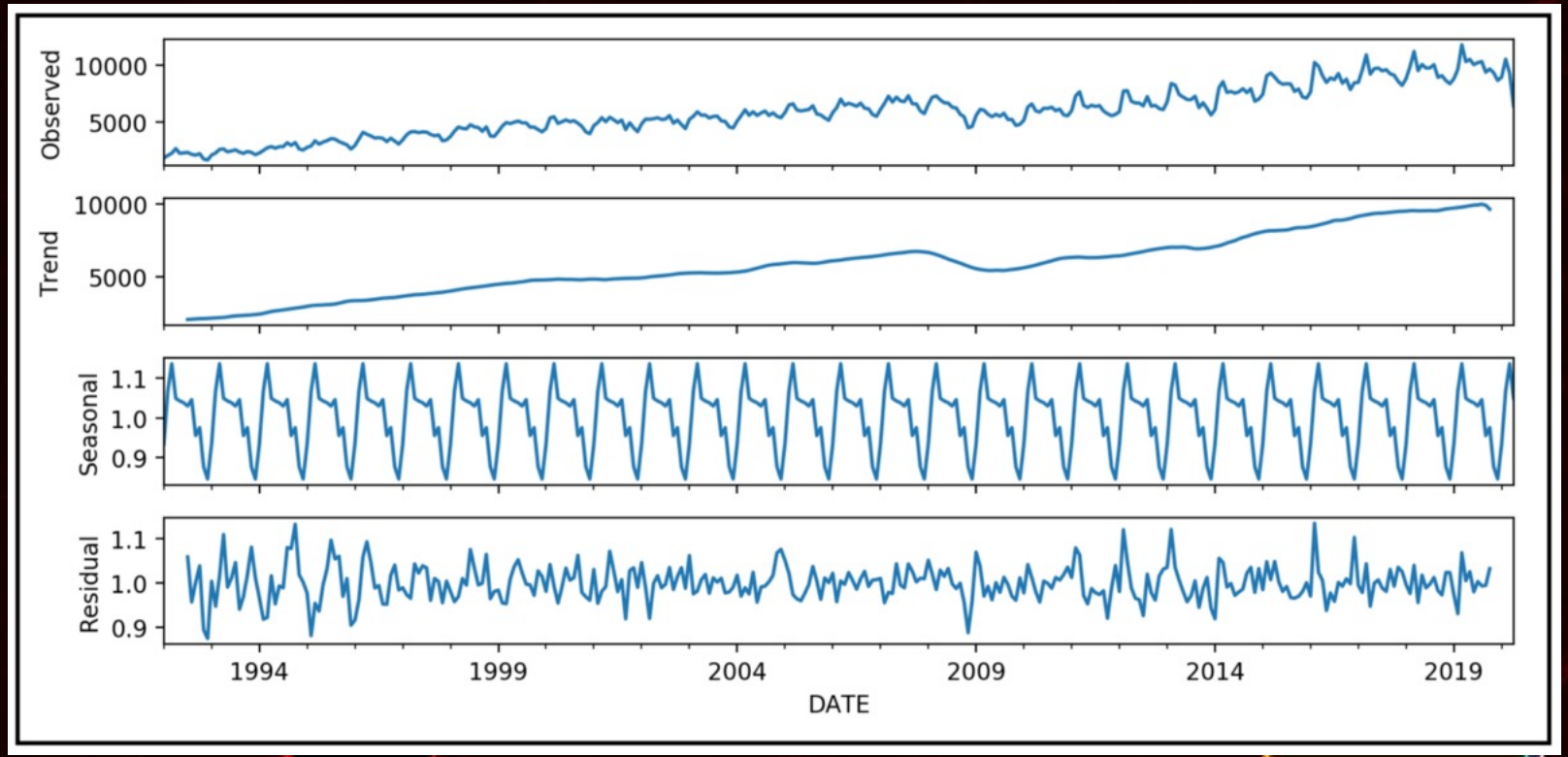
Modell típusok

Hagyományos statisztikai modellek	Gépi tanulási algoritmusok	Neurális hálók
- Exponenciális simítás - Holt-winters-módszer - ARIMA és társai	- Lineáris modellek - Fa alapú modellek - SVM	- RNN - LSTM - GRU - Transformer

