

E-recept lze tedy k 1. 1. 2018 hodnotit jako funkční, nepřinášející žádné benefity a zdržující.

3. Jaké jsou perspektivy využití e-preskripce

Zcela jistě je na prvním místě dosažení rychlejší funkčnosti a bezpečnosti preskripce, tedy nezdržování mnoha kliky, využití informace o předepsaných lécích, jejich cenách a jejich pro pacienta bezpečná editace.

Lékový záznam se musí stát klíčovou součástí zdravotnické dokumentace. Hodnotit tedy lze, že e-recept je cesta správným směrem. Jeho provázanost s dalším zdravotnickým softwarem je zatím bohužel malá. Již dnes však lékařům ukazuje, jakým směrem se elektronizace zdravotnictví bude vyvíjet a jak bezpečné zdravotnictví může být.

WEBOVÁ APLIKACE POSKYTUJÍCÍ PODPORU SPRÁVNÉHO VÝBĚRU A REALIZACE STATISTICKÉ METODY

Lubomír Štěpánek, Čestmír Štuka, Martin Vejražka

Anotace

Každý badatel čas od času čelí problému správného výběru statistické metody, která by měla být aplikována, aby vhodně analyzovala jeho data. Naší snahou je zjednodušit tento proces ve smyslu vytvoření webové aplikace, která nabízí pomocnou ruku výzkumníkovi, který řeší statistickou analýzu ve své práci.

Aplikace je vytvořena pomocí jazyka R a jednoho z jeho balíčků, Shiny, a je volně dostupná online. R je otevřený programovací jazyk a prostředí věnované zejména statistickým výpočtům a následným grafickým náhledům.

Aplikace je dostupná na R-serveru 1. lékařské fakulty Univerzity Karlovy na http://shiny.statstat.cz:3838/statisticke_nastroje/ a je založena na vývojovém diagramu navigujícím k aplikovatelné statistické metodě. Integrovaný vývojový diagram umožňuje výzkumníkovi vybrat nejvhodnější statistickou metodu ve smyslu průchodu diagramem krok po kroku tak, jak výzkumník dělá rozhodnutí odpovídáním na kladené otázky. Po určení nejvhodnější metody je přímým odkazem přesměrován na záložku aplikace, která nabízí provedení dané metody nad daty, která badatel do aplikace nahrál.

Poskytnutí uživatelsky přátelské aplikace a vývojového diagramu mapujícího rozhodovací proces potenciálních statistických metod a vybírající tu nejvhodnější považujeme za cestu, jak tento proces zjednodušit. Přímé spojení – pomocí odkazu – mezi návrhem metody a záložkou s její realizací může zlepšit uživatelský komfort.

Klíčová slova

statistické výpočty, jazyk R, Shiny, webová aplikace, biostatistika, statistická podpora, statistické rozhodování, vývojový diagram

1. Úvod

Úloha výběru nejvhodnější statistické metody pro účely analýzy dat badatele – v závislosti na charakteru jím nasbíraných dat a jím postulovaných hypotéz – je stále stěžejní fází přípravy a realizace výzkumné práce. Vhodnost volby statistické metody pro účely dané daty a hypotézami současně ovlivňuje to, zda bude publikace obsahující realizaci dané analýzy nakonec přijata recenzním řízením k publikování v odborném periodiku. Badatelé obvykle neprocházejí systematickou přípravou stran aplikované statistiky, proto je mnohdy výběr metody pro statistické zpracování daných dat (podle daných hypotéz) ovlivněn výzkumníkovou erudicí, empirií, technickými možnostmi, možná i zvyklostmi pracoviště nebo jeho subjektivní oblíbeností některých statistických přístupů.

Ve snaze zjednodušit celý uvedený proces byla navržena a implementována webová aplikace, jež podává pomocnou ruku badateli, který čelí otázce výběru a vypracování statistického modelu ve své analýze. Grafické rozhraní

a workflow práce s aplikací je založeno na použití vývojového diagramu, který badatele naviguje k výběru a použití kontextově vhodné statistické metody.

Nástrojů pro statistickou analýzu dat nejen v biomedicíně je na trhu dostupná celá řada [1]. Jde buďto o „klikací“ software typu IBM SPSS [2], Dell Statistica [3], SAS [4], Minitab [5] či do jisté míry i MS Excel [6] a mnohé další, anebo o pokročilejší nástroje umožňující vlastní skriptování, programování a řešení rozsáhlých a komplexních úloh – například prostředí Matlab [7] a Stata [8] či jazyk R se svými základními knihovnami [9] a jazyk Python s knihovnami SciPy a NumPy [10], ale i mnohé další.

