

ČESKÉ PANELOVÉ ŠETŘENÍ DOMÁCNOSTÍ

Dokumentace k datům z kontaktních formulářů

Vlna 1 (2015) – vlna 4 (2018)

Mgr. Michaela Kudrnáčová

Sociologický ústav AV ČR, v. v. i.

www.promenyceskespolecnosti.cz

promeny.spolecnosti@soc.cas.cz

Verze 3.0 (21. 11. 2019)



Obsah

1.	Obsah kontaktních formulářů	2
1.1.	Časy kontaktu (<i>month, day, hour, min</i>).....	2
1.2.	Výsledek kontaktu (<i>conout</i>)	2
1.3.	Důvody odmítnutí (<i>conref</i>).....	3
1.4.	Typ obydlí (<i>houstp</i>).....	4
1.5.	Kvalita bydlení v oblasti (<i>haqual</i>)	4
1.6.	Srovnání kvality bydlení (<i>hqualco</i>)	4
2.	Další proměnné v datovém souboru.....	5
3.	Domácnosti zařazené do dat.....	7
4.	Domácnosti vyšetřené pomocí CAWI	9
5.	Údaje o tazatelích.....	10
6.	Verze datových souborů	11

Přehled tabulek

Tabulka 1:	odvozené proměnné a proměnné z dalších zdrojů.....	6
Tabulka 2:	počty domácností v datech z kontaktních formulářů	8
Tabulka 3:	obsah datových souborů s charakteristikami tazatelů.....	10
Tabulka 4:	vydané verze datových souborů z kontaktních formulářů.....	11
Tabulka 5:	vydané verze datových souborů s údaji tazatelů	11

1. Obsah kontaktních formulářů

Za účelem monitorování průběhu terénních prací pořizovali tazatelé Českého panelového šetření domácností (CHPS) záznamy o okolnostech kontaktování domácností do kontaktního formuláře v CAPI administračním rozhraní. Při každém pokusu o kontakt na vybraných adresách bylo úkolem tazatele zaznamenat měsíc (*month*), den (*day*), hodinu (*hour*) a minutu (*min*) kontaktu, dále výsledek kontaktu (*conout*) a v případě odmítnutí také jeho důvod (*conref*). Tazatelé měli rovněž určit typ bydlení, v němž domácnost žije (*houstp*), zhodnotit kvalitu bydlení v dané oblasti (*haqual*) a srovnat kvalitu bydlení domácnosti s okolím (*hqualco*).

1.1. Časy kontaktu (*month, day, hour, min*)

Proměnné obsahují údaje o čase, v němž se odehrál pokus o kontakt daného pořadí. Hodnota 99 značí, že dle ostatních informací zadaných do kontaktního formuláře existuje evidence o návštěvě domácnosti tazatelem, ale čas kontaktu nebyl zaznamenán nebo byla hodnota ztracena kvůli technickým problémům při ukládání a přenosu dat.

Hodnota 95 značí, že kontakt daného pořadí měl formu přivání ke CAWI dotazování, jehož se domácnost zúčastnila.

V prvním a druhém roce výzkumu zapisovali tazatelé agentury STEM/MARK pouze měsíc a den kontaktu, nikoli hodinu a minutu. V těchto proměnných obsahují proto domácnosti agentury STEM/MARK chybějící hodnoty (*system missing*).

1.2. Výsledek kontaktu (*conout*)

Tazatelé kódovali výsledek kontaktu na adrese do následujících kategorií:

Výsledek kontaktu:	
1	Získán kompletní rozhovor
2	Domácnost odmítla rozhovor
3	Rozhovor započat, ale nedokončen
4	Domluveno dotazování na jiný termín
5	Jazyková bariéra – žádný člen neumí česky
6	Neschopnost se zúčastnit (nemocní, staří apod.)
7	Nikdo nebyl nezastižen
8	Nezastižen nikdo dospělý
9	Adresa neexistuje / nenalezena
10	Adresa není byt

Výsledek kontaktu daného pořadí se váže vždy k času kontaktu (*month až min*) stejného pořadí. Hodnota 99 značí, že tazatel výsledek kontaktu nezaznamenal nebo údaj byl ztracen v důsledku technických problémů.

