

## Ako sa nemá kritizovať a ako sa má dôkladne si overovať tvrdenia

Martin Kanovský

Fakulta sociálnych a ekonomických vied, Univerzita Komenského, Bratislava

Veľmi nerád reagujem na akýkoľvek druh kritiky, ktorý sa nesústreďí výhradne na vecné aspekty a ponára so do osobných sebaaprezentácií, sebaokiadzaní a dehonestácií oponenta. Nevadí mi ostrejší tón: problémom takýchto kritik (a kritikov) je to, že ich presvedčenie o vlastnej výnimočnosti zhoršuje úsudok a vedie k fantazijnému vnímaniu reality – ak niečo hoci aj bije do očí, pre nich to neexistuje. To sa v plnej miere prejavilo aj v kritickej stati Soukupa.

Začnime teda niečím, čo bije do očí absolútne: v Soukupovom texte sa pravidelne opakujú osobné poukazy, ako bol „v československém prostoru první, kdo o těchto modelech detailněji referoval“, ako o nich učil, ako písal, aký je „dlouholetý učitel a autor textů z oblasti sociálněvědní analýzy dat“. Ja nemám ani najmenší dôvod o tom pochybovať, ani to nijako nespochybňujem, no je mi záhadou, prečo si Soukup myslí, že to je vecne relevantné. Oveľa závažnejšie však je, ak z toho vznikne guru efekt, ktorý vedie k nepravdivým osobným denunciaciám oponenta. Soukup sa takto vyjadruje o mojom vedeckom profile: „...medailonek autora, kde píše, že se kromě antropologie věnuje využívání robustních a bayesovských metod v sociálních vědách. Při náhledu na publikační a výukový profil autora se mi nic podobného nepodařilo nalézt.“ Pozrime sa teda na štandardné vedecké databázy, ako sú Web of Science a Scopus, ako aj na verejný portál Google Scholar, prípadne verejnú evidenciu publikačnej činnosti Univerzity Komenského či slovenský centrálny register publikačnej činnosti CREPČ (osobne si neviem predstaviť, kde inde by ktokoľvek mohol niečo hľadať o publikačnom profile), za ostatných 5 rokov (2015 – 2019). V databázach WoS a Scopus sa nachádza za toto obdobie 17 mojich publikačných jednotiek, z toho 7 (WoS), resp. 8 (Scopus) obsahuje už v názvoch výrazy „robust (IRT) psychometric analysis“ a podobné výrazy, pričom aj zvyšné štúdie sú založené na štatistických analýzach. V databáze Google Scholar sa za roky 2015 – 2019 nachádza 23 mojich publikačných jednotiek, z toho 10 má tiež priamo v názve „psychometric analysis“ a pod., vrátane dvoch monografií: „Robustné štatistické analýzy v sociálnych vedách“ (Bratislava, SASA 2016) a „Sebasúcit a sebakritickosť: Psychometrická analýza meracích nástrojov“ (spoluautorka J. Halamová, Bratislava, UK 2017). Je ťažké uveriť, že Soukup sa nepozrel ani do jedinej publikačnej databázy – ak to urobil a napriek tomu toto verejne vyhlásil, jeho dehonestácie nielen výrazne prekračujú rámec vedeckej etiky, ale dostávajú sa na hranu právnej kvalifikácie. Ja som však napriek tomu ochotný veriť, že si nič neoveril – no v tomto prípade ide o takú zarážajúcu neschopnosť overovania dát, že ani

pozícia „dlouholetého učitele a autora textů z oblasti sociálněvědní analýzy dat“ nestojí na pevných základech.

