# UNIVERZITA KARLOVA

FAKULTA SOCIÁLNÍCH VĚD

Institut Sociologických studií

Katedra Sociologie


# Návrh metodologické optimalizace volebního modelu Median na základě poznatků Czech Household Panel Study

Diplomová práce


Autor práce: Bc. Michal Kunc

Studijní program: Sociologie

Vedoucí práce: PhDr. Martin Buchtík, Ph.D.

Konzultant: Mgr. et Mgr. Daniel Prokop



Rok obhajoby: 2019

**Prohlášení**

1. Prohlašuji, že jsem předkládanou práci zpracoval samostatně a použil jen uvedené prameny a literaturu.
2. Prohlašuji, že práce nebyla využita k získání jiného titulu.
3. Souhlasím s tím, aby práce byla zpřístupněna pro studijní a výzkumné účely.

V Praze dne 9. 5. 2019 Michal Kunc

## Bibliografický záznam

KUNC, Michal. *Návrh metodologické optimalizace volebního modelu Median na základě poznatků Czech Household Panel Study.* Praha, 2019. 65 s. Diplomová práce (Mgr.). Univerzita Karlova, Fakulta sociálních věd, Institut Sociologických studií. Katedra Sociologie. Vedoucí diplomové práce PhDr. Martin Buchtík, Ph.D.

**Rozsah práce:** 108 183 znaků včetně mezer bez abstraktu a příloh

## Klíčová slova

sekundární analýza dat, konfirmační analýza, volební chování, volební model, předvolební výzkumy, volební účast, minulé volební chování, metodologie

## Keywords

secondary data analysis, confirmatory analysis, voting behavior, likely voter model, pre-election surveys, voter turnout, prior voting behavior, methodology

## Název práce

Návrh metodologické optimalizace volebního modelu Median na základě poznatků Czech Household Panel Study

**Poděkování**

Rád bych poděkoval zejména Danielu Prokopovi, bez jehož výzvy ke spolupráci a konzultace by tato práce nemohla nikdy vzniknout, a Martinovi Buchtíkovi, bez jehož vedení by tato práce pravděpodobně nebyla nikdy dokončena. Děkuji také společnosti Median za poskytnutí datového souboru a technického zázemí. Děkuji také Sociologickému ústavu AV ČR za poskytnutí datového souboru a potřebné dokumentace.

# Obsah

1. Úvod ... 3
1.1. Cíle práce ... 4
1.2. Struktura práce ... 4
2. Volební modely a jejich parametry ... 5
2.1. Volební model ... 6
2.2. Důležité parametry ve volebních modelech ... 7
2.2.1. Volební účast ... 7
2.2.2. Záměr volby preferované strany ... 10
2.2.3. Minulá volba a vliv uplynulého času ... 11
2.3. Proč posuzuji volební model s reálnými výsledky voleb? ... 12
3. Sekundární analýza dat ... 13
3.1. Data ... 13
3.1.1. Czech Household Panel Study ... 14
3.1.2. Omnibus Median ... 16
3.2. Metody analýzy ... 17
3.2.1. Volební účast, záměr volební účasti a minulá volební účast (CHPS) ... 17
3.2.2. Preferovaná strana (CHPS) ... 22
3.2.3. Minulé volební chování (Omnibus Median) ... 23
3.3. Výsledky Sekundární analýzy dat a jejich diskuse ... 25
3.3.1. Volební účast, záměr volební účasti a minulá volební účast ... 25
3.3.2. Preferovaná strana ... 28
3.3.3. Minulé volební chování – vliv uplynulého času ... 28
4. Navrhovaný volební model ... 30
4.1. Srovnávané výstupy ... 31
4.1.1. Model MEDIAN ... 32
4.1.2. Model CVVM2017 ... 33

4.1.3. Model CVVM2018 ..... 33
4.1.4. Model KANTAR ..... 35
4.2. Konstrukce navrhovaného volebního modelu ..... 35
4.2.1. Jak v navrhovaném volebním modelu fungují parametry ..... 36
4.2.2. Účast ..... 38
4.2.3. Jistota záměru volby zvažované strany ..... 39
4.2.4. Minule volená strana a vliv času uplynulého od minulých voleb ..... 40
4.3. Kritéria srovnávání volebních modelů ..... 42
4.4. Posouzení navrhovaného volebního modelu (výsledky) ..... 43
5. Diskuse navrhovaného modelu ..... 46
6. Závěry ..... 48
7. Summary ..... 51
8. Použitá literatura ..... 52
Projekt diplomové práce ..... 58
Přílohy ..... 64

# 1. Úvod

Výstupy předvolebních výzkumů jsou jedny z mediálně nejviditelnějších a pravidelně diskutovaných produktů výzkumných agentur[1] v České republice, zejména v měsících předcházejících volbám. Někteří aktéři je kritizují, jiní se podle nich před volbami rozhodují, některé politické subjekty na nich staví svou předvolební komunikaci. Výstupy předvolebních výzkumů se kromě toho, jaké konkrétní údaje prezentují, liší často i tím, jakým způsobem tyto produkty (ač nominálně stejného druhu) vytvářejí. Do těchto detailů tvorby výstupů předvolebních výzkumů jejich příjemci z řad široké veřejnosti většinou nemají podrobný vhled. Přesto tyto informace prostřednictvím médií přijímají a v mnoha případech na ně také reagují.

Tato dynamika vkládá do rukou tvůrců výstupů předvolebních výzkumů určitou moc, a tím pádem i zodpovědnost. To považuji za důležitý důvod, proč by představitelům agentur, které se předvolebním výzkumům věnují, mělo záležet na tom, jaké jsou vlastnosti jimi publikovaných výstupů. Nikoliv pouze kvůli zachování vlastní expertní kredibility a plnění standardů profesních organizací (přestože ani důležitost těchto imperativů nechci v žádném případě podhodnocovat), ale i kvůli jejich možnému vlivu na společnost, politiku, a v důsledku i stav demokratických institucí.

Tato diplomová práce mi umožňuje se pod vedením mého vedoucího a mého konzultanta věnovat možným způsobům, jak konstrukci jednoho z výše zmíněných výstupů předvolebních výzkumů, konkrétně volebního modelu agentury Median, optimalizovat na základě poznatků sekundární analýzy dat z Českého panelového šetření domácností (CHPS) a omnibusového šetření společnosti Median.

Tomuto tématu se kromě výše zmíněných důvodů v diplomové práci věnuji i proto, že moci během studia spolupracovat na optimalizaci nástroje, který je současně skutečně používán v aplikovaném výzkumu, navíc s příslibem možného využití některých poznatků této práce v praxi, považuji za skvělou příležitost, jak napsat diplomovou práci, která může být nejen věcně podnětná, ale i reálně prakticky užitečná. Práce navíc propojuje poznatky z analýzy dat z akademického panelového výzkumu s konkrétním produktem

[1] Považuji za vhodné předem zmínit, že Centrum pro výzkum veřejného mínění není výzkumná agentura, ale výzkumné oddělení Sociologického ústavu AV ČR, v.v.i. Pokud v této práci mluvím o agenturách, mám tím, pro nedostatek lepšího obecného pojmu, na mysli i CVVM.

v aplikovaném agenturním předvolebním výzkumu, což považuji za vhodné nastavení pro mou magisterskou specializaci aplikovaného výzkumu a jeho metodologie.

## 1.1. Cíle práce

Hlavním cílem této diplomové práce je **optimalizovat konkrétní hodnoty parametru účasti, parametru jistoty záměru volby preferované strany a teoretických četností váhové proměnné minulého volebního chování**. Smyslem práce tedy není navrhnout konceptuálně nový způsob konstrukce volebního modelu, ale optimalizovat předem stanovené parametry modelu na základě poznatků ze sekundární analýzy dat. K dosažení hlavního cíle práce je potřeba postupně splnit následující tři na sebe navazující podcíle: 1) Pomocí sekundární analýzy dat z CHPS a omnibusu agentury Median ověřit hypotézy o výše zmíněných parametrech, 2) Nastavit konkrétní způsob zohlednění poznatků ze sekundární analýzy dat při konstrukci navrhovaného volebního modelu a 3) Porovnat výsledky navrhovaného způsobu konstrukce volebního modelu s výsledky dalších vybraných postupů (vypočítaných na jednom datovém souboru), a s výsledky voleb do PSP ČR v roce 2017.

## 1.2. Struktura práce

Vzhledem k tomu, že tato práce se na cestě ke svému hlavnímu cíli věnuje mnoha dílčím úkolům, považuji za vhodné věnovat v této úvodní kapitole několik odstavců její struktuře. Text jsem se sice snažil strukturovat tak, aby na sebe jednotlivé informace logicky navazovaly, a zároveň plynuly ke společným závěrům, v některých částech může být její členění nicméně stále poněkud neintuitivní. Pro větší přehlednost a jistotu, že kvůli struktuře práce nedojde k pominutí některých důležitých informací, ji nyní krátce popíšu.

Prvním velkým tematickým celkem textu je teoretická kapitola *2. Volební modely a jejich parametry*, kde se věnuji některým důležitým aspektům zkoumaných konceptů diskutovaných v odborné literatuře, a na jejich základě odůvodňuji konkrétní nastavení výzkumných problémů této práce.

Po ní následuje kapitola *3. Sekundární analýza dat*. V ní nejdříve popisuji datové zdroje, které v sekundární analýze dat zkoumám. Následně se věnuji popisu použitých metod sekundární analýzy – ty jsou (podobně jako každý větší celek textu této práce) děleny podle jednotlivých zkoumaných parametrů volebního modelu (volební účast, preferovaná strana, minulé volební chování). V každé z těchto částí popisuji příslušné hypotézy

sekundární analýzy dat a konkrétní metody analýzy vedoucí k jejich ověření. V závěru kapitoly (opět postupně po jednotlivých parametrech modelu) popisuji výsledky těchto analýz, které zároveň rovnou diskutuji.

Další velká kapitola hlavní části textu práce se věnuje samotnému navrhovanému (optimalizovanému) volebnímu modelu (*4. Navrhovaný volební model*). Na jejím začátku nejdříve popisuji způsob konstrukce všech modelů, s jejichž výsledky budu výsledky navrhovaného modelu srovnávat (MEDIAN, CVVM2017, CVVM2018 a KANTAR). Následuje detailní popis způsobu konstrukce navrhovaného modelu, který ve výše zmíněných parametrech vychází z výsledků sekundární analýzy dat v této práci, opět dělený podle jednotlivých parametrů. V další části kapitoly vymezuji celková kritéria srovnávání výsledků jednotlivých volebních modelů. Poté srovnávám výsledky všech srovnávaných modelů s výsledky voleb do PSP ČR v roce 2017, a poměřuji je podle stanovených kritérií. Následuje diskuse kombinovaných poznatků ze sekundární analýzy a srovnávání modelů.

Práci ukončuje výčet závěrů, které jsem na základě poznatků z předchozích kapitol učinil jak pro možné využití navrhovaného optimalizovaného volebního modelu agenturou Median, tak pro další předvolební výzkumy a analýzy související s přípravou jejich produktů.

# 2. Volební modely a jejich parametry

V této kapitole se postupně věnuji některým teoretickým (a metodologickým) konceptům důležitým pro následnou sekundární analýzu dat a konstrukci volebního modelu. Cílem této kapitoly není podat vyčerpávající informace o probíhající teoretické debatě všech konceptů zmiňovaných v této diplomové práci, ale vyjasnit relevantní aspekty těch konceptů a konstruktů, které jsou zásadní právě pro dále popsanou sekundární analýzu dat a na jejím základě navržený způsob výpočtu volebního modelu. Toto ukotvení jednotlivých konceptů zkoumaných v analýze a aplikovaných ve volebním modelu by zároveň mělo sloužit k tomu, aby se analytická část této práce nestala pouhým cvičením v exploraci dat a iterativním nastavováním parametrů bez konceptuální a teoretické opory v odborné literatuře.

## 2.1. Volební model

Přestože konstrukci volebního modelu se věnuji až v pozdějších částech této práce, v rámci teoretického kontextu považuji za vhodné pojmem volebního modelu začít, neboť se jedná o klíčový koncept, a konstrukce optimalizovaného volebního modelu je zároveň hlavním cílem této práce k jehož naplnění postupně směřují všechny cíle další.

Volební model je specifický produkt předvolebních výzkumů, který odhaduje výsledek hypotetických voleb v době sběru dat[2]. Označení i vymezení pojmu volebního modelu, jakožto i dalších produktů předvolebních výzkumů, se v českém kontextu objevilo a ustálilo díky dohodě o jednotné terminologii, kterou iniciovala profesní organizace SIMAR. (Lebeda, Krejčí, and Leontiyeva 2004) Ta by měla pro účely této práce posloužit jako základní pojmový rámec pro chápání problematiky volebních modelů. Podstatně lepší porozumění pojmu volební model získáme, pokud ho kromě pozitivního vymezení zároveň porovnáme s dalšími produkty předvolebních výzkumů a vyjasníme tak charakteristiky, které s nimi přes tematickou příbuznost nesdílí – alespoň tedy v intencích výše zmíněné jednotné terminologie.

Volební model není volební prognóza, která si klade za cíl co nejlépe předpovědět výsledek skutečných voleb na základě dat získaných v období před volbami – co do použití statistických metod se zpravidla jedná o mnohem komplexnější produkt, než je volební model. Volební model se sice také snaží odpovídat na otázku výsledků voleb, jedná se ale stále o hypotetické volby v období sběru dat.

Zároveň volební model nejsou ani stranické preference, které se týkají celé populace, a nejde ani o voličské preference, které se sice týkají pouze skupiny (případných) voličů, ale stejně jako ty stranické zkoumají právě preference – tedy rozložení preferovaných politických stran mezi respondenty, a neberou v potaz nerozhodnuté voliče nebo různou míru pravděpodobnosti účasti u voleb apod.[3] (Lebeda, Krejčí, and Leontiyeva 2004) Z metodologické perspektivy je tím pádem zásadní charakteristikou volebního modelu fakt, že nezohledňuje pouze preference jednotlivých respondentů, ale i další proměnné,

[2] Srov. s např. https://cvvm.soc.cas.cz/cz/caste-dotazy/4581-co-jsou-to-stranicke-preference-volicske-preference-a-volebni-model

[3] Druhem produktu, který ve výčtu výše nezmiňuji, ale mohlo by přeci jen dojít k jeho záměně za volební model, jsou stranické sympatie, které zkoumají rozložení sympatií (tedy postojů) respondentů k jednotlivým poltickým stranám. (Lebeda, Krejčí, and Leontiyeva 2004, 8) Jsou dobrým příkladem toho, že přestože se může výzkum nebo jeho výstup týkat politických stran nebo kandidátů, nemusí nijak explicitně zkoumat volební chování, ani jeho případné záměry.

jako jsou záměr volební účasti a u složitějších volebních modelů i více zvažovaných stran daného respondenta apod. Volební modely v ČR v současné době publikuje více agentur, a každá z nich používá pro jejich výpočet vlastní postup. Tyto způsoby už (v obecnější rovině s odvoláním na neveřejný charakter konkrétních postupů konstrukce volebního modelu) porovnala například (Chábová 2013)[4].