Statisztikai módszerek

Hagyományos statisztikai modellek

- Exponenciális simítás
- Holt-winters-módszer
- ARIMA és társai

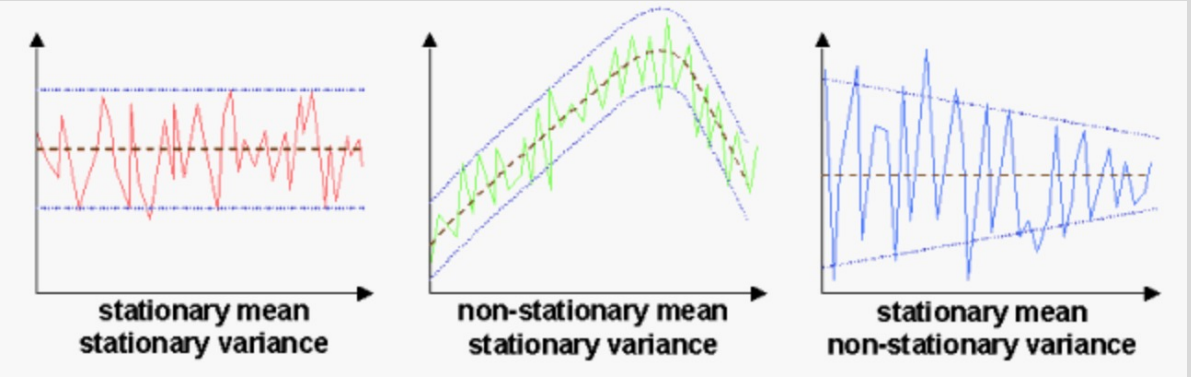


Statisztikai módszerek

Hagyományos statisztikai modellek

- Exponenciális simítás
- Holt-winters-módszer
- ARIMA és társai

- Legtöbbször stacionárius idősort feltételeznek
  - Állandó átlag
  - Állandó szórás

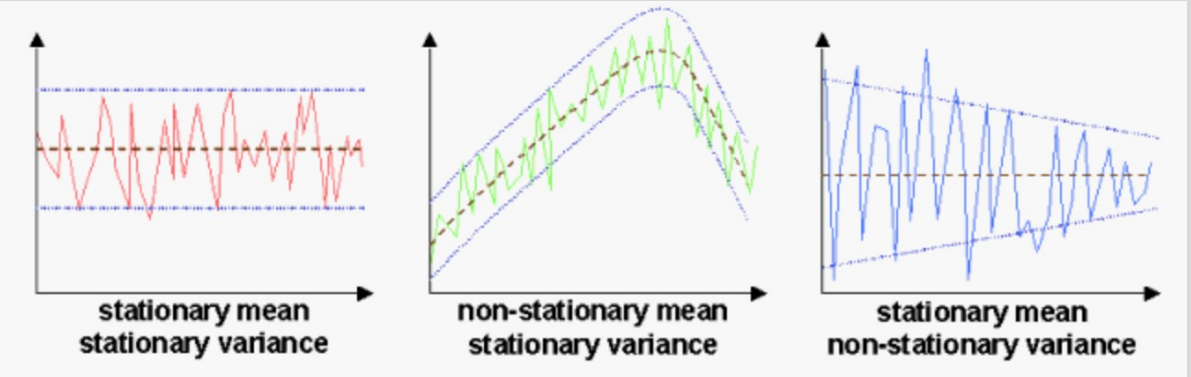


Statisztikai módszerek

Hagyományos statisztikai modellek

- Exponenciális simítás
- Holt-winters-módszer
- ARIMA és társai

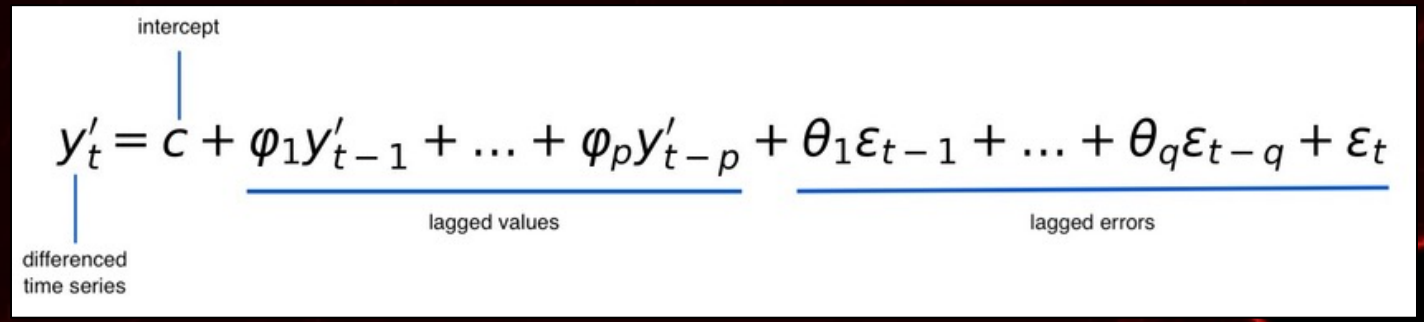
- Legtöbbször stacionárius idősort feltételeznek
  - Állandó átlag
  - Állandó szórás