Týmto by sme mohli skončiť, pretože ak niekto nie je schopný overiť si elementárne veci, a/alebo nepravdivo osobne denuncuje a dehonestuje oponenta, nemá zmysel diskutovať. Napriek tomu sa vyjadrim k niektorým Soukupovým tvrdeniam, pretože aj pri nich sa prejavuje jeho zarážajúco laxná nevšímavosť a tendencia ustavične sa uchýľovať k osobným dehonestáciám. Začnem Buishandovým U-testom časového zlomu. Soukup píše: „...autor sice udáva vzorce (neudáva odkud je převzal, z původního textu nejsou, tam je symbolika jiná)”. Tak po prvé: na s. 433 svojho článku uvádzam hneď pri spomenutí U-testu dve citácie (Buishand 1984; Lee 2010), odkiaľ som vzorce prevzal, takže „neudáva odkud je převzal“ je zjavným klamstvom. Po druhé, Soukup sa žiaľ pri čítaní Buishandovho pôvodného článku dostal iba po s. 2, kde (bez akejkoľvek referencie) plagiuje Buishandov príklad a vydáva ho za svoj vlastný (Buishand, s. 52: „change in the gauge datum... or due to a station relocation“; Soukup: „změna přístroje či stanoviště“), no ďalej už nečítal, pretože by mu predsa muselo byť jasné, že vzorce, ktoré uvádzam vo svojom článku (s. 433), sú priamo prevzaté z Buishandovho článku, no Buishand ich neuviedol všetky na jednom mieste, ale na viacerých miestach ďalej v svojom texte, čo by Soukup (aspoň dúfam) určite zistil, keby si preštudoval Buishandov článok celý a nepodľahol (ako zvyčajne) sklonu k osobnej dehonestácii oponenta. Nebudem tu vzorce prepisovať, ale uvediem v podobe obrázkov:

Obr. č. 1: Buishand, rovnica (1), s. 52 (hore); Kanovský, vzorec 1, s. 433 (dolu)

$$y_i = \begin{cases} \mu + \epsilon_i, & i = 1, \dots, m \\ \mu + \Delta + \epsilon_i, & i = m + 1, \dots, n \end{cases}$$

$$x_i \begin{cases} \mu + \epsilon_i & i = 1, \dots, m \\ \mu + \Delta + \epsilon_i & i = m + 1, \dots, n \end{cases}$$

Obr. č. 2: Buishand, rovnica (3), s. 52 (vľavo); Kanovský, vzorec 2, s. 433 (vpravo)

$$S_k^* = \sum_{i=1}^k (y_i - \bar{y}) \quad S_k = \sum_{i=1}^k (x_i - \hat{x})$$

Obr. č. 3: Buishand, rovnica (7), s. 53 (hore); Kanovský, vzorec 4, s. 433 (dolu)

$$D_y^2 = n^{-1} \sum_{i=1}^n (y_i - \bar{y})^2$$

$$D_x = \sqrt{n^{-1} \sum_{i=1}^n (x_i - \hat{x})^2}$$

Obr. č. 4: Buishand, rovnica (8), s. 54 (hore); Kanovský, vzorec 3, s. 433 (dolu)

$$U = [n(n+1)]^{-1} \sum_{k=1}^{n-1} (S_k^*/D_y)^2$$

$$U = [n(n+1)]^{-1} \sum_{k=1}^{n-1} \left(\frac{S_k}{D_x}\right)^2$$

Jasne vidíme, že síce podľa Soukupa tieto vzorce „z pôvodného textu nejsou“, no keďže Buishand tak trochu staromódne predpokladal, že jeho text sa bude čítať celý, nielen jeden a pol strany, a uviedol svoje vzorce až na s. 52 až 54, unikli Soukupovmu na overovanie dát takému citlivému oku. A malý dovetok: Buishand používa index  $y$  a ja som použil (tak ako mnou citovaní Lee a Pohlert) index  $x$ , pretože Buishand uviedol aj rozšírenie rovnice (1) na regresný model v rovnici (2), kde používa index  $x$ . Samozrejme, z toho nevyplýva, že „tam je symbolika jiná“, a takisto to nevyplýva ani z obr. č. 3, kde Buishand umocnil ľavú stranu rovnice a iní autori (vrátane mňa) odmocnili pravú stranu. Týmto však určite nechcem vyslovovať akékoľvek podozrenie, že Soukup sa dostal na tretiu či dokonca štvrtú stranu Buishandovho textu.

