

STATISTICKÉ DNY 2023 NA BŘEHU TICHÉ ORLICE

19. – 21. KVĚTNA 2023
PENZION MÍTKOV

Abstrakty

Pan(dem)ické milníky podle statistika z Karlína

Arnošt Komárek (MFF UK, Praha)

Jak se vyvíjela pandemie onemocnění covid-19, vyvíjely se též přístupy a možnosti jednotlivých aktérů k „boji“ s ní. Zjednodušeně lze říci, že jaro 2020 a boj s první fakticky nevlnou covidu (mluvíme-li o českých zemích) řídila exponenciála pána z pojišťovny, který přišel a potom už nepřišel ve spojení s červeným svetrem pána, který byl a už není ministrem. Po letní pauze a „vítězství“ nad covidem přišlo čekání na spásnou vakcínu zpestřené zlatým věkem tzv. matematických modelů kmene náčelníka Stolová hora, behaviorálního ekonoma a příležitostného dramatika či matematika z Oxfordu. Hlavním cílem tohoto období bylo pomocí pečlivých matematických výpočtů vybalancovat seznam sortimentu a služeb, jenž je povoleno prodávat či nabízet, správně nastavit normu na výšku stropů a v závěru též správně nastavit systém rotací školních dětí, který minimalizuje přenos viru mezi covidem neohroženými školáky a současně maximalizuje přenos viru směrem do seniorní ohrožené populace. Vyvrcholením této éry (během které nám virus poprvé významněji zmutoval z původní wuchanské varianty do varianty britské, korektně zvané alfa) byla blokáda okresních hranic a nutnost při sportu dělat kolečka, čímž se dodrželo nařízení b..... Mezitím dochází ke dvěma významným změnám ve vybavení bojovníků s pandemií:

- (1) testy na stopy po přítomnosti viru SARS-CoV-2 se staly široce dostupnými,
- (2) na trh byla uvedena nová genová terapie, jež se měla stát prevencí onemocnění covid-19, prý snad i prevencí nakažení virem SARS-CoV-2, která se z čistě obchodních důvodů začala nazývat vakcinací.

Záměna pojmu „(někdy v minulosti) nakažený“ za pojem „nakažlivý“, ignorace zavedených laboratorních postupů, základů imunologie a téměř 300 let starého matematicko-pravděpodobnostního vzorce dala postupně vzniknout konceptu „bezinfekčnosti“ a obchodu s propustkami (nebo odpustky?). V průběhu léta 2021 probíhá marný "boj o protilátky" spolu s nesmyslným hledáním jejich „protektivní hladiny“. A to vše v době, kdy nejenom základy imunologie platné do roku 2020, ale i v ČR dostupná data jasně ukazují, že nejde o hladinu, ale o předchozí přežití setkání s virem. Z věhlasných matematických modelářů se průběžně stávají zkušení biostatistickí, kteří i nadále znásilňují observační data, nyní navíc i Sira Davida Coxe, místo efektů opatření nyní hodnotí účinnost tzv. vakcín. S podzimem 2021 přichází na scénu mutace delta, tisíce dětí uzavíraných nezákonně, mnohdy opakovaně a hlavně úplně zbytečně v domácím vězení. Heslem doby se stává více testovat, lépe trasovat, více očkovat. Mění se vláda, nikoliv však přístup ani osoby a obsazení na postech řídicích důstojníků štábu boje s pandemií. Přichází omikron, poslanecký boj o pandemický zákon, bez kterého prý nelze pandemii v žádném případě porazit, ale po jeho podpisu prezidentem Zemanem (21. 2. 2022) pandemie v Českých zemích zcela nečekaně již 24. února 2022 končí. Nebo nekončí? Cílem příspěvku je pokus o shrnutí dějů doby minulé (v obecné společnosti o to zatím není zájem) a poskytnutí základu pro následnou diskuzi nad covidovými ozvěnami.



