

Užívání statistické a věcné významnosti v časopise Pedagogická orientace a Pedagogika v posledních deseti letech: pohled statistika

Petr Soukup

Univerzita Karlova v Praze, Fakulta sociálních věd

Redakci zasláno 9. 1. 2016 / upravená verze obdržena 31. 3. 2016 /
k uveřejnění přijato 16. 4. 2016

Abstrakt: Cílem příspěvku bude poukázat na problémy při využívání koncepcí statistické a věcné významnosti v posledním desetiletí ve dvou předních pedagogických časopisech, konkrétně v časopise *Pedagogika* a *Pedagogická orientace*. Text po stručném představení tématu obsahuje popis provedené obsahové analýzy dvou pedagogických časopisů a poté přechází k výsledkům provedeného výzkumu ve třech oblastech: nesprávná použití statistické analýzy, její nesprávná interpretace a používání měr věcné významnosti (*effect sizes*). V závěru autor kromě tradičního shrnutí výsledků a diskuse jejich limitů nabízí možná zlepšení pro publikační i vzdělávací praxi v české pedagogické komunitě. Provedený výzkum ukázal, že kvantitativně zaměřených textů je poměrně málo (ve dvou pedagogických časopisech bylo publikováno 84 textů za 10 let). Více než polovina těchto textů používá koncepci statistické významnosti, i když to prezentovaná data neumožňují. Poměrně málo častá jsou pochybení při interpretaci statistické významnosti, mezi nejrozšířenější patří mechanická práce s daty (nejčastěji užívání hvězdiček). V rozporu se současnými doporučeními se ve zkoumaných časopisech užívá poměrně málo měr věcné významnosti, nadto se užívají jen nejjednodušší míry věcné významnosti. Alarmující je též, že zhruba ve třetině analyzovaných článků chyběla věcná interpretace výsledků.

Klíčová slova: statistická významnost, věcná významnost, míry věcné významnosti, chybná interpretace

Statistická a věcná významnost jsou v oblasti kvantitativního zpracování téměř nepostradatelné. Bohužel se čím dál více setkáváme s mechanickou aplikací statistické významnosti (autoři pouze konstatují, že rozdíl je statisticky významný, resp. že souvislost je statisticky významná). Věcná interpretace, resp. zhodnocení věcné významnosti mnohdy v textech absentuje. Nadto se koncepce statistické významnosti používá často i v případech, pro něž není určena (resp. matematicky odvozena), tj. pro jiné než náhodné výběry nebo

randomizované experimenty (srov. Blahuš, 2000). V Česku dosud, s výjimkou jediného textu (Cuberek & Frömel, 2011), chybí empirická reflexe popsaných problémů a zhodnocení jejich výskytu v jednotlivých disciplínách. Obdobně chybí diskuse o možných doplněních publikačních standardů v předních českých časopisech. Právě tyto mezery (jistě ne všechny) se snaží zacelit předložený text. S ohledem na omezený rozsah se zaměřuje na následující:

- (1) zmapování chybného užívání statistické významnosti (tj. na případy, v nichž získaná data neodpovídají požadavkům na náhodný výběr resp. randomizovaný experiment);
- (2) zmapování chybné interpretace výsledků statistické významnosti;
- (3) zmapování využívání měr věcné významnosti;
- (4) zmapování frekvence věcné interpretace výsledků a interpretace měr věcné významnosti.

V jednotlivých částech jsou popsána zkoumaná kritéria v oblasti (1)–(4) s využitím odkazů na detailnější pojednání v odborné literatuře a dále jsou prezentovány výsledky jednotlivých mapování. Primárním záměrem článku je zmapování situace v oblasti pedagogiky. Pro tento účel byly vybrány dva přední časopisy, konkrétně *Pedagogika* a *Pedagogická orientace*. Pro možnost srovnání byl zvolen *Sociologický časopis*. Důvodem výběru byla zejména blízkost metodologických přístupů, často i výzkumných témat (minimálně sociologie vzdělání a sociální stratifikace mají velice blízko k tématům zkoumaným v oblasti pedagogiky). Cílem textu rozhodně není vyzvedávat či hanět jednotlivé časopisy či autory (proto také vše je prezentováno pomocí hromadných zjištění ve formě procent, a nikoli na úrovni jednotlivých článků, resp. jejich autorů). Uváděné problémy tedy nejsou citovány na konkrétních případech, ale obecně popsány, což ovšem není v tomto případě zcela v souladu s publikační etikou. Autor zároveň nabízí zájemcům celý svůj datový významový arch v anonymní podobě, tj. bez uvedení jmen autora článku, roku a čísla vydání pro následné analýzy. Samozřejmě je také možné opakovaným výzkumem prověřit zjištění prezentovaná v tomto článku.

1 Stručný popis výzkumu

Pro účely zmapování užívání statistické a věcné významnosti byly použity všechny články využívající primární kvantitativní analýzu za období

2005–2014 v časopise *Pedagogika* (celkem 46 textů) a časopise *Pedagogická orientace* (celkem 38 textů). Pro srovnání je využito 90 textů ze srovnatelného období ze *Sociologického časopisu* (jde opět o všechny kvantitativní studie otištěné ve stejném desetiletí v *Sociologickém časopise*). U všech textů byla provedena obsahová analýza (popis analytických kategorií je uveden dále), která sledovala užívání statistické a věcné významnosti dle výše zmíněných kritérií. V rámci jednotlivých článků byly sledovány jednotlivé analýzy prováděné odlišnými statistickými technikami (maximální počet zaznamenaných technik byl tři). Bylo by tedy možné analyzovat výsledky i na úrovni jednotlivých analýz, pro jednoduchost však v dále prezentovaných výsledcích je analytickou jednotkou *článek*. Na základě obsahové analýzy byl poté proveden záznam do datové matice a následné zpracování proběhlo ve formě popisné statistiky (jsou zkoumány všechny články, není tedy prostor pro využívání statistických testů, které slouží k zobecňování). Na okraj dodejme, že kromě základních indikátorů mapujících užívání statistické a věcné významnosti bylo zaznamenáno i číslo a rok vydání, autor, pracoviště prvního autora a rok, kdy byl výzkum v analyzovaném článku proveden (rok sběru dat). Tyto proměnné mohou sloužit zejména pro další detailní analýzy.

Pro možnost opakování výzkumu, případně jeho provedení na jiných textech, je namístě popsat podrobněji kategorie používané pro obsahovou analýzu prezentovanou v tomto článku.

