

## **Kohortní analýza jako alternativa panelového výzkumu (Tomáš Katrňák, 20/9/2009)**

Zkoumáme-li sociální, kulturní či politické změny, nebo analyzujeme-li vývoj vybraného sociálního fenoménu, musíme použít data z longitudinálních výzkumů. Jedním z nich je panelový výzkum. V panelu jsou jedni a ti samí jedinci v určitém časovém intervalu opakovaně dotazováni na jednu a tu samou věc. Rozlišením respondentů podle věku, studujeme v panelu fenomény podle roku narození respondentů a jejich stárnutí. Panelový výzkum tím rozšiřujeme o podstatnou dimenzi. V takovém případě je však nezbytné vyrovnat se nejen s efekty vysvětlujících proměnných (a jejich změnou v čase) na vysvětlovanou (závisle) proměnnou, ale také s efekty kohort (vymezených rokem narození respondentů), věku (biologického stárnutí respondentů) a doby (roky sběru dat), které jdou ruku v ruce. Takový typ výzkumu označujeme intrakohorní panelový výzkum (Ruspini, 2002).

Modely pro rozlišení efektů věku, doby a kohort narozených v odlišných obdobích nejsou v intrakohortním panelovém výzkumu zatím zdokumentované. Jako badatelské téma jsou výzvou pro následující léta (srov. Yang, 2007; 2008a). Jediný způsob, jak odlišit v panelových datech efekt věku od efektů doby a kohort narozených v jiných letech, zůstává prosté procentuální srovnání. Například věkovou skupinu 25-27 let a 15-17 let srovnáme po deseti letech trvání panelu (kdy skupina 15-17 let dosáhne věku 25-27 let) a rozdíly mezi nimi interpretujeme jako efekt věku. Nebo změny ve věkové skupině 15-17 let po deseti letech trvání panelu srovnáme se změnami v té samé věkové skupině v populaci a pokud nejsou tyto změny odlišné, můžeme je interpretovat jako efekt doby (čas deseti let trvání panelu). Či rozdíly mezi věkovou skupinou 15-17 let a tou samou věkovou skupinou po deseti letech trvání panelu lze interpretovat jako efekt kohorty narozené před 15 až 17 lety (po deseti let trvání panelu narozené před 25-27 lety) na zkoumanou problematiku. Všechny tyto procentuální komparace jsou ovšem problematické, jelikož procentuální změna v datech je způsobena jak změnou ve zkoumaném fenoménu, tak výběrovou chybou nevyhnutelnou pro vzorek respondentů z populace.

Máme-li k dispozici výběrová šetření z jednotlivých let, která obsahují informace o zkoumané proměnné, můžeme použít jako alternativu k intrakohortnímu panelovému výzkumu *kohortní analýzu*. Počátky této analýzy spadají do první čtvrtiny 20. století (v demografii analýza porodnosti a úmrtnosti; v epidemiologii analýza nemocnosti a specifické úmrtnosti) (Mason, Wolfinger, 2002). V polovině 60. let. 20. století Norman Ryder (1965) publikoval esej *The Cohort as a Concept in the Study of Social Change* a zavedl koncept kohorta narozených (*birth cohort*) jako explanační nástroj sociální změny do sociologie. Od té

doby byla napsána řada metodologických i věcně zaměřených studií, rozpracovávajících nebo využívajících kohortní analýzu v sociologických výzkumech, které ukázaly výhody, jakožto i nevýhody tohoto přístupu k datům (srov. např. Mason, Mason, Winsborough, Poole, 1973; Hobcraft, Menken, Preston, 1982; Fienberg, Mason, 1985; Wilmonth, 1990; Glenn, 1994; 1999; Alvin, McCammon, 1999; Yang, 2008b).

Kohortní analýza usiluje o vysvětlení zkoumané problematiky pomocí rozdílů mezi kohortami narozenými v jiné době (čas, od kdy jsou jedinci součástí společnosti), při zohlednění rozdílů mezi věkovými skupinami (čas, jak dlouho jsou součástí společnosti) a dobou (roky, v nichž je zkoumaná problematika analyzována) (Mason, Wolfinger, 2002). Jedná se o analýzu sociální, politické, kulturní nebo ekonomické změny z hlediska kohort, které se narodily v jiných letech, mají tudíž jinou historickou zkušenost a jsou nositeli jiných sociálních fenoménů (Glenn, 1977; 2005). Používá se pro ni také označení analýza věku, období a kohort (*Age-Period-Cohort analysis*, zkráceně *APC analysis* nebo také *APC models*). Nejedná se o statistickou techniku, ale o specifický přístup k datům, která – aby byla kohortní analýza použitelná – musejí obsahovat alespoň dva údaje ze tří: informaci o letech, z nichž data pocházejí, a informaci o věku jedinců zkoumané populace.<sup>1</sup>

Na rozdíl od (intrakohortního) panelového šetření, v kohortní analýze nemáme údaje o stejných jedincích v čase, ale pracujeme s proměnnými (z populace či vzorků) z jednotlivých let (není tedy pravděpodobné, že jsou dotazováni jedni a ti samí jedinci). Kohortní analýza není o jedincích (a změnách u nich), jak je tomu v panelovém výzkumu, ale o sociálních celcích, které jedinci reprezentují – o věkových skupinách, obdobích nebo kohortách narozených v jiných obdobích. Například víme-li, že ve věkové skupině 25-27 let preferovalo ODS v roce 1995 30 % respondentů a v roce 2005 byla preference ODS ve věkové skupině 35-37 let rovněž 30 %, kohortní analýza nám řekne, že nedošlo ke změně preferencí během deseti let u této kohorty narozené v letech 1968-1970 (jedná se o informaci o *čisté* změně/stabilitě v populaci). To ovšem neznamená, že reprezentanti této kohorty nezměnili své politické preference. Jak je změnili, kolik z nich se odvrátilo od ODS, kolik z nich naopak začalo ODS preferovat, zjistíme panelovým výzkumem, který je schopen identifikovat individuální změnu (konceptualizovanou oproti populační změně v panelovém výzkumu jako *hrubá* změna).

V sociální vědě jsou dnes čím dál dostupnější reprezentativní data z jednotlivých let za jednotlivé země. Sociální vědci a vědkyně v současnosti již nejsou odkázáni na práci s daty

---

<sup>1</sup> Oba tyto údaje jsou součástí každého výběrového šetření, v němž je výzkumnou jednotkou jedinec.

z jednoho časového okamžiku, ale mohou z hlediska zkoumaných proměnných odpovídající data spojovat a analyzovat v nich efekty věku, doby a kohort na zkoumanou problematiku. Ve výběrových (surveyových) datech za jednotlivé roky máme informace o narozených kohortách, pokud jsou přítomné informace o věku respondentů (kohorta narozených rovná se rok sběru dat mínus věk). Modely pro odlišení efektu věku, doby a kohort byly v průběhu posledních 30. let v sociální statistice zdokumentované a jsou poměrně snadno aplikovatelné. Zabýváme-li se tedy dnes vývojem vybraného sociálního fenoménu v čase, je kohortní analýza adekvátní alternativou k panelovému šetření.

V následujícím textu nejprve vymezím základní koncepty kohortní analýzy. Ukáži, jakou roli hrají kohorty narozené v jiné době při sociologické explanaci sociálních procesů a představím kohortní výměnu jako podstatu sociální změny. Ve zbylé části textu se budu zabývat metodologií kohortní analýzy. Zaměřím se na standardní kohortní tabulku a budu se zabývat problémem, který komplikuje identifikaci efektů věku, období a narozených kohort v datech (identifikační problém). Způsoby řešení tohoto problému budu prezentovat na datech (příkladech) z české společnosti, jimiž jsou v prvním případě data populační (tištěné v pravidelných přehledech *Pohyb obyvatelstva* Českým statistickým úřadem) a ve druhém případě data výběrová (oba příklady se týkají živě narozených dětí v České republice). Nakonec obě řešení, včetně výsledků kohortní analýzy, srovnám a budu je diskutovat. Jelikož kohortní analýza není zatím u nás příliš rozšířená, cílem tohoto textu je nejen představit tuto techniku jako nástroj analýzy sociální změny k panelovému šetření, ale také přispět k její popularizaci v českém sociálněvědném bádání.

### **Věk, období a kohorty narozených jedinců**

V kohortní analýze zohledňujeme při explanaci sociálního fenoménu věk, období a příslušnost ke kohortě narozené v určitém roce. Jedná se o tři věcně odlišitelné koncepty, které jsou v realitě (u každého člověka) provázané. To je ústřední problém při identifikaci jejich efektů na vysvětlovanou proměnnou. Metodologickou otázkou je, jak odlišit kohortní efekt od efektů věku a doby, a který z těchto efektů nejsilněji poznamenává změny ve zkoumané problematice.