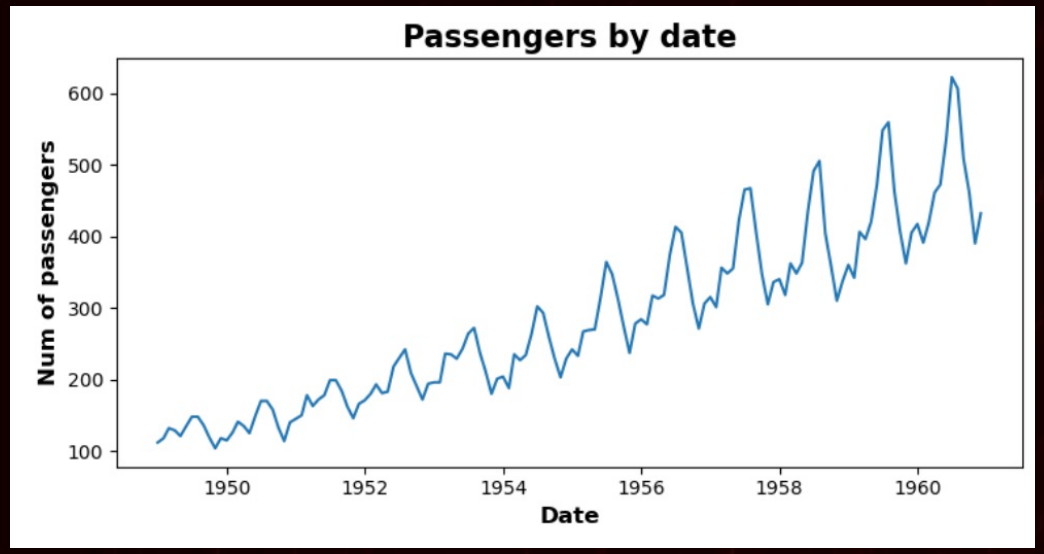
Statisztikai módszerek

ARIMA (p, d, q)

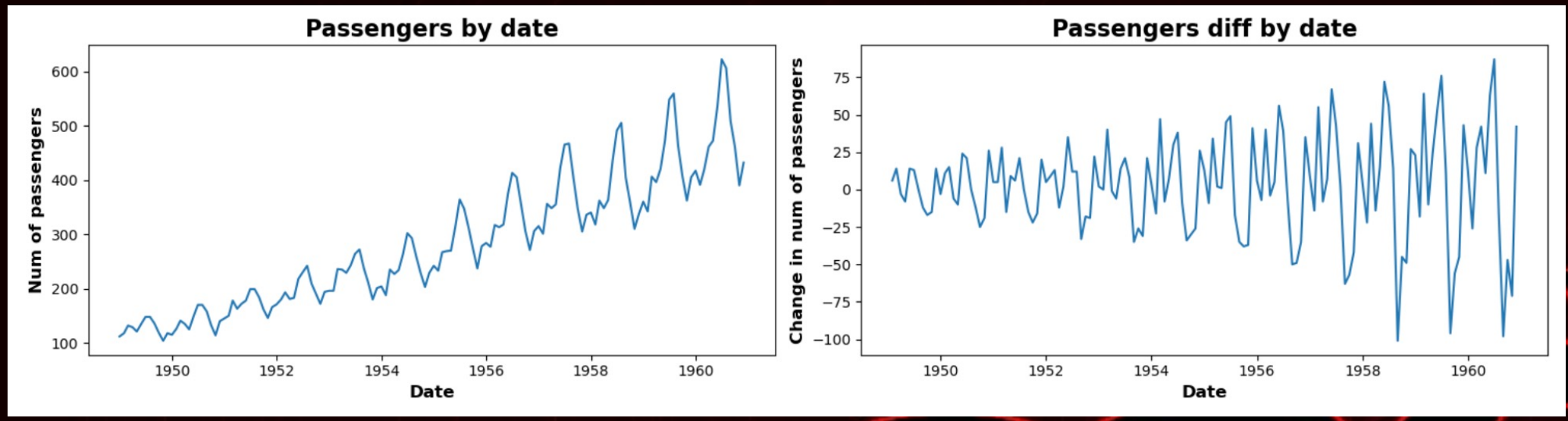
- AR – autoregressive(p)
- I – integrated (d)
- MA – moving average (q)