(A) Kategorie sledované na úrovni článku

Na úrovni celého článku byly zaznamenány situace, kdy data nespĺňují požadavek náhodného výběru či randomizovaného experimentu. Teoreticky lze nalézt mnoho situací, v nichž běžné postupy, a nadto klasické vzorce implementované v používaném software, nelze užít, případně je nutné vzorce modifikovat. Soukup a Rabušic (2007) konkrétně popsali 6 takových situací:

- data pochází z censu (máme data za celou populaci);
- data pochází ze záměrného výběru (zejména kvótního);
- data pochází z malého výběru (cca do 30 jednotek);
- podíl vybraných jednotek je velký ve vztahu k velikosti populace;
- v datech je potřeba pro zobecnění používat váhy;
- data pochází z výběru s extrémně velkým počtem vybraných jednotek či s daty spojenými z různých datových souborů.

Pro účel záznamu problematické podoby použitých dat v analyzovaném článku bylo umožněno zapsat až tři nedostatky, nicméně v praxi vždy stačil pouze jediný.

(B) Kategorie sledované na úrovni jednotlivých statistických analýz

(Ba) Chybná interpretace výsledků statistické významnosti či její mechanické užívání

Problémů, které lze v této oblasti sledovat, je velice mnoho (srov. Soukup, 2010 a literaturu tam uvedenou). Při obsahové analýze článků byla pozornost věnována jen těmto projevům:

- mechanická práce se statistickými procedurami (prezentace hvězdiček u výsledků; „stepwise“ postupy pro výběr nejlepší sady prediktorů apod.);
- interpretace statistické významnosti ve smyslu „statisticky významné = důležité“;
- pro data nebylo užito statistických testů, i když data to umožňují;
- slovně se užívá místo *statisticky významné* jen výraz „významné“, nebo výraz „signifikantní“;
- v interpretaci je ignorováno výsledku statistického testu, resp. interpretace je v rozporu s jeho výsledkem.

Pro každou analýzu bylo možné zaznamenat více projevů chybné či mechanické interpretace.

(Bb) Věcná interpretace výsledků

S ohledem na diskuse o statistické významnosti a její nesprávné či mechanické aplikaci je často poukazováno na skutečnost, že autoři běžně přetisknou statistické tabulky ze softwaru, poté interpretují statistickou významnost a tím diskuse výsledků končí. Proto další (nejvíce diskutabilní) kategorií v rámci obsahové analýzy bylo sledování, zda každá jednotlivá statistická analýza kromě komentáře k výsledkům statistického testu obsahuje věcný komentář, tj. typicky zhodnocení velikosti koeficientů, velikosti průměrů či procent, zhodnocení míry vysvětleného rozptylu modelem apod. Výsledkem kódování byl toliko záznam, zda zkoumaná statistická analýza má věcné vyhodnocení či nikoliv. Samozřejmě by bylo možné ještě posuzovat úroveň této interpretace, ale to by bylo značně subjektivnější než posouzení, zda v textu

je (kvalitativně jakákoli) věcná interpretace, nebo zda chybí. Netřeba zdůrazňovat, že pro rozvoj vědeckého poznání je věcná interpretace (nadto kvalitní) velice důležitá.

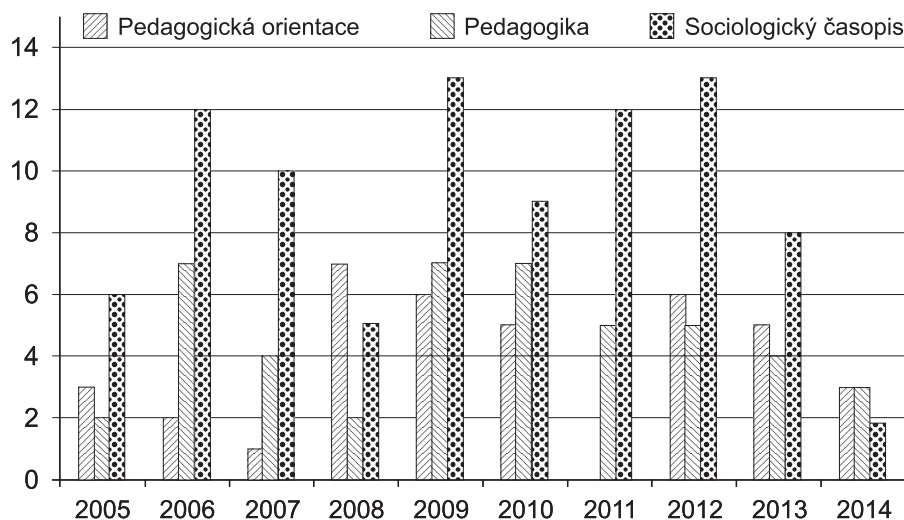
(Bc) Používání měr věcné významnosti a jejich interpretace

V rámci posledního bloku sledovaných kategorií se u každé statistické analýzy zaznamenal výskyt použité míry věcné významnosti (reálný přehled měr používaných v textech, srov. kap. 4 tohoto článku). S ohledem na skutečnost, že pro některé statistické analýzy lze využívat více měr věcné významnosti, byla kódována maximálně trojice použitých měr (v praxi však vystačovala většinou jediná). V návaznosti na vyhodnocení věcné interpretace výsledků (viz popis výše) bylo sledováno, zda jsou interpretovány použité míry věcné významnosti, tedy zda autor kromě přetištění hodnoty používá slovní popis hodnot míry věcné významnosti, např. „vysoká korelace“, „velká hodnota Cohenova d “ apod. Je zřejmé, že mezi kategorií, která sleduje interpretace měr věcné významnosti, a věcnou interpretací je vztah implikace, tj. pokud autor použil míru věcné významnosti a interpretoval ji, je to též vyhodnoceno jako provedení (byť mnohdy částečné) věcné interpretace analýzy.