*Věk (age)* indikuje nejen biologický proces stárnutí člověka, ale je spojen také s životní etapou. Označuje fázi života, v níž se člověk zrovna nachází. Jedná se o individuální proměnnou, typickou pro jednoho konkrétního člověka. Je to indikátor fyziologických změn člověka, jeho životních zkušeností, sociálních rolí a sociálního postavení ve společnosti. Lidé stejného věku tvoří věkovou skupinu (*age group*).

*Období (period)* je čas, v němž žijeme. Efekt doby působí na všechny společenské skupiny. Pokud společnost prochází ekonomickými, sociálními, politickými nebo kulturními změnami (války, ekonomické krize, změna politického režimu, pandemie), nejsou tyto změny zaměřené na určité věkové, sociální nebo ekonomické segmenty společnosti, ale mění společnost jako celek.

Jedinci patřící do *kohorty narozené* ve stejném roce (*birth cohort*) prochází společenským systémem dohromady (jejich biologické stáří je totožné). Osvojují si typy sociálního jednání ve stejné době. Jsou vystaveni společenským změnám, historickým událostem, sociálním tlakům a incentivám ve stejném věku (Glenn, 1977; 2005). Tím se odlišují od reprezentantů kohort narozených v jiných letech. Lidi narozené ve stejném roce obvykle spojuje specifická historická zkušenost. Mají podobný způsob myšlení a obvykle i jednání. Zažili stejnou politickou, ekonomickou nebo kulturní událost ve stejném věku. Jednu kohortu taková událost ovlivní, jiná kohorta zůstane vůči ní indiferentní, protože ji zažila v jiném životním období a vnímá ji odlišně.

Původní termín kohorta (*cohort*) odkazuje k římské vojenské jednotce. Ve slovnících bývá definována jako skupina vojáků či bojovníků (Glenn, 1977). Pokud dnes hovoříme v sociální vědě o kohortě, máme většinou na mysli kohortu jedinců narozených ve stejném kalendářním roce či jinak vymezeném časovém období. Kohorta nicméně nemusí být nutně vymezená rokem narození, ale mnohem širěji. Například jako skupina mužů a žen, kteří prošli určitým sociálním hnutím, nebo uzavřela sňatek, či získala vysokoškolské vzdělání v určité době. Kohortou může být ale také skupina tučňáků ohrožená lidskou aktivitou, či typ automobilů mající specifickou konstrukční vadu. Vždy záleží na kritériu, které k vymezení a pojmenování určitého souboru zvolíme – nikoliv pouze na časovém údaji (Glenn, 1977; 2005). I když je kohortní analýza aplikovatelná na jakkoliv vymezené kohorty, je spojená převážně s analýzou kohort jedinců narozených v jiných obdobích.<sup>2</sup>

### **Proč je odlišení věku, období a kohort v analýze dat důležité?**

Nositelé sociálních fenoménů vyrostli (obvykle) v různých dobách. Jedná se o představitele kohort narozených v jiných letech. V případě, že zavedeme čas do zkoumané problematiky, je nezbytné zohlednit nejen jeho efekt, ale také kohortní efekt. V sociologické

<sup>2</sup> Jako synonymum pro kohortu narozených se někdy používá termín generace. Oproti kohortě, explicitně vymezené dobou narození, se jedná však o širší a vágnější koncept (více k tomu srov. Ryder, 1965; Glenn, 2005; Mannheim, 2007). Konceptem generace odlišujeme buď rodiče od jejich potomků nebo jedince narozené ve stejném roce (např. generace babyboomu) nebo jedince vymezené určitým typem jednání (např. facebooková generace).

terminologii se tento jev označuje jako kohortní výměna (*cohort replacement*). Dříve narozené kohorty odcházejí. Jejich místo zaujímají později narozené kohorty. Jejich reprezentanti mají odlišnou životní zkušenost (vyrůstali v jiné historické době). Kohortní výměna je tudíž podstatou společenské změny (působí na proměnnu zkoumaného fenoménu). V případě, že v analýze kohortní efekt neodlišíme od efektu doby (roků) a navíc ještě od efektu věku (biologického stárnutí), efekty všech tří proměnných jsou v analýze spojené a není jasné, zdali roky sběru dat, kohortní výměna, či biologické stárnutí lidí ovlivňují změny ve zkoumané problematice. A to může být zdrojem mnoha interpretačních chyb.

Kohortní výměna je mnoha autory považována za příčinu sociální změny ve stabilních demografiích. Náhradou starších věkových skupin mladšími se západní demokratické společnosti permanentně mění, ať už se jedná o oblast sociálních nebo vzdělanostních nerovností, porodnosti, úmrtnosti nebo sociální mobility. Podle Rydera (1965) musíme takovou změnu chápat jako strukturální (kompoziční) a zároveň permanentní. Nejedná se o (náhlou) změnu v jednání jedince, ale o změnu v rovině společenské (populační), kdy se v agregátu jedinců proměňují typy jednání. Kohorty narozené v jiných letech se od sebe liší obsahem školního vzdělání, socializací v rámci vrstevnické skupiny a historickou zkušeností. Reprezentanti později narozených kohorty si s sebou přinášejí specifické (a v mnoha případech inovativní) pojetí kulturního a sociálního dědictví, čímž se odlišují od reprezentantů dříve narozených kohort. Pokud by kohortní výměna neproměňovala společnost, kohorty podle roku narození by se nelišily. Postoje, názory, vnímání a typy jednání by byly stejné jak u mladých, tak starých lidí. Rok narození nediferencoval sociální a kulturní život. V západních demokratických společnostech se to prakticky neděje. Drtivá většina sociologických průzkumů ukazuje, že sociální fenomény se liší podle doby narození. Není-li příčinou sociální a kulturní proměny náhlý (revoluční) zvrát, kohortní (kompoziční) výměnu bychom měli při explanaci vývoje většiny sociálních fenoménů brát v potaz, konstatuje Ryder (1965).

### **Standardní kohortní tabulka**

Data pro kohortní analýzu jsou obvykle prezentována v kontingenční tabulce, nazývané jako *standardní kohortní tabulka* (Mason, Mason, Winsborough, Poole, 1973; Glenn, 1977). Jedná se o prezentaci zkoumané problematiky (vyjádřené v závisle proměnné) podle věku (obvykle v řádcích tabulky) a roků (obvykle ve sloupcích tabulky). Tabulka 1 prezentuje míry plodnosti dle věku žen v letech 1990 až 1995 v České republice (vyjádřenou

v procentech).<sup>3</sup> Údaje v řádcích tabulky ukazují, jak se závisle proměnná mění v čase (v letech) podle věku. Například u žen ve věku 22 let míra plodnosti klesá mezi roky 1990 až 1995 z 18.11 % na 10.94 %. Sloupce kohortní tabulky ukazují změnu závisle proměnné podle věku v jednotlivých letech. Například v roce 1990 nejvyšší míra plodnosti připadala na ženy ve věku 21. let., v roce 1995 to bylo na ženy ve věku 23 let. A konečně na diagonálách kohortní tabulky vidíme změnu v závisle proměnné u kohort vymezených rokem narození žen. Například u kohorty, která se narodila v roce 1972 (rok minus věk: 1990 - 18) je míra plodnosti nejvyšší v roce 1992 a v dalších letech klesá (v tabulce označena jako kohorta 18). U výrazně dříve narozených kohort – například v roce 1968 (kohorta 14) nebo v roce 1958 (kohorta 4) – je míra plodnosti nejvyšší v roce 1990 a ve sledovaném období se pouze snižuje.

*zde vlož tabulku 1*

Známe-li rozpětí věku (nebo počet věkových skupin  $V$ ) a počet roků ( $R$ ), počet narozených kohort ( $K$ ) dopočítáme podle vzorce:  $K = R + V - 1$ . V tabulce 1 máme k dispozici 25 narozených kohort (reprezentovaných čísly v margináliích tabulky). Tím, že řádky tabulky ukazují vývoj závisle proměnné podle věku v čase, umožňují identifikovat kohortní výměnu (Ryder 1965; Glen 1977). Představitelé věkových skupin se v letech mění a tím se také mění povaha zkoumaného fenoménu (závisle proměnné). Sloupce kohortní tabulky nabízejí *interkohortní* komparaci. Zkoumaný fenomén srovnáváme s ohledem na rok narození jeho nositelů. *Intrakohortní* srovnání nabízejí diagonály kohortní tabulky. V jednotlivých letech (čase) vidíme vývoj závisle proměnné u kohort, který můžeme mezi kohortami komparovat.<sup>4</sup> Graf 1 ukazuje tento vývoj a jeho kohortní komparaci na základě dat tabulky 1. Součástí grafu jsou vybrané kohorty narozené mezi lety 1967 až 1977 (kohorta 13 až 23). Míra plodnosti od roku 1990 do roku 1995 u žen ve věku 21. let klesá. Tento pokles je výrazný u kohort narozených v letech 1971 až 1974. U dříve narozených kohort (v letech 1969 a 1970) dochází k tomuto poklesu až ve věku 22. let ženy. Mezi 19. a 20. rokem ženy se u nich naopak míra plodnosti zvyšuje. Společenská, ekonomická a politická změna, kterou česká společnost prochází po roce 1989, sice působí plošně, nicméně protože se z hlediska

<sup>3</sup> Jedná se o poměr počtu živě narozených dětí ženám v určitém věku ke střednímu stavu žen v tom samém věku v jednotlivých letech. Míra plodnosti se standardně vyjadřuje jako promile (na 1000 žen), v tomto případě je však prezentována z důvodů analytických v procentech (na 100 žen).