Výhodou „klikacích“ nástrojů je jejich uživatelská přívětivost, obvykle dobrá stabilita a mnohdy i solidní podpora koncových uživatelů, nevýhodou je pak relativní uzavřenost daná konečným počtem nabízených metod a postupů, a tedy nemožnost řešit velmi komplexní a různé individuální úlohy. Část nástrojů je vyvíjena komerčně; zpoplatnění licencí může být limitací pro jejich širší použití, zejména mezi uživateli-jednotlivci, kam patří i vysokoškolští pregraduální či postgraduální studenti, kteří obzvláště často řeší analytické úlohy v rámci svého disertačního výzkumu.

Snaha o integraci nějaké konkrétnější podoby podpory správného výběru vhodné statistické metody, která se nejlépe hodí pro badatelova data a hypotézy, je naopak relativně vzácná. Produkty obvykle pouze nabízejí (mnohdy podrobný) manuál k ovládání uživatelského rozhraní – ten však sám o sobě může poskytovat kvalitní doménové informace přímo o statistických a analytických metodách a příkladech jejich použití. Vyžaduje ale často relativně dlouhý čas a velké úsilí k nastudování určité skupiny metod a následně k výběru té z nich, která je pro danou analytickou situaci nejvhodnější.

Cílem práce autorů je poskytnout dobře dostupný, nekomerční nástroj pro co možná nejsnazší výběr a následnou aplikaci statistické metody, která je vhodná pro výzkumný záměr a hypotézy badatele. Cestou k takovému nástroji je dle autorů vývoj aplikace dostupné online a nevyžadující instalaci ani licenci. Repertoárem implementovaných a nabízených metod je aplikace zaměřena na oblast biomedicíny, přesah s jinými obory, např. psychologií, sociologií či ekonomikou, je však nezanedbatelný. Práce s aplikací by měla být intuitivní a neměla by vyžadovat nutnost pročítání dlouhých manuálů, které jsou obvykle dodávány ke komerčním statistickým balíkům. Těmto ideám se snaží aktuální a dále popsaná verze vyvíjené aplikace co nejvíce přiblížit.

2. Aplikace statisticke_nastroje

Aplikace je volně dostupná online, nevyžaduje instalaci ani žádnou licenci. Jádro aplikace je napsáno v jazyce R. Uživatelské rozhraní je přátelské a splňuje požadavky kladené na současné webové aplikace; je rozčleněno do několika záložek s různými funkcionalitami a featurami. Implementované metody pokrývají běžnou statistickou metodiku používanou zejména v biomedicíně, a to především na úrovni závěrečných prací pregraduálních studentů. Jazykem uživatelského rozhraní a ovládacích prvků aplikace je čeština.

2.1 Jazyk R a jeho role v aplikaci

Jazyk R je „free-as-in-beer“ a „free-as-in-speech“, tj. je zdarma ke stažení a je možné svobodně upravovat jeho zdrojový kód [9], nebo ho dokonce i přes původní licenci dále použít komerčně. R je jednak programovací jazyk, jednak prostředí pro statistické výpočty a grafiku a je široce využíváno mezi statistiky, ekonometry či biologií [9]. Dalšími výhodami používání R je silná komunita kolem jazyka (a tedy snadné nalezení odpovědí na dotazy v rámci inforatických online fór), rychlý vývoj balíčků s novými funkcionalitami, možnost snadného vytváření online aplikací nad kódem R (což je i případ představené aplikace) či snadné sázení TeXových dokumentů přímo v R. Dle některých statistik portálu kaggle.com je R v oblasti analýzy dat, data science nebo machine learning celosvětově vůbec nejpoužívanějším nástrojem [11]; v této souvislosti se o jazyku R někdy hovoří jako o lingua franca oboru data science a je použit jako hlavní analytický nástroj například v pracích [12, 13, 14, 15, 16, 17].

Offline napsaný kód v jazyce R byl s použitím balíčku Shiny nahrán online na server dedikovaný pro R-kové výpočty. Balíček Shiny je také napsán v R a poskytuje online framework pro R-kové skripty. Aplikace statisticke_nastroje je dostupná na R-serveru 1. lékařské fakulty Univerzity Karlovy na adrese

http://shiny.statet.cz:3838/statisticke_nastroje/

Užití jazyka R zajišťuje reliabilitu výpočtů prováděných zmíněnou aplikací a rovněž umožňuje díky možnému zapojení dalších R-kových balíčků do aplikace použít veškerých výpočetních možností tohoto nástroje.