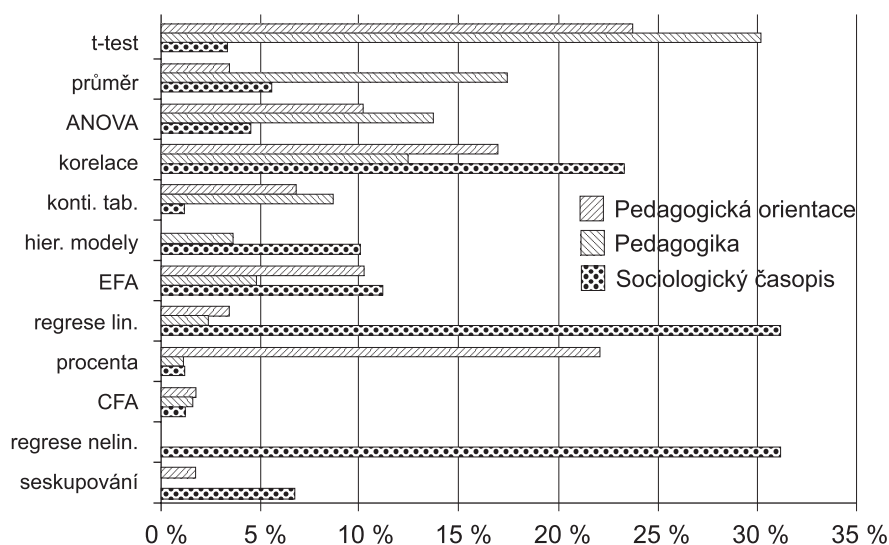
1.1 Základní popis výzkumného souboru

Dříve než představíme základní výsledky studie, je dobré se zaměřit na základní popis zkoumaného souboru článků. Tento popis následně pomůže chápat určité odlišnosti ve výsledcích pro pedagogické časopisy ve srovnání se *Sociologickým časopisem*). Z pohledu rozložení článků s kvantitativní orientací v časopisech *Pedagogika* a *Pedagogické orientace* (viz obrázek 1) lze konstatovat, že v posledních deseti letech nedochází v české pedagogické komunitě ani k expanzi, ani k útlumu kvantitativní produkce. Při detailnějším pohledu (není zobrazeno v obrázku) je navíc možné konstatovat, že někdy dochází v jednom čísle (často monotematickém) ke kumulaci několika kvantitativně laděných textů a poté zcela chybí (srov. též v obrázku 1 rok 2011 v *Pedagogické orientaci*). Z hlediska autorů převažují drtivě autoři z českých pedagogických fakult, cca 15 % produkce je od slovenských kolegů (opět z pedagogických fakult). Z hlediska jazyka byl za posledních 10 let jediný text s kvantitativní analýzou v angličtině. Pro srovnání uveďme (není zobrazeno v obrázku), že ve sledovaném období byla anglicky v *Sociologickém časopise* více než třetina kvantitativních analýz, z hlediska autorství je zhruba polovina produkce *Sociologického časopisu* z dílny výzkumníků působících na Sociologickém ústavu AV ČR.

Pro snazší čtení dále uvedených výsledků o užívání měr věcné významnosti a věcné interpretaci výsledků je dobré mít přehled o portfoliu užívaných statistických technik. Tento přehled nabízí obrázek 2. Z tohoto obrázku je patrné, že v pedagogických časopisech je nejčastěji užíván *t-test* (případně jeho neparametrické obdoby), *průměry* a *procenta*. Pokročilejší vícerozměrné statistické techniky (například *strukturní modely*, *hierarchické lineární modely*) jsou používány v těchto časopisech velmi zřídka (nadto většinou autory, kteří přišli do pedagogického prostředí z jiných oborů). Ze srovnání s údaji za *Sociologický časopis* plyne jednoznačně, že v tomto časopise je naopak užívání vícerozměrných statistických technik častější, nadto i jejich škála je pestřejší (srov. zejména podíly pro *lineární* i *nelineární regresní analýzu* v obrázku 2).



Obrázek 1. Počty článků s kvantitativní analýzou dat v časopisech *Pedagogická orientace*, *Pedagogika* a v *Sociologickém časopise* v posledních 10 letech (2005–2014).



Obrázek 2. Relativní četnost statistických technik používaných v člancích s kvantitativní analýzou dat v časopisech *Pedagogická orientace*, *Pedagogika* a v *Sociologickém časopise* v posledních 10 letech (2005–2014).

2 Nesprávná užívání statistické významnosti (nevhodná data)

První oblastí, na kterou se zaměřila obsahová analýza článků, bylo využívání statistické významnosti v případech, pro které není určena. Připomeňme, že klasické postupy spojené s koncepcí statistické významnosti (testování statistických hypotéz a výpočet intervalů spolehlivosti) se užívají ve dvou základních situacích:

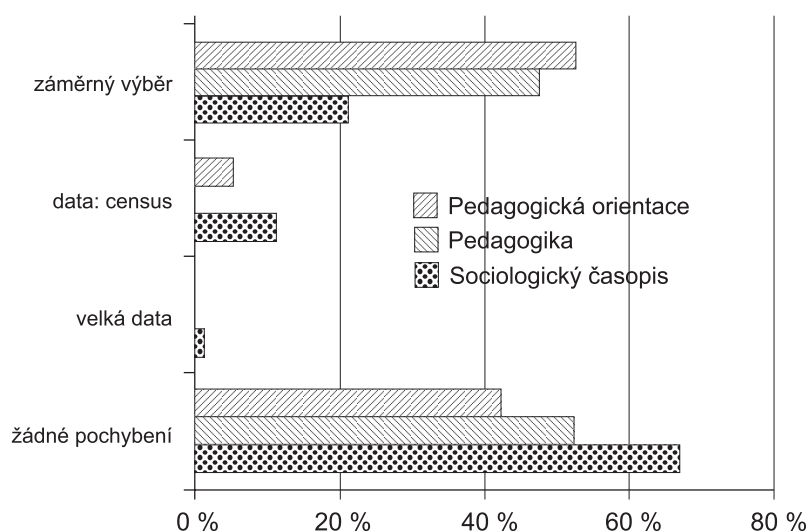
(a) data jsou náhodným (pravděpodobnostním) výběrem z populace, (b) data vznikla skrze randomizovaný experiment.

Příklady obou těchto situací lze na stránkách časopisů *Pedagogika*, resp. *Pedagogická orientace* nalézt. Statistická významnost slouží k zobecnění výsledků na populaci – případ (a), resp. k zobecnění na jiné než experimentem testované úrovně faktorů – případ (b). V rámci šesti kategorií popsanych v metodologické části byly v analyzovaných textech objeveny jen tři problémové kategorie:

- data pochází z úplného zjišťování (cenzu);
- data pochází ze záměrného výběru (nejčastěji lze očekávat kvótní);
- data jsou velká, tj. obsahují řádově mnoho tisíc jednotek a více.

V oblasti pedagogiky se nadto situace komplikuje, protože ne všichni autoři článků užívají standardní statistické názvosloví. V české a slovenské pedagogické komunitě je jedním z nejužívanějších pojmů pro *záměrný výběr* tzv. *dostupný výběr*, tento pojem byl logicky podřazen pod výše uvedenou kategorii (b) pro záměrné výběry. Podíly nesprávného užívání statistické významnosti ve sledovaných třech časopisech uvádí obrázek 3.