## 2.2. Důležité parametry ve volebních modelech

Pokud tedy volební model bere kromě stranických preferencí v úvahu i další proměnné, a způsob jeho konkrétní konstrukce se může lišit, pojďme v dalších odstavcích probrat některé důležité parametry, které do toho výpočtu vstupují (konkrétně ty, jejichž použití se v této práci snažím optimalizovat) a využít následující část této kapitoly k popisu některých jejich teoretických a metodologických aspektů, které by měly sloužit jako základ pro nastavení hypotéz popsaných v kapitole *3. Sekundární analýza dat* a jejich konkrétního využití v navrhovaném modelu (viz *4.2 Konstrukce navrhovaného volebního modelu*)[5].

### 2.2.1. Volební účast

První dvě oblasti optimalizace výpočtu modelu Median se týkají volební účasti. Během voleb musí voliči učinit dvě zásadní rozhodnutí – zda se voleb účastní, a který subjekt v nich budou volit. Motivace volební účasti je nicméně základní vlastností obou těchto rozhodnutí. (Lebeda et al. 2007, 63) Volební účast neovlivňuje výsledky pouze tím, že podmiňuje existenci platných hlasů jako takových. Celková míra volební účasti může ovlivnit i rozložení volebních zisků jednotlivých zúčastněných stran. Strany, které mají mezi svými voliči nízkou míru stranické identifikace, mohou na celkové vysoké volební účasti vydělat, strany, jejichž elektorát je silněji disciplinovaný, ale vyšší mírou celkové volební účasti na výsledcích získat nemusí, a mohou potenciálně i ztratit, neboť většina jejich voličů by se voleb zúčastnila tak či tak. (Linek 2012)

[4]Jedná se o užitečné srovnání, které poskytuje výčet nejdůležitějších vlastností publikovaných volebních modelů. Od dob jeho publikace se ale změnila povaha výběru respondentů v omnibusu společnosti Median, který už není pravděpodobnostní, ale kvótní. (MEDIAN 2018, 3) Přestože přechod od pravděpodobnostního ke kvótnímu výběru respondentů by mohl být označen za sestupný vývoj kvality dat, v případě předvolebních výzkumů jsou např. podle (Lebeda et al. 2007, 43) výhody pravděpodobnostního oproti kvótnímu výběru nepřesvědčivé, a rozdíl v nákladech výzkumu převažuje. Tuto změnu v metodice sběru dat zde zmiňuji proto, že se jedná o jeden ze dvou hlavních datových zdrojů této práce. Tato práce zároveň oproti srovnání v (Chábová 2013) popisuje i konkrétní způsoby konstrukce jednotlivých srovnávaných volebních modelů ve větší míře podrobnosti.

[5] V této kapitole se nebudu věnovat konkrétním způsobům výpočtů – ty podrobně rozebírám v kapitolách *Srovnávané výstupy* a *Konstrukce navrhovaného volebního modelu.*

V České republice je díky existenci stálých seznamů voličů vedených podle volebního zákona velmi dobrý přehled o volební účasti. Proto zde není potřeba provozovat povolební šetření, která se snaží volební účast stanovit, jako například v USA, kde podobné seznamy voličů nefungují. V těchto povolebních šetřeních se nicméně ukázalo, že respondenti často nepravdivě reportují volební účast, přestože se voleb nezúčastnili. Přestože k misreportingu účasti může docházet oběma směry (nepravdivá účast i nepravdivá neúčast), odborná literatura má celkově spíše za to, že po volbách respondenti misreportují téměř vždy pouze nepravou účast. Specifickým případem jsou volby, které jsou menšího významu, nebo spojené dohromady s jinými (významnějšími) volbami v ten samý den, které pak respondenti nejsou schopni v paměti odlišit, a reportují tím pádem nepravdivou neúčast. (Adamany and Dubois 1975; Adamany and Shelley 1980)

Jedním z faktorů, které tento druh misreportingu ovlivňují, je sociální desirabilita, tedy reportování chování, o kterém jsou respondenti přesvědčení, že je v příslušné sociální skupině očekáváno/preferováno. (Ansolabehere and Hersh 2012) Tento druh misreportingu je logicky silněji zastoupený ve skupinách, které cítí vyšší sociální tlak na svou volební účast. Které skupiny to jsou, nicméně silně záleží na lokálním kontextu, může se jednat o národnostní menšiny (Bernstein, Chadha, and Montjoy 2001), respondenty s vyšším stupněm vzdělání (Silver, Anderson, and Abramson 1986) nebo zájmem o politiku[6]. (Sciarini and Goldberg 2016)

Přiznat se k neúčasti u voleb může být pro některé respondenty náročnou a citlivou situací, proto byla také vyšší míra tohoto misreportingu zaznamenána během dotazování tváří v tvář s tazatelem (CAPI) oproti metodám, kde tazatel není fyzicky přítomen a respondent má větší míru soukromí. (Stocké 2007) Jedním z projevů výše popsaného jevu je nadhodnocování míry účasti v povolebních studiích. To je ale částečně způsobené i tím, že voliči jsou obecně nadreprezentováni ve výzkumech – lidé, kteří se účastní voleb, mají menší pravděpodobnost odmítnutí participace ve výzkumu. (Sciarini and Goldberg 2015; Durand, Blais, and Vachon 2002)

Misreporting volební účasti zmiňuji v kontextu této práce proto, že v sekundární analýze dat CHPS dále používám reportovaný ukazatel volební účasti k tomu, abych posoudil vliv

[6] Tento druh misreportingu můžeme vysvětlovat i kognitivní disonancí, kdy respondenti nechtějí narušovat svou vlastní představu o svém volebním chování a politických postojích, viz *2.2.3. Minulá volba a vliv uplynulého času.*

záměru volební účasti a reportované účasti u minulých voleb na účast u reálných voleb. Je velmi nepravděpodobné, že by se žádný z výše popsaných jevů neprojevil i na proměnné reportované účasti u voleb v datech CHPS. Tu nicméně nepoužívám v kontextu této práce k tomu, abych odhadoval míru volební účasti v populaci, ale proto, abych analyzoval některé faktory, které tuto účast ovlivňují – pro takový druh analýzy reportovaná data o volební účasti postačují[7]. (Sigelman 1982)

#### *2.2.1.1. Záměr volební účasti*

Volební účast je ve volebním modelu zastoupena dvěma proměnnými. První z nich je záměr volební účasti, tedy uzavřená otázka, kde respondent (typicky ve 4 kategoriích)[8] indikuje, jestli by se zúčastnil případných voleb konaných v období dotazování. Taková otázka je důležitou částí prakticky všech volebních modelů publikovaných v současné době v ČR[9], nejčastěji v takové podobě, že pokud respondent v této otázce vyloučí účast, dotazování jeho preferencí už nepokračuje[10].

Záměr volební účasti se tedy ve volebních modelech používá k rozdělení respondentů na pravděpodobné voliče a pravděpodobné nevoliče, a jedná se o dobrý indikátor volební účasti (Bolstein 1991), i když zároveň musíme mít na paměti, že se jedná pouze o indikátor, nikoliv o ten samý koncept. Proto se v prvních hypotézách sekundární analýzy dat zaměřuji v této práci právě na vztah mezi záměrem volební účasti a reportované volební účasti po volbách – abych ho mohl následně v konstrukci volebního modelu zohlednit tak, aby díky špatné klasifikaci respondentů na voliče a nevoliče nedocházelo k nepřiměřeným pokřivením výsledků modelu. (srov. s Erikson, Panagopoulos, and Wlezien 2004)

#### *2.2.1.2. Minulá volební účast*

Na účast u reálných voleb má nicméně kromě záměru volební účasti a osobních charakteristik respondenta vliv také jeho chování v minulosti. (Granberg and Holmberg 1990) Jedním takovým chováním je účast u minulých voleb. Účast u voleb do určité míry

[7] Vyšší míru response voličů oproti nevoličům se pak při konstrukci modelu snažím kompenzovat imputací dat z follow-upu třetí vlny CHPS pomocí logistické regrese, viz *0*
*Doplnění chybějících hodnot účasti 2017 pomocí logistické* regrese.

[8] Kromě uzavřené otázky na záměr volební účasti je v zahraničí používána například pravděpodobnostní (otevřená) otázka, kde respondent odpovídá číslicí reprezentující procenta pravděpodobnosti jeho volební účasti. (Delavande and Manski 2010)

[9] Viz např. *4.1 Srovnávané výstupy*.

[10] Výjimkou je například agentura STEM, která sice takové respondenty vyřazuje z volebního modelu, nicméně jejich stranické preference dále dotazuje, viz např. (STEM 2019)

vytváří u voličů volební návyk, tzn. podstatně zvyšuje pravděpodobnost jejich účasti u následujících voleb. (Cutts, Fieldhouse, and John 2009) Tento efekt rozhodně není všeprostupující – účast je nadále ovlivněna osobními charakteristikami (ne)voličů, rozhodováním v daném politickém kontextu apod., efekt minulé volební účasti ale platí dále. U specifického typu tzv. habituálních voličů může síla tohoto návyku dospět tak daleko, že pro jejich volební účast už nejsou rozhodující osobní postoje, záměry, nebo rozhodnutí, ale právě návyk volební účasti. (Aldrich, Montgomery, and Wood 2011)

(Denny and Doyle 2009) ve své práci stanovili, že účast u voleb zvyšuje šanci účasti u následujících voleb od 13 %. Přestože by bylo lákavé takové konkrétní číslo rovnou přejmout a aplikovat, ověřuji raději vliv minulé účasti na reportovanou účast u voleb do PSP ČR v roce 2017 vzhledem k možné lokální nebo situační specificitě v sekundární analýze dat, a to jak samostatně, tak v kombinaci se záměrem volební účasti, viz *3.2.1 Volební účast, záměr volební účasti a minulá volební účast (CHPS)* – výsledný parametr ve volebním modelu pak nastavuji podle výsledků této analýzy.

### 2.2.2. Záměr volby preferované strany

Druhou zkoumanou oblastí konstrukce volebního modelu v této diplomové práci je jistota volby zvažované strany. Vzhledem k tomu, že politická strana je de facto jednotkou analýzy prezentovaných výstupů volebních modelů, jedná se společně s volební účastí o další zásadní parametr volebních modelů. Proměnná preferované strany je v předvolebních výzkumech většinou dotazována otázkou, která navazuje na otázku záměru volební účasti, a respondent si v ní vybírá stranu, kterou by v případě účasti u příslušných (hypotetických) voleb pravděpodobně volil.

Přestože se taková otázka nabízí jako nejlogičtější indikátor voličské preference (respektive si osobně nejsem schopen představit logičtější indikátor), stále v sobě její použití ve volebním modelu skrývá některé obtíže. Jak vyplývá z České volební studie, přibližně třetina všech voličů se o volbě konkrétní strany rozhoduje až v posledním měsíci před volbami, necelá pětina všech voličů pak v několika dnech před volbami. (Linek 2018) Bez ohledu na období, kdy data pro volební model sbíráme, se budeme tedy vždy potýkat s určitou částí respondentů, jejichž volební preference není krystalizována.

Nekonzistence volené strany a deklarované strany může být vyšší u voličů s nižší mírou stranické identifikace (Linek 2009, 204), i u voličů, kteří se rozhodují v souladu s valenční

teorií volebního chování, kdy si stranu vybírají podle toho, jak velkou vnímají podobnost svých postojů ke konkrétním problémům s postoji dané strany, a nakolik zároveň vnímají danou stranu jako kompetentní tyto problémy řešit (Linek 2010, 11–12). Pokud si takovou stranu nebude moci respondent z kandidujících subjektů vybrat, anebo mu budou připadat všechny v tomto ohledu srovnatelné, nebude mít motivaci se voleb vůbec účastnit – svou volební účast nebude spatřovat jako smysluplnou a efektivní. (Linek 2011)

Z výše uvedeného vyplývá, že deklarovanou volební preferenci nelze samu o sobě brát jako plně stabilní indikátor ani pro volební chování v reálných volbách, ale vzhledem k nízké míře krystalizace ani vzhledem ke konstrukci volebního modelu. Proto se v druhé části kapitoly sekundární analýzy dat snažím nastavit obecnou míru shody preferované strany a volené strany u respondentů v období před volbami (viz *3.2.2 Preferovaná strana (CHPS)*), kterou následně používám jako vodítko pro nastavení parametru jistoty volby preferované strany, tedy nástroje, kterým se volební model agentury Median a volební model navrhovaný v této práci snaží vypořádat s výše zmíněnými obtížemi[11].

### 2.2.3. Minulá volba a vliv uplynulého času

Jedním z nástrojů, které některé české výzkumné agentury používají při tvorbě svých modelů, je vážení dat (kromě jiných charakteristik i podle minulého volebního chování respondenta, tedy v případě voleb do PSP ČR podle toho, jestli se minulých voleb účastnil, a kterou stranu volil. (MEDIAN 2018; TNS Kantar 2019) Dělají to proto, aby dosáhly reprezentativity i v této charakteristice – respektive, pokud se na reprezentativitu díváme jako na odpovídající rozložení charakteristik v souboru a v populaci (Krejčí 2008), chtějí, aby minulé volební chování bylo jednou z těchto podobných charakteristik.

