

Počiatkové šírenie epidémie

Keywords: kombinatorika, pravdepodobnosť a štatistika, štatistika, zpracovanie dát, aritmetický priemer, geometrický priemer, regresná analýza, GeoGebra

Najznámejším prípadom epidémie z posledného obdobia je pandémia vírusového ochorenia COVID-19, ktoré je spôsobené koronavírusom SARS-CoV-2. Táto epidémia vznikla v decembri 2019 v meste Wu-chan (provincia Chu-pej, centrálna Čína). Začiatok epidémie COVID-19 v Európe sa datuje na január 2020. Prvé potvrdené prípady boli hlásené vo Francúzsku dňa 24. januára 2020. Išlo o troch pacientov, ktorí sa nedávno vrátili z Číny, kde už bolo ochorenie v exponenciálnej fáze šírenia. Do Českej republiky sa ochorenie rozšírilo pred 1. marcom 2020, keď boli potvrdené prvé tri prípady infekcie. K 18. marcu 2020 bolo evidovaných už 464 prípadov. Z matematického hľadiska má šírenie infekčného ochorenia v ranom štádiu obvykle exponenciálny charakter. V neskorších fázach dochádza k spomaleniu rastu a ďalej už má iný priebeh (lineárny, logistický apod.).

Typický priebeh epidémie

Na základe získaných dát sa pokúsime namodelovať funkciu : počet nakazených osôb v závislosti od počtu dní od začiatku epidémie. V tabuľke sú poskytnuté hodnoty ktoré vyjadrujú počet nakazených v jednotlivých dňoch.

Deň (n)	Počet nakazených (a_n)
1	3
2	3
3	5
4	6
5	9
6	20
7	27
8	33
9	39
10	64
11	95
12	117
13	142
14	190
15	299
16	384

Poznámka. Jedná sa o reálne dáta z Českej republiky od 1. marca 2020.

Úloha 1. Vypočítajte pomer počtu nakazených osôb v jednom dni vzhľadom na predchádzajúci deň.

Riešenie. Označíme q_n pomer nakazených v dni n ku počtu nakazených v predchádzajúcom dni $n - 1$. Potom platí vzťah

$$q_n = \frac{a_n}{a_{n-1}}, \quad n \geq 2.$$

Takto môžeme vypočítať všetky údaje a doplniť tabuľku.

Deň (n)	Počet nakazených (a_n)	Pomer počtu nakazených
1	3	—
2	3	1,000
3	5	1,667
4	6	1,200
5	9	1,500
6	20	2,222
7	27	1,350
8	33	1,222
9	39	1,182
10	64	1,641
11	95	1,484
12	117	1,232
13	142	1,214
14	190	1,338
15	299	1,574
16	384	1,284

Vypočítanú hodnotu pomeru počtu nakazených v jednom dni vzhľadom na predchádzajúci deň môžeme interpretovať ako rýchlosť šírenia ochorenia. Pri presne exponenciálnom raste alebo poklese by tento pomer vyšiel konštantný. My pracujeme iba s „nameranými“ dátami, takže hodnota pomeru vychádza iba približne. Pozrime sa preto na tento pomer podrobnejšie.

Úloha 2. Vypočítajte aritmetický a geometrický priemer pomeru počtu nakazených. Ktorý z nich je v tomto prípade vhodnejší?

Riešenie. Aritmetický priemer $\bar{q}$ je daný vzorcom

$$\bar{q} = \frac{q_2 + \dots + q_{16}}{15} = 1,40731.$$

Podobne pre geometrický priemer G platí

$$G = \sqrt[15]{q_2 \cdot \dots \cdot q_{16}} = 1,38191.$$

Lepšie je v tomto prípade použiť geometrický priemer.

Aritmetický priemer môže byť ovplyvnený extrémnymi hodnotami (pre nás napríklad hodnota q_6), čo môže skresliť interpretáciu priemerného rastu alebo poklesu. Geometrický priemer naopak tento vplyv zmierňuje. Navyše v prípade exponenciálneho rastu (ak sa obmedzíme na základný tvar funkcie $k \cdot a^x$) sa hodnoty znásobujú stále tou istou hodnotou – teda konštantným koeficientom rastu. Geometrický priemer presne vystihuje túto násobnú povahu zmien. Napríklad, ak máme dve po sebe idúce merania s pomermi