Z obrázku 3 plyne, že ve dvou sledovaných časopisech z oblasti pedagogiky se pohybuje nesprávné užívání statistické významnosti okolo 55 % z publikovaných článků, v *Sociologickém časopise* zhruba na úrovni jedné třetiny z publikovaných textů. Z obrázku tedy jasně vyplývá, že je v ČR publikováno poměrně hodně textů, ve kterých autoři používají koncepci statistické významnosti, i když jejich data toto neumožňují. Samozřejmě, že důsledky této praxe jsou v různých člancích, resp. analýzách různě závažné. Zřejmě nejproblematictější jsou srovnávací analýzy (národní, a ještě více mezinárodní), v nichž jsou skrze statistické testy srovnávány dva a více záměrné (dostupné) výběry a je vyvozen meritorní závěr. To, že je závěr spíše ovlivněn výběrem (a jeho systematickým zkreslením) než reálně působícími faktory, už ale bohužel většina autorů v diskusi nerozvádí a nepoučený čtenář může snadno podlehnout falešnému závěru.



Obrázek 3. Relativní četnost používání nevhodných dat pro statistické testování v člancích s kvantitativní analýzou dat v časopisech *Pedagogická orientace*, *Pedagogika* a v *Sociologickém časopise* v posledních 10 letech (2005–2014).

3 Chybná interpretace výsledků statistické významnosti

Po zhodnocení vhodnosti užívání koncepce statistické významnosti se zaměříme na její správné využití, tj. využití v souladu s poznatky statistické teorie. Cílem nebude posuzovat nuance v používání testů (tj. například kdy použít *dvouvýběrový t-test* a kdy jeho neparametrický protějšek, tj. *M-W test*), ale cílem bude obecné zhodnocení interpretace statistické významnosti. Problémů, které lze v této oblasti řešit, je opět velice mnoho (srov. Soukup, 2010 a literaturu tam uvedenou), při obsahové analýze článků byla pozornost věnována jen kategoriím popsaným výše v metodologické části.

Pro lepší pochopení výsledků stručně vysvětlíme, co je problematické na jednotlivých projevech popsaných v metodologické části. Z popisu vynecháme nepoužití statistických testů pro data, která toto umožňují, a ignorování výsledků testů, resp. interpretace v rozporu s jejich výsledky.

3.1 *Mechanická aplikace při interpretaci statistických testů, resp. využívání mechanických postupů*

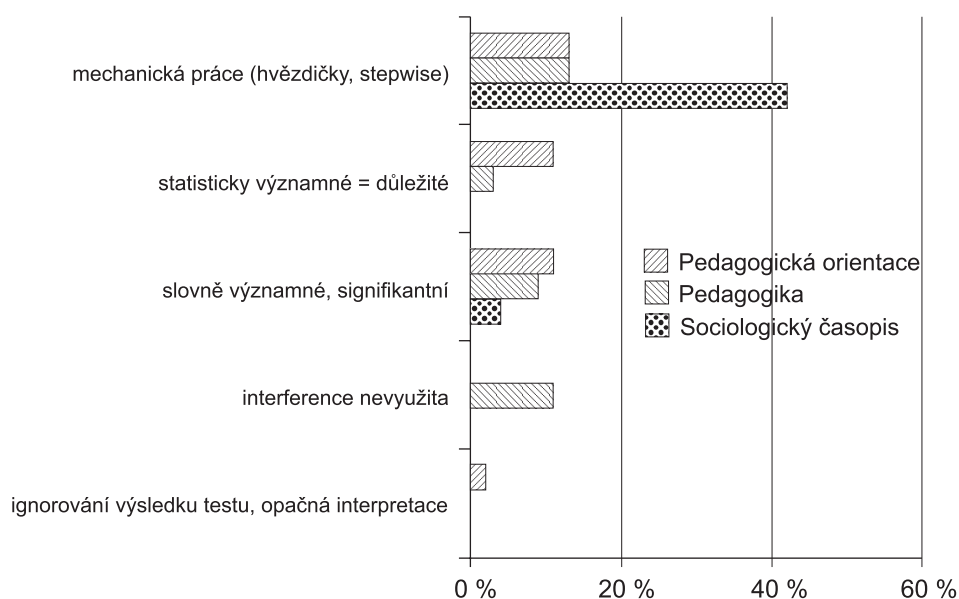
Statisticy a metodologové se obecně brání automatizovanému a mechanickému použití statistických nástrojů. Důvodem je zejména skutečnost, že se pak statistika stává nástrojem, kdy se výpočet (mnohdy i na špatně připravených datech) získá stiskem jednoho tlačítka a autorem výpočtu není autor textu, ale automatizovaná procedura, která samozřejmě problému věcně nerozumí a sleduje jen naprogramovaný algoritmus. Jako nejčastější prohřešky se udává používání hvězdiček pro označení statisticky významných výsledků. Problémy spojené s hvězdičkami v tabulkách s výsledky jsou mnohé. První problém je, že autor zveřejní jen hvězdičky a kvůli tomu neuvede přesné hodnoty pravděpodobností chyb prvního druhu (nesprávného zamítní nulové hypotézy). Hvězdička tak může nahradit hodnotu 0,049, stejně jako hodnotu 0,000049. Hvězdičky dále vedou často k tomu, že autor analýzy je spokojen s tím, že má krásnou tabulku plnou hvězdiček (tj. zhodnocení statistické významnosti) a nemá potřebu se již věnovat věcné interpretaci výsledků (tj. zhodnocení jejich věcné významnosti). Hvězdičky tak fakticky odstraňují potřebu přemýšlet nad výsledky, i když je to zcela nezamýšlený důsledek používání této notace (původním záměrem zcela jistě bylo šetřit místo ve výstupech ze statistických softwarů a zpřehlednění jejich výsledků). Posledním argumentem proti užívání hvězdiček, posledním nikoliv svou důležitostí, je

nejednotná praxe v užívání hvězdiček. Někteří autoři (resp. některé softwary) užívají jednu hvězdičku pro výsledky statisticky významné na hladině 0,05, někdy se nadto užívají dvě hvězdičky pro hladinu 0,01 a tři pro 0,001. Ale můžeme se setkat i s autory, kteří užívají jednu hvězdičku pro hladinu 0,1, někdy se užívají pro tuto hladinu křížky a praktik je ještě mnohem více. Obdobně statistici nemají rádi mechanické přístupy, které slouží k nalezení „nejlepšího“ modelu – typicky je to například algoritmus „stepwise“ užívaný v lineární regresi. Důvody proti užívání těchto přístupů jsou minimálně dva. Automatický výběr nejvhodnějších proměnných, které mají být v modelu využity, je veden pouze kritériem statistické významnosti. Může se tak stát, že náš model sice obsahuje všechny statisticky významné proměnné, ale věcně nedává smysl, resp. věcně zajímavé proměnné byly automatickým algoritmem vyřazeny. Druhým argumentem proti užívání procedur pro automatické budování modelů je již popsaná tendence k nepřemýšlení nad výsledky a jejich kritickému zhodnocení („spočítala to přeci kvalitní procedura, tak proč bych měl něco zpochybňovat“).