Do proměnných *conout* byly při zpracování dat doplněny následující kódy:

11	Domácnost nebyla tazatelem navštívena	kód ponechán pouze jako finální výsledek v proměnné <i>outcome</i> , protože pro nenavštívené domácnosti nebyl kontaktní formulář pořízen (tzn. na těchto adresách neproběhl pokus o osobní kontakt)
12	Domácnost nereagovala na CAWI výzvu	(od W2) kód pro nevyšetřené domácnosti, které byly v předchozím roce dotázány pomocí CAWI a nebyly proto vydány k provedení osobního rozhovoru
13	Změna adresy, nedotázáno	(od W2) kód doplněn na základě informace tazatele o změně adresy celé domácnosti
14	Dotázáno, vyřazeno na základě kontrol	domácnost byla dotázána, avšak následné kontroly odhalily pochybení tazatele nebo jiné nedostatky ve vyšetření domácnosti

1.3. Důvody odmítnutí (*conref*)

Jestliže domácnost odmítla účast ve výzkumu, vybírali tazatelé hlavní důvod odmítnutí z kategorií níže.

Jaký byl hlavní důvod odmítnutí účasti celé domácnosti?	
1	Nedostatek času (členové domácnosti stále zaneprázdnění)
2	Špatné načasování, výjimečná situace v domácnosti (např. právě odjíždějí, nemocný člen, návštěva)
3	Nezájem / nedůvěra k výzkumům obecně
4	V nedávné minulosti na nějaký výzkum odpovídali
5	Předchozí špatná zkušenost s výzkumy
6	Nezájem o téma tohoto výzkumu (nezajímá ho téma apod.)
7	Nedůvěra k tomuto výzkumu (nemá dost informací, vadí mu téma apod.)
8	Obava z nedostatku anonymity / zneužití informací / poskytování osobních údajů
9	Nedostatečná motivace – malá odměna apod.
10	Strach z tazatele (strach otevřít, strach z podomního prodeje)
11	Některý člen domácnosti „zakázal“ účast
12	Jiné, doplňte
13	Odmítnutí bez uvedení důvodu

Důvod odmítnutí daného pořadí se váže vždy k výsledku kontaktu (*conout*) stejného pořadí.

Při zpracování dat byly slovní odpovědi v polootevřené otázce *conref* zařazeny do odpovídající uzavřené kategorie. Hodnota 99 značí, že tazatel důvod odmítnutí nezaznamenal nebo hodnota byla v důsledku technických problémů ztracena. Pokud byla tazatelem vybrána kategorie 12 a zároveň chyběla navazující volná odpověď, proměnná *conref* byla začištěna na hodnotu 99. Tento postup byl zvolen, protože převažující část uvedených volných odpovědí bylo možné zakódovat do uzavřených možností.

1.4. Typ obydlí (*houstp*)

V jakém typu obydlí domácnost žije?

- 1 Zemědělská usedlost
 - 2 Samostatný dům
 - 3 Dvojdomek
 - 4 Řadový dům
 - 5 Byt v domě s více obytnými jednotkami
 - 6 Jiné, doplňte:
 - 7 Nevím
-

1.5. Kvalita bydlení v oblasti (*haqual*)

Jak hodnotíte kvalitu budov/obydlí v této oblasti?

- 1 Velmi dobrá
 - 2 Dobrá
 - 3 Uspokojivá
 - 4 Špatná
 - 5 Velmi špatná
-

1.6. Srovnání kvality bydlení (*hqualco*)

Jak hodnotíte kvalitu obydlí domácnosti ve srovnání s kvalitou budov/obydlí v této oblasti?

- 1 Výrazně vyšší
 - 2 Spíše vyšší
 - 3 Stejná
 - 4 Spíše nižší
 - 5 Výrazně nižší
-

Hodnoty 99 v proměnných *houstp*, *haqual* a *hqualco* značí, že tazatel odpověď na otázku nezaznamenal nebo hodnota byla kvůli technickým problémům ztracena.

