

Univerzita Karlova

Filozofická fakulta

Katedra Sociologie

Diplomová práce

Bc. Filip Rybín

Sampling Fusion: Ekvivalence měření tří módů sběru dat

Sampling Fusion: Measurement invariance of three modes of data
collection

Praha 2021

Vedoucí práce: PhDr. Jiří Vinopal, Ph.D.

Prohlášení:

Prohlašuji, že jsem diplomovou práci vypracoval samostatně, že jsem řádně citoval všechny použité prameny a literaturu a že práce nebyla využita v rámci jiného vysokoškolského studia či k získání jiného nebo stejného titulu.

V Praze, dne 29. July 2021

Filip Rybín

Poděkování

Děkuji svému vedoucímu práce, PhDr. Jiřímu Vinopalovi, Ph.D., za podporu a flexibilitu po celou dobu vzniku této práce. Zároveň děkuji agentuře NMS Market Research, jmenovitě Mgr. Ondřeji Veisovi, Mgr. Tereze Antoňové, Mgr. Petru Chlubnovi a RNDr. Jiřímu Votinskému za pravidelné konzultace ve všech fázích této práce a při realizaci sběru dat. V neposlední řadě děkuji Móně za důvěru a neutuchající podporu.

Věnováno babičce a dědovi.

V Praze, dne 29. July 2021

Klíčová slova (česky)

Smíšené módy sběru dat, ekvivalence měření, politická participace, politická preference, CAWI, CATI, pravděpodobnostní výběr, nepravděpodobnostní výběr, efekt módu

Klíčová slova (anglicky):

Mixed mode design of data collection, measurement invariance, political participation, political preferences, CAWI, CATI, probability sampling, nonprobability sampling

Abstrakt (česky)

Dotazníkové šetření je klíčovým nástrojem zkoumání společenských jevů, stejně jako celá společnost ale prochází vývojem a je potřeba ho podrobovat neustálým analýzám za účelem jeho zlepšení. Přístup smíšených módů sběru dat lze poté využít k hodnocení a srovnání jednotlivých módů na úrovni ekvivalence měření za účelem poukázání na jejich míru přesnosti měření a zhodnocení podobnosti či rozdílnosti. Skrze analýzu politické participace, preference a přesnosti měření pro tři rozdílné módy sběru dat – CATI, CAWI a CAWI s oporou výběru na sociální síti Facebook byly definovány jasné rozdíly jednotlivých módů a poukázáno na vhodnost jejich kombinace pro využití silných stránek, které nabízejí, při současné minimalizaci chyb, které jsou s nimi spojeny.

Abstract (in English):

The questionnaire survey is a key tool for examining social phenomena, but like society as a whole, it is undergoing development and needs to be subjected to constant analysis in order to improve it. The approach of mixed modes of data collection can then be used to evaluate and compare individual modes at the level of measurement invariance in order to point out their degree of measurement accuracy and to evaluate similarity or difference. Through the analysis of political participation, preference and measurement accuracy for three different modes of data collection - CATI, CAWI and CAWI with sampling frame of social network Facebook, clear differences of individual modes were defined and the recommendation of their combination was stated in order to use the strengths they offer, while minimizing the errors associated with them.

OBSAH

1 ÚVOD	3
1.1 CÍL PRÁCE	4
1.2 VÝZKUMNÁ OTÁZKA A HYPOTÉZY	4
1.3 STRUKTURA DIPLOMOVÉ PRÁCE	5
2 TEORETICKÁ ČÁST	7
2.1 SMÍŠENÉ MÓDY SBĚRU DAT	7
2.2 EKVIVALENCE, SROVNATELNOST A PŘESNOST DAT	8
2.3 TEORIE VÝZKUMNÉ CHYBY	9
2.3.1 Celková chyba měření	10
2.4. KOMPLEMENTÁRNÍ TEORIE CHYB.....	14
2.4.1 Nesystematické zdroje chyb.....	14
2.4.2. Systematické zdroje chyb.....	15
2.4.3 Efekty nesystematických a systematických chyb.....	15
2.4.3 Střední kvadratická chyba.....	16
2.5 PROBLEMATIKA (NE)PRAVDĚPODOBNOSTNÍCH VÝBĚRŮ	16
2.5.1 Krátká historie módů sběru dat.....	16
2.5.2 Přesnost dat u pravděpodobnostních a nepravděpodobnostních výběrů.....	17
2.5.3 Snižující se návratnost.....	18
2.5.4 Vliv módů sběru dat na míru návratnosti.....	18
2.5.5 Váha jednotlivých módů sběru dat v aplikovaném výzkumu veřejného mínění a výzkumu trhu.....	19
2.6 SHRNUÍ TEORETICKÉ SEKCE – VLIV MÓDU	21
3 METODOLOGICKÁ ČÁST	23
3.1 DESIGN VÝZKUMU	23
3.1.1 Telefonické dotazování (CATI)	23
3.1.2 Online dotazníkové šetření na panelu (CAWI).....	24
3.1.3 Online dotazníkové šetření na sociální síti Facebook (Social)	25
3.1.4 Struktura výběrových souborů	25
3.2 ANALYTICKÉ PŘÍSTUPY K EKVIVALENCE MĚŘENÍ.....	29
3.2.1 Role hypotéz v analýze ekvivalence měření	30

3.3 ANALYTICKÝ POSTUP	31
4 ANALYTICKÁ ČÁST	34
<i>4.1.1 Analýza přesnosti měření deklarované politické preference</i>	<i>34</i>
<i>4.1.1 Analýza přesnosti měření deklarované politické participace</i>	<i>50</i>
4.2 ANALÝZA FAKTORŮ OVLIVŇUJÍCÍCH POLITICKOU PARTICIPACI	52
<i>4.2.1 Analýza faktorů ovlivňujících politickou participaci - shrnutí</i>	<i>63</i>
4.3 DOPROVODNÉ ANALÝZY MÓDŮ	64
5 ZÁVĚR	66
6 DISKUZE NAD VÝSLEDKY PRÁCE	68
7 SEZNAM POUŽITÝCH ZDROJŮ	69

1 Úvod

Empirický sociovědní aparát je klíčový pro percepci, výzkum a analýzu společenských jevů, tedy společnosti jako takové. Z tohoto důvodu je nutné podrobit samotné metodologické postupy sběru dat průběžným analýzám s cílem zhodnocení kvality dat, reliability, validity a reprezentativity, které poskytují.

Dotazníkové šetření patří k integrální součásti celého výzkumného aparátu. Tato diplomová práce má za cíl srovnat tři módy sběru dat – telefonické dotazování (CATI) s pravděpodobnostním prostým náhodným výběrem, online dotazníkové šetření s nepravděpodobnostním kvótním výběrem a oporou dat v online panelu (CAWI na panelu respondentů) a online dotazníkové šetření s nepravděpodobnostním kvótním výběrem a oporou dat na sociální síti Facebook (Social, CAWI na sociální síti Facebook).

Srovnání probíhá na úrovni ekvivalence dat volebních preferencí a participace z podzimu roku 2020 a reflektuje tak současnou situaci ovlivněnou pandemií COVID 19 a její dopad na dostupné módy sběru. Do výběru použitých módů se tak propisuje nemožnost použití CAPI sběru stejně jako dichotomie tradičních opor výběru (náhodný prostý pravděpodobnostní výběr) ve srovnání s oporami využívající postupnou digitalizaci společnosti (kvótní výběr v panelu respondentů a sběr dat na sociálních sítích).

Výzkum, na kterém tato práce stojí, je předvolební výzkum Krajských voleb 2020, provedený v druhé polovině září téhož roku. Následná analýza poté porovnává data získaná smíšeným módem sběru dat za účelem jasného zmapování jejich vzájemných rozdílů jak z hlediska výsledných dat, tak finanční a časové náročnosti každé metody.

1.1 Cíl práce

Cílem této diplomové práce je přinést poznání o ekvivalenci měření na úrovni tří rozdílných módů sběru dat, přičemž jeden z těchto módů využívá prostý pravděpodobnostní výběr (CATI) a další dva nepravděpodobnostní kvótní výběr (CAWI a Social). Ačkoli se jedná o analýzu srovnatelnosti, která má zodpovědět na otázku, zdali se jednotlivé výběrové soubory (potažmo data sebraná třemi různými módy) liší, a ideálně poukázat i na případné zdroje těchto odlišností, nejedná se o vyčerpávající studii ekvivalence měření, která rozkrývá kauzální vztahy mezi jednotlivými chybami měření a výběrovými soubory samotnými. Přínos této práce leží v zasazení do kontextu reálného každodenního problému ve výzkumu veřejného mínění a trhu, kdy data využitá pro tuto práci pochází z reálného výzkumu provedeného agenturou NMS Market Research. Závěry této práce mají poukázat na přítomnost základních rozdílů mezi módy sběru dat skrze analýzu ekvivalence měření, jedná se tedy o důležité téma z hlediska praxe ve výzkumu veřejného mínění a trhu a zjištění mají sloužit právě k rozvoji empirických výzkumných nástrojů.

Téma ekvivalence měření je samo o sobě často přehlíženo, ačkoli je pro vývoj empirické sociologie klíčové. Důvodů je několik, jedním z hlavních je vysoká nákladnost čistě metodologického výzkumu, který sám o sobě nepřináší žádné substantivní výsledky, ale *jen* validuje módy sběru a metodologii výzkumu.

V Českém prostředí se takovému výzkumu věnuje např. PhDr. Petra Anýžová (Raudenská), Ph.D., zároveň v současné době vzniká komplexní ekvivalenční výzkum pod vedením Mgr. Martina Lakomého, Ph.D. na FSS MU, jehož výsledky zatím bohužel nejsou k dispozici. V rámci kvalifikačních prací se tématu věnovala například i Mgr. Kristýna Pospíšilová z FF UK a Mgr. Jarmila Pilecká z FSV UK. V neposlední řadě je nutno podotknout, že pokud už ekvivalenční výzkum na téma smíšeného módu sběru dat vzniká, jedná se převážně o výzkum módů postavených na pravděpodobnostních výběrech. Z toho důvodu je dílčím cílem této diplomové práce rozšířit ekvivalenční výzkum právě pro nepravděpodobnostní sběr dat.

1.2 Výzkumná otázka a hypotézy

Základní výzkumnou otázkou této diplomové práce je otázka ekvivalence dat získaných pomocí 3 rozdílných módů sběru dat – Do jaké míry jsou data získaná skrze CATI, CAWI na panel (CAWI) a CAWI na sociální síti Facebook (Social) srovnatelná?

Analýza dat proběhla s přihlédnutím na charakter dat a téma výzkumu na úrovni otázek, škál a baterií se zaměřením na politickou participaci a preferenci. Výsledné nulové hypotézy jsou následující:

H₀₁: Data sebraná odlišnými módy sběru dat, tedy CATI, CAWI Panel (CAWI) a CAWI Facebook (Social), jsou ekvivalentní z hlediska politické preference.

H₀₂: Data sebraná odlišnými módy sběru dat, tedy CATI, CAWI Panel (CAWI) a CAWI Facebook (Social), jsou ekvivalentní z hlediska politické participace.

H₀₃: Data sebraná odlišnými módy sběru dat, tedy CATI, CAWI Panel (CAWI) a CAWI Facebook (Social), stejnou měrou reprezentují skutečný stav z hlediska politické participace a preference (jsou tedy stejně validní).

1.3 Struktura diplomové práce

Tato práce je dle zvyklosti empirických výzkumů členěna na 5 základních sekcí: úvod, teoretická část, metodologická část, analytická část a závěrečné shrnutí s diskuzí. Přechod mezi teoretickou a empirickou částí je plynulý, zejména díky širokému rozsahu použitých teoretických konceptů, díky kterým je možné data v analytické části interpretovat a navrhnout v rámci závěrečného shrnutí možné kauzální vztahy a postup pro navazující výzkumy. Stejně tak přechod mezi metodologickou a analytickou částí. Právě šíře použitých přístupů a neexistující konsenzus z hlediska nejvhodnějšího metodologického postupu je způsoben již zmiňovaným nedostatkem podobně zaměřených prací a exploračním charakterem práce.

V úvodu práce jsou zmíněny cíle práce a jejich relevance pro obor sociologie, včetně aplikovaného výzkumu na úrovni výzkumu veřejného mínění a výzkumu trhu. Zároveň obsahuje definici testovaných nulových hypotéz.

Teoretická část seznamuje čtenáře se základními koncepty: definuje smíšené módy sběru dat a jejich využití ve výzkumu, představuje koncept ekvivalence měření, staví na teorii měření kvality a chybovosti dat pro definici hlavních rozdílů jednotlivých analyzovaných módů a v neposlední řadě se věnuje metodologické problematice na úrovni pravděpodobnostních a nepravděpodobnostních metod výběru.

Metodologická část představuje postup sběru dat u všech tří módů, následně samotný datový soubor a analytické přístupy k jeho analýze. Obhájí jejich použití a přínos a navazuje na východiska teoretické části, na kterých staví.

V analytické části jsou představeny výstupy použitých analýz a nástrojů. Na několika úrovních testuje hypotézy definované v úvodu práce a interpretuje důvody k jejich potvrzení či zamítnutí.

Závěr práce obsahuje shrnutí výsledků a následnou diskuzi nad výsledky analýzy. Poukazuje na charakteristiky provedeného výzkumu a diskutuje navrhované navazující postupy pro výzkum ekvivalence rozdílných módů sběru dat.

2 Teoretická část

2.1 Smíšené módy sběru dat

Výzkum této diplomové práce spadá do kategorie výzkumů využívajících smíšené módy sběru dat (v anglosaské literatuře je využíváno označení *mixed-mode* nebo *multiple-mode*). V základu se jedná o takové výzkumy, které kombinují několik módů sběru dat, a to v různé fázi výzkumu a/nebo na různých podskupinách výzkumného vzorku.

Historicky vzato došlo k prvnímu smíšenému módu sběru již na konci 18. století (de Heer et al., 1999, s. 29). K první empirické komparaci při využití několika různých módů sběru dat za účelem jejich srovnání došlo v roce 1967 (Hochstim, 1967, s. 976). Na tomto místě je nutné zmínit hlavní teoretickou z hlediska smíšených módů sběrů dat, Edith de Leeuw. Ta poukazuje na primární účel a výhody využití smíšených módů sběru v rámci jednoho výzkumu – právě tento výzkumný design umožňuje využívat silných stránek jednotlivých módů (Leeuw et al., 2008, s. 301). Při dobře navrhnutém designu výzkumu, zohledňujícím téma a cílovou skupinu, lze vhodnou kombinací módů sběru dat docílit snížení výzkumné chyby. Zároveň je díky kombinaci různých módů sběru možné docílit i efektivity na úrovni času a prostředků, kde různé módy sběru nabízejí různé vlastnosti z hlediska efektivity (de Leeuw, 2005, s. 234).

Nejčastější rozdělení designu výzkumu zahrnující smíšený mód sběru dat je definován kombinací těchto módů (Buchtík, 2012, s. 132):

1. Souběžný způsob sběru dat: v tomto případě je více módů sběru dat využito v jednu chvíli v rámci designu výzkumu.
2. Následný způsob sběru dat: jednotlivé módy sběru dat po sobě následují.

Nabízí se tak možnost kombinace módů specifických pro rekrutaci cílové skupiny a samotný sběr. Další možností je kombinace různých módů během samotného sběru dat – paralelně může část cílové skupiny respondentů vyplňovat dotazník v rozdílných módech nebo může každý jednotlivý respondent vyplnit část dotazníku v jednom módu a zbytek v jiném. Posledním a v dnešní době stále relativně sporadicky využívaným příkladem je situace, kdy si respondent sám zvolí, jaký mód sběru dat preferuje, a ten poté následuje. Přínosem takového přístupu je potenciální zvýšení návratnosti a snížení chyby spojené s výpadkem návratnosti (Dillman, 2020, s. 18).

Jak již toto základní rozdělení ukazuje, smíšené módy sběru dat lze využít i jiným způsobem, shodným s cíli této práce. Výzkum spojený se smíšenými módy sběru dat lze rozdělit na tři hlavní proudy (Pilecká, 2016, s. 12):

1. Výzkum zaměřený primárně na kognitivní procesy při dotazování (Bowling, 2005, s. 281).
2. Výzkum zaměřený primárně na vliv módu na odpovědi a chování respondenta (Randolph et al., 2006, s. 195).
3. Výzkum zaměřený primárně na ekvivalenci měření, dopad vlivu módu na samotné měření a kvalitu dat včetně analýzy response rate a reprezentativity výzkumu (Klausch et al., 2013, s. 227).

Tento výzkum jasně spadá do třetí kategorie a jedná se zároveň o nejméně zastoupený výzkum dle předloženého třídění. Vysvětlení tohoto relativního nezájmu ze strany akademické obce je nasnadě – ekvivalenci je nutné vztáhnout k měření kvality dat. K jejímu měření je možné využít mnoho teoretických a metodologických přístupů, nicméně právě kombinací srovnání kvality měření mezi jednotlivými módy dochází k interferencím a je obtížné měření kvality očistit od ostatních efektů. Výzkumníci tak raději zkoumají dílčí efekty jednotlivých módů než jejich komplexní zhodnocení (Buchtík, 2012, s. 134). Ale i poukázání na dílčí efekty je přínosné a potvrzení či vyvrácení ekvivalence dat skrze analýzy prvního a druhého stupně je validní metodologický postup pro odhalení rozdílnosti či shodnosti jednotlivých výběrových souborů (Anýžová, 2013, s. 41).

Smíšené módy sběru dat jsou díky potenciálu snižování výzkumné chyby skrze vhodnou kombinací módů sběru dat hojně využívanou alternativou k výzkumům využívající pouze jeden mód sběru dat. Výzkum ekvivalence jednotlivých módů je ale opomíjené výzkumné téma, zejména kvůli své komplexitě a možné interferenci způsobené kombinací módů a dalších faktorů, které mohou výzkum ovlivnit, a v důsledku toho komplikují jednoznačné poukázání na přímý vliv testovaných módů sběru dat (Jäckle et al., 2010, s. 18).

2.2 Ekvivalence, srovnatelnost a přesnost dat

Cílem této práce je srovnat tři odlišné módy sběru dat. Z tohoto důvodu je nutné jasně konceptualizovat problematiku srovnatelnosti, ekvivalence a následně i přesnosti dat. Pojem ekvivalence je v oblasti sociologie, psychologie, statistiky nebo například i vzdělávání používán pro popis stálosti, neměnnosti a srovnatelnosti měření. Ekvivalence (v anglosaském

kontextu je využíván pojem invariance) samotná se dělí na tři základní úrovně (Anýžová, 2013, s. 31):

1. Srovnatelnost teoretického konceptu
2. Srovnatelnost položek, které daný koncept měří
3. Srovnatelnost jejich škál měření (tj. kategorií odpovědí).

Ekvivalence teoretického konceptu je tématem pro srovnání studií napříč zeměmi, kde jazykové a kulturní odlišnosti mohou ovlivnit samotný význam zkoumaného konceptu (více v následující kapitole, která se mimo jiné zabývá chybou validity). V této práci byly nicméně všechny tři módy sběru aplikovány v rámci jedné země a tato ekvivalence není pro níže představený výzkum klíčová.

Jedná se právě o ekvivalenci na úrovni dat, potažmo měření, která je pro tuto práci klíčová. Zde primárně dochází k testování srovnatelnosti měření skrze metrickou a skalární ekvivalenci. Metrická ekvivalence popisuje měřicí nástroj ve formě škály, kde jsou použity stejné jednotky i rozsah škály, nicméně subjektivní chápání (a hodnocení) se na škálách mezi skupinami (v tomto případě módy) liší. Skalární ekvivalence poté popisuje takové měření, kde je efekt subjektivního chápání a hodnocení mezi skupinami odstraněn.

V neposlední řadě je kromě teorie srovnatelnosti dat nutné čtenáře seznámit také s konceptem přesnosti dat, kterou lze uchopit jako důležitý ukazatel kvality dat. V kontextu této práce přesnost měření označuje blízkost hodnot získaných v rámci jednotlivých výběrových šetření vůči pravdivé (či reálné) hodnotě (Krejčí, 2008, s. 24).

2.3 Teorie výzkumné chyby

V návaznosti na definici využití smíšeného módu sběru dat je potřeba představit teoretický rámec přístupu k chápání výzkumné chyby.

V první řadě je nutné vymezit definici chyb, které jsou obecně s kvantitativním sběrem a zpracováním dat spojovány a jejich analýza je pro srovnání jednotlivých módů sběru nezbytná. V tomto ohledu je jedna z nejvíce klíčových teorií teorie Celkové chyby měření (Total Survey Error), poprvé definovaná v roce 1989 (Groves, 1989, s. 15). Tento teoretický zastřešující rámec popisuje rozdíl ve statistické hodnotě mezi populací a odhadem téže statistické hodnoty získané skrze výběrové šetření (Biemer & Lyberg, 2003, s. 36).

Cílem teorie celkové chyby měření je zohlednit veškeré aspekty, které mají vliv na výslednou kvalitu výběrového šetření (Groves & Lyberg, 2010, s. 849).

Při jakémkoli výzkumu je nutné zajistit dvě základní podmínky:

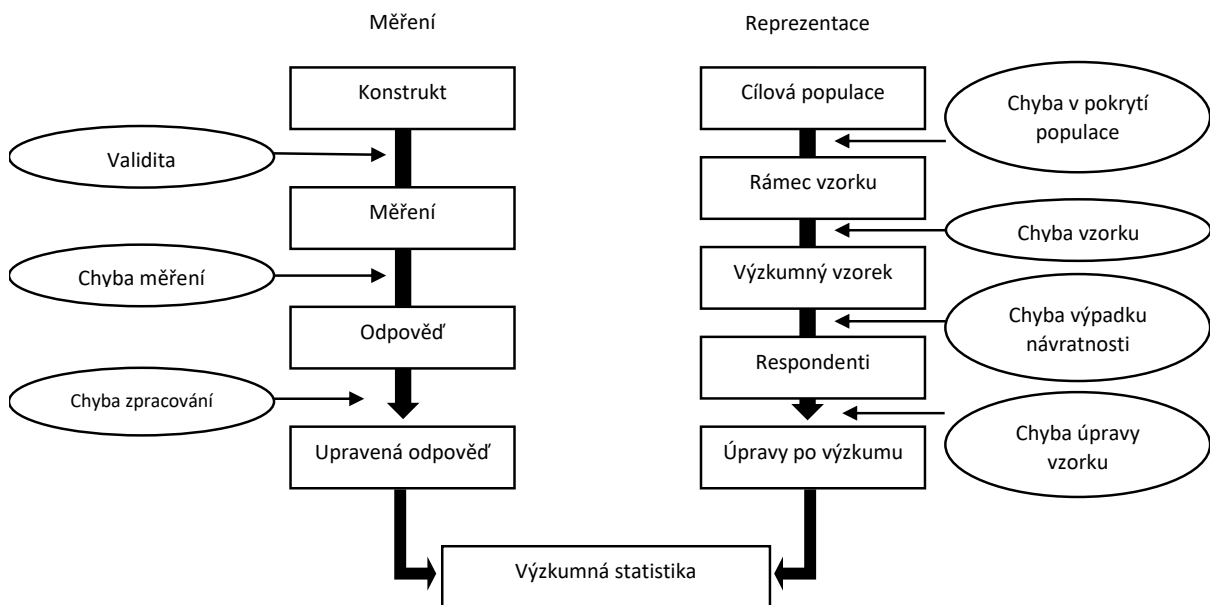
1. Odpovědi získané od respondenta musí korespondovat s jeho/jejími vlastními charakteristikami
2. Výzkumný vzorek musí být svými charakteristikami zástupný za cílovou populaci

Při nedodržení či odchýlení od těchto dvou podmínek vzniká výzkumná chyba (Groves, 2004, s. 40), která popisuje sumu všech chyb, které vznikají během designu výzkumu, sběru dat, práce s daty a jejich následnou analýzou (Biemer, 2016, s. 2). V neposlední řadě jsou tyto chyby navázané i na samotný mód sběru, který ovlivňuje samotné nastavení výzkumného instrumentu (Groves, 2004, s. 47).