$$q_1 = 2, \quad q_2 = 0,5,$$

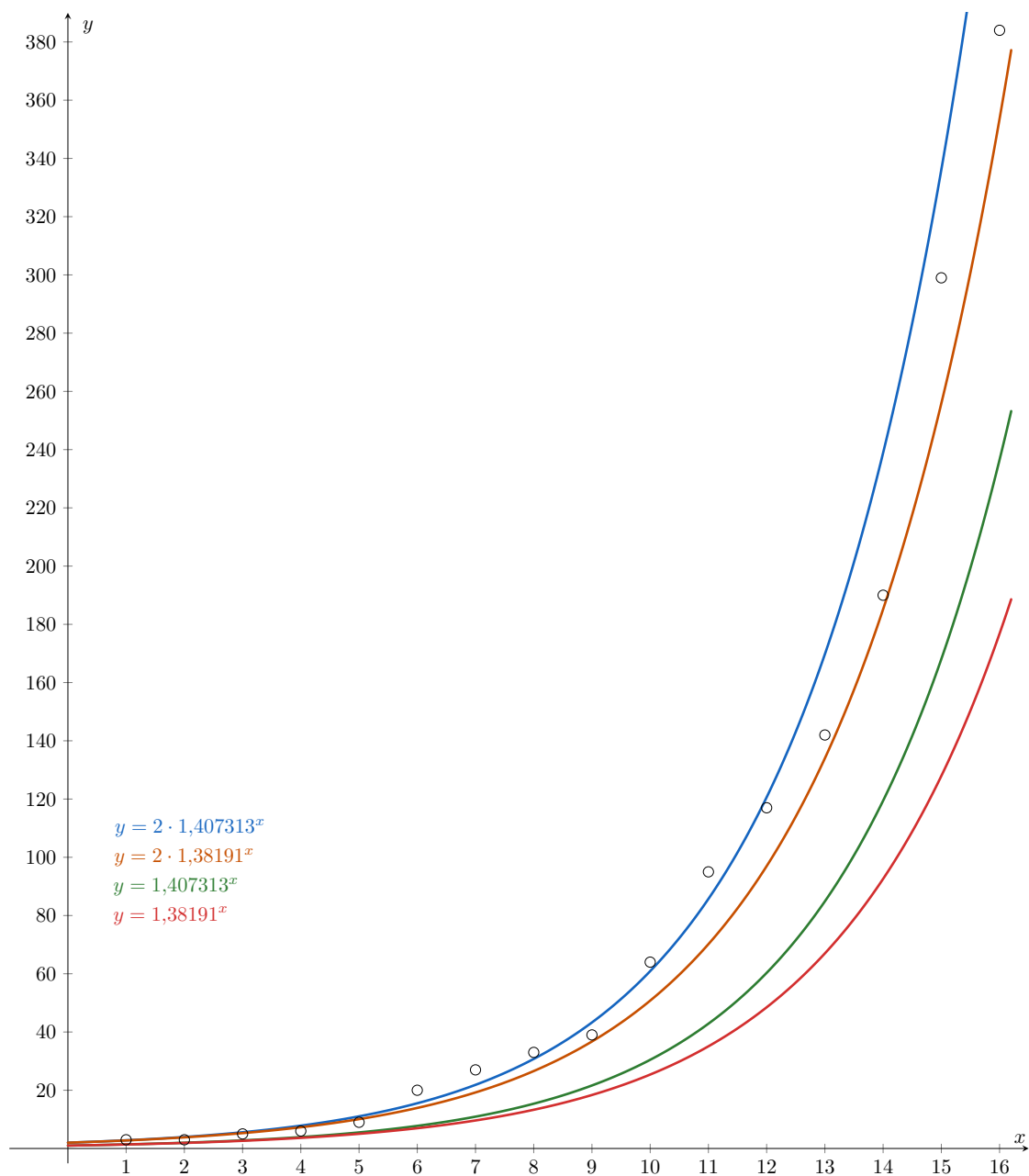
aritmetický priemer $\bar{q} = 1,25$ by naznačoval rast, zatiaľ čo geometrický priemer $G = 1$. správne ukazuje, že výsledný čistý rast je nulový.

Úloha 3. Navrhňte funkciu, ktorá by aproximovala počet nakazených v jednotlivých dňoch. Vytvorte graf v Geogebre (alebo inom programe).

Results matter!

Riešenie. Metódou pokus–omyl môžeme zostaviť exponenciálnu funkciu, ktorá vhodne reprezentuje naše dáta, napríklad použitím vypočítaných priemerov. Použijeme obvyklé značenie používané aj v Geogebre a na obrázku nižšie. Na grafe x označuje čas v dňoch a y počet nakazených.

