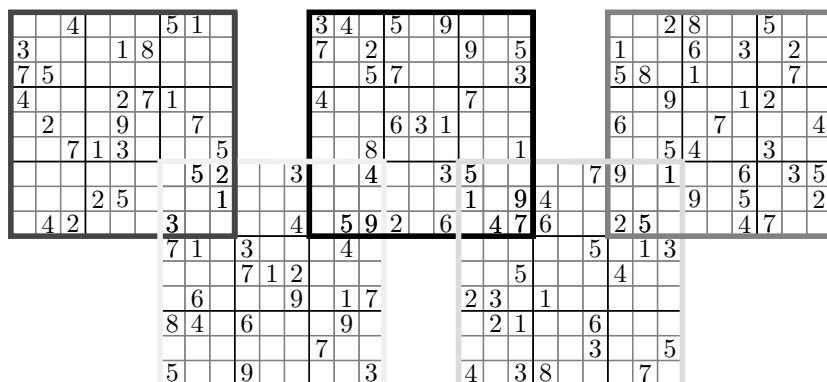


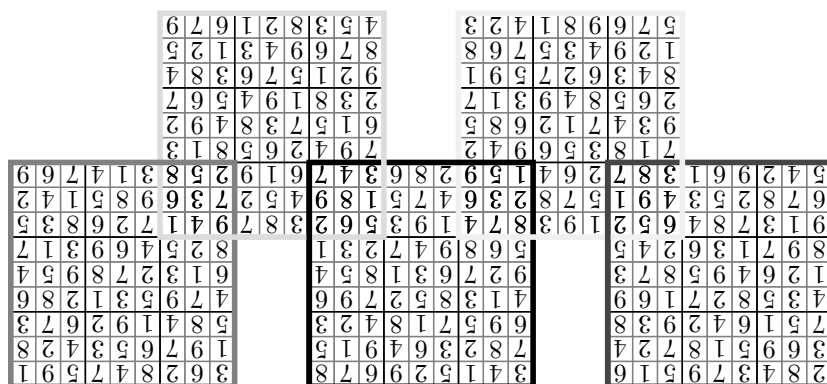
ŠŤASTNÝ NOVÝ ROK 2024! POUR FÉLICITER 2024!

Redakce



Pozvánka na přednášku Pavla Stríže *Sudoku s překryvy* na Statistických dnech 2024 v Perné pod Pálavou. Název sudoku by byl: Těšíme se na XXXIII. olympijské hry v Paříži.

Pro zájemce přikládáme i řešení.



VLIV COVIDU NA VÝSLEDKY SPORTOVNÍCH UTKÁNÍ: ZHORŠILY PRÁZDNÉ OCHOZY BILANCI DOMÁCÍCH TÝMŮ?

IMPACT OF COVID ON RESULTS OF SPORTS MATCHES: DID SPECTATORS ABSENCE DECREASE THE ADVANTAGE OF HOME TEAMS?

Petr Volf

Adresa: Institute of Information Theory and Automation AS CR, Pod vodárenskou věží 4, 182 08 Praha 8

E-mail: volf@utia.cas.cz

Abstrakt: 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 tento jev se týká i jiných sportů, a nejen míčových. Tato výhoda může souviset se znalostí prostředí, s psychickým nastavením („doma by se mělo vyhrát“), ale také 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 2020/2021. 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, srovnávám prostě jen počty domácích výher (proher, remíz), nebo počty vstřelených a obdržených branek, po jednotlivých kolech soutěže, bez rozlišování týmů. Ale přesto nejprve připomenu základní modely pro statistický popis sportovních utkání.

Abstract: The home advantage is a phenomenon having evidently a strong impact on results of sports matches. Therefore the “home advantage” parameter or component is standardly a part of statistical models describing (and forecasting) sports results. I will concentrate here to the soccer and ice-hockey, however this phenomenon concerns other sports as well. One initiator of this advantage is the presence of home spectators. That is why the time of Covid pandemic, when at least one season was played without (or with strong limitation of) spectators, gives a chance to compare the „Covid” season with the others and, possibly, to discover a significant decrease of home advantage impact. This paper uses just rather simple statistical tools, nevertheless, with the aid of basic statistical tests the changes of home teams results (regarding both the gained points and scored goals) have been detected.

1. Úvod, modely zohledňující domácí výhodu

Mám teď na mysli sporty jako fotbal, hokej, kde výsledek je určen na základě počtu branek. Cílem je zpravidla modelovat výsledky jednotlivých utkání, za účelem jejich predikce, tj. i s ohledem na spolehlivé sázení na ně. Takže je třeba zahrnout parametry jednotlivých týmů zohledňující jejich sílu, i další faktory mající vliv na zápas. Nejjednodušší by bylo vytvořit např. trino-mický model pro výsledky výhra, remíza, prohra, třeba logistického typu, který umožňuje popsat i onen vliv různých okolností. Ale od začátku byly také vyvíjeny modely pro skóre zápasů. Samozřejmě, předpověď skóre moc spolehlivá není, hlavním cílem je ale stále předpovídat vítěze.

Nejjednodušší model je třeba následující založený na dvou nezávislých Poissonových náhodných veličinách (např. Maher, 1986, model 2): Označme X_{ij} a Y_{ij} počet gólů domácích a hostů v zápase mezi týmem i (domácí) a j (hosté). Pak

$$\begin{aligned} X_{ij} &\sim \text{Poisson}(\lambda_1 = \alpha_i \beta_j \gamma), \\ Y_{ij} &\sim \text{Poisson}(\lambda_2 = \alpha_j \beta_i), \end{aligned} \quad (1)$$

kde $\alpha_i, \beta_i > 0$, α_i charakterizuje útočnou sílu týmu i , β_i pak sílu obrany. No a γ je parametr domácí výhody. Parametry α_i, β_i musí být vhodně normovány (kvůli jednoznačnosti). Takže i zde se pochopitelně počítá s výhodou domácího prostředí, očekává se $\gamma > 1$. V článku Maher (1986), kde zpracovává data z ročníku 1971/1972 anglické Premier Division, přímo odhad γ není uveden, dá se spočítat z jeho tabulek, vychází kolem 1,38.

Maher si také všímá nesouladu četností výsledků s malým počtem gólů s nezávislým modelem, a navrhuje použít 2-rozměrný Poissonův model. Ten je vybudován takto: Nechť X_1, X_2, X_3 jsou 3 nezávislé veličiny s Poissonovým rozdělením s parametry $\lambda_1, \lambda_2, \lambda_3$. Pak 2-rozměrné Poissonovo rozdělení je tvořeno veličinami $Y = X_1 + X_3, Z = X_2 + X_3$. Jejich kovariance je λ_3 , a mají marginální Poissonova rozdělení s parametry $\lambda_1 + \lambda_3, \lambda_2 + \lambda_3$. Z konstrukce plyne, že korelace je nezáporná. Maherův odhad korelace je 0,2, a model prokazuje skutečně lepší fit.

V dalším článku Dixon, Coles (1997) analyzují sezónu 1995/1996 a 4 nejvyšší anglické fotbalové soutěže (92 týmů). Parametr domácí výhody γ je opět přibližně 1,38, a zároveň ukazují, že je prakticky konstantní během celé sezóny. To znamená, že také sledovali, jak se parametry mění během roku. Nesoulad dat s nezávislým Poissonovým modelem řeší pak použitím „inflat-ed“ modelu, čili směsí základního modelu s distribucí dávající jinou váhu

POZVÁNKA NA STATISTICKÉ DNY 2024 INVITATION TO THE STATISTICAL DAYS 2024

Redakce



Česká statistická společnost, ve spolupráci s Ústavem matematiky, Fakulta strojního inženýrství, VUT, pořádá ve dnech 24. – 26. května 2024 v Penzionu Janota (<https://www.penzionjanota.cz/>) tradiční třídní Statistické dny vyzývající k diskuzím nad tématy spojenými s teorií, aplikací i výukou statistiky. Tentokrát mohou účastníci přednést své příspěvky a přispět k rozpravám obklopeni Pálavskou přírodou v blízkosti okouzlujících vinných sklepů obce Perná.

Zaměření konference. Statistické dny budou zaměřeny na prezentaci výzkumných výsledků, poznatků a zkušeností z vývoje, aplikací a výuky statistických metod z oblasti:

- statistické metody pro řízení jakosti a metrologii,
- modelování spolehlivosti systémů a procesů,
- stochastické modely pro průmyslové a neprůmyslové aplikace,
- optimalizace, diagnostika a hodnocení rizik,
- výuka statistiky,
- statistický software.