<sup>4</sup> Ve standardní kohortní tabulce nemohou být nikdy všechny kohorty narozených prezentovány v celé své šíři (období). Některé kohorty jsou zachyceny pouze v jednom časovém okamžiku, jiné v řadě časových okamžiků (roků). Transformace kohortní tabulky, kdy kohorty narozených jsou v řádcích nebo sloupcích tabulky podle roků nebo věku, tento problém neřeší.

počtu narozených dětí dotýká kohort žen narozených v jiných letech, její efekt je u nich časově zpožděný.

*zde vlož graf 1*

Podle Rydera (1965) je uspořádání dat podle věku a období do podoby kontingenční tabulky důležité pro věcné pochopení významu kohort narozených v jiných letech (ve srovnání s věkovými skupinami a obdobími) na zkoumanou problematiku. Problémem kohortní tabulky nicméně je, že v procentuálních distribucích podle věku a roků kohortní efekt na zkoumanou problematiku neidentifikujeme. I když máme oprávnění domnívat se, že tento efekt existuje, v kohortní tabulce je promíchán buď s efektem věku nebo období. Rozdíly v procentuálních distribucích podle věkových skupin můžeme interpretovat jako rozdíly dané věkem, ale také jako rozdíly mezi kohortami narozenými v jiných letech. Podobně jsou interpretovatelné rozdíly mezi roky – mohou být výsledkem efektu věku, stejně jako kohortního efektu.

Pokud bychom připustili, že jeden z těchto efektů na závisle proměnnou neexistuje (například efekt věku), procentuální rozdíly, které v kohortní tabulce vidíme, by byly interpretovatelné pouze hlediska zbylých dvou proměnných (kohort a roků). Problémem ovšem je, že z kohortní tabulky nepoznáme, která z proměnných (a jestli vůbec) nemá na závisle proměnnou vliv. To je nutné buď vědět předem, nebo to předpokládat. Zjistit to lze však pouze z analýzy všech tří proměnných dohromady.

### **Problém empirické identifikace efektů věku, období a kohort**

I když jsou na teoretické rovině efekty věku, období a kohort pochopitelné, odlišení těchto efektů v datech je technicky problematické. Data ve standardní kohortní tabulce můžeme mít jednak v podobě tabulkových četností a jim odpovídajících sdružených procent (v takovém případě pracujeme s jednou kohortní tabulkou), nebo jako četnosti zkoumané události (narození, úmrtí, počet nemocných, zločinnost) oproti exponované populaci (populace vystavené danému jevu) v daném poli kohortní tabulky (v takovém případě máme k dispozici dvě kohortní tabulky – jednu pro výskyt jevu a jednu pro exponovanou populaci). Model pro odlišení efektů věku, období a kohort zapíšeme (v podobě lineárně regresní rovnice) pro standardní kohortní tabulku následovně:

$$D_{ij} = F_{ij} / P_{ij} = \mu + \alpha_i + \beta_j + \gamma_c + \varepsilon_{ij}, \quad (1)$$

kde  $D_{ij}$  označuje poměr výskytu jevu  $F_{ij}$  oproti exponované populaci  $P_{ij}$  v  $i$ -té věkové kategorii ( $i = 1, \dots, I$ ) a  $j$ -tém roce ( $j = 1, \dots, J$ ) (v tabulce 1 počet živě narozených dětí podle věku žen v jednotlivých letech oproti počtu žen podle jejich věku v těch samých letech);  $\mu$  označuje posunutí (intercept) (pro tabulku 1 průměrnou hodnotu počtu živě narozených v případě, že efekt věku, roků a kohort je nulový);  $\alpha_i$  označuje efekt věků (v každém  $i$ -tém řádku kohortní tabulky);  $\beta_j$  označuje efekt roků (pro každý  $j$ -tý sloupec kohortní tabulky);  $\gamma_c$  označuje kohortní efekt (pro každou  $c$ -tou kohortu na diagonále kohortní tabulky) a konečně  $\varepsilon_{ij}$  označuje odchylku (chybu) jednotlivých pozorování od regresní přímky, kdy předpokládáme  $E(\varepsilon_{ij}) = 0$ .<sup>5</sup>

Pokud je věk vymezen stejnou jednotkou (intervalem) jako období (např. kalendářním rokem), existuje mezi věkem, roky a narozenými kohortami (perfektní) lineární závislost. Model zapsaný v rovnici 1 pak nelze odhadnout (rovnice má nekonečně mnoho řešení). Kohorty narozené v různých letech jsou funkcí roků a věku, věk je funkcí roků a narozených kohort, a roky jsou funkcí narozených kohort a věku. Informace o dvou proměnných dostačují k tomu, abychom určili hodnotu třetí (zbylé) proměnné (kohorta = rok – věk). Například víme-li, že žena má 21. let v roce 1993, víme také, že reprezentuje kohortu narozenou v roce 1972 (kohorta 18 v tabulce 1). Nebo víme-li, že se například žena narodila v roce 1968 (kohorta 14 v tabulce 1), víme také, že v roce 1994 měla 26. let. Tento problém se nazývá identifikační problém (Blalock, 1966; Fienberg, Mason, 1985) a je příčinou potíží s určením, jaká část vysvětlované proměnné připadá na vrub věku, roků a kohort narozených v odlišných letech.

Jedná se o poměrně typický problém sociální vědy. Na teoretické rovině jsme schopni analyticky odlišit (a také zdůvodnit) koncepty věku, období a narozených kohort, empiricky je však obtížné je identifikovat kvůli jejich provázanosti.<sup>6</sup> Každý z těchto efektů má své teoretické opodstatnění, každý z nich určitým způsobem formuje lidské jednání. Jejich

<sup>5</sup> K identifikaci takového modelu je nezbytné parametry buď centrovat (tzv. ANOVA omezení), kdy suma jednotlivých parametrů  $\alpha_i$ ,  $\beta_j$ , a  $\gamma_c$  odpovídá číslu 0, nebo můžeme zvolit alternativní omezení, nazývané jako dummy omezení, kdy jednu variantu z každé proměnné věku, období a kohort zvolíme jako referenční kategorii. V prvním případě výsledky interpretujeme vzhledem k průměrné hodnotě proměnných, ve druhém případě je interpretujeme vzhledem k referenčním kategoriím proměnných (více k tomu srov. Řeháková, 2008).

<sup>6</sup> Podobné je to například v sociálně stratifikačním výzkumu se strukturální (vynucenou) a čistou (intenční) sociální mobilitou. Obě tyto mobility jsou teoreticky zdůvodněné a analyticky pochopitelné, nicméně v datech je od sebe *ex-post* odlišit v dokončeném sociálně mobilním pohybu člověka není možné. Obě mobility spolu empiricky koexistují, pravděpodobně se posilují, z dat ovšem nelze poznat, jakou část z mobilního procesu způsobují změny v zaměstnanecké struktuře a jakou úsilí člověka (více k tomu srov. Goldthorpe, 2007).



oddělení je tudíž pro badatele výzvou již po více jak 40. let. V současnosti existuje několik způsobů řešení tohoto problému.

### **Řešení empirické identifikace efektu věku, období a narozených kohort**

Literatura k odlišení efektů věku, roků a narozených kohort je na poli demografie, sociologie, biostatistiky a epidemiologie poměrně rozsáhlá (srov. např. Fienberg, Mason, 1979, 1985; Osmond, Gardner, 1982; Kupper, Janis, Karmou, Greenberg, 1985; Clayton, Shiffers, 1987; Hobcraft, Menken, Preston, 1982; Wilmont, 1990; Holford, 1992; Robertson, Boyle, 1998; Fu, 2000; O'Brien 2000). Nebudeme se zde zabývat všemi těmito řešeními (k jejich přehledu srov. Kupper, Janis, Karmous, Greenberg, 1985; Robertson, Gandini, Boyle, 1999). Představíme si pouze dvě poslední a tedy nejslibnější z nich.<sup>7</sup> A to jednak pro populační (demografická) data a jednak pro výběrová (surveyová) data. Obě tato řešení budou demonstrována na reálných datech z české společnosti.