Soukup ďalej píše: „Test pro body zlomu, ale meritorně slouží pro jiné typy dat, než užívá autor (dlouhé meteorologické řady) a snaží se nalézt jeden bod, kdy se řada z určité průměrné úrovně dostane na jinou.“ Je síce faktom, že sa tento test používa (aj) pre dlhé meteorologické rady, no používa sa aj pre krátke časové rady. Keby Soukup čítal Buishandov článok celý, zistil by, že Buishand odvodil (tabuľka č. 3, s. 54) kritické hodnoty pre U-test pre rôzne veľkosti časových radov, počnúc  $n = 10$ . Testy časových zlomov sa používajú aj pre kratšie časové rady, napríklad počet mrazivých dní v Mníchove v rokoch 1957 – 1968 ( $n = 12$ ), pozri Pohlert 2017, alebo behaviorálno-ekologické časové rady za 12 mesiacov (Bru et al. 2011). Soukup teda bez akejkoľvek znalosti a vecnej argumentácie iba deklaruje, že výsledky U-testu sú chybné alebo že stačí pohľad na graf a zdravý sedliacky rozum. Niečo podobné si istotne mysleli a vraveli aj skladníci a účtovníci v Guinesse, keď Gosset zaviedol t-test na porovnávanie priemerov pri veľmi malých vzorkách. Buishandov U-test rieši netriviálny problém, a Buishand by sa neobťažoval v rokoch 1982 a 1984 odvodzovať pomocou (vtedy bez počítačov s veľkou výpočtovou silou)

pomerne náročných Monte Carlo simulácií kritické hodnoty pre časové rady s 10, 20 alebo 30 hodnotami. Soukup má zrejme veľmi dobré intuície praktického výskumníka (k tomu sa ešte dostaneme), ale jeho limitované matematické znalosti a nepozornosť vyvolaná snahou o osobné útoky spôsobujú, že niektoré dôležité vecné problémy mu unikajú. Príklad: Soukup podozrivo chvatne prechádza neparametrický Mann-Kendallov test, ktorý som použil na overenie monotónneho trendu (nie časového zlomu). Keby chcel relevantne namietat', mohol by sa spýtať, prečo Kanovský nepoužil aj na overenie časového zlomu neparametrický test (napr. Pettittov test, Pettitt 1979). Soukup však uvádza aj vecnú námietku, respektíve skôr obavu: „Je namieste se ptát, jaká byla síla použitých statistických testů (osobně odhaduji max. 20-30 %)”. Súhlasím, že ide o relevantnú otázku – avšak rozhodne by som ju neponechával na „osobný odhad“ Soukupa. V mojom článku (Tabuľka č. 2) má uvedené v prvom stĺpci sprava rozsahy za každú vysokú školu, vie, že ide o 10 údajov, všetky použité testy má k dispozícii (teraz už aj tie, ktoré podľa neho „nejsou“ v pôvodnom texte), rovnako ako Buishandove kritické hodnoty pre úrovne  $\alpha$  0,01, 0,05, 0,10 v tabuľke č. 1 (Buishand 1984: 54), takže mu nič nemôže brániť, aby uskutočnil Monte Carlo simuláciu a určil štatistickú silu týchto testov pre malé vzorky. Tak ako sa mi nepozdávajú osobné útoky, nepozdávajú sa mi ani osobné odhady.

