

# Odhady sociologických parametrů

Vít Tuček

PMF University of Zagreb, MFF UK

13. listopadu 2020



1. Odhad průměrného počtu kontaktů za týden pro epidemiologické modely.



1. Odhad průměrného počtu kontaktů za týden pro epidemiologické modely.
2. Co o nás říká pohyb našeho chytrého telefonu?



Počet kontaktů za týden je klíčová epidemiologická proměnná.



Počet kontaktů za týden je klíčová epidemiologická proměnná.

*Odhadnout ji lze ze sociologického šetření.*



Počet kontaktů za týden je klíčová epidemiologická proměnná.

*Odhadnout ji lze ze sociologického šetření.*

Ale než dostanu čísla za poslední týden, tak musím čekat dva týdny.



# Problém

Počet kontaktů za týden je klíčová epidemiologická proměnná.

*Odhadnout ji lze ze sociologického šetření.*

Ale než dostanu čísla za poslední týden, tak musím čekat dva týdny.

*Použij data Googlu o mobilitě obyvatel jako proxy proměnnou.*



- ▶ Život během pandemie - PAQ Research
  - ▶ aktualizováno dvakrát do měsíce
  - ▶ sociologický výzkum na 3000 respondentech





- ▶ Život během pandemie - PAQ Research
  - ▶ aktualizováno dvakrát do měsíce
  - ▶ sociologický výzkum na 3000 respondentech
- ▶ COVID-19 Community Mobility Reports - Google
  - ▶ denní granularita, aktualizováno několikrát týdně
  - ▶ odhad založený na sdílení polohy zapnuté v OS Android

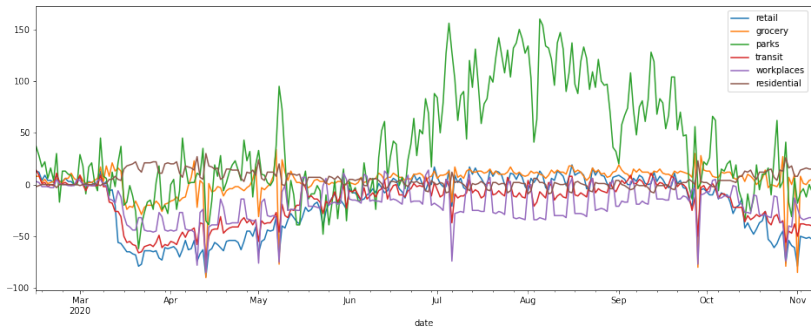


- ▶ Život během pandemie - PAQ Research
  - ▶ aktualizováno dvakrát do měsíce
  - ▶ sociologický výzkum na 3000 respondentech
- ▶ COVID-19 Community Mobility Reports - Google
  - ▶ denní granularita, aktualizováno několikrát týdně
  - ▶ odhad založený na sdílení polohy zapnuté v OS Android

Poděkování: Daniel Prokop, Eva Hromádková, Milan Zajíček



# Google: COVID-19 Community Mobility Reports



## Podproblém

Data z Googlu mají několik dní zpoždění a rád bych měl odhad týdenního průměru hned. Už je úterý a Google mi pořád ukazuje čísla jen do minulého pátku!



## Podproblém

Data z Googlu mají několik dní zpoždění a rád bych měl odhad týdenního průměru hned. Už je úterý a Google mi pořád ukazuje čísla jen do minulého pátku!