#### *Řešení identifikace efektů APC u populačních dat*

V případě populačních (často agregovaných) dat máme většinou k dispozici standardní kohortní tabulku (věk v řádcích tabulky, období ve sloupcích tabulky a kohorty na diagonálách tabulky), což znamená, že máme omezené možnosti pro manipulaci s intervaly věku a období (fixně dány).

Na začátku 70. let byl identifikační problém řešen pomocí volby více omezení (kontrastů) u věku, období nebo kohort (Mason, Mason, Winsborough, Poole, 1973; Fienger, Mason, 1979; 1985; Mason, Smith, 1985). V praxi to vypadalo tak, že všechny tři proměnné byly pojímány jako kategoriální. U jedné z nich se namísto jednoho kontrastu (referenční kategorie, k níž jsou ostatní kategorie proměnné interpretovány) zvolila rovnost dvou kontrastů (např. vek1 = vek2 nebo rok1 = rok2). Tím se zrušila perfektní multikolinearita mezi proměnnými a efekty věku, období a kohort na vysvětlovanou proměnnou mohly být identifikovány. Pro toto řešení se ustálil název *početní model věku, období a kohort (age-period-cohort accounting model)*. Problémem je, že výsledky tohoto modelu se liší podle zvolených kontrastů (referenční kategorií) a tudíž jsou ambivalentní. Bez dodatečné informace o distribuci alespoň jednoho z efektů věku, období nebo kohort na zkoumanou proměnnou není možné rozhodnout, který model s ohledem na zvolené kontrasty přijmout. V případě, že se jeden ze tří efektů ukazuje jako zanedbatelný, lze jej z modelu odstranit a

<sup>7</sup> Nejslibnější proto, že v sobě kumulují poznání všech předcházejících a dosud známých řešení.

odhadnout model pouze se zbylými dvěma efekty (k tomuto postupu srov. např. Breen, Jonsson 2007). To ovšem není adekvátní řešení identifikačního problému.

Od poloviny 80. let minulého století sílí volání po speciální odhadovací metodě (funkci), pomocí níž by byly všechny tři efekty v datech zcela jednoznačně identifikovány. (srov. Holford, 1983; 1991; 1992; Kuplet et al. 1983; 1985; Robertson, Gandini, Boyle, 1999). Reakcí na tuto výzvu byl návrh tzv. skutečné odhadovací metody *IE* (*intrinsic estimator*) efektů věku, období a kohort, v níž není arbitrárnost výsledků modelových parametrů již přítomna (Fu, 2000; 2008; Fu, Hall, 2006). Podstatou této metody je, že neomezujeme vybrané kategorie proměnných, ale v rámci odhadu parametrů modelu je omezena geometrická orientace jejich vektorů v parametrickém prostoru. Tím se stávají odhady parametrů na sobě nezávislémi.<sup>8</sup> Odhadovací metoda *IE* byla na populačních datech několikrát použita a validita jejích výsledků byla otestována jak v komparaci s původním přístupem rovnosti vybraných kontrastů (početní model věku, období a kohort), tak na datech, v nichž efekty věku, období a narozených kohort na vysvětlovanou proměnnou byly známy (srov. Yang, Fu, Land, 2004; Yang, Fu, Shulhofer-Wolf, Land, 2008). V současnosti je tento přístup posledním, a proto velmi slibným řešením identifikace efektů věku, období a kohort na populačních datech.

Odhadovací *IE* metodu použijí na data z Českého statistického úřadu. Data se týkají živě narozených dětí a otázka, na níž budu s jejich pomocí odpovídat zní: Jaký je efekt věku, období a kohort žen na narození dítěte v České republice od konce druhé světové války do současnosti?

Tabulka 2 ukazuje počet živě narozených dětí podle věku ženy v letech 1948 až 2007 v České republice (tabulkové  $N = 8411973$ ). Tabulka 3 ukazuje počty žen podle věku v letech 1948 až 2007 v České republice (tabulkové  $N = 122079044$ ).<sup>9</sup> Obě tabulky mají uspořádání standardní kohortní tabulky (věk v řádcích, období ve sloupcích tabulky). Jak věk (29 kategorií), tak roky (60 kategorií) jsou prezentovány v intervalu jednoho roku. Na diagonálách obou tabulek identifikujeme 88 kohort žen narozených v rozdílných letech.<sup>10</sup> Z dat lze spočítat míry plodnosti pro jednotlivé kohorty žen. Těmito deskriptivními charakteristikami se zde nebudeme zabývat a přejdeme rovnou k odhadům k sobě navzájem

<sup>8</sup> K algebraickému, geometrickému a verbálnímu popisu tohoto odhadu srov. Yang, Fu, Shulhofer-Wolf, Land (2008).

<sup>9</sup> Za přípravu těchto dat děkuji Bc. Zdeně Lechnerové.

<sup>10</sup> Kohorta = rok + věk - 1 ( $88 = 60 + 29 - 1$ ).

kontrolovaných efektů věku, roků a kohort žen na počty narozených dětí pomocí metody *IE*.<sup>11</sup>

*zde vlož tabulku 2 a 3*

Velikost těchto efektů ukazují grafy 2, 3 a 4.<sup>12</sup> Efekt věku na narození dítěte má v České republice parabolický průběh (graf 2). Do 22. či 23. roku ženy věk působí pozitivně na narození dítěte. Poté – do 34. roku ženy – věk narození dítěte oslabuje a od 35. roku ženy věk působí negativně na narození dítěte (bez ohledu na to, v kterém roce a jaké kohortě žen se dítě narodí).

Oproti věku efekt období (roků) odráží společenské změny a kohortní efekt jejich chápání a reakci na ně. Graf 3 ukazuje, že roky 1948-1956 mírně zvyšují počty narozených dětí. Poté období socialismu počty narozených dětí oslabuje až do roku 1990. Výkyvy v tomto trendu vidíme letech 1963 a 1964 a od roku 1970 do roku 1974, kdy se počet narozených dětí zvyšuje. První polovina 90. let 20. století, doba významných politických, sociálních a ekonomických změn, znamená významné snížení počtu narozených dětí v České republice a od roku 1996 se trend obrací. Doba začíná působit pozitivně na narození dítěte (při kontrole věku a kohort žen v modelu).

Kohortní efekt na narození dítěte se u žen narozených mezi lety 1904 až 1991 v České republice podobá periodě (graf 4). U žen narozených v letech 1904 až 1930 kohortní efekt snižuje počty narozených dětí. Počínaje kohortami narozenými v letech 1931 a 1932, které začínají rodit po konci druhé světové války, kohortní efekt začíná zvyšovat počty narozených dětí až ke kohortám žen narozených v letech 1972 a 1973. U následujících kohort (ženy narozené v letech 1974 až 1987, které rodí po roce 1989) kohortní efekt začíná opět snižovat počty narozených dětí a u kohort žen narozených v letech 1988 až 1991 kohortní efekt začíná opět zvyšovat počty narozených dětí.

*zde vlož graf 2, 3 a 4*

<sup>11</sup> Odhad modelu APC pomocí metody IE je velmi snadné provést ve statistickém programu Stata, kde autoři této metody připravili *adofile apc\_ie* (*adofile* je automaticky spustitelný soubor). Součástí tohoto souboru je i možnost odhadu početního modelu efektů věku, období a kohort, pojmenovaného zde jako *CGLIM (constrained generalized linear models estimator)* s volbou totožnosti jakýchkoliv dvou kontrastů u věku, období nebo kohort. Komparace výsledků obou přístupů je pak nabíledni.

<sup>12</sup> Stata syntaxt modelu: *apc\_ie f, age(age) period(period) cohort(cohort) family(poisson) link(log) exposur e(pop) scale(x2)*. Velikost modelové konstanty je  $-4.97$  (SE = 0.02;  $z = -281.70$ ); Log likelihood modelu je  $-26916.32362$ .

Všechny tři efekty se ukazují jako významné při explanaci vývoje počtu narozených dětí v České republice v letech 1948 až 2007. I když je období socialismu spojeno s představou pro rodinných a pro populačních opatření (více k tomu srov. Rabušic, 1990), statistické odlišení efektů věku, období a narozených kohort žen ukázalo, že nikoliv doba (s výjimkou roků 1963 a 1964 a první poloviny 70. let minulého století),<sup>13</sup> ale po sobě narozené kohorty žen a tedy jejich „zkušenost“ s tehdejšími společenskými uspořádáními působily pozitivně na počty narozených dětí v letech 1948 až 1989. U každé mladší kohorty žen byla plodnost soustředěna do stále užšího a brzkého věkového rozmezí, které šlo ruku v ruce s nemožností realizace v jiných sférách společenského života (studium, práce, cestování, pracovní kariéra), než byla rodina (Možný, Rabuši, 1998). Socialistická společnost tak prostřednictvím omezených životních šancí působila prostřednictvím kohortní výměny pozitivně na porodnost. Nemůžeme to ovšem interpretovat jako výsledek jejich pro populačních a pro rodinných opatření, ale jako důsledek socialistického režimu omezujícího životní šance mladých lidí, dělících se do kohort podle doby narození.