3.2 Záměna statistické a věcné významnosti

Zcela nešťastné bývá, že někteří autoři zaměňují statistickou a věcnou významnost a výsledky statisticky významné označí jako výsledky věcně významné. Nicméně tato tendence má dvě úrovně. První, závažnější, je 100% záměna, při níž autor hovoří o výsledcích statistické významnosti a ty interpretuje ve smyslu praktické důležitosti výsledků (přitom prakticky mohou být získané výsledky zcela nezajímavé). Slabší verzí prohřešku je, pokud autor používá místo výrazu *statisticky významné* výrazu „důležité“ (někdy „významné“), ale interpretuje statistickou významnost. Problémem tohoto postupu je, že nezalého čtenáře snadno zmate, a ten pak výsledek toliko statisticky významný považuje za výsledek významný věcně. Tento druhý prohřešek samozřejmě plyne z jazykové blízkosti slova významný a důležitý a ze snahy o úsporu místa v textu (tj. místo sousloví „statisticky významný“ se užije jen druhé slovo). Zde lze doporučit buď užívat plný dvouslovný výraz, nebo (a možná lépe) užívat slova „zobecnitelný“, čímž se vyhneme dvojznačnosti a každému je jasné, že jsme objevili výsledek platný pro základní soubor, ale o jeho věcné významnosti nesdělujeme nic.

Po popisu problematiky se věnujme výskytu jednotlivých nevhodných postupů pro interpretace statistické významnosti ve třech zkoumaných časopisech (viz obrázek 4).



Obrázek 4. Relativní četnost výskytu chyb v interpretaci statistické významnosti v člancích s kvantitativní analýzou dat v časopisech *Pedagogická orientace*, *Pedagogika* a v *Sociologickém časopise* v posledních 10 letech (2005–2014).

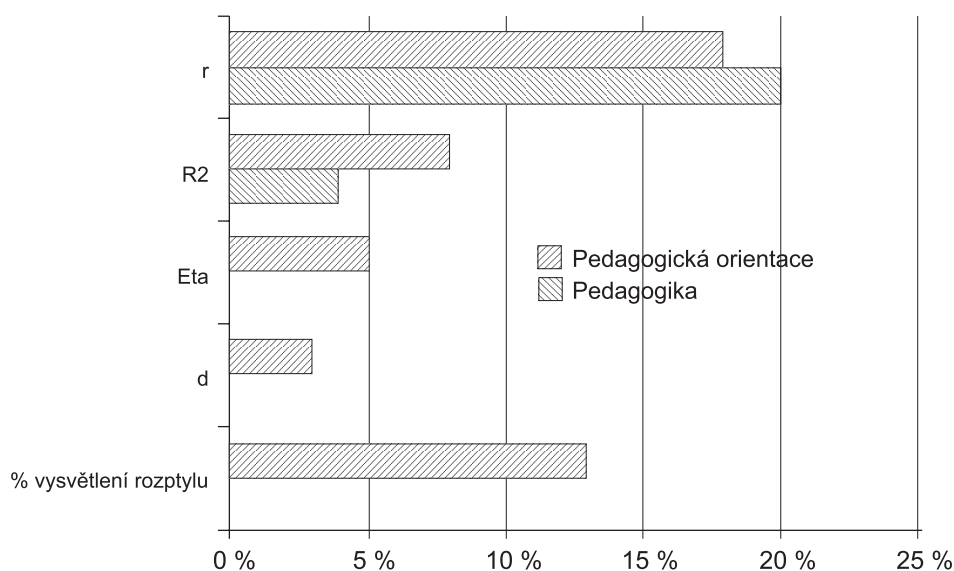
Z obrázku 4 plyne, že nejčastější je mechanická aplikace statistické významnosti (cca 13 % zkoumaných článků v pedagogických časopisech a více než 40 % v *Sociologickém časopise*). Rozdíl lze vysvětlit odlišným spektrem technik – hvězdičky se typicky užívají v korelační a regresní analýze, „stepwise“ postup v regresní analýze. A právě tyto analýzy jsou mnohem častější na stránkách *Sociologického časopisu* ve srovnání s pedagogickými časopisy (srov. obrázek 2). Méně časté je používání výrazů „významné“, „důležité“ či „signifikantní“ namísto plného výrazu *statisticky významné*. Nejméně častá je pak záměna statistické a věcné významnosti, i když v *Pedagogické orientaci* se tento prohřešek objevil stejně často (v 11 % článků) jako mírnější prohřešek ve formě slovního pochybení. Obecně lze konstatovat, že chybná interpretace statistické významnosti je ve dvou zkoumaných pedagogických časopisech málo častá, což může být mj. způsobeno poměrně omezeným spektrem užívaných statistických technik, tj. využíváním zejména těch jednodušších (srov. obrázek č. 1 a komentář k němu uvedený).

4 Využívání měr věcné významnosti

V posledních zhruba dvaceti letech zejména ve světové psychologii a pedagogice sílí snaha po zavedení nového standardu interpretace výsledků kvantitativních analýz skrze využívání měr věcné významnosti (*effect sizes*) a jejich interpretaci (viz další část tohoto článku). Tyto snahy vyústily v doporučení pro publikace, zejména APA manuál (2010) pro psychologii a AERA (2014) pro pedagogiku, nicméně není dosud standardem ani ve všech zahraničních časopisech, aby byl explicitně formulován požadavek na počítání měr věcné významnosti a jejich interpretaci. V ČR s výjimkou časopisu *Česká kinantropologie* není autorovi tohoto textu známo, že by některý časopis popisovanou praxi požadoval.

Seznam měr věcné významnosti v současnosti již čítá několik desítek, základní seznámení s nimi podává česky psaný článek Soukupa (2013) a literatura v něm uvedená.

S ohledem na malé rozšíření znalostí o mírách věcné významnosti lze čekat, že užití těchto instrumentů bude v pedagogickém výzkumu poměrně řídké, nadto že složitější míry věcné významnosti (většinou komunity neznámé) budou užívány maximálně okrajově. Toto očekávání ostatně potvrzuje výsledek obsahové analýzy článků, jenž je zobrazený v obrázku 5.