Důvodem, proč se v rámci optimalizace volebního modelu agentury Median věnuji i vážení podle minulého volebního chování je fakt, že odchylky v rozložení proměnné minulého volebního chování (oproti populaci, což v tomto kontextu znamená oproti výsledkům příslušných voleb) nejsou způsobeny pouze metodologií sběru dat, tedy pouze např. chybou výběru, nebo náhodnými chybami, ale také tím, že respondenti v proměnných

[11] Agentura Median ve svých předvolebních výzkumech umožňuje respondentům jmenovat více zvažovaných stran (stejně jako např. CVVM). Přestože nastavení parametru jistoty záměru volby preferované strany nepřímo nastavuje (díky omezení hodnot váhy všech zvažovaných stran každého respondenta na součet rovný právě 1) i váhu ostatních zvažovaných stran v modelu (viz *4.1.1 Model MEDIAN*), věnuji se v této kapitole a v sekundární analýze dat pouze první preferované straně. Důvodem je datový zdroj Czech Household Panel Study, kde jsou respondenti dotazováni pouze na jednu preferovanou stranu, čímž je stanoven i rozsah možné sekundární analýzy těchto dat na toto téma.

minulého volebního chování reportují volené subjekty do určitě míry nepravdivě. (Waldahl and Aardal 1982) U špatně vybaveného (*recall*) minulého volebního chování se možné důvody většinou dělí na tři skupiny, které jsou částečně shodné s důvody misreportingu volební účasti výše. Prvním důvodem jsou problémy s pamětí, kdy si respondent díky času uplynulému od vybavované události jednoduše není schopen na své volební chování správně vzpomenout, a nesprávně se pak vztáhne k události, k níž nedošlo, druhým je sociální desirabilita, kdy voliči reportují svou volbu nepravdivě, aby vyhověli sociálnímu tlaku na určitý typ jednání. (Belli et al. 1999, 90–91) Třetím podnětným důvodem pro nepravdivě reportované minulé volební chování je jev kognitivní disonance, kdy respondenti nepravdivě reportují své volební chování, aby vyhověli představě sebe sama jako politicky konzistentního jedince, respektive tuto představu nenarušovali. (Waldahl and Aardal 1982)

Tyto vlivy nejsou tak závažné, aby kvůli nim dotazování respondentů na minulé volební chování pozbývalo smyslu. (Dassonneville and Hooghe 2016) Pokud proměnnou minulé volby ale používáme k vážení dat, bylo by vhodné s výše popisovanými chybami nenakládat jako s chybou výběru (a vyrovnávat je prostřednictvím vážení), ale jako potenciálním projevem příslušné společenské situace. Proto se v kapitole *3.2.3* sekundární analýzy snažím odlišit, které odchylky minulého volebního chování od výsledku voleb jsou projevem výběru respondentů společnosti Median (a měly by být narovnány vážením), a které byly způsobené spíše vlivem času uplynulého od voleb (a měly by být v datech zachovány). Na základě poznatků z této analýzy následně upravuji váhovou proměnnou minulého volebního chování pro navrhovaný volební model.

## 2.3. Proč posuzuji volební model s reálnými výsledky voleb?

V závěru této kapitoly bych se rád věnoval ještě otázce způsobu posuzování navrhovaného volebního modelu, která je, vzhledem k nastavení hlavního cíle, důležitá pro legitimitu nastavení celé práce coby návrhu metodologické optimalizace. Pokud jsme na základě výše zmíněné jednotné terminologie stanovili, že volební model není volební prognóza, je potřeba vysvětlit, proč v kapitole *4.4 Posouzení navrhovaného volebního modelu (výsledky)* srovnávám navrhovaný volební model (a další způsoby výpočtu srovnávaných volebních modelů) s výsledky reálných voleb – nemá být přeci jejich předpovědí.

Nejedná se jen o terminologický spor vymezení pojmu volební model, ale také o součást již existující debaty. Zda za předpokladu, že víme, že média a veřejnost přistupují

k výstupům předvolebních výzkumů jako k predikcím výsledků voleb (respektive je k finálním výsledkům reálných voleb vztahují), měli bychom se tomuto snažit přizpůsobit a posuzovat volební modely i v jejich vztahu k výsledkům voleb, nebo na tento úhel pohledu nepřistupovat a spíše se proti němu vymezit, jako to dělá např. CVVM[12]? Výsledky volebních modelů či volebních preferencí na druhou stranu srovnává kromě původního návrhu volebního modelu společnosti Median v (Prokop 2012) například také (Lebeda 2003).

Přestože si uvědomuji, že srovnáváním výsledků volebních modelů s výsledky voleb se v důsledku nevyhnutelně stavím na první zmíněnou stranu sporu (přizpůsobování pohledu veřejnosti), myslím, že se mi výzkumným nastavením této práce daří tuto teoretickou debatu do značné míry obejít. Jak je vyjasněno v kapitole *3.1 Data*, datový soubor, z něhož vypočítávám srovnávané volební modely, byl společností Median sbírán bezprostředně před příslušnými volbami.[13] Stejně tak v sekundární analýze dat z CHPS omezuji datové soubory na respondenty dotazované v posledních dvou měsících před volbami, abych redukoval velikost časového okna mezi sběrem dat a datem reálných voleb.

Pokud má tedy volební model vypovídat o současném stavu v období sběru dat,[14] a zároveň je toto období situací blízko samotným volbám, je výsledek příslušných voleb podle mě vhodným (resp. nejvhodnějším dostupným) zdrojem informací o přesnosti volebního modelu reprezentujícího tuto společenskou situaci. Toto omezení dat v kombinaci se snahou o zohlednění různé míry nejistoty v rozhodování voličů je pak způsobem, kterým se snažím vypořádat s riziky plynoucími mj. z tohoto teoreticko-metodologického sporu, zejména redukcí času, kdy může docházet k zásadním změnám podstatně měnícím předvolební situaci, jako tomu bylo např. v (Benoit 1998).

# 3. Sekundární analýza dat

## 3.1. Data

Vzhledem k tomu, že hlavním výzkumným designem této diplomové práce je sekundární analýza dat, je výběr datových souborů, které jsou v rámci této práce analyzovány,

[12] Viz např. https://cvvm.soc.cas.cz/cz/caste-dotazy/4581-co-jsou-to-stranicke-preference-volicske-preference-a-volebni-model nebo https://cvvm.soc.cas.cz/cz/caste-dotazy/4580-predpovida-volebni-model-vysledky-voleb

[13] Výsledky volebního modelu z tohoto datového souboru nebyly publikovány z důvodu informačního embarga před volbami.

[14] Viz např. https://cvvm.soc.cas.cz/cz/caste-dotazy/4580-predpovida-volebni-model-vysledky-voleb

zásadním krokem. Pokud by práce stála na primárním výzkumu, mohli bychom udělat poměrně velké množství rozhodnutí, kterými bychom přispěli lepší kvalitě dat, a takové konceptualizaci a operacionalizaci jednotlivých pojmů a proměnných, která by přímo odpovídala našim cílům a zvoleným metodám. V případě sekundární analýzy dat je tento vztah v podstatě obrácený.

Všechna výše zmíněná rozhodnutí už za nás učinil někdo jiný a my je v rámci sekundární analýzy společně se zvolenými datovými soubory přijímáme bez možnosti je nějak ovlivnit, a s předem omezeným rozsahem informací, které z dat můžeme vytěžit. Proto musíme být jednak schopni ospravedlnit, proč jsme dané datové soubory zvolili, a které metody analýzy jsou vzhledem k našim cílům relevantní, a vzhledem k charakteru dat vhodné, na tyto soubory použít. Zároveň bychom neměli zapomínat na otázku, jestli tato sekundární analýza přináší podstatná zjištění, a není pouhou reakcí na existenci dostupných dat. (Murphy and Schlaerth 2010, 381) Tomuto se, doufám, daří v práci předcházet aplikací výsledků sekundární analýzy na konkrétní v praxi využívaný volební model. Sekundární analýza dat tak není samoúčelná – její výsledky se v diplomové práci snažím bezprostředně aplikovat. Proto se tedy ve zbytku této kapitoly postupně věnuji: 1) použitým datovým zdrojům, 2) použitým metodám analýzy a 3) diskusi jejích výsledků. Analyzovanými vztahy jsou: 1) vztah volební účasti, záměru volební účasti a minulé volební účasti, 2) vztah volební preference a volené strany a 3) vliv času uplynulého od voleb na proměnnou minulé volby.

### 3.1.1. Czech Household Panel Study

V rámci sekundární analýzy používám pro získávání poznatků dva zdroje dat. Prvním z nich je datový soubor z šetření CHPS, jehož analýza bude sloužit k dosažení prvního a druhého podcíle[15]. Czech Household Panel Study (CHPS) je české panelové šetření domácností, které v České republice v jednotlivých vlnách probíhá už od roku 2015. Cílovou populací tohoto šetření je neinstitucionalizovaná populace České Republiky. Domácnosti, které se šetření účastní, jsou vybírány dvoustupňovým stratifikovaným náhodným výběrem (Kudrnáčová 2019, Verze 4.0:5), jehož základní jednotkou je hospodařící domácnost (přestože základní jednotkou analýzy v této diplomové práci je individuální respondent – viz dále).

[15] K pro tuto práci nezbytným datům z Czech Household Panel Study mi zprostředkoval přístup Sociologický ústav AV ČR, za což děkuji.

Data českého panelového šetření domácností se získávají mixed-mode sběrem. Část dat je sbírána metodou CAPI, tedy vyplněním dotazníku během rozhovoru s tazatelem za asistence počítače nebo tabletu, část dat je ale sbírána pomocí papírového dotazníku, který respondent sám vyplňuje (SAQ). Do této metody spadají nejen klasické dotazníky, ale i časové snímky (Kudrnáčová 2019, Verze 4.0:24), v nichž respondent zaznamenává své denní aktivity, nebo samostatným vyplňováním dotazníků respondenty v prostředí internetového prohlížeče (CAWI). Sběr dat CHPS byl realizován společně dvěma výzkumnými agenturami – Median a STEM/MARK.

Větší počet používaných metod sběru dat určitě vyplývá zejména z množství a rozmanitosti témat, na něž byli členové zúčastněných domácností dotazováni. Zároveň se ale mixed-mode sběry často používají pro zlepšení návratnosti, jak zmiňuje (Millar and Dillman 2011) nebo (Dillman et al. 2009). Tomu u CHPS nasvědčuje i fakt, že jsou v některých případech ty samé otázky sbírány několika různými metodami. Já se nicméně při sekundární analýze dat CHPS zabývám politickými postoji a volebním chováním, které byly dotazovány v rámci tzv. „individuálních dotazníků“, a jednotkou analýzy v nich tak zůstává individuální respondent, přestože samotné šetření je zaměřené (i v jednotce výběru) na domácnosti.

V analýze mě budou zajímat otázky, které se dotýkají voleb do PSP ČR, které proběhly v říjnu 2017. Proto budu zkoumat pouze data ze třetí vlny CHPS, která probíhala od 21. 6. do 30.10.2017. (Kudrnáčová 2019, Verze 4.0:5) Data všech analyzovaných proměnných ze třetí vlny byla sebrána buďto metodou CAPI, nebo metodou CAWI (N=4635). Specifickou částí analyzovaných dat, která mi vůbec umožňuje potřebné analýzy provádět, je follow-up třetí vlny, který proběhl po volbách do PSP ČR v říjnu a listopadu 2017, a doptává se na některé otázky ohledně volebního chování v čase těchto voleb. Data analyzovaných proměnných z follow-upu byla získána většinou metodou CAWI, a v některých případech telefonickým rozhovorem s respondentem – CATI (N=3693). Pro účely sekundární analýzy dat používám datové soubory ze třetí vlny a z follow-upu sloučené. Pro analýzu proměnných záměru volební účasti používám data pouze těch respondentů, kteří byli ve třetí vlně dotazováni v období od 14. 8. 2017 (33. týden) do 19. 10. 2017 (poslední den pře volbami do PSP ČR), abych omezil množství času, které měli respondenti mezi třetí vlnou a follow-upem na změnu názorů. Konkrétní proměnné popisuji v příslušných částech sekce *3.2 Metody analýzy*.

### 3.1.2. Omnibus Median

Druhým zdrojem dat pro sekundární analýzu používaným v této diplomové práci je omnibusové šetření agentury Median. Většina jeho dat (85–95 %) je sbírána metodou CAPI na základě kvótního výběru, kdy tazatelé sami vybírají respondenty podle jim přidělených charakteristik. Datový soubor považuji po účely této práce za vhodný – přestože kvótní výběr je metodicky méně sofistikovaný, organizačně méně náročný a nákladný než např. pravděpodobnostní výběr u CHPS, neznamená automaticky méně kvalitní data. (Cumming 1990) Skupiny respondentů, které se nedaří tímto výběrem zastihnout, jsou dotázány metodou CAWI přes online panel společnosti Median[16]. (MEDIAN 2018, 3) Jak vyplývá z označení samotného šetření, jedná se o sběr dat, který se pravidelně opakuje (resp. probíhá průběžně v oddělených vlnách), a jeho výzkumný nástroj – dotazník – má více oddělených tematických částí. Jednou z nich je i předmět druhé části sekundární analýzy dat v této diplomové práci, sekce otázek věnovaná politickým postojům a volebnímu chování.

Konkrétně zkoumám data ze dvou vln Omnibusu Median, „povolební" vlny sebrané v listopadu a prosinci 2013, tedy bezprostředně po volbách do PSP ČR v roce 2013, a druhé, a těsně „předvolební" vlny sebrané v září a říjnu 2017[17]. Tyto vlny jsou vybrány záměrně tak, aby u první vlny byla během sběru co nejmenší časová vzdálenost od voleb, a u druhého naopak co největší časová vzdálenost od voleb, o kterých otázky na volební chování v minulosti referují. Proto by se mělo v rámci Omnibusu Median jednat o ta nejvhodnější data pro zkoumání vlivu času na míru zapomínání (viz dále).

Kvalita dat je zajištěna zázemím a metodikou profesionální výzkumné agentury[18], a oba datové soubory Medianu jsou bezprostředně blízké období voleb a umožňují tak dělat některé závěry z analýzy, které bych např. z analýzy o měsíc dříve sebraných dat mohl dělat jen s podstatně většími obtížemi. Navíc k nim existují předchozí (a následující

[16] K datům z omnibusu Median se mi nepodařilo najít zpracovanou dokumentaci, a proto při citování metodiky, znění otázek apod. z těchto dat používám jako hlavní zdroj tiskovou zprávu k volebnímu modelu Median z října 2018. Fakt, že chybí podrobná dokumentace, je podle mého názoru způsoben zejména tím, že se nejedná o akademický výzkum (jako například u CHPS), ale o výzkum realizovaný komerční výzkumnou agenturou. Přesto jsou ale příslušné výstupy o volebním chování publikovány veřejně, a hlavní parametry výzkumného nastavení a sběru jsme schopni sledovat z příslušných tiskových zpráv a samotných dat. Konkrétní proměnné opět popíšu v příslušných částech sekce Metody analýzy této kapitoly.