Díky volné licenci jazyka R je možné takovou licenci adoptovat i v aplikaci a nabízet ji tedy volně dostupnou online na zmíněné adrese.

2.2 Uživatelské rozhraní

Uživatelské rozhraní aplikace odpovídá současným nárokům na webové utility (Obr. 1).

Roli back-endu hraje v aplikaci již zmíněné R, roli front-endu pak R-kový balíček Shiny, dále HTML, CSS a javascript, díky kterým je aplikace uživatelsky komfortnější.

Posledně jmenovaný umožňuje mít v aplikaci busy indikátor (symbol přesýpacích hodin při zaneprázdnění aplikace) a další funkcionality jako pop-up vyskakovací hlášky, thumbnailové texty vyskakující při přejíždění („hover-over“) nad některými aktivními texty apod. Aplikace je obdobě vybavena počítadlem návštěv, kdy návštěvou je rozuměno jedno připojení na adresu aplikace (včetně obnovení dané stránky). Díky implementované třídě CSS zvané bootstrap, kterou lze k R-kové shiny aplikaci připojit, je aplikace funkční i na mobilních zařízeních a je zajištěn její responzivní design.

Aplikace používá češtinu jako jazyk volby, neboť v tuto chvíli je cílíme na česky mluvící výzkumníky a studenty.

Layout aplikace tvoří několik záložek, které zleva doprava reflektují očekávané work-flow práce s aplikací.

První záložka je úvodní, obsahuje informace o účelu a fungování aplikace a zmenšeninu vývojového diagramu s doporučeným postupem vhodné volby finální statistické metody dle vstupních dat a hypotéz (Obr. 1).

Druhá záložka umožňuje nahrání vlastních uživatelských dat ve formátu .csv nebo .txt. Volba parametrů podle vstupního formátu dat zajistí jejich správné načtení. Eventuálně je možné zvolit pro edukační účely i jeden z vestavěných datasetů (Obr. 2).

Třetí záložka zleva pak představuje vývojový diagram, který na základě zodpovězení některých vybraných otázek týkajících se vstupních dat nebo hypotéz nabádá k volbě určité statistické metody. Tyto metody jsou posléze implementovány na dalších záložkách. Pro pohodlí uživatele je možné se na danou záložku přesunout poklepekem na ikonu pojmenovanou stejně jako koncový uzel diagramu s doporučenou metodou (Obr. 3).

Další záložky představují již implementace konkrétních metod odpovídajících názvům záložek (Obr. 4) a slouží k realizaci metody na data nahraná pomocí druhé záložky.

2.3 Vývojový diagram

Vývojový diagram integrovaný do aplikace a dostupný na třetí záložce zleva pomáhá učinit správná statistická rozhodnutí v závislosti na použitých datech a zamýšlených statistických hypotézách. Design diagramu vychází jednak ze zásad „dobré praxe“ problematiky, jednak ze zkušeností hlavního autora v této oblasti; korektnost závěrů vývojového diagramu lze porovnat s doporučeními, která pro daný typ dat a hypotéz navrhuje např. [18].

Klíčovou komponentou aplikace je zmíněný vývojový diagram (Obr. 5). Umožňuje výzkumníkovi, který již nasbíral data a formuloval výzkumnou hypotézu nebo hypotézy, vybrat pokud možno nejvíce odpovídající statistickou metodu, a to tak, že výzkumník projde diagramem v naznačeném směru (shora dolů) na základě odpovědí na pokládané otázky. Ve chvíli, kdy se výzkumník dostane postupným (a optimálně správným) odpovídáním na otázky

v diagramu až k jeho koncovému uzlu, je zde připraven odkaz na záložku, která vede k dané statistické metodě umožňující posléze provést celou analýzu.

Jsou-li do aplikace nahrána online adekvátní data, analýza proběhne a uživateli je poté nabídnuta možnost stáhnout offline výstupy analýzy – tabulky se sumáři, diagramy s infografikou.

Hlavní rozhodovací uzly vývojového diagramu vyžadují zodpovězení otázek týkajících se typů a struktur nasbíraných dat, empirického rozložení dat, počtu porovnávaných výběrů a toho, zda by analýza měla porovnávat střední hodnoty nebo míry variability dat.