STATISTICKÉ DNY 2023 NA BŘEHU TICHÉ ORLICE

19. – 21. KVĚTNA 2023

PENZION MÍTKOV

Cesta dobou covidovou po boku statistiků

Zuzana Krátká (Sdružení mikrobiologů, imunologů a statistiků – SMIS)

Měla jsem to štěstí, že na cestu covidovou epochou jsem se vydala po boku matematiků a statistiků vyzbrojených nadhledem a odvahou. Domnívám se, že klíčové pro nás ve SMIS bylo zorientovat se v neznámém terénu, zvolit vhodnou strategii a spolehnout se na znalosti z oborů imunologie, virologie a epidemiologie. Zpětně mi přijde, že obor vakcinologie byl jediným, jehož učení dostalo na frak a uvidíme, jestli to nebude začátek jeho konce. Oproti většinové společnosti, která poslouchala jen informace předkládané experty FDA, CDC a EMA, my jsme naše poznání založili na vlastních "průzkumných vrtech". Zpětně žasnu, jak dobře nám posloužily, a hlavně jak se díky nim dařilo zvolit správný směr. Proto jsme uměli mezi stovkami vědeckých publikací s protichůdnými závěry najít ty správné... Moje otázka na statistiky je: "Jak velký musí být průzkumný vrt, aby se o jeho výsledky dalo opřít?" Přišlo mi, že nečekaně malý. Jihočeská studie nám potvrdila mortalitu, havlíčkobrodská studie studentů byla svými výsledky velmi blízko Prosecu, břevnická studie zase potvrdila dlouhodobost protilátek i efekt reinfekcí. Ani jeden z našich vrtů nás nezklamal...

Druhé poučení je, že "švec se má držet svého kopyta". Úspěch našeho týmu spočíval v tom, že si experti různých oborů byli vzájemně oporou a inspirací. Velmi rychle jsem si odpustila jakékoliv komentování publikovaných statistických dat, raději jsem požádala matematiky o analýzu publikovaných studií. Také matematici si imunologii nechali popsat a vysvětlit od imunologů, aby něco nepochopili špatně. Vypadá to jako triviálnost, ale bohužel v posledních letech vidím snahu lékařů i přírodovědců data spočítat svépomocí - vždyť je stačí jen "nasypat do Statistiky" a ono to velmi rychle samo spočítá "signifikanci". Jenže zdravotníci až na výjimky vůbec statistice nerozumí a nechápou rozdíly mezi různými testy, podstatu normálního rozložení, závislostí mezi proměnnými a důsledky různých bias. A už vůbec nerozumí nejrůznějším grafům a grafickým znázorněním, kterých jsou publikace plné. Takže ve finále to skončí tím, že si v publikacích jen přečtou abstrakt a diskusi a akceptují nekriticky jakoukoliv větu, kterou pochopí.

Jak z toho ven? Psát publikace tak jednoduše, aby je pochopil žák gymnázia. Používat grafická znázornění mnohdy lépe ukazující vztahy mezi proměnnými než nicneříkající krabicové grafy. A používat selský rozum raději než metaanalýzy a obrovské počty dat. Život je strašně jednoduchý... a úspěch tkví v detailu.



STATISTICKÉ DNY 2023 NA BŘEHU TICHÉ ORLICE

19. – 21. KVĚTNA 2023
PENZION MÍTKOV

Diskrétní model epidemie a diskrétní informace

Lucia Houfková

Na začátku epidemie nikdo z nás nevěděl, co nás čeká, a hodně lidí věřilo tomu, že 14 dní zůstaneme doma a vše bude opět ok. Během této doby jsem se vrátila k diskrétním modelům časových řad, kde jsem už během března 2020 měla silné podezření, že tehdy zmiňované zplošťování křivky bude záležitost let a že po lockdownu přijde zase růst. Jenže se o tom nemluvilo a já hledala chybu v modelu. Chyba byla ale v časové ose oficiálně prezentovaných grafů, které diskrétně neukazovaly, co přijde po zrušení lockdownu, diskrétně neukazovaly, jak dlouho lockdown musí trvat a diskrétně zanedbávaly náklady lockdownu a to nejen finanční, ale i osobní na promarněném čase života zejména dětí a mladých lidí, u kterých riziko spojené s onemocněním covid bylo nízké. Vše se počítalo pouze na úmrtí v daném čase. Nežítí kvůli lockdownu bylo zanedbáno stejně jako se zanedbala současná hodnota nákladů spojených s "posunutím" epidemie do budoucna a nezohledňovalo se ani riziko toho, jaké očkování a kdy a jestli vůbec se v dostatečném množství povede vyrobit.

