

---

## Transformace kategorické proměnné na proměnnou kontinuální na příkladu návštěvnosti bohoslužeb\*

ANTONÍN PALEČEK\*\*

Fakulta sociálních studií, Masarykova univerzita, Brno

### The Transformation of a Categorical Variable into a Continuous Variable: The Example of Church Attendance

**Abstract:** Many surveys include variables on the frequency of respondents' church attendance. Usually these are a categorical, ordinal type of variable. However, in a quantitative analysis of the inter- and intragenerational reproduction of church attendance categorical variables may have some disadvantages and limitations. This article discusses and tests the possibilities of a different methodological approach, where ordinal variables indicating the frequency of church attendance are transformed into numerical-type variables. More simply put, respondents are asked how many times a year they attend church and this information is expressed in a new variable. There are a few analytical advantages to this numerical variable. First, the information on the frequency of church attendance is not divided up into multiple categories and can be transposed for use in multivariate analysis. Second, even if there are only a small number of cases in the data set there is no need to reduce the categories of variables. Third, the results of the analysis can be presented as a single numerical value. The individual examples in this analysis focus narrowly on the rates of inter- and intragenerational reproduction of church attendance, and the analytical possibilities of numerical variables on attendance rates are tested on data from the Religion module of the International Social Survey Programme 2008.

**Keywords:** frequency of church attendance, methodology of quantitative religious research, numerical variable, intergenerational and intragenerational transmission

*Sociologický časopis / Czech Sociological Review, 2017, Vol. 53, No. 4: 593–615*

*<https://doi.org/10.13060/00380288.2017.53.4.359>*

---

\* Tento text byl podpořen projektem specifického výzkumu na Fakultě sociálních studií Masarykovy univerzity „Společnost a její proměny: kvalitativní a kvantitativní pohled (MUNI/A/1182/2016)“.

\*\* Veškerou korespondenci posílejte na adresu: Mgr. Antonín Paleček, Katedra sociologie, Fakulta sociálních studií Masarykovy univerzity, Joštova 10, 602 00 Brno, e-mail 98347@mail.muni.cz.

## Úvod

Proměnné, které poskytují informace o návštěvnostech bohoslužeb, patří v českém kvantitativním výzkumu religiozity mezi nejčastěji využívané [Váně, Hásová 2014], ze dvou základních důvodů. Za prvé, návštěvnost na bohoslužbách jako prakticky uskutečňované náboženské chování [Glock 1962] je spolu s podíly lidí hlásících se k náboženskému vyznání řazena mezi indikátory tradiční religiozity. Prostřednictvím poklesu hodnot těchto indikátorů je sledován proces sekularizace společnosti. Za druhé, proměnné obsahující informace o četnosti účasti jedince na bohoslužbách v kostelích jsou k dispozici v datech z mnoha výběrových šetření. Pro příklad. Proměnné uvádějící informace o četnostech účastí dospělého respondenta v době sběru dat i v jeho formativním období jsou obsaženy ve výběrových šetřeních World/European Value Survey (WVS/EVS), European Social Survey (ESS) i International Social Survey Programme (ISSP). Vlny ISSP 1991, 1998 a 2008 poskytují navíc tyto údaje i v případech rodičů respondenta v jeho dětství. Pro sekundární analýzu jsou proto k dispozici mnohé datové soubory z různých výběrových šetření i časových období.

Ve všech výše zmíněných výběrových šetřeních je návštěvnost bohoslužeb uváděna formou proměnných ordinálního typu s velkým počtem kategorií. Například v případě výběrového šetření ISSP je to devět kategorií, na škále od nikdy až po vícekrát týdně (viz kap. Transformace proměnné), spolu s možnostmi neví a neodpověděl. Takovéto rozsáhlé škály umožňují podrobnou analýzu měr návštěvnosti dospělých jedinců, těchto jedinců v dětství i jejich rodičů. Dále popis míry návštěvnosti v případech konkrétně definovaných skupin (např. dle regionu, pohlaví, věku, období atd.), srovnání měr návštěvnosti mezi těmito dvěma nebo více skupinami a jejich možná vysvětlení nebo také zkoumání úspěšnosti předávání míry návštěvnosti z generace na generaci. Je však třeba poukázat i na některé analytické problémy, se kterými se může výzkumník setkat. Velký počet kategorií proměnné na jedné straně umožňuje detailní analýzu sledovaného stavu nebo procesu, ale na straně druhé zvyšuje nároky na dostatečný počet případů v datovém souboru. Při třídění vyšších řádů tak zpravidla dochází k nedostatečnému zastoupení absolutních četností v některých buňkách vícerozměrných tabulek. Ve většině empirických studií nebývá použita celá kategorizace, ale počet kategorií je redukován v závislosti na výzkumné otázce a návazně potřebě ověření hypotéz. Sloučení některých původních kategorií je doprovázeno nárůstem absolutních četností v nově definovaných kategoriích, ale současně také dochází ke ztrátě části informace, kterou obsahovala právě hranice mezi sloučenými kategoriemi. Pokud výzkumník potřebuje pracovat s celou škálou kategorizace, při současné možnosti užití vícerozměrné analýzy, pak je postaven před konkrétní metodologický problém a otázku jeho možného řešení.

Tato stať vznikla v reakci na potřebu metodologického řešení výše uvedené problematiky v konkrétním případě sledování míry úspěšnosti přenosu návštěvnosti bohoslužeb mezi generacemi. Přesněji, v náhledu úspěšnosti přenosu mezi rodiči a jejich potomky ve formativním období – v průběhu intergenerační

transmise – a následně mezi tímto formativním obdobím a dospělostí jedince – v průběhu intragenerační transmise. Pokud by bylo cílem studie zjištění inter a intrageneračního přenosu určité pravidelnosti v návštěvnosti, například pravidelné účasti na bohoslužbách alespoň každý měsíc nebo každý týden, pak je namístě redukce proměnné do požadovaného počtu kategorií. Při sloučení některých kategorií následně dochází k nárůstu absolutních četností v nově definovaných kategoriích a tím narůstají i možnosti užití vícerozměrných analýz. Pokud si však studie klade otázku, v jaké míře se zvýšila nebo opačně poklesla návštěvnost například v průběhu intergenerační transmise, to znamená kolikrát více nebo méně dítě navštěvovalo bohoslužby oproti svým rodičům, pak je s ohledem na potřebu práce s neredukovanými informacemi nutno použít celé dostupné kategorizace proměnných návštěvnosti rodičů i návštěvnosti potomka. Tímto následně vzniká otázka po vhodném metodologickém postupu, který by umožnil analyticky pracovat se všemi devíti kategoriemi proměnných návštěvnosti rodičů i potomků a přitom při limitovaném počtu případů v datovém souboru by mohly být použity i vícerozměrné analýzy.

Cílem této stati je navrhnout a testovat takovýto možný metodologický postup. Škály proměnných návštěvností bohoslužeb o devíti kategoriích, které jsou u velkých výběrových šetření obvyklé, obsahují nejen velmi podrobné informace o četnostech návštěv, ale i určitý časový údaj. Například návštěvnost v kategoriích 1krát měsíčně nebo 1krát týdně by bylo možno převést na návštěvnost 12krát nebo 52krát ročně. Proto bude na základě všech devíti kategorií proměnné návštěvnosti bohoslužeb a v nich obsažených časových údajích vytvořena nová číselná proměnná, která bude obsahovat informaci, kolikrát se jedinec účastní bohoslužeb v průběhu jednoho roku. Možnosti a omezení této transformované číselné proměnné budou dále podrobněji rozvedeny a následně v analytické části testovány na mezinárodním datovém souboru ISSP 2008 – Religion.

Je třeba upozornit, že tato stať je ryze metodologická. Jejím cílem je pouze představení a testování transformované číselné proměnné návštěvnosti bohoslužeb na konkrétních příkladech přenosu měr návštěvnosti v průbězích inter a intragenerační transmise. Možné příčiny poklesů úspěšnosti přenosu religiozity a tím i procesu sekularizace již byly shrnuty v mnoha publikacích [např. Lužný 1999; Nešpor, Václavík a kol. 2008; Vido 2011]. Proto, i když bude na praktických příkladech sledována změna návštěvnosti bohoslužeb v závislosti na některých vysvětlujících faktorech, nebudou vlivy těchto prediktorů spojovány s teoretickými východisky. Zájemce o sekularizační teorie lze odkázat na výše zmíněné zdroje.

## **Metodologické otázky**

Nejprve je třeba zohlednit otázku, proč zkoumat pokles religiozity ve společnosti prostřednictvím změn, ke kterým dochází v průbězích inter a intragenerační transmise. Z důvodů neustálé obměny lidské populace je i přetrvávání religio-

zity ve společnosti závislé na úspěšnosti jejího předávání z generace na generaci [Myers 1996]. Popis mechanismu přenosu religiozity, spolu s teoretickým vysvětlením míry jeho úspěšnosti, byl již publikován v předchozích pracích [např. Paleček, Vido 2014]. Rodiče prostřednictvím primární náboženské socializace předávají svoji religiozitu v určité míře svým dětem, kdy hovoříme o míře přenosu v průběhu intergenerační transmise. Tyto děti následně dospívají a svoji primárně socializovanou religiozitu si opět v určité míře mohou udržet v průběhu svého dospělého života, kdy hovoříme o míře přenosu během intragenerační transmise. Tito dospělí jedinci, jejichž religiozita se může lišit od religiozity jejich rodičů, se stávají potenciaálními rodiči, kteří opět primárně socializují své potomky, a takto se kruh přenosu uzavírá. Proto je míra i skladba religiozity v současné společnosti odvislá od míry úspěšnosti přenosu v obou transmisích. Při sledování změn četnosti návštěvnosti bohoslužeb, například v průběhu intergenerační transmise, může být četnost návštěv dětí shodná s četností návštěv jejich rodičů – v tom případě dochází k plné reprodukci –, četnost návštěv dětí může být nižší – dochází k sekularizaci – nebo naopak vyšší – dochází k revitalizaci [Paleček 2015]. Stejně procesy jsou i součástí následné intragenerační transmise. V grafickém znázornění se jedná o linii rodiče–děti–dospělí potomci (stávající se opět potenciaálními rodiči). Důsledkem změn v obou transmisích je proces sekularizace nebo revitalizace. Proto je při analýze sekularizačních nebo revitalizačních procesů nutno sledovat míry návštěvnosti na začátcích i koncích obou transmisí. To znamená návštěvnost rodičů, jejich dětí a těchto dětí v dospělosti.