2.3.1 Celková chyba měření

Koncept celkové chyby měření je definován životním cyklem výběrového šetření od prvotního návrhu dotazníku až po analýzu sebraných dat, jak je zobrazeno níže ve Schématu 1.

Schéma 1: Životní cyklus výzkumu z pohledu kvality a chyb



Zdroj: (Groves, 2004, s. 48)

Ve schématu jsou chyby v oválech, zatímco jednotlivé kroky výzkumu v kvádrech. Zatímco chyby spojené s měřením (levá větev Schématu 1) staví na roli individuálních respondentů, chyby spojené s reprezentací (pravá část Schématu 1) staví na statistických operacích. V následující sekci bude všech sedm chyb postupně popsáno, čímž dojde i k definici těch, které jsou klíčové pro tento výzkum.

2.3.1.1 Chyba validity

Chyba validity vzniká na základě rozdílu mezi výzkumným konstruktem a samotným měřením. Staví na teoretickém oboru psychometrie, který původně zkoumá a měří psychické jevy a jejich vnímání (Kaplan & Saccuzzo, 2017, s. 115). V praxi se tak jedná o vztah mezi skutečností, pravdivou hodnotou zkoumané proměnné a jejím obrazem např. v dotazníkovém šetření. Jedná se tedy o chybu vzniklou při snaze měřit samotný konstrukt, případně korelovat se skutečnou hodnotou (Groves, 2004, s. 50). Tato chyba může vznikat například při neúspěšném procesu konceptualizace a operacionalizace proměnných, kdy se na dotazovaný koncept neptáme vhodnými otázkami, tím pádem nezískáváme naměřené hodnoty náležící ke skutečné pravdivé hodnotě.

2.3.1.2 Chyba měření

Jamile je konstrukt, který má za úkol výzkum měřit, definován, vzniká další chyba, a to mezi měřením a odpovědí respondenta. Tato chyba je rozdílem mezi změřenou a očekávanou hodnotou a je spojována s reliabilitou výzkumu. Jedná se o podhodnocení či nadhodnocení dané hodnoty, např. při výzkumu spojeném s drogovou problematikou, kdy respondent může považovat kladnou odpověď na otázku „Užil jste v posledním měsíci nějaké drogy?“ za nežádoucí a raději tak zvolí odpověď „Ne“, i když se nejedná o pravdivou odpověď. Tato chyba je také navázána na mód výzkumu, kde míra anonymity respondenta může zvýšit či naopak snížit potřebu odpovídat tak, jak se očekává, nebo jak je společensky vnímáno jako žádoucí. Zároveň se liší podle zvoleného typu dotazování, kdy u osobního či telefonického dotazování může být tato potřeba vyšší, zatímco u online dotazníkového šetření dochází k jejímu snížení (Revilla, 2010, s. 151).

Pro tento výzkum je zajímavá i další úroveň ovlivňující tuto chybu, a to cílení výzkumu skrze platformu Facebook. Ačkoli je respondentovi garantována anonymita, samotný fakt, že pozvánku k výzkumu zahlédl skrze svůj osobní facebookový profil, může vést k navýšení zmíněné potřeby odpovídat na otázky tak, jak respondent považuje za společensky žádoucí

(např. z hlediska volební participace). Právě sociální desirabilita (snaha přizpůsobit svou odpověď společenským normám a mínění společnosti) je nedílně spjata s rolí tazatele, respektive jeho přítomností či nepřítomností v procesu sběru dat (Chylíková, 2011, s. 193–194).

Dalším aspektem spojeným s chybou měření je kognitivní zátěž, kterou mohou rozdílné módy sběru způsobovat. Pokud je mód autonomní, tedy respondent při odpovídání na otázky nemá k dispozici tazatele, je tato zátěž vyšší. Naopak při přítomnosti tazatele je tato zátěž nižší, jelikož je respondentovi tazatel k dispozici a v jednotlivých krocích kognitivního procesu, od porozumění otázce až k poskytnutí odpovědi, mu může pomoci (Schwarz, 2007, s. 283–284). Kognitivní zátěž tak může způsobit systematickou chybu měření (Jäckle et al., 2010, s. 18).

2.3.1.3 Chyba zpracování

Chyba zpracování vzniká po sebrání dat a jejich následném zpracování, a to skrze rozdíl v hodnotě naměřené a hodnotě očekávané. Dochází k ní při kontrole interní logiky dotazníku a úpravě odpovědi či jejímu odstranění z důvodu její domnělé nevalidity nebo také při chybném zakódování otevřené otázky způsobené lidskou chybou. Příkladem první z možností je odpověď na otázku „Kolik litrů alkoholu denně vypijete?“, která dosahuje velmi vysoké hodnoty spojené s pravidelnou frekvencí, ačkoli doprovodné statistiky respondenta nenapovídají nebo dokonce dokazují, že se nejedná o alkoholika. Při zpracování dat tak z hlediska zachování interní logiky a koherence dat může dojít ke smazání této odpovědi z datové matice, ačkoli je zcela validní a je způsobena tím, že respondent je zaměstnán jako ochutnávač piva.

2.3.1.4 Chyba v pokrytí populace

Následující čtyři chyby pokrývají chyby ovlivňující statistické operace na rozdíl od jednotek výběru. Pokrytí populace pracuje s předpokladem, že ideální výběrové šetření vybírá respondenty do výzkumného vzorku, který je všemi charakteristikami (a následně i všemi statistickými výpočty) shodný s cílovou skupinou, která je v těchto ohledech shodná se základní populací. K chybě poté dochází, pokud jedinec či skupina jedinců z cílové skupiny nemá možnost být vybrána do výzkumného vzorku. K tomuto může dojít například pokud respondent nemá telefon a mód výzkumu je telefonické dotazníkové šetření nebo respondent nemá přístup k internetu či není součástí panelu při online dotazníkovém šetření. Ačkoli se počty jedinců bez telefonu či přístupu k internetu stále snižují (Roser et al., 2015), jedná se o zásadní zdroj výzkumné chyby (Revilla, 2010, s. 151), zejména pokud se oslovená a

neoslovená skupina respondentů systematicky liší ve svých charakteristikách. Pokud se ale nejedná o systematické odchylky, je chyba způsobena odlišnými charakteristikami jednotlivých respondentů z výběrové populace, která může být výzkumem oslovena a která nemůže být výzkumem oslovena, vyřešena skrze agregaci statistik na celý výzkumný vzorek. Náhodné odchylky v tomto případě mají totiž tendenci vzájemně se vyrušit a na celkové makro úrovni zásadně neovlivňují kvalitu šetření (Groves, 2004, s. 52).

2.3.1.5 Chyba vzorku

Chyba vzorku je navázána na časovou a finanční náročnost výzkumů a popisuje úskalí, kde mají konkrétní respondenti nižší šanci na to být zahrnuti do výzkumného vzorku. Pokud má tato chyba systematický charakter, ohrožuje validitu dat. Pokud je nahodilá, ale dosahuje vysokého rozptylu (sampling variance), může být ohrožena reliabilita výzkumu (Groves, 2004, s. 57).

2.3.1.6 Chyba výpadku návratnosti

Výběrová šetření v českém kontextu bojují od pádu komunismu se stále nižší návratností (Krejčí, 2007, s. 567). Návratnost označuje počet dokončených dotazníků oproti velikosti oslovené populace. Tato chyba je ale pro výzkum ohrožující jen a pouze v případě, kdy se respondenti, kteří neposkytnou odpovědi, systematicky odlišují svými charakteristikami od těch, kteří odpovědi poskytnou.

Snaha o minimalizaci této chyby může být v určitých případech i kontraproduktivní (prokázáno empiricky a na základě různých módů sběru dat, např. Robert M. Groves, 2002, s. 254; Keeter et al., 2000, s. 147), snižování výpadku návratnosti je totiž časově i finančně velmi náročné a pokud nedochází k systematickým odchylkám ve dvou výše zmíněných vzorcích, je toto navyšování návratnosti pro kvalitu výzkumu bezpředmětné.

2.3.1.7 Chyba úpravy vzorku

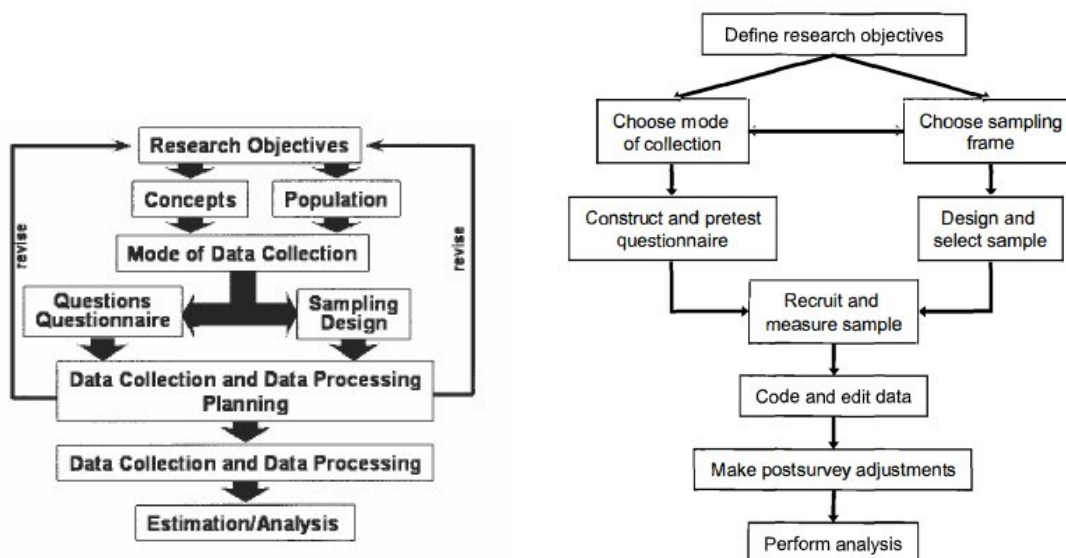
Poslední z chyb spadajících do teoretického konceptu celkové chyby měření vzniká při úpravě dat na základě výzkumného vzorku. K tomu může dojít, pokud se vzorek liší od základní populace ve svých popisných charakteristikách, například při snaze o reprezentativitu vzorku skrze jeho převážení. To samotné pozmění datovou strukturu a může vést k závažným

rozdílům na úrovni statistických výpočtů a interpretaci jejich hodnot (Soukup & Rabušic, 2007, s. 392).

2.4. Komplementární teorie chyb

Další teoretický koncept, který shodou okolností využívá stejné jméno Celková chyba měření a je komplementárním ke konceptu v předchozí kapitole, byl definován v roce 2003 (Biemer & Lyberg, 2003, s. 34). Tento koncept staví na obdobném vnímání výzkumného procesu, jak ukazuje Schéma 2.

Schéma 2: Srovnání výzkumného procesu, nalevo Biemer & Lyberg, napravo Groves



Zdroj: (Biemer & Lyberg, 2003, s. 27; Groves, 2004, s. 47)

Biemer a Lyberg ve svém uchopení analýzy kvality výzkumu a měření výzkumné chyby akcentují rozdělení na systematické a nesystematické zdroje chyb.

2.4.1 Nesystematické zdroje chyb

První ze dvou forem chyb je nesystematický rozptyl naměřených hodnot. Jedná se o chybu způsobenou v jednotlivých měřeních na úrovni respondentů agregovanou na kumulativní úroveň výzkumů. Pokud se respondent odchýlí svou odpovědí od předpokládané hodnoty, dá vzniknout pozitivní či negativní odchylce. V běžné situaci dojde na makro úrovni celého

výzkumu k vzájemnému vyrušení těchto odchylek a výsledná naměřená hodnota se výrazně neodlišuje od hodnoty očekávané – chyba je tím pádem velmi nízká (Biemer & Lyberg, 2003, s. 46). Navazujícím konceptem je pak koncept reliability. U výběrových šetření s malou nesystematickou chybou způsobenou nízkým či vzájemně se nulujícím rozptylem hodnot se očekává vysoká reliabilita a naopak.

Příkladem může být výzkum zabývající se stravováním ve fast foodech. Při dotazu „Kolikrát jste za poslední půl rok navštívil jakoukoli provozovnu fast food občerstvení?“ může respondent založit svou odpověď na nepřesné aproximaci. Namísto skutečného počtu uskutečněných návštěv vyjde z předpokladu, že za poslední týden navštívil fast food třikrát a znásobí tuto hodnotu 24, což je hrubý a nepřesný odhad počtu týdnů za půl roku. Výsledná hodnota je pak pozitivně či negativně odchýlena od skutečné hodnoty.

2.4.2. Systematické zdroje chyb

Druhým a do jisté míry navazujícím zdrojem chyb jsou pak chyby systematické. Ty vznikají při konzistentním vychýlení měřených hodnot od hodnot očekávaných. V praxi se tak jedná o dlouhodobé vychýlení (ať už pozitivního rázu, tj. když pozitivní odchylky svým počtem systematicky převyšují ty negativní, či naopak negativního rázu, kdy negativní odchylky systematicky převyšují pozitivní) (Biemer & Lyberg, 2003, s. 47).

Jako příklad systematické chyby může posloužit chyba měření z předchozí kapitoly. Při odpovědi na stejnou otázku stravování ve fast food restauracích může respondent zvažovat společensky preferovanou odpověď a systematicky tak podhodnotit jím poskytnuté hodnoty.

2.4.3 Efekty nesystematických a systematických chyb

Výše zmíněné chyby samozřejmě nepříznivě ovlivňují kvalitu výzkumu. Na úrovni ovlivňování statistických operací a výpočtů pak mají rozdílné druhy chyb různý vliv (Biemer & Lyberg, 2003, s. 49–50).

U lineárních statistických operací, tj. průměrů a proporcí, hraje stěžejní roli v ovlivnění kvality výsledků systematická chyba. U nesystematických chyb způsobených rozptylem naměřených hodnot se pracuje s předpokladem, že se na úrovni celkového datového souboru navzájem vyruší.

U nelineárních statistických operací, jako jsou korelace a regresní analýzy, nelze jednoduše vyvodit, který druh chyb je pro výsledné hodnoty operací více škodlivý. Oba dva typy chyb, tedy nesystematické a systematické chyby, mají různé negativní efekty a je tak v zájmu výzkumníka vynaložit snahu o jejich minimalizaci.

2.4.3 Střední kvadratická chyba

Závěrečným přínosem k uchopení chyby od Biemera a Lyberga je její matematické vyjádření. K tomu používají ukazatel střední kvadratické chyby (Mean Square Error) (Biemer & Lyberg, 2003, s. 62). Jedná se o reprezentaci všech zdrojů chyb ve výběrovém šetření, který stojí na součtu systematických chyb a nesystematickém rozptylu měřených hodnot.

V neposlední řadě autoři zmiňují, že snaha o minimalizaci hodnoty střední kvadratické chyby souvisí i s rovnováhou mezi opatřeními, které ji redukuje, a celkovou cenou a časovou náročností výzkumu.

2.5 Problematika (ne)pravděpodobnostních výběrů

Předposlední sekce teoretické části práce se zabývá problematikou metod výběru. Výše zmíněné teorie chyb, jejich komplementární části a doplnění totiž primárně staví čistě na pravděpodobnostních výběrech, ať už prostých náhodných či (vícestupňových) skupinkových (Groves, 2004, s. 57). Nicméně empirická část této práce analyzuje celkem tři módy sběru dat, přičemž jen jeden z nich (telefonické dotazníkové šetření, CATI) využívá pravděpodobnostní výběr. Zbývající dva módy sběru staví na kvótním nepravděpodobnostním výběru respondentů.

2.5.1 Krátká historie módů sběru dat

Pro nastavení kontextu následující sekce je nezbytné kompaktně shrnout historii výzkumných metod a módů sběru dat. V samotných počátcích současného kvantitativního výzkumu před první světovou válkou nebyly výzkumy prováděny formou výběrových šetření. Jejich zaměření bylo na úrovni případových studií a nevyužívali možností pravděpodobnostních výběrů (Alastalo, 2008, s. 28–29). První pokusy o generalizaci výsledků se ale datují již do

sedmnáctého století, kdy došlo k odhadu velikosti populace Londýna na základě parciálních dat na omezeném vzorku obyvatel (Bethlehem, 2009, s. 6).

Zlatá éra kvantitativního výzkumu přichází v období po druhé světové válce a trvá až do konce šedesátých let minulého století. Dominantní paradigma stojící za výzkumem je funkcionalismus, používané módy jsou zejména osobní dotazování a dotazníkové šetření rozesílané poštou (Alastalo, 2008, s. 32). Právě v této době vznikají dodnes využívané teorie kvality měření a chyb, postavené na základech aktuálně využívaných pravděpodobnostních výběrů. V osmdesátých letech je osobní dotazování nahrazeno telefonickým, v devadesátých letech přichází další inovace – nahrazení tazatele přednahrávanými otázkami (IVR CATI) (Leeuw et al., 2008, s. 2).

V posledních několika desetiletích můžeme v komunitě výzkumníků pozorovat diskuzi o nových metodách výběrů s odkazem na zvyšující se náklady a snižující se návratnost pravděpodobnostních výběrů, které byly a jsou impulsem pro hledání alternativních způsobů výběrů v poli nepravděpodobnostních metod. Jednou z navrhovaných cest představuje kvótní výběr, jako kvazireprezentativní vzorek vytvořený na základě dobře navrženého schématu nepravděpodobnostního výběru (Yang & Banamah, 2014, s. 29–30).

Konkrétně kvótní online sběr, který je porovnáván v této práci, se dostává do popředí až s rozvojem potřebné technologie a dostatečné penetraci populace přístupem k internetu (Dillman et al., 2014, s. 58). Zároveň se v současnosti nacházíme v době, kdy má výzkumník k dispozici doposud nejširší nabídku výzkumných možností z hlediska módů sběru dat a způsobů, jak nadesignovat samotný výzkum. Dopadem je i heterogenita výzkumných postupů a možností na úrovni jednotlivých regionů a zemí, která dosahuje zatím nejvyšší úrovně (Leeuw et al., 2008, s. 2).

Je tedy jasné, že v současnosti neexistuje optimální řešení, výzkumný mód, které by skloubil nízkou úroveň výzkumné chyby, dostatečné pokrytí populace bez výpadků návratnosti a efektivitu z hlediska časové a finanční náročnosti sběru samotného.

2.5.2 Přesnost dat u pravděpodobnostních a nepravděpodobnostních výběrů

Z hlediska přesnosti dat vzniká debata na úrovni komparace pravděpodobnostního a nepravděpodobnostního výběru. V současné době ale neexistuje mezi akademiky jednohlasný konsenzus. Existují empirické studie, které přisuzují výzkumům využívajícím

pravděpodobnostní výběry větší přesnost výsledků (Erens et al., 2014, s. 5; MacInnis et al., 2018, s. 728). Na opačné straně se ale ukazuje, že například v předvolebních průzkumech dopadají lépe nepravděpodobnostní výběry (Ansolabehere & Rivers, 2013, s. 321; Wang et al., 2015, s. 16). Oba tyto přístupy jsou tedy využívány a jejich přesnost lze z makro hlediska považovat za *srovnatelnou*.

2.5.3 Snižující se návratnost

Klíčovým tématem, které rezonuje akademickou a výzkumnou sférou a je úzce spojené s módy sběru a typy výběru, je postupné globální snižování návratnosti a s tím spojené riziko nárůstu systematické chyby spojené s výpadkem návratnosti, zejména u pravděpodobnostních výběrů.

V českém kontextu se tomuto tématu věnuje mimo jiné sociolog Jindřich Krejčí a popisuje úpadek návratnosti v období od devadesátých let minulého století do současnosti. Identifikuje dva hlavní důvody (Krejčí, 2007, s. 563):

1. Postupný nárůst objemu realizovaných výzkumů, zejména v oblasti výzkumu trhu a v marketingu. Populace je tak saturovaná množstvím výzkumů a zákonitě je i častěji odmítá.
2. Změna v životním stylu společnosti, kdy je stále obtížnější respondenty zastihnout.

Stejná situace platí i na globální úrovni (Williams & Brick, 2018, s. 207–208) a její řešení, stejně jako problematika ekvivalenčního výzkumu smíšených módů, je nesmírně komplexní a vyžaduje neprodlenou pozornost výzkumníků (Dillman, 2020, s. 15).

2.5.4 Vliv módů sběru dat na míru návratnosti

Problém snižující se návratnosti se přímočaře nepropisuje do nepravděpodobnostních sběrů – jmenovitě kvótního online sběru na panelu respondentů. Například Český národní panel měl v roce 2021 zhruba 60000 aktivních členů a tento objem se meziročně zvyšuje (ČNP, 2021). Díky stále rostoucímu objemu výzkumů jsou respondenti v panelu neustále vytěžováni a návratnost CAWI dotazníků tak zůstává stabilní (Keusch, 2015, s. 186), ačkoli je obecně nižší, než u pravděpodobnostních metod výběru (Porter & Whitcomb, 2004, s. 56).

Při bližším pohledu lze identifikovat hlavní faktor stabilní návratnosti a tím je incentiva (finanční či jiná odměna) za vyplněný dotazník, která je klíčovým faktorem ovlivňujícím návratnost

(Dillman, 2020, s. 22; Millar & Dillman, 2011, s. 265). Z tohoto hlediska lze očekávat, že módy sběru dat poskytující incentivu (v tomto případě CAWI) dosáhnou vyšší návratnosti než módy sběr dat bez ní (CATI a Social).

Velkou otázkou je ale systematická chyba spojená s chybou pokrytí populace (část společnosti je stále offline, zároveň část online populace není součástí panelu a může vykazovat jiné charakteristické rysy oproti panelové populaci). Touto problematikou se akademici zabývají (Keusch, 2015, s. 186; Stiglbauer et al., 2011, s. 186) a jedním z cílů této práce je zhodnotit, zdali je předložené riziko spojené se systematickou chybovostí v oblasti pokrytí populace reálné ve svých důsledcích a ovlivňuje přesnost dat získaných skrze nepravděpodobnostní kvótní online dotazování na panelu respondentů oproti výsledkům získaných skrze alternativní módy sběru dat.

2.5.5 Váha jednotlivých módů sběru dat v aplikovaném výzkumu veřejného mínění a výzkumu trhu

Téma váhy jednotlivých módů sběru dat je potřeba uvést do kontextu. Online kvótní dotazníkové šetření je v současné době velmi populární mód sběru dat v globálním sociologickém výzkumu (Keusch, 2015, s. 183) i v aplikovaném výzkumu trhu nejen v českém kontextu. Na ústupu jsou naopak ve výzkumu trhu módy osobního a telefonického dotazování, u digitálních módů je očekáván v budoucnu další růst (SIMAR, 2021a, s. 14–15).