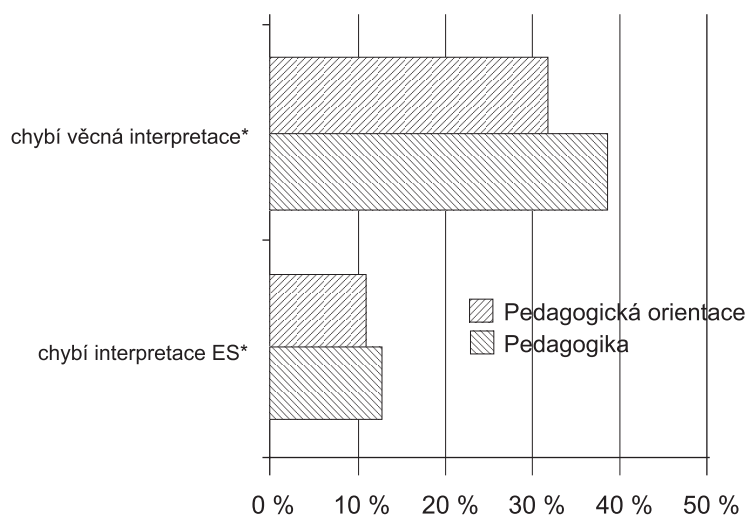


Obrázek 5. Relativní četnost používání jednotlivých měr věcné významnosti v článcích s kvantitativní analýzou dat v časopisech *Pedagogická orientace* a *Pedagogika* v posledních 10 letech (2005–2014).

Nejčastěji užívanou mírou v časopise *Pedagogická orientace* i *Pedagogika* v posledních 10 letech je *korelační koeficient* (cca pětina článků). Trochu překvapivě na druhém místě v časopise *Pedagogická orientace* je *procento vysvětleného rozptylu v explorační faktorové analýze*, nicméně tuto skutečnost lze snadno vysvětlit vyšším výskytem této analytické techniky v tomto časopise (srov. obrázek 2). S ohledem na skutečnost, že *regresní analýza* byla ve sledovaných pedagogických časopisech užívána velmi málo, není překvapením řídké užití klasického *indexu determinace* (R^2). Naopak by bylo možno očekávat, že pokud autoři užívají často testu pro dvě nezávislé skupiny, užijí pro jeho vyhodnocení *Cohenovo d* (Hendl, 2012; Soukup, 2013). Toto použití ale bylo naopak velice řídké, vyskytlo se v jediném článku v *Pedagogické orientaci*. Obdobně řídké je užívání míry *Eta* (resp. Eta^2), i když analýza rozptylu je použita poměrně často (opět srov. obrázek 2) a tento ukazatel bývá běžně obsažen v učebnicích i výkladu této statistické techniky. Souhrnně můžeme konstatovat, že užívání měr věcné významnosti je v pedagogických časopisech poměrně řídké, nadto se užívají jen nejjednodušší míry věcné významnosti.

5 Věcná interpretace výsledků a interpretace měr věcné významnosti

V návaznosti na doporučení užívat míry věcné významnosti při vyhodnocování kvantitativních dat je doporučováno, aby byly hodnoty těchto měr interpretovány a obecně aby byla vždy přítomna věcná interpretace získaných výsledků. Toto doporučení zní banálně, ale zahraniční analýzy ukazují, že mnoho textů končí výpočtem statistických testů a jejich vyhodnocením a věcná stránka výsledků se již ponechává bez povšimnutí. Je proto namístě otázka, jaká je praxe v českém prostředí. Míru nepoužívání věcné interpretace, resp. interpretace měr věcné významnosti shrnuje obrázek 6.



Obrázek 6. Relativní četnost věcné interpretace výsledků a měr věcné významnosti v článcích s kvantitativní analýzou dat v časopisech *Pedagogická orientace* a *Pedagogika* v posledních 10 letech (2005–2014).

Pozn.: *Mimimálně 1x v článku chybí.

Z obrázku vyplývá poměrně nelichotivý výsledek pro pedagogické časopisy, protože cca ve třetině článků chyběla v posledních deseti letech věcná interpretace výsledků alespoň u jedné provedené analýzy (hůře dopadl časopis *Pedagogika*). Oproti tomu v *Sociologickém časopise* (není zobrazeno v obrázku) ve sledovaném období se tento problém vyskytl jen u desetiny článků. Rozdíl lze opět částečně vysvětlit odlišným spektrem používaných statistických technik. V pedagogických časopisech jsou užívány spíše jednodušší techniky a u nich může autor snáze podlehnout dojmu, že interpretace není potřebná. V případě měr věcné významnosti lze konstatovat, že pokud je již autoři užijí (viz předchozí část článku a obrázek 5), většinou je interpretují (nečiní tak cca v desetině článků). Samozřejmě, že není dobrou vizitkou časopisů a ani jejich recenzentů, že nevyžadují interpretaci výsledků a postačí jim jen „žonglování“ s daty a prezentace statistických testů.

6 Shrnutí a diskuse výsledků

Připomeňme, že tento text si kladl za cíl následující:

- (1) zmapování chybného užívání statistické významnosti (tj. případy, kdy získaná data neodpovídají požadavkům na náhodný výběr, resp. randomizovaný experiment);
- (2) zmapování chybné interpretace výsledků statistické významnosti;
- (3) zmapování využívání měr věcné významnosti;
- (4) zmapování frekvence věcné interpretace výsledků a interpretace měr věcné významnosti.

Vše bylo omezeno na posledních deset (2005–2014) produkce dvou pedagogických časopisů (*Pedagogická orientace* a *Pedagogika*) a jednoho časopisu ze sociologie (*Sociologický časopis*). Kvůli tomu je samozřejmě těžké zobecňovat a popisovat, jak funguje česká pedagogická výzkumná komunita, česká sociologická komunita atd. Nicméně dva časopisy byly vybrány pro jejich jasný reprezentativní charakter a dlouhodobé působení (*Pedagogika*, resp. *Sociologický časopis*), druhý pedagogický časopis (*Pedagogická orientace*) pak byl vybrán pro možnost doplnění výsledků z oblasti pedagogického výzkumu, dále proto, že existoval po celé zkoumané období a vycházelo v něm poměrně mnoho kvantitativních analýz. V tomto ohledu neměl autor textu možnost výběru jiného časopisu. Samozřejmě, že by bylo žádoucí provést analýzu i pro další časopisy (např. *Orbis Scholae*, *Studia paedagogica*, *e-Pedagogium*), aby bylo zmapování komplexnější. Obdobně bude do budoucna žádoucí provést analýzu prestižních zahraničních časopisů zaměřených na *educational science*, aby bylo zřejmé, jak si stojí český pedagogický výzkum v mezinárodním srovnání.