Podrobné informace o návštěvnosti dospělých respondentů, těchto respondentů v dětství a jejich rodičů obsahuje například mezinárodní výběrové šetření ISSP – Religion nebo český výzkum Detradicionalizace a individualizace náboženství 2006, a to v proměnných o devíti kategoriích. Potřebné proměnné jsou tedy k dispozici. Dalším krokem je otázka, jak s těmito proměnnými při sledování změn v průběhu určité transmise analyticky pracovat. Použít všechny informace obsažené v devíti kategoriích, počet těchto kategorií redukovat v závislosti na cíli analýzy, nebo hledat další možná metodologická řešení? Jako konkrétní příklad využití celé kategorizace byly do modelu v Tabulce 4a zavedeny dvě nezávislé a jedna závislá číselná proměnná, které obsahují roční četnosti návštěv bohoslužeb v případech obou rodičů a jejich potomků v dětství. Pokud by byl tento vztah zkoumán prostřednictvím vícerozměrné tabulky a při použití původních proměnných s devíti kategoriemi, pak by vznikla tabulka o rozměrech  $9 \times 9 \times 9 = 729$ . I při absolutní homogenitě rozložení a žádných ztracených případech (missing values) by při počtu případů v ISSP 2008 pro ČR = 1512 připadalo na jednu buňku této tabulky v průměru 2,1 případů, vypočteno  $1512/729 = 2,1$ . Avšak jak je patrné již z rozložení proměnné v Grafu 1, zastoupení v jednotlivých kategoriích se výrazně liší. Proto je možno předpokládat, že by absolutní četnosti v takovéto vícerozměrné tabulce byly v mnoha jejích buňkách velice nízké nebo nulové. S takto nízkými hodnotami absolutních četností je problematické statisticky pracovat. V sociálně stratifikačním výzkumu je tento problém limitován

velikosti výběrového vzorku mnohdy řešen slučováním datových souborů z více výběrových šetření, jehož výsledkem je nárůst počtu případů až na několik desítek tisíc [např. Simonová 2007]. Pro analýzu návštěvnosti bohoslužeb je však rozhodující, že pokud bychom chtěli sledovat například přenos návštěvnosti mezi rodiči a jejich potomky v ČR, pak bychom mohli sloučit pouze data DIN 2006<sup>1</sup> a vlny ISSP 1998 a 2008, protože pouze tato tři výběrová šetření obsahují proměnné sledující i návštěvnost rodičů.

Metodologické přístupy k řešení problematiky limitovaného počtu případů lze rozdělit zhruba do dvou směrů. Jednak jsou sledována zastoupení ve všech kategoriích, tj. třídění I. řádu, které neklade vysoké nároky na velikost výběrového vzorku [např. Hamplová 2000: 17, graf 1 a 2], případně při použití nižšího počtu třídících kritérií, například genderu [Hamplová 2011]. Aby však bylo možno zkoumat sociální stavy a procesy hlouběji, bylo by třeba, aby na základě teoretických východisek vstupoval do analýzy potřebný počet prediktorů. Při vysokém počtu kategorií sledované proměnné i vyšším počtu použitých třídících kritérií (například ve vícerozměrných tabulkách) však zásadně klesá i zastoupení absolutních četností v jednotlivých buňkách tabulek. Tímto je užití kategoriálních proměnných v analýzách limitováno počtem případů v datovém souboru jako celku i v jednotlivých kategoriích závislé i nezávislé proměnných. Důsledkem pak může být problém zobecnění analytických výsledků z výběrového vzorku na cílovou populaci. Při měření vztahů mezi dvěma nebo více kategoriálními proměnnými by bylo možno použít hodnoty asociací/korelací, ale v těchto případech jsou statistickými výstupy pouze hodnoty vztahů mezi proměnnými, bez možnosti rozboru těchto vztahů v jejich jednotlivých kategoriích.

Druhým často používaným metodologickým řešením je slučování některých kategorií dané proměnné. Ordinální proměnná návštěvnost bohoslužeb s devíti kategoriemi umožňuje vytvářet nové kategorie dle cíle studie. V závislosti na kulturním kontextu může být populace rozdělena na pravidelné a nepravidelné účastníky bohoslužeb, kdy je tato pravidelnost nastavena na účast například každý měsíc nebo každý týden. Při sledování přenosu návštěvnosti v průběhu konkrétní transmise by prostřednictvím takového konceptu pravidelné návštěvnosti byla sledována míra reprodukce kulturně definované pravidelnosti. Na druhé straně je výsledkem tohoto řešení také nárůst zastoupení absolutních četností v jednotlivých (sloučených) kategoriích pravidelné a nepravidelné návštěvnosti. Ovšem důsledkem redukce kategorií proměnné do nižšího počtu je také ztráta části informací, které byly obsaženy v původní kategorizaci. Přesněji, pokud by byla, například z dat ISSP 2008, vytvořena nová dichotomická proměnná s kategoriemi pravidelná návštěvnost – alespoň jednou měsíčně – a nepravidelná návštěvnost – méně než jednou měsíčně [např. Paleček 2011], pak by bylo v první kategorii této redukované proměnné sloučeno pět původ-

---

<sup>1</sup> Výzkumné šetření Detradicionalizace a individualizace náboženství pod vedením Dušana Lužného.

ních kategorií: 1krát měsíčně, 2–3krát měsíčně, skoro každý týden, každý týden a vícekrát týdně. Tato redukce ve svém důsledku znamená jednak ztrátu informace přítomné v původní kategorizaci. Za druhé je možno si představit, jakou heterogenitou se bude vyznačovat kategorie proměnné, která slučuje jedince navštěvující bohoslužby v rozmezí 1krát měsíčně až vícekrát týdně. Dále je třeba se ptát: Pokud by byla výše popsaná dichotomická kategorizace použita například ve vysvětlující proměnné, tak s jakým rozdílem ovlivňuje postoje nebo jednání člověka (jako vysvětlovanou proměnnou) jeho návštěvnost bohoslužeb každý měsíc, tj. 12krát ročně a vícekrát týdně, tj. více než 52krát ročně? Dalším problémem je konceptualizace pravidelné návštěvnosti. Očekávaná pravidelnost účastí na bohoslužbách se může lišit mezi jednotlivými kulturami, ale i uvnitř těchto kultur v průběhu času [Filipi 2001]. K určitým zkreslením dochází i při zkoumání přenosu návštěvnosti například mezi rodiči a jejich potomky. Pro příklad opět s použitím dichotomické proměnné v kategoriích návštěvnost alespoň 1krát měsíčně a méně než 1krát měsíčně. Pokud by se rodič účastnil bohoslužeb vícekrát týdně a jeho potomek 1krát měsíčně, pak by oba ve výše zmíněných dichotomických proměnných spadali do stejné kategorie. To znamená, že by byla zjištěna reprodukce mezi rodičem a jeho potomkem, ačkoliv jejich četnost účasti byla výrazně rozdílná, tj. vícekrát týdně  $\geq$  52krát ročně a 1krát měsíčně = 12krát ročně. Naopak, pokud se rodič účastnil bohoslužeb vícekrát ročně, uvažujme například 6krát ročně, a jeho potomek 1krát měsíčně, tj. 12krát ročně, pak by byl při použití uvedené dichotomické kategorizace mezi rodičem a jeho potomkem zjištěn nárůst návštěvnosti, tj. revitalizace v této transmisi. Přitom mezi ročními návštěvnostmi 12krát ročně a více než 52krát ročně je mnohem vyšší rozdíl než mezi návštěvnostmi 12krát ročně a 6krát ročně. Přesnost měření přenosu návštěvnosti v určité transmisi narůstá s počtem sledovaných kategorií. Oproti dichotomické proměnné lze očekávat přesnější výsledky například při použití kategorizace každý týden, každý měsíc, méně než každý měsíc a nikdy. S nárůstem počtu kategorií však klesá zastoupení absolutních četností v buňkách tabulek a tím je limitován počet použitých třídících kritérií a také narůstá počet výsledných hodnot a následně i interpretační složitost. Konceptualizace určité pravidelné návštěvnosti není bezproblémová ani při srovnávání měř návštěvnosti mezi více skupinami, například společnostmi s rozdílnými kulturami nebo náboženskými konfesemi s rozdílnými požadavky vůči svým věřícím.

