

PERNÉ STATISTICKÉ DNY 2024 POD PÁLAVOU

24. – 26. KVĚTNA 2024

Abstrakty

Sto let (nejen) Shewhartových regulačních diagramů

Ondřej Vencálek (Univerzita Palackého v Olomouci, PŘF)

V květnu 2024 si připomínáme 100 let od chvíle, kdy Walter A. Shewhart navrhl použití regulačních diagramů při statistické kontrole kvality. V příspěvku připomeneme základní principy, z nichž regulační diagramy vychází. Přiblížíme rovněž okolnosti tohoto převratného návrhu.

Implementace regulačních diagramů v praxi

Tomáš Jurczyk (Spotfire)

Shewartovy regulační diagramy jsou jedním z nejpoužívanějších nástrojů pro kontrolu kvality. V tomto příspěvku se nebudeme bavit o teorii na pozadí, ale zamyslíme se nad implementačními aspekty. Navržením správného grafu pro náš proces totiž všechno teprve začíná. Co vůbec potřebujeme ze softwarového hlediska, abychom dostali z diagramů maximum? Co když máme stovky či tisíce parametrů ke sledování? A co dělat, když už v datech narazíme na problém výrobního procesu? Diskuzi těchto otázek podpoříme ukázkou v softwaru.

Neurčitost rekonstrukce funkcionálních trajektorií při diskretním pozorování s šumem

David Kraus (Masarykova univerzita, Ústav matematiky a statistiky)

Příspěvek se věnuje situaci, kdy funkcionální data (například křivky) jsou pozorované v diskretních argumentech s šumem. Ukážeme si, jak se flexibilně odhadují komponenty modelu (střední funkce, kovarianční funkce, varianční funkce pozorovacího šumu) včetně efektů kovariát na ně. Dále se zaměříme zejména na rekonstrukci (predikci) individuálních realizací funkcí na základě jejich řídkých pozorování. Ukážeme si tradiční přístup ke konstrukci predikčních intervalů a pásů, uvidíme, že často nefunguje moc dobře, a ukážeme si vylepšený postup.

Několik poznámek k závislosti náhodných jevů

Zdeněk Karpíšek, Tomáš Pospíšil (FSI VUT v Brně)

Příspěvek je zaměřený na míry závislosti náhodných jevů, kterými jsou jejich korelace a regrese. Jde o pojmy, se kterými se posluchač základního kurzu z teorie pravděpodobnosti a matematické statistiky takřka jistě nepotká, a které nejsou ani v převážné většině učebnic. Jedná se přitom o charakteristiky závislosti náhodných jevů, které se snadno počítají a mají zajímavé vlastnosti. Navíc mohou nalézt uplatnění ve statistice, když pravděpodobnosti náhodných jevů odhadneme pomocí pozorovaných relativních četností jejich nastoupení.



Praktické využití Random Forests, TreeNet, CART v Minitabu

Vilém Patloka (SC&C Partner spol. s r.o.)

Problém optimálního hledání ztracené osoby formulovaný pomocí pravděpodobnosti, i s užitím Bayesova vzorce

Petr Volf (ÚTIA AV ČR)

Dostal se mi do ruky jakýsi zárodek článku, viz [1], kde je stručně tento problém nastíněn. Autor je zřejmě obeznámen s praxí takového hledání (ve Velké Británii, ať už v krajině či na moři), jsou tam i odkazy na články na toto téma. Ale matematická formulace, byť jednoduchá, potřebuje zpřesnit. Situace je tato: Hledá se osoba, ví se, že je v určitém území, to je za účelem hledání rozděleno na víc oblastí. Každá tato oblast má pravděpodobnost, P_j , že v ní ona osoba je, tyto pravděpodobnosti se tedy sčítají do 1. Zároveň má každá oblast pravděpodobnost, Q_j , že při hledání se tam ona osoba najde, pokud tam je. Čili po neúspěšném hledání v oblasti j se P_j zmenší (a to třeba pomocí Bayesova pravidla), zároveň se ostatní P_k zvětší. A to je vlastně celé. Úloha je minimalizovat střední dobu, za kterou se osoba najde (liší-li se doby prohledávání), nebo minimalizovat střední počet hledání do jejího nalezení. Zůstanu u této poměrně jednoduché formulace, i když by se dala usložit třeba časy přesunů, změnou Q_j , třeba přerozdělením prostředků hledání, a podobně. Zajímavé by mohly být analogie takové úlohy v jiném oboru, třeba hledání závad v rámci analýzy spolehlivosti(?).