[17] Dokumentaci ani výstupy těchto dvou vln bohužel nelze přímo citovat, protože z nich nebyl (vzhledem k velké blízkosti k volbám), veřejně publikován výstup, přestože byla v rámci omnibusu sebrána.

[18] K datům mi přístup poskytla společnost Median, bez čehož by, stejně jako v případě CHPS, tato diplomová práce nemohla v této podobě vzniknout. Pro mě jako studenta magisterského oboru se jedná o skvělý datový soubor k analýze, který bych svépomocí jen těžko mohl nahradit.

i budoucí) vlny sběru, které probíhaly za použití shodné metodiky, a pokud se tím pádem v této diplomové práci podaří učinit nějaký závěr, nemělo by být složité ho vyzkoušet aplikovat jak na předchozích, tak budoucích vlnách omnibusu.

## 3.2. Metody analýzy

V této části diplomové práce postupně popíšu postupy, kterými jsou v rámci sekundární analýzy dat splněny jednotlivé podcíle práce. Následně také podrobněji rozeberu způsob, kterým je zkonstruován navrhovaný volební model, který je (jako „návrh metodologické optimalizace“) hlavním výstupem této práce.

Níže popsané hypotézy (a stejně tak hypotézy v dalších podcílích této kapitoly) jsou věcně spíše triviální a nijak zvlášť překvapivé. Smyslem jejich ověřování v této práci nicméně není ověření samotné, ale ověřením vztahu analyzovaných proměnných jednak ospravedlnit jejich použití v navrhovaném volebním modelu, a jednak na základě výsledků analýzu nastavit u jednotlivých proměnných konkrétní hodnoty na nich založených parametrů modelu. V každém podcíli sekundární analýzy dat se nejdříve vhodným statistickým testem snažím ověřit, že je v datech přítomen v hypotézách formulovaný vztah mezi proměnnými, a následnou deskriptivní analýzou těchto proměnných získat potřebné údaje pro nastavení parametrů popsané v kapitole *4.2 Konstrukce navrhovaného volebního modelu.* Přikládám také tabulku, která pro přehlednost shrnuje, které datové soubory jsou používány pro jakou část sekundární analýzy dat. (*Tabulka 1*)

| Oblast analýzy | Datové soubory |
|---|---|
| Volební účast (…) | 3. vlna CHPS (2017)<br>+follow-up |
| První zvažovaná strana | 3. vlna CHPS (2017)<br>+follow-up |
| Minulé volební chování | Omnibus Median:<br>vlna XI/XII 2013<br>vlna IX/X 2017 |

*Tabulka 1: Oblasti analýzy a použité datové soubory*

### 3.2.1. Volební účast, záměr volební účasti a minulá volební účast (CHPS)

V prvním podcíli sekundární analýzy dat si kladu za úkol zjistit, jaký je v datech CHPS vztah mezi proměnnou deklarovaného záměru účastnit se hypotetických voleb týden od

dotazování v době před volbami[19], účastí u voleb do PSP ČR v roce 2013[20] a reportovanou volební účastí v době po volbách do PSP v roce 2017.[21] Závislou proměnnou je zde reportovaná účast u voleb u voleb do PSP ČR v roce 2017 z follow-upu CHPS. Jedná se o dichotomickou proměnnou (účast/neúčast).[22] Hypotézy pro tento podcíl sekundární analýzy dat jsou následující:

H1a: Respondenti, kteří deklarovali záměr účastnit se hypotetických voleb do PSP ČR konaných týden od dotazování, ve follow-upu potom reportovali účast u voleb do PSP ČR v roce 2017 častěji než respondenti, kteří záměr se zúčastnit nedeklarovali.

H1b: Respondenti, kteří reportovali účast u voleb do PSP ČR v roce 2013, ve follow-upu potom reportovali účast u voleb do PSP ČR v roce 2017 častěji než respondenti, kteří reportovali neúčast u voleb do PSP ČR v roce 2013.

H1c: Respondenti, kteří deklarovali záměr účastnit se hypotetických voleb do PSP ČR konaných týden od dotazování, **a zároveň** reportovali účast u voleb do PSP ČR v roce 2013, potom ve follow-upu reportovali účast u voleb do PSP ČR v roce 2017 častěji než respondenti, kteří sice deklarovali záměr účastnit se hypotetických voleb konaných týden od dotazování, ale **zároveň** reportovali neúčast u voleb do PSP ČR v roce 2013.

Zatímco hypotézy H1a a H1b si kladou za cíl ospravedlnit a nastavit použití deklarovaného záměru účasti u hypotetických voleb a minulé volební účasti pro vytvoření parametru modelu, hypotéza H1c ospravedlňuje nastavení tohoto parametru na základě kombinace jejich hodnot u každého respondenta (nikoliv jako dvou proměnných odděleně, viz kapitola *4.2 Konstrukce navrhovaného volebního modelu*). U hypotézy H1a statistiku vypočítávám v softwaru SPSS jako součást jednoduché kontingenční tabulky, u hypotéz H1b a H1c potom jako součást trojité kontingenční tabulky (three-way crosstab), kde deklarovaný záměr volební účasti slouží jako kontrolní proměnná.

---

[19] Záměr volební účasti u případných voleb je v individuálním dotazníku CHPS dotazován otázkou: *„Představte si, že by příští týden byly volby do Poslanecké sněmovny. Šel/Šla byste volit?“* Rozhodně ano / Spíše ano / Spíše ne / Nemá volební právo / Nevím, nemohu posoudit / Odmítl(a) (CHPS 2018, 69)

[20] Minulá účast u voleb do PSP ČR v roce 2013 je dotazována otázkou *„Ve dnech 25.-26. října 2013 se konaly volby do Poslanecké sněmovny. Když hovoříme s lidmi o těchto volbách, zjišťujeme, že se jich řada lidí nezúčastnila. A co Vy, zúčastnil(a) jste se těchto voleb?“* Ano / Ne, neměl(a) jsem volební právo / Ne, z jiného důvodu / Nevím, nemohu posoudit / Odmítl(a) (CHPS 2018, 68)

[21] Ta je ve follow-up dotazníku zastoupena otázkou *„Ve dnech 20. a 21. 10. 2017 se konaly volby do Poslanecké sněmovny. Zúčastnil(a) jste se těchto voleb?“* Ano / Ne / Nevím, nechci uvést. Otázku cituji z neveřejné verze dotazníku follow-upu třetí vlny.

[22] Respondenti s odpovědí „Nevím, nechci uvést“ nebyly v případě této otázky v datech přítomni.

Pro ověření výše popsaných hypotéz používám Pearsonův Chi-kvadrát test nezávislosti, neparametrickou statistiku vhodnou pro testování hypotéz v kategorických datech. Tento test (jako všechny ostatní procedury popsané v této diplomové práci) vypočítám v softwaru SPSS. Jeho důležitými předpoklady jsou vzájemná nezávislost jednotlivých pozorování[23], vzájemná výlučnost jednotlivých kategorií v tabulce a očekávaná četnost větší nebo rovná 5 alespoň v 80 % všech buněk tabulky. (McHugh 2013, 143–44) Aby byl zachován předpoklad dostatečně velkých očekávaných četností ve všech buňkách analyzované kontingenční tabulky, vstupují do mé analýzy pouze ti respondenti, kteří ani v proměnné záměru volební účasti ani v proměnné minulé volební účasti **neuvedli**, že neví, nebo odmítají odpovědět.[24][25] Nulovou hypotézou testu nezávislosti pak je, že srovnávané proměnné jsou na sobě nezávislé. (Agresti and Wackerly 1977, 111) Zvolenou hladinou významnosti pro tento test (a všechny ostatní statistické testy v této diplomové práci) je standardně užívaná hodnota 0,05. (Soukup 2010, 80) Jako statistiku pro posouzení těsnosti tohoto vztahu použiju statistiku Cramerovo V. (Cohen 1988, 222)

Tento postup by za jiných okolností pro získání kýžených hodnot pro zohlednění záměru účasti u voleb při konstrukci navrhovaného volebního modelu stačil. Jak už ale může být evidentní z rozdílu počtů respondentů ve třetí vlně a ve follow-upu zmíněných v sekci Data této kapitoly, ve follow-upu chybí oproti třetí vlně podstatná část odpovědí respondentů.[26]

---

[23] Chi-kvadrát test se tak nehodí pro závislé výběry, panelová data apod. Já v tomto případě sice srovnávám proměnné z 3.vlny a follow-upu CHPS, nicméně se nejedná o vícenásobné měření shodných proměnných (protože minulá volební účast, deklarovaný záměr účasti i reportovaná účast jsou různé proměnné), takže předpoklad nezávislosti pozorování zůstává zachován.

[24] U proměnné minulé účasti jsem toto rozhodnutí učinil proto, že očekávané četnosti respondentů, kteří nevěděli nebo odmítli odpovědět (celkem N=56 respondentů s platnou odpovědí reportované volební účasti v roce 2017), byly v některých kategoriích příliš malé, a tak zneplatňovaly celou proceduru Chi-kvadrát testu nezávislosti. Abych se ujistil, že si mohu dovolit tyto respondenty z analýzy vyřadit, překódoval jsem proměnnou minulé účasti na dichotomickou (respondenty s platnými odpověďmi a respondenty, kteří nevědí nebo odmítli), a tu jsem následně porovnal v kontingenční tabulce s proměnnou reportované účasti u voleb do PSP ČR v roce 2017. Pearsonův test nezávislosti indikoval, že mezi těmito skupinami respondentů vzhledem k jejich reportované účasti není statisticky signifikantní rozdíl (p=0,494), a proto považuji za oprávněné je zanedbat. Toto rozhodnutí následně zohledňuji i při konstrukci volebního modelu, viz kapitole *4.2 Konstrukce navrhovaného volebního modelu*. U proměnné záměru volební účasti zanedbávám respondenty, kteří nevěděli nebo odmítli odpovědět (N=18 respondentů s platnou odpovědí reportované volební účasti v roce 2017) proto, že tyto skupiny v datovém souboru omnibus Median (kvůli jinému výčtu možností odpovědí v dotazníku) neexistují.

[25] I po výše popsaném omezení do analýzy vstupujících případů měla jedna analyzovaná buňka v kategorii záměru „Rozhodně ano“ očekávanou četnost menší než 5 (ale stále větší než 1). Omezením dat ale zároveň došlo ke zmenšení rozměrů tabulky na 2x2, při které je Chi-kvadrát test robustnější a není potřeba tento předpoklad striktně dodržovat. (Camilli and Hopkins 1978)

[26] Kdyby byly tyto chybějící hodnoty vzhledem k analyzovaným proměnným rovnoměrně rozložené, mohli bychom je zanedbat, nicméně relativní četnost chybějících hodnot ve follow-upu je významně vyšší

Vztah záměru účasti a reportované účasti už přímo na tomto datovém souboru zkoumal (Linek 2018), nicméně s předpokladem, že nižší návratnost follow-upu ve follow-upu nemá na tento vztah vliv. Já se tento možný problém s nadreprezentací účastníků voleb ve follow-upu třetí vlny CHPS snažím řešit následujícím způsobem.

Po ověření hypotéz výše (ty proběhnou na původním datovém souboru), tedy ještě před nastavením hodnot příslušného parametru volebního modelu, doplním imputací chybějících hodnot na základě logistické regrese. Protože se nejedná vyloženě o součást způsobu konstrukce navrhovaného modelu, ale spíše přípravy potřebných dat, zmiňuji pro úplnost i tento proces včetně jeho výsledků raději v této kapitole v následujících několika odstavcích.

#### *3.2.1.1. Doplnění chybějících hodnot účasti 2017 pomocí logistické regrese*

Že se ve follow-up sběru nepodaří dosáhnout 100% response, není překvapivé, retence je v panelových šetřeních ovlivněna řadou faktorů a jedná se o důležité téma jakéhokoliv dlouhodobě probíhajícího šetření. (srov. s Radler and Ryff 2010) Zároveň také volební chování spadá pod citlivé údaje (ESOMAR 2011, 14), a proto na něj respondenti mohou odmítnout odpovědět. Vzhledem k tomu, že nechci kvůli chybějícím hodnotám zavrhnout celou zamýšlenou analýzu, nabízí se mi několik způsobů řešení. Zanedbat problém, nebo se ho pokusit řešit vážením dat, jsem se rozhodl neudělat. Namísto toho využívám informací o těch respondentech, které znám, k tomu, abych na jejich základě odhadl informace o respondentech, které mi chybí, konkrétně odhadnutím těchto chybějících hodnot pomocí logistické regrese, která mimo jiné umožňuje u dichotomické závislé proměnné na základě hodnot nezávislých proměnných odhadovat (modelovat) pravděpodobnosti hodnot této závislé proměnné. Logistická regrese sice neumožňuje všechna využití, která bychom měli k dispozici např. u lineární regrese pro spojité proměnné (Mood 2010), k vypočítání pravděpodobností v rámci této analýzy je ale dostačující.