Jednotlivé empirické studie se v počtech sledovaných kategorií shodují nebo i liší. Čtenář se setkává s četnostmi návštěv bohoslužeb redukovány do dichotomických proměnných například alespoň jednou týdně a méně [Polack 2008; Ruiter, van Tubergen 2009], více než jednou měsíčně a méně [Need, Evans 2001], alespoň jednou měsíčně a méně [Lambert 2004]. Také s redukcí do tří či čtyř kategorií, alespoň jednou měsíčně, méně a nikdy [Hamplová, Řeháková 2009], 2krát měsíčně a častěji, několikrát ročně, nanejvýš 1krát ročně a nikdy [Nešpor 2010: 62]. Všechny tyto typy redukcí mají svá teoretická opodstatnění. Jsou plně relevantní při sledování stavu nebo vývoje blíže definované pravidelné návštěvnosti,

například každý měsíc nebo každý týden. Výsledkem analýz jsou pak podíly lidí pravidelně navštěvujících bohoslužby nebo síla vlivů teoreticky definovaných faktorů, které tuto pravidelnost ovlivňují. Pokud by byla sledována také míra úspěšnosti přenosu definované pravidelné návštěvnosti mezi generacemi [např. Paleček 2011], pak by byly výsledkem analýzy podíly jedinců, kteří tuto pravidelnost reprodukovali nebo nereprodukovali. Pokud je však cílem studie zjistit změny v četnostech účastí na bohoslužbách, například mezi rodiči a dětmi, pak by bylo s ohledem na omezení možných zkreslení nutno použít celou škálu o devíti kategoriích. Jak již bylo pojednáno výše, s nárůstem počtu kategorií vysvětlované i vysvětlujících proměnných při limitovaném počtu případů také klesá možnost použití vícerozměrných analýz. Dále je nutno upozornit na další možná úskalí redukce do nižšího počtu kategorií. V případech různých počtů nových kategorií nebo rozdílnosti hranic mezi těmito kategoriemi v nových (rekódovaných) proměnných, které jsou užívány v různých empirických pracích, vyvstává i problém komparace analytických výsledků mezi těmito pracemi. Jen pro příklad, bylo by zavádějící tvrzení, že 40 % jedinců navštěvujících bohoslužby více než jednou týdně je méně než 60 % navštěvujících bohoslužby více než jednou měsíčně. Při výběru vhodného metodologického postupu, jehož prostřednictvím je možno zkoumat míru přenosu návštěvnosti v průběhu inter a intragenerační transmise, tak vyvstává otázka, jakým způsobem zjišťovat nárůst nebo pokles návštěvnosti v průběhu konkrétní transmise, a to s využitím celé škály návštěvnosti – celé informace, která je k dispozici v datových souborech.

### **Transformace proměnné**

Při úvahách o možném řešení problémů a otázek, které byly pojednány v předchozí části, vychází tato stať z již dříve publikovaných metodologických postupů. Jedná se například o empirickou studii zjišťující četnosti kontaktů vnoučat s jejich prarodiči, ve které autor sleduje skutečněnou četnost kontaktů během posledních 12 měsíců [Geurts, Poortman, Dykstra 2009]. Proměnná četnost kontaktů nabývá v této studii hodnot v intervalu 0 až 365, protože 12 měsíců tvoří jeden rok, který má standardně 365 dnů. Tento metodologický postup se jeví vhodným i při zkoumání přenosu návštěvnosti mezi generacemi.

Časovou informaci, která vychází z kalendářního rozdělení roku, obsahují i kategorie proměnné návštěvnost bohoslužeb. Datový soubor ISSP 2008 nabízí tyto proměnné v devíti kategoriích: Nikdy, méně než 1krát ročně, asi 1krát až 2krát ročně, vícekrát ročně, asi 1krát měsíčně, 2krát až 3krát měsíčně, skoro každý týden, každý týden a vícekrát týdně. Při transformaci této kategoriální proměnné do proměnné číselného typu vycházejme z kalendářního uspořádání času. Každý rok má 12 měsíců a 52 týdnů. Proto je možno se ptát, kolikrát jedinec navštíví bohoslužby v průběhu jednoho kalendářního roku. Na základě této otázky a rozdělení kategorií původní kategoriální proměnné je každé kategorii přiřazena číselná hodnota četnosti návštěv bohoslužeb, ke které dochází v průběhu jednoho roku:

Nikdy = 0; méně než 1krát ročně = 0,5; 1krát až 2krát ročně = 1,5; vícekrát ročně = 6; 1krát měsíčně = 12; 2krát až 3krát měsíčně = 30; skoro každý týden = 41; každý týden = 52; a vícekrát týdně = 104. Jak je patrné, transformace kategorií do číselných hodnot, tj. četností návštěv v průběhu jednoho roku, není u všech kategorií zcela bezproblémová. Zatímco například kategorii nikdy je bezproblémově přiřazena hodnota 0, podobně kategoriím měsíčně a týdně 12 a 52, pak v některých dalších kategoriích je nutno tuto hodnotu logicky odhadovat. Jedná se o kategorie méně než 1krát ročně, která je kódována na hodnotu 0,5krát ročně, protože sousedící kategorie nikdy má hodnotu 0 a 1krát ročně znamená hodnotu 1. Takto je odhadnutá hodnota tvořena jako logicky odvozený střed. Podobně je postupováno u dalších odhadovaných hodnot. Výjimkou je kategorie vícekrát týdně = 104krát ročně, která nemá druhou sousední kategorii nebo možnou variantu. V případě ČR by tento odhad neměl způsobit velké zkreslení, protože účast vícekrát týdně uvádí méně než 1 % dospělé populace (viz Graf 1). Prostřednictvím této transformace tak v datových souborech vznikají nové, číselné proměnné, které například v datech ISSP 2008 obsahují roční četnosti návštěv respondenta v současnosti, v jeho formativním období i četnosti návštěv jeho matky a otce.

Pro připomenutí, u ordinální proměnné je možno zjistit, zda je jedna kategorie vyšší/nížší oproti jiným, ale ne o jakou hodnotu, tuto hodnotu známe pouze u proměnné intervalové [Disman 1993: 188–195]. Transformovaná číselná proměnná roční návštěvnost bohoslužeb obsahuje informaci, kolikrát ročně dochází k účasti na bohoslužbách. Interval je vymezen hodnotou roční návštěvnosti o 1krát více nebo méně oproti sousední hodnotě. V nové, transformované číselné proměnné je v případě respondenta samozřejmě obsaženo tolik variant hodnot, kolik kategorií měla původní kategoriální proměnná, tzn. v případě ISSP 2008 je to devět variant hodnot. Ovšem v původní proměnné vyjadřovalo číselné označení kategorií (1–9) pouze vzestupný či sestupný trend těchto kategorií. V nové, transformované proměnné bude číselná hodnota obsahovat informaci, kolikrát ročně k návštěvám bohoslužeb dochází, a to v rozsahu 0–104krát ročně v hodnotách celých čísel. V případě jedince v jediné hodnotě a v případě skupiny v hodnotě průměrné roční návštěvnosti v této skupině, například v určité zemi, u matek, jejich dětí apod. V případě průměrné návštěvnosti definované skupiny může tato průměrná návštěvnost teoreticky nabývat jakékoliv hodnoty v rozsahu 0–104.

Při výpočtu rozdílu mezi četnostmi návštěv dvou jedinců, například mezi rodičem a jeho potomkem, vzniká nová, rozdílová proměnná, obsahující hodnotu rozdílu návštěvnosti mezi tímto rodičem a jeho potomkem. Mezi rodiči a potomky pak v průměrných hodnotách, tj. v hodnotách mobility v průběhu intergenerační transmise. Tato rozdílová proměnná nabývá hodnot v intervalu od –104 do +104. Takovéto rozdílové proměnné mohou být vypočteny i na základě dalších dělicích kritérií, například dle populací jednotlivých zemí, genderu, období, věku, denominační příslušnosti atd. Při analýzách by pak mohly být tyto vypočtené rozdílové proměnné použity jako proměnné vysvětlované i vysvětlující.



Pro interpretace analytických výsledků je také přínosné, že na rozdíl od původní proměnné o devíti kategoriích nebo její redukce do tří či čtyř kategorií jsou stavy i změny návštěvnosti vyjádřeny pouze jednou hodnotou. Jedinou číselnou hodnotou lze například popsat roční návštěvnost jedince nebo průměrnou roční návštěvnost definované skupiny, například země, rodičů, potomků atd., nebo rozdíl průměrných hodnot ročních návštěvností mezi dvěma subjekty, například rodiči a jejich potomky. V modelech lineární regrese je také jedinou hodnotou zjišťován vliv změny nezávislé proměnné na proměnnou závislou, o jakou hodnotu roční účasti se změní závislá číselná proměnná při změně prediktoru. Z četností roční návštěvnosti nebo rozdílů v těchto četnostech lze počítat například průměry, meziskupinové i vnitroskupinové rozptyly, korelace, odhady parametrů lineární regrese atd.

Jednou číselnou hodnotou lze vyjádřit změny i v případě závislé ordinální proměnné, a to v modelech ordinální logistické regrese [viz Řeháková 2000]. Ovšem i v těchto modelech jsou na závislou i nezávislé proměnné kladeny určité požadavky, především požadavek na test rovnoběžnosti a na dostatečný počet případů ve výběrovém vzorku, protože model ordinální logistické regrese pracuje s nezávislými proměnnými jako s proměnnými nominálního typu. Pro srovnání, pokud bychom například modelovali vliv návštěvnosti matek na primárně socializovanou návštěvnost jejich dětí, byly by s použitím obou transformovaných proměnných roční návštěvnosti výsledkem lineární regrese dva parametry, jeden regresní koeficient a jedna konstanta. A právě tento regresní koeficient je odhadem, o kolikrát ročně více se dítě zúčastňuje bohoslužeb, pokud se návštěvnost matky změní o hodnotu 1krát ročně. Při použití ordinální logistické regrese by byla jedna z devíti kategorií návštěvnosti matky použita jako referenční kategorie a výsledkem ordinální regrese by pak bylo osm hodnot a jedna hodnota koeficientu.