[1] Andre Clark: Notes on the Implications of Ignoring Bayes' Rule in Search and Rescue Practice in the UK. Qeios, CC-BY 4.0 · Article, February 9, 2024.

Objektivní a subjektivní hledání závislosti v datech

Libor Žák (Ústav matematiky FSI VUT v Brně)

Při hledaných závislostech v datech, se zaměřujeme na vhodný popis proměnných a následně souvislost mezi nimi. Ve statistických metodách vycházíme z „objektivních“ vlastností proměnných – naměřené hodnoty a jejich interpretace nezávisí (neměla by záviset) na subjektu. Pak pomocí vhodných statistických metod hledáme možné závislosti. Při subjektivním přístupu předpokládáme, že data i jejich zpracování ovlivňuje subjekt. Předpokládá se, že subjekt je „expert“ pro danou oblast zkoumání. Chování experta budeme simulovat pomocí fuzzy množin a metod přibližného uvažování. Pokud v datech (procesu) je reálná závislost, měly by tyto dva přístupy vést k jejímu nalezení.

Stochastické diferenciální rovnice a neuronové sítě

Matej Benko (Ústav matematiky FSI VUT v Brně)

Geometrické interpretácia McKean-Vlasovej stochastickej diferenciálnej rovnice ako krivka s najväčším klesaním funkcionálu v priestore pravdepodobnostných mier s Wassersteinovou vzdialenosťou a jej využitie a aplikácia v hľadani koeficientov neuronových sietí.



Extremální index a jeho odhady

Michal Fusek, Jan Holešovský (VUT v Brně)

Extremální index je důležitým parametrem pro vyhodnocení extrémních hodnot v časových řadách, protože řídí shlukování extrémů a charakterizuje krátkodobou (lokální) závislost v časové řadě. V posledních letech se objevilo několik nových odhadů extrémálního indexu, např. odhad založený na cenzorovaném nebo useknutém rozdělení pravděpodobnosti. Cílem tohoto příspěvku je představit a porovnat vybrané odhady extrémálního indexu.

Lokální závislost extrémních hodnot při vyhodnocení časových řad

Jan Holešovský, Michal Fusek (Fakulta stavební VUT v Brně)

Z teorie extrémních hodnot je známo, že při pozorování stacionárních řad dochází ke shlukování extrémních hodnot. Před samotným vyhodnocením extrémních událostí je tak zpravidla nutné provést identifikaci shluků pozorování nad nějakou vysokou prahovou hodnotou u_n . Sílu lokální závislosti extrémních hodnot je přitom možné popsat tzv. extrémálním indexem. Identifikace shluků tedy zahrnuje určení vhodného odhadu tohoto parametru. Většina odhadů extrémálního indexu je založena na předpokladu omezující lokální závislost řady. Tímto předpokladem obvykle bývá $D^k(u_n)$ podmínka, která limituje vazbu mezi časově vzdálenými pozorováními a kontroluje tak tendenci procesu vykazovat neextrémní chování v rámci daného shluku. Konkrétně bývá často uvažována podmínka řádu 2, tj. $D^2(u_n)$ podmínka. Odhad extrémálního indexu založený na podmínce daného řádu však nemusí být vhodný pro další procesy, které vyhovují podmínce jiné. Ignorování tohoto faktu obvykle vede k významnému nárůstu vychýlení odhadů. V poslední době byly představeny některé odhady, jež jsou založeny na podmínce obecného řádu k . Jejich vhodnost a další vlastnosti jsou dány hodnotami pomocných parametrů. V současnosti však není k dispozici téměř žádná vhodná metoda, jak platnost $D^k(u_n)$ podmínky daného řádu ověřit. Již dříve byly navrženy postupy založené na grafickém vyhodnocení, tyto ale často vedou k subjektivním závěrům. Jiné postupy byly odvozeny pro vybrané odhady extrémálního indexu, jejichž vlastnosti mnohdy nepatří k nejlepším. Cílem tohoto příspěvku je shrnutí stávajících postupů a představení některých nových metod pro vyhodnocení platnosti dané podmínky lokální závislosti pomocí cenzorovaného a useknutého rozdělení.