2.4 Implementované metody

Statistické metody, které jsou implementovány (Tab. 1), pokrývají jednak parametrické testy jako t-testy, F-testy nebo rodinu ANOVA a další; jednak neparametrické metody jako Wilcoxonovy testy, Kruskal-Wallisův test nebo Friedmanův test a další. Metody jsou přímo linkovány z koncových bodů diagramu k záložkám, které pak výpočet dané metody na vložených datech provádí.

parametrické metody	neparametrické metody
t-testy	Wilcoxonovy testy
F-testy	Kruskal-Wallisův test
χ^2 -testy	Friedmanův test
AN(C)OVA	...
Shapiro-Wilkův test	

Tabulka 1 – Přehled implementovaných parametrických a neparametrických metod v aplikaci *statisticke_nastroje*.

3. Diskuze

Přínos aplikace je momentálně předpokládán zejména pro uživatele z řad bakalářských a magisterských studentů různých oborů (bio)medicíny, kteří hledají relativně jednoduchý, přesto použitelný nástroj pro vypracování statistické analýzy do závěrečné práce.

Výhodou pro ně může být, že je aplikace zcela volně přístupná, není omezena žádnou licencí. Dalším plusem může být relativně velký prostor pro rychlé doplnění (doimplementování) chybějící, nicméně žádané metody. Rozsah metod, které je možné do aplikace doimplementovat a zahrnout, je limitován v podstatě jen možnostmi jazyka R a možnostmi všeobecného webového rozhraní založeného na HTML5, CSS3 a javascriptu.

Příkladem uveďme dva klasické problémy, se kterými studenti „bojují“ a které aplikace překlenuje – (i) zpracovává-li student analýzu pouze s využitím MS Excelu® (když mu např. jiný nástroj není z nějakého důvodu dostupný), brzy

zjistí, že v MS Excel® není možné vykreslit krabicový diagram. Mnohdy (ii) rovněž studenti v rámci analýzy kategoriálních dat potřebují analyzovat pomocí testu hypotézy (zde χ^2 -testu nebo Fisherova přesného testu) kontingenční tabulku o rozměrech daleko větších než pouze 2×2 pole, jak je nabízí tradiční komerční statistické balíčky. Vznikl tak i mýtus, že Fisherův přesný test existuje pouze pro čtyřpolní tabulku 2×2 a pro větší není vůbec definován. Oběma nesnázím se lze s použitím předvedené aplikace celkem snadno vyhnout.

4. Další vývoj

Projekt je stále ve vývoji a může být interaktivně updatován v návaznosti na zpětnou vazbu uživatelů. Zpětná vazba uživatelů je autory již zaznamenána. Do budoucna je v plánu implementovat (semi)automatické doporučení statistické metody už na bázi charakteru dat, které uživatel aplikaci nabídne a eventuálně (před)připraví. Nabízí se rovněž překlad aplikace do anglického jazyka.

V úvahu připadá též doplnění záložky pro (pre)processing dat. Vzhledem k unikátnosti každého předzpracování dat je obtížné vytvořit jednotný „klikací“ přístup – záložka s (pre)processingem by tedy měla podobu textové arey, kam by se vkládal kód s R-kovou syntaxí a ten by data svou exekucí předzpracoval. Kód v tomto případě dovoluje využít prakticky neomezené možnosti úpravy dat nástrojem R, zatímco „klikací“ rozhraní by bylo omezeno jen počtem procedur daných počtem tlačítek, na než lze poklepat.

5. Závěr

Kvalita statistického zpracování výzkumné práce v odborném sdělení má bezesporu vliv na celkový dojem z publikace. Prvním krokem každé statistické analýzy je volba správného statistického modelu, který odpovídá vstupním datům a hypotézám badatele. Tento krok je však trikový a mnohdy netriviální i pro zkušeného výzkumníka. Uživatelsky přátelská aplikace, která byla vyvinuta, se snaží tento krok co nejvíce zjednodušit. Propojení mezi doporučením konkrétní statistické metody, vyplývající z vývojového diagramu, a záložkou, která analýzu danou metodou reálně provede, může zlepšit uživatelský komfort aplikace.