Teraz sa pozrime na viacúrovňové (multilevel) modely. Úplne v úvode sa zdá, že Soukup tematizuje skutočne relevantný problém, ktorý vzniká pri testovaní populačných dát viacúrovňovými modelmi: „I pokud bychom připustili spíše okrajový názor o superpopulacích a užívání statistických testů pro celopopulační data (autor recenzního příspěvku jej nesdílí a preferuje pro tyto situace popisnou statistiku)...“ Zdieľam so Soukupom názor, že používanie viacúrovňových modelov pre celopopulačné dáta je problematické – teoretický a metodologický základ pre tieto modely predpokladá, že pre jednotky druhej (či tretej alebo n-tej) úrovne ide o ich náhodný výber z populácie: jednotlivci s viacnásobnými odpoveďami, triedy, školy, regióny, krajiny. Ak dátový súbor obsahuje prakticky všetky alebo dokonca reálne všetky možné jednotky (celopopulačné testovanie na všetkých školách v danej krajine, veľké prieskumy OECD typu PISA, TIMSS, PIRLS, európske prieskumy hodnôt vo všetkých európskych krajinách, volebné výsledky vo všetkých okrskoch a okresoch, atď.), nedajú sa výsledky takýchto modelov štatisticky generalizovať na hypotetickú populáciu a frekventistické pojmy ako štatistická významnosť, štandardné chyby odhadov, intervaly spoľahlivosti a podobne nie sú použiteľné (výrazne iná situácia je pri použití bayesovských modelov, no tu nie je priestor na zachádzanie do detailov). Dost' ma teda prekvapilo, keď sa ich v prípade môjho viacúrovňového modelu zrazu Soukup nástojčivo dožaduje: „I když se autor v předchozích analýzách zarytě drží výsledků statistické významnosti

(P hodnot), u víceúrovňového modelu je čtenář nenajde. Nenajde v zásadě žádný výsledek sloužící ke zobecnění (standardní chyby odhadů, hodnoty testových kritérií či intervaly spolehlivosti).“

Toto je však drobnost: ovela zásadnejším problémom je, že Soukup (zrazu z ničoho nič opúšťajúci prvú osobu a skrývajúci sa za „čtenáře“) priznáva, že nerozumie modelu multivariačnej t-distribúcie: „Náznak rovnice je na straně 434, ale nejde o rovnici konkrétního analyzovaného modelu, ale o zcela obecnou rovnici, nadto operující s obecnými pojmy (subjekt), maticemi a vektory, které jsou pro čtenáře zcela nesrozumitelné.“ Po prvé, nejde o žiadny náznak, ale o kompletnú rovnicu aj s popisom, a po druhé, ak Soukup nerozumie algebre matic a vektorov pre multivariačnú t-distribúciu, prečo by to mala byť moja chyba a prečo mi to vyčíta? Ak tomu nerozumie, nech si naštuduje tam citovanú literatúru (Nadarajah –Kotz 2008; Lange et al. 1989; Pinheiro et al. 2001) – podotýkam, že viac než prvé strany. Táto hrubá neznalosť a nepochopenie multivariačnej t-distribúcie spôsobuje, že sa ustavične dožaduje indexov a testov, typických pre viacúrovňové lineárne modely s Gaussovou distribúciou: „chybí hodnoty klasických ukazatelů jako je ICC či  $R^2$ “. Tieto indexy sa počítajú na základe (kondicionálnych a reziduálnych) rozptylov a v prípade t-distribúcie je pri ich výpočte potrebné počítat so stupňami voľnosti a škálovými váhami, ako som uviedol v pôvodnom článku: „Výhodou multivariačnej t-distribúcie v porovnaní s multivariačne normálnou (Gaussovou) distribúciou je najmä to, že každý subjekt má svoju vlastnú škálu  $\tau_i$ , ktorá sa nedá pozorovať a musí byť imputovaná z údajov. Tieto odlišné škály vedú k odlišným váham pri odhade populačných parametrov: niektoré subjekty (tie s vyššími reziduálmi) budú mať menšiu váhu pri určovaní odhadov maximálnej vierohodnosti, a to umožňuje tomuto modelu vyhnúť sa devastujúcemu vplyvu extrémnych hodnôt (outliers)“ (s. 434). Hodnoty ICC a  $R^2$  by som síce dokázal vypočítať manuálne pre tento daný model s jeho stupňami voľnosti, no bez generalizácie zohľadňujúcej stupne voľnosti (tá nebola zatiaľ pre multivariačné modely s t-distribúciou podľa mojich vedomostí preskúmaná a odvodená), špecifické hodnoty ICC a  $R^2$  pre tento model by boli zavádzajúce, pretože by sa nedali porovnať s hodnotami pre iné modely s inými stupňami voľnosti. Tu je jasne vidieť, že úplná neznalosť matematických základov štatistického modelovania vedie k absurditám.