Současný obrázek, byť značně kusý, není zcela lichotivý. Výzkumníci publikují kvantitativně zaměřených textů poměrně málo (ve dvou pedagogických časopisech bylo publikováno 84 textů, což je méně, než bylo publikováno za stejné období v *Sociologickém časopise*), nadto téměř veškerá produkce je jen v českém jazyce a kvůli tomu nedochází k výměně výsledků a zkušeností (jediný text byl za deset let anglicky, v *Sociologickém časopise* byla anglicky třetina kvantitativní produkce). Více než polovina kvantitativně zaměřených textů, které byly publikovány v posledních deseti letech v časopisech *Pedagogická orientace* a *Pedagogika*, používá koncepci statistické významnosti, i když to prezentovaná data neumožňují. Naopak poměrně málo častá

jsou pochybení při interpretaci statistické významnosti, mezi nejrozšířenější patří mechanická práce s daty (skrze užívání hvězdiček, „stepwise“ postupy atd.). V rozporu se současnými doporučeními se ve zkoumaných časopisech užívá poměrně málo měř věcné významnosti, nadto se užívají jen nejjednodušší míry věcné významnosti. Alarmující je též, že zhruba ve třetině analyzovaných článků chyběla věcná interpretace výsledků.

7 Závěrečná doporučení

Bývá zvykem končit článek shrnutím výsledků a případnou diskusí jeho slabín a naznačením možného dalšího výzkumného snažení. Tento zvyk naplnila předchozí část (byť značně stručná s ohledem na možný maximální rozsah textu). Nicméně tento text má ještě jednu ambici, a to navrhnout řešení ne zcela lichotivého stavu kvantitativně orientovaného pedagogického výzkumu. Dále prezentovaná doporučení vychází ze zkušeností autora a neodrážejí dříve prezentované výzkumné výsledky. Na tyto výsledky však věcně navazují, protože se snaží řešit problémy, které provedený výzkum identifikoval u pedagogické produkce v ČR. Konkrétně jde o doporučení pro zlepšení publikační činnosti, ale též pro zlepšení v oblasti výuky kvantitativní metodologie.

7.1 *Doporučení pro zlepšení publikovaných výsledků kvantitativní analýzy*

Z výše uvedených popisů i výsledků obsahové analýzy lze vyvodit poměrně jednoznačná doporučení pro publikační činnost. Konkrétně je třeba provést tyto kroky:

- (1) zavedení publikačních standardů, které zohlední poslední novinky mezinárodních asociací, tj. AERA (2014) a APA (2010);
- (2) vyvinutí tlaku na recenzenty, aby posuzovali dodržování publikačních standardů a v případě jejich nedodržování vraceli texty k přepracování; rozšíření okruhu recenzentů kvantitativně laděných textů;
- (3) zvýšení tlaku na redakční rady, aby se bránily snaze publikovat co nejvíce čísel jimi vedeného časopisu na úkor jeho kvality;
- (4) zvážit vybudování datového archivu pro shromažďování dat z pedagogických výzkumů.

Ad (1) Publikační standardy časopisů by měly být výrazně rozšířeny o doporučení inspirovaná například poslední verzí manuálu APA (2010). Konkrétní

rozšíření s ohledem na provedenou analýzu by mělo obsahovat minimálně tyto části:

- Zákaz používání statistických testů a dalších nástrojů ke zobecňování (např. intervaly spolehlivosti) pro data, která tyto postupy neumožňují (nenáhodné výběry, úplná zjišťování).
- Důraz na tvůrčí práci se statistickými technikami a vyhnutí se jejich mechanickému užití.
- Nepoužívání výrazů „důležité“, „významné“ či „signifikantní“ pro výsledky, které jsou statisticky významné. Obecně lze doporučit užívat zcela neutrální výraz „zobecnitelný“, díky kterému lze předejít mnoha nedorozuměním.
- Doporučit používání měr věcné významnosti a jejich interpretace včetně doporučených vodítek.
- Požadavek na věcnou interpretaci výsledků.

Ad (2) V oblasti recenzního řízení je jistě také mnoho možností ke zlepšení. První možností, jak vylepšit recenzní řízení v oblasti kvantitativní analýzy, je rozšíření potenciálního okruhu recenzentů. V České republice jsou minimálně stovky statistiků, kteří jsou schopni posoudit korektnost práce s kvantitativními daty a případně mohou autorům v recenzním posudku naznačit, kde mají chyby a jak je napravit.¹ Pomohlo by též rozšíření produkce v anglickém jazyce, protože by se zvýšil okruh možných recenzentů o kvalitní odborníky ze zahraničí, kde je často úroveň kvantitativních analýz vyšší. A samozřejmě ze strany redakcí bude nutné dbát na to, aby recenzenti posuzovali dodržení publikačních standardů a aby jejich případné nedodržení mělo za následek vrácení textu k přepracování.²

Ad (3) Současný tlak na publikace skrze RIV body vede zcela jednoznačně k upřednostňování kvantity před kvalitou. Proto se časopisy snaží mít co nejvíce čísel a doslova chrlí články na vědecký trh. Toto je zcela bez pochyb naprosto zničující strategie. Samozřejmě by bylo možné produkci například

¹ Autor osobně zastává názor, že by bylo zcela vhodné, aby pedagogičtí výzkumníci již při přípravě svých textů se statistiky spolupracovali a vyhnuli se tak případné blamáži v recenzním řízení.

² Zde autor vstupuje na velice tenký led, ale doporučuje, aby i pedagogické časopisy po vzoru např. *Sociologického časopisu* za práci recenzenta platily odměnu, protože pak lze daleko více od recenzentů žádat (nekvalitní či pozdě dodaný posudek je možné nezaplatit).

zdvojnásobit, ale znamenalo by to, že by jen narostl počet publikací, které nemají žádný užitek. Počet kvalitních publikací by se tímto krokem ještě snížil. Zde je namístě apelovat na všechny, kdo vyjednávají na úrovni univerzit s politickými reprezentanty České republiky, aby jim vysvětlili, že současný systém hodnocení vědy je zcela zvrácený a jeho (snad původně nezamýšlené) zvrácené důsledky se již projevují. Bohužel ze současné novely vysokoškolského zákona již vypadl návrh pětiletého financování vysokých škol, které by honbu za publikačními výsledky (bez ohledu na jejich kvalitu) snad ještě mohlo zastavit. Naopak již pozorujeme nárůst počtu predátorských časopisů (srov. Knecht, 2014) a lze jen čekat, kdy tato predátorská hrozba zachvátí i českou pedagogickou obec. Lépe ani nespekulovat, zda se na dráhu predátorského časopisu nevydá i některý z českých časopisů.