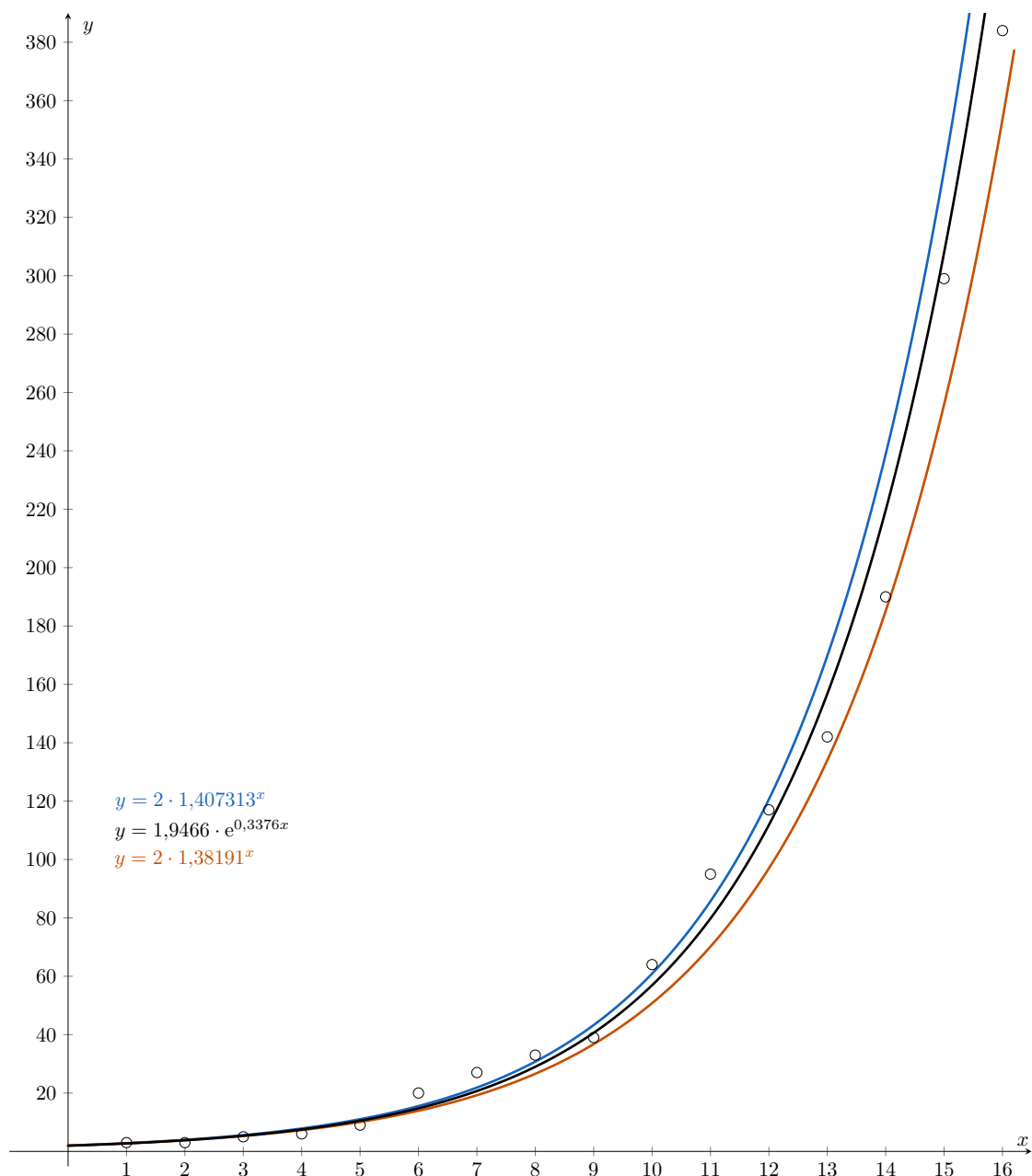
S aritmetickým priemerom dostaneme funkciu $y = 1,40731^x$, s geometrickým priemerom funkciu $y = 1,38191^x$. Tieto funkcie sú od zobrazených bodov pomerne vzdialené, preto ich mierne upravíme: $y = 2 \cdot 1,40731^x$, $y = 2 \cdot 1,38191^x$. Grafy navrhovaných funkcií sú zobrazené na príslušnom obrázku. Samozrejme, je možné získať aj presnejšie exponenciálne funkcie prispôbené dátam.



