

Statistické modelování pomocí reálných dat

SMIS smis-lab.cz/2021/06/07/statisticke-modelovani-pomoci-realnych-dat

7. 6. 2021

V posledním roce jsme byli konfrontováni s celou řadou tzv. matematických modelů šíření viru SARS-CoV-2. Kromě snahy predikovat budoucí vývoj se některé z těchto modelů snaží též vyvozovat kauzální (příčinné) závěry ohledně efektů nejrůznějších opatření. Bohužel velká část přístupů (a téměř všechny, jež se objevily v českém mediálním prostoru) trpí mnoha zásadními neduhy. Jednotlivé modely stojí na předpokladech, kterých není zrovna málo a které obvykle nelze ověřit žádnými skutečnými daty. Data, kterými jsou tyto modely „krmeny“ potom vykazují hrubé nereprezentativnosti a nekonzistence, jež znemožňují získané závěry jakkoliv zobecňovat na celou populaci. Jedna ilustrace za všechny: aktuálně si např. [deník echo24](#) povšimnul, že ještě před dvěma týdny vládní predikce vývoje epidemie nepočítala s tím, že by týdenní incidence nových nákaz klesla k současným hodnotám pod 25/100 000 dříve než na konci letních prázdnin. Ale jedna další chybná predikce přeci ničemu nevádí, jedeme dál...

Postupně se nicméně objevují v odborné literatuře přístupy hodnocení vývoje epidemie, pod kterými jsou podepsáni odborní statistici, resp. lidé, kteří byli neoddiskutovatelnými experty na modelování náhodných přírodních jevů (jakým epidemie bezesporu je) již dříve než v roce 2020. Jedním takovým je prominentní britský statistik [Simon Wood](#), jehož přístup hodnocení vývoje epidemie byl přijat k publikaci v prestižním [odborném časopise Biometrics](#). Woodovy výsledky pro vývoj epidemie ve Velké Británii (a efekt místních opatření) ve zjednodušené formě jsou ke zhlédnutí na [jeho webových stránkách](#). Krása Woodova přístupu spočívá v jednoduchosti (jak tomu bývá i v jiných oblastech. Wood nepotřebuje předpokládat téměř nic a jako vstupní data používá pouze údaje o úmrtích ve spojitosti s nemocí COVID-19. Tato data jsou ze své podstaty reprezentativní pro populaci zájmu (protože jsou prakticky úplná) a v

rámci jednoho státu též konzistentní (jejich vypovídací hodnota se v čase příliš nemění. Ani jednou z těchto vlastností nedisponují např. rutinně reportovaná data o počtech pozitivně testovaných. Těch několik málo předpokladů Woodova modelu lze navíc pomocí reálných dat validovat. Zaslouhou Roberta Straky s přispěním Arnošta Komárka a několika dalších kolegů se můžete [ZDE](#) podívat, jak dopadá Woodova analýza a hodnocení skutečného, nikoliv jenom počítačovou simulací domnělého, efektu nejrůznějších opatření nad daty z České republiky.

[Statistické modelování pomocí reálných dat | Straka, Komárek](#)
[Stáhnout](#)

