

## Odpoověď na diskusní příspěvek O. Špačka „Jak neargumentovat při transformaci proměnné“

Děkuji Ondřeji Špačkovi za jeho zájem o pojednávané téma i jeho připomínky, které by v některých případech mohly být inspirativní pro další práci. Proto je také třeba blíže rozebrat některé Špačkovy argumenty, aby nedocházelo k nedorozuměním.

Špaček uvádí: „*Paleček zpochybňuje užití metod založených na práci s kategoriickými daty (např. logistická regrese) na základě obtížnosti jejich aplikace či limitů spojených s nedostatečným zastoupením méně častých kategorií.*“ K žádnému „zpochybnění“ statistických metod pracujících s kategoriálními proměnnými v mém textu nedošlo. Jak si čtenář může ověřit, v případě kategoriálních proměnných návštevnost bohoslužeb jsem pojednával riziko nedostatečného zastoupení v některých buňkách vícerozměrných tabulek, ztráty části informace při redukcí proměnné do nižšího počtu kategorií a interpretační rozsáhlost, danou velkým počtem výsledných hodnot. Každá statistická procedura je vytvořena pro určité typy proměnných (např. logistická a lineární regrese), proto nelze „zpochybnovat“ užitost některých z nich jen proto, že pracují s jinými typy proměnných.

Špaček upozorňuje, že „*autor ‚zapomíná‘ na přibližnost, která byla do kvantifikace nutně vložena při transformaci, a dále pracuje s kvantifikovanými údaji jako s exaktními hodnotami*“. Děkuji Špačkovi za tuto připomínku. Ano, některé hodnoty transformované proměnné jsou logicky odhadovány a tyto odhady mohou vnášet v určité míře zkreslení do analytických výsledků. Ovšem zkreslení vnáší i redukce kategoriálních proměnných do nižšího počtu kategorií, protože se tím ztrácí část původní informace. Ideální typy metodologických postupů v praxi neexistují. Jak jsem uvedl ve své stati, volba metodologického řešení je v rukou výzkumníka. Je také třeba upozornit na další možné zdroje zkreslení analytických výstupů. Například ne všichni respondenti zařazení náhodným výběrem do sběru dat jsou ochotni spolupracovat, ne všichni zodpoví všechny otázky, může docházet k efektu social desirability atd. Otázku, v jaké míře se na zkreslení analytických výsledků může podílet odhadování některých hodnot roční návštevnosti bohoslužeb oproti dalším zdrojům zkreslení, nelze v tomto okamžiku zodpovědět. Pro zpracování této tematiky by byla potřebná další rozsáhlá práce.

Špaček dále argumentuje: „*Autor například referuje o tom, že díky kvantifikaci může ukázat, že průměrná roční návštevnost bohoslužeb je v ČR 5,5... Údaj o průměru tak naprosto zakrývá fakt, že více než 60 % respondentů nenavštívuje bohoslužby vůbec, dalších téměř 20 % méně než dvakrát za rok a do oné ‚průměrné‘ kategorie spadá pouze 6,5 % respondentů. Tedy stěží se jedná o průměr jakožto vyjádření typičnosti.*“ Pokud bychom chtěli zjistit nejtýpější četnost návštevnosti, pak by opravdu stačilo uvést, v jaké kategorii ordinální proměnné je nejvyšší zastoupení. Ovšem průměrná hodnota roční návštevnosti vypovídá o míře návštevnosti v celé populaci,

**Tabulka 1. Rozložení kategorií původní proměnné ve čtyřech redukovanych kategoriích v případě ČR**

	Nikdy	Nejvýše 2× ročně	Několikrát ročně až 1× měsíčně	2× měsíčně a častěji
Nikdy	100 %	–	–	–
Méně než 1× za rok	–	54 %	–	–
1–2× za rok	–	46 %	–	–
Několikrát za rok	–	–	75 %	–
1× za měsíc	–	–	25 %	–
2–3× za měsíc	–	–	–	26 %
Skoro každý týden	–	–	–	30 %
Každý týden	–	–	–	36 %
Několikrát za týden	–	–	–	8 %

Zdroj: ISSP 2008.

Poznámka:  $n = 1\,484$ . Kendallovo tau-c = 0,754, sig < 0,00.

nikoliv o „typické“ návštěvnosti. Podobně, například pokud je v populaci ČR průměrný věk 42 let, pak je to věk průměrný a neznamená to, že 42letý člověk je typickým Čechem. Podíl „typických“ návštěvníků bohoslužeb v populaci a průměrná roční návštěvnost v populaci jsou různé náhledy, z nichž každý zodpovídá jinou výzkumnou otázku.

Špaček na základě své kategorizace proměnné návštěvnost v tabulce 1 srovnává průměrnou roční návštěvnost s rozložením návštěvnosti v ČR, ve Francii a Japonsku a uvádí: „Srovnání průměrů tak v daném případě nelze popsat lépe než přesné srovnání nepřesných hodnot, respektive hodnot, které pouze málo vypovídají o relevantní komplexitě řešených dat.“ Zde opět dochází k nedorozumění. Ano, mezi uvedenými zeměmi jsou rozdíly v rozložení četnosti návštěvnosti v jednotlivých kategoriích. Tyto rozdíly mohou být dány například rozdílnou kulturou, ale i rozdíly v denominační skladbě v těchto zemích. V různých náboženských věroukách má účást na bohoslužbách v kostelích rozdílný význam. Jak jsem uvedl například v závěru své stati: „První přístup vychází z analýzy zastoupení ve všech kategoriích proměnné. Tento přístup je vhodný pro detailní popis procentuálních podílů jedinců, kteří spadají do jednotlivých kategorií.“ (s. 613) Tedy pro podrobný popis struktury návštěvnosti. Struktura návštěvnosti a průměrná roční návštěvnost jsou dva rozdílné koncepty. Proměnná roční návštěvnost neobsahuje informace o struktuře návštěvnosti. Hodnota průměrné roční návštěvnosti je vypočtenou charakteristikou celé populace, a proto nelze souhlasit se Špačkovým tvrzením: „Srovnání průměrů tak v daném případě nelze popsat lépe než přesné srovnání nepřesných hodnot.“