Náš příklad ukázal, že odhady efektů věku, roků a kohort metodou *IE* nejsou ambivalentní, jsou smysluplné a logicky interpretovatelné. V odhadovací metodě *IE* mají v současnosti sociální vědci a vědkyně silnou a přesvědčivou metodu pro analýzu vývoje sociálních fenoménů v čase (při zohlednění efektů věku a narozených kohort). Kohortní analýzu bychom měli na základě této metody považovat za adekvátní alternativu k panelovému výzkumu.

#### *Řešení identifikace efektů APC u výběrových dat*

Výběrová (surveyová) data mají jiný charakter než data populační. Zatímco data populační vypovídají o zkoumaném fenoménu obvykle (tabulkovými) četnostmi (oproti exponované populaci), data výběrová vypovídají o vzorku respondentů a jejich sociálních charakteristikách. Populační data jsou vyčerpávající (charakterizují celou populaci bez výběrové chyby), informace v nich jsou však v drtivé většině případů omezené na několik málo proměnných. Výběrová data poznamenává výběrová chyba, obsahují však informace o řadě proměnných, jejichž vliv lze s ohledem na zkoumanou problematiku analyzovat. To je jejich výhodou oproti populačním datům. Máme-li k dispozici řadu vzorků z jednotlivých let, data umožňují nejen identifikovat efekt času, věku a narozených kohort, ale také efekt dalších

<sup>13</sup> V obou těchto obdobích byly implementovány opatření tehdejší státní politiky. V první polovině 60. let byl jednak zpřísněn potravní zákon a byla prodloužena mateřská dovolená z 18 na 22 měsíců. V první polovině 70. let se jednalo o celý komplex pronatalitních opatření (více k tomu srov. Rabuši, 1990; Aleš, 1992; Možný, Rabušic, 1998).

vysvětlujících proměnných (kontrolovaný pro věk, roky a kohorty) na závisle proměnnou. To lze udělat pomocí standardní regrese (je-li závisle proměnná spojitá), pokud nezvolíme pro věk a období stejné časové intervaly. Touto jednoduchou úpravou se sice vyhneme identifikačnímu problému (věk již není perfektní funkcí roků a kohort), nicméně Yangová a Land (2006) ukazují, že v takovém případě nelze dosáhnout jednoho z předpokladů regresního modelu, jímž je nezávislost residuí. Lineární regresní model (jako aproximace dat a procesů v nich) pak není příliš spolehlivý a v interpretaci dat nám nepomůže.

Z tohoto důvodu Yangová (2006) a Yangová a Land (2006, 2008) formulují hierarchický (víceúrovňový) model věku, období a kohort (*hierarchical age-period-cohort model, HAPC*), který řeší tento problém. Ve výběrových datech informace o věku, rocích sběru dat a narozených kohortách nemají stejný ontologický status. Věk je individuální proměnná, která charakterizuje každého jednotlivého respondenta. V datech ji Yangová a Land (2006, 2008) navrhují držet jako spojitou proměnnou, jejíž efekt by měl být analyzován dohromady s ostatními vysvětlujícími proměnnými. Oproti tomu rok (období) a narozená kohorta jsou nadindividuální (kontextuální) proměnné. Nejedná se o vlastnost jednoho konkrétního jedince, ale o charakteristiku skupiny, do které jedinec přináleží. Efekt těchto proměnných lze považovat za kvalitativně odlišný od efektů individuálních proměnných. Zatímco věk indikuje biologické stáří člověka (a míru sociální zkušenosti či mentální zralosti), období a kohorty zachycují vůči člověku vnější sociální tlaky. Působí na nositele zkoumaného fenoménu jiným způsobem než věk. Yangová a Land (2008) navrhují tyto dva druhy proměnných a jejich efekty na lidské jednání proto v analýze dat oddělit. V duchu víceúrovňových modelů (k tomu viz Soukup, 2006) konceptualizují individuální proměnné jako efekty první úrovně (věk a další vysvětlující proměnné) a skupinové proměnné (období a kohorty) jako efekty vyšší, druhé úrovně, jež jsou na efektech první úrovně nezávislé.

Tímto způsobem – jako hierarchický model věku, období a kohort – řeší identifikační problém u výběrových dat.<sup>14</sup> Ve druhé (kontextuální) úrovni proměnných spolu koexistují dvě proměnné: období a kohorty – hierarchický model věku, období a kohort musí být v takovém případě specifikován jako *model náhodných křížených efektů (Cross-Classified Random-Effects Model – CCREM)*. Náhodný (kontextuální) efekt každého období na vysvětlovanou proměnnou je identifikován jako průměrný napříč všemi kohortami a náhodný (opět kontextuální) efekt kohort je identifikován jako průměrný napříč všemi roky (obdobími). Protože jsou analytické efekty roků a kohort oddělené (a nezávislé) od efektu věku, není

<sup>14</sup> Návrh tohoto řešení, i když bez možnosti jeho realizace pomocí odpovídajícího statistického programu, viz také Fienberg, Mason (1985)

nutné pracovat se stejnými věkovými, časovými a kohortními intervaly.<sup>15</sup>

Pro ilustraci tohoto přístupu použiji data z tabulky 2 a 3. Odečtením četností tabulky 2 od četností tabulky 3 dostaneme informaci o počtech dětí, které se nenarodili podle věku v jednotlivých letech (tabulkové  $N = 122079044 - 8411973 = 113667071$ ). Roky jsem agregoval do 30 kategorií po 2 letech (z původních 60 kategorií pro léta 1948-2007). Data jsem poté transformoval do individuální podoby, kde jsem v každém řádku měl informaci o věku ženy, období (vymezení intervalem 2 let) a zdali k narození dítěte došlo (ano/ne). S ohledem na věk a závisle proměnnou (narození dítěte) jsem v každém období provedl náhodný výběr (30 náhodných výběrů) o velikosti jednoho promile původní dat z daném období. Data za jednotlivá období jsem poté sloučil do jedné matice (celkové  $N$  vzorku = 122257). Odečtením věku od roků jsem v tomto vzorku identifikoval 87 kohort žen (první kohorta se narodila v roce 1905 a poslední v roce 1991), které jsme sloučil po 3 letech do 29 kohort (první kohorta 1905/1907, druhá kohorta 1908/1910.... poslední kohorta 1989/1991).

Otázka, na kterou nyní odpovídám, je stejná jako v případě populačních dat. Jaký je efekt věku, období a kohort na narození dítěte v České republice? Oproti populačním datům mám nyní dispozici náhodný výběr, jinak specifikovanou závisle proměnnou (binární, nikoliv počet živě narozených dětí z populace žen) a s roky a narozenými kohortami pracuji jako s agregáty delších časových úseků.

Závisle proměnná (narození) má dvě kategorie (ano/ne). Pro její explanaci pomocí nezávisle proměnných věku, období a kohort zvolím model binární logistické regrese s náhodnými kříženými efekty období a kohort na modelovou konstantu. Rovnice modelu je následující:

$$\text{logit}(y_{ijk}) = \beta_{0jk} + \beta_1 \text{věk}_{ijk} + \beta_2 \text{věk}_{ijk}^2 + \varepsilon_{ijk}, \quad (2) \quad \text{úroveň 1}$$

$$\beta_{0jk} = \gamma_0 + \mu_{0j} + \nu_{0k}, \quad (3) \quad \text{úroveň 2}$$

$$\text{logit}(y_{ijk}) = \gamma_0 + \beta_1 \text{věk}_{ijk} + \beta_2 \text{věk}_{ijk}^2 + \mu_{0j} + \nu_{0k} + \varepsilon_{ijk}, \quad (4) \quad \text{kombinace úrovně 1 a 2}$$

kde  $i$  označuje ženu ( $i = 1, 2, \dots, n_{jk}$ ) v rámci jednotlivých období  $j$  ( $j = 1, \dots, 30$ ) a kohort  $k$  ( $k = 1, \dots, 29$ ). V tomto víceúrovňovém modelu je přirozený logaritmus narození oproti nenarození dítěte u ženy  $i$  (logit) modelován jako funkce věku a kvadratického věku (předpoklad nelineárního vztahu mezi věkem a narozením dítěte) v rámci každého roku  $j$  a

<sup>15</sup> Yangová a Land (2008) navrhuje věk vymezovat po roce, období (pokud to data umožňují) rovněž po roce a kohorty delším časovým intervalem pěti let.

kohorty  $k$ . Konstanta  $\beta_{0,jk}$  je směrnice, neboli průměrný přirozený logaritmus šance na narození dítěte v roce  $j$  a kohortě  $k$ ;  $\beta_1$  je fixní efekt věku na logit narození dítěte;  $\varepsilon_{ijk}$  je náhodný (individuální) efekt v rámci první úrovně;  $\gamma_0$  je hlavní modelový průměr (logit narození pro všechny ženy ve vzorku),  $\mu_{0,j}$  je náhodný (residuální) efekt období  $j$  (tedy efekt roků na logit narození) a  $\nu_{0,k}$  je náhodný efekt kohorty  $k$  (tedy efekt kohort na logit narození). Průměrný logit narození dítěte v rámci kohort  $k$  pak dostaneme jako  $\beta_{0,j} = \gamma_0 + \mu_{0,j}$  a v rámci roků  $j$  jako  $\beta_{0,k} = \gamma_0 + \nu_{0,k}$ .