Avšak analytické užití číselných proměnných (proměnných spojitých, intervalových nebo ordinálních s velkým počtem kategorií) má také svoji metodologickou problematiku. Je to především často zmiňovaný požadavek na normální rozložení zkoumané proměnné, tzn. kdy je distribuce symetricky centrována okolo průměru (Gaussova křivka) [např. Field 2000]. V případě transformované číselné proměnné návštěvnost bohoslužeb současné české populace však nelze takovouto normalitu rozložení očekávat, protože jak bude znázorněno v Grafu 1, větší část obyvatelstva se bohoslužeb vůbec neúčastní [srov. Hamplová 2000: graf 1]. V transformované proměnné, která nabývá hodnot v intervalu 0 až 104 je tak četnost jedinců nezúčastňujících se bohoslužeb v krajním bodu tohoto intervalu, tj. na hodnotě 0. Oproti ČR by však bylo možno normalitu rozložení očekávat v zemích s vysokou mírou tradiční religiozity, například v Polsku, Irské republice, Irsku nebo USA, kde většina obyvatelstva navštěvuje bohoslužby každý týden a nikdy jen malá část [viz Hamplová 2000: graf 3]. Návštěvnost 1krát týdně se v těchto zemích v transformované proměnné rovná hodnotě 52, tzn. uprostřed intervalu 0 až 104. Je možno se ptát: Pokud by v případě ČR nebyla normalita

rozložení prokázána, bylo by toto zjištění nepřekonatelnou překážkou při analytickém využití transformované číselné proměnné roční návštěvnosti bohoslužeb?

Proměnné indikující sociální charakteristiky se svým rozložením od normálního rozložení mnohdy liší. Z teoretického hlediska by takovéto případy znamenaly nemožnost použití statistik založených na parametrických testech. Oproti tomu někteří autoři uvádějí, že pokud je ve výběrovém vzorku vyšší počet případů než 100, pak je možno použít i statistické postupy, které normalitu rozložení vyžadují [De Vaus 2002]. Požadavek na normalitu rozložení měl svoji důležitost v dobách, kdy datové soubory obsahovaly jen několik desítek případů. Datový soubor ISSP 2008 obsahuje pro ČR 1512 případů. Ovšem nelze také opomíjet, že při užití třídění vyšších řádů se limit  $> 100$  případů týká každé kategorie vysvětlující proměnné. Dnes již klasická učebnice statistických postupů uvádí čtyři základní předpoklady pro statistickou inferenci při odhadování parametru lineární regrese [Agresti, Finlay 1999: 326].

1. Vztah mezi proměnnými  $Y$  a  $X$  je lineární.
2. Standardní chyba se neliší mezi jednotlivými hodnotami proměnné  $X$ .
3. Distribuce proměnné  $Y$  je normální ve všech hodnotách  $X$ .
4. Výběrový vzorek byl vybrán náhodným výběrem.

Autoři dále uvádějí: „Nejdůležitější je předpoklad 1 a 4,“ a k tomu ještě přidávají: „pro velké výběrové vzorky je předpoklad normálního rozložení nedůležitý“ [Agresti, Finlay 1999: 327]. Vzorek respondentů v ISSP i v mnoha dalších výběrových šetřeních, jako například WVS/EVS, byl vybírán víceúrovňovým pravděpodobnostním výběrem, čímž je splněna podmínka č. 4. Datový soubor ISSP 2008, který bude užit v analytických příkladech, obsahuje pro ČR 1512 případů, pro každou další zúčastněnou zemi je to vždy více než 900. Proto lze podle výše citovaných autorů v takovýchto velkých výběrových vzorcích považovat podmínku normálního rozložení proměnné za méně důležitou. Linearita vztahu mezi závislou a nezávislou proměnnou (podmínka č. 1) bude zjištěna až z výstupních hodnot modelu lineární regrese. Na základě výše citovaných autorů by tedy odchylka od normálního rozložení transformované proměnné návštěvnosti bohoslužeb neměla být překážkou při užití této proměnné i v pokročilých statistických postupech. Transformované číselné proměnné návštěvnosti bohoslužeb nabývají tolik hodnot, kolik jich měly původní ordinální proměnné, ze kterých byly s pomocí časového údaje konstruovány. „Má-li ordinální proměnná dostatečný počet kategorií, říkává se alespoň sedm, můžeme s ní pracovat jako se spojitou.“ [Řeháková 2000] Například v modelech lineární regrese, ve kterých odvozujeme kvalitu modelu v závislosti na hodnotě „ $R$ “ a možnost zobecnění výsledků z výběrového vzorku na celou populaci na základě statistické signifikance.

Nakonec je možno uvažovat, jaké statistické výsledky by mohlo přinést použití výše představené transformované číselné proměnné. Při popisu hodnot četnosti roční návštěvnosti nebo hodnot rozdílových proměnných lze říci, o kolik

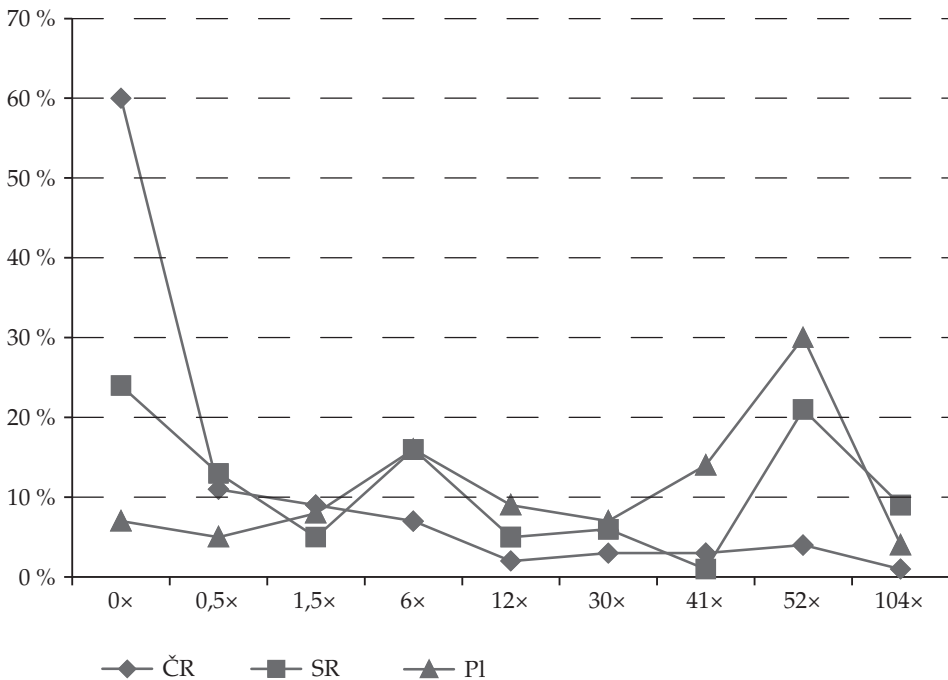
nebo kolikrát se liší hodnoty průměrů četností roční návštěvnosti v závislosti na vysvětlujících proměnných, například mezi populacemi jednotlivých zemí, mezi náboženskými denominacemi, mezi věkovými kategoriemi, v závislosti na genderu, mezi rodiči a jejich potomky a i následně mezi těmito potomky v dětství a v dospělosti atd. Z výsledků hodnot regresních koeficientů lze také zjistit, v jaké hodnotě se změni vysvětlovaná proměnná četnost roční návštěvnosti, pokud se vysvětlující proměnná změní o jeden bod, tj. o jednu návštěvu ročně. Nakonec, možnosti komparací hodnot četností ročních návštěvností nekončí u porovnávání výsledků v jedné empirické práci. Přínosem mohou být i bezproblémové komparace statistických výsledků z více výběrových šetření, která obsahují odlišné kategorizace proměnné indikující návštěvnost bohoslužeb. Analytické možnosti a postupy při užití číselných proměnných jsou podrobně popsány v mnoha publikacích [např. Agresti, Finlay 1999; Disman 1993; Fox 2008; Hamplová 2005; Hendl 2004; Kreft, Leeuw 1998; Norušis 1998; Soukup 2006]. Pro práci s IBM SPSS lze doporučit například publikaci *Discovering Statistics Using SPSS for Windows* [Field 2000].

### **Analytické příklady**

V této části budou představeny některé možnosti užití transformované proměnné na konkrétních příkladech analýzy míry úspěšnosti přenosu návštěvnosti mezi generacemi. Nejprve bude zkoumáno rozložení transformované proměnné a testována jeho normalita. Druhý příklad se zaměřuje na porovnání průměrných hodnot ročních návštěvností a v posledních dvou příkladech bude sledován přenos míry návštěvnosti v průbězích inter a intragenerační transmisi. Pro možnost srovnání bylo spolu s vysoce sekulární ČR vybráno více religiózní Slovensko a vysoce religiózní Polsko [viz Hamplová 2000].