Vítaná jsou ale i témata předchozích let jako lékařství, sociologie, ekonomie, finančníctví atd. Finální podoba programu bude stanovena na základě počtu a charakteru přihlášených příspěvků.

Registrace. <https://www.statspol.cz/STATDNY2024/registrace.php>

Důležité termíny. Registrace, účastnický poplatek, abstrakt a příspěvek do 14. března 2024. Konference je možné se zúčastnit i bez vlastního příspěvku. Příspěvky i abstrakt lze posílat na adresu ondrej.vencalek@upol.cz.

Kontaktní email. Martina.Litschmannova@vsb.cz.

Aktuální informace a program budou zveřejňovány na:

<https://www.statspol.cz/STATDNY2024>



Statistické dny v roce 2022: Výlet na Roštejn.



Statistické dny v roce 2023: Sobotní večer.

několika výsledkům, (konkrétně, pravděpodobnosti 0:0, 1:1 jsou v „nezávislém“ modelu podceněny, naopak 0:1, 1:0 se vyskytují méně než z něj vyplývá).

Rodina modelů je opravdu bohatá. Třeba v Volf (2009) jsou modelem pro vývoj skóre 2 zároveň běžící číselné procesy, které se vzájemně ovlivňují prostřednictvím kovariát (v rámci Coxova regresního modelu, jehož parametry také mohou být proměnné v čase). Kovariáty odrážejí stavy týmů i průběžný stav zápasu.

Pokud jde třeba o lední hokej, tam vychází parametr domácí výhody nižší. Např. při analýze dat z české hokejové Extraligy v ročníku 2007/2008 jsem kdysi (opět v rámci modelu 2 nezávislých Poissonových rozdělení pro počet gólů) odhadl $\gamma = 1,23 (1,17, 1,29)$ (Volf, ICPM 2008). Porovnání domácí výhody u více sportů uvidíme hned v další části, v tabulce na Obr. 1.

2. Jak jednoduše měřit domácí výhodu?

Inspirací může být článek Pollard (1986). Ten se soustřeďuje právě na kvantifikaci výhody domácího prostředí, v rámci určité soutěže či ročníku, bez toho, aby modeloval vývoj utkání, sílu týmů a pod. Na začátku formuluje několik možných příčin takové výhody, některé se dají měřit, některé hůř:

1. Divácká kulisa, povzbuzování.
2. Znalost domácího prostředí (domácí hráči mají i zázemí, na které jsou zvyklí).
3. Taktika, domácí chtějí vyhrát.
4. Strannost rozhodčích (i nechtěná, způsobená atmosférou, reakcí diváků).
5. Další psychické faktory (+ i –, třeba i zvýšená nervozita).
6. Únava hostů z cesty, atd.

Dále se Pollard zabývá poměrně jednoduchou metodou: prostě spočte relativní četnost výher domácích týmů, nebo v případě, kdy je víc možností než prostá výhra/prohra, spočte poměrnou část bodů získaných domácími týmy. A pokud je statisticky významně nad 50 %, je to důkaz o existenci domácí výhody a zároveň i její číselné ohodnocení. Pro některé soutěže jsou jeho výsledky v následující tabulce.

V podstatě použijí také takové jednoduché porovnávání, nyní s cílem porovnat bilanci domácích za Covidu (tj. v podstatě bez diváků) s normálními sezónami. Ještě se zmíním o jedné studii vyhodnocující domácí výhodu, od autorů z KMA ZČU v Plzni (Marek, Vávra, 2020). Ti navrhli použít rozdíly

Table 1. Home advantage in different professional team sports.

Sport	League	Seasons	Record of home teams				Home advantage		
			Played	Won	Drawn	Lost	Per cent	s.e.(%)	P*
North America									
Baseball	AL and NL	1982-84	6316	3388	—	2928	53.6	0.6	<0.001
US football	NFL	1982-84	574	314	3	257	55.0 ^b	2.1	<0.01
Ice hockey	NHL	1981/2-83/4	2520	1328	361	831	59.9 ^c	0.9	<0.001
Basketball	NBA	1981/2-83/4	2829	1790	—	1039	63.3	0.9	<0.001
Soccer	NASL	1982-84	512	334	—	178	65.2	2.1	<0.001
England									
Cricket	County Cham- pionship	1981-83	478	162	189	127	56.1 ^b	2.9	<0.05
Soccer	FL Division 1	1981/2-83/4	1386	695	350	341	63.9 ^d	1.5	<0.001

*One-sided test of proportion; null hypothesis of no home advantage.

^bDrawn games excluded.^cTwo points for a win, one for a draw.^dThree points for a win, one for a draw.

Obrázek 1: Tabulka 1 z Pollard (1986).

skóre ze zápasů každé dvojice týmů, i, j , doma i venku, a vytvořit trinomickou veličinu C s hodnotami $-1, 0, 1$ takto:

$$C_{ij} = \text{sign}[(h_i - a_j) - (a_i - h_j)],$$

kde h_i, a_i jsou počty gólů vstřelených týmem i doma (home) a venku (away) v tomto „dvojzápase“ týmů i, j . Tyto hodnoty pak sečetli a zprůměrovali za jednotlivý ročník soutěže, pak případně za více ročníků téže soutěže. Zpracovali takto dost velké množství dat, z 19 evropských fotbalových soutěží (i nižších), za ročníky 2007/2008 až 2016/17, počet dvojic zápasů byl přes 36 000. Je otázka, zda a proč je tento způsob bilance výhodný. Metoda používá poměrně detailní informaci (skóre zápasů), kterou ale nevyužije naplno. Pro domácí výhodu zde svědčí, když je střední hodnota veličiny C významně větší než nula.

3. Analýza vlivu Covidu na fotbalovou Fortuna ligu

Nejdřív se podívám na českou 1. fotbalovou ligu. Budu porovnávat ročníky 2018/2019 až 2022/2023 bez 2020/2021 právě s tímto ročníkem. A to sice základní části těchto ročníků, bez nadstavby. Výsledky zápasů z oněch ročníků lze nalézt na www.sport.cz/sekce/fotbal-ceska-1-liga-vysledky-4.

Jaro 2020 se vlastně dohrálo s diváky, neboť liga byla od poloviny března do konce května přerušena, zatímco ročník 2020/2021 začal s omezením na 2000 diváků, v září se rychle přešlo na maximum 500 a v říjnu byly všechny soutěže přerušeny. Od listopadu se pokračovalo, ale kompletně bez diváků, k pozvolnému uvolnění došlo až v květnu 2021. Nadstavba se už nehrála.

Každá základní část měla 16 týmů, hrál každý s každým doma i venku, takže bylo 30 kol po osmi zápasech. Až právě na sezónu 2020/2021, ta měla 18

Druhý večer byl oproti tomu prvnímu velmi odlehčený. Nápojovou část tentokrát zařizovali kolegové z VUT, kteří mají pro víno očividně talent (ač se sommeliér Matej letošního ročníku bohužel neúčastnil). A co je hlavní – žádné násobení už v těch propočtech nefalšovalo. Vyprávěné historky publikovatelné i nepublikovatelné byly proloženy písněmi Zdeňka Fabiána, jehož nevyčerpatelný repertoár nás provázel večerem až do pozdních hodin.

Závěrečné dopoledne bylo věnováno velice různorodým tématům. Marie Budíková povídala o míře predikce a úrovni sebehodnocení žáků při řešení matematických úloh, to vše zabalené do logaritmicko-lineárního modelu. Čas dodržet musela, Ladislav Budík totiž svůj příspěvek o modifikovaném LN5 rozdělení a jeho použití v hydrologii a klimatologii prezentoval až po Elišce Cézové (*Vyhodnocování statistických dat pomocí programu Python*), na SJM tedy tentokrát nikdo nezískal ani minutu. Závěrečná přednáška Statistických dnů 2023 patřila Pavlu Střížovi, jehož nadšení pro pravděpodobnost a kombinatoriku nemohl popřít nikdo ze zúčastněných.

Svíčková je zřejmě typické zakončovací jídlo statistických konferencí. Nebo to byla náhoda? Ze dvou pozorování se závěry dělají těžko, snad bude mít autor příležitost sbírat data i v budoucnu. Už teď vypadá příští setkání velmi slibně, kolegové z VUT chtějí prohloubit už tak nesmazatelnou stopu, kterou na Statistických dnech zanechali, a chopili se organizace ročníku 2024. Lokality nemohli vybrat lépe – své kolegy totiž pozvou k úpatí Pavlovských vrchů do malebné obce Perná. Že by se soutěž *Poznej víno!* konala znovu a v nesrovnatelně větší kvalitě podpořené lokálními produkty? A sesadí někdo Mateje Benka z trůnu, pokud by se znovu účastnil? Na to si budeme muset počkat.