Díky velkému nasazení finančního a lidského kapitálu se povedlo v rekordně krátké době dostat na trh očkování, které se po krátkém testování velmi rychle nasadilo a nestandardně se doporučovalo/nutilo bez zohlednění individuálních benefitů a rizik. Klinické testy vakcín byly po 5 měsících na žádost výrobců vakcín předčasně ukončeny těsně předtím, než výrobci zveřejnili informace o rychlém mizení účinku vakcín, čehož si museli všimnout už dříve.... Mnoho informací zůstalo diskrétních, nebo se raději nezjišťovaly vůbec. Jiné informace zase sociální sítě začaly diskrétně mazat a omezovat dosah, aby se nemohly šířit mezi lidmi. Některé informace si cestu našly samy, jiné už nejspíš zjistit nepůjde.

Covidové desatero

Tomáš Fürst (PřF UP Olomouc, SMIS)

Covidová response bude nakonec mít tři fáze:

1. Dočasná ztráta soudnosti
2. Velké Covidové Ticho
3. Wiedergutmachung

Dnes stojíme na přelomu druhé a třetí fáze. Kapitáni covidové response a mainstreamová média se momentálně usilovně snaží předstírat, že covid už není téma, vše je za námi, nějaké chyby se možná udělaly, ale jinak jsme vše společně zvládli. Jenže mnohé bývalé konspirační teorie, dis- a misinformace se mezitím ukázaly býti pravdivé a zpod hladiny se zvolna vynořuje ledovec lidí postižených lockdowny, uzávěry škol, cenzurou, persekucí a zejména experimentální mRNA genovou terapií. Velké Covidové Ticho je tedy jen dočasné a v následujícím období bude třeba, abychom zjistili příčiny této dočasné globální ztráty soudnosti a vyvodili z ní nějaké poučení. K tomu se ovšem musíme shodnout na několika základních faktech ohledně koronaviru a covidu, abychom vykolíkovali prostor, ve kterém lze vést racionální a daty podloženou debatu. Ve své přednášce se pokusím zapíchnout deset takových kolíků.



STATISTICKÉ DNY 2023 NA BŘEHU TICHÉ ORLICE

19. – 21. KVĚTNA 2023

PENZION MÍTKOV

Úskalí zachycení, analýzy a prognózy vývoje ekonomických ukazatelů (nejen) ve světle covid-19

Jakub Fischer (Vysoká škola ekonomická v Praze, Fakulta informatiky a statistiky)

Příspěvek se věnuje vývoji sociálněekonomických ukazatelů ve světle pandemie covid-19. Upozorňuje na překážky, k nimž dochází při rychlých zlomech v ekonomickém vývoji, a to jak v oblasti zachycení vývoje ekonomických ukazatelů (problémy při použití indexů typu SOPR), tak při jejich analýze (neplatnost běžně používaných předpokladů). Kromě jiného se zaměříme na vývoj HDP a dopad změny odvětvové struktury, vývoj výdělků a vysvětlení rozdílnosti vývoje mezd ve mzdové a platové sféře, a na problémy spojené s prognózováním meziročního vývoje cenové hladiny.

Výzvy při aktualizaci regionálních cenových hladin

Jana Fischerová (Vysoká škola ekonomická v Praze), Petr Musil, Jaroslav Kahoun

Náš zájem o toto téma začal již kolem roku 2010. Pomocí upravené metody EKS byly odhadnuty regionální cenové hladiny pro kraje České republiky postupně za roky 2007, 2012 a 2017. Nyní stojíme před úkolem aktualizovat regionální cenové hladiny znovu. Jsme na samém počátku, a tedy začínáme nejprve diskuzí o volbě dalšího „ideálního“ referenčního roku a analýzou dostupných datových zdrojů. Tento příspěvek si klade za cíl zdůraznit některé z výzev, které jsou s tímto cílem spojené, a zamyslet se nad jejich možným řešením. Mimo jiné je potřeba vzít v úvahu změny v ekonomice v posledních letech, způsobené například pandemií nemoci covid-19, válkou na Ukrajině, zvyšováním cen energií nebo změnami trendů v oblasti cenotvorby.