Ve druhém ročníku zjišťovali tyto údaje pouze tazatelé agentury STEM/MARK a domácnostem agentury MEDIAN byly v datech z daného roku přiděleny chybějící hodnoty (system missing).

Domácnostem kontaktovaným od druhého ročníku pouze prostřednictvím CAWI je v uvedených proměnných přiřazena hodnota 95 (pouze CAWI).

Při zpracování dat byly slovní odpovědi v polootevřené otázce *houstp* zařazeny do odpovídající uzavřené kategorie.

2. Další proměnné v datovém souboru

Datové soubory obsahují unikátní identifikační čísla domácností (*hid*) pro spojení s dalšími typy daty. Zahrnuty jsou vždy identifikátory ze současné a všech předchozích vln.

Při zpracování dat bylo do souborů doplněno několik dalších proměnných odvozených z otázek v kontaktním formuláři nebo z dalších zdrojů informací o průběhu šetření.

Proměnná *coform* určuje, zda jsou pro danou domácnost dostupné odpovědi na otázky z kontaktního formuláře (*month – hqualco*). V některých případech tazatelé kontaktní formuláře nevyplnili nebo data byla ztracena kvůli technickým problémům. Domácnostem bez údajů z kontaktního formuláře je v proměnné *coform* přiřazena hodnota 0 a proměnné *month* až *hqualco* obsahují chybějící hodnoty (system missing). Domácnosti pokračující z předchozího roku v CAWI mají v proměnné *coform* vždy hodnotu 1.

Hodnoty proměnné *result* indikují, zda domácnost byla v příslušném ročníku výzkumu dotázána (1) či nikoli (0). Jestliže byla domácnost dotázána, ale následné kontroly ukázaly na pochybení tazatele nebo se vyskytly jiné problémy (např. nebyl dokončen ani CAPI dotazník pro domácnost, základní sociodemografické údaje o respondentech z PAPI dotazníků byly v rozporu s údaji z CAPI), je domácnosti přiřazena hodnota 2. Tyto domácnosti nabývají v proměnné *outcome* hodnoty 14.

Agenturu, která měla vyšetřit danou domácnost, udává proměnná *agen*.

Mód dotazování je uveden v proměnné *qmode*. Rozlišeno je zde CAPI (1) a CAWI (2) dotazování. Nevyšetřeným domácnostem je v této proměnné přidělen kód 0.

Proměnná *convert* nabývá hodnotu 1 u těch domácností, které byly přizvány ke CAWI dotazování po odmítnutí CAPI rozhovoru nebo neúspěšných pokusech o osobní kontakt tazatelem. Tato proměnná je dostupná v datech až od druhého ročníku výzkumu, protože v prvním roce nebyla o CAWI konverzi pořizována systematická evidence na úrovni jednotlivých domácností.¹ Od druhého ročníku výzkumu nabízela agentura MEDIAN možnost přechodu na CAWI jako plošnou konverzní strategii zejména ve větších městech, agentura SM dávala tuto nabídku pouze ve výjimečných případech.

Proměnná *intid* obsahuje číslo tazatele, který dotázal příslušnou domácnost. U domácností konvertovaných na CAWI a nedotázaných domácností je uvedeno číslo přiděleného tazatele. Pokud byla domácnost v předchozím roce vyšetřena prostřednictvím CAWI, v následujícím roce jí byla poštou nebo elektronicky zaslána výzva ke CAWI dotazníku bez pokusu o návštěvu tazatelem. V těchto případech je v proměnné *intid* uveden kód 999 996. Pokud se nepodařilo některou z oblastí obsadit tazatelem, příslušné domácnosti jsou v proměnné *intid* značeny kódem 999 998.

Finální výsledek kontaktu udává proměnná *outcome*, jejíž hodnoty jsou odvozeny z proměnných *conout*. Pokud pro danou domácnost nebyl k dispozici kontaktní formulář (*coform* = 0) a domácnost nebyla dotázána (*result* = 0), je jako výsledný kontakt v proměnné *outcome* uvedeno „domácnost nebyla

¹ Údaj o úspěšnosti CAWI konverze pro první ročník je v dokumentaci k hlavnímu datovému souboru odhadnut z celkových počtů odmítajících a nekontaktovaných domácností.

navštívena“ (11). Jestliže domácnost dotázána byla (*result* = 1), výsledný kontakt byl v proměnné *outcome* doplněn na „získán kompletní rozhovor“ (1).