Organizátorům konferencí v Telči i penzionu Mítkov patří velký dík za skvěle odvedenou práci a obnovení tradice Statistický dnů. Vášnivě diskuze, zajímavé příspěvky, objevování zákoutí ČR i společné večery plné smíchu a hudby – za to všechno stálo se znovu potkat!

Reference

- [1] SMOLJAK, Ladislav. *Dobytí severního pólu Čechem Karlem Němcem 5. dubna 1909*. Praha: Paseka, 1993. Divadlo Jára Cimrmana. ISBN 80-851-9252-7. cit. 24

a rozebrat je z různých stran. Celou covidovou pandemii z pohledu vědy krásně shrnul Michal Kulich a ukázal jasné rozdíly mezi tím, co vědecká komunita dělat měla a co ve skutečnosti dělala. Páteční sekci uzavřel novinář Petr Šourek příspěvkem o vlivu a práci médií v době covidu.

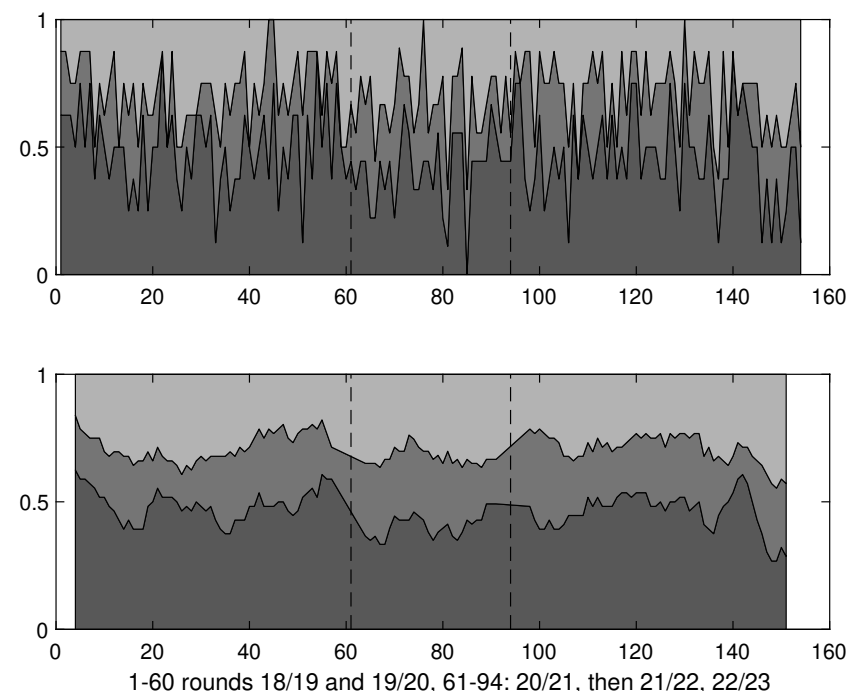
Po večeri se účastníci sešli opět v přednáškovém sále – je čas diskuze! Plynuje se navázalo na prezentace odpoledních řečníků, některá témata se dál rozvinula (např. kritické přijímání informací, vzdělávací systém...), celkově došlo k zajímavému zhodnocení posledních let i nástinu toho, co by bylo potřeba udělat – konkrétní kroky však nezazněly. Až případná další krize ukáže, jak a zda se (nejen) statistická obec poučila z let 2020–2022 a jak zareaguje.

V sobotu ráno jsme se po bohaté snídani, kterou nám majitelé penzionu přichystali, sešli znovu v hlavní místnosti celé konference. Venku se začínalo dělat krásně, my však měli až do oběda nabitý program. Jakub Fischer nám osvětlil, jak to vlastně bylo s vlivem covidu na ekonomiku, následně se dotkl i financování vysokých škol, což vzhledem k proporcí účastníků z univerzitní půdy vyvolalo značný ohlas. Jana Fischerová pak i přes zkrácený časový prostor zvládla popsat výzvy v regionalizaci některých ekonomických ukazatelů (čas vyhrazený na přednášku zřejmě také spadá pod SJM¹, žádné rozbroje proto nevznikly). Následoval Jaromír Antoch s přednáškou o znárodněném dotazování. Po kávě jsme se od Petra Volfa dozvěděli něco o vlivu domácí podpory při sportovních utkáních – covid totiž přinesl unikátní příležitost zjistit, jaké výkony týmy podávají, když jsou sice na domácím hřišti, avšak bez fanoušků. Poslední příspěvek tohoto dne byl věnován analýze opotřebení leteckých motorů, které se věnuje Zuzana Hübnerová z VUT.

Krásné počasí naštěstí vydrželo, odpoledne byl totiž v plánu výlet „ve stopách Jana Ámose Komenského“ – ten totiž část života strávil v nedalekém Brandýse nad Orlicí. Kromě výhledu ze zříceniny hradu Brandýs nám byl nabídnut také nevšední zážitek – cesta „Labyrintem světa a rájem srdce“, tedy přírodním bludištěm u Komenského obelisku². Dostat se do jeho středu byl oříšek, dostat se z něj zpátky ven možná ještě větší výzva. Výprava se o to snažila rozdělena na malé skupinky – často se stávalo, že jedna skupinka šla „poslední možnou“ a „zaručenou“ cestou, kterou už se ven přece musí dostat, načež se srazila s jinou, smutně kývající skupinkou. Tak tudy zase ne? Ach jo. Naštěstí se z bludiště podařilo dostat všem zúčastněným. Po zpáteční procházce vyhládko opravdu všem, v penzionu se o nás však náramně starali.

¹Společné jmění manželů.

²Více se o tomto zajímavém místě můžete dozvědět zde:
<http://www.brandys-ve-svete.cz/labyrint/labyrint.php>.



Obrázek 2: Proporce domácích výher (spodní, nejtmaší pole), remíz, proher (horní, nejsvětleší pole) po jednotlivých kolech v sezónách 2018/2019 až 2022/2023. Dole je totéž vyhlazeno klouzavým oknem šířky 7.

týmů, takže se hrálo 34 kol po 9 zápasech. Nezabývám se jednotlivými týmy, nejmenší jednotkou je pro mne 1 odehrané kolo, takže bilanci kol 2020/2021 jsem přepočítal jakoby na 8 zápasů. Ještě zde podotknu, že grafy mají formu časových řad (časem je index kola, což většinou odpovídá periodě 1 týden), ale jde v podstatě o realizaci nezávislých náhodných veličin (stejně rozdělených v každé z obou skupin). To bylo v obou skupinách otestováno testy nezávislosti (série nad a pod mediánem, série nahoru a dolů).

Nejdřív jsem se zabýval počty výher domácích, remíz a výher hostů v jednotlivých kolech. Graf těchto počtů je na Obrázku 2. V dolní části je vždy zprůměrováno nejbližších 7 kol. Z obrázku se nedá usoudit na nějaký výrazný

pokles bilance domácích, to se dá udělat až pomocí porovnání proporcí, nejlépe analýzou kontingenční Tabulky 1.

	Výhry	Remízy	Prohry	Součty	Proporce		
Ostatní	458	227	275	960	0,4771	0,2364	0,2865
2020/2021	126	80	100	306	0,4118	0,2614	0,3268
Součty	584	307	375	1266			

Tabulka 1: Četnosti a relativní četnosti domácích výher, remíz, proher.

Z této tabulky nedostaneme statisticky významný výsledek, p -hodnota příslušného chí-kvadrát testu nezávislosti v kontingenční tabulce je 0,1341. Pokud však vezmeme domácí výhry proti ostatním výsledkům, rozdíl proporcí 0,4771 proti 0,4118 už je statisticky významný, test nezávislosti v 2×2 kontingenční tabulce má p -hodnotu 0,0459. Takže lze aspoň takto považovat tezi o zhoršené bilanci v domácích zápasech za prokázanou.

Pokud vycházím jen z bilance výher, remíz a proher, mohu je ohodnotit body (3 body za výhru, 1 bod za remízu, 0 za prohru) a tedy porovnávat body získané domácími týmy v obou obdobích. Obrázek 3 ukazuje, jak to vypadalo: V každém kole jsem spočítal počet bodů získaných domácími, relativně na 1 zápas. Takto získaná veličina má základní charakteristiky v Tabulce 2, a test normalitu jejího rozdělení nezamítá (použil jsem Lillieforsův test, tj. „zprůsňený“ Kolmogorovův Smirnovův test, zvlášť pro sezónu 2020/2021 a pro ostatní ročníky).

Sezóna	Průměr	Směr. odchylka	Medián	n je počet kol
Ostatní sezóny	1,6677	0,4582	1,6250	120
2020/2021	1,4967	0,3966	1,5000	34

Tabulka 2: Charakteristiky počtu bodů získaných domácími za 1 zápas.