#### *Rozložení transformované proměnné*

Rozložení transformované číselné proměnné roční návštěvnost bohoslužeb v současných dospělých populacích ČR, Slovenska a Polska je zobrazeno v Grafu 1. Z tvarů křivek je již na první pohled patrné, že se rozložení proměnné může blížit normálnímu rozložení pouze v případě Polska. V ČR většina populace (60 %) nenavštěvuje bohoslužby vůbec a na Slovensku jsou nejvyšší podíly také bez účasti (24 %) a dále účastníků se 52krát ročně, tj. 1krát týdně (21 %). V případě ČR se tvar křivky výrazně odlišuje od normálního rozložení. Tento tvar by se mohl přiblížit normálnímu rozložení, pokud by poklesly podíly s nulovou účastí a narostly podíly s účastí okolo středu intervalu 0 až 104, jak je tomu v případě Polska. Normalita rozložení je také zjišťována prostřednictvím středních hodnot proměnné, tzn. prostřednictvím průměru a mediánu, nebo podle tvaru křivky, tzn. podle hodnot šikmosti a špičatosti.

**Graf 1. Rozložení proměnných roční návštěvnosti bohoslužeb v dospělých populacích ČR, SR a Polska (hodnoty podílů v %)**

Zdroj: ISSP 2008.

Poznámka: N: ČR = 1484, SR = 1116, PI = 1251.

Tyto hodnoty jsou uvedeny v Tabulce 1. Současná dospělá česká populace se účastní bohoslužeb v průměru 5,5krát ročně, ale střední hodnotou neboli mediánem, dělicím případy na shodnou část vyšších a nižších četností, je nulová roční účast (viz také Graf 1). Proto jsou i hodnoty šikmosti a špičatosti<sup>2</sup> v absolutních hodnotách výrazně vyšší než 1. V případě Slovenska již hodnota šikmosti jen mírně převyšuje hodnotu 1 a špičatost je 0,7. Z vybraných zemí tak kritéria normálního rozložení plně splňuje nejlépe populace Polska, kde se hodnota průměru (29,8) prakticky rovná hodnotě mediánu (30). Tato normalita rozložení je pak dále indikována hodnotami šikmosti a špičatosti blížícími se nule. V ČR a na Slovensku tedy transformovaná číselná proměnná návštěvnost bohoslužeb ne-

<sup>2</sup> Hodnoty šikmosti a špičatosti mohou nabývat kladných i záporných hodnot. Hodnoty blízké 0 vypovídají o normálním rozložení, normalita rozložení je předpokládána v intervalu  $-1$  až  $1$ .

**Tabulka 1. Hodnoty indikátorů normality rozložení proměnných roční návštěvnosti bohoslužeb v ČR, SR a Polsku**

	ČR	SR	PI
Počet případů	1484	1116	1251
Průměr	5,5	23,7	29,8
Medián	0	6	30
Šikmost	3,6	1,3	0,7
Špičatost	15,2	0,7	0,3

Zdroj: ISSP 2008.

vykazuje normální rozložení. Mohla by vyvstat otázka, zda v těchto případech transformovanou proměnnou používat v pokročilejších statistických procedurách. V případě ISSP 2008 byly výběrové vzorky respondentů vybírány vícestupňovým náhodným výběrem z cílových populací a tyto vzorky obsahují více než 1000 případů pro každou zemi. Takto jsou splněny dva základní předpoklady i pro užití pokročilých statistických postupů [např. Agresti, Finlay 1999]. Použití transformované proměnné v případě ČR, kde se výrazně odlišuje od normálního rozložení, v případě Slovenska, kde se k normalitě blíží, i v případě Polska, kdy rozložení splňuje kritéria normality, bude testováno v následujících praktických příkladech.

#### *Příklad porovnání průměrů*

V Tabulce 2 jsou uvedeny hodnoty průměrných četností ročních návštěvností dospělých jedinců, těchto jedinců v jejich dětství a jejich rodičů. Opět jsou sledovány případy zemí, které byly zahrnuty již do Grafu 1. Průměrná roční návštěvnost současné dospělé populace je nejnižší v ČR, a to 5,5krát ročně. Oproti situaci v ČR je na Slovensku i v Polsku míra návštěvnosti více než čtyřnásobně vyšší. Hodnoty rozdílů mezi těmito zeměmi je třeba dále dopočítat a prokázat, zda lze rozdíly zjištěné ve výběrových vzorcích zobecňovat na celé populace sledovaných zemí (viz Tabulka 3). Následně budou analyzovány vztahy mezi návštěvností rodičů a jejich potomků ve formativním období – v průběhu intergenerační transmise (viz Tabulka 4) – a mezi tímto formativním obdobím a dospělostí jedince – v průběhu transmise intragenerační (viz Tabulka 5).

Vypočtené hodnoty rozdílů průměrných návštěvností bohoslužeb mezi sledovanými zeměmi jsou uvedeny v Tabulce 3. Současná dospělá populace ČR navštěvuje bohoslužby v průběhu jednoho roku oproti Slovensku v průměru o –18,1 návštěv méně a oproti Polsku je to ještě méně, tj. o –24,3. Nejnižší rozdíl je mezi Slovenskem a Polskem, o –6,2 roční návštěvnosti Slovenska oproti Polsku. V této

**Tabulka 2. Popis průměrné roční návštěvnosti bohoslužeb dospělé populace, této populace v dětství a jejich rodičů v ČR, SR a Polsku**

	Dospělost	Dětství	Matky	Otcové
ČR	5,5	10,7	13,3	9,3
SR	23,7	33,7	37,3	24,6
Pl	29,8	47,1	45	39,7

Zdroj: ISSP 2008.

**Tabulka 3. Porovnání rozdílů v průměrných ročních návštěvnostech bohoslužeb mezi ČR, SR a Polskem**

	Rozdíl průměrů	Sig	Interval spol. 95 %
ČR–SR	-18,1	< 0,00	-20,5–15,8
ČR–Pl	-24,3	< 0,00	-26,5–22,0
SR–Pl	-6,2	< 0,00	-8,6–3,7

Zdroj: ISSP 2008.

Poznámka: Anova – Posthoc Test Gabriel.

dimenzi tradiční religiozity je ČR oproti Slovensku i Polsku výrazně sekulárnější. Vyšší míra sekularity byla zjištěna i při použití kategoriální proměnné [srov. např. Hamplová 2000]. Jak již však bylo zobrazeno v Grafu 1 a následně zjištěno v Tabulce 1, kritéria normálního rozložení proměnných návštěvností, ze kterých byly počítány rozdíly v průměrných mírách návštěvnosti mezi zeměmi, splňuje pouze Polsko. Na druhou stranu je však patrné, že intervaly spolehlivosti (95%) těchto rozdílů nejsou ani v jedné zemi příliš široké. Pro příklad, -20,5 až -15,8 mezi ČR a SR. Proto je v návaznosti na hodnoty statistické významnosti možno říci, že se hodnoty rozdílů v průměrných ročních návštěvnostech pohybují s více než 95% pravděpodobností právě v těchto intervalech. Ve všech třech zemích byly vzorky vybírány pravděpodobnostním výběrem a mají více než 1000 případů, čímž jsou splněny základní podmínky pro užití číselných proměnných ve statistické analýze [Agresti, Finlay 1999]. Statistická signifikance těchto rozdílů ( $\text{sig} < 0,05$ ) byla vypočtena pomocí testu typu Gabriel, který je konstruován právě pro použití proměnných, jejichž rozložení se odlišuje od normálního rozložení [Field 2000]. Proto jsou zjištěné rozdíly roční návštěvnosti bohoslužeb, přesněji v jejich intervalech spolehlivosti 95 %, zobecňovány na celé cílové populace.

Pro srovnání možností kategoriálních a nových, transformovaných číselných proměnných. V Tabulce 3 byl například rozdíl mezi průměrnou roční návštěvností v ČR a v SR vyjádřen jedinou hodnotou, pro statistickou inferenci intervalem, přičemž nedošlo ke ztrátě části informace, která by nastala při redukci

kategorií kategoriálních proměnných do nižšího počtu. Je pravdou, že rozdíly v návštěvnostech lze zjišťovat i při použití původních proměnných o devíti kategoriích. Ovšem při porovnávání všech devíti kategorií by výsledek srovnání mezi dvěma zeměmi obsahoval také devět výsledných hodnot. Na druhou stranu, použití všech devíti kategorií, výběru některých z nich nebo jejich redukováného počtu umožňuje srovnání návštěvnosti mezi zeměmi v každé kategorii odděleně. Výběr konkrétních kategorií nebo kategorií pro sloučení však nemusí být bez problémů. Pro příklad lze použít procentuální hodnoty podílů z populace (viz Graf 1), které navštěvují bohoslužby v jednotlivých kategoriích. Na první pohled je patrné, že průběhy křivek nejsou souběžné. To znamená, že se mezi zeměmi liší i rozdíly v procentuálních podílech jedinců v jednotlivých kategoriích. Například v ČR navštěvuje bohoslužby 12krát ročně, tzn. 1krát měsíčně, 2 % jedinců z populace a na Slovensku 5 %. To je buďto o 3 % méně nebo 0,4krát méně. Ale v kategorii návštěvnosti 52krát ročně, tzn. každý týden, je tento rozdíl  $4 - 21 = -17$ , tzn. o 17 % méně nebo  $4/21 = 0,19$ , tzn. asi 0,2krát méně. Rozdíly v návštěvnosti se mezi zeměmi v každé kategorii liší. Které kategorie tedy pro srovnání návštěvnosti mezi zeměmi vybrat nebo které při redukcí do nižšího počtu kategorií vybrat pro sloučení? Při použití transformované číselné proměnné není nutno tuto otázku řešit. Výsledkem srovnání dvou průměrů roční návštěvnosti na bohoslužbách mezi dvěma zeměmi je vždy pouze jedna číselná hodnota, přičemž nehrozí riziko ztráty části informace, které by vzniklo při redukcí kategoriální proměnné například do dichotomického typu. Odpadá také nutnost konstrukce redukováné kategorizace a teoretického opodstatnění nově vzniklých kategorií. Na druhé straně jsou však při použití číselné proměnné analytickými výsledky pouze průměry návštěvnosti za celé populace nebo jejich části, bez možnosti podrobnějšího srovnání rozdílů v mírách návštěvnosti v jednotlivých kategoriích.