Územní identifikace adresy jsou obsaženy v proměnných *c_msize* (velikost místa bydliště), *c_region* (kraj) a *c_county* (okres). Údaje za první rok výzkumu se vztahují k adrese v době provedení výběru. V dalších vlnách údaje odpovídají bydlišti při dotazování v předchozím roce.

V datovém souboru z prvního ročníku výzkumu jsou zahrnuty proměnné *strata* a *psu*, které udávají příslušnost domácnosti ke stratu při stratifikovaném výběru a výběrový bod.

Tabulka 1: odvozené proměnné a proměnné z dalších zdrojů

w(x)_hid	unikátní identifikační číslo domácnosti z vlny (x)	W1 +
coform	zda jsou pro domácnost dostupné údaje z kontaktního formuláře	W1 +
result	zda byla domácnost úspěšně dotázána (1), nedotázána (0) nebo dotázána, avšak vyřazena z dat (2)	W1 +
agen	agentura	W1 +
qmode	zda se domácnost zúčastnila osobního rozhovoru (1) nebo prostřednictvím CAWI (2) nebo nebyla vyšetřena (0)	W1 +
convert	pokus o konverzi z CAPI na CAWI ano (1) či ne (0)	W2 +
intid	číslo tazatele (tentýž tazatel má stejné číslo v datech ze všech let výzkumu)	W1 +
outcome	konečný výsledek kontaktu	W1 +
refusal	konečný důvod odmítnutí	W1 +
c_msize	velikost místa bydliště domácnosti v době výběru/dotazování v předchozím roce	W1 +
c_region	kraj bydliště domácnosti v době výběru/dotazování v předchozím roce	W1 +
c_county	okres bydliště domácnosti v době výběru/dotazování v předchozím roce	W1 +
strata	strata (1 až 58)	W1
psu	výběrový bod (1 až 1 275)	W1

3. Domácnosti zařazené do dat

Data z **1. ročníku** obsahují údaje o **12 810** domácnostech. Kromě 12 750 domácností z výběrového souboru je zahrnuto 60 úspěšně vyšetřených domácností, které představují druhé a další hospodařící domácnosti ve vybraných bytových jednotkách (tzn. jednu bytovou jednotku obývá více než jedna hospodařící domácnost).

Nedotázané další domácnosti v bytě nejsou do dat z kontaktního formuláře zahrnuty. O těchto domácnostech se nevedla systematická evidence a informace o jejich počtu vychází z proměnné *ohous* v CAPI dotazníku pro domácnost (za předpokladu, že v bytě je pouze jedna další domácnost). V dokumentaci k hlavnímu datovému souboru jsou nevyšetřené další domácnosti v bytě zahrnuty do kategorie „domácnost nebyla navštívena“ v přehledu o výsledcích kontaktů.

Během sběru dat v první vlně bylo 312 adres vyhodnocených jako nebytové nahrazeno dalšími adresami ze stejného výběrového bodu v hrubém výběrovém souboru. Nové adresy nejsou v datech z kontaktních formulářů obsaženy společně s původními adresami, ale původní adresy nahrazují.

Soubor z **2. ročníku** zahrnuje **5 109** domácností, z toho 5 093 domácností vyšetřených v první vlně (z 5 159 dotázaných nebylo 67 CAWI domácností vydáno, protože k nim neexistoval kontaktní údaj nebo byl chybně pořízen; naopak byla vydána a úspěšně vyšetřena jedna domácnost vyřazená kvůli pochybení tazatele z dat prvního ročníku) a 16 domácností vzniklých rozdělením.

Do souboru z **3. ročníku** je zařazeno **4 158** domácností, z toho všech 4 147 dotázaných z druhé vlny a 11 domácností vzniklých rozdělením. Soubor ze **4. ročníku** zahrnuje **3 626** domácností. Z toho 3 616 domácností vyšetřených ve třetí vlně a 10 vzniklých rozdělením.