Ad (4) Po vzoru sociologicky a politicky orientovaných výzkumů, které jsou již téměř dvacet let v Česku ukládány do Českého datového archivu Sociologického ústavu,³ a každý výzkumník má tudíž volný přístup k datům jiných výzkumníků, lze doporučit pro oblast pedagogického výzkumu založení obdobné instituce, případně spolupráci se stávajícím Českým datovým archivem. Pokud chce výzkumník uložit data do archivu, musí splnit poměrně přísné požadavky, tj. jeho data musí mít garantovanou kvalitu. Platí ale také, že pokud výzkumník získal data za veřejné prostředky (granty, dotace pro vysoké školy či výzkumné ústavy), pak data nejsou jeho majetkem, ale veřejným majetkem, ke kterému by měli mít (volný) přístup všichni. Do budoucna by tedy bylo žádoucí zajistit archivaci dat z nově vzniklých studií a otevřít možnost jejich používání pro ostatní. Díky tomu by bylo možné též přezkoumat analýzy konkrétního autora a podrobit je případně kritické analýze, což by jistě zvýšilo kvalitu prováděných analýz.

7.2 Doporučení pro výuku kvantitativní analýzy

Výše uvedená doporučení mají samozřejmě omezený charakter. Působí na etablované výzkumníky, kteří mají málo času na své další vzdělávání a navíc mají své zažité postupy, které je těžké změnit.⁴ Lze proto s nadějí hledět do budoucna a doufat ve změny, které přinesou další generace výzkumnic a výzkumníků. Aby tomu tak bylo, musí tito mladí dostat slušné základy. Konkrétně to znamená následující:

³ <http://archiv.soc.cas.cz/>

⁴ V češtině pro to máme i pěkné přísloví, ale pro jeho genderovou nekorektnost jej zde autor necituje.

- (1) Výuku musí vést ti, kdo sledují nové trendy v používání technik pro kvantitativní analýzu a nejsou jim cizí doporučení uvedená v předchozí části věnované úpravám publikačních standardů časopisů.
- (2) Výuka musí být nadto vedena dle učebních textů, které zohledňují novinky, a to je v ČR poměrně problém. Ze všech učebnic bude vyhovující snad jen Hendlova (2012).
- (3) Ve výuce musí být kladen důraz místo teoretických pouček na praktické postupy a musí se ideálně kombinovat statistický rozměr problému s věcným rozměrem. Pokud se toto nedodrží, odchází studenti s pocitem, že analýza je hotová tím, že se vypočte test a vyhodnotí se, opomíjí pak ale věcnou stránku problému. Ideální je, když si studenti na svém vlastním výzkumu vyzkouší jak formulaci věcných otázek, tak jejich vyhodnocení skrze svá data.
- (4) Ve výuce je také žádoucí používat reálné datové soubory ze vzdělávacích studií. V tomto jsou možnosti téměř bezbřehé, vždyť studie typu PISA či TIMSS jsou volně k dispozici pro každého zájemce (srov. Soukup, 2012). Pokud se v budoucnu podaří archivovat data z českých studií, samozřejmě bude potřebné užívat data z ČR, která budou studentům bližší.
- (5) V oblasti výuky by bylo též vhodné spojit síly jednotlivých pedagogických pracovišť (zřejmě skrze ČAPV) a zavést systém pravidelnějšího vzdělávání v oblasti složitějších postupů kvantitativní analýzy. Pro počátek lze využít platformu doktorských seminářů na pravidelných konferencích ČAPV, ale do budoucna je namístě uvažovat o vzniku letních (zimních) škol, které by postupně doplňovaly vzdělání v oblastech kvantitativní analýzy, které běžně na pedagogických fakultách není možné zajistit. Odborníků (nejen v ČR) je jistě dostatek, jde jen o to, jak tyto odborníky využít pro vzdělání (budoucích) pedagogických výzkumníků.

Literatura

- APA. (2010). *Publication manual of the American Psychological Association*. Washington: American Psychological Association.
- AERA. (2014). *Publications manual*. Dostupné z <http://www.aera.net/Portals/38/docs/Publications/AERA%20Publications%20Manual.pdf?timestamp=1415307712543>
- Blahuš, P. (2000). Statistická významnost proti vědecké průkaznosti výsledků výzkumu. *Česká kinantropologie*, 4(2), 53–71.
- Cuberek, R., & Frömel, K. (2011). K problematice výzkumného výběru a testování nulové hypotézy. *Československá psychologie*, 55(5), 468–477.
- Hendl, J. (2012). *Přehled statistických metod: analýza a metaanalýza dat*. Praha: Portál.

- Knecht, P. (2014). O strašáku excelentního publikování a predátorských časopisech. *Universitas – revue Masarykovy univerzity*, 47(3), 3–8.
- Soukup, P. (2010). Nesprávná užívání statistické významnosti a jejich možná řešení. *Data a výzkum – SDA Info*, 4(2), 77–104.
- Soukup, P. (2012). Mezinárodní výzkumy v oblasti vzdělávání. In J. Krejčí & J. Leontiyeva (Eds.), *Cesty k datům: zdroje a management sociálněvědních dat v České republice* (s. 287–304). Praha: Sociologické nakladatelství, Sociologický ústav AV ČR.
- Soukup, P. (2013). Věcná významnost výsledků a její možnosti měření. *Data a výzkum – SDA Info*, 7(2), 125–148.
- Soukup, P., & Rabušic, L. (2007). Několik poznámek k jedné obsesi českých sociálních věd – statistické významnosti. *Sociologický časopis/Czech Sociological Review*, 43(2), 379–395.

Autor

PhDr. Ing. Petr Soukup, Univerzita Karlova v Praze, Fakulta sociálních věd, U Kříže 8, 158 00 Praha 5, e-mail: soukup@fsv.cuni.cz

Usage of statistical and substantive significance in Pedagogická orientace and Pedagogika in the last 10 years: Statistical insight

Abstract: The aim of this paper is to highlight the problems in the use of statistical and substantive significance in the last decade in two leading educational journals, particularly in the journal *Pedagogika* and *Pedagogická orientace*. After a brief introduction the article offers a description of the content analysis performed on two pedagogical journals. The results of the research are divided into three areas: the incorrect use of statistical analysis, its incorrect interpretation and the application of measures of substantive significance (effect sizes). In conclusion the author in addition to traditional summary of results and discussion of their limitations offers possible improvements for publishing and educational practice in the Czech educational community. The research showed that quantitative research is not very widespread (two pedagogical journals published only 84 quantitatively oriented articles). More than half of these texts uses the concept of statistical significance for data that do not allow this practice. Errors in the interpretation of statistical significance are quite rare, the most common include mechanical work with the data (most often use asterisks). Contrary to current guidelines Czech educational scientists use quite few measures of substantive significance (effect sizes) and if so they use only the simplest ones. It is also alarming that about a third of the articles analyzed lacked substantive interpretation of the results.

Keywords: statistical significance, substantive significance, effect sizes, misinterpretation