Na Obrázku 3 vidíme průměr i kontrolní linii (ve vzdálenosti 2 a 3 směrodatné odchylky od průměru) odpovídající této veličině v sezónách mimo 2020/2021. Pokud udělám t -test hypotézy rovnosti středních hodnot, příslušná p -hodnota je 0,0501. „Hrubší“ Wilcoxonův test na rovnost mediánů pak dá p -hodnotu 0,0674, čili v obou případech je výsledek na hranici zamítnutí hypotézy rovnosti.

Kromě počtu výher či bodů jsem se také podíval na skóre zápasů, zase vždy dohromady přes celé kolo. Počet gólů vstřelených domácími a hosty

fikace (odrůda, ročník, vinařství) byly účastníkům promítány na plátně pod čísly 1–4. Za platnou odpověď byla tedy brána libovolná permutace těchto čísel, správná byla však jen jedna. Vítězem se stal Matej Benko z brněnského VUT, který nad rámec své specializace (Veltlínské zelené) správně určil i zbylé tři vzorky a první ročník této soutěže tak ovládl.

Nedělní dopoledne bylo věnováno dalším praktickým příspěvkům. Po Elišce Cézové jsme si poslechli i včerejšího šampiona (*Využití entropie při posuzování modelu pohybu lietadla*), po coffee breaku si vzali slovo kolegové z MFF UK. Zdeněk Hlávka se snažil najít cestu, jak zbohatnout na kryptoměnách, Jaromír Antoch zase obeznámil účastníky s tím, jak se dá měřit komfort autosedačky.

Celé víkendové setkání jsme zakončili výbornou svíčkovou, zbývalo jen rozloučení a poděkování organizátorům. A že bylo za co děkovat! Proč to tak dlouho trvalo?

2. Statistické dny na břehu Tiché Orlice – květen 2023

Po vydařené podzimní akci jistě nejednoho člena naší společnosti zajímalo, zda budeme muset na příští setkání čekat dalších 8 let. Nebo se po telčském úspěchu dočkáme brzy dalšího shledání? Odpovědi na tyto otázky byly zodpovězeny již koncem února roku 2023. Tehdy totiž dorazila pozvánka na další ročník Statistických dnů, tentokrát konaných 19. – 21. 5. v penzionu Mítkov, který stojí v sevření Tiché Orlice a jedné z nejvytěžovanějších železničních tratí u nás, která není tichá ani trochu. Někteří účastníci byli z nové intenzity setkávání tak nadšení, že na vlak do nedaleké Chocně vyrazili dokonce s týdenním předstihem!

Plánovaný program sliboval zajímavé příspěvky a možná i bouřlivou debatu – první den byl totiž věnován covidu. Celé to odstartoval Arnošt Komárek se svým pohledem na uplynulé roky a přehledem pandemických milníků. Navázala na něj Zuzana Krátká, která přiblížila svoji spolupráci se statistiky, snahu o rozšíření našeho poznání o covidu, ale také nepříjemnosti, které z toho plynuly. O modelování vývoje pandemie, špatné interpretaci modelů ze strany povolaných osob a také algoritmech jedné sociální sítě přednášela Lucia Houfková, autorka stránky *COVID zápisky* na oné sociální síti.

Po krátké pauze u kávy se slova ujal Tomáš Fürst a v 10 bodech shrnul, co jsme se o covidu dozvěděli a „nedá se zpochybnit“. Komentáře z publika u pár bodů naznačily, že zpochybnit se některá tvrzení rozhodně dají. Ale tak je to správně a proto tu jsme – cílem tohoto bloku nebylo prezentovat neprůstředné pravdy a vzájemně si je odkývat, nýbrž poukázat na mnohé jevy, problémy a rozhodnutí této zvláštní etapy lidstva, diskutovat o nich

a vykročit do nového dne tím nejlepším způsobem. I přes časnou hodinu se přidala také čile vypadající skupina kolegů z Brna – přece „Večer šohaj, ráno šohaj“! Zájemci si také po cestě ze snídaně mohli zakoupit burčák v jednom ze stánků právě začínajícího trhu na onom proslaveném náměstí. Lokalita, termín, organizátoři mysleli na všechno!

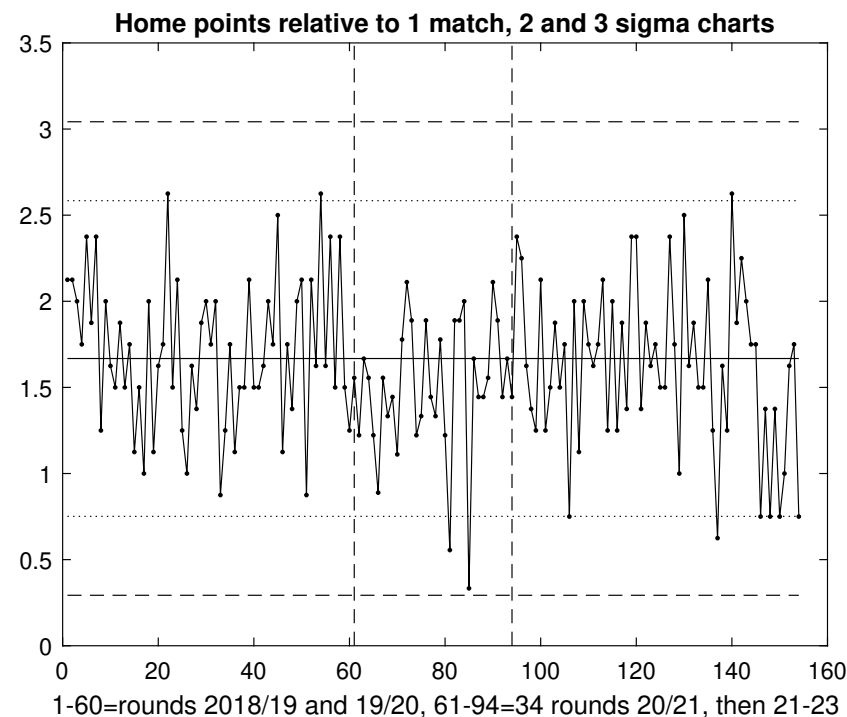
Dopoledne se se svým příspěvkem týkajícím se úmrtnostních tabulek představil Jirka Novák z ČSÚ, podnětnou diskuzi (mj. o odpadové politice státu a obcí) vyvolal velmi zajímavý příspěvek Marka Omelky, sekci pak uzavírala Daniela Jarušková povídáním o variabilitě predikční chyby v lineární regresi. Po krátkém coffee breaku jsme měli možnost vyslechnout ryze praktickou přednášku Katariny Krausové osvětlující možnosti využití softwaru SAS.

Po obědě vyrazila celá skupina na plánovaný výlet – cílem byl nedaleký hrad Roštejn pocházející ze 14. století. Část výletníků se po prohlídce vydala na turistickou trasu a aktivní odpoledne si protáhla, zbytek se vrátil do města – např. rozněžnit chuťové pohárky v některé z výborných kaváren, užít si centrum, případně si projít zámecký park a vychodit tak alespoň část kalorií ze dvou sněžených cheesecaků a flat white (jen příklad, nijak nesouvisející s osobní zkušeností autora).

Následovala společná večeře, po níž nastala jediná drobná krize našeho víkendového setkání. Zásoby vína, které byly k dispozici pro oba večery, se nějakým způsobem zvládnou spotřebovat během večera prvního. Jistojistě nešlo o důsledek přehnané spotřeby – to by si páteční diskuze neudržela tak vysokou kvalitu až do nočních hodin. Šlo spíše o chybu při plánování – i ti nejlepší organizátoři se mohou přepočítat. V literatuře můžeme koneckonců najít podobné případy:

Náčelník: *To jsem vám právě chtěl říct, abyste si na to dali pozor, kdybyste někdy plánovali nějakou takovou cestu, aby se vám nestalo, jako nám tedko, že já jsem ty zásoby přesně propočítal a pořád jsem si říkal, až to sečteš, nezapomeň to znásobit dvěma. A v tom zmatku, jak jsme balili a vy jste mi do toho pořád mluvili, tak jsem na to zapomněl. Takže vypočtený to bylo správně, ale falšíruje nám v tom jenom to násobení. To jen tak pro příště. [1, s. 80–81]*