Tabulka 1: Výzkumné metody podle realizovaného obrátu v ČR, vybrané evropské země a svět, 2019, výzkum trhu

Výzkumná metoda	Svět	Německo	USA	Polsko	ČR
Telefonický výzkum (CATI)	8 %	18 %	7 %	9 %	9 %
Osobní dotazování (PAPI, CAPI)	5 %	12 %	1 %	21 %	12 %
Online kvantitativní výzkum (CAWI)	17 %	42 %	16 %	19 %	24 %
Mobile/Smartphone online	9 %	8 %	10 %	0 %	6 %
Audience Measurement	4 %	0 %	5 %	6 %	3 %
Online Traffic/Web analytic	8 %	0 %	9 %	1 %	0 %
Automated digital/electronic	7 %	5 %	8 %	22 %	24 %

Social media monitoring	4 %	0 %	5 %	0 %	1 %
Ostatní kvantitativní výzkum	2 %	6 %	0 %	10 %	5 %
Celkem kvantitativní výzkum	64 %	91 %	61 %	88 %	84 %
Skupinové diskuze (FGD)	3 %	3 %	2 %	5 %	7 %
Hlubkové rozhovory (IDI)	2 %	2 %	1 %	3 %	5 %
Online kvalitativní metody	3 %	1 %	4 %	0 %	1 %
Mobile kvalitativní metody	1 %	0 %	1 %	0 %	0 %
Etnografie	1 %	1 %	1 %	0 %	1 %
Online komunity	3 %	1 %	4 %	1 %	1 %
Ostatní kvalitativní výzkum	1 %	1 %	0 %	0 %	0 %
Celkem kvalitativní výzkum	14 %	9 %	13 %	9 %	15 %
Ostatní, sekundární analýza dat	22 %	0 %	26 %	3 %	1 %

Zdroj: (ESOMAR, 2020; SIMAR, 2021a, s. 15)

Jak Tabulka 1 jasně ukazuje, České republika je z hlediska využívání online kvótního dotazování nad světovým průměrem, například i před USA. Tento mód sběru je tak důležité nepodceňovat, predikce budoucího vývoje se ubírají směrem k větší autonomii respondenta z hlediska vyplňování dotazníků a smíšenému módu sběru s integrací digitálních možností (Dillman, 2020, s. 38).

Zároveň se dá předpokládat, že odvětví výzkumu trhu se bude snažit o prosazování inovací v oblasti online módů sběru dat, zejména díky jejich efektivitě v oblasti nákladů finančních i časových (Dillman et al., 2014, s. 92). Koneckonců je v zájmu tohoto odvětví, aby byly opory výběru, například online panely, co nejvíc shodné ve svých charakteristikách se základní populací a získaná data co nejvíce kvalitní – z hlediska validity, reliability i reprezentativity. Postupně se rozšiřující přístup k internetu na úrovni celé společnosti a její postupná digitalizace nabízí ideální podmínky.

Na druhou stranu nelze upřít v současnosti stále silnou pozici i módu osobního (PAPI, CAPI) a telefonického (CATI) dotazování. Dle Tabulky 1 byl objem sumy těchto výzkumů v roce 2019 srovnatelný s online výzkumy na úrovni České republiky a překonal online výzkumy v Polsku. Tyto výzkumy skýtají výhodu v oblasti pokrytí celé populace a zároveň jsou v některých oblastech (například výzkum nových prototypů produktů, car clinic výzkum) v současné chvíli nenahraditelné.

Výzkum s oporou dat na úrovni sociálních sítí zatím nemá signifikantní roli z hlediska výzkumu veřejného mínění a trhu. Jeho role do budoucna je tak v současné chvíli nejasná, nicméně závěry této práce mohou pomoci k objasnění jeho silných a slabých stránek.

2.6 Shrnutí teoretické sekce – vliv módu

V předešlých kapitolách byla představena definice smíšených módů sběru a definice konceptu ekvivalence měření a přesnosti měření. Dále byl čtenář seznámen s teorií hodnocení kvality výzkumu skrze systematické a nesystematické chyby, které mohou v celém výzkumném designu vznikat. V neposlední řadě bylo poukázáno na problematiku pravděpodobnostních a nepravděpodobnostních výběrů a váhy jednotlivých módů sběru dat v aplikovaném výzkumu veřejného mínění a výzkumu trhu.

Před vstupem do metodologické sekce se nabízí shrnutí doposud představených teoretických východisek a jejich vliv na jednotlivé módy sběru dat. Pro tento účel byla vytvořena tabulka znázorňující srovnání jednotlivých módů – CATI, CAWI Panel a CAWI Facebook.

Tabulka 2: Srovnání módů sběru dat mezi sebou, vycházející z uvedených teoretických konceptů

Charakteristiky módu	CATI	CAWI Panel	CAWI Facebook
	Náhodně generovaná telefonní čísla (RDD)	Online panel	Sociální síť Facebook
Opora výběru			
Vliv tazatele	Ano, nižší než u CAPI	Ne, autonomní dotazování	Ne, autonomní dotazování
Kognitivní zatížení	Nižší	Vyšší	Vyšší
Sociální desirabilita	Vyšší	Nižší	Nižší
Pokrytí populace	Vyšší	Nižší	Nižší
Potenciál pro chybu v pokrytí populace	Nižší	Vyšší	Vyšší
Potenciál pro chybu výpadku návratnosti	Vyšší	Nižší	Vyšší
Předpoklad přesnosti měření	Vyšší	Nižší	Nižší

Klesající návratnost	Ano, dlouhodobě	Stabilní	Neznámá
Náklady (finanční, časové)	Vysoké	Nízké	Nízké

Zdroj: Vlastní

Před samotným závěrečným shrnutím teoretické části je nutno akcentovat, že veškeré chyby měření, které jsou v Tabulce 2 uvedeny a které mohou stát za přítomností či nepřítomností rozdílů z hlediska ekvivalence měření zkoumané v analytické části, jsou chyby na úrovni jednotlivých módů sběru dat, nikoli chyby společné pro všechny módy sběru dat. Ekvivalence měření se totiž zaměřuje právě na poukázání rozdílů mezi jednotlivými výběrovými soubory, nikoli na hledání společných chyb měření napříč.

Z tohoto porovnání (Tabulka 2) vychází potenciálně větší míra systematické chyby měření u online sběru na uživatelských sítích Facebook (Social), zejména v oblasti chyby v pokrytí populace a chyby výpadku návratnosti. U telefonického dotazování poté hrozí vyšší riziko u chyby měření, zejména z hlediska sociální desirability, kde vliv tazatele může ovlivnit odpověď respondenta (Chylíková, 2011, s. 193). Naopak nepřítomnost tazatele u CAWI sběrů zvyšuje kognitivní zátěž, tedy riziko systematické chyby měření. Přesnost měření je klíčovým tématem v této práci a díky riziku systematické chyby pokrytí populace se dá předpokládat, že bude u online kvótních nepravděpodobnostních výběrů nižší než u pravděpodobnostního telefonického dotazování. V neposlední řadě je třeba upozornit na to, že veškeré chyby měření lze systematicky snižovat (v ideálním případě se přiblížit k jejich eliminaci), nicméně se jedná o velmi náročný proces, který může být i kontraproduktivní vzhledem k vynaloženým prostředkům (Groves, 1989, s. 6). Časová a finanční efektivita módu sběru dat je tedy podstatným kritériem pro jejich srovnání a jedná se o aspekt, ve kterém dlouhodobě vede kvótní online dotazníkové šetření, ačkoli standardně nabízí respondentům incentivu za vyplnění dotazníku, čímž na druhou stranu značně stimuluje návratnost.

3 Metodologická část

Metodologická část práce se dělí na tři základní podsekce. První podsekce čtenáře seznámí s výzkumným designem všech tří módů sběru dat a představí strukturu každého výběrového souboru před a po převážení dat. Druhá část rekapituluje výzkumnou otázku a hypotézy a dále rozvíjí pojem ekvivalence měření na úrovni analytických postupů jejímu potvrzení či vyvrácení. Poslední podsekce poté představuje spojnici mezi metodologickou a analytickou částí a definuje postupy a analytické nástroje a jejich relevanci vzhledem k zaměření práce.

3.1 Design výzkumu

Sběr dat pro všechny tři módy probíhal shodně na území České republiky v 39. týdnu v roce 2020, tedy od 21. do 27. září, přesně týden před volbami do krajských zastupitelstev. Data byla sebrána agenturou NMS Market Research.

Výzkum byl primárně zaměřen na predikci výsledků krajských voleb, základní populace byla stanovena na občany České republiky starší 18 let, kteří mohou volit v nadcházejících volbách a pocházejí ze všech třinácti krajů České republiky vyjma Hlavního města Prahy, kde v roce 2020 neprobíhaly volby do krajských zastupitelstev.

Dotazník byl ve všech třech módech sběru shodný na úrovni znění otázek a možností odpovědí, lišil se jen v rámci pokynů pro respondenty a tazatele tak, aby reflektoval mód sběru využívající tazatele.

3.1.1 Telefonické dotazování (CATI)

Telefonické dotazování (CATI) probíhalo skrze prostý náhodný výběr, oporou výběru byla databáze náhodně vygenerovaných čísel (tento způsob výběru se také nazývá RDD – random digit dialing). Dotazování provádělo celkem pět proškolených tazatelů. Respondenti byli kontaktováni jak v dopoledních, tak v odpoledních a večerních hodinách. Tento mód sběru dat byl vybrán záměrně kvůli porovnání s nepravděpodobnostními výběry, zároveň probíhající epidemie COVID19 značně ztěžovala využití jiného typu pravděpodobnostního náhodného módu sběru dat ve formě osobního dotazování. Dotazování mohl úspěšně dokončit každý respondent starší 18 let s českým občanstvím, který mohl na podzim roku 2020 volit v krajských volbách.

3.1.2 Online dotazníkové šetření na panelu (CAWI)

Online dotazníkové šetření na panelu respondentů probíhalo skrze nepravděpodobnostní kvótní výběr. Oporou výběrů byl Český národní panel, jehož popisu je věnována následující podkapitola. Kvótní předpis byl nastaven dle dat Českého statistického úřadu z roku 2017 a 2018 na úrovni provázané kvóty pohlaví a věku, vzdělání, velikosti místa bydliště a regionu. Tento předpis byl upraven pro účely výzkumu pro obyvatele všech krajů krom Hlavního města Prahy. Dotazování probíhalo formou online dotazníku na vlastní dotazovací platformě NMS Market Research, do které obdrželi respondenti pozvánku do e-mailové schránky. Dotazník bylo možné vyplnit kdykoli v průběhu sběru dat, včetně možnosti přerušit dotazování a vrátit se k němu posléze. Plnění kvót bylo průběžně sledováno, skrze cílení v panelu a uzavírání jednotlivých kvótních skupin byl průběžně sběr upravován za cílem přiblížení struktury výběrového vzorku z hlediska její reprezentativity k obecné populaci České republiky ve věku 18 let a více. Za vyplnění dotazníku nenáležela respondentovi žádná incentiva.

3.1.2.1 Český národní panel

Český národní panel je opora výběru, online panel respondentů patřící do skupiny Evropských národních panelů (panely operují v Polsku, na Slovensku, v Maďarsku a v Bulharsku) a spoluvlastněný agenturou NMS Market Research (spolu s agenturou STEM/MARK a Nielsen Admosphere). Panel v současné době disponuje databází 60 000 aktivních respondentů a průměrnou mírou návratnosti 67 % (ČNP, 2021).

Rekrutace do panelu probíhá skrze online i offline kanály. Identita každého respondenta je v rámci registrace do panelu fyzicky ověřena (telefonicky či poštou). Panel dlouhodobě implementuje mechanismy pro kontrolu kvality respondentů. Respondenti mohou být označeni jako nespolehliví například pokud je jejich doba průchodu dotazníkem signifikantně kratší než medián této hodnoty na úrovni všech respondentů či pokud neposkytují validní odpovědi v otevřených otázkách. Takoví respondenti jsou varováni a opakované porušování standardů kvality může vést až k jejich vyřazení z panelu (ČNP, 2021).

Panel disponuje základními sociodemografickými charakteristikami každého respondenta. Zároveň obsahuje i množství subpanelů. Je tedy možné cílit na konkrétní skupiny respondentů jak na základě sociodemografie, tak na základě dalších charakteristik dle subpanelů (například mediální chování, finanční situace, trávení volného času) (ČNP, 2021).

Respondenti jsou za úspěšně dokončený dotazník (včetně dotazníku, který byl ukončen kvůli naplnění kvót či respondent nesplňoval podmínky cílení) odměněni ve formě bodů, které mohou směnit za peníze, odměny nebo je věnovat na charitativní účely. Každý respondent může dle standardů SIMARU dokončit maximálně 2 výzkumy týdně, respektive 36 ročně (SIMAR, 2021b, s. 6).

3.1.3 Online dotazníkové šetření na sociální síti Facebook (Social)

Poslední z módů sběru je do jisté míry experimentální. Ačkoli již v minulosti proběhly výzkumy na specifických sociálních a komunitních sítích (Brickman Bhutta, 2012, s. 57; Wang et al., 2015, s. 3), nejedná se o standardně využívaný mód sběru dat. Část dotazování probíhala stejně jako u online dotazníkového šetření na panelu, tedy formou online dotazníku na vlastní dotazovací platformě NMS Market Research. Výběr respondentů byl nepravděpodobnostní kvótní a probíhal skrze cílenou reklamní kampaň na síti Facebook. Sociální síť Facebook dovozovala v době sběru cílit pouze skrze kombinaci údajů o kraji, věku a pohlaví. Na kvótu vzdělání a velikosti místa bydliště nebylo možné proaktivně cílit, jen uzavřít případně saturované kategorie na úrovni dotazníku a respondenty, kteří do nich spadají, následně z dotazníku vyřadit. To vedlo k přesběru či nedosbírání dílčích kategorií na úrovni vzdělání a velikosti místa bydliště. Za vyplnění dotazníku nenáležela respondentovi žádná incentiva.

3.1.4 Struktura výběrových souborů

Jak bylo v předešlých podkapitolách zmíněno, jednotlivé módy sběru dat se liší na mnoha úrovních, přičemž klíčovým rozdílem pro strukturu vzorku je typ výběru. Tabulka 3 ukazuje rozložení jednotlivých sociodemografických ukazatelů u jednotlivých módů sběru dat spolu s kvótním předpisem.

Tabulka 3: Struktura výběrových souborů a kvótní předpis

	Kvótní předpis	CATI	CAWI	Social
Pohlaví				
Muži	49 %	52 %	49 %	49 %
Ženy	51 %	48 %	51 %	51 %

Věk				
18-24	8 %	5 %	11 %	12 %
25-34	16 %	10 %	17 %	16 %
35-44	20 %	17 %	18 %	20 %
45-54	17 %	17 %	17 %	16 %
55-64	15 %	16 %	18 %	16 %
65+	24 %	35 %	20 %	21 %
Vzdělání				
Základní	15 %	8 %	11 %	5 %
Střední bez maturity	34 %	26 %	34 %	28 %
Střední s maturitou	33 %	40 %	35 %	44 %
Vysokoškolské	17 %	26 %	19 %	23 %
Velikost místa bydliště				
Do 1 999 obyvatel	30 %	30 %	27 %	21 %
2 000 až 4 499 obyvatel	14 %	11 %	15 %	13 %
5 000 až 19 999 obyvatel	21 %	23 %	23 %	19 %
20 000 až 49 999 obyvatel	14 %	13 %	15 %	15 %
50 000 až 99 999 obyvatel	9 %	9 %	10 %	14 %
100 000 a více obyvatel	11 %	14 %	11 %	18 %
Kraj				
Středočeský	15 %	15 %	16 %	10 %
Jihočeský	7 %	6 %	7 %	9 %
Plzeňský	6 %	6 %	8 %	10 %
Karlovarský	3 %	2 %	3 %	5 %
Ústecký	9 %	9 %	9 %	7 %
Liberecký	5 %	6 %	5 %	7 %
Královéhradecký	6 %	7 %	6 %	5 %
Pardubický	6 %	5 %	5 %	5 %
Vysočina	6 %	4 %	6 %	5 %
Jihomoravský	13 %	13 %	12 %	13 %
Olomoucký	7 %	6 %	6 %	7 %
Zlínský	6 %	7 %	6 %	6 %
Moravskoslezský	13 %	14 %	11 %	12 %

Zdroj: Vlastní

Kvótní předpis je standardně využíván v agentuře NMS Market Research pro online kvótní dotazníkové šetření i další módy sběru dat a udává kvóty v daných sociodemografických ukazatelích. Jedná se o pohlaví provázané s věkem, vzděláním, velikostí místa bydliště a krajem. V tomto případě je předpis upraven tak, aby reflektoval rozložení vzorku a jeho reprezentativitu vůči obecné populaci České republiky bez obyvatel Hlavního města Prahy. Kvótní předpis byl vytvořen na základě údajů Českého statistického úřadu z let 2017 a 2018.

Tabulka 3 jasně poukazuje na několik klíčových rozdílů na úrovni jednotlivých módů sběru dat. U CATI sběru skrze prostý náhodný výběr jsou nadreprezentováni respondenti z nejstarší věkové kategorie. U CAWI a Social sběru naopak i přes kvótní nepravděpodobnostní výběr dochází k jejich mírné podreprezentovanosti. Na úrovni nejvyššího dosaženého vzdělání je poté viditelný rozdíl u kategorie základního a vysokoškolského vzdělání. Skupina respondentů se základním vzděláním je podreprezentována ve všech třech výběrových vzorcích, nicméně nejsignifikantnější je tento rozdíl mezi skutečným rozdělením a rozdělením ve vzorku u Social módu sběru dat. Naopak nadreprezentováni jsou vysokoškoláci, a to opět u všech třech módů sběru dat. U pohlaví, velikosti místa bydliště a kraje jsou rozdíly minimální a zanedbatelné.

Z tohoto důvodu byly všechny tři výběrové soubory převáženy. Pro určení váhy byl použit datový soubor Českého statistického úřadu ze čtvrtého čtvrtletí roku 2019 s provázaným rozdělením vzdělání a věku (Český statistický úřad, 2020). Aplikace vah pouze na úrovni těchto dvou proměnných byla zvolena ze čtyř hlavních důvodů:

1. Opora pro vytvoření vah u těchto dvou provázaných proměnných byla v době sběru tou nejaktuálnější vydanou a nejlépe tak reflektovala vzdělanostní rozložení dle věku v české společnosti.
2. Ostatní proměnné nevykazují obdobnou míru rozdílu oproti použitému kvótnímu předpisu.
3. Věk a vzdělání jsou klíčovými faktory pro politickou participaci (Linek, 2015, s. 87) a proto musí být jednotlivé výběrové soubory ve struktuře těchto proměnných ekvivalentní pro potřeby provedení analýzy.
4. V neposlední řadě vzniká vždy při vážení dat riziko spojené s vytvořením rozdílů na úrovni statistických výpočtů a interpretaci jejich hodnot (Soukup & Rabušic, 2007, s. 392), to může vést k paradoxní situaci, kdy zapojením více vah a následným převážením dat analyzujeme diametrálně odlišný výběrový soubor na úrovni konzistence dat a latentních vzorců, které je ovlivňují výměnou za jeho větší reprezentativitu.

Výsledná struktura převážených výběrových souborů je zobrazena v Tabulce 4. Převážením dle provázaných kvót na vzdělání a věk došlo k mírnému posunu i v rozložení ostatních

sociodemografických ukazatelů, nicméně jak již bylo zmíněno výše, odchylky mezi módy sběru dat jsou marginální a pro účely následujících analýz zanedbatelné.

Tabulka 4: Struktura vážených výběrových souborů

	CATI	CAWI	Social
Pohlaví			
Muži	51 %	50 %	48 %
Ženy	49 %	50 %	52 %
Věk			
18-24	8 %	8 %	8 %
25-34	16 %	16 %	16 %
35-44	19 %	19 %	19 %
45-54	18 %	18 %	18 %
55-64	15 %	15 %	15 %
65+	25 %	25 %	25 %
Vzdělání			
Základní	11 %	11 %	11 %
Střední bez maturity	34 %	34 %	34 %
Střední s maturitou	35 %	35 %	35 %
Vysokoškolské	20 %	20 %	20 %
Velikost místa bydliště			
Do 1 999 obyvatel	32 %	27 %	22 %
2 000 až 4 499 obyvatel	12 %	15 %	13 %
5 000 až 19 999 obyvatel	23 %	22 %	19 %
20 000 až 49 999 obyvatel	13 %	16 %	15 %
50 000 až 99 999 obyvatel	9 %	10 %	13 %
100 000 a více obyvatel	12 %	11 %	17 %
Kraj			
Středočeský	13 %	14 %	10 %
Jihočeský	7 %	6 %	8 %
Plzeňský	5 %	6 %	10 %
Karlovarský	3 %	3 %	4 %
Ústecký	9 %	9 %	7 %
Liberecký	6 %	5 %	7 %
Královéhradecký	7 %	6 %	5 %

Pardubický	6 %	6 %	6 %
Vysočina	5 %	5 %	5 %
Jihomoravský	13 %	12 %	13 %
Olomoucký	6 %	7 %	6 %
Zlínský	7 %	6 %	6 %
Moravskoslezský	13 %	14 %	13 %

Zdroj: Vlastní

3.2 Analytické přístupy k ekvivalence měření

Druhá sekce metodologické části práce se zaměřuje na analytické přiblížení konceptu ekvivalence měření a následně přibližuje relevanci zvolených výzkumných hypotéz.

Jak již bylo uvedeno v teoretické části, ekvivalence měření (či srovnatelnost měření) je koncept, který je využíván ke zkoumání neměnnosti výzkumných postupů (Anýžová, 2013, s. 30). Ekvivalence měření je využívána například u srovnání mezinárodních výzkumů či u longitudinálních výzkumů, nicméně lze ji využít i u analýzy srovnatelnosti rozdílných módů sběru dat (Pilecká, 2016, s. 31). Právě v tomto ohledu dokáže analýza ekvivalence měření, zaměřená na rozdílné módy sběru, poukázat na chyby spojené s danými módy, které ovlivňují výsledky jinak na první pohled srovnatelných výzkumů (Davidov et al., 2014, s. 13).

Z hlediska analytických nástrojů a přístupů lze analýzu ekvivalence měření rozdělit na dvě větve, přístupy:

1. První přístup staví ekvivalenci měření na úroveň analýz prvního a druhého stupně, zaměřuje se na základní otázku, zdali jsou data sebraná různými *způsoby* (ať už v případě sběru mezi zeměmi, v čase či skrze jiný mód) srovnatelná (Anýžová, 2013, s. 41).
2. Pokud se ukáže, že data srovnatelná jsou, nebo je potřeba hlubší analýza, nabízí se pokročilé statistické metody analýzy ekvivalence měření. Mezi tyto mnohorozměrné statistické analýzy patří například strukturní modelování (MGCF), Multitrait-Multimethod (MTMM), technika modelování proporcionálních šancí (Jäckle et al., 2010, s. 8), teorie odpovědi na otázku (IRT) či analýza latentních tříd (LCM), (Anýžová, 2013, s. 44–49; Davidov et al., 2014, s. 16–18; Vinopal, 2008, s. 86).

První větev analytického přístupu ke zkoumání ekvivalence měření je tedy základním ukazatelem ekvivalence na úrovni srovnávání četností či průměrů. Tato práce staví částečně analýzu právě na těchto postupech a ačkoli nevyužívá plně mnohorozměrných statistických

analýz z druhé větve přístupů ke zkoumání ekvivalence, staví zároveň i na zkoumání latentních faktorů ovlivňujících politickou participaci.