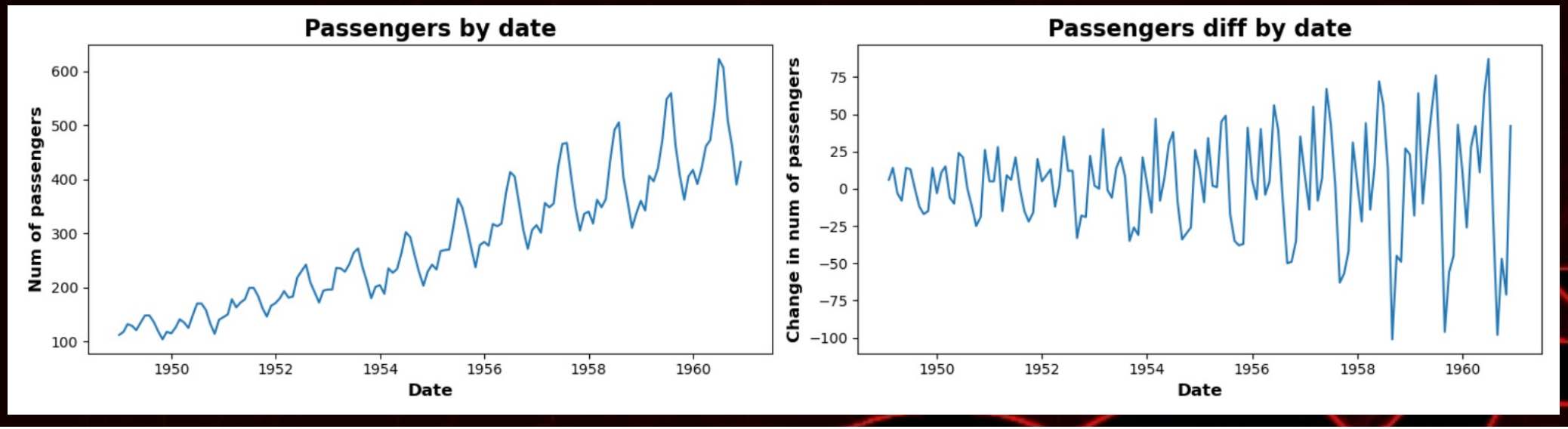
Statisztikai módszerek - differenciálás

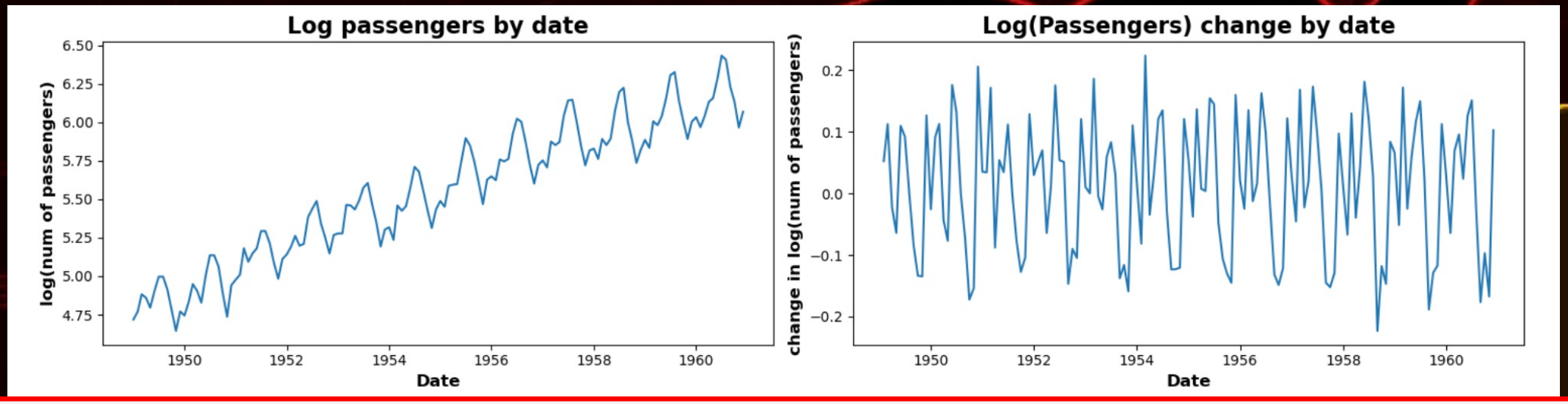


Statisztikai módszerek - differenciálás



Statisztikai módszerek - differenciálás





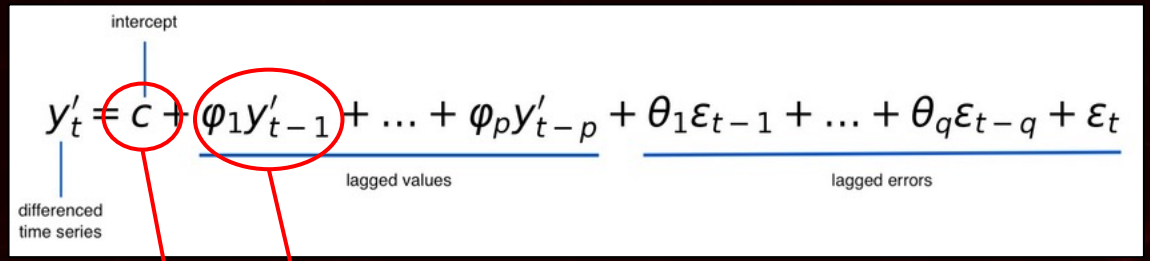
Statisztikai módszerek

AR