Do dat z kontaktních formulářů jsou druhé a další části rozdělených domácností začleněny pouze, pokud byly dotázány. O nevyšetřených dalších částech rozdělených domácností nebyla vedena systematická evidence pomocí kontaktních formulářů a jejich počet je v dokumentaci k hlavnímu datovému souboru odhadnut z proměnné *rstat* (rezidenční status členů) v CAPI dotazníku pro domácnost za předpokladu, že domácnost s odchozími nezemřelými členy se rozdělila na dvě poloviny. V dokumentaci k hlavnímu datovému souboru jsou nevyšetřené druhé poloviny rozdělené domácnosti zahrnuty do kategorie „domácnost nebyla navštívena“ v přehledu o výsledcích kontaktů.

Počty domácností zahrnutých do dat z kontaktního formuláře a považovaných za vydané v dokumentaci k hlavnímu datovému souboru shrnuje *tabulka 2*. Sloupec „výsledek kontaktu“ značí typ výsledku kontaktu, do nějž jsou v dokumentaci k hlavnímu datovému souboru zařazeny odhadované počty nevyšetřených dalších domácností v bytě a rozdělených domácností a dále náhradní domácnosti za nebytové adresy v první vlně. Ačkoli se ne všechny nedotázané části rozdělených domácností přesunuly na novou adresu², jsou z důvodu zachování jednoduchosti ve výsledcích kontaktu započítány do kategorie 13 (změna adresy, nedotázáno).

² Podle odhadu na základě proměnných *rstat* a *asame* se na novou adresu přesunulo 95 % nevyšetřených částí rozdělených domácností.

Tabulka 2: počty domácností v datech z kontaktních formulářů

	W1	W2	W3	W4	výsledek kontaktu
čistý výběrový soubor	12 750	-	-	-	-
další domácnost v bytě, vyšetřená	60	-	-	-	-
další domácnost v bytě, nevyšetřená *)	147	-	-	-	11 (nenavštívena)
náhradní adresy	312	-	-	-	10 (adresa není byt)
vyšetřené v předchozí vlně (<i>result</i> = 1)	-	5 092	4 147	3 616	-
vyřazené v předchozí vlně (<i>result</i> = 2)	-	1	0	0	-
oddělené domácnosti, vyšetřené	-	16	11	10	-
oddělené domácnosti, nevyšetřené *)	-	155	145	108	13 (změna adresy)
CELKEM v KF datech	12 810	5 109	4 158	3 626	
CELKEM vydané adresy	13 269	5 264	4 303	3 734	

*) Odhad na základě CAPI dotazníku pro domácnost.

4. Domácnosti vyšetřené pomocí CAWI

V prvním roce výzkumu byl domácnostem odmítajícím osobní účast a dlouhodobě nezastiženým domácnostem zaslán dopis s nabídkou vyplnění dotazníku na webu obsahující unikátní přístupový kód do CAWI mutace dotazování. Pokud domácnost nabídku CAWI využila, je jako výsledek kontaktu (*conout*) následujícího po posledním osobním kontaktu uvedeno „získání kompletního rozhovoru“ (1) a do odpovídajících časů kontaktu (*month* až *min*) je vložen kód 95. Vyplnění po internetu je řazeno vždy jako poslední kontakt, ačkoli reálně mohla po vyplnění CAWI následovat návštěva tazatele, který včas nedostal informaci o proběhnutím CAWI dotazování.

U neúspěšných CAWI konverzí není v proměnných *month* až *min* a *conout* obsažen záznam o CAWI nabídce a jako finální výsledek kontaktu (*outcome*) je uveden jeden z kódů vybraných tazatelem. V prvním roce výzkumu nebyla zajištěna systematická evidence o domácnostech oslovených s nabídkou CAWI dotazování, a v datech z kontaktního formuláře proto není obsažena indikace pokusu o konverzi na CAWI.