Znáhodněné dotazování

Jaromír Antoch (MFF UK, Praha)

Ve své přednášce se soustředím na tzv. RRT (Randomised Response Techniques). Soustředím se přitom na odhadování populačního úhrnu.



STATISTICKÉ DNY 2023 NA BŘEHU TICHÉ ORLICE

19. – 21. KVĚTNA 2023

PENZION MÍTKOV

Vliv Covidu na výsledky sportovních utkání: Zhoršily prázdné ochozy bilanci domácích týmů?

Petr Volf (ÚTIA AV ČR)

Součástí modelů, kterými se snažíme popsat a predikovat vývoj a výsledky sportovních utkání, bývá i komponenta či parametr zohledňující výhodu domácího prostředí. Budu se zabývat fotbalem či hokejem, ale jistě se to týká i jiných sportů, a nejen míčových. Tato výhoda může souviset se znalostí prostředí, s psychickým nastavením („doma by se mělo vyhrát“), ale hlavně ji vytvářejí fandící diváci. A právě „doba covidová“ umožňuje porovnání, jak silný tento vliv je. Bez diváků či se silným omezením jejich počtu se hrála celá sezóna 2020/21. Porovnávám proto úspěšnost domácích týmů v této sezóně s dvěma sezónami před i po. A to dost jednoduchým způsobem. Nemodeluji síly obran či útoků jednotlivých týmů, jak se to běžně dělá (viz literatura), srovnávám prostě jen počty domácích výher (proher, remíz), nebo počty vstřelených a obdržených branek, po jednotlivých kolech soutěže, bez rozlišování týmů. Zatímco průzkumová a grafická analýza ukazuje na zjevný pokles úspěšnosti domácích týmů (i když úspěšnost je stále nadpoloviční), výsledky statistických testů (hypotéz „sezóna 20/21 se neliší“ v určité charakteristice, se snahou je zamítnout) jsou většinou jen „na hraně“ (typicky s p-hodnotou kolem 0,1). Nakonec, i tento výsledek je cenný, dá se interpretovat i tak, že domácí výhoda nespočívá jen v podpoře obecenstva na stadionu (zvláště když na většinu utkání české fotbalové ligy chodí běžně jen pár tisíc diváků).

Literatura

Dixon, M.J., Coles, S. G. (1997). Modelling association football scores and inefficiencies in the football betting market. Appl. Stat., 46, 265-280.

Karlis, D., Ntzoufras, I. (2003). Analysis of sports data by using bivariate Poisson models. J. R. Stat. Soc. Ser. D, 52, 381-394.

Maher, M.J. (1982). Modelling association football scores. Stat. Neerlandica, 36, 109-118.

Analýza opotřebení turbomotorů leteckých motorů

Zuzana Húbnerová (FME VUT Brno), Jaroslav Juračka

V současnosti nejsou letecké motory kontrolovány pouze v pravidelných intervalech, ale jejich měřitelné parametry mohou být díky digitalizaci sledovány i během každého letu. Sběr dat a následná průběžná vyhodnocování jsou příslibem včasného odhalení počínajících poškození či zanášení a napomáhají tak k prodloužení servisních intervalů motoru či životnosti. Vzhledem ke komplexnosti vazeb mezi sledovanými parametry motoru, různými provozními podmínkami a profily letu není možné aplikovat běžné metody vyhodnocování monitorovaných parametrů. V příspěvku je navržen přístup k odhalení změn ve stavu motoru vycházející z metod regresní analýzy.