Na analýzu a plánování občerstvení do dalších let bude ještě času dost – teď se situace musela hlavně řešit! Ondra Vencálek se toho coby předseda ujal a uspořádal rychlou výpravu do nejmenovaného obchodního řetězce s cílem zásoby doplnit. Po cestě uzrála myšlenka, že bychom z krize mohli i těžit a večer obohatit unikátním zážitkem pro všechny účastníky. Takto vznikla první soutěž v ochutnávání vína v historii naší společnosti. Účast byla hojná, cílem bylo přiřadit ke čtyřem vzorkům správnou láhev, jejichž základní speci-

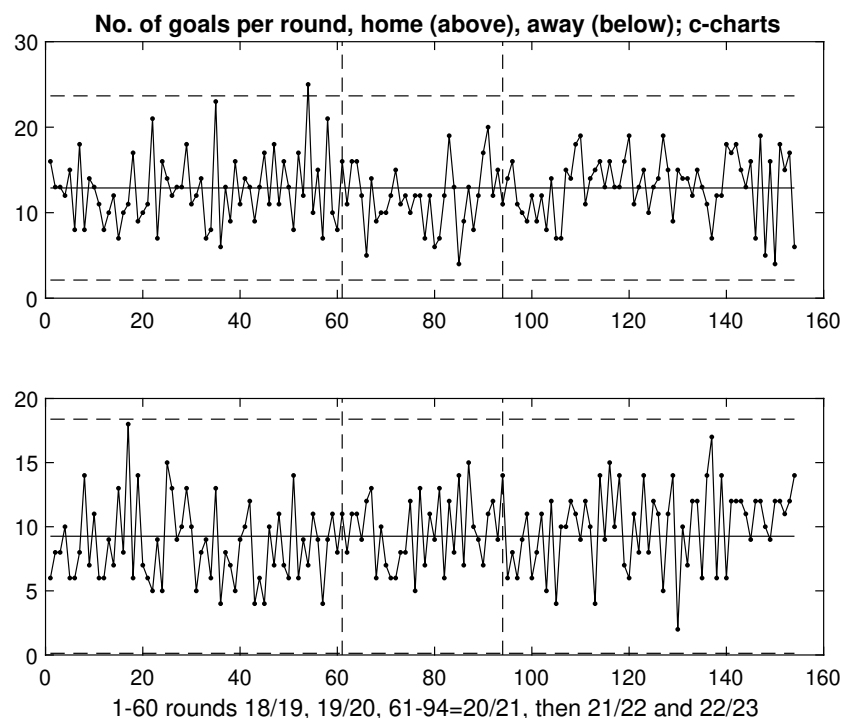


Obrázek 3: Průměrný počet bodů získaný domácím týmem za 1 zápas, po jednotlivých kolech.

během jednotlivých kol ukazuje Obrázek 4 (opět, počty gólů v každém kole v sezóně 2020/2021 byly brány jako zaokrouhlení 8/9 skutečného počtu). Obě skupiny dat (2020/2021 a ostatní) v obou grafech lze modelovat jako nezávislé Poissonovy náhodné veličiny. Tabulka 3 ukazuje jejich charakteristiky.

Sezóna	Průměr doma	Průměr venku	Std doma	Std venku
Ostatní	12,8917	9,2583	3,9211	3,2032
2020/2021	11,7647	9,7059	3,7178	2,7362

Tabulka 3: Charakteristiky počtu vstřelených gólů za jedno kolo.



Obrázek 4: Počty gólů vstřelených domácími (nahore) a hosty (dole) v jednotlivých kolech.

Obrázek 4 sice naznačuje jakési rozdíly, snížené počty gólů doma a zvýšené venku, výraznější je to pro domácí góly, zatímco pro góly venku se nepodařilo prokázat statisticky významnou odlišnost. Jednostranné testy rovnosti středních hodnot na základě CLV použité pro Poissonovo rozdělení vedly k následujícím p -hodnotám:

$$p = 0,0471 \text{ pro góly domácích, } p = 0,2287 \text{ pro góly hostů.}$$

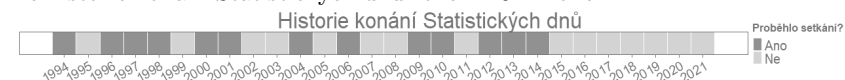
Obrázek 4 jsem doplnil kontrolními liniemi odpovídajícími řadě Poissonových veličin, z dat bez ročníku 2020/2021.

Jak jsem zmínil, někteří autoři používají 2-rozměrné Poissonovo rozdělení, pro góly vstřelené a obdržené během jednoho zápasu. Ta používaná verze má jedno omezení, a to že korelace je kladná. Zde šlo o počty gólů (domácích

STATISTICKÉ DNY 2022 A 2023 STATISTICAL DAYS 2022 AND 2023

Tomáš Chupář

Tradice Statistických dnů pořádaných ČStS započala již v roce 1994. Od té doby se až do roku 2014 setkávali kolegové a odborníci poměrně pravidelně – během této éry stihli celkem 13 setkání! Ještě nedávno se však mohlo zdát, že historie konání Statistických dnů rokem 2014 končí.



Co se stalo? Proč to dlouhé „sucho“? Odpověď na tuto otázku mají možná zkušenější kolegové, pro novější členy naší společnosti však Statistické dny byly pouze záložkou na webových stránkách v sekci ‚Konference‘ obsahující pár vzpomínek na poměrně vzdálenou minulost. V loňském roce to bylo již 8 let od poslední události tohoto typu. 8 let! Pro srovnání – autor tohoto textu v době posledního setkání (23. 6. 2014 v Ústí nad Labem) teprve končil povinnou školní docházku.

1. Statistické dny v Telči – září 2022

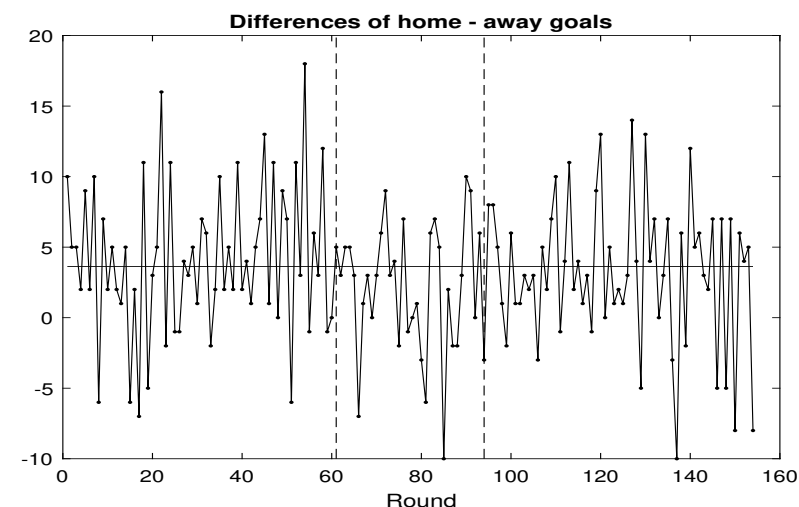
Pozvánka na akci konanou začátkem podzimu 2022 tak přišla jako blesk z čistého nebe. A aby toho nebylo málo, kromě opětovného shledání kolegů a přátel (v některých případech po mnoha letech) byla lákadlem také lokalita – malebné město Telč, jehož historické jádro je od roku 1992 na seznamu UNESCO. Na okraji tohoto historického jádra je také Univerzitní centrum Masarykovy univerzity, které od pátku 23. do neděle 25. září 2022 statistiky hostilo.

Páteční program začínal až večer v Švejk restaurantu, který se nachází jen pár kroků od dějiště konference a zajišťoval obědy i večere naší početné skupině. Večerní blok přednášek si po úvodním slovu našeho předsedy rozdělili Zdeněk Fabián a Marie Hušková, po jejich vystoupení se přešlo k neformálnímu družení. A že si bylo co říct, po těch letech... Nejvytrvalejší skupinka kolegů z VUT diskutovala až do pozdních nočních hodin, kdy je už autor (té noci klíčník přednáškového sálu) musel poprosit o vyklizení prostor, ač nerad. Tak to dopadá, když se mezi ročníky Statistických dnů dělají osmileté prodlevy.

Snídaně si každý z účastníků zařizoval individuálně, někteří šli cestou nadpozemského luxusu – jak jinak také nazvat hotelové snídaně v Café Telč. V 8 hodin jsme se mohli přidat k hostům ubytovaným v tomto podniku

dopady změn v energetice hodnotil Lubomír Lízal. Zkušenosti ze zahraničí prezentoval Frank Lenzmann – zástupce Energy Research Center of the Netherlands. Florian Sobieczsky ze Software Competence Centra v rakouském Hagenbergu pak hovořil o optimalizovaném řízení nejenom domácího vytápění. V rámci workshopu se rovněž diskutovalo o dotačních příležitostech v energetice.