3.2.1 Role hypotéz v analýze ekvivalence měření

Po základní definici výzkumných přístupů k měření ekvivalence je nutné vrátit se zpět na začátek a uvést do kontextu výzkumné hypotézy.

H₀1: Data sebraná odlišnými módy sběru dat, tedy CATI, CAWI Panel (CAWI) a CAWI Facebook (Social), jsou ekvivalentní z hlediska politické preference.

H₀2: Data sebraná odlišnými módy sběru dat, tedy CATI, CAWI Panel (CAWI) a CAWI Facebook (Social), jsou ekvivalentní z hlediska politické participace.

H₀3: Data sebraná odlišnými módy sběru dat, tedy CATI, CAWI Panel (CAWI) a CAWI Facebook (Social), stejnou měrou reprezentují skutečný stav z hlediska politické participace a preference (jsou tedy stejně validní).

První dvě hypotézy se v tomto případě zaměřují na ekvivalenci na úrovni politické preference a politické participace. Pokud se tedy ukáže, že jednotlivé módy sběru dat jsou v těchto ohledech srovnatelné, jejich naměřené hodnoty se mezi sebou neliší, případně neliší signifikantně, budou tyto nulové hypotézy potvrzeny.

Poslední hypotéza se zaměřuje na validitu dat. Porovnává tak naměřené hodnoty u jednotlivých módů sběru oproti skutečné pravdivé hodnotě.

V neposlední řadě je na tomto místě nutné vzhledem k hypotézám upřesnit názvosloví. Pojem politické participace je v této práci a na datech získaných z výzkumu NMS Market Research používán jako synonymum pro volební účast (ať už predikovanou či deklarovanou pro již proběhlé volby). Politická preference je poté používána jako synonymum pro voličskou preferenci, tedy model, do kterého vstupují pouze voliči, kteří deklarují svoji volbu z hlediska strany, které chtějí dát či již dali svůj hlas.

3.3 Analytický postup

Jak se tedy dostat k odpovědi na výzkumnou otázku o srovnatelnosti jednotlivých módů sběru dat? Ačkoli byl výzkum provedený agenturou NMS Market Research na podzim roku 2020 zaměřen na krajské volby, nelze pro analýzu využít naměřené deklarované hodnoty pro jednotlivé strany v jednotlivých krajích kvůli nedostatečné velikosti vzorku. Analýza tedy probíhá na úrovni agregovaných dat za celý soubor. Z tohoto hlediska se nabízejí následující analytické postupy:

První analýza staví na deklarované volbě ve volbách do poslanecké sněmovny v roce 2017. Z pohledu politických výzkumů se jedná o voličskou preferenci (CVVM, Sociologický ústav AV ČR, b.r.) v tomto případě je ale deklarovaná vůči již proběhlým volbám a dále je v textu práce označována jen jako politická preference. Analýza staví na následující otázce:

Zúčastnil(a) jste se posledních voleb do poslanecké sněmovny, které proběhly v říjnu 2017?
Pokud ano, koho jste volil/a?

Vyberte jednu z uvedených možností.

- 1 Voleb jsem se nezúčastnil/a
 - 2 ANO 2011
 - 3 Občanská demokratická strana (ODS)
 - 4 Česká pirátská strana (Piráti)
 - 5 Svoboda a přímá demokracie (SPD)
 - 6 Česká strana sociálně demokratická (ČSSD)
 - 7 Komunistická strana Čech a Moravy (KSČM)
 - 8 KDU-ČSL
 - 9 TOP 09
 - 10 Starostové a nezávislí (STAN)
 - 11 Svobodní
 - 12 Strana zelených
 - 13 Jiná strana nebo hnutí
 - 14 Nemohl/a jsme volit, nebylo mi ještě 18
 - 15 Nevím, nepamatuji si, koho jsem volil/a
 - 16 Odmítám odpovědět
-

Na rozdíl od volebních modelů či vícekrokových voličských preferencí staví tato volební preference jen na jedné otázce. Jako chybějící odpovědi (missing values) jsou označeny

odpovědi 14 až 16. Do modelu politické preference tedy vstupují jen odpovědi 2 až 13, tedy jen ti respondenti, kteří deklarovali participaci na volbách do poslanecké sněmovny v roce 2017 a určili stranu, kterou volili. Analýza politické participace poté probíhá na úrovni srovnání výsledků za jednotlivé módy sběru dat a určení signifikance rozdílů mezi nimi.

Dalším krokem je porovnání těchto hodnot s reálnými volebními zisky jednotlivých stran ve volbách do poslanecké sněmovny v roce 2017. Tato skutečná pravdivá hodnota bere v potaz jen výsledky za kraje obsažené v analýze, tedy všechny kraje kromě Hlavního města Prahy. Tato analýza slouží k hodnocení přesnosti (a do jisté míry i validity) jednotlivých měřících nástrojů a navazuje na ni srovnání absolutních odchylek, kvadratických odchylek a adjustovaných odchylek jednotlivých módů od skutečné pravdivé hodnoty.

Doprovodnou analýzou je srovnání i se třemi volebními modely publikovanými v době měření (tedy v září roku 2020), respektive jejich průměrem. Jedná se o volební modely CVVM (Pilnáček, 2020), STEM (STEM, 2020) a Trendy Česka od agentury Kantar (Šabotník et al., 2020). Tyto modely mají pouze indikativní roli, jelikož sběr probíhal na celorepublikové úrovni a jedná se o skutečné volební modely. Slouží ale jako dokreslení aktuální společenského politického diskurzu a posunu voličů a voličských základů mezi lety 2017 a 2020.

Na stejné otázce z dotazníku staví i další analýza, jmenovitě analýza politické participace. Opět se jedná o deklarovanou hodnotu vzhledem k volbám do poslanecké sněmovny v roce 2017. Účast ve volbách je zde kódována jako jakákoli odpověď v rozmezí 2 až 13, jako neúčast je poté brána odpověď 1 a chybějící hodnoty jsou odpovědi 14 až 16. Nechybí srovnání mezi jednotlivými módy a se skutečnou pravdivou hodnotou volební účasti za všechny regiony kromě Hlavního města Prahy. Doprovodně jsou dodány i výsledky z výše zmíněných volebních modelů, které ale staví shodně s politickou preferencí na hypotetické situaci, kdy by volby proběhly následující týden.

Druhým krokem je analýza latentních vzorců ovlivňujících politickou participaci v krajských volbách 2020 a volbách do poslanecké sněmovny 2017. Tato analýza staví na předchozích výzkumech, které jasně definovaly určující faktory mající vliv na politickou participaci jedince v českém kontextu. Nástrojem pro provedení analýzy je binární logistická regrese a do analýzy vstupují závislé proměnné deklarované proběhlé politické participace ve volbách do poslanecké sněmovny 2017 (definovaná výše) a deklarované nadcházející politické participace ve volbách do krajských zastupitelstev v roce 2020:

Chystáte se volit v nadcházejících krajských volbách, které proběhnou 2. a 3. října 2020?

Vyberte jednu z uvedených možností.

- 1 Rozhodně ano
 - 2 Spíše ano
 - 3 Spíše ne
 - 4 Rozhodně ne
-

Jako politicky participující jsou definováni respondenti, kteří vybrali možnosti 1 nebo 2. Neparticipující pak 3 nebo 4. Mezi nezávislé proměnné vstupuje do modelu logistické regrese pohlaví, věk, vzdělání, zaměstnání (jako zaměstnání jsou definováni zaměstnanci a podnikatelé), velikost místa bydliště a předchozí či budoucí deklarovaná politická participace.

Politická participace ve volbách do krajských zastupitelstev vychází z následující otázky:

Zúčastnil(a) jste se posledních Krajských voleb v roce 2016?

Vyberte jednu z uvedených možností.

- 1 Ano, volil/a jsem
 - 2 Ne, nevolil/a jsem
 - 3 Nemohl/a jsme volit, nebylo mi ještě 18
 - 4 Nevím, nepamatuji si
-

Jako voliči byli definováni respondenti, kteří zvolili možnost 1. Nevoliči poté ti, kteří zvolili možnost 2. Možnost 3 až 4 byla kódována jako chybějící hodnota a nevstupovala do analýzy. Tato analýza má za cíl poukázat na konzistenci dat a přítomnost či absenci vlivu určujících faktorů na politickou participaci, jejichž vliv je v české společnosti prokázán.

V neposlední řadě závěr analytické části porovnává doprovodné statistiky časové a finanční efektivity jednotlivých módů sběru dat.

Analýzy byly zpracovány v programu SPSS a interním nástroji NMS Market Research. Z tohoto důvodu jsou v některých výstupech analýz vyznačovány p hodnoty v kategoriích dle hladiny významnosti testu a v některých je p hodnota uvedena číselně. Agentura NMS Market Research je členem sdružení SIMAR a ESOMAR a analyzovaná data byla sebrána v souladu se standardy kvality obou sdružení.

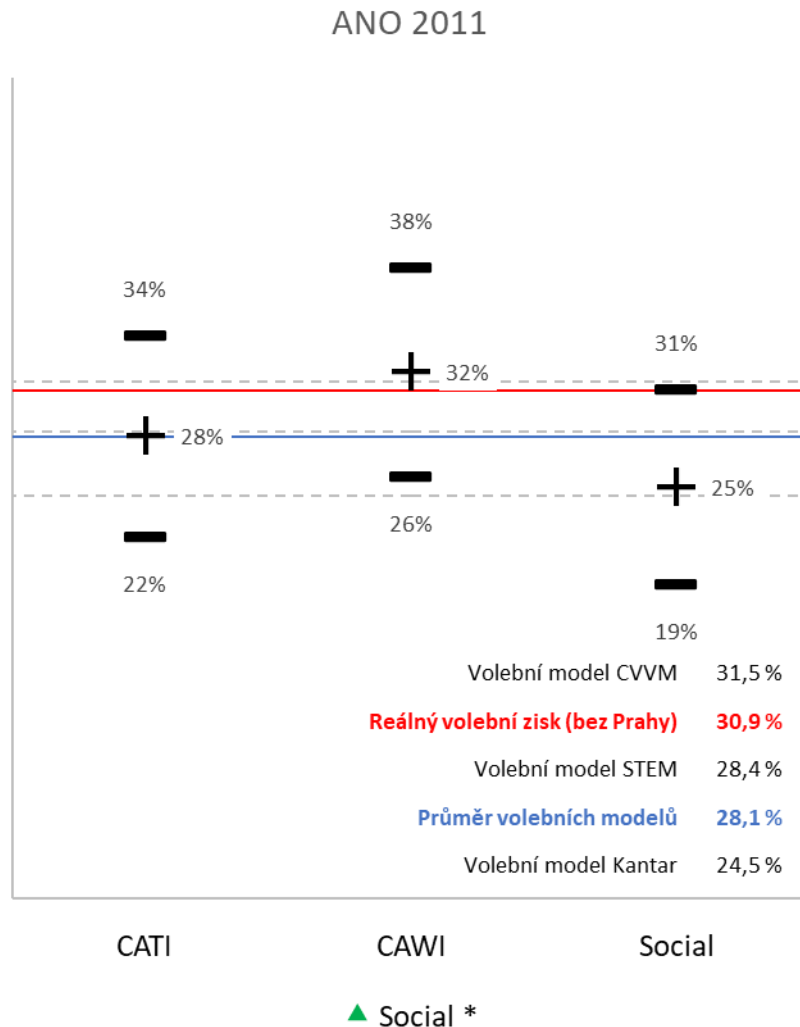
4 Analytická část

Tato část je rozdělena do čtyř základních částí a jejím cílem je potvrdit či vyvrátit testované hypotézy a zodpovědět tak na výzkumný problém této práce, zdali jsou jednotlivé módy sběru dat ekvivalentní. V první části je čtenáři představena analýza přesnosti měření na úrovni jednotlivých módů sběru dat skrze srovnání politických preferencí a politické preference. V druhé části je testována ekvivalence měření jednotlivých módů sběru dat pomocí analýzy faktorů ovlivňujících politickou participaci, a to na latentní úrovni dat, tedy korelací jednotlivých faktorů se závislou proměnnou, kterou je v tomto případě politická participace. Třetí podsekcce analytické části se poté věnuje dalším doprovodným analýzám a nakonec poslední část představuje syntézu dosavadních zjištění a předchází závěrečné diskuzi v poslední kapitole této práce.

4.1.1 Analýza přesnosti měření deklarované politické preference

První analýzou je srovnání deklarované politické preference ve volbách do poslanecké sněmovny v roce 2017.

Graf 1: Politická preference dle módů, volby do poslanecké sněmovny 2017, ANO 2011



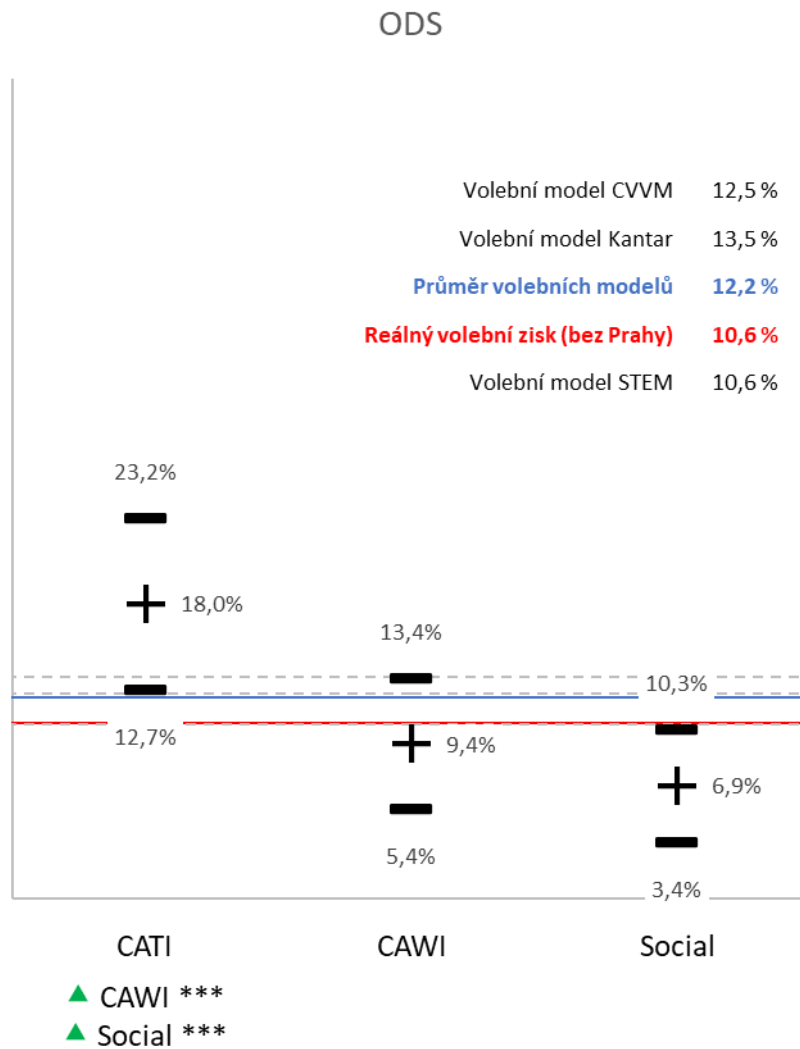
Zdroj: Vlastní

Legenda ke Grafům 1 až 12

- | | |
|---------------------------------------|--|
| - Horní interval spolehlivost (95 %) | ▲ Signifikantně vyšší hodnota (χ^2 test) |
| + Naměřená hodnota | * Na hladině testu 5 % |
| - Spodní interval spolehlivost (95 %) | ** Na hladině testu 1 % |
| | *** Na hladině testu 0,1 % |

CATI N=206, CAWI N=658, Social N=467

Graf 2: Politická preference dle módů, volby do poslanecké sněmovny 2017, ODS



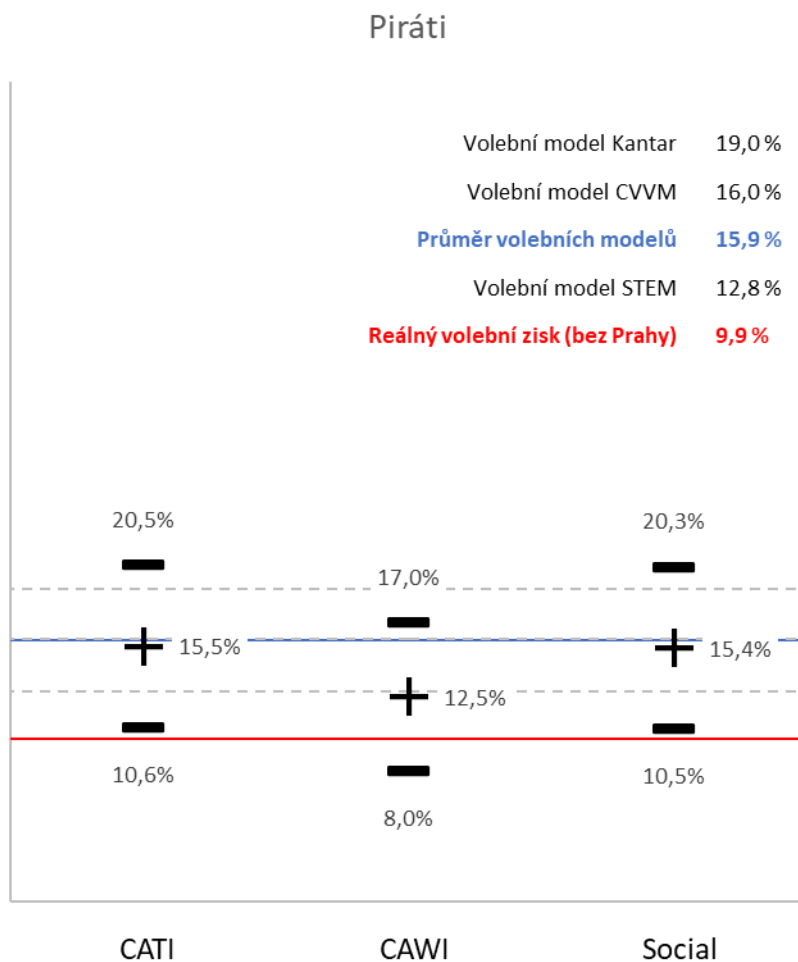
Zdroj: Vlastní

Legenda ke Grafům 1 až 12

- | | |
|---------------------------------------|--|
| - Horní interval spolehlivost (95 %) | ▲ Signifikantně vyšší hodnota (χ^2 test) |
| + Naměřená hodnota | * Na hladině testu 5 % |
| - Spodní interval spolehlivost (95 %) | ** Na hladině testu 1 % |
| | *** Na hladině testu 0,1 % |

CATI N=206, CAWI N=658, Social N=467

Graf 3: Politická preference dle módů, volby do poslanecké sněmovny 2017, Piráti



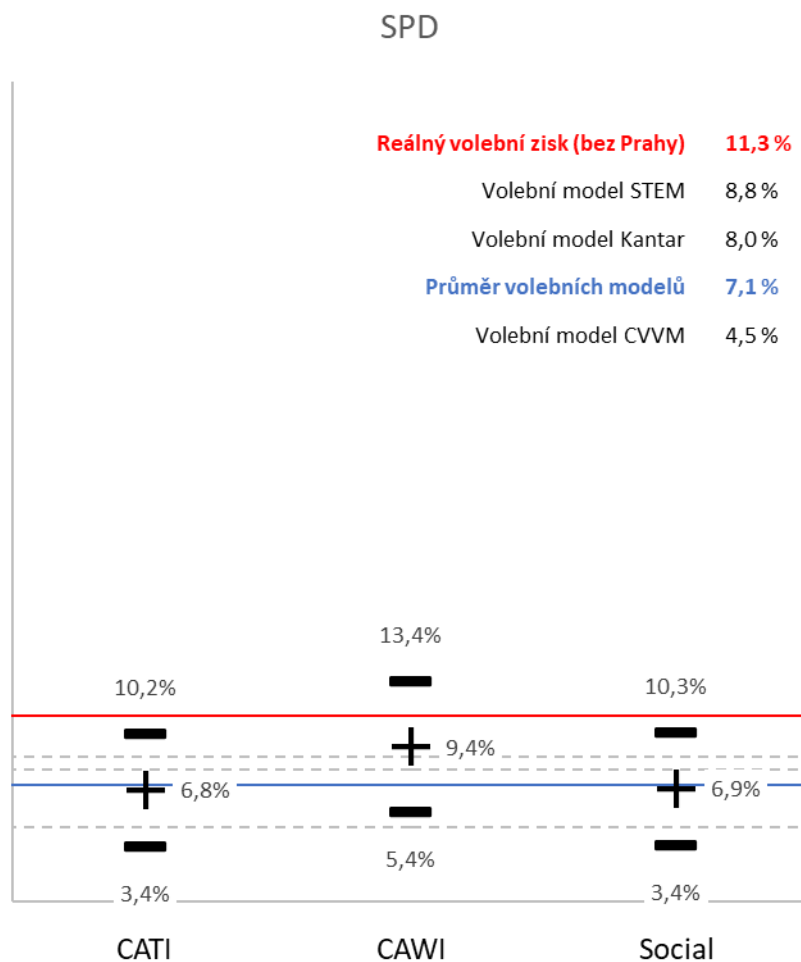
Zdroj: Vlastní

Legenda ke Grafům 1 až 12

- | | |
|---------------------------------------|--|
| - Horní interval spolehlivost (95 %) | ▲ Signifikantně vyšší hodnota (χ^2 test) |
| + Naměřená hodnota | * Na hladině testu 5 % |
| - Spodní interval spolehlivost (95 %) | ** Na hladině testu 1 % |
| | *** Na hladině testu 0,1 % |

CATI N=206, CAWI N=658, Social N=467

Graf 4: Politická preference dle módů, volby do poslanecké sněmovny 2017, SPD



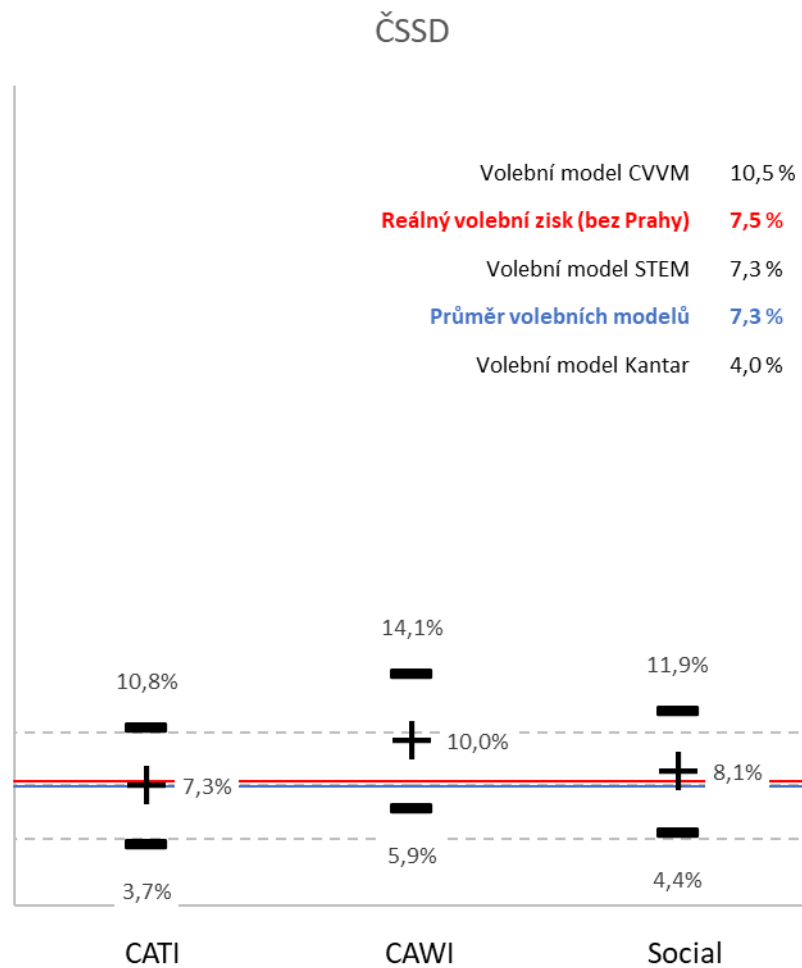
Zdroj: Vlastní

Legenda ke Grafům 1 až 12

- | | |
|---------------------------------------|--|
| - Horní interval spolehlivost (95 %) | ▲ Signifikantně vyšší hodnota (χ^2 test) |
| + Naměřená hodnota | * Na hladině testu 5 % |
| - Spodní interval spolehlivost (95 %) | ** Na hladině testu 1 % |
| | *** Na hladině testu 0,1 % |