Literatura

- [1.] ALI, Zulfiqar a SBala BHASKAR. *Basic statistical tools in research and data analysis*. *Indian Journal of Anaesthesia* [online]. 2016, 60(9), 662. Dostupné z: doi:10.4103/0019-65049.190623
- [2.] IBM, Corp. *IBM SPSS Statistics for Windows, Version 22.0*. B.m.: Armonk, NY: IBM Corp. 2013
- [3.] DELL, Software. *Statistica, Version 13.0*. 2015
- [4.] SAS INSTITUTE INC. *SAS 9.4 Language Reference: Concepts*. Cary, NC, USA: SAS Institute Inc., 2013. ISBN 1612905641, 9781612905648.
- [5.] MINITAB, Inc. *Minitab 17 Statistical Software*. B.m.: State College, PA. 2010

- [6.] MICROSOFT. *MS Office Excel, version 16*. 2016
- [7.] THOMPSON, C. M. a L. SHURE. *Global Optimization Toolbox: User's Guide*. B.m.: MathWorks. 2011
- [8.] STATA CORP. *Stata Statistical Software: Release 14* [online]. 2015. ISBN 1881228541. Dostupné z: doi:10.2307/2234838
- [9.] R CORE TEAM. *R: A Language and Environment for Statistical Computing* [online]. Vienna, Austria: R Foundation for Statistical Computing, 2016. Dostupné z: <https://www.R-project.org/>
- [10.] FOUNDATION, Python Software. *Python Language Reference, version 2.7* [online]. 2017. Dostupné z: <http://www.python.org>
- [11.] Tools Used By Competitors. *Kaggle.com* [online]. 2017 [cit. 2018-03-02]. Dostupné z: <https://www.kaggle.com/wiki/DataScientistTools>
- [12.] MARTINKOVÁ, Patricia, Lubomír ŠTĚPÁNEK, Adéla DRABINOVÁ, Jakub HOUDEK, Martin VEJRAŽKA a Čestmír ŠTUKA. *Semi-real-time analyses of item characteristics for medical school admission tests* [online]. In: 2017-9-24, s. 189–194 [cit. 2018-03-02]. DOI: 10.15439/2017F380. Dostupné z: <https://fedcsis.org/proceedings/2017/drp/380.html>
- [13.] KASAL, Pavel — FIALA, Patrik — ŠTĚPÁNEK, Lubomír, et al. *Application of image analysis for clinical evaluation of facial structures*. In: *MEDSOFT ...: sborník příspěvků*. Praha: Dům techniky ČSVTS, 1989–. 2015, s. 64–70. ISSN: cnb001577313.
- [14.] KASAL, Pavel — SHALA, Laura — ŠTĚPÁNEK, Lubomír — MĚŠŤÁK, Jan. *Model posuzování funkční morfologie výrazů obličeje*. In: *MEDSOFT ...: sborník příspěvků*. Praha: Dům techniky ČSVTS, 1989–. 2016, s. 30–37. ISSN: cnb001577313.
- [15.] KASAL, Pavel — MĚŠŤÁK, Jan — ŠTĚPÁNEK, Lubomír, et al. *Analýza obrazu pro hodnocení efektu rhinoplastiky*. In: *MEDSOFT ...: sborník příspěvků*. Praha: Dům techniky ČSVTS, 1989–. 2014, s. 90–96. ISSN: cnb001577313.
- [16.] FIALA, Patrik — KASAL, Pavel — ŠTĚPÁNEK, Lubomír — MĚŠŤÁK, Jan. *Analýza obrazu v hodnocení symetrie obličeje*. In: *MEDSOFT ...: sborník příspěvků*. Praha: Dům techniky ČSVTS, 1989–. 2014, s. 36–40. ISSN: cnb001577313.
- [17.] PAZDÍREK, Filip — LEFFLER, Jan — ŠTĚPÁNEK, Lubomír, et al. *Dlouhodobé výsledky po radikálních resekcích pro adenokarcinom žaludku*. *Gastroenterologie a hepatologie*, 2016, roč. 70, č. 5, s. 402–407. ISSN: 1804-7874. DOI: 10.14735/amgh2016402.
- [18.] HENDL, Jan. *Přehled statistických metod: analýza a metaanalýza dat*. Praha: Portál, 2015. ISBN 978-80-262-0981-2.

Kontakt

MUDr. Lubomír Štěpánek

Oddělení biomedicínské statistiky

Ústav biofyziky a informatiky

1. lékařská fakulta

Univerzita Karlova

email: lubomir.stepanek@lf1.cuni.cz

RNDr. Čestmír Štuka, MBA, PhD.

Ústav biofyziky a informatiky

1. lékařská fakulta

Univerzita Karlova

MUDr. Martin Vejražka, PhD.

Ústav lékařské biochemie a laboratorní
diagnostiky

1. lékařská fakulta

Univerzita Karlova a Všeobecná

fakultní nemocnice v Praze