Otázku přesnosti měření lze dokumentovat i na příkladu srovnání struktury návštěvnosti ve Špačkově kategorizaci oproti kategorizaci původní proměnné v ISSP 2008 (viz Tabulka 1).

Z hodnot v Tabulce 1 je logicky patrné, že při redukcí původní proměnné návštěvnost bohoslužeb, obsažené v datech ISSP 2008, do čtyř kategorií, které Špaček použil ve svém diskusním příspěvku, došlo ke ztrátě části původní informace. Také je třeba upozornit, že každá nová kategorizace by měla být předem teoreticky odůvodněna, což Špaček neučinil. Dále, například nová kategorie „několikrát za rok až jednou měsíčně“ je z původní kategorizace ze 75 % zastoupena jedinci, kteří navštěvují bohoslužby „několikrát za rok“<sup>6</sup> a z 25 % jedinci, kteří bohoslužby navštěvují jednou za měsíc. Přitom návštěvnost například 3krát za rok může znamenat návštěvy kostelů pouze při svátcích nebo společenských událostech. Tedy návštěvnost danou spíše kulturními tradicemi. Na rozdíl od pravidelné návštěvnosti jednou za měsíc, která může vycházet z náboženského přesvědčení. Může se tak jednat nejen o dva typy tradiční religiozity, ale i jejich zastoupení v nové Špačkově kategorizaci je výrazně rozdílné. Pokud by byla tato kategorie použita jako součást vysvětlující proměnné, pak by nastaly problémy s validitou měření, protože by nebyly rozlišeny vlivy kulturních tradic a vlivy náboženského přesvědčení. Bylo by pracováno s aritmetickým průměrem síly těchto dvou vlivů. Ano, v případě slučování kategorií pracujeme s průměry, aniž si to přímo uvědomujeme. Lze pouze shrnout, kategorické proměnné, v datech nejčastější, jsou při analytické práci důležité, ovšem tvorba nových kategorií musí být teoreticky zdůvodněna. Kontinuální proměnná roční návštěvnost žádné informace o struktuře návštěvnosti neobsahuje. V určité míře je to její výhoda i omezení.

Špaček dále srovnává hodnoty  $R^2$  v modelech vícenásobné lineární regrese s vysvětlovanou proměnnou roční návštěvnost dítěte a prediktory návštěvnost matky a otce a uvádí: „*Přímé srovnání přitom ukazuje, že pokud v regresních modelech představených na str. 608 použijeme proměnné kvantifikované prostřednictvím pořadového čísla kategorie (tj. 1, 2, 3... 9), tak ve všech případech jejich prediktivní schopnosti významně vzrostou (viz Tabulka 2).*“ Ano, při použití ordinálních proměnných je v modelu vícenásobné lineární regrese – oproti kontinuálním proměnným – hodnota  $R^2$  vyšší, asi o desetinu. Důvodem je, že hodnota  $R^2$  je dána souběhem průběhů proměnných zavedených do modelu, velikostí hodnot rozptylu. V kvantifikovaných proměnných roční návštěvnost jsou intervaly mezi hodnotami o 1krát ročně. U ordinálních proměnných návštěvnost vyjadřují hodnoty pouze pořadové číslo kategorií, mezi nimiž nejsou v počtech návštěv shodné intervaly. Ano, při použití ordinálních proměnných lze ve vícenásobné lineární regresi získat vyšší hodnoty  $R^2$ , ale to je vše. Jinou výpovědní hodnotu v těchto modelech proměnné, mezi jejichž kategoriemi nejsou shodné intervaly, prostě mít nemohou. Také je

---

<sup>6</sup> Předchozí kategorií je 1–2krát za rok, proto je počáteční hodnota této kategorie 3krát za rok.

nutno upřesnit, že kontinuální proměnná roční návštěvnost má počátek na hodnotě nula (kategorie „Nikdy“) a původní ordinální proměnná návštěvnost má první kategorii návštěvnosti „Nikdy“ na pořadovém čísle jedna.

Na závěr své reakce musím konstatovat, že diskusní příspěvek Ondřeje Špačka na mě již svým názvem „Jak neargumentovat při transformaci proměnné“ působí spíše jako názorový postoj. Mnohem raději bych četl diskusní příspěvky, i negativní, které se snaží s menším či větším úspěchem posouvat hranice sociologické imaginace. Ne všechny inovace se uchytí, některé myšlenky se ukážou být „slepou uličkou“ nebo se ukážou jako nepotřebné a upadnou v zapomnění. Výsledkem ne každé střely na branku je gól, ale bez těchto střel by góly nikdy nepadaly. Užitečná hodnota kontinuálních proměnných roční návštěvnost se ukáže nebo také neukáže až po určité době.

*Antonín Paleček*