CATI N=206, CAWI N=658, Social N=467

Graf 5: Politická preference dle módů, volby do poslanecké sněmovny 2017, ČSSD



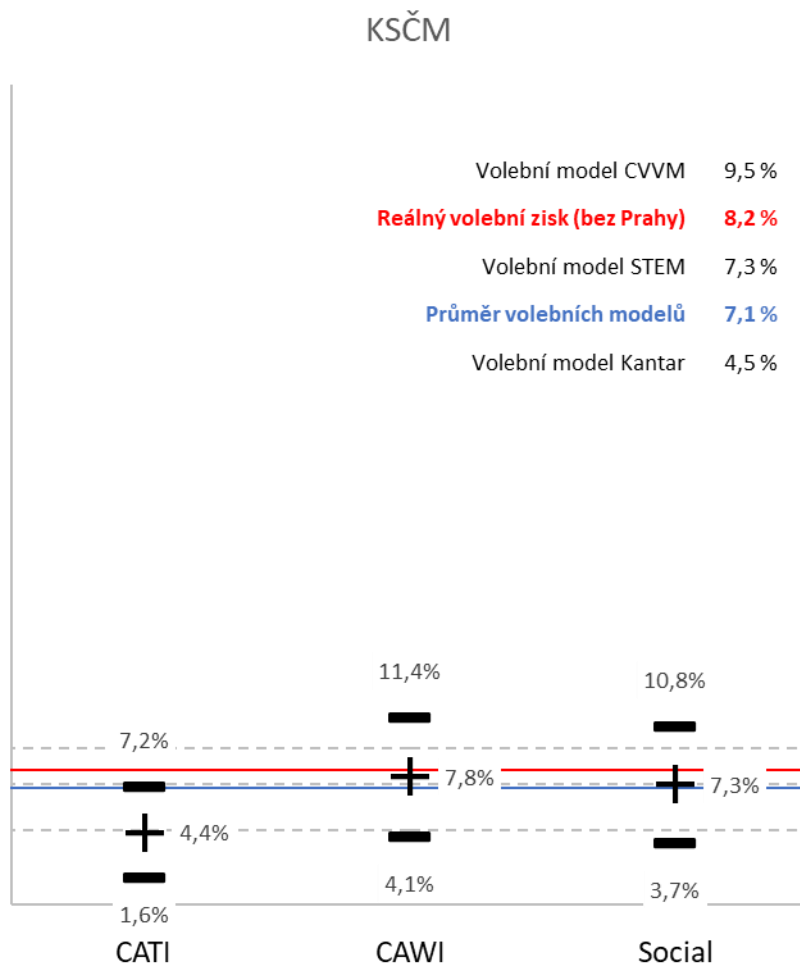
Zdroj: Vlastní

Legenda ke Grafům 1 až 12

- | | | | |
|---|-------------------------------------|-----|--|
| - | Horní interval spolehlivost (95 %) | ▲ | Signifikantně vyšší hodnota (χ^2 test) |
| + | Naměřená hodnota | * | Na hladině testu 5 % |
| - | Spodní interval spolehlivost (95 %) | ** | Na hladině testu 1 % |
| | | *** | Na hladině testu 0,1 % |

CATI N=206, CAWI N=658, Social N=467

Graf 6: Politická preference dle módů, volby do poslanecké sněmovny 2017, KSČM



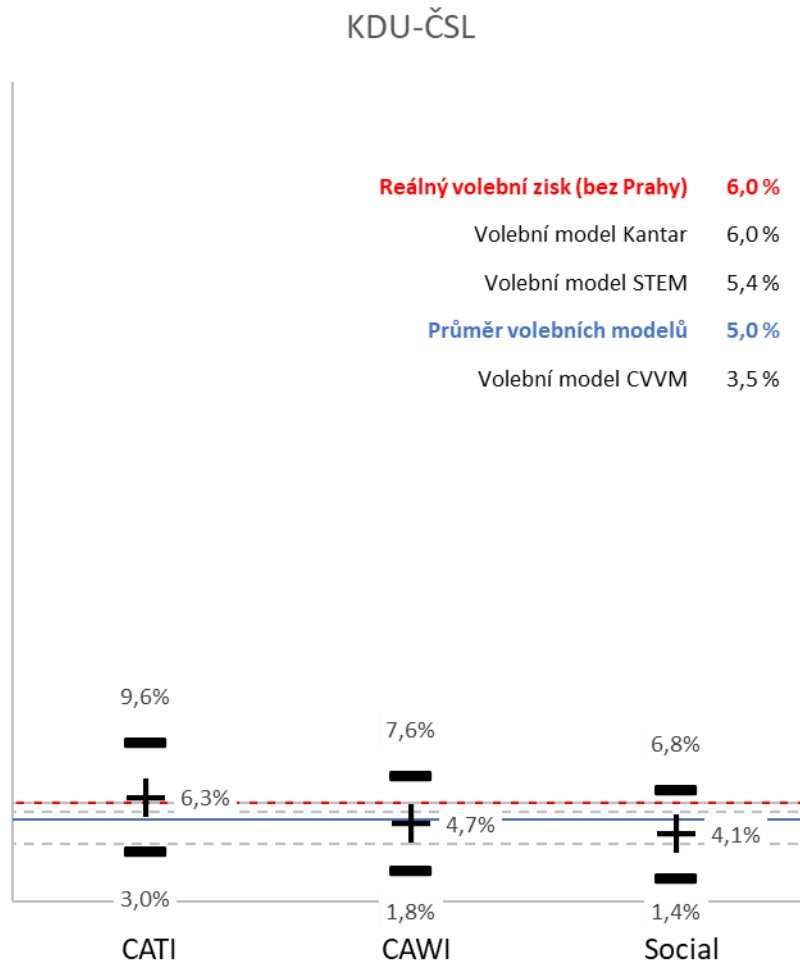
Zdroj: Vlastní

Legenda ke Grafům 1 až 12

- | | |
|---------------------------------------|--|
| - Horní interval spolehlivost (95 %) | ▲ Signifikantně vyšší hodnota (χ^2 test) |
| + Naměřená hodnota | * Na hladině testu 5 % |
| - Spodní interval spolehlivost (95 %) | ** Na hladině testu 1 % |
| | *** Na hladině testu 0,1 % |

CATI N=206, CAWI N=658, Social N=467

Graf 7: Politická preference dle módů, volby do poslanecké sněmovny 2017, KDU-ČSL



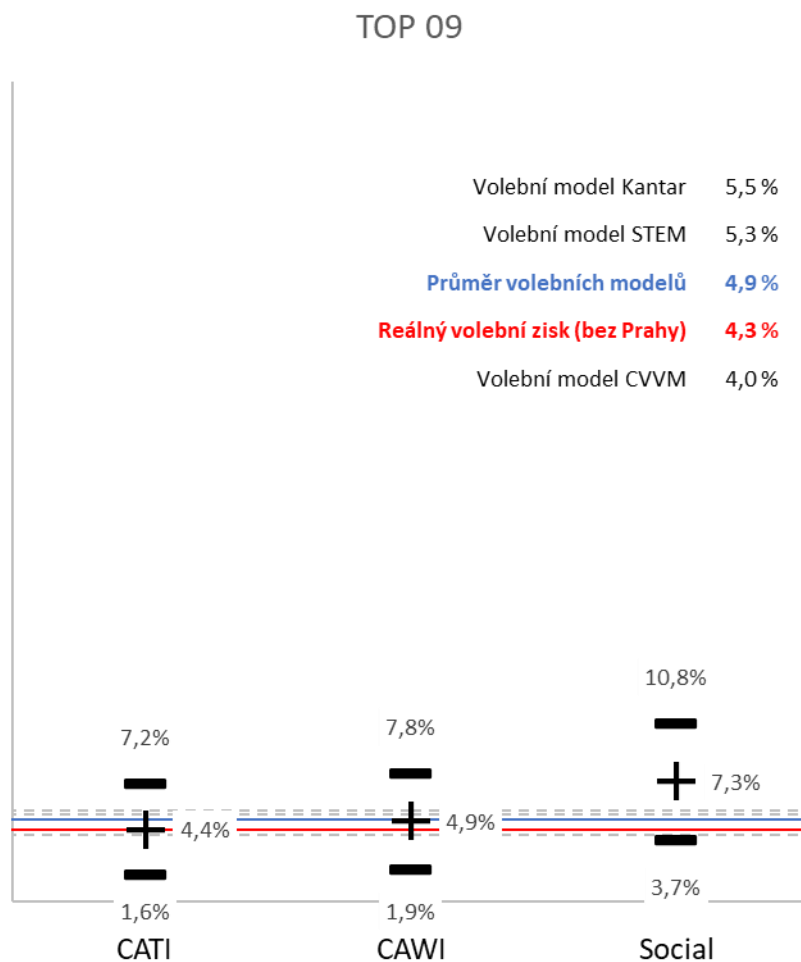
Zdroj: Vlastní

Legenda ke Grafům 1 až 12

- | | |
|---------------------------------------|--|
| - Horní interval spolehlivost (95 %) | ▲ Signifikantně vyšší hodnota (χ^2 test) |
| + Naměřená hodnota | * Na hladině testu 5 % |
| - Spodní interval spolehlivost (95 %) | ** Na hladině testu 1 % |
| | *** Na hladině testu 0,1 % |

CATI N=206, CAWI N=658, Social N=467

Graf 8: Politická preference dle módů, volby do poslanecké sněmovny 2017, TOP 09



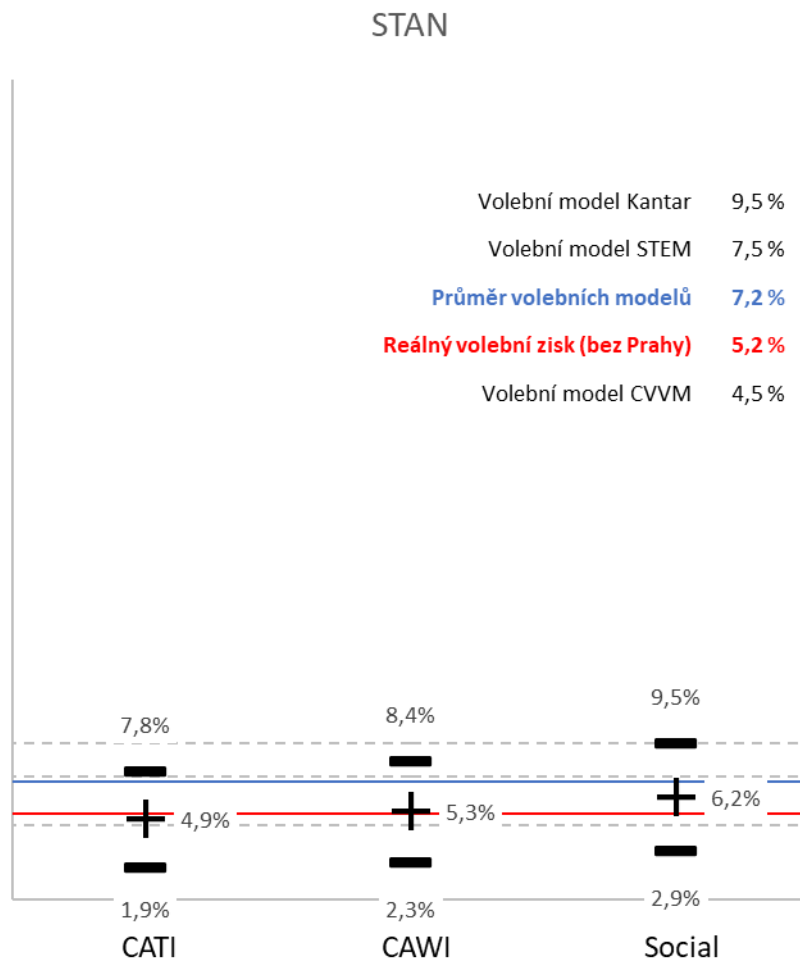
Zdroj: Vlastní

Legenda ke Grafům 1 až 12

- | | |
|---------------------------------------|--|
| - Horní interval spolehlivost (95 %) | ▲ Signifikantně vyšší hodnota (χ^2 test) |
| + Naměřená hodnota | * Na hladině testu 5 % |
| - Spodní interval spolehlivost (95 %) | ** Na hladině testu 1 % |
| | *** Na hladině testu 0,1 % |

CATI N=206, CAWI N=658, Social N=467

Graf 9: Politická preference dle módů, volby do poslanecké sněmovny 2017, STAN



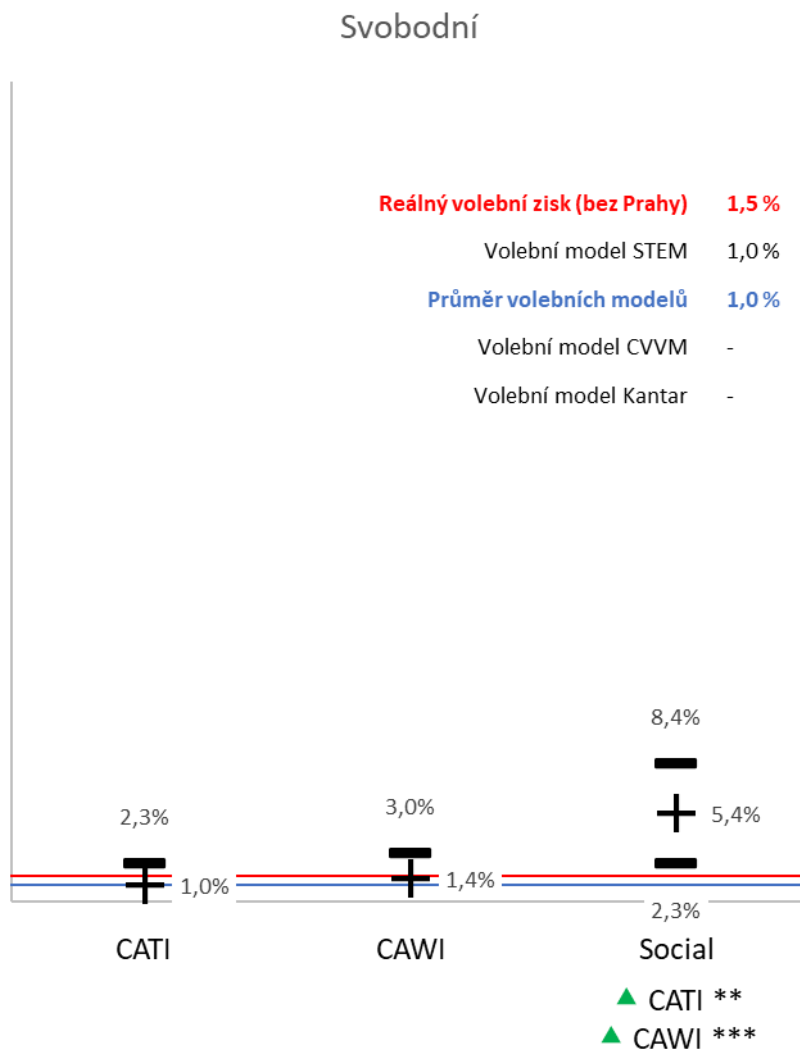
Zdroj: Vlastní

Legenda ke Grafům 1 až 12

- | | |
|---------------------------------------|--|
| - Horní interval spolehlivost (95 %) | ▲ Signifikantně vyšší hodnota (χ^2 test) |
| + Naměřená hodnota | * Na hladině testu 5 % |
| - Spodní interval spolehlivost (95 %) | ** Na hladině testu 1 % |
| | *** Na hladině testu 0,1 % |

CATI N=206, CAWI N=658, Social N=467

Graf 10: Politická preference dle módů, volby do poslanecké sněmovny 2017, Svobodní



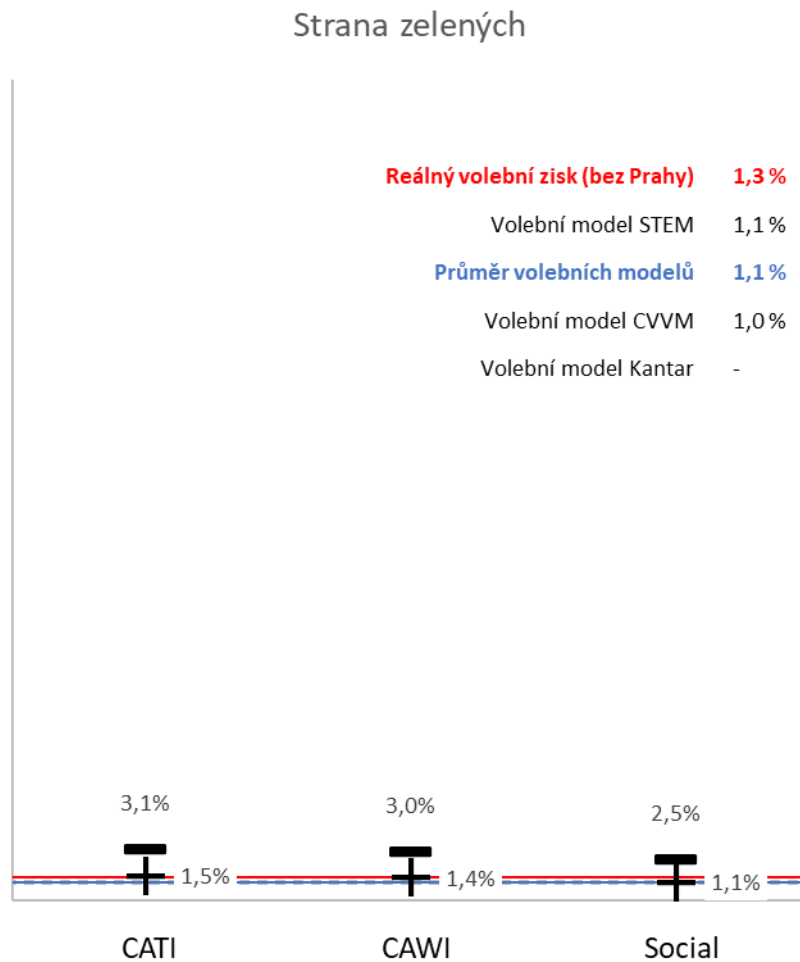
Zdroj: Vlastní

Legenda ke Grafům 1 až 12

-	Horní interval spolehlivost (95 %)	▲	Signifikantně vyšší hodnota (χ^2 test)
+	Naměřená hodnota	*	Na hladině testu 5 %
-	Spodní interval spolehlivost (95 %)	**	Na hladině testu 1 %
		***	Na hladině testu 0,1 %

CATI N=206, CAWI N=658, Social N=467

Graf 11: Politická preference dle módů, volby do poslanecké sněmovny 2017, Strana zelených



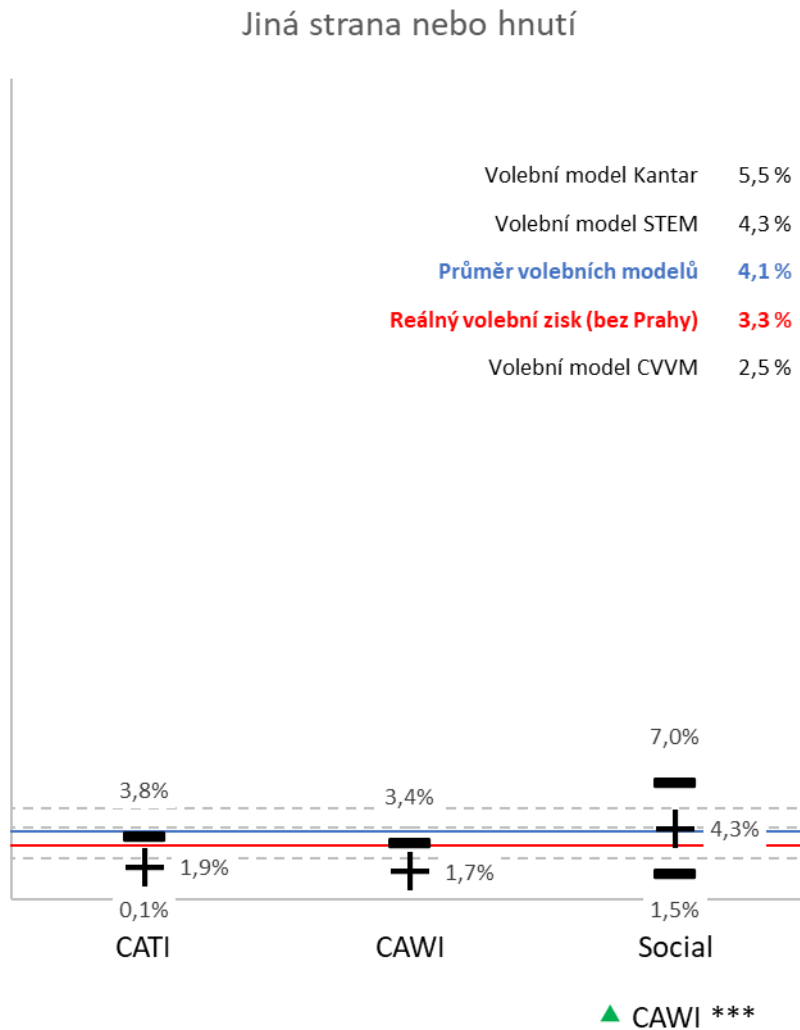
Zdroj: Vlastní

Legenda ke Grafům 1 až 12

- Horní interval spolehlivost (95 %)	▲ Signifikantně vyšší hodnota (χ^2 test)
+ Naměřená hodnota	* Na hladině testu 5 %
- Spodní interval spolehlivost (95 %)	** Na hladině testu 1 %
	*** Na hladině testu 0,1 %

CATI N=206, CAWI N=658, Social N=467

Graf 12: Politická preference dle módů, volby do poslanecké sněmovny 2017, Jiná strana nebo hnutí



Zdroj: Vlastní

Legenda ke Grafům 1 až 12

- | | |
|--|--|
| - Horní interval spolehlivosti (95 %) | ▲ Signifikantně vyšší hodnota (χ^2 test) |
| + Naměřená hodnota | * Na hladině testu 5 % |
| - Spodní interval spolehlivosti (95 %) | ** Na hladině testu 1 % |
| | *** Na hladině testu 0,1 % |

CATI N=206, CAWI N=658, Social N=467

Grafy 1 až 12 ukazují naměřené hodnoty politických preferencí (na základě voličských preferencí, zobrazené hodnoty tedy ukazují podíl hlasů jednotlivých stran na základě deklarační voličů, kteří volili ve volbách a vybavili si svou volbu) spolu s intervaly spolehlivosti.