Obr. 1: Navrhnuté exponenciálne funkcie metódou pokus omyl

Úloha 4. Použitím regresnej analýzy údajov sa získa vhodnejšia funkcia, ktorá popisuje priebeh počtu nakazených, a to v tvare $y = 1,9466 \cdot e^{0,3376x}$. Porovnajte vo Geogebre (alebo inom vhodnom softvéri) vašu navrhnutú funkciu s touto funkciou. Vypočítajte hodnoty všetkých funkcií pre dni 14 až 16, výsledky zaokrúhlite na celé čísla. Porovnajte ich s údajmi v tabuľke. *Poznámka.* Exponenciálnu funkciu získanú regresnou analýzou možno vytvoriť v tabuľkovom procesore alebo v Geogebre. V Geogebre sa zadávajú body z tabuľky príkazom: `RegreseExponencialni({(1,3), (2,3), (3,5), ..., (16,384)})`.

Riešenie. Všetky tri funkcie sú zobrazené na nasledujúcom obrázku. Pohľadom vidíme, že na začiatku sú funkcie veľmi podobné, ale pre ďalšie hodnoty funkcia $y = 1,9466 \cdot e^{0,3376x}$ vyzerá ako najvhodnejšia na popis daných dát.



Obr. 2: Porovnanie navrhnutých funkcií s funkciou podľa regresnej analýzy

Túto skutočnosť môžeme kvantifikovať pomocou údajov z tabuľky:

Deň (n)	Počet nakazených (a_n)	Hodnoty podľa funkcie $y = 2 \cdot 1,40731^x$	Hodnoty podľa funkcie $y = 2 \cdot 1,38191^x$	Hodnoty podľa funkcie $y = 1,9466 \cdot e^{0,3376x}$
14	190	239	185	220

Deň (n)	Počet nakazených (a_n)	Hodnoty podľa funkcie $y = 2 \cdot 1,40731^x$	Hodnoty podľa funkcie $y = 2 \cdot 1,38191^x$	Hodnoty podľa funkcie $y = 1,9466 \cdot e^{0,3376x}$
15	299	336	256	308
16	384	473	354	432

Ako ukazovateľ presnosti, s akou naše funkcie reprezentujú dané dáta, môžeme použiť súčet absolútnych odchýlok od skutočných hodnôt:

$$|239 - 190| + |336 - 299| + |473 - 384| = 175$$

$$|185 - 190| + |256 - 299| + |354 - 384| = 92$$

$$|220 - 190| + |308 - 299| + |432 - 384| = 87.$$

Často sa však pre hodnotenie presnosti používajú druhé mocniny (štvorce) rozdielov, pretože umocňovanie znižuje vplyv malých rozdielov (menších ako 1) a naopak zväčšuje vplyv veľkých odchýlok:

V tomto prípade dostávame

$$(239 - 190)^2 + (336 - 299)^2 + (473 - 384)^2 = 11691$$

$$(185 - 190)^2 + (256 - 299)^2 + (354 - 384)^2 = 4174$$

$$(220 - 190)^2 + (308 - 299)^2 + (432 - 384)^2 = 3285.$$

Z tohto vyplýva, že ako najpresnejšia sa ukazuje tretia krivka, čo nie je náhoda, keďže táto krivka je výsledkom regresie, ktorá minimalizuje daný typ chyby.

Úloha 5. Bude sa epidémia šíriť neustále podľa navrhutej funkcie? Čo môže ovplyvniť jej ďalší priebeh?

Riešenie. Pre zachovanie exponenciálneho priebehu epidémie by sa podmienky šírenia ochorenia nesmeli meniť v čase. V praxi však dochádza k zmene správania populácie – ľudia začínajú používať ochranné prostriedky, minimalizujú sociálne kontakty alebo sú očkovaní, čo vedie k zníženiu rýchlosti rastu počtu nakazených. Epidémia tak dosiahne vrchol a počet nakazených začne klesať. Pre trvalý exponenciálny rast by zároveň musel existovať neobmedzený počet jedincov, ktorí môžu byť infikovaní.

Literatúra

- European Centre for Disease Prevention and Control *Novel Coronavirus: three cases reported in France*. Dostupné z https://www.flickr.com/photos/ecdc_eu/50321985653/in/dateposted/ [cit. 08.08.2024]
- Ministerstvo zdravotníctví ČR. *Onemocnění aktuálně - COVID 19*. Dostupné <https://onemocneni-aktualne.mzcr.cz/api/v2/covid-19> [cit. 30.8.2023]