STATISTICKÉ DNY 2023 NA BŘEHU TICHÉ ORLICE

19. – 21. KVĚTNA 2023
PENZION MÍTKOV

Logaritmicko-lineární modely v pedagogické praxi

Marie Budíková (Ústav matematiky a statistiky Přírodovědecké fakulty Masarykovy univerzity)
Eva Nováková (Katedra matematiky Pedagogické fakulty Masarykovy univerzity)

V příspěvku prezentujeme vybrané výsledky výzkumného šetření realizovaného na Katedře matematiky PedF MU. Šetření se zúčastnilo 311 žáků 4. a 5. ročníku základních škol v Jihomoravském kraji. Cílem bylo zjistit míru predikce a úroveň sebehodnocení žáků při řešení rutinních a nerutinních slovních matematických úloh a analyzovat vztahy mezi mírou predikce, správností řešení a úrovní sebehodnocení. Na získaná data byly aplikovány logaritmicko – lineární modely, které umožňují systematicky analyzovat vztahy mezi dvěma a více kategoriálními veličinami a modelovat četnosti ve vícerozměrných kontingenčních tabulkách. Analýzy potvrdily naše očekávání, že existuje silná závislost mezi úspěšností řešení, mírou predikce a úrovní sebehodnocení.

Abstract: In this paper, we present selected results of a research investigation carried out at the Department of Mathematics, Faculty of Education, MU. 311 4th and 5th-grade pupils of the South - Moravian region primary schools participated in the investigation. The aim was to determine the level of prediction and the level of self-assessment of pupils in solving routine and non-routine mathematical word problems and to analyze the relationships between the level of prediction, the correctness of the solution, and the level of self-assessment. Log-linear models were applied to the data obtained, allowing us to systematically analyze the relationships between two or more categorical variables and model frequencies in multivariate contingency tables. The analyses confirmed our expectation that there is a strong relationship between solution success, prediction rate, and level of self-assessment.

Vyhodnocování statistických dat pomocí programu Python

Eliška Cézová (ČVUT, Praha)

Vyhodnocování statistických dat pomocí programu Python.



STATISTICKÉ DNY 2023 NA BŘEHU TICHÉ ORLICE

19. – 21. KVĚTNA 2023

PENZION MÍTKOV

Využití rozdělení mLN5 v hydrologii a klimatologii

Ladislav Budík (Český hydrometeorologický ústav, pobočka Brno)

V příspěvku je popisováno modifikované 5parametrické logaritmicko-normální rozdělení mLN5, které vznikne transformací normálního rozdělení. V současnosti se již využívá v hydrologické praxi k modelování křivky překročení hydrologických měření jako rozšíření možností rozdělení LN3. Ukazuje se, že na rozdíl od jiných dosud používaných rozdělení lze rozdělení mLN5 s úspěchem použít i pro modelování křivky překročení denních srážkových úhrnů pro všechny dny, tedy i dny beze srážek, kde jsou nulové hodnoty. Po jisté úpravě rozdělení mLN5 velmi dobře prokládá i data vykazující symetrické rozložení, např. průměrné denní teploty vzduchu a lze ho použít i pro modelování křivek překročení takových veličin, které jsou ohraničená zdola i shora, jako jsou průměrné denní vlhkosti vzduchu.

V příspěvku je popsána i nová metoda odhadu parametrů křivky překročení, tzv. trojúhelníková metoda. Ta respektuje skutečnost, že existují chyby nejen v naměřených datech, ale také nepřesnosti v určení pravděpodobnosti překročení daných hodnot.

Abstract: The paper describes a modified 5-parameter log-normal distribution mLN5 obtained by transforming the normal distribution. It is already used in hydrological practice to model the exceedance curve of hydrological measurements as an extension of the capabilities of the LN3 distribution. Unlike other distributions, the mLN5 distribution can also be successfully used to model the daily rainfall exceedance curve for all days, i.e., even zero-precipitation days. After some modification, the mLN5 distribution also fits very well data symmetrically distributed, e.g., average daily air temperatures, and can be used to model exceedance curves for variables bounded from below and above, such as average daily humidity.

The paper also describes a new method for estimating the parameters of the exceedance curve, the so-called triangular method. This method respects the fact that there are errors in the measured data and inaccuracies in determining the probability of exceedance of given values.

(Mé) strasti a slasti rekreační kombinatoriky

Pavel Stříž

V příspěvku zmíním nedávné úspěchy a výzvy kombinatoriky, především v oblasti deskových a logických her. Poukáži na experiment generování pseudonáhodných čísel $[0, N]$ po cifrách. Na prozkoumání bude k dispozici zdrojový kód pro Python i R. Během příspěvku zmíním řešené a otevřené otázky z této oblasti rekreační matematiky/kombinatoriky.