Zároveň grafy zobrazují i skutečnou pravdivou hodnotu politické participace ve volbách do poslanecké sněmovny v roce 2017 (za všechny kraje s výjimkou Hlavního města Prahy). Další zobrazenou hodnotou jsou jednotlivé volební modely ze září roku 2020, tedy z období, kdy probíhal sběr dat. Jejich účel je indikativní nastínění současné situace pro upozornění na případný posun v politickém diskurzu a struktuře podpory jednotlivým stranám v české společnosti oproti volbám do poslanecké sněmovny v roce 2017. Je ale nutné mít na zřeteli tři základní body, které podtrhují indikativnost těchto hodnot:

1. Jedná se o volební modely (nikoli voličské preference, které jsou výzkumem použitým v této práci analyzovány).
2. Jednotlivé volební modely pracují s mírně odlišným počtem uvedených stran a hnutí, jmenovitě se jedná o Stranu svobodných občanů (Svobodní), Stranu zelených a tím pádem i hodnotu pod položkou Jiné strany a hnutí.
3. Volební modely staví na celorepublikových datech, zatímco analyzovaná data sebraná před volbami do krajských zastupitelstev v roce 2020 pracují jen s kraji mimo Hlavní město Prahu.

V neposlední řadě se liší i mód sběru dat a metoda výběru jednotlivých volebních modelů. Výzkum CVVM sbíral data skrze nepravděpodobnostní kvótní výběr a osobní dotazování (kombinace CAPI a PAPI) (Pilnáček, 2020, s. 7). Výzkum Trendy Česka agentury Kantar použil prostý náhodný pravděpodobnostní výběr skrze vytáčení náhodných čísel (RDD) a telefonické dotazování (CATI) (Šabotník et al., 2020). Výzkum STEMu shodně s výzkumem CVVM využil nepravděpodobnostní kvótní výběr a kombinaci CAPI a PAPI.

Hodnoty jednotlivých volebních modelů jsou v grafech 1 až 12 zobrazeny samostatně i ve formě neváženého průměru (ačkoli nepracovaly všechny volební modely s naprosto přesným počtem sebraných dotazníků, pro účely následující analýzy jsou brány mezi sebou jako stejně významné) za všechny modely dohromady. V případě grafu 10 až 12 se liší počet modelů, které přispívají do celkového průměru dle stran a hnutí, které modely zahrnují samostatně a které zahrnují do kategorie Jiná strana nebo hnutí.

Z analýzy jednotlivých politických preferencí a srovnání se skutečnou hodnotou volebního zisku ve volbách do poslanecké sněmovny v roce 2017 vyplývá vliv tazatele (a chybu s tím spojenou) a sociální desirability na sebraná data. Jmenovitě se jedná o případ u krajně pravicových a levicových stran a hnutí. SPD, ačkoli celkově podreprezentována, má nejvíce vychýlenou hodnotu právě u CATI sběru (Graf 4). Obdobná situace platí i u strany

Komunistická strana Čech a Moravy, kde CAWI a Social mód sběru vykazují jen drobnou odchylku od skutečné hodnoty, ale u CATI sběru je tato strana silně podhodnocena (Graf 6).

Na úrovni chí-kvadrát testu diskretních rozdělení měření za jednotlivé módy sběru dat vycházejí signifikantní rozdíly (na minimální hladině testu 5 %) v Grafu 1, 2, 10 a 12. Celkově vzato tedy se tedy naměřené hodnoty v jednotlivých módech liší jen částečně.

Tabulka 5: Analýza přesnosti jednotlivých módů sběru dat skrze srovnání hodnot odchylek od reálného volebního zisku

	CATI	CAWI	Social
Suma absolutních hodnot odchylek (rozdíl naměřených a skutečných hodnot)	27,1 %	14,3 %	29,9 %
Indexace	0,91	0,48	1,00
Suma hodnot kvadratických odchylek (rozdíl naměřených a skutečných hodnot)	130,7	27,5	114,7
Indexace	1,00	0,21	0,88
Suma adjustovaných odchylek (poměr odchylky modelu a reálné hodnoty)	323,3	199,3	579,2
Indexace	0,56	0,34	1,00

Zdroj: Vlastní

Tabulka 5 ukazuje srovnání přesnosti jednotlivých módů sběru dat skrze porovnání naměřených a skutečných pravdivých hodnot. U prostého součtu absolutních hodnot odchylek dopadl nejhůře Social mód sběru dat, přičemž nepřesnost u CATI měření je obdobná. CAWI mód sběru dat pak dosahuje přibližně poloviční hodnoty z hlediska prosté sumy jednotlivých odchylek.

Pokud dojde k umocnění odchylek a jejich následné sumarizaci za účelem penalizace vyšších hodnot u jednotlivých dílčích odchylek na úrovni jednotlivých stran, CAWI mód sběru dat vychází nejlépe a jeho hodnota sumy kvadratických odchylek je zhruba jen pětina oproti CATI sběru.

U adjustovaných odchylek, které zohledňují velikost odchylky vůči zisku hlasů jednotlivých stran a počítají tak závažnost stejné odchylky u méně volených stran oproti stranám více

voleným, dosahuje nejvyšší přesnosti CAWI mód sběru dat (zhruba třetinová hodnota sumy adjustovaných odchylek oproti nejméně přesnému módu), následovaný CATI módem (zhruba poloviční hodnota).

Tabulka 6: Analýza přesnosti jednotlivých módů sběru dat skrze srovnání hodnot odchylek od volebních modelů

	CATI	CAWI	Social
Suma absolutních hodnot odchylek (rozdíl naměřených a průměru volebních modelů)	16,0 %	20,4 %	20,9 %
Indexace	0,77	0,98	1,00
Suma hodnot kvadratických odchylek (rozdíl naměřených a průměru volebních modelů)	53,2	55,4	69,8
Indexace	0,76	0,79	1,00
Suma adjustovaných odchylek (poměr odchylky modelu a průměru volebních modelů)	257,8	286,4	623,0
Indexace	0,41	0,46	1,00

Zdroj: Vlastní

Stejný postup byl aplikován i pro indikativní srovnání naměřených hodnot oproti průměru tří volebních modelů ze září 2020. V tomto případě vychází na všech třech úrovních srovnání jako nejpresnější CATI mód sběru dat. Rozdíly na úrovni indexace ale nejsou ve srovnání s porovnáním s reálným volebním ziskem stran (Tabulka 5) tak razantní. Suma kvadratických odchylek a adjustovaných odchylek dopadla pro CATI a CAWI mód podobně. Je ale jasné, že pokud zohledníme situaci a voličskou strukturu v období měření (září 2020) a nikoli reálné volební výsledky z roku 2017, změní se přesnost jednotlivých módů sběru dat.

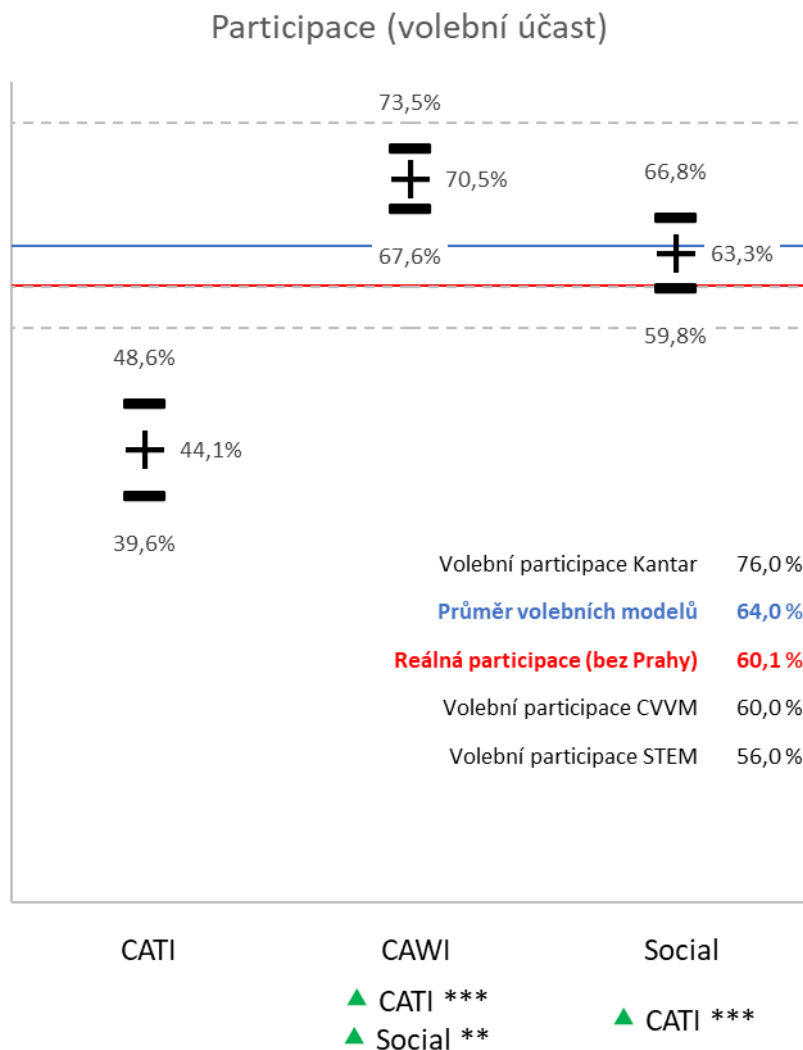
Všechna tato zjištění jasně vedou k zamítnutí první nulové hypotézy:

H01: Data sebraná odlišnými módy sběru dat, tedy CATI, CAWI Panel (CAWI) a CAWI Facebook (Social), jsou ekvivalentní z hlediska politické preference.

4.1.1 Analýza přesnosti měření deklarované politické participace

Dalším krokem je obdobná analýza na úrovni politické participace ve volbách do poslanecké sněmovny v roce 2017.

Graf 13: Politická participace dle módů, volby do poslanecké sněmovny 2017



Zdroj: Vlastní

Legenda

- Horní interval spolehlivost (95 %)	▲ Signifikantně vyšší hodnota (χ^2 test)
+ Naměřená hodnota	* Na hladině testu 5 %
- Spodní interval spolehlivost (95 %)	** Na hladině testu 1 %
	*** Na hladině testu 0,1 %

CATI N=467, CAWI N=933, Social N=738

V tomto případě jsou patrné rozdíly v jednotlivých módech měření na úrovni chí-kvadrát testu diskrétních rozdělení hodnot měření. Hodnoty CATI měření se významně odlišují od obou zbývajících módů, stejně tak se významně liší hodnoty CAWI a Social módů sběru dat.

Tabulka 7: Analýza přesnosti jednotlivých módů sběru dat skrze srovnání hodnot odchylek od reálné volební účasti

	CATI	CAWI	Social
Absolutní hodnota odchylky (rozdíl naměřené a skutečné hodnoty)	16,0 %	10,5 %	3,2 %
Indexace	1,00	0,66	0,20
Kvadratická odchylka (rozdíl naměřené a skutečné hodnoty)	254,4	109,5	10,4
Indexace	1,00	0,43	0,04
Adjustovaná odchylka (poměr odchylky modelu a průměru volebních modelů)	0,27	0,17	0,05
Indexace	1,00	0,66	0,20

Zdroj: Vlastní

Při srovnání odchylek na úrovni politické participace je nejpresnějším módem sběru dat Social. Naopak největších odchylek dosahuje CATI sběr. Stejná situace platí i pro srovnání naměřené míry volební participace s průměrem tří volebních modelů z roku 2020. V tomto případě je Social sběr ještě přesnější, stejně tak i CAWI.

Tabulka 8: Analýza přesnosti jednotlivých módů sběru dat skrze srovnání hodnot odchylek od volebních modelů

	CATI	CAWI	Social
Absolutní hodnota odchylky (rozdíl naměřené a skutečné hodnoty)	20,0 %	7,0 %	1 %
Indexace	1,00	0,33	0,04
Kvadratická odchylka (rozdíl naměřené a skutečné hodnoty)	395,6	42,6	0,5
Indexace	1,00	0,11	0,00
Adjustovaná odchylka (poměr odchylky modelu a průměru volebních modelů)	0,31	0,10	0,01
Indexace	1,00	0,33	0,04

Zdroj: Vlastní

Závěrem první analytické části lze na základě provedených analýz zamítnout i druhou a třetí nulovou hypotézu:

H02: Data sebraná odlišnými módy sběru dat, tedy CATI, CAWI Panel (CAWI) a CAWI Facebook (Social), jsou ekvivalentní z hlediska politické participace.

H03: Data sebraná odlišnými módy sběru dat, tedy CATI, CAWI Panel (CAWI) a CAWI Facebook (Social), stejnou měrou reprezentují skutečný stav z hlediska politické participace a preference (jsou tedy stejně validní).

4.2 Analýza faktorů ovlivňujících politickou participaci

Druhý analytický přístup se zaměřuje na politickou participaci a faktory, které jí ovlivňují. Nástrojem této analýzy je binární logistická regrese, která v několika modelech odhaluje vliv nezávislých faktorů na závislou proměnnou politické participace. Jedná se tak o analytický nástroj k získání znalosti ohledně vnitřní konzistence jednotlivých datových souborů skrze

poukázání na roli latentních vlivů ovlivňujících politickou participaci a slouží jako opora k prohloubení rozhodnutí o zamítnutí druhé nulové hypotézy:

H02: Data sebraná odlišnými módy sběru dat, tedy CATI, CAWI Panel (CAWI) a CAWI Facebook (Social), jsou ekvivalentní z hlediska politické participace.

Mezi klíčové faktory ovlivňující politickou participaci v českém kontextu patří vzdělání, věk, předchozí politická participace a částečně i zaměstnanost (saturující socioekonomický status), a to jak na celorepublikové úrovni, tak na krajské (Bláha et al., 2017, s. 53; Linek, 2015, s. 87; Linek & Petrušek, 2018, s. 533–534; Prokop, 2020, s. 9–10). Všechny tyto faktory působí na politickou participaci pozitivně a v celé své šířce.

Níže představené modely logistické regrese jsou vždy analyzovány v párech, politická participace je totiž zkoumána na úrovni deklarované budoucí volební účasti ve volbách do krajských zastupitelstev v roce 2020 (závislá proměnná prvního regresního modelu) a následně i na deklarované předešlé volební účasti v krajských volbách v roce 2017 (závislá proměnná druhého regresního modelu).

Tabulka 9a: Párový model 1 - Politická participace v krajských volbách 2020 (binární logistická regrese), model s vlivem pohlaví, věku a vzdělání

	CATI			CAWI			SOCIAL		
	B	Exp(B)	Sig.	B	Exp(B)	Sig.	B	Exp(B)	Sig.
Pohlaví (žena)	0,14	1,16	0,55	-0,49	0,61	0,004	-0,19	0,82	0,316
Věk (18-24 let)			0,000			0,009			0,058
25-34 let	1,88	6,53	0,006	0,91	2,49	0,035	-0,03	0,97	0,963
35-44 let	1,46	4,31	0,029	0,76	2,14	0,073	-0,07	0,93	0,908
45-54 let	1,58	4,85	0,020	0,78	2,19	0,066	0,21	1,23	0,743
55-64 let	2,41	11,17	0,001	0,94	2,56	0,031	-0,37	0,69	0,552
65 a více let	2,63	13,87	0,000	1,43	4,17	0,001	0,55	1,73	0,378
Vzdělání (základní)			0,000			0,001			0,000
Střední bez maturity	0,62	1,87	0,098	0,56	1,75	0,036	0,00	1,00	0,992
Střední s maturitou	1,60	4,94	0,000	0,80	2,23	0,003	0,59	1,80	0,079
Vysokoškolské (včetně VOŠ)	2,25	9,46	0,000	1,25	3,48	0,000	1,65	5,20	0,000
Zaměstnaní	-	-	-	-	-	-	-	-	-
Velikost místa bydliště (do 1999 obyvatel)			-			-			-
2000-4999 obyvatel	-	-	-	-	-	-	-	-	-
5000-19999 obyvatel	-	-	-	-	-	-	-	-	-
20000-49999 obyvatel	-	-	-	-	-	-	-	-	-

50000-99999 obyvatel	-	-	-	-	-	-	-	-	
100000 a více obyvatel	-	-	-	-	-	-	-	-	
Účast ve volbách do KZ 2016	-	-	-	-	-	-	-	-	
Účast ve volbách do PS 2017	-	-	-	-	-	-	-	-	
N	467			933			738		
Negelkerke R ²	0,174			0,06			0,086		
Úspěšnost klasifikace	3,2 (74,4 - 77,6)			0,2 (77,6 - 77,7)			0 (80,8 - 80,8)		

Zdroj: Vlastní

Tabulka 9b: Párový model 1 - Politická participace ve volbách do poslanecké sněmovny 2017 (binární logistická regrese), model s vlivem pohlaví, věku a vzdělání

	CATI			CAWI			SOCIAL		
	B	Exp(B)	Sig.	B	Exp(B)	Sig.	B	Exp(B)	Sig.
Pohlaví (žena)	-0,71	0,49	0,00	-0,88	0,41	0,000	-0,50	0,60	0,001
Věk (18-24 let)			0,014			0,000			0,158
25-34 let	1,85	6,34	0,049	0,66	1,94	0,113	-0,16	0,85	0,763
35-44 let	1,85	6,36	0,048	0,61	1,85	0,138	-0,41	0,66	0,416
45-54 let	2,01	7,49	0,032	1,12	3,07	0,008	-0,40	0,67	0,437
55-64 let	2,75	15,65	0,004	1,52	4,58	0,000	-0,30	0,74	0,563
65 a více let	2,24	9,38	0,016	1,80	6,03	0,000	0,14	1,15	0,782
Vzdělání (základní)			0,000			0,000			0,000
Střední bez maturity	0,38	1,46	0,367	-0,06	0,94	0,815	-0,11	0,89	0,689
Střední s maturitou	1,23	3,43	0,003	0,46	1,59	0,082	0,26	1,30	0,367
Vysokoškolské (včetně VOŠ)	1,81	6,13	0,000	0,92	2,52	0,002	0,84	2,32	0,009
Zaměstnaní	-	-	-	-	-	-	-	-	-
Velikost místa bydliště (do 1999 obyvatel)			-			-			-
2000-4999 obyvatel	-	-	-	-	-	-	-	-	-
5000-19999 obyvatel	-	-	-	-	-	-	-	-	-
20000-49999 obyvatel	-	-	-	-	-	-	-	-	-
50000-99999 obyvatel	-	-	-	-	-	-	-	-	-
100000 a více obyvatel	-	-	-	-	-	-	-	-	-
Účast ve volbách do KZ 2016	-	-	-	-	-	-	-	-	-
Účast ve volbách do KZ 2020	-	-	-	-	-	-	-	-	-
N	467			933			738		
Negelkerke R ²	0,16			0,127			0,065		
Úspěšnost klasifikace	10,4 (55,9 - 66,3)			3 (68,1 - 71)			1,3 (61,4 - 62,7)		

Zdroj: Vlastní

První dva modely (Tabulka 9a a 9b) srovnávají vliv základních sociodemografických ukazatelů pohlaví, věku a vzdělání (nezávislé proměnné) na politickou participaci v jednotlivých volbách (závislá proměnná). Dle výše představené literatury a předešlých výzkumů lze očekávat signifikantní pozitivní vliv věku a vzdělání a naopak nesignifikantní vliv pohlaví.

Vliv věku v tomto případě nejlépe diskriminuje CATI mód sběru dat. Nabývá signifikance ve všech kategoriích a jeho vliv je nejsilnější u dvou nejstarších skupin respondentů a to shodně v obou párových modelech. U CAWI módů sběru dat je poté signifikantní pozitivní vliv věku také přítomný, nicméně ne ve všech kategoriích a vliv oproti kontrastní kategorii (v tomto případě respondenti ve věku 18-24 let) je nižší, než u CATI dat. Zatímco u CATI mají respondenti spadající věkem do poslední kategorie (65 a více let) téměř čtrnáctkrát, respektive devětkrát větší šanci v případě druhého modelu, že budou volit oproti respondentům v nejmladší kategorii. U CAWI je tato šance pouze čtyřnásobná, respektive šestinásobná v případě druhého modelu. U Social módu sběru dat je čitelný pozitivní vliv u poslední věkové kategorie (pozitivní hodnota exponovaných regresních koeficientů $\text{Exp}(B)$ je v tomto případě větší než 1), ale jedná se o nesignifikantní vliv.

Přesuňme se k vlivu vzdělání. Zde shodně všechny tři datové soubory náležící k jednotlivým módům sběru prokazují signifikantní vliv vzdělání u vysokoškolsky vzdělaných respondentů. Tento vliv je u obou párových modelů největší u CATI sběru. V rámci prvního modelu (vliv na politickou participaci v krajských volbách 2020, Tabulka 9a) lze sledovat tendenční nárůst (čím vyšší úroveň vzdělání, tím větší pozitivní vliv) vlivu vzdělání u CATI a CAWI módů sběru dat. U druhého modelu (vliv na politickou participaci v parlamentních volbách 2017, Tabulka 9b) je tento tendenční vliv patrný jen u CATI sběru. Z hlediska důkazu signifikance vlivu vzdělání ve všech jeho kategoriích v prvním modelu lépe dopadl CAWI sběr (signifikance u všech kategorií), v druhém modelu pak CATI sběr (signifikance u tří ze čtyř kategorií včetně kontrastní kategorie).

Poslední nezávislou proměnnou vstupující do párové logistické regrese je pohlaví, které nemá vykazovat signifikantní vliv na politickou participaci. V obou regresních modelech k tomuto náznaku dochází. U prvního modelu pouze u CAWI sběru, u druhého pak napříč všemi třemi módy sběru dat. Ačkoli se jedná o poměrně malý vliv z hlediska porovnání změny šance se změnou pohlaví, poukazuje tato signifikance na prokázání falešného vlivu faktoru na závislou proměnnou politické participace, tedy na chybu na úrovni měření. Na tomto místě je ještě vhodné zmínit, že ačkoli nebyla žádná ze tří struktur výběrového souboru převažována dle pohlaví, distribuce této proměnné je ve všech třech souborech srovnatelná jak mezi sebou, tak vůči základnímu souboru (Tabulka 3 a 4).

Posledním krokem je srovnání pseudo koeficientu determinace (v tomto případě Nagelkerkeho R^2) a úspěšnosti klasifikace (srovnání procenta správně zařazených případů před a po aplikování nezávislých proměnných v rámci modelu) jednotlivých módů sběru dat v obou modelech. U prvního modelu fungují zvolené sociodemografické ukazatele nejlépe u CATI dotazování dle saturace koeficientu determinace. U druhého modelu je tato situace pro CATI

stejná, CAWI dotazování ale dosahuje vyšších hodnot a rozdíl s CATI tak vyrovnává. U úspěšnosti klasifikace také vede CATI. Celkově vzato lze konstatovat, že při dokazování relevance efektu pohlaví, věku a vzdělání na politickou participaci vychází nejlépe CATI, následně CAWI a nejhůře Social mód sběru dat.