#### *Příklad intergenerační transmise*

V Tabulce 2 byly uvedeny průměrné hodnoty ročních návštěvností v případech dospělých jedinců. V dalších sloupcích této tabulky jsou také uváděny hodnoty ročních návštěvností ve formativním období těchto jedinců i v případech jejich rodičů. Na základě četností návštěvnosti rodičů a jejich potomků ve formativním období<sup>3</sup> bude zjišťována míra přenosu návštěvnosti mezi těmito rodiči a potomky, tzn. v průběhu intergenerační transmise.

V Tabulce 4a jsou uvedeny odhady parametrů modelů vícenásobné lineární regrese, zobrazujících vztah mezi roční návštěvností rodičů a jejich potomků v případech ČR, Slovenska a Polska. Výpovědní hodnotu tohoto modelu lze odhadovat na základě hodnoty R, tzn. hodnot Pearsonových korelací mezi prediktory a vysvětlovanou proměnnou. Tato hodnota je nejvyšší v případě ČR ( $R = 0,823$ )

<sup>3</sup> V dotazníkové otázce vymezeno věkem potomka 11–12 let.

**Tabulka 4a. Odhady parametrů lineární regrese: Vliv roční návštěvnosti rodičů na návštěvnost jejich potomků ve formativním období v ČR, SR a Polsku**

	ČR	SR	PI
Pearsonovo R	0,823	0,769	0,596
Návštěvnost matky (B)	0,6	0,44	0,41
Návštěvnost otce (B)	0,26	0,39	0,18
Konstanta	1,1	7,7	21,4

Zdroj: ISSP 2008.

Poznámka: N: ČR = 1209, SR = 1037, PI = 1120. Regresní koeficienty sig < 0,00.

F: ČR = 1268,9, SR = 748,5, PI = 304,9, sig < 0,00.

Levenův test: ČR F = 13,2, SR F = 4,4, PI F = 6,2, sig = 0,000.

a nejnižší v případě Polska ( $R = 0,596$ ). Tyto hodnoty korelací by bylo možno vypočítat i z původních ordinálních proměnných, ale model lineární regrese s transformovanými číselnými proměnnými umožňuje sledovat vztahy mezi návštěvností rodičů a jejich dětí mnohem podrobněji. Například, v ČR s nárůstem návštěvnosti matky o 1krát ročně narůstá primárně socializovaná návštěvnost jejího dítěte přibližně o 0,6krát ročně. Oproti tomu nárůst návštěvnosti otce také o 1krát ročně zvyšuje návštěvnost jeho potomka oproti matce méně, tj. o 0,26krát ročně. Podobně lze sledovat vliv návštěvnosti rodičů i u obou dalších zemí. Z odhadnutých parametrů v Tabulce 4 lze vyčíst také další informace. Za prvé, všechny hodnoty regresních koeficientů jsou menší než 1. To znamená, že při přímém přenosu návštěvnosti v linii rodič–dítě dochází k neúplné reprodukci. Za druhé, přímý vliv návštěvnosti matky je významnější oproti vlivu návštěvnosti otce. A nakonec za třetí, do rovnice vícenásobné lineární regrese  $y = a + b_1x_1 + b_2x_2$ , vstupuje kromě již zmíněných regresních koeficientů návštěvnosti matky ( $b_1$ ) a otce ( $b_2$ ) také hodnota konstanty ( $a$ ). Je zřetelné, že se sledované země v těchto hodnotách konstant výrazně liší. Jak by však bylo možno tyto hodnoty konstant<sup>4</sup> interpretovat? Pokud by například v ČR matka i otec nenavštěvovali bohoslužby vůbec, tj. na hodnotě 0 roční účasti, pak lze zapsat  $1,1 + 0,6 \times 0 + 0,26 \times 0 = 1,1$ . S touto četností roční návštěvnosti byli v průměru primárně socializováni potomci rodičů, kteří nenavštěvovali bohoslužby vůbec.<sup>5</sup> To znamená, že návštěvy těchto dětí na bohoslužbách v kostelích musely ovlivňovat faktory mimo nukleární rodinu, mimo rodiče, například členové širší rodiny, náboženská výuka na školách nebo aktivity církví. A právě v síle vlivu těchto faktorů ležících mimo nukleární rodinu (hodnotách konstant) se ČR výrazně liší oproti Slovensku a ješ-

<sup>4</sup> Konstanta je hodnotou, ve které přímka lineární regrese protíná osu y. Proto může být odhad této hodnoty zkeslen velikostí rozptylu.

<sup>5</sup> Do rovnice lze zapsat jakoukoliv četnost účasti rodičů v rozmezí 0 až 104krát ročně.



**Tabulka 4b. Hodnoty pravděpodobností: Vliv pravidelné nebo nepravidelné návštěvnosti rodičů na pravidelnou návštěvnost potomků ve formativním období v ČR, SR a Polsku**

	ČR	SR	PI
Oba rodiče bez návštěvnosti	3 %	13 %	54 %
S návštěvností jen matka	62 %	79 %	91 %
Oba rodiče s návštěvností	89 %	94 %	99 %

Zdroj: ISSP 2008.

Poznámka: N: ČR = 1205, SR = 1042, PI = 1116.

Likelihood Ratio: ČR = 845,8, SR = 650,6, PI = 174,1, sig < 0,00.

tě výrazněji oproti Polsku. V ČR navyšovaly faktory mimo nukleární rodinu návštěvnost dítěte v průměru o 1,1 návštěv ročně, na Slovensku o 7,7 a v Polsku o 21,4. Vyšší pokles návštěvnosti v ČR v průběhu intergenerační transmise je tak v nezanedbatelné míře dán také absencí nebo malým vlivem primárně nábožensky socializačních faktorů mimo nukleární rodinu [viz např. Paleček 2015].

Jak je patrné z Tabulky 4a, lze v modelu lineární regrese odhadovat vliv návštěvnosti rodičů a dalších faktorů mimo nukleární rodinu na návštěvnost dětí prostřednictvím dvou hodnot regresních koeficientů a jedné hodnoty konstanty. Dále nelze opomíjet význam hodnoty  $R^2$ , která signalizuje přesnost predikce modelu, založenou na velikosti rozptylů, a po vynásobení stem uvádí, jaká část rozptylu vysvětlované proměnné je dána vlivem prediktorů. V ČR lze návštěvnosti rodičů vysvětlovat přibližně 68 % návštěvnosti jejich dětí, na Slovensku 59 % a v Polsku 35 %. Zbytek do 100 % je dán vlivem faktorů mimo nukleární rodinu, který byl zjištěn již při interpretaci hodnot konstant. Míra rozptylu okolo regresní přímky, a tedy i vhodnost regresního modelu je zjišťována také prostřednictvím analýzy rozptylu, hodnoty  $F$  a její statistické signifikance. V případech všech tří zemí jsou hodnoty  $F$  statisticky významné, sig = 0,000 (viz Tabulka 4a). Jedním z dalších požadavků, který úzce souvisí s přesností odhadnutých parametrů lineární regrese, je homoskedasticita. To znamená, že velikost rozptylu proměnné  $y$  by se neměla výrazněji lišit mezi jednotlivými hodnotami proměnné  $x$ . Tento podmíněný rozptyl je zjišťován například prostřednictvím Levenova testu. Na základě výsledků Levenova testu a jeho signifikance (sig = 0,000) jsou však modely v Tabulce 4 zatíženy významnou měrou heteroskedasticity, která zkresluje hodnoty odhadnutých parametrů. V tomto okamžiku vyvstává otázka, jakým způsobem prakticky ověřit výpovědní hodnotu modelu v Tabulce 4a nebo jakými jinými způsoby lze sledovat vliv návštěvnosti v rodině původu na primárně socializovanou návštěvnost dětí.