Chybějící hodnoty odhaduji pomocí softwaru SPSS, kde používám proceduru Binary Logistic Regression. Do procedury vstupují všechny případy (respondenti), kteří odpověděli na otázku ohledně záměru účasti u hypotetických voleb konaných týden od dotazování. Závislá proměnná je dichotomická – deklarovaná účast u voleb do PSP ČR

u respondentů, kteří nedeklarovali záměr zúčastnit se voleb než už těch, kteří deklarovali záměr zúčastnit se voleb (45 % oproti 35 % dle pořadí, Mann-Whitney p=0.000).

v roce 2017 (účast/neúčast). Nezávislými proměnnými, které používám pro odhadnutí chybějících hodnot, a které můžeme obecně shrnout jako 1) sociodemografické ukazatele: nejvyšší dosažené vzdělání respondenta[27], ekonomické postavení respondenta[28] a věk respondenta[29], 2) proměnné spojené s volebním chováním: záměr účasti u hypotetických voleb konaných týden od dotazování[30], volební účast u voleb do PS ČR v roce 2013[31], a zájem o politiku[32]. (výběr srov. s Tabery and Šamanová 2012) Vzhledem k tomu, tyto že nezávislé proměnné (kovariáty) určitě nelze považovat za kardinální, nastavím je v nabídce analýzy jako kategorické. (IBM Corp. 2015, 4)

Během analýzy se mi nepodařilo najít závažnou kolinearitu[33], tudíž přepokládám splnění předpokladu, že nezávislé proměnné nesmí být silně závislé mezi sebou (Blalock 1963). Test dobré shody modelu indikuje, že nastavenou proceduru logistické regrese lze na tato data aplikovat[34]. Zvolenými prediktory se pak podařilo vysvětlit téměř polovinu v datech přítomného rozptylu.[20]

Samotnou imputaci jsem provedl na základě uložených hodnot pravděpodobností (hodnoty spojitě 0 až 1 včetně) příslušnosti ke třídě 1 (účast) vypočítaných procedurou logistické regrese. Imputace se týkala pouze respondentů, kterým ve follow-upu chyběla hodnota proměnné deklarované účasti. Pokud hodnota výše zmíněné pravděpodobnosti u respondenta byla rovna 90 % a více, byla respondentovi imputována hodnota účasti. U všech pravděpodobností nižších než 90 % byla respondentovi imputována neúčast. Hraniční hodnota 90 % byla stanovena iterativně na základě relativní četnosti volební účasti ve vážených imputovaných datech oproti reálné volební účasti ve volbách do PSP

[27] Rekódováno na 4 kategorie: Základní škola / SŠ bez maturity / SŠ s maturitou / Vysoká škola

[28] Rekódováno na 6 kategorií: Zaměstnanec / Podnikatel / Nezaměstnaný / Důchodce / Student / Jiné, v domácnosti, mateřská dovolená

[29] Rekódováno na 7 kategorií: 15–24 let / 25–34 let / 35–44 let / 45–54 let / 55–64 let / 65–74 let / 75 let a více

[30] 6 kategorií: Rozhodně ano / Spíše ano / Spíše ne / Rozhodně ne / Nemá volební právo / Nevím, nemohu posoudit

[31] Rekódováno na 3 kategorie: Účast / Neúčast / Neví, nebo neměl(a) volební právo

[32] 5 kategorií: Velmi se zajímám / Spíše se zajímám / Spíše se nezajímám / Vůbec se nezajímám / Nevím, nemohu posoudit

[33] Tu jsem posuzoval skrze VIF (Variance Inflation Factor), který indikuje, nakolik je rozptyl regresního koeficientu navýšen díky kolinearitě. (Wuensch 2019, 8) V tomto případě byly VIF pro všechny prediktory menší než 1,8. To znamená, že směrodatné odchylky všech regresních koeficientů jsou vždy méně než 1,34 krát (tzn. druhá odmocnina rozptylu) větší, než by byly v případě, že by prediktor nebyl korelován s ostatními prediktory.

[34] Hosmer and Lemeshow Test Chi-square=1,355; p=0,995. (Paul, Pennell, and Lemeshow 2012) Podrobněji viz *Přílohy*.

ČR 2017. Zatímco volební účast u voleb do PSP ČR v roce 2017 byla 60,8 %, v neimputovaných datech byla relativní četnost reportované volební účasti 77,8 %, a v datech doplněných imputací potom 64,9 %[35].

### 3.2.2. Preferovaná strana (CHPS)

Pokud při dotazování respondent deklaruje, že plánuje ve volbách volit konkrétní politickou stranu, tuto samotnou odpověď nemůžeme automaticky považovat za neomylný indikátor jeho budoucího volebního chování. To je konec konců i hlavním důvodem, proč se některé agentury (například v případě dvou z volebních modelů, s nimiž svůj navrhovaný model srovnávám, modelů Median a CVVM2017 viz *4.1 Srovnávané výstupy*) ptají respondentů na více zvažovaných stran, někdy s možností indikace preferované strany. S touto skutečností se snaží na základě poznatků z CHPS vypořádat i druhý podcíl sekundární analýzy dat v této diplomové práci. Hypotézou, kterou ověřuji, a která nám slouží k získání konkrétních hodnot pro konstrukci volebního modelu, je:

H2: Respondenti, kteří před volbami do PSP ČR v roce 2017 deklarovali, že by v hypotetických volbách konaných týden od dotazování volili určitou politickou stranu, po těchto volbách reportovali, že tuto stranu volili, častěji než respondenti, kteří tuto preferenci před volbami nedeklarovali.

Hypotéza sama o sobě je opět poměrně intuitivní, ale jejím hlavním účelem je i zde poskytnout vodítko pro optimalizaci výpočtu volebního modelu[36]. Proměnné v hypotéze se vztahují k dotazníkovým položkám *„Představte si, že by příští týden byly volby do Poslanecké sněmovny. Šel/Šla byste volit?“* (CHPS 2018, 68), viz též část analýzy *3.2.1 Volební účast, záměr volební účasti a minulá volební účast (CHPS)*, a *„Kterou stranu jste volil(a)?“*[37] (vztahuje se k volbám do PSP ČR v roce 2017). Z těchto dvou proměnných jsem vypočítal novou proměnnou shody záměru a volby – pokud se shodovala strana

[35] Legitimní kritikou tohoto postupu je fakt, že během doplňování chybějících hodnot v datech CHPS pomocí logistické regrese nemám jistotu, že ta skupina respondentů, která se nezúčastnila follow-upu, není ve svém volebním chovní nějakým dalším podstatným způsobem odlišná od té, která se follow-upu zúčastnila. Takový rozdíl bych touto logistickou regresí nebyl schopen postihnout. Je bohužel vlastností jakéhokoliv výzkumu, který probíhá ve více vlnách, že jen málokdy dosahuje retence všech respondentů. (Olsen 2005) Proceduru logistické regrese volím právě proto, abych co nejvíce využil dostupné informace pro odhad těch nedostupných. Informace, které se v datech nenacházejí, bohužel v rámci sekundární analýzy postihnout nelze. Nedoplnění chybějících hodnot, ani použití jiné metody odhadu by tento problém ale neodstranilo.

[36] Shodu záměru volby strany a reportované volené strany už podle jednotlivých měsíců sběru zkoumal (Linek 2018), já zde nicméně pro účely konstrukce modelu vypočítávám shodu v datech omezených na výše definované období sběru.

[37] Otázku cituji z neveřejného dotazníku follow-upu třetí vlny CHPS.

deklarovaná před volbami s reportovanou volenou stranou ve follow-upu, dostal respondent hodnotu 1, v opačném případě hodnotu 0. Do analýzy vstupují jen ti respondenti dotazovaní v posledních dvou měsících před volbami do PSP ČR v roce 2017, kteří uvedli konkrétní stranu (nikoliv „jiné strany“) jak v záměru volby, tak ve follow-up dotazníku vlastní volby. Z analýzy jsem zároveň vyřadil respondenty, kteří v původním dotazníku uvedli záměr volit koalici KDU-ČSL/STAN – ta byla před volbami zrušena, a proto není reálné, aby takový respondent mohl avizovanou stranu volit.

V případě tohoto podcíle analýzy nedoplňuji chybějící hodnoty proměnných, zejména proto, že do analýzy vstupují jen respondenti, kteří nevylučují svou účast u hypotetických voleb, a podíl chybějících odpovědí ve follow-upu je tím pádem menší, a také proto, že stran, které si respondent může vybrat, je poměrně mnoho – odhadování chybějících hodnot by bylo výrazně obtížnější než v případě dichotomické proměnné reportované účasti. Pro konstrukci samotného navrhovaného modelu je finálním výstupem poměr voličů, kteří reportují, že nakonec volili tu samu stranu, kterou před volbami deklarovali záměr volit. Konkrétní způsob aplikace viz kapitola *4.2 Konstrukce navrhovaného volebního modelu.*

### 3.2.3. Minulé volební chování (Omnibus Median)

Pro třetí podcíl sekundární analýzy dat používám jiný datový soubor – následující odstavce už se tedy týkají dat z omnibusového šetření agentury Median. Cílem této části sekundární analýzy je zjistit, jak se v omnibusu Median u jednotlivých politických stran liší odchylka četností reportované volby v minulých volbách do PSP ČR (2013) od výsledků příslušných voleb 1) v datech sebraných bezprostředně po volbách do PSP ČR 2013 od 2) analogické odchylky v datech sebraných bezprostředně před volbami do PSP ČR 2017.

Analyzuji data vážená[38] podle socio-demografických charakteristik (pohlaví, věk, dosažené vzdělání, kraj, velikost místa bydliště a ekonomické postavení respondenta). V analýze mě zajímá i ta část odchylky, která může být způsobena over-coverage respondentů volících konkrétní politické strany, a proto zde (na rozdíl třeba od analýzy CHPS, a přestože je to praxí při konstrukci volebního modelu agentury Median) data nevážím ani podle volební účasti ani podle výsledků minulých voleb. Minulé volební chování je v omnibusu Median

[38] Všechny váhové proměnné jsem vypočítal pomocí softwaru *Weighting* vlastněného společností Median. Ten používá iterativní metodu vážení, jejímž kritériem je konvergence celé tabulky (posuzovaná testem Chí-kvadrát), a vychází z diplomové práce J. Neustadta na MFF UK. (1997)

dotazováno otázkou *„Zúčastnil(a) jste se posledních voleb do Poslanecké sněmovny, které proběhly na konci letošního října? Pokud ano, koho jste volil(a)?“*[39] v případě sběru dat v roce 2013 a *„Kterou stranu nebo hnutí jste volil(a) při volbách do Poslanecké sněmovny, které proběhly v říjnu 2013?“* během sběru dat v roce 2017. (MEDIAN 2017, 12)

Proměnná minulé volby může nabývat hodnot všech politických stran, které jsou v době dotazování zastoupené v poslanecké sněmovně/Ostatních politických stran/Neúčasti u voleb, nebo si nepamatuje/Respondent neměl v příslušné době volební právo (typicky proto, že mu ještě nebylo 18 let).

Z obou datových souborů (2013 a 2017, viz sekci Data této kapitoly) vytvořím frekvenční tabulku reportovaného volebního jednání u voleb do PSP ČR v roce 2013, jejíž základem budou všichni respondenti, kteří deklarují, že se daných voleb účastnili, a reportují, kterou stranu v nich volili. Oboje tyto četnosti potom srovnávám s výsledky skutečných voleb do PSP ČR v roce 2013. (Český Statistický Úřad 2013) Zaměřuji se jednak na odchylky obou četností od reálných výsledků, ale zároveň také rozdíl mezi těmito dvěma odchylkami samotnými.

Pro tyto odchylky totiž předpokládám dva hlavní zdroje. Prvním zdrojem je způsob výběru respondentů – nemám zde na mysli pouze výběrovou chybu, ale zejména možnou o over-coverage voličů jednotlivých politických stran, a analogicky under-coverage u jiných (srov. s Lipps, Pekari, and Roberts 2015) – toto u předvolebních výzkumů některých agentur může vysvětlovat např. konzistentní nadhodnocování stranických preferencí konkrétních stran oproti jejich preferencím ve výstupech jiných agentur). U tohoto prvního zdroje tedy předpokládám, že tvoří většinu odchylek v datech sebraných bezprostředně po volbách do PSP ČR v roce 2013 (protože respondenti ještě neměli čas zapomenout, koho volili). Druhým předpokládaným zdrojem je vliv uplynulého času – tam naopak předpokládám, že má největší vliv v datech z roku 2017 sebraných těsně před volbami, protože míra zapomínání se zvětšuje s přibývajícím časem. (Sudman and Bradburn 1973) Rozdíl, o který se bude odchylka z roku 2017 lišit oproti odchylce z roku 2013, pro mě bude indikativní tomu, jak silně a kterým směrem funguje u konkrétní politické strany vliv uplynulého času. Vzhledem ke kombinaci těchto dvou zdrojů chyby v novějším datovém souboru pak předpokládám hypotézu:

[39] Citováno z neveřejného archivovaného dotazníku poskytnutého společností Median.

H3: Odchylka stranických preferencí dotazovaných v září a říjnu 2017 od výsledků voleb do PSP ČR v roce 2013 je větší, než odchylka stranických preferencí dotazovaných v listopadu a prosinci 2013 od výsledků voleb do PSP ČR v roce 2013.

Odchylky zkoumám odděleně pro jednotlivé politické strany, protože u nich předpokládám různé vlivy faktorů jako je např. míra stranické identifikace. (Linek 2009, 193–94) Statistiky, jejichž prostřednictvím budu posuzovat hypotézu H3 jsou: Suma čtverců odchylek, Maximální absolutní odchylka u strany, Průměrná absolutní odchylka u stran, Maximální relativní odchylka u strany a Průměrná relativní odchylka u stran přičemž „odchylkou" zde mám na mysli rozdíl relativní četnosti konkrétní strany v daném datovém souboru od podílu volebního zisku té samé strany ve volbách do PSP ČR v roce 2013. Tyto statistiky používám (kromě jiných) zároveň k finálnímu srovnání mnou navrhovaného volebního modelu, kde je také podrobněji popisuji (viz kapitola *4.3 Kritéria srovnávání volebních modelů*).