Energy Days 2023 se zúčastnilo celkem 82 účastníků. Česká statistická společnost prostřednictvím příspěvku na dopravu podpořila účast řady studentů. Organizátory tak může mrzet snad pouze nezájem politické reprezentace (zváni byli členové podvýboru pro energetiku Poslanecké sněmovny Parlamentu ČR) a široce pozvaných médií.



Obrázek 5: Rozdíly počtu gólů vstřelených domácími a hosty v jednotlivých kolech.

a hostů) vždy za celé kolo, korelace byla jednak hodně malá (statisticky nevýznamná), a navíc záporná.

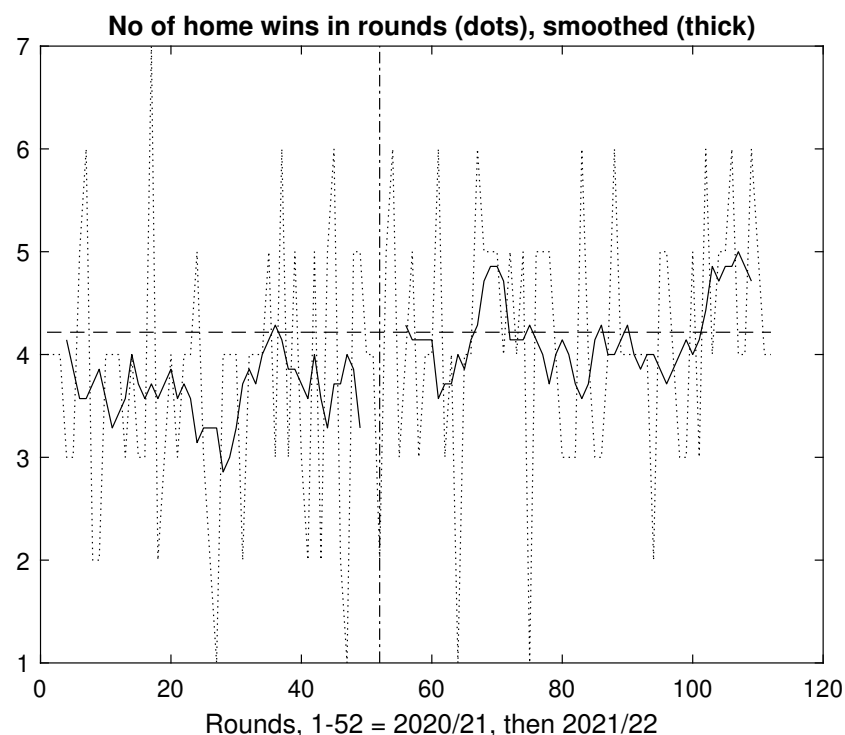
Ještě jsem zkusil i další model, a to rozdíl dvou nezávislých Poissonových rozdělení, výsledkem je tak zvané **Skellamovo rozdělení** (viz třeba opět Volf, ICPM 2008). Na Obrázku 5 jsou vidět rozdíly počtu gólů vstřelených domácími a hosty po jednotlivých kolech, průměry těchto rozdílů jsou 2,0588 pro 2020/2021 a 3,6434 pro ostatní (této hodnotě odpovídá i centrální linie v grafu).

Opět je možné použít alespoň přibližný test na rovnost příslušných středních hodnot, založený na CLV: Jsou-li X, Y nezávislé Poissonovy n.v. s parametry λ_1, λ_2 , tak $S = X - Y$ má $E(S) = \lambda_1 - \lambda_2$ a rozptyl $\text{var}(S) = \lambda_1 + \lambda_2$, čili vhodná testovací veličina je

$$Z = (\bar{S}_1 - \bar{S}_2) / \sqrt{\text{var}(S_1)/n_1 + \text{var}(S_2)/n_2},$$

kde $\bar{S}_1, \bar{S}_2$ jsou ony průměry rozdílů uvedené výše, odhady rozptylů jsou (jak naznačeno výše) součty oněch průměrů z Tabulky 3, zvláště v řádce pro 2020/2021 a v řádce pro ostatní, n_1, n_2 odpovídající počty hodnot, zde tedy počty kol v 2020/2021 a v ostatních sezónách.

Nyní takto provedený jednostranný test dal $p = 0,0164$, neboli prokázal, že v sezóně 2020/2021 se významně snížil rozdíl mezi počty gólů vstřelených domácími a hosty. Nicméně tento rozdíl je stále kladný, a to statisticky významně, test může opět být založen na podobné veličině $Z_1 = (\bar{S}_1 - 0)/\sqrt{\text{var}(S_1)/n_1}$, která má hodnotu 2,5908 a tedy odpovídající p -hodnota v rámci standardního normálního rozdělení je 0,0048.



Obrázek 6: Počty domácích výher v jednotlivých kolech hokejové extraligy (tečkovaně), vyrovnané pro každou sezónu klouzavým oknem šířky 7 (plné křivky).

KONFERENCE ENERGY DAYS 2023 ENERGY DAYS 2023 CONFERENCE

Ondřej Vencálek

E-mail: ondrej.vencalek@upol.cz

Ve dnech 10. – 11. listopadu 2023 proběhl v prostorách Českého statistického úřadu v Praze (10.11.) a ČVUT v Praze (11.11.) mezinárodní workshop Energy Days 2023. Organizátoři (Česká statistická společnost a Akademie věd ČR) tak po vynucené několikaleté pauze navázali na pět předchozích ročníků konaných v letech 2015 až 2019, viz <https://energy-workshop.nipax.cz/>.

Problematika energetiky se v poslední době dostává čím dál více do popředí zájmu široké veřejnosti, ať už bezprostředně v souvislosti s rostoucími (a následně klesajícími) cenami energií, či v souvislosti s politicky prosazovanými snahami o využívání ekologicky šetrných zdrojů energie, které s sebou přinášejí nemalé náklady a technické výzvy/problémy, a v neposlední řadě také v souvislosti se soběstačností, respektive závislostí na zdrojích surovin ze zahraničí. Všechny tyto aspekty byly v rámci programu workshopu diskutovány.

Hned první příspěvek, který přednesl místopředseda ČSÚ Jaroslav Sixta, se týkal nedávno velmi intenzivně sledované statistiky „průměrných“ cen energií. Metodika odhadu průměru se v době turbulentních změn cen energií stala předmětem kritiky a byla následně upravována. Mimo jiné se potvrdila známá skutečnost, že nelze mezi sebou porovnávat hodnoty z různých zemí, ke kterým se došlo různými způsoby.

Následovaly další příspěvky, které byly stejně pestré, jako bylo pestré složení řečníků. Zastoupeni byli akademici s matematicko-statistickým zaměřením, fyzici z Akademie věd ČR či ČVUT v Praze, ale také odborníci z praxe. První skupinu reprezentoval např. Miloš Kopa z MFF UK, který hovořil o predikcích cen energií v Británii, nebo Emil Pelikán z AV ČR s příspěvkem o využití umělé inteligence v energetice. Jakub Dostál posluchačům vysvětloval, proč energetiku 21. století považuje za obor statistiků, a ukázal zajímavé predikce. Neméně zajímavé byly i příspěvky fyziků, např. o využití malých jaderných reaktorů (Vladimír Wagner) či o současném stavu využití solární energie v ČR i ve světě (Antonín Fejfar a Vítězslav Benda).

Se svými příspěvky však vystoupili také lidé z praxe, např. Pavel Janeček (Karma Český Brod), který hodnotil současný stav evropské energetické infrastruktury a trhu, či Hynek Beran (Cygni), který se podělil o zkušenosti s komunitní energetikou. Zajímavá debata se otevřela po příspěvku Ondřeje Mamuly (ČVUT Praha) o hybridních zdrojích energie. Některé ekonomické

Mikulášské setkávání ČStS má za sebou 17 ročníků. Když Gejza Dohnal psal nadšeně o vydařeném předvánočním setkání roku 2019, které se poprvé uskutečnilo na VŠE v Praze, a které mělo (téměř) rekordní účast, v daleké Číně už se šířil virus, který měl brzo ovlivnit nejen konání tohoto předvánočního setkání [2]. Mikuláš se stejně jako řada dalších akcí přesunul v roce 2020 do online prostoru. A přežil. V následujících dvou letech, kdy jsme se vrátili do Karlína (MFF UK) i na Žižkov (VŠE Praha), byl Mikuláš „hybridní“ – účastníků se sešlo méně, zato někteří využili možnost zúčastnit se online. Letošní Mikuláš definitivně opustil virtuální prostor a vrátil se do reality. Věřím, že to bylo správné rozhodnutí. Hojnou účast (21 účastníků) vnímám jako blýskání na lepší Mikulášské časy.