Niektoré záležitosti sú iba výsledkom terminologických nejasností. Termíny „fixed effects“ a „random effects“ predstavujú technické termíny: súhlasím, že ide o koeficienty (keby som chcel nasledovať Soukupa, pohrdlivo a teatrálne by som upozornil, že to nie sú ani koeficienty, ale odhady koeficientov). Je úplne zrejmé, že netvrdím, že tieto koeficienty (alebo odhady) v prípade vysokých škôl sú náhodné v zmysle štatistickej pravdepodobnosti: práve preto používam v slovenčine nie celkom šťastný termín „random efekty“, no ide

o úplne bežne používaný pojem. Oveľa vhodnejší by bol v bayesovskej terminológii používaný termín „skupinové efekty“. No ani mne nie je jasné (hoci to nezatrucujem, iba zmieňujem) Soukupovo rozlíšenie „pro koeficienty proměnných první úrovně lze volit mezi fixními a pevnými efekty“, no spolieham sa, že sa to dozviem v jeho dosiaľ nevydanej knihe. A opäť si neodpustím výstrel k sebe samému: Soukup by sa mohol veľmi oprávnene spýtať, prečo Kanovský najskôr ukazuje, že viaceré časové rady v prípade niektorých univerzít ukazujú signifikantné časové zlomy, no potom sa púšťa do lineárneho viacúrovňového modelu s multivariačnou t-distribúciou: ak sú niektoré časové rady nelineárne (napr. kvadratické), asi ťažko bude dobrým modelom pre ne linárny model, hoci aj s t-distribúciou. Takýto nelineárny model (ako uvádzajú aj Pinheiro et al. 2001) by bol podstatne zložitejší: napríklad polynomiály sú citlivé na extrémne hodnoty a penalizované kubické krivky (splines) sú náročné na interpretáciu – príslušné vzorce a ich výklad by boli pre „čtenáře“ asi „zcela nesrozumitelné“, najmä ak by sa nechcel obťažovať preštudovať si referovanú literatúru. A to ešte by niekto mohol hnidopišsky trvať na preverení, či napríklad daná t-distribúcia je vôbec symetrická a nie napríklad šikmá.

Napriek dehonestáčnemu tónu vítam poukaz na slabú diskusiu vecnej signifikancie v mojom pôvodnom článku (to som aj uviedol v slabých stránkach pôvodného článku). Ako som však uviedol aj tam, potrebovali by sme podstatne detailnejšie čiastkové preskúmanie, čo sa vlastne deje na jednotlivých vysokých školách. Niektoré vysoké školy mohli klesnúť vo výskumnej efektívite preto, lebo už nemali veľký zahraničný grant, ktorý v minulom roku mali, iné stúpnuť, lebo prudko zvýšili počet karentovaných článkov. Index výskumnej efektívnosti v mojom článku je iba pomer medzi pomerne zložitým kompozitným údajom (výskumný podiel vysokej školy, daný kombináciou rôznych údajov) a počtom zamestnancov.

Nie celkom rozumiem výhrade, že nezohľadňujem „relatívnosť“ výsledkov: to sa mi zdá byť natoľko triviálne, že som to ani nespomenul – každý pomer je bytostne relatívny, ide o zabudovanú matematickú vlastnosť. Je úplne zrejmé, že môžu nastať situácie, že vysokej škole klesne výskumná efektivita, aj keď sú jej absolútne výsledky presne rovnaké ako v predošlom roku, alebo aj naopak, výskumná efektivita môže zostať približne rovnaká, aj keď absolútne výsledky danej vysokej školy stúpia či klesia. Takéto výhrady sú absurdné: ak sa napríklad analyzujú volebné výsledky prostredníctvom percentuálneho podielu získaných hlasov, je predsa triviálne, že politická strana môže získať vyšší či nižší percentuálny podiel hlasov, či už bude mať počet hlasov voličov vyšší, nižší alebo rovnaký (v závislosti napríklad od účasti). Ešte vypuklejšim by sa to stalo v prípade počtu mandátov, kde by sa brali do úvahy prahy (napr. 5 %) a skrutíniá.