## 3.3. Výsledky Sekundární analýzy dat a jejich diskuse

### 3.3.1. Volební účast, záměr volební účasti a minulá volební účast

Jako první se zaměřím na hypotézu H1a, a tedy vztah mezi deklarovaným záměrem volební účasti v případných volbách a reportované účasti u voleb do PSP ČR v roce 2017. Při pohledu na kontingenční tabulku (*Tabulka 2*) je evidentní, že podíl reportované účasti se zmenšuje od skupiny respondentů rozhodnutých se voleb zúčastnit (96 %) směrem k respondentům rozhodnutým se voleb nezúčastnit (39 %). Celková míra účasti (87 %) přibližně odpovídá míře účasti u skupiny respondentů se záměrem se voleb spíše zúčastnit (88 %).

| | Volby 2017 | | | |
|---|---|---|---|---|
| Záměr | Neúčast | Účast | Celkem | N |
| Rozhodně ano | 4 % | 96 % | 100 % | 907 |
| Spíše ano | 12 % | 88 % | 100 % | 235 |
| Spíš ne | 34 % | 66 % | 100 % | 99 |
| Rozhodně ne | 61 % | 39 % | 100 % | 128 |
| Celkem | 13 % | 87 % | 100 % | 1369 |

*Tabulka 2: Volební účast a záměr volební účasti*

Nulovou hypotézu testu nezávislosti, tedy že proměnná záměru volební účasti a reportované volební účasti jsou nezávislé můžeme zamítnout (Chi-kvadrát=384,2; p=0,000), a tím pádem přijmout a doplnit o konkrétní hodnoty původně stanovenou

alternativní hypotézu H1a. Respondenti, kteří deklarovali záměr účastnit se hypotetických voleb do PSP ČR konaných týden od dotazování, ve follow-upu potom reportovali účast u voleb do PSP ČR v roce 2017 častěji (96 % u „Rozhodně ano“ a 88 % u „Spíše ano“) než respondenti, kteří záměr se zúčastnit nedeklarovali (66 % u „Spíše ne“ a 39 % u „Rozhodně ne“). Mezi oběma proměnnými je poměrně silná statisticky významná závislost (Cramerovo V=0,52; p=0,000). Nyní použijeme otestovanou proměnnou záměru volební účasti jako kontrolní proměnnou při analýze vztahu mezi volební účastí u voleb do PSP ČR v roce 2013 a reportovanou účastí v roce 2017.

| | | Volby 2017 | | | |
|---|---|---|---|---|---|
| Záměr | Minulá účast | Neúčast | Účast | Celkem | N |
| Rozhodně ano | Ano | 3 % | 97 % | 100 % | 869 |
| | Ne | 21 % | 79 % | 100 % | 38 |
| Spíše ano | Ano | 8 % | 92 % | 100 % | 186 |
| | Ne | 27 % | 73 % | 100 % | 49 |
| Spíše ne | Ano | 18 % | 82 % | 100 % | 39 |
| | Ne | 45 % | 55 % | 100 % | 60 |
| Rozhodně ne | Ano | 35 % | 65 % | 100 % | 20 |
| | Ne | 66 % | 34 % | 100 % | 108 |
| Celkem | Ano | 5 % | 95 % | 100 % | 1114 |
| | Ne | 47 % | 53 % | 100 % | 255 |
| | Celkem | 13 % | 87 % | 100 % | 1369 |

*Tabulka 3: Volební účast, minulá volební účast a záměr volební účasti (kontrolní proměnná)*

Nejdříve popíšu vztah reportované účasti v roce 2017 a proměnné minulé volební účasti celkově – ten je statisticky významný a opět poměrně silný (viz *Tabulka 4*). Zatímco relativní četnost reportované účasti u těch, kdo říkají že v roce 2013 byli u voleb je 95 %, u těch, kdo se v roce 2013 nezúčastnili je to 53 %. Na základě výše zmíněných výsledků statistického testu můžeme tvrdit, že se podařilo potvrdit i hypotézu H1b o vztahu mezi minulou účastí v roce 2013 a reportovanou účastí v roce 2017. V tom případě mě ale nyní zajímá, jestli zmíněný vztah obstojí i pokud použijeme záměr účasti u případných voleb jako kontrolní proměnnou (tzn. když respondenty zároveň rozdělíme na skupiny podle záměru).

Při pohledu na horní část tabulky (*Tabulka 3*) je viditelné, že: 1) výše popsaný vztah mezi minulou účastí a účastí reportovanou v roce 2017 zůstal ve všech skupinách podle záměru volební účasti zachován. Zároveň vidíme, že 2) také ve **všech** skupinách podle záměru volební účasti došlo k té samé změně: oproti celkově četnosti reportované účasti u voleb

(viz *Tabulka 2*) v dané skupině je reportovaná účast u těch, kdo se zúčastnili voleb v roce 2013, větší, a u těch, kdo se nezúčastnili, menší. Například v případě skupiny „Rozhodně ne“ se četnost reportované účasti v roce 2017 těch kdo se zúčastnili minulých voleb (65 %) prakticky vyrovná celkově četnosti skupiny „Spíše ne“, u které jsme pro záměr nekontrolovali (66 %). Když navíc tyto informace zkombinujeme s tabulkou výsledků testů nezávislosti (viz *Tabulka 4*), vidíme, že tento vztah zůstal statisticky významný ve všech skupinách respondentů podle záměru volební účasti.

| Záměr (kontrola) | Chi-kvadrát | p | Cramerovo V | p |
|---|---|---|---|---|
| Rozhodně ano | 32,914 | 0,000 | 0,190 | 0,000 |
| Spíše ano | 12,601 | 0,000 | 0,232 | 0,000 |
| Spíše ne | 7,671 | 0,006 | 0,278 | 0,006 |
| Rozhodně ne | 6,699 | 0,010 | 0,229 | 0,010 |
| Celkem | 325,685 | 0,000 | 0,488 | 0,000 |

*Tabulka 4: Hodnoty Chi-kvadrát testu a Cramerova V pro kontingenční tabulku volební účasti a minulé volební účasti s kontrolou pro záměr volební účasti*

Tím pádem se nejen podařilo potvrdit i hypotézu H1c, ale navíc ještě analogický vztah platí ve všech skupinách podle záměru účasti. Obě zkoumané proměnné tak mají významný vztah s reportovanou účasti jak samostatně, tak v interakci. O té nám ještě může napovědět elaborace na základě Cramerova V. To ve vrstvené tabulce dosahuje nižších hodnot než v původní, nevrstvené. To indikuje buďto sporný vztah, anebo jednu z analyzovaných proměnných jako intervenující – to z těchto samotných statistik nelze s určitostí odlišit. Nicméně vzhledem k tomu, že existuje určitý kauzální vztah mezi minulou volbou a záměrem jít k volbám v současné době (Denny and Doyle 2009), považuji za smysluplné považovat proměnnou záměru volební účasti ve vztahu mezi minulou účastí u voleb v roce 2013 a reportovanou účastí u voleb v roce 2017 za intervenující.

Za důležitý poznatek do diskuse z této části sekundární analýzy dat považuji to, že respondenti, kteří deklarují záměr rozhodně se nezúčastnit případných voleb, mají podle CHPS 40% pravděpodobnost se reálných voleb zúčastnit – tato pravděpodobnost navíc vzrostla na 65 % u těch z nich, kteří se zúčastnili voleb v roce 2013. Všechny volební modely, které v této práci srovnávám, nejen že tyto respondenty ve volebním modelu zanedbávají, ale např. u společnosti Median se jich kvůli filtru v dotazníku už dále nedotazují na zvažované politické strany – jedná se o podstatnou skupinu (de facto)

nerozhodnutých voličů, které kvůli nastavení dotazníků nelze do modelu zahrnout. Tuto praxi mi na základě poznatků ze sekundární analýzy přijde oprávněné změnit.

### 3.3.2. Preferovaná strana

| | Shoda záměru a volby | |
|---|---|---|
| | Čet. | % |
| Neshoda | 360 | 37 % |
| Shoda | 612 | 63 % |
| Celkem | 972 | 100 % |

*Tabulka 5:Shoda záměru volby a reálně volené strany*

Respondenti, u nichž se zamýšlená volená strana shoduje s reportovanou volenou stranou, jsou v analyzovaném souboru zastoupeni více (63 %), než respondenti, u nichž se záměr s volbou neshoduje (37 %)[40]. Záměr volené strany je tedy častěji spojený i s volbou dané strany. Jak už bylo řešeno výše, hypotéza H3 byla spíše banální a není velkým úspěchem, že se ji podařilo prokázat[41]. Její výsledek nicméně využívám při konstrukci navrhovaného volebního modelu (viz kapitola *4.2 Konstrukce navrhovaného volebního modelu*).

### 3.3.3. Minulé volební chování – vliv uplynulého času

Třetí a poslední částí sekundární analýzy dat, jejíž výsledky budu nyní prezentovat, je vliv času uplynulého od příslušných minulých voleb na proměnnou minulé volby. V následující tabulce (*Tabulka* 6) prezentuji kromě výsledků voleb do PSP ČR i rozložení deklarované minulé volby ve volbách do PSP ČR v roce 2013, a to jak z omnibusu Median sbíraného bezprostředně po těchto volbách (Median 2013) a bezprostředně před následujícími o necelé čtyři roky později (Median 2017). Vedle sloupce relativních četností proměnné minulé volby každého z obou sběrů je jejich příslušná odchylka, tedy rozdíl mezi četností

[40] Pro jistotu jsem pro posouzení hypotézy použil ještě jednoduchý neparametrický binomiální test vypočítaný v softwaru SPSS, který posuzuje, jestli se pozorované četnosti výběrového souboru statisticky liší od očekávaných při binomiálním rozložení s pravděpodobnosti 0,5 – tedy v rozložení, kde mají obě hodnoty stejnou pravděpodobnost (p=0,000).

[41] Tato část sekundární analýzy dat v sobě ale také obsahovala jeden konkrétní příklad toho, jak výzkumné nastavení sběru dat může ovlivnit potenciál jejich sekundární analýzy. Zatímco u možnosti „jiná strana“ doplňuje podle dotazníku tazatel do textového pole kterou jinou stranu měl respondent na mysli, u otázky je zároveň také možnost „ostatní výroky“ u které už ale toto textové pole přítomné není, a tazatel je v dotazníku přímo instruován jako „ostatní výroky“ zaznamenat „odpovědi, které nejsou přímým označením politické strany (např. respondent řekne, že by volil „Babiše“, nebo „nácky“)“. (CHPS 2018, 68) U takového respondenta se už v datech nedozvíme, o jaký výroky se jednalo – přesto ale tvrdím, že právě u výroků jako „Babiše“, nebo např. „Okamuru“ by bylo možné s dostatečnou mírou jistoty dovodit, o které politické subjekty se jedná a upřednostnit tak indikativní hodnotu odpovědi oproti striktní shodě. Takto bohužel přicházíme o desítky až stovky respondentů, jejichž odpovědi by mohli v důsledku např. ještě zkvalitnit zde navrhovaný volební model. Analýzu to nicméně neznemožnilo a jedná se spíše o detailní poznámku než o zásadní problém.

konkrétní strany v daném sběru a jejího volebního výsledku. Poslední sloupec tabulky pak obsahuje rozdíl těchto odchylek. Z tabulky je patrné, že minulí voliči ČSSD byli oproti výsledkům minulých voleb v datovém souboru nadreprezentováni už přímo po volbách (odchylka +6,16 p.b.) – taková odchylka je způsobena pravděpodobně výběrovou chybou nebo jinou vlastností sběru. Nadreprezentované bylo v této vlně i ANO 2011 (+3,20 p.b.), podreprezentováno naopak hnutí Úsvit (-3,06 p.b.). Největší rozdíl mezi oběma sběry se nachází u KSČM a ODS. Zatímco KSČM je v novějším sběru zastoupeno podílem o 5,63 procentních bodů menším, zastoupení ODS stouplo o 3,32 p.b.

| | Volby 2013 | Median 2013 | Odch. (p. b.) | Median 2017 | Odch. (p. b.) | Rozdíl odch. |
|---|---|---|---|---|---|---|
| ČSSD | 20,46 % | 26,63 % | 6,16 | 25,29 % | 4,82 | -1,34 |
| ANO 2011 | 18,66 % | 21,87 % | 3,20 | 23,32 % | 4,66 | 1,45 |
| KSČM | 14,91 % | 14,00 % | -0,92 | 8,37 % | -6,55 | -5,63 |
| TOP 09 | 12,00 % | 12,10 % | 0,10 | 12,88 % | 0,88 | 0,78 |
| ODS | 7,73 % | 7,91 % | 0,18 | 11,23 % | 3,50 | 3,32 |
| Úsvit | 6,89 % | 3,83 % | -3,06 | 3,80 % | -3,09 | -0,03 |
| KDU-ČSL | 6,78 % | 5,81 % | -0,97 | 6,79 % | 0,01 | 0,98 |
| Strana zelených | 3,20 % | 2,42 % | -0,78 | 0,89 % | -2,31 | -1,53 |
| Česká pirátská strana | 2,66 % | 2,16 % | -0,50 | 3,69 % | 1,03 | 1,53 |
| Svobodní | 2,47 % | 1,18 % | -1,29 | 1,54 % | -0,92 | 0,36 |
| jiná strana | 4,22 % | 2,09 % | -2,13 | 2,19 % | -2,03 | 0,10 |

*Tabulka 6: Odchylky minulé volby v omnibusu Median XI/XII 2013 a IX/X 2017 od výsledků voleb do PSP ČR v roce 2013*

Zkoumaná hypotéza se nicméně nezabývá konkrétními politickými stranami, ale statistikami odchylek četností proměnné v obou sběrech od výsledků voleb v roce 2013. Ty shrnuje *Tabulka 7*. Všechny sledované statistiky indikují, že odchylka deklarované minulé volby u voleb do PSP ČR v roce 2013 se mezi oběma datovými soubory zvětšila, a tedy že můžeme na základě stanovených kritérií přijmout hypotézu H3 a vycházet z těchto údajů při návrhu optimalizace výpočtu volebního modelu (viz kapitola *4.2 Konstrukce navrhovaného volebního modelu*). Nejmenší poměrný rozdíl nastává u maximální absolutní odchylky, kde nicméně nešlo o nárůst odchylky té samé strany, ale o výměnu strany na prvním místě ve velikosti absolutní odchylky (z ČSSD na KSČM).

| | Median 2013 | Median 2017 |
|---|---|---|
| Suma čtverců odchylek | 66,47 | 121,71 |
| Maximální absolutní odch. | 6,16 | 6,55 |
| Průměrná absolutní odch. | 1,75 | 2,71 |
| Maximální relativní odch. | 52 % | 72 % |
| Průměrná relativní odch. | 24 % | 35 % |

*Tabulka 7: Vybrané statistiky odchylek minulé volby omnibusu Median XI/XII 2013 a IX/X 2017 od výsledků voleb do PSP ČR v roce 2013*

Zejména velký rozdíl záporným směrem u KSČM (-5,63 p.b.) si ale podle mého názoru ještě zaslouží podrobnější komentář. Pro interpretaci tohoto poklesu mi přijdou adekvátní dva hlavní důvody. Prvním z nich je změna demografického profilu respondentů (během uplynulých let od minulých voleb). Vzhledem k tomu, že průměrný věk voličů KSČM je vyšší, než u jiných stran[42], nabízí se interpretace, že určitá část respondentů, kteří v minulých volbách volila KSČM, ubyla ze základní populace výzkumu omnibus Median – ten má horní věkové omezení 79 let. (Prokop 2019) K úbytku respondentů tedy nemusí dojít pouze z důvodu např. úmrtí, ale zejména kvůli jejich přesunu do skupiny respondentů nedosažitelných agenturní tazatelskou sítí společnosti Median. Druhou tezí, která se nabízí pro interpretaci tohoto rozdílu, ale tentokrát ji lze aplikovat i na změny u dalších ze sledovaných politických stran, je že misreporting minulé volby v určité míře může kopírovat vývoj volebních preferencí. Za dobu od voleb do PSP ČR v roce 2013 by to odpovídalo poklesu preferencí i četností minulé volby u ČSSD, nárůstu u ODS nebo i nárůstu u Pirátské strany. Předpoklad u takového vztahu by tedy byl, že respondenti misreportují svou minulou volbu na základě svých preferencí v době dotazování – ať už proto, že na svou minulou volbu zapomněli, nechtějí se přiznat ke změně své preference, nebo naopak k tomu, že v minulých volbách volili konkrétní stranu (nebo že se jich ani nezúčastnili)[43].