Tabulka 10a: Párový model 2 - Politická participace v krajských volbách 2020 (binární logistická regrese), model s vlivem pohlaví, věku, vzdělání, zaměstnání a velikosti místa bydliště

	CATI			CAWI			SOCIAL		
	B	Exp(B)	Sig.	B	Exp(B)	Sig.	B	Exp(B)	Sig.
Pohlaví (žena)	0,10	1,11	0,69	-0,42	0,66	0,015	-0,22	0,80	0,254
Věk (18-24 let)			0,036			0,001			0,234
25-34 let	1,43	4,18	0,071	0,85	2,33	0,053	0,11	1,12	0,861
35-44 let	0,86	2,36	0,262	0,58	1,78	0,180	0,18	1,20	0,780
45-54 let	1,18	3,24	0,132	0,60	1,83	0,164	0,39	1,47	0,545
55-64 let	1,88	6,55	0,019	0,88	2,41	0,044	-0,28	0,75	0,651
65 a více let	1,97	7,18	0,016	1,62	5,03	0,000	0,41	1,50	0,520
Vzdělání (základní)			0,000			0,004			0,000
Střední bez maturity	0,46	1,58	0,283	0,48	1,61	0,078	-0,01	0,99	0,978
Střední s maturitou	1,57	4,78	0,000	0,73	2,06	0,008	0,62	1,86	0,072
Vysokoškolské (včetně VOŠ)	2,14	8,48	0,000	1,10	3,01	0,001	1,67	5,31	0,000
Zaměstnání	-0,16	0,85	0,642	0,50	1,64	0,015	-0,44	0,64	0,082
Velikost místa bydliště (do 1999 obyvatel)			0,384			0,541			0,636
2000-4999 obyvatel	-0,28	0,75	0,469	-0,16	0,85	0,552	0,11	1,12	0,717
5000-19999 obyvatel	0,22	1,25	0,503	-0,45	0,64	0,055	0,43	1,54	0,152
20000-49999 obyvatel	-0,46	0,63	0,242	-0,15	0,86	0,568	0,29	1,34	0,344
50000-99999 obyvatel	0,61	1,83	0,270	-0,12	0,88	0,687	0,21	1,23	0,521
100000 a více obyvatel	-0,21	0,81	0,617	-0,29	0,75	0,319	0,48	1,62	0,138
Účast ve volbách do KZ 2016	-	-	-	-	-	-	-	-	-
Účast ve volbách do PS 2017	-	-	-	-	-	-	-	-	-
N	467			933			738		
Negelkerke R ²	0,170			0,074			0,099		
Úspěšnost klasifikace	1,8 (74,9 - 76,7)			0,5 (77,6 - 78)			0,2 (80,8 - 80,9)		

Zdroj: Vlastní

Tabulka 10b: Párový model 2 - Politická participace ve volbách do poslanecké sněmovny 2017 (binární logistická regrese), model s vlivem pohlaví, věku, vzdělání, zaměstnání a velikosti místa bydliště

	CATI			CAWI			SOCIAL		
	B	Exp(B)	Sig.	B	Exp(B)	Sig.	B	Exp(B)	Sig.
Pohlaví (žena)	-0,77	0,46	0,00	-0,88	0,42	0,000	-0,52	0,60	0,001
Věk (18-24 let)			0,071			0,000			0,355
25-34 let	1,63	5,13	0,098	0,63	1,88	0,133	-0,14	0,87	0,795
35-44 let	1,68	5,34	0,088	0,58	1,79	0,163	-0,45	0,64	0,391
45-54 let	1,79	6,02	0,069	1,07	2,91	0,012	-0,39	0,68	0,454
55-64 let	2,44	11,52	0,014	1,51	4,55	0,000	-0,31	0,73	0,551
65 a více let	2,04	7,66	0,043	1,82	6,19	0,000	0,12	1,13	0,811
Vzdělání (základní)			0,000			0,000			0,000
Střední bez maturity	0,22	1,24	0,625	-0,07	0,93	0,793	-0,14	0,87	0,638
Střední s maturitou	1,01	2,76	0,019	0,46	1,59	0,085	0,26	1,29	0,378
Vysokoškolské (včetně VOŠ)	1,59	4,89	0,001	0,93	2,54	0,002	0,87	2,39	0,008
Zaměstnání	-0,09	0,92	0,766	0,05	1,05	0,787	-0,03	0,97	0,899
Velikost místa bydliště (do 1999 obyvatel)			0,810			0,570			0,868
2000-4999 obyvatel	0,15	1,16	0,670	0,30	1,35	0,216	-0,16	0,86	0,566
5000-19999 obyvatel	0,16	1,17	0,582	0,08	1,09	0,688	-0,13	0,88	0,601
20000-49999 obyvatel	0,46	1,58	0,185	0,42	1,52	0,080	-0,26	0,77	0,309
50000-99999 obyvatel	0,36	1,43	0,350	0,12	1,13	0,655	-0,11	0,90	0,688
100000 a více obyvatel	0,31	1,36	0,401	0,19	1,21	0,469	-0,31	0,73	0,226
Účast ve volbách do KZ 2016	-	-	-	-	-	-	-	-	-
Účast ve volbách do KZ 2020	-	-	-	-	-	-	-	-	-
N	467			933			738		
Negelkerke R ²	0,145			0,132			0,070		
Úspěšnost klasifikace	12,1 (54,8 - 66,9)			3,9 (68,1 - 71,9)			0,3 (61,3 - 61,5)		

Zdroj: Vlastní

Druhá dvojice párových modelů přidává do logistických regresí vliv zaměstnanosti a velikosti místa bydliště. Zatímco velikost místa se neřadí k faktorům, které signifikantně pozitivně ovlivňují politickou participaci, socioekonomický status ano. V tomto případě se tedy dá očekávat, že zaměstnanost, která saturuje socioekonomický status, bude mít pozitivní signifikantní vliv na politickou participaci.

Začněme s revidováním vlivu nezávislých proměnných z předešlého párového modelu. Falešný negativní vliv pohlaví zůstal signifikantní pro všechny tři módy sběru dat u druhého párového modelu (vliv na politickou participaci ve volbách do poslanecké sněmovny 2017, Tabulka 10b) a pro CAWI sběr u prvního párového modelu (vliv na politickou participaci ve volbách do krajských zastupitelstev 2020, Tabulka 10a). V obou případech se ale jedná jen o

velmi mírný vliv na samotnou politickou participaci. Vliv věku u prvního modelu úspěšně v některých kategoriích odhalilo CATI i CAWI, u druhého modelu pak lépe dle prokázání signifikance CATI. Social mód v tomto ohledu u obou párových modelů zaostává. U vzdělání prokazují signifikantní vliv vysokoškolského vzdělání všechny tři módy u obou modelů, lépe tyto pravidelnosti ale demonstruje opět CATI a CAWI sběr oproti Social módu sběru dat.

Pozitivní a signifikantní vliv zaměstnání byl prokázán jen u prvního modelu a to u CAWI sběru. Signifikance vlivu velikosti místa bydliště na politickou participaci není zřetelná v žádném z šesti analyzovaných případů. Pseudo koeficient determinace i úspěšnost klasifikace je opět největší u CATI módu sběru dat.“

Tabulka 11a: Párový model 3 - Politická participace v krajských volbách 2020 (binární logistická regrese), model s vlivem pohlaví, věku, vzdělání, zaměstnání, velikosti místa bydliště a předchozí politické participace

	CATI			CAWI			SOCIAL		
	B	Exp(B)	Sig.	B	Exp(B)	Sig.	B	Exp(B)	Sig.
Pohlaví (žena)	0,22	1,24	0,423	-0,23	0,79	0,255	-0,22	0,80	0,325
Věk (18-24 let)			0,305			0,238			0,268
25-34 let	0,72	2,05	0,378	0,54	1,72	0,278	0,11	1,12	0,869
35-44 let	0,03	1,03	0,966	0,19	1,20	0,707	0,52	1,69	0,453
45-54 let	0,34	1,41	0,673	-0,06	0,94	0,908	0,37	1,45	0,595
55-64 let	0,77	2,16	0,357	0,00	1,00	0,999	-0,32	0,73	0,642
65 a více let	0,97	2,63	0,251	0,58	1,79	0,251	0,09	1,09	0,894
Vzdělání (základní)			0,001			0,908			0,015
Střední bez maturity	0,46	1,58	0,317	0,10	1,11	0,746	-0,13	0,88	0,734
Střední s maturitou	1,21	3,36	0,010	-0,02	0,98	0,962	0,32	1,37	0,419
Vysokoškolské (včetně VOŠ)	1,70	5,48	0,002	0,16	1,18	0,663	1,04	2,82	0,034
Zaměstnání	-0,01	0,99	0,974	0,57	1,76	0,016	-0,53	0,59	0,061
Velikost místa bydliště (do 1999 obyvatel)			0,394			0,497			0,561
2000-4999 obyvatel	-0,26	0,77	0,519	-0,26	0,77	0,399	-0,01	0,99	0,983
5000-19999 obyvatel	0,28	1,32	0,412	-0,50	0,61	0,068	0,43	1,53	0,197
20000-49999 obyvatel	-0,49	0,61	0,238	-0,20	0,82	0,515	0,34	1,40	0,332
50000-99999 obyvatel	0,60	1,83	0,295	-0,25	0,78	0,473	0,06	1,06	0,874
100000 a více obyvatel	-0,14	0,87	0,761	-0,53	0,59	0,111	0,51	1,67	0,152
Účast ve volbách do KZ 2016	1,02	2,76	0,000	2,10	8,14	0,000	1,68	5,36	0,000
Účast ve volbách do PS 2017	0,90	2,46	0,001	0,91	2,48	0,000	0,71	2,04	0,003
N		467			933			738	
Negelkerke R ²		0,265			0,363			0,286	
Úspěšnost klasifikace		6 (74,9 - 80,9)			5,5 (77,6 - 83,1)			2,5 (80,8 - 83,3)	

Zdroj: Vlastní

Tabulka 11b: Párový model 3 - Politická participace ve volbách do poslanecké sněmovny 2017 (binární logistická regrese), model s vlivem pohlaví, věku, vzdělání, zaměstnání, velikosti místa bydliště a politické participace

	CATI			CAWI			SOCIAL		
	B	Exp(B)	Sig.	B	Exp(B)	Sig.	B	Exp(B)	Sig.
Pohlaví (žena)	-0,88	0,42	0,00	-0,93	0,39	0,000	-0,66	0,52	0,000
Věk (18-24 let)			0,557			0,005			0,475
25-34 let	0,89	2,43	0,395	0,25	1,29	0,618	-0,21	0,81	0,727
35-44 let	0,98	2,67	0,344	0,14	1,16	0,774	-0,40	0,67	0,494
45-54 let	1,03	2,80	0,324	0,68	1,98	0,183	-0,77	0,46	0,196
55-64 let	1,48	4,41	0,159	1,02	2,78	0,047	-0,39	0,68	0,512
65 a více let	1,09	2,98	0,305	1,01	2,76	0,048	-0,34	0,71	0,561
Vzdělání (základní)			0,010			0,008			0,430
Střední bez maturity	0,20	1,22	0,662	-0,76	0,47	0,017	-0,23	0,79	0,490
Střední s maturitou	0,69	2,00	0,128	-0,38	0,68	0,239	-0,11	0,90	0,754
Vysokoškolské (včetně VOŠ)	1,19	3,29	0,015	-0,01	0,99	0,983	0,19	1,20	0,620
Zaměstnání	0,05	1,05	0,859	-0,08	0,93	0,732	-0,01	0,99	0,973
Velikost místa bydliště (do 1999 obyvatel)			0,724			0,280			0,522
2000-4999 obyvatel	0,21	1,23	0,574	0,43	1,54	0,127	-0,39	0,68	0,212
5000-19999 obyvatel	0,16	1,18	0,573	0,27	1,31	0,266	-0,32	0,73	0,255
20000-49999 obyvatel	0,57	1,78	0,117	0,65	1,92	0,018	-0,46	0,63	0,118
50000-99999 obyvatel	0,27	1,30	0,512	0,17	1,19	0,570	-0,36	0,70	0,246
100000 a více obyvatel	0,39	1,47	0,320	0,23	1,26	0,442	-0,55	0,58	0,060
Účast ve volbách do KZ 2016	1,01	2,76	0,000	2,04	7,67	0,000	2,15	8,61	0,000
Účast ve volbách do KZ 2020	0,96	2,61	0,001	0,92	2,52	0,000	0,74	2,09	0,002
N	467			933			738		
Negelkerke R ²	0,239			0,394			0,31		
Úspěšnost klasifikace	13,9 (54,8 - 68,7)			12,1 (68,1 - 80,2)			14,8 (61,3 - 76,1)		

Zdroj: Vlastní

Předposledním párovým modelem je model logistické regresní analýzy, který k již aplikovaným nezávislým proměnným přidává politickou participaci. U prvního modelu (vliv na politickou participaci ve volbách do krajských zastupitelstev 2020, Tabulka 11a) se jedná o deklarovanou politickou participaci v předešlých krajských volbách v roce 2016 a předešlých volbách do poslanecké sněmovny v roce 2017. U druhého modelu (vliv na politickou participaci ve volbách do poslanecké sněmovny 2017, Tabulka 11b) se jedná o deklarovanou politickou participaci v předešlých volbách do krajských zastupitelstev v roce 2016 a deklarovanou predikci politické participace v nadcházejících volbách do krajských zastupitelstev v roce 2020. U obou těchto proměnných lze očekávat silný signifikantní pozitivní vliv na politickou participaci.

Při rekapitulaci vlivu předešlých proměnných došlo k jeho oslabení. Signifikantní zůstal v případě prvního modelu u CATI módů sběru dat jen efekt vzdělání a ne v celé své šíři, ačkoli lze sledovat i pozitivní vliv chybějící kategorie (středoškolské vzdělání bez maturity) na nesignifikantní úrovni. U druhého modelu je u CATI sběru situace obdobná, kromě signifikantního pozitivního vlivu vzdělání a nesignifikantního pozitivního vlivu věku je ale patrný i falešný signifikantní vliv pohlaví.

V rámci CAWI módu sběru dat a prvního modelu zůstal z předchozích modelů signifikantní pouze vliv zaměstnání. Nelze sledovat tendenční pozitivní vliv v případě vzdělání a ani věku. U druhého modelu je prokázán signifikantní vliv vzdělání, ale pouze na úrovni kontrastní kategorie a středoškolského vzdělání bez maturity, která má ale negativní vliv na politickou participaci. Tento efekt je patrný i u předešlých modelů politické participace v krajských volbách v roce 2017 (Tabulka 9b, 10b) v rámci CAWI módu sběru dat a jedná se o další falešný vliv, jelikož vzdělání by mělo působit tendenčně pozitivně ve všech svých kategoriích (Linek, 2015, s. 87). Dále je patrný signifikantní vliv věku a falešný signifikantní vliv pohlaví.

Social mód sběru dat v prvním modelu rozlišuje signifikantní pozitivní vliv vzdělání v nejvyšší kategorii. V druhém modelu pak opět poukazuje na falešný signifikantní vliv pohlaví.

Nově přidané nezávislé proměnné mají shodně signifikantní a pozitivní vliv napříč všemi šesti analyzovanými případy. U CAWI a Social módu sběru dat má větší vliv vždy předchozí participace ve volbách do krajských zastupitelstev z roku 2016, u CATI sběru je vliv všech tří případů politické participace v obou modelech vyrovnaný.

U obou modelů došlo k nárůstu pseudo koeficientu determinace, zejména pak CAWI a Social módů sběru dat, které v rámci těchto párových modelů poprvé překonali CATI modely. Ke zlepšení došlo i u úspěšnosti klasifikace, u druhého párového modelu Social sběr překonal z hlediska úspěšnosti CATI sběr. Lze tedy konstatovat, že přidáním faktoru politické participace jako nezávislé proměnné bylo dosaženo nejvíce objasňujících modelů a vliv těchto nově přidaných proměnných výrazně pomohl zejména modelům postaveným na výběrových souborech sebraných metodou CAWI a Social.

Tabulka 12a: Model 4 - Politická participace v krajských volbách 2020 (binární logistická regrese), model s vlivem předchozí politické participace

	CATI			CAWI			SOCIAL		
	B	Exp(B)	Sig.	B	Exp(B)	Sig.	B	Exp(B)	Sig.
Pohlaví (žena)	-	-	-	-	-	-	-	-	-
Věk (18-24 let)			-			-			-
25-34 let	-	-	-	-	-	-	-	-	-
35-44 let	-	-	-	-	-	-	-	-	-
45-54 let	-	-	-	-	-	-	-	-	-
55-64 let	-	-	-	-	-	-	-	-	-
65 a více let	-	-	-	-	-	-	-	-	-
Vzdělání (základní)			-			-			-
Střední bez maturity	-	-	-	-	-	-	-	-	-
Střední s maturitou	-	-	-	-	-	-	-	-	-
Vysokoškolské (včetně VOŠ)	-	-	-	-	-	-	-	-	-
Zaměstnaní	-	-	-	-	-	-	-	-	-
Velikost místa bydliště (do 1999 obyvatel)			-			-			-
2000-4999 obyvatel	-	-	-	-	-	-	-	-	-
5000-19999 obyvatel	-	-	-	-	-	-	-	-	-
20000-49999 obyvatel	-	-	-	-	-	-	-	-	-
50000-99999 obyvatel	-	-	-	-	-	-	-	-	-
100000 a více obyvatel	-	-	-	-	-	-	-	-	-
Účast ve volbách do KZ 2016	1,38	3,98	0,000	2,07	7,95	0,000	1,74	5,70	0,000
Účast ve volbách do PS 2017	1,05	2,85	0,000	0,92	2,52	0,000	0,79	2,21	0,001
N	467			933			738		
Negelkerke R ²	0,206			0,341			0,243		
Úspěšnost klasifikace	2,3 (74,4 - 76,7)			5 (77,6 - 82,6)			0,5 (80,8 - 81,2)		

Zdroj: Vlastní

Tabulka 12b: Model 4 - Politická participace ve volbách do poslanecké sněmovny 2017 (binární logistická regrese), model s vlivem politické participace

	CATI			CAWI			SOCIAL		
	B	Exp(B)	Sig.	B	Exp(B)	Sig.	B	Exp(B)	Sig.
Pohlaví (žena)	-	-	-	-	-	-	-	-	-
Věk (18-24 let)			-			-			-
25-34 let	-	-	-	-	-	-	-	-	-
35-44 let	-	-	-	-	-	-	-	-	-
45-54 let	-	-	-	-	-	-	-	-	-
55-64 let	-	-	-	-	-	-	-	-	-
65 a více let	-	-	-	-	-	-	-	-	-
Vzdělání (základní)			-			-			-
Střední bez maturity	-	-	-	-	-	-	-	-	-
Střední s maturitou	-	-	-	-	-	-	-	-	-
Vysokoškolské (včetně VOŠ)	-	-	-	-	-	-	-	-	-
Zaměstnaní	-	-	-	-	-	-	-	-	-
Velikost místa bydliště (do 1999 obyvatel)			-			-			-
2000-4999 obyvatel	-	-	-	-	-	-	-	-	-
5000-19999 obyvatel	-	-	-	-	-	-	-	-	-
20000-49999 obyvatel	-	-	-	-	-	-	-	-	-
50000-99999 obyvatel	-	-	-	-	-	-	-	-	-
100000 a více obyvatel	-	-	-	-	-	-	-	-	-
Účast ve volbách do KZ 2016	1,03	2,81	0,000	2,00	7,37	0,000	2,06	7,86	0,000
Účast ve volbách do KZ 2020	1,05	2,85	0,000	0,92	2,52	0,000	0,79	2,21	0,001
N	467			933			738		
Negelkerke R ²	0,149			0,327			0,273		
Úspěšnost klasifikace	8 (55,9 - 63,9)			10,7 (68,1 - 78,8)			13,5 (61,4 - 74,9)		

Zdroj: Vlastní

V návaznosti na párové modely zahrnující veškeré sociodemografické ukazatele a politickou participaci (Tabulka 11a, 11b) se nabízí poslední sada párových modelů. Tentokrát do analýzy vstupuje jako nezávislá proměnná pouze politická participace, ať už deklarovaná minulá (vliv na politickou participaci ve volbách do krajských zastupitelstev 2020, Tabulka 12a), tak kombinace deklarované minulé a deklarované predikce budoucí participace (vliv na politickou participaci ve volbách do poslanecké sněmovny 2017, Tabulka 12b).

Došlo k zachování signifikantních a pozitivních vlivů nezávislých proměnných na úrovni všech šesti případů. Shodně s předchozím párovým modelem (Tabulka 11a, 11b) zůstala zachována i míra velikosti vlivu, u CATI módu sběru dat je patrná vyrovnanost napříč nezávislými proměnnými, zatímco u CAWI a Social má dominantní pozici z hlediska pozitivního vlivu na politickou participaci předešlá participace ve volbách do krajských zastupitelstev z roku 2016.

Zajímavé je i srovnání pseudo koeficientů determinace, které navazují na zjištění u předešlé dvojice modelů (Tabulka 11a, 11b). Pro CATI sběr je vliv nezávislých proměnných ve formě politické participace méně vyčerpávající z hlediska hodnoty Negelkerkeho R^2 než u zbylých dvou módů sběru dat. To samé platí v případě CAWI u obou modelů i v rámci úspěšnosti klasifikace, v případě Social sběru pak jen u druhého modelu (Tabulka 12b).

4.2.1 Analýza faktorů ovlivňujících politickou participaci - shrnutí

Jak bylo v předchozích modelech prokázáno, primární pozitivní signifikantní vliv na politickou participaci má předchozí a deklarovaná budoucí politická participace. Signifikantní pozitivní vliv vzdělání je patrný nejčastěji a konzistentně je patrný jen u CATI sběru, zejména v případě modelů, do kterých vstupuje větší množství nezávislých proměnných. Pozitivní vliv věku v celé šíři všech kategorií je na nesignifikantní úrovni přítomen u CATI a CAWI módu sběru dat v rámci robustnějších modelů. Falešně negativní a signifikantní vliv pohlaví poté v tomto modelu vykazují všechny tři sběry. Velikost místa bydliště, která dle předchozích výzkumů signifikantně neovlivňuje volební participaci je poté falešně signifikantně pozitivní v jednom z modelů u CAWI módu sběru dat v kategorii respondentů z měst o velikosti 20000-49999 obyvatel. Signifikantní vliv zaměstnanosti se podařilo prokázat pouze v případě CAWI sběru a to i v rámci modelu, do kterého vstupovaly všechny proměnné. Tato zjištění lze nalézt i v Tabulce 13.

Tabulka 13: Shrnutí faktorů se signifikantním vlivem na politickou participaci dle módu sběru dat

	CATI	CAWI	Social
Pohlaví	Ano	Ano	Ano
Věk	Ano, v celé šíři	Ano, částečně	Ano, částečně
Vzdělání	Ano, částečně	Ano, v celé šíři	Ano, v celé šíři
Zaměstnanost	Ne	Ano	Ne
Velikost místa bydliště	Ne	Ano, částečně	Ne
Politická participace	Ano	Ano	Ano

Zdroj: Vlastní

Při analýze párových modelů politické participace ve volbách do krajských zastupitelstev v roce 2020 a voleb do poslanecké sněmovny v roce 2017 se jednotlivé módy sběru dat jasně

odlišovaly. Dominantní ovlivňující proměnné byly vždy shodně předchozí či deklarovaná budoucí politická participace, tím ale podobnost napříč módy sběru dat končí.