Debatovať však je možné iba vtedy, ak existuje snaha niečo sa dozvedieť, niečo si zistiť a naštudovať a nie iba chuť výsmešne dehonestovať. Napríklad: „Proč jsou jako tvůrčí zaměstnanci bráni jen ti, co mají celý úvazek?“ Pretože iba tí si môžu zaregistrovať publikáciu, čo je uvedené v metodike, na ktorú odkazuje pôvodný článok. Alebo táto pasáž: „Zvláštní je, že závěr... zcela odhlíží od výsledků analýz... a vypadá jako když si autor vyřizuje své účty s tvůrčí slovenské vzdělávací politiky. Namísto vědecké argumentace však nastupují zcela normativní a nepodložená tvrzení, což je pro článek v odborném časopise zcela nepřijatelné.“ Uvediem príslušnú pasáž z pôvodného článku: „Treba však uviesť, že spôsob, ako sa inštitucionálne vykonávajú zmeny vo financovaní, je veľmi nešťastný: ak sa zmení metodika a teda aj kritériá financovania, vysoké školy nedostanú nijaký priestor na zmenu svojich stratégií, pretože zmeny sa zavedú okamžite a platia spätne – napríklad metodika rozpisu dotácií na rok 2017 sa zmení, no táto zmena platí pre indikátory za roky 2014 a 2015, takže vysoké školy nemôžu nijako reagovať: dokonca aj keby táto metodika zostala nezmenená pre rok 2018, do úvahy sa budú brať indikátory za roky 2015 a 2016, ktoré tiež nie je možné nijako zmeniť. Takýto spôsob zavádzania zmien je svetovým unikátom, je vysoko neštandardný a vo svojich dôsledkoch absurdný“. Nechápem, čím a ako si vybavujem účty – Soukup je prvý akademický zamestnanec sveta, ktorému sa zdá dokonale v poriadku, že zmena metodiky financovania v roku 2017 spätne platí pre indikátory z rokov 2014 a 2015, a upozornenie na to považuje za normatívne, neprijateľné a vybavovanie účtov.

Záverom by som ešte dodal, že v tejto polemike v žiadnom prípade nemienim pokračovať v akejkoľvek forme. Zároveň pána Soukupa vyzývam, aby sa zdržal akýchkoľvek lživých a dehonestáčných verejných prejavov o absencii mojej publikačnej činnosti.

#### LITERATÚRA

- BUIHAND, T. A., 1984: Tests for Detecting a Shift in the Mean of Hydrological Time Series. *Journal of Hydrology* 73, s. 51-69.
- BRU, N. – BIRITXINAGA-ETHCART, E. – d'AMICO, F., 2011. Detection of Significant Changes in Short Time Series. 19th International Congress on Modelling, <https://hal.inria.fr/hal-00868113/>
- LANGE, K. L. – LITTLE, R. J. A. – TAYLOR, J. M. G., 1989: Robust Statistical Modeling Using the T-Distribution. *Journal of the American Statistical Association* 84, s. 881-896.
- LEE, T.-S., 2010: Change-Point Problems: Bibliography and Review. *Journal of Statistical Theory and Practice* 4, s. 643-662.
- NADARAJAH, S. – KOTZ, S., 2008: Estimation Methods for the Multivariate t Distribution. *Acta Applicandae Mathematicae* 102, č. 1, s. 99-118.

- PETTITT, A. N., 1979: A Non-Parametric Approach to the Change Point Problem.  
Journal of the Royal Statistical Society Series C, Applied Statistics 28, s. 126-135.
- PINHEIRO, J. C. – LIU, C. H. – WU, Y. N., 2001: Efficient Algorithms for Robust Estimation in Linear Mixed-Effects Models Using a Multivariate T-Distribution.  
Journal of Computational and Graphical Statistics 10, s. 249-276.