# 4. Navrhovaný volební model

V této kapitole se věnuji hlavnímu cíli této diplomové práce, a tím je konkrétní návrh výpočtu volebního modelu (metodologická optimalizace na základě poznatků sekundární

[42] Podle http://showme.median.cz/snemovni-volby/#/ (naposledy navštíveno 4. 5. 2019) byl v období od května do října 2017 průměrný věk voličů KSČM 61 let – např. u ANO 2011 potom 53 let, a např. u ODS 46 let.

[43] Tuto interpretaci bohužel na těchto dvou datových souborech nelze dostatečně ověřit, ale jedná se o velmi zajímavý směr pro budoucí výzkumnou práci, který detailněji popíšu v závěrech této diplomové práce.

analýzy dat), a srovnání jeho výstupů s výstupy dalších modelů. Nejdříve jmenuji a stručně popisuji jednotlivé další volební modely (způsoby konstrukce volebního modelu), se kterými budu výstupy mnou navrhovaného postupu srovnávat. Poté podrobněji rozepisuji způsob, kterým budu na základě výše popsaných poznatků model konstruovat. Kapitolu uzavírá srovnání výsledného modelu.

## 4.1. Srovnávané výstupy

Pokud se v této práci snažím navrhnout metodologickou optimalizaci, měl bych být taky schopen argumentovat, že a proč je můj postup pro danou aplikaci vhodnější než jiné stávající postupy, nebo alespoň v čem se od ostatních postupu liší, a z jakého důvodu. Proto výstup navrženého způsobu výpočtu volebního modelu srovnávám s výstupy čtyř jiných způsobů výpočtu[44] – převzatých z jiných používaných volebních modelů a pro porovnatelnost vypočítaných na těch samých datech (omnibus Median září/říjen 2017). Porovnávám nakolik, a jak konkrétně, se výstupy těchto volebních modelů liší od výsledků voleb do PSP ČR v roce 2017. (Český Statistický Úřad 2017) Nyní podrobněji popíšu všechny čtyři srovnávané modely. Přestože zevrubně už byly tyto modely popsány a srovnány (Chábová 2013), pro konstrukci srovnatelných výstupů je potřeba popsat (a co nejlépe replikovat) konkrétní metodiky výpočtu. Proto v následujících odstavcích popisuji způsoby výpočtu jednotlivých srovnávaných výpočtů a případné odchylky od jejich původní metodiky a způsob řešení z nich plynoucích problémů. Kromě popisu konstrukce srovnávaných modelů pro přehlednost přikládám i dvě tabulky, které shrnují a srovnávají dva jejich nejdůležitější parametry. (*Tabulka 8*; *Tabulka 9*)

Konstrukce těchto srovnávaných modelů není hlavním cílem této diplomové práce. Proto je následující část spíše popisná než kritická. Zároveň zde považuji za nutné ještě jednou upozornit, že výstupy, které zde srovnávám pojmenované např. „model CVVM“ či „KANTAR“ neodpovídají hodnotou výsledků žádné konkrétní tiskové zprávě CVVM nebo TNS KANTAR, protože se jedná o výstupy vypočítané na datech z omnibusu Median. Tento postup byl zvolen, aby srovnávané výsledky modelu nebyly zaneseny vlivem různých metodik a různých období sběru dat. Ani výstupy modelu Median neodpovídají

[44] Volebním modelem, který ze srovnání záměrně vynechávám, je volební model publikovaný agenturou STEM. (viz např. STEM 2019) Přestože se jedná o kredibilní model, který se svými charakteristikami zároveň podstatně liší od ostatních srovnávaných modelů, a tím pádem by byl velmi zajímavým referenčním bodem, nelze ho v této analýze do srovnání zařadit. Konstrukce modelu agentury STEM totiž využívá proměnné jako je zájem respondenta o politické dění nebo vazba k preferované straně, které nejsou v omnibusu Median přítomny, a tak není možné model STEM ve srovnání uspokojivě replikovat.

žádné tiskové zprávě, neboť výstupy vypočítané z tohoto konkrétního datového souboru nebyly publikovány z důvodu informačního embarga před volbami.

| Záměr volební účasti | MEDIAN | CVVM 2017 | CVVM 2018 | KANTAR |
|---|---|---|---|---|
| Rozhodně ano | 1 | 1 | 1 | 1 |
| Spíše ano | 0,7 | 1 | 0,75 | 1 |
| Spíše ne | 0,2 | 0* | 0,25 | 1 |
| Rozhodně ne | 0* | 0* | 0* | 0* |

**respondent vůbec nevstupuje do modelu*

*Tabulka 8: Parametr záměru volební účasti ve srovnávaných modelech (názvy kategorií podle modelu MEDIAN)*

| Jistota záměru volby preferované strany | Parametr jistoty preferované strany: MEDIAN | CVVM 2017 | CVVM 2018 | KANTAR |
|---|---|---|---|---|
| Zcela rozhodnutý | 1 | 1 | 1 | 1 |
| Skoro jistý | 0,8 | 1 | 0,8 | 1 |
| Nepříliš jistý | 0,7 | 1 | 0,5 | 1 |
| Vůbec si není jistý | 0,6 | 1 | 0,33 | 0* |
| Neuvádí preferovanou stranu | 1 / počet zvažovaných | 0* | 0,33 | 0* |

**respondent vůbec nevstupuje do modelu*

*Tabulka 9: Parametr jistoty preferované strany ve srovnávaných volebních modelech (názvy kategorií dle modelu MEDIAN)*

### 4.1.1. Model MEDIAN

První srovnávaný výstup nazývám model MEDIAN. Jedná se o způsob, kterým agentura Median před volbami do PSP ČR v roce 2017 vypočítávala svůj publikovaný volební model. Výstupy z něho jsou prezentovány například v níže citovaných tiskových zprávách. Model byl navržen v diplomové práci konzultanta (Prokop 2012) V současné době společnost Median publikuje volební model v ještě mírně odlišné podobě.

Do modelu vstupují pouze ti respondenti, kteří **nevyloučili účast u voleb** v otázce *„Pokud by se nyní konaly volby do Poslanecké sněmovny, šel/šla byste volit?"* (MEDIAN 2018, 12) Podobně jako v modelu CVVM2018 (viz dále) je k odpovědi „Určitě ano" přiřazena ve výpočtu modelu pravděpodobnost 100 %, a následně už odlišně 70 % a 20 % u možností „spíše ano" a „spíše ne". Dále dostanou respondenti otevřenou otázku na **první zvažovanou stranu**: *"Kterou stranu nebo hnutí byste nejspíše volil(a), pokud byste v současnosti šel / šla k volbám do Poslanecké sněmovny?"* (MEDIAN 2018, 12) Kromě první zvažované strany se tazatel respondenta ptá i na **další možné zvažované strany**:

"Můžete prosím uvést všechny strany, o jejichž volbě byste uvažoval(a)“ (MEDIAN 2018, 12) – výčtová otázka s rotujícím seznamem možných odpovědí. Pokud respondent **uvedl preferovanou stranu**[45], dostane u ní otázku na **jistotu volby** *„Představte si, že by se v horizontu jednoho měsíce opravdu konaly volby do Poslanecké sněmovny. Nakolik jste si jistý/á, že byste v případě své účasti volil(a) právě tuto stranu?“* (MEDIAN 2018, 12) Pokud je si zcela jistý, je této **straně přiřazena pravděpodobnost** 100 %. Pokud si je skoro jistý, je to 80 %, a dále 70 % a 60 % u odpovědí „nepříliš jistý“ a „vůbec si nejsem jistý“ dle pořadí. V takovém případě se doplněk do 100 % rovnoměrně dělí mezi všechny ostatní zvažované strany. Pokud ale respondent **neuvedl preferovanou stranu**, dělí se 100 % rovnoměrně mezi všechny jím zvažované strany.

K tomu jsou data ještě **zvážena** podle věku, pohlaví, kraje, velikosti místa bydliště, dosaženého vzdělání, pracovního statusu, intenzitu užívání internetu a volebního chování v minulých volbách do PSP ČR. Mnou navrhovaný model z modelu částečně vychází, rozdíl podrobněji popisuji v sekci „Konstrukce navrhovaného volebního modelu“ dále.

### 4.1.2. Model CVVM2017

Volební model, který Centrum pro výzkum veřejného mínění publikovalo v roce 2017, byl ve srovnání např. s modelem agentury Median, ale i se současným modelem samotného CVVM, poměrně jednoduchý. (srov. s Chábová 2013, 31) Do volebního modelu vstupují pouze respondenti, kteří **indikovali záměr účastnit se** hypotetických voleb odpovědí „Určitě ano“ nebo „Spíše ano“ v otázce *„Představte si, že by příští týden byly volby do Poslanecké sněmovny. Šel byste volit?“* (CVVM 2017, 2), a zároveň **deklarovali záměr volit konkrétní stranu**. Výstup volebního modelu CVVM2017 pak jejich stranické preference. Model do srovnání zahrnuji zejména proto, abych srovnával volební modely o různé míře komplexity výpočtu. Jak preferovanou stranu (viz *4.1.1 Model MEDIAN*), tak záměr volební účasti jsem na analyzovaných datech schopen dobře replikovat.

### 4.1.3. Model CVVM2018

Třetím volebním modelem, se kterým budu svůj návrh srovnávat, je model, který publikuje CVVM v současné době, konkrétně od roku 2018. Do tohoto modelu vstupují jen ti

[45] Pokud respondent zvažuje volbu stran, je mu položena ještě uzavřená otázka „Vybral jste více stran, jejichž volbu byste v současnosti zvažoval. Můžete mi prosím říct, kterou z nich byste v současněji zvažoval(a) nejsilněji a je největší šance, že byste ji volil(a)?“. Seznam možností odpovědí pak odpovídá všem zvažovaným stranám. Tento detail není evidentní z dotazníku publikovaného v (MEDIAN 2018, 12) a dospěl jsem k němu až na základě explorace analyzovaného datového souboru.

respondenti, kteří **nevyloučili účast u voleb** v otázce *„Představte si, že by příští týden byly volby do Poslanecké sněmovny. Šel byste volit? Rozhodně ano, spíše ano, spíše ne, rozhodně ne, nebo nevíte?“* (CVVM 2018, 2) Pokud respondent odpoví „rozhodně ano“, je jeho odpovědím přiřazena pravděpodobnost 100 %. V případě odpovědí „spíše ano“ a „spíše ne“ je to pak 75 % a 25 % dle pořadí. Těmto respondentům pak tazatel položí otevřenou otázku *„Pokud byste k volbám šel, kterou stranu byste volil?"* (CVVM 2018, 2) Pokud nějakou politickou stranu v této otevřené otázce jmenuje, dostane ohledně této strany ještě otázku ohledně **jistoty své volby**: *„Nakolik jste si jistý, že byste volil právě tuto stranu? Zcela jistý, skoro jistý, nepříliš jistý, vůbec si nejste jistý.“* (CVVM 2018, 2) (Pokud je si zcela jistý, mají jeho odpovědi v tomto parametru pravděpodobnost 100 %. Jakákoliv jiná opověď znamená **ještě otevřenou otázku** *„Vzhledem k tomu, že si nejste svou volbou zcela jistý, jaké (další) strany byste zvažoval(a)?“* na níž může jako odpověď uvést až dvě strany. Pokud je si u první strany skoro jistý, dostane tato první strana pravděpodobnost 80 % a zbylých 20 % se dělí mezi další (až) dvě strany jmenované v druhé otázce. Pokud si je nepříliš jistý, přiřadí se první straně 50% pravděpodobnost, a pokud si není jistý vůbec, 33 %. Doplněk do 100 % je vždy rozdělen mezi ostatní strany. Pokud ale respondent v **první otázce žádnou konkrétní stranu neuvede**, dostane stále otevřenou otázku **na další zvažované strany**, ale kterákoliv strana, kterou jmenuje, dostane vždy přidělenou pravděpodobnost 33 %. (CVVM 2018, 3) Jednotlivé pravděpodobnosti se i v tomto modelu ve výsledku násobí. Novější model CVVM2018 je více komplexní než model CVVM2017, to ale samo o sobě neznamená, že se jedná o lepší nástroj. I proto zahrnuji do srovnání obě dostupné podoby volebního modelu Centra pro výzkum veřejného mínění.

Hlavní odlišností modelu CVVM2018 od mnou navrhovaného modelu a modelu MEDIAN je, že u respondentů, kteří si **nevybrali preferovanou stranu** (a tak každá jimi zvažovaná strana získala váhu 33 %), **nemusí součet vah všech zvažovaných stran být roven 1**. Zatímco celková váha respondenta je v modelu Median ovlivňována (kromě váhové proměnné, viz *4.1.1 Model MEDIAN*) jen deklarovaným záměrem účasti, v případě modelu CVVM může být celková váha **všech stran zvažovaných** jedním respondentem zredukována ještě i na základě míry jeho nejistoty[46].