Fotografie ze setkání a jednotlivé příspěvky jsou zájemcům k dispozici na stránce <https://www.statpol.cz/konference/mikulassky-den/>.

Reference

- [1] Malý, M. (2008). Mikulášský statistický den 2007, *Informační bulletin ČStS*, 19(1), s. 34. cit. 19
- [2] Dohnal, G. (2020). Mikulášské setkání členů České statistické společnosti, *Informační bulletin ČStS*, 31(1), s. 12–15. cit. 20



Účastníci semináře v respíriu karlínské budovy MFF UK.

4. Výsledky pro hokejovou Tipsport extraligu

Pro českou hokejovou extraligu jsem provedl jen zkrácenou analýzu, na základě dat – výsledků zápasů, které jsou k dispozici na webových stránkách <https://www.sport.cz/sekce/hokej-extraliga-vysledky-23>.

Porovnával jsem tytéž sezóny jako pro fotbal, resp. jejich základní části. Česká hokejová extraliga má (zpravidla) 14 týmů, hraje se čtyřikrát každý s každým, tj. 52 kol po sedmi zápasech. Výjimkou byl ročník 2021/2022, kde bylo 15 týmů, takže se hrálo 60 kol (také po sedmi zápasech, 1 tým měl vždy volno). Základní část sezóny 2019/2020 se stihla dohrát před vypuknutím pandemie (play-off se pak už nehrálo). V sezóně 2020/2021 to bylo opět tak, že do října 2020 byly počty diváků omezeny, říjen se vynechal a od listopadu se hrálo zcela bez diváků.

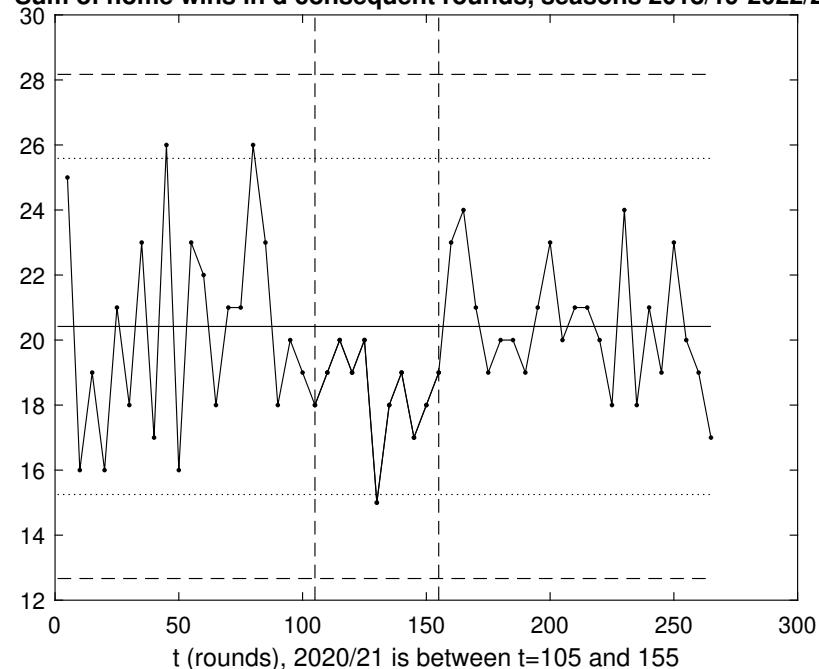
Porovnával jsem nyní jen počty vítězství domácích, po jednotlivých kolech (s tím, že jsem zahrnul i vítězství v prodloužení či po nájezdech). Obrázek 6 porovnává tyto počty v kolech sezón 2020/2021 a 2021/2022. Je vidět, že v průměru je počet vítězství domácích menší v „covidové“ sezóně. Průměry jsou 3,6538, 4,2167 (tomu odpovídá centrální linie v grafu). Po vydělení sedmi dostanu relativní četnosti 0,5220 a 0,6024. Test jejich rovnosti (v 2×2 kontingenční tabulce, viz 1. část Tabulky 4) vede k zamítnutí této hypotézy, s p -hodnotou 0,0235.

Pokud porovnávám sezónu 2020/2021 se všemi čtyřmi ostatními, dostanu z oněch 4 průměr 4,0833 vítězství domácích na 1 kolo, tj. relativní četnost 0,5833, dohromady z 216 kol $\times$ 7 zápasů. I v příslušné kontingenční tabulce (2. část Tabulky 4) hypotézu nezávislosti zamítám, s p -hodnotou 0,0337.

Nabídnou ještě jedno porovnání: Sečetl jsem počty domácích vítězství vždy za 5 kol (tj. za kola 1–5, 6–10, atd.), čili z celkem 268 kol jsem dostal 53

Sezóna	Dom. výhry	Dom. prohry	Součty
2020/2021	190	174	364
2021/2022	253	167	420
Součty	443	341	784
2020/2021	190	174	364
Ostatní	882	630	1512
Součty	1072	804	1876

Tabulka 4: Četnosti domácích výher a proher.

Sum of home wins in d consequent rounds, seasons 2018/19-2022/23

Obrázek 7: Počty domácích vítězství za $d = 5$ kol po sedmi zápasech. Linie jsou průměr a ± 2 sigma, 3 sigma pásy, počítané bez ročníku 2020/2021.

hodnot, z toho 10 hodnot odpovídalo sezóně 2020/2021. Výsledek je na Obrázku 7. Jeden bod tak odpovídá součtu 35 výsledků Bernoulliho pokusů. Při $n = 35$ už s touto veličinou mohou zacházet jako s normální, a snažím se ukázat, že ve zkoumané sezóně má jiné charakteristiky. Graf to i potvrzuje. Průměr pro sezónu 2020/2021 byl 18,2690, pro ostatní 20,4186. Směrodatné odchylky byly 1,5055 a 2,5841, mediány 19 a 20.

F -test zamítl rovnost rozptylů, s $p = 0,0458$. Hypotéza rovnosti středních hodnot byla pak testována pomocí přibližného t -testu pro různý rozptyl, s $p = 0,0015$, ale i testování shody mediánů testem Wilcoxonova prokázalo jejich nerovnost (s $p = 0,0122$). Takže tímto způsobem se podařilo prokázat, že v sezóně 2020/2021 byla bilance vítězství domácích celků statisticky významně horší než v sezónách ostatních. Ale i za nepřítomnosti diváků působí

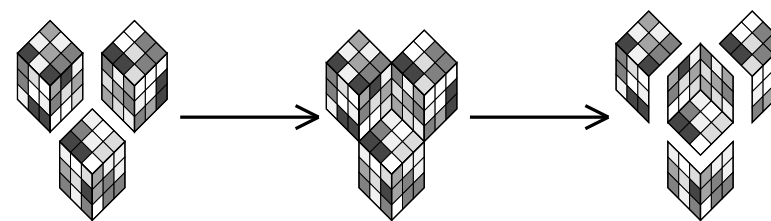
MIKUKLÁŠSKÉ BLÝSKÁNÍ MEETING ON THE SAINT NICHOLAS' DAY

Ondřej Vencálek

E-mail: ondrej.vencalek@upol.cz

Od dnes již legendárního prvního Mikuklášského setkání České statistické společnosti (ČStS), které se uskutečnilo 6. prosince 2007 v Balbínově poetické hospůdce [1], letos uplynulo už 16 let. Od té doby se čeští statistici scházejí na Mikuláše (plus minus pár dní – jak to máme v našem oboru rádi) každoročně. Letošní setkání bylo již sedmnácté. Mikuláš se po dvou letech vrátil na Karlín a protože po (zřejmě více než dvou) letech napadl v Praze sníh, byla atmosféra letošního setkání vskutku sváteční.

Účastníci setkání vyslechli pět odborných příspěvků. Program zahájil Jan Kalina, který s hlubokým zaujetím pro historii našeho oboru připomněl nejen dvě významné osobnosti – Johna Graunta a Adolpha Queteleta, od jejichž úmrtí uplyne v příštím roce 350, resp. 150 let – ale i řadu dalších velikánů. Následoval příspěvek Ondřeje Vencálka, který se zamýšlel nad 11% meziročním poklesem počtu narozených dětí v ČR zaznamenaným v roce 2022, a ukázal rovněž vizualizaci dat o zemřelých v Anglii srovnávající riziko úmrtí ve skupinách dle věku a očkovacího statusu. Vizualně nejzajímavější příspěvek však bezpochyby prezentoval až další řečník – Pavel Stríž. Ten už dopředu avizoval speciální Mikuklášské téma: Rubikovy-Čertovy kostky. Posluchače seznámil s vizuálním klamem – tzv. čertovou kostkou – vznikajícím při zobrazování tří k sobě přiléhajících Rubikových kostek (viz obrázek). A ukázal, jak snadno se stane, že takto vzniklá čertova kostka je „neplatná“ (stejně obarvená Rubikova kostka neexistuje).