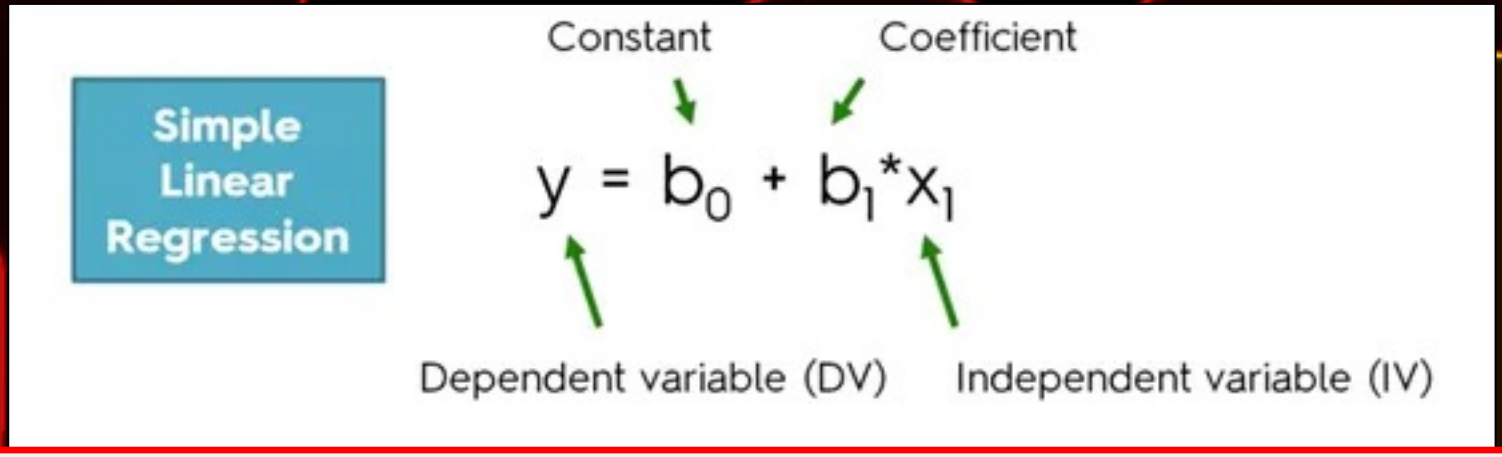
MA

ARIMA (p, d, q)

- AR – autoregressive(p)
- I – integrated (d)
- MA – moving average (q)



Pred = b0 + b1*x1 + b2*x2 … + bn*xn



Statisztikai módszerek

ARIMA (p, d, q)

- AR – autoregressive(p)
- I – integrated (d)
- MA – moving average (q)