Tabulka 4 ukazuje statistické charakteristiky tohoto modelu. V první části tabulky vidíme efekt věku a kvadrátu věku (dohromady parabolický efekt) na logit závisle proměnné (přirozený logaritmus narození versus nenarození), včetně intervalů spolehlivosti (95% hladina významnosti) a modelovou konstantu. Druhá část tabulky ukazuje variabilitu (směrodatnou odchylku) konstanty napříč kohortami narozených žen a obdobími, včetně intervalů spolehlivosti (95% hladina významnosti). Mezi kohortami konstanta variuje ve větší míře než mezi roky (explanační potenciál závisle proměnné je vyšší). To, ve kterém roce se žena narodí, silněji poznamenává narození či nenarození jejího dítěte v České republice než rok, v němž se dítě narodí.

Grafy 5, 6 a 7 ukazují odhadnuté modelové efekty pro věk a kvadrát věku, období a kohorty žen na logit narození dítěte, včetně intervalů spolehlivosti (95% hladina významnosti). Graf 5 (stejně jako graf 2) ukazuje, že efekt věku na narození dítěte je nesilnější okolo 22 či 23 roku života ženy a poté šance na narození dítěte věk snižuje. V grafu 3 jsme viděli, že období 1957 až 1990 s výkyvy v polovině 60. a první polovině 70. let 20. století působilo negativně na narození dítěte. Tyto výkyvy jsou v grafu 6 zřejmé, nicméně trend oslabování šancí na narození dítěte mimo tyto výkyvy není z grafu 6 patrný. Po roce 1991 je již trend v obou grafech stejný: do konce poloviny 90. let minulého století značné oslabení šancí na narození dítěte a od té doby čas působí ve prospěch narozených dětí.

Poslední komparaci trendů nabízí graf 7 a 4. Z grafu 4 víme, že počínaje kohortami žen, které začaly rodit po druhé světové válce (narodili se v roce 1931 a později), kohortní efekt na narození dítěte roste až ke kohortám žen, které začaly rodit po roce 1989 (narodili se v roce 1974 a později). Poté kohortní efekt prudce oslabuje a teprve u kohort narozených na konci 80. let 20. století a později začíná opět růst. I když graf 7 ukazuje v kohortních efektech podobné trendy jako graf 4, není růst křivky pro kohorty narozené mezi 30. lety až polovinou 70. let 20. století tak explicitní. Pokles kohortního efektu poté je již v obou grafech stejný.

Jeho vzrůst u nejpozději narozených kohort žen však graf 7 nedokumentuje. Nicméně velké rozpětí intervalu spolehlivosti u nejpozději narozených kohort žen znamená, že tomu takto může být (existuje zde však velká míra nejistoty, daná nízkým počtem narozených dětí u žen těchto kohort ve výběrových datech).

*zde vlož graf 5, 6 a 7*

I v tomto případě analýza dat pomocí hierarchického (binárně logistického) modelu s kříženými náhodnými efekty období a kohort na modelovou konstantu ukázala, že všechny tři proměnné (věk, období a kohorty narozených žen) nejsou při explanaci narozených dětí v České republice zanedbatelné. Základní struktura ve výběrových datech, i když závisle proměnná byla jiného charakteru a období a kohorty měly jiné intervaly, odpovídá struktuře populačních dat, kterou jsme odhalili pomocí metody *IE*. Tato komparace nejen ukázala, že hierarchický (víceúrovňový) přístup je řešením identifikačního problému kohortní analýzy, pracujeme-li s výběrovými daty, ale také naznačila, že existuje-li v datech dominantní struktura, lze ji identifikovat více adekvátně zvolenými statistickými technikami.

## **Závěr**

Zkoumáme-li dnes vývoj sociálních fenoménů, kohortní analýza je alternativou k panelovému výzkumu. Kohortní analýzu není nutné oproti panelovému výzkumu dopředu nijak zvlášť plánovat. Máme-li k dispozici data za jednotlivé roky, která obsahují informaci o věku zkoumaných (a pochopitelně informaci o zkoumané problematice), máme rovněž informace o narozených kohortách. Kohortní analýza je v takovém případě použitelná. Pokud se jedná o data populační (obvykle jde o vyjádření jevu v podobě četností ze zkoumané populace), text ukázal, že výzkumníci a výzkumnice mají v současnosti k dispozici odhadovací metodu *IE*. Tato metoda neambivalentně identifikuje efekty věku, narozených kohort a období na zkoumanou problematiku, i když mezi těmito proměnnými existuje perfektní multikolinearita (údaj jedné proměnné je vždy výsledkem údajů zbylých dvou proměnných). V případě, že pracujeme s výběrovými daty za jednotlivé roky, adekvátním řešením je použití hierarchického (víceúrovňového) modelu s náhodnými kříženými efekty období a kohort na modelovou konstantu. V tomto případě jsou efekty období a kohort pojímány jako kvalitativně odlišné (jedná se o skupinové efekty) oproti efektu věku, který je definován jako individuální (typický pro každého jednotlivého člověka), což je umožňuje nejen od sebe oddělit, ale také v empirických datech identifikovat.



I když byla v tomto textu představena zatím dvě poslední řešení pro odlišení efektů věku, období a kohort narozených, je nezbytné konstatovat, že konečné řešení identifikačního problému není demografickou, epidemiologickou nebo sociologickou komunitou zatím jednomyslně akceptováno. Text ukázal, že významným posunem na cestě k tomu řešení je pojmání populačních a surveyových (výběrových) dat jako kvalitativně odlišných. I když se odhadovací metoda *IE* jeví jako velmi slibné řešení pro populační data, je nezbytné provést další testy a zevrubně tento způsob odhadu efektů věku, období a narozených kohort otestovat. Víceúrovňový model s kříženými efekty období a kohort na modelovou konstantu se zdá být jako nezpochybnitelné řešení v případě surveyových dat. Problém tohoto řešení však spočívá v tom, že zavedením složitějších interakcí a dalších proměnných do modelu narážíme na limity statistického poznání a znalostí, jak model odhadnout. Zvláště interakce mezi efekty věku, období a kohort jsou výzvou pro budoucnost (například identifikace kohortě specifického efekt věku). Teoreticky se dnes sice už ví, že dané interakce by mělo být možno odhadnout (Yang, 2007; 2008a), nicméně prakticky (pomocí dostupných statistických programů) nejsou zatím modely s těmito interakcemi na poli statistiky nejen identifikované, ale ani zdokumentované. S ohledem na tuto skutečnost budou následující léta v badání modelů kohortní analýzy velkým příslibem. Doufejme, že jak se budou rozvíjet víceúrovňové modely s kříženými efekty, budeme svědky nejen metodologických, ale i věcných posunů na poli bádání statistických modelů pro kohortní analýzu.