Od druhého ročníku výzkumu nebyly adresy domácností, které se v předchozím roce zúčastnily prostřednictvím CAWI, vydávány tazatelům k provedení osobního rozhovoru, ale poštou nebo elektronicky byly opět přizvány k vyplnění dotazníku na internetu. Jako výsledek kontaktu (*conout*) prvního pořadí je u těchto domácností uvedeno buď „získání kompletního rozhovoru“ (1), nebo „nereagování na CAWI výzvu“ (12). V časech kontaktů prvního pořadí je vložena hodnota 95.

Domácnosti, kterým bylo ve druhém a dalším roce výzkumu nabídnuto CAWI dotazování v rámci konverzní strategie, jsou rozlišeny v proměnné *convert* hodnotou 1. Z důvodu zachování konzistence s datovým souborem z prvního ročníku není neúspěšná CAWI konverze zanesena jako kontakt posledního pořadí v proměnných *month* až *cmin* a *conout*.

Pokud bylo domácnosti ve druhé a další vlně nabídnuto přímo CAWI dotazování a adresa tedy nebyla přidělena tazateli k osobnímu dotazování, je v proměnných *houstp*, *haqual* a *hqualco* uvedena hodnota 95.

5. Údaje o tazatelích

Datové soubory z dotazníků a kontaktních formulářů obsahují unikátní identifikační číslo tazatele (*intid*), které zůstává stejné ve všech vlnách, na nichž příslušný tazatel pracoval. Pomocí unikátního identifikačního čísla tazatele lze k souborům připojit základní charakteristiky tazatelů.

Základní informace o tazatelích jsou dostupné v samostatných souborech připravených pro jednotlivé vlny výzkumu. Mezi tyto charakteristiky patří pohlaví (*insex*), rok narození (*inbirthy*), nejvyšší dosažené vzdělání (*ineduc*) a délka praxe u agentury (*inexp*) vyjádřená v kategoriích. Nejvyšší dosažené vzdělání a délka praxe odpovídají stavu při začátku sběru dat v daném ročníku výzkumu.

Interní čísla tazatelů platná v rámci agentur byla v datových souborech nahrazena novými náhodnými identifikačními čísly.

Do souborů s charakteristikami tazatelů jsou zahrnuti jak tazatelé, kteří do výzkumu přispěli alespoň jednou dotázanou domácností, tak tazatelé, kteří žádnou z přidělených domácností nedotázali.

Tabulka 3: obsah datových souborů s charakteristikami tazatelů

intid	číslo tazatele (tentýž tazatel má stejné číslo v datech ze všech let výzkumu)
agen	agentura
insex	pohlaví tazatele
inbirthy	rok narození tazatele
ineduc	nejvyšší dosažené vzdělání tazatele
inexp	praxe tazatele u dané agentury

6. Verze datových souborů

Datové soubory z kontaktních formulářů v prvním a druhém ročníku CHPS byly ke dni 18. 2. 2019 vydány ve dvou verzích. Ve verzích 1.1 byla upravena proměnná *c_msize* tak, aby se kódování velikosti obce řídilo stejným číselníkem jako v hlavních datových souborech. Soubor z druhé vlny zahrnuje nově navíc územní identifikace domácnosti, která byla z důvodu pochybení tazatele vyřazena z dat první vlny, ale ve druhém roce byla přesto vydána a úspěšně dotázána.

Datový soubor z kontaktních formulářů ve třetím ročníku byl ke dni 18. 2. 2019 vydán v jedné verzi. Datový soubor ze čtvrtého ročníku byl ke dni 21. 11. 2019 zveřejněn v jedné verzi.

Datové soubory s údaji o tazatelích z první až čtvrté vlny byly vydány v jedné verzi.

Tabulka 4: vydané verze datových souborů z kontaktních formulářů

	1. ročník	2. ročník	3. ročník	4. ročník
v1.0	20. 2. 2018	20. 2. 2018	18. 2. 2019	21. 11. 2019
v1.1	18. 2. 2019	18. 2. 2019	-	-

Tabulka 5: vydané verze datových souborů s údaji tazatelů

	1. ročník	2. ročník	3. ročník	4. ročník
v1.0	20. 2. 2018	20. 2. 2018	18. 2. 2019	21. 11. 2019