Možným vysvětlením této skutečnosti je poté systematická chyba pokrytí populace, která v tomto případě staví kvalitu dat získaných z pravděpodobnostního náhodného výběru nad kvótní výběrové šetření na online panelu a online sociální síti (Fielding et al., 2008, s. 200).

Analýza faktorů ovlivňujících politickou participaci tak ve svém důsledku prohloubila závěry analýzy přesnosti měření politické participace z předchozí kapitoly analytické části práce a ačkoli byla druhá hypotéza již zamítnuta, nyní je díky osmi modelům logistické regrese tento závěr podruhé validován.

H02: Data sebraná odlišnými módy sběru dat, tedy CATI, CAWI Panel (CAWI) a CAWI Facebook (Social), jsou ekvivalentní z hlediska politické participace

4.3 Doprovodné analýzy módů

Předchozí kapitoly obsahující analýzu politických preferencí, politické participace a logistickou regresi faktorů ovlivňujících politickou participaci vedly k zamítnutí všech tří testovaných nulových hypotéz. Jednotlivé módy sběru dat, potažmo data samotná, se liší. Pozorný čtenář se ale v tomto místě může ptát, zdali je možné srovnat i časovou a finanční efektivitu jednotlivých módů, která je nedílným hodnotícím kritériem z hlediska celkové kvality sběru dat.

Tabulka 14: Míra návratnosti za jednotlivé módy sběru

	CATI	CAWI	Social
Kompletní rozhovory	485	1135	832
Částečně provedené rozhovory	65	9	2759
Odmítnuté rozhovory	2030	139	53
Nekontaktovaní	2454	2323	72872
Míra návratnosti	9,6 %	31,5 %	1,1 %

Zdroj: Vlastní

Začněme srovnáním u míry návratnosti. V teoretické části bylo zmíněno, že právě tento aspekt je stále častěji skloňován (nejen) u pravděpodobnostních výběrů v kontextu dlouhodobého

tendenčního poklesu. Právě návratnost silně ovlivňuje chybu výpadku návratnosti. V tomto případě je míra návratnosti u online kvótního dotazníkového šetření více než trojnásobná ve srovnání s prostým náhodným výběrovým šetřením formou CATI. Kvótní nepravděpodobnostní sběr na sociální síti je specifický ze své podstaty a míra návratnosti není přímo porovnatelná s ostatními módy sběru dat kvůli diametrálně odlišné formě cílení a oslovení respondentů.

Druhým kritériem je množství dokončených rozhovorů (v tomto případě dotazníků), které byly provedeny ve stejném časovém období jednoho týdne. CAWI a Social (i přes velmi malou míru návratnosti) módy sběru dat překonaly CATI sběr, přičemž skrze CATI nedošlo k nasbírání ani 500 kompletních dotazníků.

Třetí kritérium je poté finanční náročnost, kterou ukazuje Tabulka 15. Jedná se o indexaci finančních nákladů spojených se sběrem dle jednotlivých módů. Do finanční náročnosti spadá v případě CATI sběru čas tazatelů, supervizorů, cena za telefonické spojení s respondenty a cena za generování náhodné databáze čísel. U CAWI sběru se do ceny započítává incentiva pro respondenty a čas na straně panelu se založením a cílením projektu. V případě Social módu sběru dat je poté cena určena zejména náklady spojenými s reklamními kampaněmi na Facebooku, skrze které byl dotazník distribuován mezi respondenty. Všechny ostatní náklady spojené s přípravou dotazníku a jeho nascriptováním jsou pro všechny tři módy považovány za shodné a do srovnání nevstupují.

Tabulka 15: Finanční náročnost na sběr za jednotlivé módy sběru (index)

CATI	CAWI	Social
1,00	0,63	0,35

Zdroj: Vlastní

CATI mód sběru dat je v tomto případě benchmarkem pro ostatní módy, CAWI sběr dosáhl nákladů na sběr ve velikosti 63 % vůči CATI sběru a Social sběr 35 %. Na tomto místě je nutné podotknout, že tento údaj nelze generalizovat, tento sběr dat byl specifický svým zacílením popsáním v metodologické části práce a pro kvótní sběr obsahoval náročnější kritéria na plnění kvót na úrovni jednotlivých krajů než např. běžný výzkum veřejného mínění na cílové skupině běžné reprezentativní populace celé České republiky.

5 Závěr

Závěrem práce je vhodné připomenout původní výzkumný cíl. Tato diplomová práce měla za cíl ověřit ekvivalenci měření tří rozdílných módů sběru dat a poukázat na možná faktory na úrovni chyb měření, které mohou na jednotlivé módy působit. Analýza vedla k zamítnutí všech tří nulových hypotéz:

H₀1: Data sebraná odlišnými módy sběru dat, tedy CATI, CAWI Panel (CAWI) a CAWI Facebook (Social), jsou ekvivalentní z hlediska politické preference.

H₀2: Data sebraná odlišnými módy sběru dat, tedy CATI, CAWI Panel (CAWI) a CAWI Facebook (Social), jsou ekvivalentní z hlediska politické participace.

H₀3: Data sebraná odlišnými módy sběru dat, tedy CATI, CAWI Panel (CAWI) a CAWI Facebook (Social), stejnou měrou reprezentují skutečný stav z hlediska politické participace a preference (jsou tedy stejně validní).

K analýze byly použity metody testování ekvivalence měření na úrovni analýz prvního a druhého stupně spolu s binární logistickou regresí. Tyto analytické postupy poukázaly na jasnou nesrovnatelnost měření jednotlivých módů sběru dat na úrovni politické participace, preference a přesnosti měření.

V první části analytické sekce byla testována přesnost jednotlivých módů sběru dat na základě deklarované volební preference ve volbách do poslanecké sněmovny v roce 2017. Skrze analýzu odchylek bylo jasně poukázáno na úroveň přesnosti dat, kde vyšel nejlépe mód sběru dat skrze online kvótní dotazníkové šetření při srovnání se skutečnou pravdivou hodnotou. Při srovnání s průměrem volebních modelů vydaných v době měření bylo ale poukázáno i na vysokou validitu CATI módu sběru dat. Doprovodnou informací bylo poukázání na vliv tazatele, resp. sociální desirabilitu, která je nedílně spjatá s chybou měření, kde autonomní módy sběru dat nedosahovaly zkreslení výsledků u volební preference krajně pravicových a levicových stran a hnutí jako u CATI sběru.

Navazující sekce podrobila stejné analýze politickou participaci. V tomto srovnání dosáhl největší přesnosti Social mód sběru dat a to jak v porovnání se skutečnou pravdivou hodnotou, tak v porovnání s průměrem tří volebních modelů ze září 2020. Naopak největší odchylky v obou případech srovnání politické participace vykazoval CATI sběr.

Třetí část měla za cíl poukázat na interní koherenci dat a latentní vzorce, potažmo korelaci závislé proměnné a nezávislých proměnných v binární logistické regresi. Ta odhalovala vliv sociodemografických ukazatelů a předchozí či deklarované budoucí politické participace na závislou proměnnou politické participace. Výsledky jednotlivých modelů pak byly porovnány s validovanými výsledky ostatních výzkumů provedenými v kontextu České republiky. Ačkoli žádný z modelů na základě dat sebraných pomocí tří rozdílných módů plně neobstál a neodhalil veškeré faktory ovlivňující politickou participaci na signifikantní úrovni, konzistentně nejsrovnatelnější výsledky podával CATI sběr, zatímco Social sběr vykazoval největší inkonzistence. A priori vysvětlení je navázané na chybu v pokrytí populace, kde opory výběru pro CAWI i Social sběr mohou vykazovat systematicky odlišné charakteristiky základního souboru.

Doprovodné analýzy se poté primárně zaměřily na efektivitu jednotlivých módů sběru dat. Časově i finančně je nejméně efektivní CATI sběr. Vykazuje také nižší míru návratnosti nežli CAWI mód sběru dat a stoupá tak riziko spojené s chybou výpadku návratnosti.

Syntéza těchto poznatků není jednoznačná. Každý z módů podrobených analýze se liší, má své přednosti i nedostatky a je zatížen inklinací k rozdílným výzkumným chybám. Na praktickém příkladě z aplikovaného výzkumu v oblasti marketingu a výzkumu trhu a veřejného mínění se ale ukazuje, že je žádoucí jednotlivé módy kombinovat, data sbírat skrze smíšený mód sběru a minimalizovat tak slabiny jednotlivých módů při současné maximalizaci jejich benefitů.

6 Diskuze nad výsledky práce

Hlavní přínos této práce spočívá v analýze reálných dat sebraných agenturou NMS Market Research, které reflektují každodenní proces sběru dat u aplikovaného výzkumu v oblasti výzkumu veřejného mínění a výzkumu trhu. Data staví na předvolebním výzkumu z roku 2020 a testují ekvivalenci měření jednotlivých módů sběru dat, skrze které výzkum probíhal.

Testované hypotézy a provedené analýzy jasně poukázaly na rozdíly jak na úrovni samotného měření hodnot, tak na rozdíly ve struktuře a konzistenci dat., respektive přítomnosti latentních faktorů ovlivňujících politickou participaci. Na tomto místě nelze jednoznačně určit, který ze srovnávaných módů sběru dat dopadl celkově nejlépe, jelikož každý byl ovlivněn jinou chybou.

Analýza politické preference poukázala na chybu měření, respektive vliv tazatele a sociální desirability u CATI sběru na úrovni nejvyšších odchylek od skutečných pravdivých hodnot u preference krajně pravicových a levicových hnutí a stran. Následná analýza latentních vlivů a konzistence vývěrových souborů naopak u CATI sběru poukázal na fakt, že dokáže nejlépe popisovat klíčové faktory ovlivňující politickou participaci. Zde je možné výsledky interpretovat skrze chybu spojenou s pokrytím vzorku, kde u CAWI a Social sběru základní populaci nahrazuje množina respondentů, kteří jsou online, a v případě CAWI jsou motivováni incentivou za poskytnutí svých dat. Při srovnání jednotlivých módů z hlediska míry návratnosti a efektivity na úrovni času a nákladů spojených se sběrem nejlépe dopadl CAWI mód sběru dat.

Výsledky tedy poukazují na potřebu sběru dat skrze smíšené módy za účelem minimalizace jednotlivých chyb, kterými jsou dílčí módy více či méně zatíženy. Lze tak využít silných stránek, které módy skýtají a minimalizovat jejich nevýhody. Podobný přístup se zdá být jako ideální vyústění debat o srovnání kvalit jednotlivých módů sběru dat (Dillman, 2020, s. 38).

V závěru je nutno poukázat na limitace této práce. Analýza samotná neměla za cíl vyčerpávajícím způsobem definovat jednotlivé chyby, které kauzálně ovlivňují kvalitu dat zkoumaných módů, ale poukázat na ekvivalenci módů samotných. Otázku, zdali je takový výzkum vůbec možný, jelikož odfiltrování souběžných vlivů chyb, módů a výběrů je minimálně velmi obtížné (Jäckle et al., 2010, s. 18), ponechám otevřenou. Svůj původní cíl poukázat na přítomnost či nepřítomnost srovnatelnosti módů skrze ekvivalenci měření práce splnila, v dalších krocích je ale určitě vhodné navázat podobným výzkumem a použít mnohorozměrné statistické analýzy pro získání hlubšího vhledu do problematiky ekvivalence měření. Zároveň by bylo vhodné použít data sebraná na celorepublikové úrovni a vyvarovat se výzkumům zaměřeným jen na specifické kraje, což byl i případ této práce vzhledem k regionálnímu zaměření voleb do krajských zastupitelstev.

7 Seznam použitých zdrojů

Alastalo, M. (2008). The History of Social Research Methods. In P. Alasuutari, L. Bickman, & J. Brannen, *The SAGE Handbook of Social Research Methods* (s. 26–41). SAGE Publications Ltd. <https://doi.org/10.4135/9781446212165.n3>

Ansolabehere, S., & Rivers, D. (2013). Cooperative Survey Research. *Annual Review of Political Science*, 16(1), 307–329. <https://doi.org/10.1146/annurev-polisci-022811-160625>

Anýžová, P. (2013). Item Equivalence in Cross-National Survey Research: Basic Definition and Possibilities for Analysis. *Data and Research – SDA Info*, 7(1), 29. <https://doi.org/10.13060/1802-8152.2013.7.1.2>

Bethlehem, J. G. (2009). *Applied survey methods: A statistical perspective*. Wiley.

Biemer, P. P. (2016). Total Survey Error Paradigm: Theory and Practice. In C. Wolf, D. Joye, T. Smith, & Y. Fu, *The SAGE Handbook of Survey Methodology* (s. 122–141). SAGE Publications Ltd. <https://doi.org/10.4135/9781473957893.n10>

Biemer, P. P., & Lyberg, L. (2003). *Introduction to survey quality*. Wiley.

Bláha, P., Havlínová, I., Maškarinec, P., & Vojtíšková, K. (2017). *Politická participace: Současné trendy, příležitosti a problémy jako inspirace pro Ústecký kraj*.

Bowling, A. (2005). Mode of questionnaire administration can have serious effects on data quality. *Journal of Public Health*, 27(3), 281–291. <https://doi.org/10.1093/pubmed/fdi031>

Brickman Bhutta, C. (2012). Not by the Book: Facebook as a Sampling Frame. *Sociological Methods & Research*, 41(1), 57–88. <https://doi.org/10.1177/0049124112440795>

Buchtík, M. (2012). Smíšené módy sběru dat v kvantitativním sociálně vědním výzkumu. *Data a výzkum – SDA info*, 6(2), 129–149.

CVVM, Sociologický ústav AV ČR. (b.r.). *Co jsou to stranické preference, voličské preference a volební model? Získáno 26. červenec 2021, z <https://cvvm.soc.cas.cz/cz/caste-dotazy/4581-co-jsou-to-stranicke-preference-volicke-preference-a-volebni-model>*

Český statistický úřad. (2020, březen 31). *Úroveň formálního vzdělání a účast v neformálním vzdělávání ve věkové skupině 15-64 let—4. Čtvrtletí 2019. Zaměstnanost a nezaměstnanost podle výsledků VŠPS - 4. čtvrtletí 2019*. <https://www.czso.cz/csu/czso/zamestnanost-a-nezamestnanost-podle-vysledku-vsps-ctvrtletni-udaje-4-ctvrtleti-2019>

ČNP. (2021). *Národní panel*. <https://www.ceskynarodnipanel.cz/>

- Davidov, E., Meuleman, B., Ciecuch, J., Schmidt, P., & Billiet, J. (2014). Measurement Equivalence in Cross-National Research. *Annual Review of Sociology*, 40(1), 55–75. <https://doi.org/10.1146/annurev-soc-071913-043137>
- de Heer, W., de Leeuw, E. D., & van der Zouwen, J. (1999). Methodological Issues in Survey Research: A Historical Review. *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique*, 64(1), 25–48. <https://doi.org/10.1177/075910639906400104>
- de Leeuw, E. D. (2005). To Mix or Not to Mix Data Collection Modes in Surveys. *Journal of Official Statistics*, 21(2), 233–255.
- Dillman, D. A. (2020). Towards Survey Response Rate Theories That No Longer Pass Each Other Like Strangers in the Night. In P. S. Brenner (Ed.), *Understanding Survey Methodology* (Roč. 4, s. 15–44). Springer International Publishing. https://doi.org/10.1007/978-3-030-47256-6_2
- Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014). *Internet, phone, mail, and mixed-mode surveys: The tailored design method* (4th edition). Wiley.
- Erens, B., Burkill, S., Couper, M. P., Conrad, F., Clifton, S., Tanton, C., Phelps, A., Datta, J., Mercer, C. H., Sonnenberg, P., Prah, P., Mitchell, K. R., Wellings, K., Johnson, A. M., & Copas, A. J. (2014). Nonprobability Web Surveys to Measure Sexual Behaviors and Attitudes in the General Population: A Comparison With a Probability Sample Interview Survey. *Journal of Medical Internet Research*, 16(12), e276. <https://doi.org/10.2196/jmir.3382>
- ESOMAR. (2020). *Global Market Research 2020*. ESOMAR B.V. <https://ana.esomar.org/documents/global-market-research-2020>
- Fielding, N., Lee, R. M., & Blank, G. (Ed.). (2008). *The SAGE handbook of online research methods*. London : SAGE.
- Groves, R. M. (1989). *Survey Errors and Survey Costs*. <https://nbn-resolving.org/urn:nbn:de:101:1-201410214330>
- Groves, R. M. (Ed.). (2002). *Survey nonresponse*. Wiley.
- Groves, R. M. (Ed.). (2004). *Survey methodology*. J. Wiley.
- Groves, R. M., & Lyberg, L. (2010). Total Survey Error: Past, Present, and Future. *Public Opinion Quarterly*, 74(5), 849–879. <https://doi.org/10.1093/poq/nfq065>
- Hochstim, J. R. (1967). A Critical Comparison of Three Strategies of Collecting Data from Households. *Journal of the American Statistical Association*, 62(319), 976–989. <https://doi.org/10.1080/01621459.1967.10500909>

- Chylíková, J. (2011). *Úvod do problematiky výzkumu citlivých témat ve výběrových šetřeních*.
- Jäckle, A., Roberts, C., & Lynn, P. (2010). Assessing the Effect of Data Collection Mode on Measurement. *International Statistical Review*, 78(1), 3–20. <https://doi.org/10.1111/j.1751-5823.2010.00102.x>
- Kaplan, R. M., & Saccuzzo, D. P. (2017). *Psychological testing: Principles, applications, & issues* (Ninth edition). Cengage Learning.
- Keeter, S., Miller, C., Kohut, A., Groves, R. M., & Presser, S. (2000). Consequences of Reducing Nonresponse in a National Telephone Survey. *Public Opinion Quarterly*, 64(2), 125–148. <https://doi.org/10.1086/317759>
- Keusch, F. (2015). Why do people participate in Web surveys? Applying survey participation theory to Internet survey data collection. *Management Review Quarterly*, 65(3), 183–216. <https://doi.org/10.1007/s11301-014-0111-y>
- Klausch, T., Hox, J. J., & Schouten, B. (2013). Measurement Effects of Survey Mode on the Equivalence of Attitudinal Rating Scale Questions. *Sociological Methods & Research*, 42(3), 227–263. <https://doi.org/10.1177/0049124113500480>
- Krejčí, J. (2007). Non-Response in Probability Sample Surveys in the Czech Republic. *Sociologický časopis*, 43 (3).
- Krejčí, J. (2008). *Kvalita sociálněvědních výběrových šetření v České republice* (Vyd. 1). Sociologické Nakl., SLON.
- Leeuw, E. D. de, Hox, J. J., Dillman, D. A., & European Association of Methodology (Ed.). (2008). *International handbook of survey methodology*. Lawrence Erlbaum Associates.
- Linek, L. (2015). Účast ve sněmovních volbách v roce 2013: Zdroje, mobilizace a motivace. *Evropské volební studie*, 10, 76–90.
- Linek, L., & Petrušek, I. (2018). Development of Educational Inequalities in Political Participation in Czech Republic and Slovakia and their Implications for Representation of Attitudes. *Sociológia - Slovak Sociological Review*, 5. <https://doi.org/10.31577/sociologia.2018.50.5.20>
- MacInnis, B., Krosnick, J. A., Ho, A. S., & Cho, M.-J. (2018). The Accuracy of Measurements with Probability and Nonprobability Survey Samples: Replication and Extension. *Public Opinion Quarterly*, 82(4), 707–744. <https://doi.org/10.1093/poq/nfy038>

- Millar, M. M., & Dillman, D. A. (2011). Improving Response to Web and Mixed-Mode Surveys. *Public Opinion Quarterly*, 75(2), 249–269. <https://doi.org/10.1093/poq/nfr003>
- Pilecká, J. (2016). *Smíšené módy sběru dat: PAPI a CAPI srovnání*. <https://is.cuni.cz/webapps/zzp/detail/165742>
- Pilnáček, M. (2020). *Volební model v září 2020*. CVVM, Sociologický ústav AV ČR. <https://cvvm.soc.cas.cz/cz/tiskove-zpravy/politicke/volby-a-strany/5285-volebni-model-v-zari-2020>
- Porter, S. R., & Whitcomb, M. E. (2004). Understanding the effect of prizes on response rates. *New Directions for Institutional Research*, 2004(121), 51–62. <https://doi.org/10.1002/ir.100>
- Prokop, D. (2020, listopad 11). *Analýza výsledků krajských voleb 2020*. PAQ Research. <https://www.paqresearch.cz/post/analyza-vysledku-krajskych-voleb-2020>
- Randolph, J. J., Virnes, M., Jormanainen, I., & Eronen, P. J. (2006). The Effects of a Computer-Assisted Interview Tool on Data Quality. *J. Educ. Technol. Soc.*, 9(3), 195–205.
- Revilla, M. (2010). Quality in Unimode and Mixed-Mode designs: A Multitrait-Multimethod approach. *Survey Research Methods*, Vol 4, 151-164 Pages. <https://doi.org/10.18148/SRM/2010.V4I3.4278>
- Roser, M., Ritchie, H., & Ortiz-Ospina, E. (2015). Internet. *Our World in Data*. <https://ourworldindata.org/internet>
- Schwarz, N. (2007). Cognitive aspects of survey methodology. *Applied Cognitive Psychology*, 21(2), 277–287. <https://doi.org/10.1002/acp.1340>
- SIMAR. (2021a). *VÝROČNÍ ZPRÁVA O STAVU VÝZKUMNÉHO TRHU 2020*. SIMAR. https://simar.cz/assets/media/files/assets/uploads/SIMAR_vyrocka_2020_final.pdf
- SIMAR. (2021b). *Výzkumný Standard 2—Kvantitativní online výzkum*. https://simar.cz/assets/media/files/standardy/Standard_02.pdf
- Soukup, P., & Rabušic, L. (2007). Some Notes on the Obsession of the Czech Social Sciences with Statistical Significance. *Czech Sociological Review*, 43(2), 379–396. <https://doi.org/10.13060/00380288.2007.43.2.06>
- STEM. (2020, září). *Volební preference*. <https://www.stem.cz/volebni-preference-stem-zari-2020/>

Stiglbauer, B., Gnams, T., & Gamsjäger, M. (2011). The Interactive Effects of Motivations and Trust in Anonymity on Adolescents' Enduring Participation in Web-Based Social Science Research: A Longitudinal Behavioral Analysis. *International Journal of Internet Science*, 29, 29–43.

Šabotník, P., Ranocha, P., Bucek, P., Horna, D., Kubíček, J., & Týmová, R. (2020). *Trendy Česka '20*. Česká Televize, Kantar. https://ct24.ceskatelevize.cz/sites/default/files/2506702-trendy_ceska_2020_vlna_63_volebni_model.pdf

Vinopal, J. (2008). *Kognitivní přístupy v metodologii výzkumných šetření—Metoda okamžité validizace*. Sociologický ústav AV ČR.

Wang, W., Rothschild, D., Goel, S., & Gelman, A. (2015). Forecasting elections with non-representative polls. *International Journal of Forecasting*, 31(3), 980–991. <https://doi.org/10.1016/j.ijforecast.2014.06.001>

Williams, D., & Brick, J. M. (2018). Trends in U.S. Face-To-Face Household Survey Nonresponse and Level of Effort. *Journal of Survey Statistics and Methodology*, 6(2), 186–211. <https://doi.org/10.1093/jssam/smx019>

Yang, K., & Banamah, A. (2014). Quota Sampling as an Alternative to Probability Sampling? An Experimental Study. *Sociological Research Online*, 19(1), 56–66. <https://doi.org/10.5153/sro.3199>