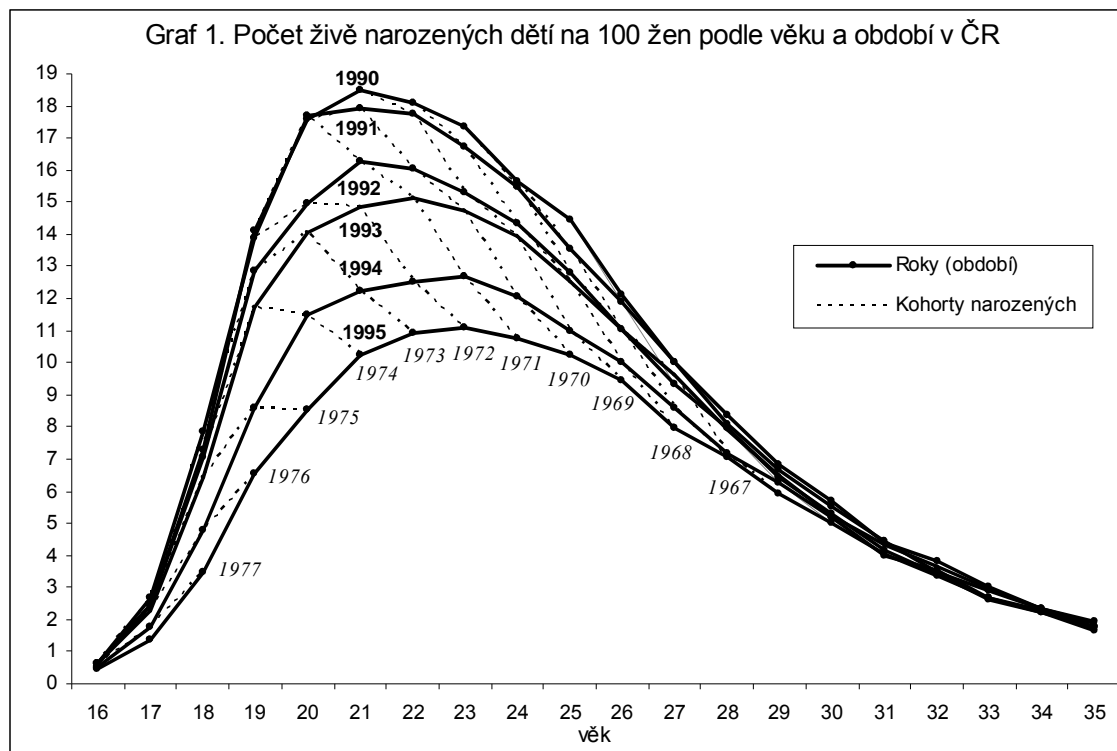
**Literatura**

- Aleš, M.: „Populační politika – ano či ne?“ *Demografie* (1992), 1, 54-55.
- Alwin, D. F. and R. J. McCammon. 1999. „Aging Versus Cohort Interpretations of Intercohort Differences in GSS Vocabulary Scores.“ *American Sociological Review* 64: 272-86.
- Blalock, H. M. 1966. „The Identification Problem and Theory Building.“ *American Sociological Review* 31 (1): 52-61.
- Breen, R. and J. O. Jonsson. 2007. „Explaining Change in Social Fluidity: Educational Equalization and Educational Expansion in Twentieth-Century Sweden.“ *Američan Journal of Sociology* 112:1775-1810.
- Clayton, D., and E. Schiffers. 1987. “Models for Temporal Variation in Cancer Rates II: Age-Period-Cohort Models.” *Statistics in Medicine* 6 (4):469–81.
- Fienberg, S. E., Mason, W. M. 1985. „Specification and Implementation of Age, Period, and Cohort Models.“ Pp. 45-88 in Mason, W. M., Fienberg, S. E. (ed.). *Cohort Analysis in Social Research*. New York: Springer.
- Fienberg, S. E.; Mason, W. M. 1979. „Identification and Estimation of Age-Period-Cohort Models in the Analysis of Discrete Archival Data.“ *Sociological Metodology* 10 (1): 1-67.
- Fu, W. J. 2000. “Ridge Estimator in Singular Design with Application to Age-Period- Cohort Analysis of Disease Rates.” *Communications in Statistics Theory and Method* 29 (2):263–78.
- Fu, W. J. 2008. “A Smoothing Cohort Model in Age-Period-Cohort Analysis with Applications to Homicide Arrest Rates and Lung Cancer Mortality Rates.” *Sociological Methods and Research* 36 (3): 327–61.
- Fu, W. J., Hall, P. 2006. “Asymptotic Properties of Estimators in Age-Period- Cohort Analysis.” *Statistics and Probability Letters* 76 (17):1925–29.
- Glenn, N. D. 1994. „Television Watching, Newspaper Reading, and Cohort Differences in Verbal Ability.“ *Sociology of Education* 67:216-30.
- Glenn, N. D. 1997. *Cohort Analysis*. London: Sage.
- Glenn, N. D. 1999. „Further Discussion of the Evidence for an Intercohort Decline in Education-Adjusted Vocabulary.“ *American Sociological Review* 64: 267-71.
- Glenn, N. D. 2005. *Cohort Analysis*. London: Sage.
- Goldthorpe, J. H. 2007. „Outline of a Theory of Social Mobilty.“ Pp. 154-185 in. Goldthorpe, J. H. (ed.). *On Sociology*. 2. vydání. Stanford: Stanford University Press.
- Hobcraft, J., J. Menken, and S. Preston. 1982. “Age, Period, and Cohort Effects in Demography: A Review.“ *Population Index* 48:4-43.
- Hobcraft, J., Menken, J., Preston, S. 1982. “Age, Period, and Cohort Effects in Demography: A Review.” *Population Index* 48 (1): 4–43.
- Holford, T. R.. 1992. “Understanding the Effects of Age, Period, and Cohort on Incidence and Mortality Rates.” *Annual Review of Public Health* 12 (1): 425–57.
- Kupper, L. L., Janis, J. M., Karmous, A., Greenberg B. G. 1985. "Statistical Age-Period-Cohort Analysis: A Review and Critique." *Journal of Chronic Disease* 38 (10): 811-830.
- Mannheim, K. 2007. „Problém generací.“ *Sociální studia* 4 (1-2): 11-44.
- Mason, K. O., Mason, W. H., Winsborough, H. H., Poole, K. 1973. „Some Methodological Issues in Cohort Analysis of Archival Data.“ *American Sociological Review* 38 (2): 242-258.
- Mason, K. O., W. H. Mason, H. H. Winsborough, and K. Poole. 1973. „Some Methodological Issues in Cohort Analysis of Archival Data.“ *American Sociological Review* 38: 242-58.

- Mason, W. H., Smith, H. L. 1985. "Age-Period-Cohort Analysis and the Study of Deaths from Pulmonary Tuberculosis." Pp. 151–228 in Mason, W. M., Fienberg, S. E. (eds.). *Cohort Analysis in Social Research*. New York: Springer-Verlag.
- Mason, W. M., Wolfinger, N. H. 2002. „Cohort Analysis.“ Pp. 2189-2194 in Smelser, N. J., Baltes, P. B. (eds.) *International Encyclopedia of the Social & Behavioral Sciences*. New York: Elsevier 2002.
- Možný, I.; Rabušic, L.: „Česká rodina, sňatkový trh a reprodukční klima.“ In. Večerník, J. ; Matějů, P.: *Zpráva o vývoji české společnosti*. Academia, Praha 1998, 92-110.
- O'Brien, R. M. 2000. "Age Period Cohort Characteristic Models." *Social Science Research* 29 (1): 123–39.
- Osmond, C., Gardner, M. J. 1982. "Age, Period, and Cohort Models Applied to Cancer Mortality Rates." *Statistics in Medicine* 1 (3): 245–59.
- Rabušic, L.: „Manifestní a latentní funkce československé populační politiky.“ *Demografie* (1990), 3, 234-238.
- Rabušic, L.: *Česká společnost stárne*. MU a Georgetown, Brno 1995.
- Řeháková, B. „Kontrasty v logistické regresi.“ *Sociologický časopis* 44 (4): 745-765.
- Robertson, C., Boyle, P. 1998. "Age-Period-Cohort Analysis of Chronic Disease Rates I: Modeling Approach." *Statistics in Medicine* 17 (12): 1305–23.
- Robertson, C., Gandini, S., Boyle, P. 1999. "Age-Period-Cohort Models: A Comparative Study of Available Methodologies." *Journal of Clinical Epidemiology* 52 (6): 569-583.
- Ruspini, E. 2002. *Introduction to Longitudinal Research*. London: Routledge.
- Ryder, N. 1965. „The Cohort As A Concept in the Study of Social Change.“ *American Sociological Review* 30 (6): 843-861.
- Soukup, P. 2006. „Proč užívat hierarchické lineární modely?“ *Sociologický časopis* 42 (5): 987-1012.
- Wilmoth, J. R. 1990. "Variation in Vital Rates by Age, Period, and Cohort." *Sociological Methodology* 20: 295–335.
- Wilmoth, J. R. 1990. „Variation in Vital Rates by Age, Period, and Cohort.“ *Sociological Methodology* 20:295-335.
- Yang, Y. 2006. "Bayesian Inference for Hierarchical Age-Period-Cohort Models of Repeated Cross-section Data." *Sociological Methodology* 36 (1): 39–74.
- Yang, Y. 2007. "Age/Period/Cohort Distinctions". Pp. 20-22 in Markides, K. S. (ed.). *Encyclopedia of Health and Aging*. London: Sage Publications.
- Yang, Y., Fu, W. J., Land, K. C. 2004. "A Methodological Comparison of Age-Period-Cohort Models: Intrinsic Estimator and Conventional Generalized Linear Models." *Sociological Methodology* 34 (1): 75–110.
- Yang, Y., Land, K. C. 2006. "A Mixed Models Approach to the Age-Period-Cohort Analysis of Repeated Cross-section Surveys, with an Application to Data on Trends in Verbal Test Scores." *Sociological Methodology* 36 (1): 75–97.
- Yang, Y. and Kenneth C. Land. 2008. "Age-Period-Cohort Analysis of Repeated Cross-Section Surveys: Fixed or Random Effects?" *Sociological Methods and Research* 36: 297-326.
- Yang, Y., Sam Schulhofer-Wohl, Wenjiang Fu, and Kenneth Land. 2008. „The Intrinsic Estimator for Age-Period-Cohort Analysis: What It Is and How to Use It?“ *American Journal of Sociology*. 113: 1697-1736.
- Yang, Y. 2008a "Age, Period, Cohort Effects." Pp. 6-10 in *Encyclopedia of the Life Course and Human Development*, Deborah Carr (ed). New York: Gale Publishing.
- Yang, Y. 2008b. „Social Inequalities in Happiness in the U.S. 1972-2004: An Age-Period-Cohort Analysis.“ *American Sociological Review* 73: 204-226.