Pro srovnání s výsledky lineární regrese jsou v Tabulce 4b uvedeny hodnoty pravděpodobností na primárně socializovanou pravidelnou návštěvnost dětí

v závislosti na pravidelné nebo nepravidelné návštěvnosti jejich rodičů. Proměnné návštěvnost matky, otce i dítěte jsou dichotomické, v kategoriích návštěvnost pravidelná (alespoň 1krát měsíčně) a nepravidelná (méně než 1krát měsíčně). Pokud se ani jeden z rodičů pravidelně neúčastnil bohoslužeb, pak byli v těchto rodinách děti primárně socializovány s pravidelnou návštěvností v ČR s 3% pravděpodobností a tato hodnota pravděpodobnosti narůstá na Slovensku na 13 % a v Polsku na 54 %. Primárně socializovaná pravidelná návštěvnost těchto dětí musela být ovlivněna faktory ležícími mimo nukleární rodinu. Oproti ČR měly tyto faktory výraznější vliv na Slovensku a ještě vyšší v Polsku, jak již bylo zjištěno prostřednictvím hodnot konstant v modelu lineární regrese v Tabulce 4a. Při použití dichotomických proměnných však může být toto zjištění zatíženo určitou mírou zkreslení. Pro příklad. Pokud dítě navštěvovalo bohoslužby každý měsíc, ale jednou s matkou a jindy s otcem, pak toto dítě spadá do průsečíku kategorií „oba rodiče bez pravidelné návštěvnosti“ a „dítě s pravidelnou návštěvností“. Přitom se na socializaci tohoto dítěte s pravidelnou návštěvností nemusely podílet žádné faktory mimo nukleární rodinu. Riziko tohoto zkreslení by sice klesalo s nárůstem počtu kategorií, ale nárůst kategorií by vytvářel další problémy. Jednak, kategorizace, která se jeví nejvhodnější pro jednu zemi, nemusí být optimální pro jiné země (viz Graf 1). Za druhé, s počtem vysvětlujících proměnných a s počty kategorií vysvětlujících i vysvětlované proměnné klesá i zastoupení absolutních četností v některých kategoriích a tím i v jednotlivých buňkách tabulky. Nedostatečné zastoupení absolutních četností pak zvyšuje rizika zkreslení analytických výsledků.

V rodinách, kde se bohoslužeb pravidelně účastnili oba rodiče, byly děti socializovány s pravidelnou návštěvností v ČR s 89% pravděpodobností a na Slovensku i v Polsku byla tato pravděpodobnost mírně vyšší (94% a 99%). To znamená, že v těchto rodinách, kde se rodiče shodovali v pravidelné návštěvnosti, docházelo i k vysoké míře reprodukce pravidelné návštěvnosti z rodičů na potomky. Pokud se bohoslužeb pravidelně účastnila pouze matka, pak byly děti těchto matek socializovány také s pravidelnou návštěvností s nižší pravděpodobností, v ČR 62%, na Slovensku 79% a v Polsku 91%. Pro srovnání, rozdílný vliv matek a otců byl v Tabulce 4a zjišťován prostřednictvím regresních koeficientů a vliv faktorů mimo nukleární rodinu prostřednictvím hodnot konstant.

Zatímco v modelu lineární regrese v Tabulce 4a byla sledována reprodukce návštěvnosti v celé škále, tak v Tabulce 4b byla zjišťována pouze reprodukce pravidelné návštěvnosti při použití dichotomických proměnných s kategoriemi pravidelná a nepravidelná návštěvnost. Pokud je cílem studie zkoumání míry reprodukce pravidelné návštěvnosti, pak má takováto dichotomizace své opodstatnění. Pokud je však studie zaměřena na celkovou reprodukci návštěvnosti v celé její dostupné škále, pak každá redukce kategorií znamená ztrátu určité části informací a tím i zkreslení analytických výsledků. S nárůstem počtu kategorií je sice ztráta původních informací nižší, ale s počtem kategorií klesá zastoupení absolutních četností v jednotlivých kategoriích i buňkách tabulky a tím opět na-

růstá riziko zkreslení výsledných hodnot. Oproti tomu je zkreslení analytických výsledků v modelu vícenásobné lineární regrese v Tabulce 4a dáno velikostí celkového rozptylu i rozdíly hodnot podmíněných rozptylů proměnné  $y$  v závislosti na proměnných  $x$ .

#### *Příklad intragenerační transmise*

V předchozí části bylo představeno možné použití transformovaných číselných proměnných na příkladu míry přenosu návštěvnosti bohoslužeb mezi generacemi rodičů a jejich potomků, tzn. v průběhu intergenerační transmise. Po skončení formativního období si člověk může svoji primárně socializovanou návštěvnost udržet během dospělosti nebo se četnost jeho účasti na bohoslužbách může také změnit. V těchto případech hovoříme o reprodukci nebo mobilitě v průběhu intragenerační transmise. Jak je patrné z hodnot v Tabulce 2, průměrná roční návštěvnost bohoslužeb dospělých jedinců se liší oproti jejich návštěvnosti v dětství. Například v ČR došlo k intrageneračnímu poklesu v průměru o  $-5,2$  návštěv ročně (vypočítáno  $5,5 - 10,7 = -5,2$ ). Na Slovensku činí tento intragenerační rozdíl  $-10$  a v Polsku  $-17,3$  návštěv ročně.

Rozdíl v četnosti návštěv bohoslužeb mezi dospělostí a dětstvím lze však vypočítat také pro každý jednotlivý případ, pro každého respondenta zvlášť a tuto hodnotu uložit do datového souboru jako novou rozdílovou proměnnou. Rozdílová proměnná obsahuje v číselné hodnotě informaci, o kolik účastí ročně v průměru narostla nebo poklesla roční návštěvnost během intragenerační transmise. Tato nová rozdílová proměnná je dále použita jako vysvětlovaná proměnná v příkladu v Tabulce 5. Protože je možno předpokládat, že hodnota této rozdílové proměnné (velikostí poklesu nebo nárůstu) bude ovlivňována nejen individuálními charakteristikami jedinců, ale i makrocharakteristikami jednotlivých zemí, bude použit víceúrovňový lineární model [viz Soukup 2006]. Do tohoto modelu v Tabulce 5 bylo zahrnuto všech 25 evropských zemí<sup>6</sup> zúčastněných ve vlně ISSP 2008. Použití víceúrovňového modelu znamená, že výsledné hodnoty bude možno zobecňovat na dospělou populaci Evropy, přičemž model respektuje rozdíly mezi jednotlivými zeměmi.

Je vhodné nejprve zjistit hodnotu koeficientu vnitrotřídní korelace [Soukup 2006] ( $ICC = 0,267$ ), která po vynásobení stem říká, že varianci použité závislé rozdílové proměnné lze přibližně z 27% vysvětlovat rozdíly mezi zúčastněnými zeměmi. Z modelu dále vyplývá, že v průběhu intragenerační transmise poklesla v průměru více návštěvnost mužů oproti ženám, a to o  $-1,5$  návštěv ročně. Nižší religiozita mužů oproti ženám je obecným faktem [Hamplová 2011] a jejich nižší úspěšnost intrageneračního přenosu návštěvnosti může tento rozdíl z části vysvětlovat. Dále, oproti celé populaci u jedinců s římsko-katolickým vyznáním

<sup>6</sup> Bylo vypuštěno Turecko, protože se od evropských zemí kulturně liší.

**Tabulka 5. Odhady parametrů hierarchického lineárního modelu: Rozdíly v návštěvnosti mezi dospělostí a formativním obdobím (25 evropských zemí)**

	Intragenerační transmise	Sig
Pohlaví muž	-1,5	< 0,00
Římsko-katolické	-6,1	< 0,00
Protestantské	1,3	< 0,00
Křesťanské ortodoxní	-1,6	0,08
Věk	-0,05	< 0,00
HDI v % oproti Norsku	-0,7	0,02

Zdroj: ISSP 2008.

Poznámka: N = 36 954, BIC = 313 335, ICC = 0,267.

klesá úspěšnost intrageneračního přenosu o -6,1 návštěv a u jedinců s křesťanským ortodoxním vyznáním o -1,6 návštěv v průběhu roku. Naopak, v případech protestantských denominací došlo k intrageneračnímu nárůstu v průměru o 1,3 návštěvy ročně. Pokles roční návštěvnosti je také v pozitivním vztahu s věkem, směrem ke starším ročníkům. S nárůstem věku<sup>7</sup> o jeden rok, je zjištěn pokles o -0,05 návštěv. Posledním prediktorem analytického příkladu v Tabulce 5 je makrocharakteristika, hodnota indexu lidského rozvoje (HDI) pro každou zemi. Tato hodnota je počítána jako procentuální podíl HDI oproti Norsku,<sup>8</sup> jehož hodnota HDI je mezi evropskými zeměmi nejvyšší. Pokud se HDI země, ve které jedinec žije, přiblíží o 1 % k Norsku, pak naroste jeho intragenerační pokles návštěvnosti v průměru o -0,7 návštěv v jednom roce. To znamená, že spolu s nárůstem HDI klesá úspěšnost intrageneračního přenosu návštěvnosti bohoslužeb. Tento negativní vztah je v souladu se zjištěními předchozích studií, ve kterých byl pokles zájmu o náboženství vysvětlován nárůstem životních jistot [Norris, Inglehard 2004].

<sup>7</sup> Intragenerační transmise je ukončena teprve úmrtím jedince, ale v datech je četnost návštěv zjišťována v době sběru dat. To znamená, že v průběhu dalšího života může ještě docházet ke změnám, i vícenásobným, že starší ročníky měly na tyto změny více času a že mladší ročníky budou mít naopak času více.

<sup>8</sup> HDI Norska je na hodnotě 0,943, což je bráno za 100 %. HDI každé další země je pak vypočteno jako procentní podíl HDI Norska. Např. hodnota HDI ČR je 0,865, což je 91,7 % HDI Norska.