*Extrapoluj data z Googlu do konce týdne.*

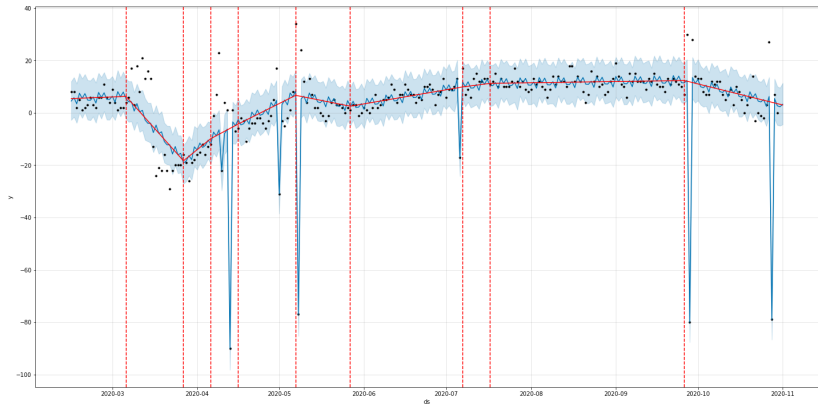


# Extrapolace - FB Prophet

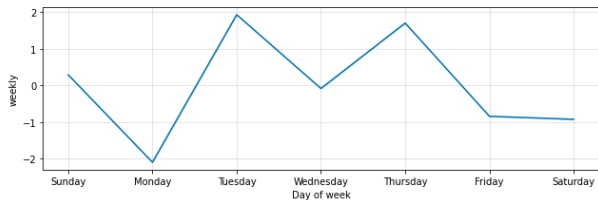
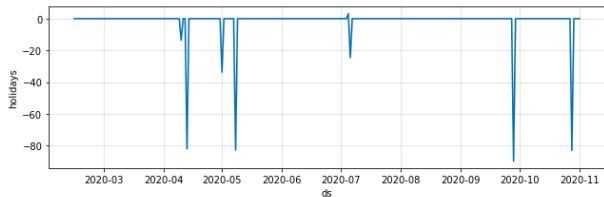
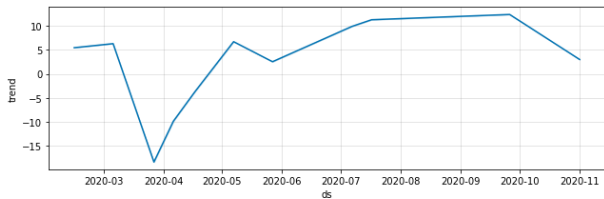
- ▶ státní svátky, uživatelem zadané významné dny
- ▶ parametr pro trade-off mezi trendovostí a sezónností
- ▶ možnost přidání externí kontrolní proměnné



# FB Prophet - grocery

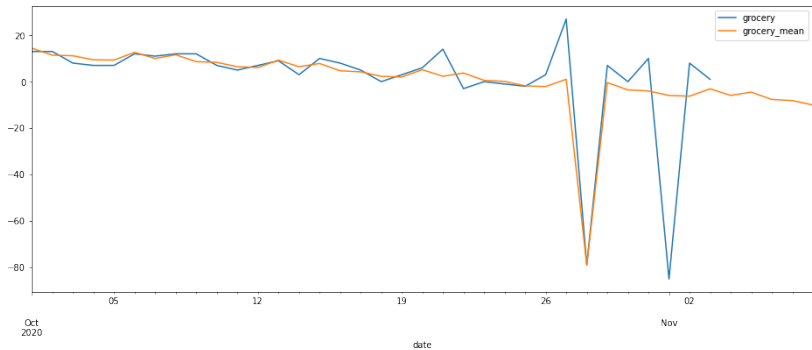


# FB Prophet - grocery





# Extrapolace - grocery



# Vztah mezi daty PAQ a Googlu

Datasets zachycují rámcově stejné aktivity.

Volíme prostou lineární regresi:

$$PAQ\_nakupovani \sim retail + grocery$$

$$PAQ\_park\_mesto \sim parks$$

$$PAQ\_MHD \sim transit$$

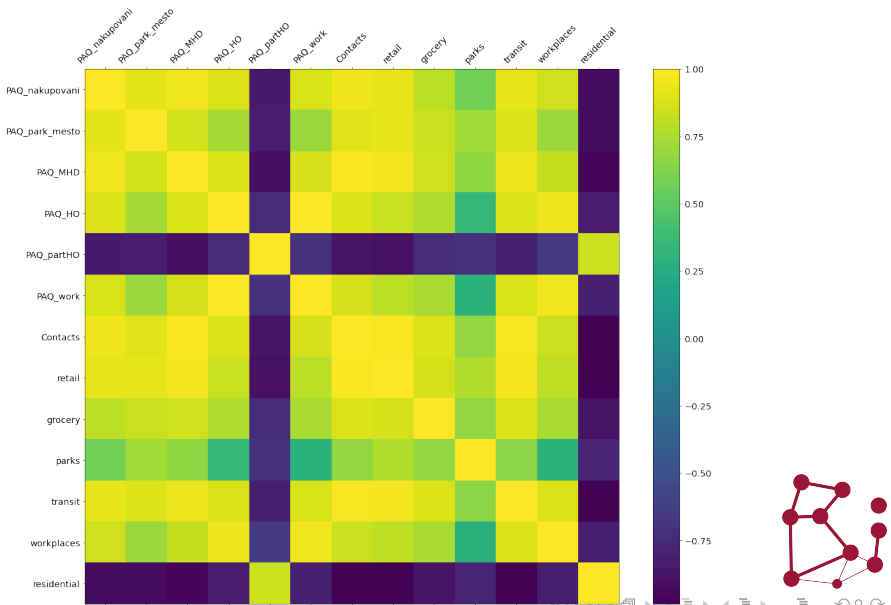
$$PAQ\_HO \sim workplaces + residential$$

$$PAQ\_partHO \sim workplaces + residential$$

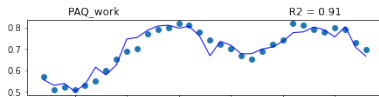
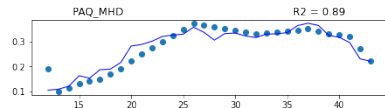
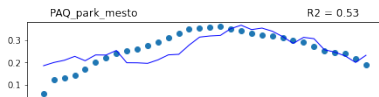
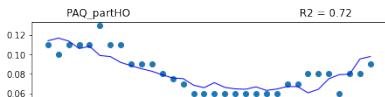
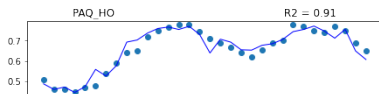
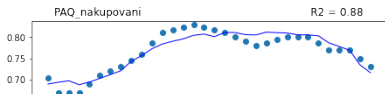
$$PAQ\_work \sim workplaces$$