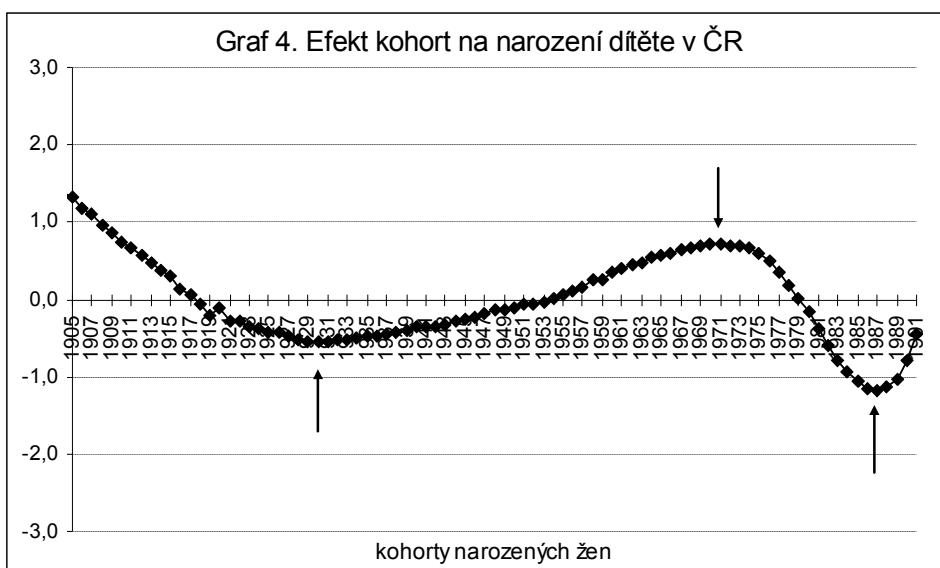
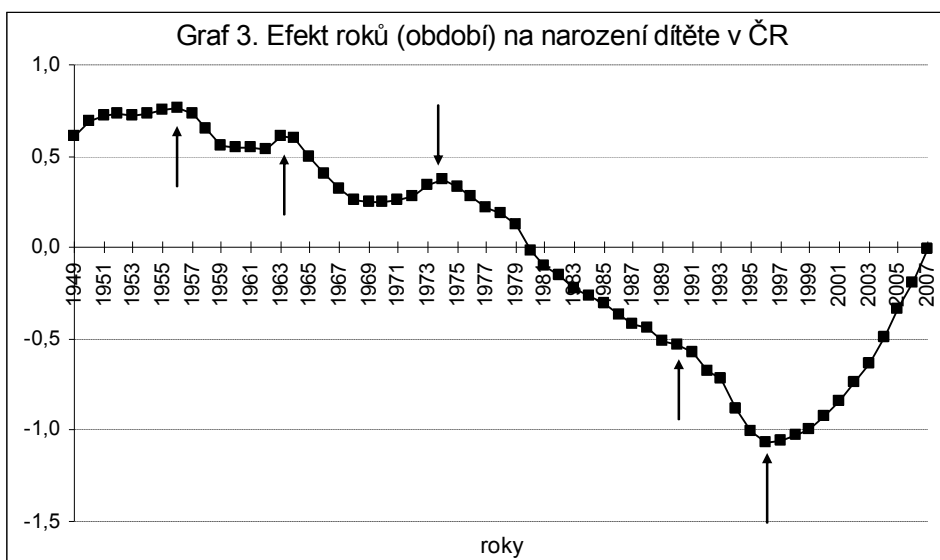
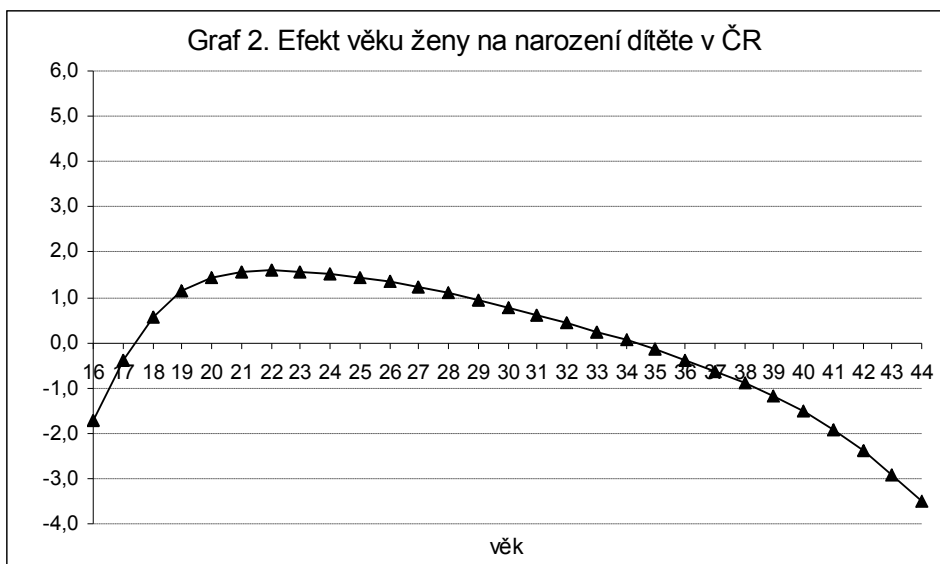
Tabulka 1. Počet živě narozených dětí na 100 žen podle věku a období v ČR

věk	období (roky)						kohorty	rok narození
	1990	1991	1992	1993	1994	1995		
16	0,60	0,64	0,62	0,61	0,49	0,46	25	1979
17	2,44	2,70	2,43	2,30	1,79	1,34	24	1978
18	7,30	7,85	7,08	6,44	4,77	3,45	23	1977
19	14,11	13,87	12,86	11,73	8,60	6,56	22	1976
20	17,60	17,69	14,96	14,05	11,51	8,53	21	1975
21	18,49	17,94	16,27	14,82	12,24	10,23	20	1974
22	18,11	17,73	16,07	15,16	12,50	10,94	19	1973
23	17,36	16,70	15,30	14,71	12,71	11,07	18	1972
24	15,66	15,46	14,33	13,96	12,07	10,74	17	1971
25	14,43	13,55	12,79	12,51	10,96	10,25	16	1970
26	12,14	11,88	11,03	11,01	10,02	9,46	15	1969
27	10,04	10,01	9,32	9,63	8,58	7,95	14	1968
28	8,39	8,06	7,95	7,88	7,19	7,08	13	1967
29	6,80	6,68	6,45	6,49	6,24	5,90	12	1966
30	5,69	5,49	5,28	5,21	5,19	4,99	11	1965
31	4,36	4,44	4,17	4,35	3,97	4,06	10	1964
32	3,80	3,45	3,44	3,64	3,52	3,33	9	1963
33	3,01	2,90	2,64	2,95	2,67	2,68	8	1962
34	2,34	2,36	2,20	2,32	2,23	2,28	7	1961
35	1,92	1,78	1,64	1,85	1,75	1,75		
kohorty	1	2	3	4	5	6		
rok narození	1955	1956	1957	1958	1959	1960		









Tabulka 4. Binárně logistický model náhodně křížených efektů období a kohort na modelovou konstantu.

```

Mixed-effects logistic regression                Number of obs    =    122257
-----
Group Variable |   No. of   Observations per Group   Integration
                |   Groups   Minimum   Average   Maximum   Points
-----+-----
      _all |         1   122257   122257.0   122257     1
    period |        30    3604    4075.2    4358     1
-----

Log likelihood = -27942.663                    Wald chi2(2)      =    2059.77
                                                Prob > chi2      =     0.0000

-----
      birth |   Coef.   Std. Err.    z    P>|z|    [95% Conf. Interval]
-----+-----
      C_age | .6847682   .0181245   37.78  0.000    .6492448   .7202916
    C_age2 | -.0143389   .0003428  -41.82  0.000   -.0150108  -.0136669
      _cons | -3.231371   .1822799  -17.73  0.000   -3.588633  -2.874109
-----

-----
Random-effects Parameters |   Estimate   Std. Err.    [95% Conf. Interval]
-----+-----
_all: Identity
      sd(R.cohort) |   .9476724   .1491123    .6961859   1.290004
-----+-----
period: Identity
      sd(_cons) |   .2028878   .0348972    .1448262   .2842266
-----

LR test vs. logistic regression:    chi2(2) =    750.81    Prob > chi2 = 0.0000

```