## **Závěr a diskuze**

V mnoha výběrových šetřeních jsou používány kategoriální proměnné s velkým počtem kategorií, které poskytují informace o četnosti návštěv bohoslužeb. Jednotlivé kategorie podrobně vypovídají nejen o míře návštěvnosti jedince, ale současně poskytují i časové údaje, tzn. v jakých časových intervalech k pravidelným návštěvám dochází nebo docházelo. Tyto skutečnosti umožňují v závislosti na cíli studie volit mezi třemi metodologickými postupy. 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. Lze například zodpovědět otázku, jak je rozložena návštěvnost v celé populaci nebo v jejích částech, a tyto údaje graficky znázornit. Při limitovaném počtu případů v datovém souboru, což se týká všech výběrových šetření, však s nárůstem třídících kritérií zpravidla dochází k nedostatečnému zastoupení absolutních četností v některých buňkách vícerozměrných tabulek. Tím je také omezována možnost použití většího počtu třídících (vysvětlujících) faktorů nebo faktorů s mnoha kategoriemi. Limitovaný počet případů v datovém souboru řeší druhý metodologický přístup, založený na redukci počtu kategorií. Některé kategorie proměnné jsou slučovány, tím dochází k nárůstu absolutních četností v nově definovaných sloučených kategoriích a následně i možnosti použití více třídících kritérií. Počet nových kategorií je odvislý od cíle studie. Tato varianta je vhodná, pokud je cílem práce zjištění a vysvětlení určité definované pravidelnosti v návštěvnosti na bohoslužbách, například alespoň 1krát měsíčně nebo týdně atd., nebo i v případě, že je práce zaměřena na úspěšnost přenosu této pravidelnosti v průběhu inter a intragenerační transmise. Na druhé straně však jakákoliv redukce znamená také ztrátu části informace, která byla obsažena v původním počtu kategorií. Pokud je třeba pracovat s celou informací, tzn. s neredukovaným počtem kategorií proměnné návštěvnost bohoslužeb, a přitom použít potřebný počet třídících (vysvětlujících) faktorů, pak může být přínosný třetí metodologický přístup, transformace kategoriální proměnné do proměnné číselné. Popis transformace, zdůvodnění volby této varianty a nakonec i testování nově vzniklé číselné proměnné návštěvnost bohoslužeb na příkladech přenosu míry návštěvnosti v průběhu inter a intragenerační transmise se stalo tématem této stati.

I transformovaná číselná proměnná má své přínosy a svá omezení. Na jedné straně je výsledkem porovnání průměrných měr návštěvnosti mezi dvěma subjekty, například zeměmi, pouze jedna číselná hodnota, aniž by byla ztracena část původní informace způsobená redukcí počtu kategorií. Tato hodnota přináší informaci, o kolik návštěv více nebo méně se v průměru první subjekt v průběhu jednoho roku účastní bohoslužeb oproti subjektu druhému. Podobně lze bez ztráty části informace i prostřednictvím malého počtu výsledných hodnot zjišťovat, v jaké míře se například přenáší návštěvnost bohoslužeb v průběhu inter a intragenerační transmise.

První dva zde uvedené metodologické postupy již byly použity v mnoha empirických studiích a na několik z nich bylo v této stati odkazováno. Třetí meto-

dologický postup, transformace kategoriální proměnné návštěvnost bohoslužeb do proměnné číselné, může být určitým doplněním těchto možných, již běžně užívaných variant metodologických přístupů. Tak jako první dva uvedené postupy i transformace proměnné do proměnné číselné má své výhody i svá omezení. Na rozdíl od kategoriálních proměnných je v případě transformovaných číselných proměnných pracováno s průměry za celou populaci nebo její části. Přínosem může být malý počet výsledných hodnot, ale na druhé straně nelze detailněji pracovat s informacemi, které jsou specifické pro jednotlivé kategorie. Je tedy na výzkumníkovi, na zaměření jeho studie, jaký metodologický postup zvolí. Cílem této stati bylo pouze představení jednoho z dalších možných řešení.

ANTONÍN PALEČEK je absolventem magisterského programu sociologie na Fakultě sociálních studií Masarykovy univerzity. V současnosti pokračuje v doktorském studiu sociologie tamtéž.

## Literatura

- Agresti, A., B. Finlay. 1999. *Statistical Methods for the Social Sciences*. 3rd ed. Prentice Hall.
- De Vaus, D. 2002. *Analyzing Social Science Data*. London: SAGE Publication.
- Disman, M. 1993. *Jak se vyrábí sociologická znalost*. Praha: Karolinum.
- Field, A. 2000. *Discovering Statistics Using SPSS for Windows: Advanced Techniques for the Beginner*. London: SAGE Publication.
- Filipi, P. 2001. *Křesťanstvo. Historie, statistika, charakteristika křesťanských církví*. Brno: Centrum pro studium demokracie a kultury.
- Fox, J. 2008. *Applied Regression Analysis and Generalized linear Models*. Los Angeles: SAGE Publication.
- Geurts, T., A. R. Poortman, P. A. Dykstra. 2009. „Contact Between Grandchildren and Their Grandparents in Early Adulthood.“ *Journal of Family Issues* 30 (12): 1698–1713, <https://doi.org/10.1177/0192513X09336340>.
- Glock, Ch. 1962. „On the Study of Religious Commitment.“ *Religious Education* 57 (4): 98–110, <https://doi.org/10.1080/003440862057S407>.
- Hamplová, D. 2000. *Náboženství a nadpřirozeno ve společnosti. Mezinárodní srovnání na základě empirického výzkumu ISSP*. Sociologický ústav AV ČR.
- Hamplová, D. 2005. „Základní principy víceúrovňových modelů.“ *SDA info* 7 (2): 1–2.
- Hamplová, D. 2011. „Náboženství a pohlaví: Proč jsou ženy religioznější než muži?“ *Sociologický časopis / Czech Sociological Review* 47 (2): 297–321.
- Hamplová, D., B. Řeháková. 2009. *Česká religiozita na počátku 3. tisíciletí. Výsledky mezinárodního programu sociálního výzkumu ISSP 2008 – Náboženství*. Praha: Sociologický ústav AV ČR, v.v.i.
- Hendl, J. 2004. *Přehled statistických metod zpracování dat: analýza a metaanalýza dat*. Praha: Portál.
- Kreft, I., J. de Leeuw. 1998. *Introducing Multilevel Modeling*. London: SAGE Publications.
- Lambert, Y. 2004. „A Turning Point in Religious Evolution in Europe.“ *Journal of Contemporary Religion* 19 (1): 29–45, <https://doi.org/10.1080/1353790032000165104>.

- Lužný, D. 1999. *Náboženství a moderní společnost*. Brno: Masarykova univerzita.
- Myers, S. M. 1996. „An Interactive Model of Religiosity Inheritance: The Importance of Family Context.“ *American Sociological Review* 61 (5): 858–866, <https://doi.org/10.2307/2096457>.
- Need, A., G. Evans. 2001. „Analysing Patterns of Religious Participation in Post-communist Eastern Europe.“ *British Journal of Sociology* 52 (2): 229–248, <https://doi.org/10.1080/00071310120044962>.
- Norris, P., R. Inglehart. 2004. *Sacred and Secular. Religion and Politics Worldwide*. Cambridge University Press, <https://doi.org/10.1017/CBO9780511791017.002>.
- Nešpor, Z. 2010. *Příliš slabí ve víře. Česká ne/religiozita v evropském kontextu*. Praha: Kalich.
- Nešpor, Z., D. Václavík a kol. 2008. *Příručka sociologie náboženství*. Praha: Sociologické nakladatelství (SLON).
- Norusis, M. J. 1998. *SPSS 8.0: Guide to Data Analysis*. Upper Saddle River: Prentice Hall.
- Paleček, A. 2011. *Reprodukce religiozity v ČR druhé poloviny 20. století*. Brno: Masarykova Univerzita, archiv závěrečných prací.
- Paleček, A. 2015. „Sekularizace v pohledu inter a intragenerační transmise: ČR ve srovnání post-komunistických zemí střední Evropy.“ *Naše společnost* 13 (2): 13–26, <https://doi.org/10.13060/1214438X.2015.2.13.231>.
- Paleček, A., R. Vido. 2014. „Náboženské vyznání v České republice z perspektivy inter a intragenerační transmise.“ *Naše společnost* 12 (2): 24–35, <https://doi.org/10.13060/1214438X.2014.2.12.124>.
- Polack, D. 2008. „Religious Change in Europe: Theoretical Considerations and Empirical Findings.“ *Social Compass* 55 (2): 168–186, <https://doi.org/10.1177/0037768607089737>.
- Ruiter, S., F. van Tubergen. 2009. „Religious Attendance in Cross-National Perspective: A Multilevel Analysis of 60 Countries.“ *American Journal of Sociology* 115 (3): 863–895, <https://doi.org/10.1086/603536>.
- Řeháková, B. 2000. „Nebojte se logistické regrese.“ *Sociologický časopis* 36 (4): 475–492.
- Simonová, N. 2007. „Vzdělanostní reprodukce v České republice od roku 1916 do současnosti: Mobilitní pohled.“ Pp. 27–42 in P. Mareš a kol. *Sociální Reprodukce a integrace: Ideály a meze*. Brno: Masarykova Univerzita.
- Soukup, P. 2006. „Proč užívat hierarchické lineární modely?“ *Sociologický časopis / Czech Sociological Review* 42 (5): 987–1012.
- Váně, J., V. Hásová. 2014. „Jak se v České republice vlastně měří náboženství?“ *Acta fakulty filozofické Západočeské univerzity v Plzni* 3/2014.
- Vido, R. 2011. *Konec velkého vyprávění? Sekularizace v sociologické perspektivě*. Brno: Centrum pro studium demokracie a kultury.

## Data

*International Social Survey Programme 2008 – Religion*. 2008. Český sociálněvědní datový archiv, Sociologický ústav AV ČR, v.v.i.