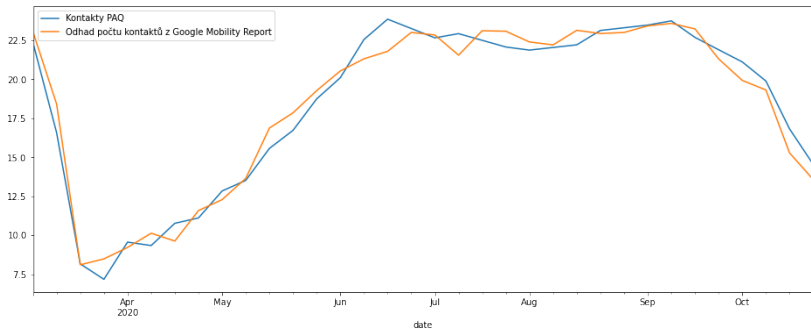
# Korelace mezi PAQ a Google



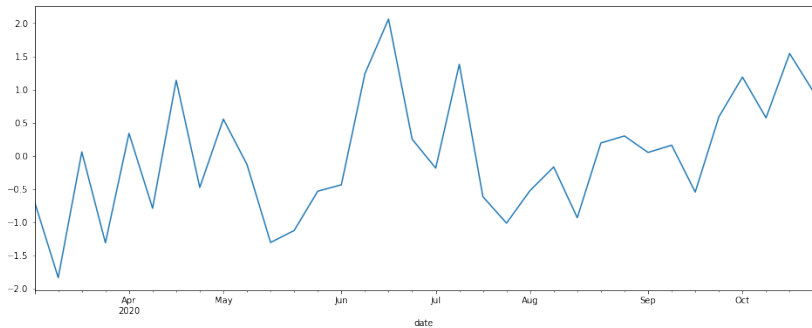
# Lineární regrese



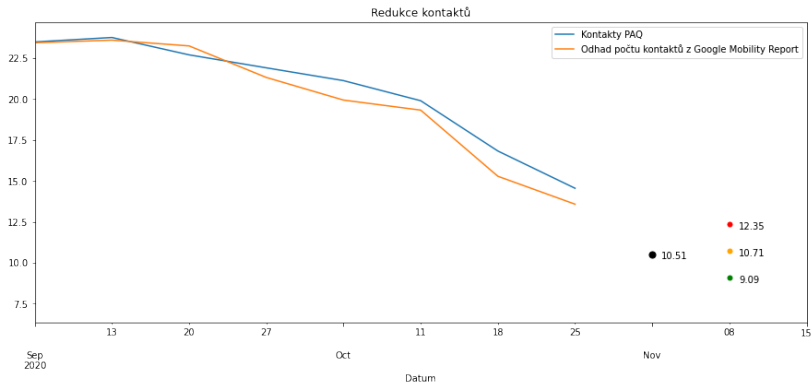
# Lineární regrese - kontakty



# Výsledek



# Výsledek



Děkuji za pozornost!





# Extrapolace - FB Prophet

- ▶ aditivní regresní model  $y(t) = g(t) + s(t) + h(t) + \epsilon_t$
- ▶ po částech lineární / logistický trend

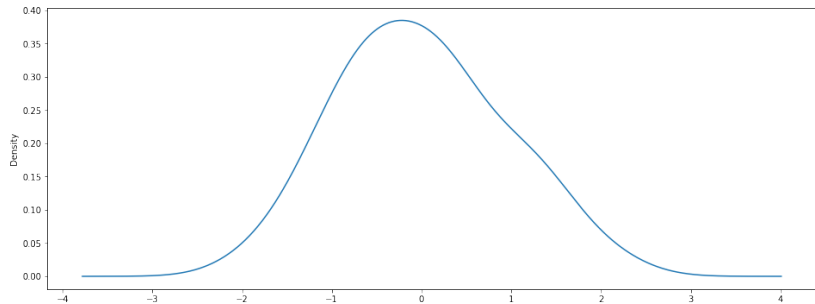
$$y = t \left( k + \sum_j (H(t - s_j)\delta_j) \right) + m + \sum_j H(t - s_j)\gamma_j$$

- ▶ týdenní sezónní složky modelované dummy proměnnými, fitované přes truncated Fourier series

$$s(t) = \sum_{n=1}^N (a_n \sin(2\pi nt/7) + b_n \cos(2\pi nt/7))$$



# Distribuce reziduí



# Autokorelace reziduí