Po krátké přestávce program pokračoval příspěvkem Kamily Fačevicové. Před jejími skrytými komponentami se nestihli ukrýt ani krabi, ani novorozenci. V závěrečném příspěvku Jaromír Antoch připomněl nedávno proběhlé Energy Days, přičemž prozradil i leccos ze „zákulísí“, tedy z pohledu organizátora.

POZVÁNKA NA KONFERENCI OSSCONF 2024 INVITATION TO THE OSSCONF 2024 CONFERENCE

Redakce



SOIT (Spoločnosť pre otvorené informačné technológie) v spolupráci s Fakulťou riadenia a informatiky Žilinskej Univerzity, CSTUG a Českou statistickou společností organizuje tento rok už 12. ročník konferencie OSSConf 2024.

Konferencia sa uskutoční v priestoroch Fakulty riadenia a informatiky v Žiline v termíne 2. júla – 4. júla 2024.

Cieľom našej konferencie je poskytnúť priestor pre informovanie o novinách vo vývoji otvoreného softvéru a otvorených technológií, o možnostiach využitia týchto nástrojov vo vede a vzdelávaní a taktiež poskytnúť priestor pre neformálne priateľské stretnutie užívateľov a priaznivcov otvorených informačných technológií.

Pri príležitosti konferencie bude vydaný zborník príspevkov. Príspevky autorov prijímame buď ako nerecenzované príspevky, z ktorých uverejňujeme len abstrakt alebo recenzované príspevky, ktoré budú posudzované dvomi nezávislými recenzentmi. Predpísaná šablóna pre písanie príspevkov v typografickom systéme $\text{\LaTeX}$ / $\text{\LaTeX}$ bude k dispozícii na webovej stránke konferencie. Príspevky prijímame v slovenskom, českom alebo anglickom jazyku.

Dôležité termíny. Recenzované príspevky do 1. júna 2024 a abstrakty nerecenzovaných vystúpení do 15. júna 2024.

Sekcie. $\text{\LaTeX}$, R a ich priatelia; Open AI; Vývoj OSS; Open GIS & Open Data; OSS vo vzdelávaní a Open Hardware.

Našu konferenciu sa snažíme, tak ako po iné roky, udržať bez platenia vložného. Pre záujemcov môžeme za úhradu zabezpečiť obedy v stravovacom zariadení Žilinskej univerzity (aktuálnu cenu uverejníme na web stránke). Účastníci, ktorí majú záujem o ubytovanie si toto vybavujú individuálne podľa pokynov zverejnených na web stránke konferencie. Predpokladaná cena zborníka bude 8–10 €.

Bližšie informácie o konferencii budeme postupne zverejňovať na stránke <http://ossconf.soit.sk/>.

domáci prostredí motivačne, relatívni četnosť domácich víťazstiev bola stále nadpoloviční, i keď pozorovaná proporcie 0,5220, z 364 utkání, není statisticky významně odlišná od 0,5; p -hodnota jednostranného testu (v rámci CLV pro binomické rozdělení) byla 0,2.

5. Závěrečné poznámky

Průzkumová a grafická analýza ukazuje zpravidla na pokles úspěšnosti domácích týmů (i když úspěšnost je stále nadpoloviční). Výsledky statistických testů (hypotéz „sezóna 2020/2021 se neliší“ v určité charakteristice, se snahou je zamítnout) jsou většinou statisticky významné, a někdy jen „na hraně“ (typicky s p -hodnotou kolem 0,1). I takový 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). I Pollard (1986) ve svém rozboru uvádí diváky jen jako jeden z faktorů výhody domácího prostředí. A podobné závěry jsem našel i v práci Wunderlich et al. (2021), která už se taky zabývá důsledky nepřítomnosti diváků během pandemie:

„More than 40,000 matches before and during the pandemic, including more than 1,000 professional matches without spectators across the main European football leagues, have been analyzed. Results support the notion of a crowd-induced referee bias: the increased sanctioning of away teams disappears in the absence of spectators with regard to fouls ($p < 0.001$), yellow cards ($p < 0.001$), and red cards ($p < 0.05$). Moreover, the match dominance of home teams decreases significantly as indicated by shots ($p < 0.001$) and shots on target ($p < 0.01$).“

Domnívám se, že počet faulů jednoho týmu souvisí s aktivitou týmu druhého (je víc obranných zákroků, i nečistých), není to tedy jen případný důsledek strannosti rozhodčích. Dále, ten významný pokles aktivity domácích týmů je trochu v rozporu s tím, že se to (podle autorů) neprojevovalo dostatečně ve výsledcích, počtu gólů a bodů:

„However, in terms of the home advantage itself, surprisingly, only a non-significant decrease has been found. Spectators thus do not seem to be the main driving factor of the home advantage.“

Soudím, že mně se, na „mých“ datech, podařilo prokázat statisticky významný pokles vlivu domácího prostředí na výsledky.

Literatura

- [1] Dixon, M. J., Coles, S. G. (1997): Modelling association football scores and inefficiencies in the football betting market. *Appl. Stat.* **46**, 265–280. *cit. 4*
- [2] Hejdušek, M. (2007): *Matematické metody pro modelování a predikci výsledků sportovních utkání*. DP na KM FJFI ČVUT v Praze.
- [3] Karlis, D., Ntzoufras, I. (2003): Analysis of sports data by using bivariate Poisson models. *J. R. Stat. Soc. D* **52**, 381–394.
- [4] Maher, M. J. (1982): Modelling association football scores. *Stat. Neerlandica* **36**, 109–118. *cit. 4*
- [5] Marek, P., Vávra, F. (2020): Comparison of Home Advantage in European Football Leagues. *Risks* **87**. <https://doi.org/10.3390/risks8030087>. *cit. 5*
- [6] Pollard, R. (1986): Home advantage in soccer: A retrospective analysis. *Journal of Sport Sciences* **4**, 237–248. *cit. 5, 6*
- [7] Rambousková, D. (2007): *Pravděpodobnostní modely a metody pro analýzu sportovních výsledků*. DP na KPMS MFF UK v Praze.
- [8] Volf, P. (2008): On model for difference of two Poisson variables. In: *Proceedings of the ICPM 2008*, TU Liberec, 34–44. *cit. 5*
- [9] Volf, P. (2009): A random point process model for the score in sport matches. *IMA Journal of Management Mathematics* **20**, 121–131. *cit. 5*
- [10] Volf, P. (2017): Competing risk model with a non-traditional application. *Informační bulletin České statistické společnosti* **28**, 1, 1–11.
- [11] Wunderlich F., Weigelt M., Rein R., Memmert D. (2021): How does spectator presence affect football? Home advantage remains in European top-class football matches played without spectators during the COVID-19 pandemic. *PLoS ONE* **16**, 3: e0248590. <https://doi.org/10.1371/journal.pone.0248590>. *cit. 15*

11. SVĚTOVÝ KONGRES VĚNOVANÝ PRAVDĚPODOBNOSTI A STATISTICĚ THE BERNOULLI-IMS 11TH WORLD CONGRESS IN PROBABILITY AND STATISTICS

Redakce

Dear colleagues,

the Bernoulli-IMS 11th World Congress in Probability and Statistics will be held from 12 to 16 August 2024 on the campus of Ruhr University Bochum, Germany. The World Congress will be the major international conference in the area of Probability and Statistics in 2024, where scientists from around the globe will gather to exchange ideas and to present the results of their most recent research. For more information about the World Congress please visit the website <https://www.bernoulli-ims-worldcongress2024.org/>.

On behalf of the local organizing committee and of the Department of Mathematics of Ruhr University Bochum, I cordially invite you to participate in this meeting, and to take the opportunity to present the results of your own research and to listen to the ideas of your colleagues. Submission of abstracts for contributed talks and posters as well as for proposals for organised contributed paper sessions is now open; see

<https://www.bernoulli-ims-worldcongress2024.org/registration>

Please note the following deadlines:

- Abstracts for contributed talks and posters: 15 February 2024 (decision before 31 March 2024)
- Proposals for organised contributed paper sessions: 30 November 2023 (decision before 15 December 2023)

For more details, please see the Call for Papers

<https://www.bernoulli-ims-worldcongress2024.org/call-for-papers>

I hope to be able to welcome many of you in August 2024 in Bochum for the 11th World Congress in Probability and Statistics.

Best regards,
Herold Dehling
Chair of the Local Organizing Committee