Možnosti výběrových šetření a statistické kontroly pro odhalení řídice se vyskytujícími závad

Jaromír Antoch (MFF UK, KPMS)

Uvažujme sériový systém skládající se z mnoha tisíc či desítek tisíc velmi podobných či prakticky identických prvků. Předpokládejme, že tento systém je pravidelně kontrolován, takže o spolehlivosti jednotlivých prvků máme alespoň částečnou informaci. Přenáška se bude zabývat možnostmi použití výběrových šetření a metod statistické kontroly jakosti pro odhalení závažně vadných prvků takového systému. Ukážeme, že pomocí těchto postupů není možné statistickými prostředky zodpovědně zodpovědět na otázku týkající se pravděpodobnosti výskytu závažných chyb jednotlivých prvků.

Integrovaný statistický informačný systém (IŠIS) - systém na zber, spracovanie a poskytovanie údajov Štatistickým úradom SR

Iveta Stankovičová, Iveta Syneková (SŠDS)



Meranie reálnej konvergenie - Disponibilné dôchodky domácností ako alterantívny indikátor

Viliam Páleník

Reálna konvergenca krajín alebo regiónov sa zvyčajne meria porovnaním hrubých domácich produktov na obyvateľa v parite kúpnej sily, čo je vhodné pri porovnávaní hospodárskej výkonnosti. Ak je cieľom porovnania životná úroveň obyvateľstva ako alternatívny indikátor je v práci diskutovaný disponibilný dôchodok domácností z hľadiska ich nápočtu v systéme národných účtov a u vybranej skupiny krajín a sledované obdobie podľa disponibilných dôchodkov domácností na obyvateľa v parite kúpnej sily a v podiele diponibilných dôchodkov domácností na hrubom národnom dôchodku.

Modely pro podporu výuky vysokoškolské statistiky

Josef Košťálek, Pavla Koťátková Stránská (VŠCHT v Praze)

Mezi hlavní problémy patří, že nulové tolerance vad klasickými statistickými postupy bohužel nelze z principu dosáhnout. Důvodem k tomuto tvrzení je na jedné straně neschopnost odhalit řídké jevy za použití klasických výběrových šetření. Druhým důvodem je nevhodnost použití klasických přejímacích postupů v případě, kdy povolené tolerance vadné výroby a dohodnutá rizika jsou velmi malé. Oběma těmito problémy se budeme v přednášce podrobně zabývat.

Postoj študentov ekonomickej fakulty k štatistike

Erika Liptáková (Ekonomická fakulta Technickej univerzity v Košiciach)

Príspevok publikuje postoje a vnímanie štatistiky z pohľadu študentov. Prieskum ich postojov bol realizovaný medzi študentmi Ekonomickej fakulty TUKE v letnom semestri akademického roku 2023/2024 na základe štandardizovaného dotazníka SATS (Survey of Attitudes Toward Statistics). Bolo vypočítané priemerné skóre dosiahnuté v šiestich komponentoch vnímania štatistiky – Pocity, Vedomosti, Význam, Obťažnosť, Záujem a Úsilie. Zistené výsledky boli porovnávané s predchádzajúcim prieskumom, ktorý sa realizoval na tej istej fakulte pred šiestimi rokmi, v akademickom roku 2017/2018.

Sudoku s překryvy

Pavel Stříž (Nakladatelství Martin Stříž)

V povídání autor představí sudoku s překryvy: jejich generování, kontrolu jedinečnosti řešení a také řešení buňku po buňce s komentářem, která hráčská strategie je použita. Mezi nástroji se objeví programy Sugen (Daniel Beer), Picat (Hakan Kjellerstrand) a Sudoku (Kyle Gough). K automatizaci autor užil Bash, Python, Lua, C a LuaLaTeX. Jádrem povídání bude si představit podmínky kladené na specifická sudoku (olympijské kruhy, logo firmy Husqvarna, měsíční kalendář a 6 sudoku na krychli). Autor se snaží najít zobecněnou metodu vypsání podmínek, postupně představí dílčí úvahy a mezikroky s jejich mantinely. Tím se povídání dostane k nejtěžším typům klasického sudoku: 3D sudoku a dvouročnímu kalendáři.