[46] Jedná se rozhodně o legitimní postup, minimálně v případě modelu navrhovaného v této diplomové práci jsem se ho nicméně rozhodl nenásledovat – záměrem práce je nastavovat parametry volebního modelu na

### 4.1.4. Model KANTAR

Aby byly ve srovnání podobně početně zastoupeny jak modely složitější (MEDIAN a CVVM2018), tak modely méně komplexní, rozhodl jsem se kromě modelu CVVM2017 do srovnání výsledků zařadit i způsob konstrukce volebního modelu, který používá společnost TNS Kantar. Do něj vstupují pouze ti respondenti, kteří **přímo nevyloučili svou účast u voleb**, zvolili **preferovanou stranu**, a zároveň **nepovažují za pravděpodobné, že by volbu své preferované strany změnili**. (TNS Kantar 2019, 8) Zatímco proměnná deklarovaného záměru účasti je v datovém souboru omnibusu Median přítomná, otázku na pravděpodobnost změny volby v podobě odpovídající modelu Kantar[47] v něm dotazována není – omnibus Median sice pevnost přesvědčení dotazuje, není ale plně překódovatelná (viz např. *4.2.3 Jistota záměru volby zvažované strany*). Na základě konzultace s konzultantem této diplomové práce jako její náhradu používám omezení modelu na respondenty, kteří uvedli preferovanou zamýšlenou stanu (protože četnostmi nejlépe odpovídá otázce na pravděpodobnost změny názoru volebního modelu Kantar). (Prokop 2019) Data ve volebním modelu KANTAR jsou také **vážena**, a to, kromě socio-demografických charakteristik (pohlaví, věk, kraj, velikost místa bydliště, vzdělání), také podle minulého volebního chování. (TNS Kantar 2019, 3)

## 4.2. Konstrukce navrhovaného volebního modelu

Kromě parametrů výpočtu modelu a váživí proměnné, které vycházejí z podcílů analýzy popsaných v předchozích částech této sekce, a jejichž příslušné výpočty budou popsány dále – záměr volební účasti, první zvažovaná volená strana volební chování v minulých volbách do PSP – bude mnou navrhovaný model vycházet z modelu agentury Median navrženém v diplomové práci konzultanta (Prokop 2012). Shodné tedy bude vážení podle socio-demografických charakteristik a zohlednění dalších zvažovaných stran (viz *4.1.1 Model MEDIAN* výše).

---

základě sekundární analýzy dat, a pro volbu takového změny postupu od optimalizovaného modelu bych potřeboval provést ještě důkladnou analýzu toho, jaký je eventuální vztah mezi přiřazením jistoty záměru volby zamýšlené strany a deklarovaným záměrem volební účasti (a potenciálně i reportovanou účastí), od které momentálně odvozuji celkovou váhu odpovědí jednoho respondenta v celém modelu.

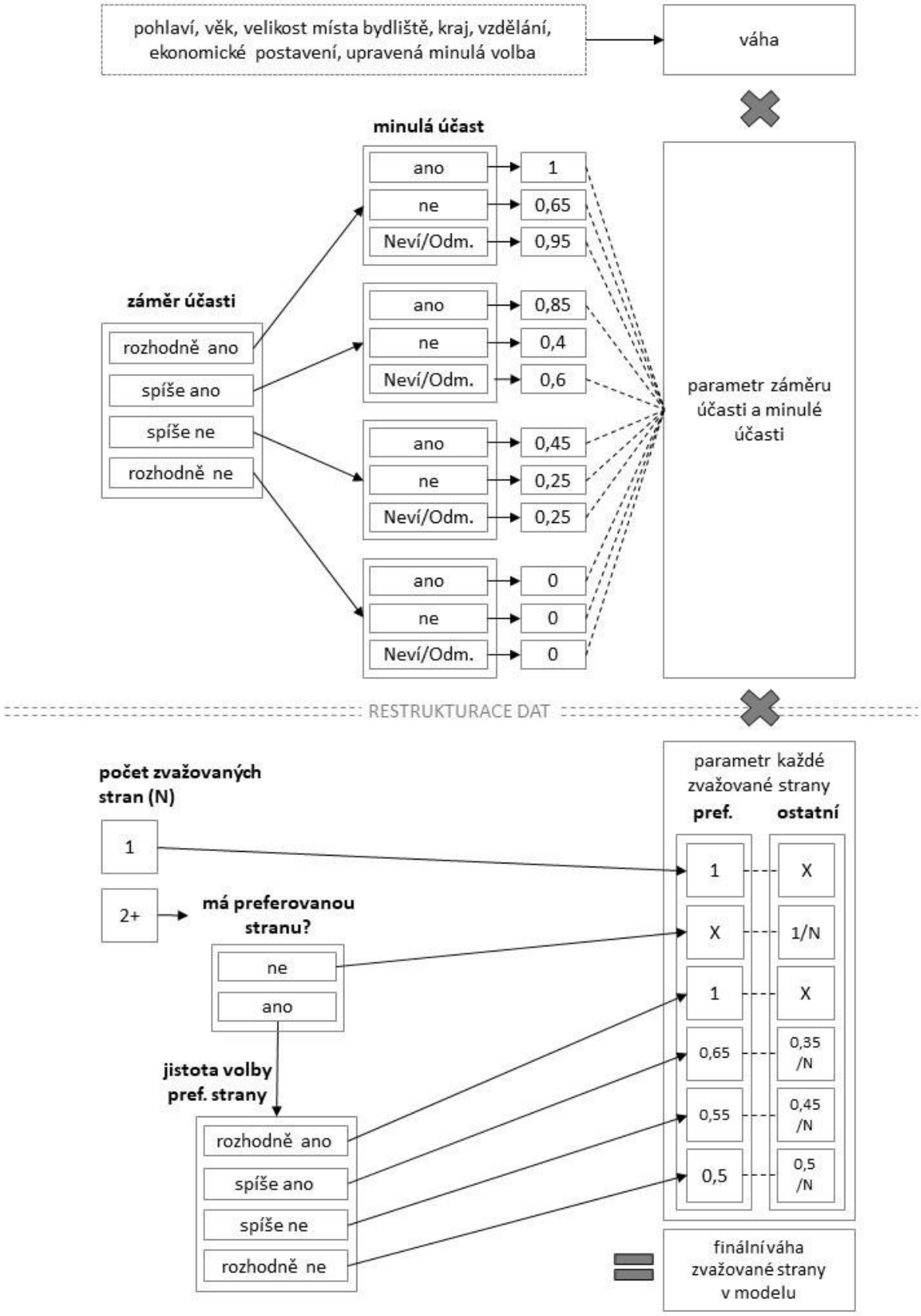
[47] Znění otázku z modelu Kantar: „*Je toto Vaše rozhodnutí volit stranu … již pevné, nebo se Vaše rozhodnutí může ještě změnit? (Jste pevně rozhodnut(a) volit tuto stranu nebo hnutí / Je pravděpodobné, že byste volil(a) tuto stranu nebo hnutí / Je možné, že Vaše rozhodnutí volit tuto stranu nebo hnutí se ještě změní / Je dost pravděpodobné, že vaše rozhodnutí volit tuto stranu nebo hnutí se ještě změní / Neví, bez odpovědi)*“ (TNS Kantar 2019, 19)

### 4.2.1. Jak v navrhovaném volebním modelu fungují parametry

Vzhledem k tomu, že v omnibusu Median mohou nerozhodnutí respondenti vybírat více zvažovaných stran, a všechny tyto zvažované strany budou u každého respondenta nějakou měrou vstupovat do navrhovaného volebního modelu, nestačí pro vytvoření kýženého výstupu pouze jedna frekvenční tabulka – před tímto krokem je nutné data nejdříve restrukturovat. Příslušný postup vychází z původního návrhu volebního modelu Median v (Prokop 2012). Protože každý respondent může vytvořit více než jednu dvojici respondent – zvažovaná strana, restrukturuji před vytvořením výstupu modelu datový soubor z „širokého“ na „dlouhý“ v softwaru SPSS pomocí procedury VARSTOCASES. (IBM Corp. 2014, 110) Výstupem této transformace je datový soubor, kde jeden řádek (případ) nereprezentuje jednoho konkrétního respondenta, ale jednu každou politickou stranu zvažovanou jedním respondentem (označenou případnou nejvyšší preferencí). V datech, z nichž počítáme výsledný navrhovaný model, zastupuje tedy každého respondenta 1 nebo více řádků, kde se počet těchto řádků u každého respondenta rovná počtu jím zvažovaných stran. Každá zvažovaná strana každého respondenta tak ve volebním modelu má:

1) Socio-demografickou váhu zohledňující minulé volební chování respondenta
2) Parametr (záměru i minulé) volební účasti respondenta
3) Parametr jistoty preference dané zvažované strany

V každém řádku se tato socio-demografická váha vynásobí s oběma parametry, a tím vznikne finální váha modelu pro danou zvažovanou politickou stranu (daný řádek). Tento postup je co do datové transformace shodný se způsobem výpočtu volebního modelu Median navrženého v (Prokop 2012), **způsob výpočtu hodnoty výše zmíněné váhy a parametrů** je ale tím, co v této práci navrhovaný způsob výpočtu odlišuje, a jedná se o součást vlastního přínosu této diplomové práce. V následujících odstavcích postupně popíšu výpočet hodnot obou parametrů a následně způsob zohlednění míry vlivu uplynulého času v proměnné minulého volebního chování. Popis jednotlivých charakteristik modelu je poměrně technický, a proto pro názornost přikládám zároveň i diagram popisující konstrukci modelu (*Obrázek 1*).




*Obrázek 1: Diagram konstrukce navrhovaného volebního modelu*

### 4.2.2. Účast

Sekundární analýza panelových dat CHPS ukázala, že pravděpodobnost skutečné účasti u voleb závisí na kombinaci deklarované účasti a minulé volební účasti. Pro stanovení hodnoty parametru záměru volební účasti v kombinaci s minulou volební účastí využívám trojitou kontingenční tabulku se shodným výběrem proměnných, ale už vytvořenou pomocí dat po imputaci chybějících hodnot. Kombinací proměnné deklarovaného záměru volební účasti[48] a proměnné minulé volební účasti rozděluji respondenty na 8 skupin. Hodnotami parametru záměru volební účasti v kombinaci s minulou volební účastí pak jsou relativní četnosti respondentů, které v těch samých 8 skupinách v imputovaných datech CHPS reportovali volební účast ve volbách do PSP ČR v roce 2017, viz *Tabulka 10*. Vzhledem k tomu, že vztah s reportovanou ve skupině respondentů, kteří **neví, nebo odmítli odpovědět** ohledně své volební účasti v roce 2013, se nepodařilo prokázat, jsou hodnoty parametru u této skupiny (4 poslední řádky tabulky) stanoveny na základě celkových relativních četností v imputovaných datech CHPS. Hodnoty byly zaokrouhleny na nejbližší násobek 0,05. U respondentů, kteří vyloučili účast u hypotetických voleb, byla hodnota parametru automaticky stanovena rovná nule, protože už tito respondenti nebyli v dotazníku Omnibus Median dále dotazováni.

| Záměr účasti 2017 | Účast 2013 | Parametr účasti a minulé volby |
|---|---|---|
| Rozhodně ano | Ano | 1 |
| | Ne | 0,65 |
| | Neví/Odmítl | 0,95 |
| Spíše ano | Ano | 0,85 |
| | Ne | 0,4 |
| | Neví/Odmítl | 0,6 |
| Spíše ne | Ano | 0,45 |
| | Ne | 0,25 |
| | Neví/Odmítl | 0,25 |
| Rozhodně ne | Ano | 0 |
| | Ne | 0 |
| | Neví/Odmítl | 0 |

*Tabulka 10: Hodnoty parametru záměru účasti a minulé účasti v navrhovaném volebním modelu*

[48]V datovém souboru CHPS, který jsem používal pro sekundární analýzu dat v první polovině této diplomové práce, byly u otázky záměru volební účasti u případných voleb i tři neproduktivní možnosti odpovědí: *Nemám volební právo / Nevím, nemohu posoudit / Odmítl(a)*. Protože v omnibusu Median tyto tři možnosti u záměru volební účasti neexistují, do následující trojité kontingenční tabulky nejsou tyto skupiny respondentů zahrnuty.

### 4.2.3. Jistota záměru volby zvažované strany

Druhým parametrem, jehož hodnoty je potřeba na základě poznatků ze sekundární analýzy dat z CHPS stanovit, je parametr jistoty volby pro preferovanou politickou stranu respondenta. Nastavení hodnoty parametru preferované strany zároveň nastavuje i hodnoty parametru u všech ostatních zvažovaných stran respondenta –doplněk parametru preferované strany do 1 dělený počtem ostatních zvažovaných stran. Sekundární analýza dat CHPS ukázala, že pouze 65 % (po zaokrouhlení na násobek 0,05) respondentů volí stejnou stranu, kterou deklarovali záměr volit 1 až 2 měsíce před volbami. Nebudu zohledňovat fakt, že u každé stran je tato pravděpodobnost uchování preference do voleb mírně odlišná, protože tím bych nevytvářel obecný postup, ale model příliš těsně navázaný na volby do PSP ČR v roce 2017. Jak tedy hodnotu 0,65 při konstrukci modelu zohledním?

Součet hodnot parametru jistoty volby zvažovaných stran má v navrhovaném modelu u každého respondenta být roven 1. Pokud respondent neuvedl preferovanou stranu, je nastavení hodnot parametru jednoduché – každá zvažovaná strana dostane přiřazenou poměrnou část, tedy hodnotu rovnou 1/počtem zvažovaných stran. Pokud ale preferovanou stanu uvede, ta by měla mít (z titulu preference) logicky vyšší váhu než ostatní zvažované strany daného respondenta. Vzhledem k tomu, že počet zvažovaných stran může být u každého respondenta jiný, dosáhneme splnění obou vše popsaných podmínek tím, že nejmenší možnou hodnotu parametru u preferované strany nastavíme rovnou 0,5. Zároveň musíme zohlednit, že filtrovací otázky v dotazníku omnibusu Median jsou nastaveny tak, že pokud u preferované strany respondent uvede, že si je její volbou „Zcela jistý“, na další zvažované strany už respondent není dotazován. Taková preferovaná strana musí vzhledem k podmínkám výše dostat přiřazenou hodnotu 1.

Manévrovací prostor, který mi v případě parametru jistoty záměru volby zvažované strany pro nastavení hodnot na základě výsledků sekundární analýzy zbyl, není velký – týká se pouze respondentů, kteří u preferované strany uvedly, že jsou si její volbou „Skoro jistí“, nebo „Nepříliš jistí“. Proto jsem hodnotu tohoto parametru u skoro jistých respondentů nastavil na 0,65 (hodnotu z výsledků sekundární analýzy dat), a hodnotu pro nepříliš jisté respondenty na polovinu 0,55, abych dodržel zmenšující se vzdálenost směrem ke spodní hranici hodnot parametru u respondentů s preferovanou stranou nastavenou rozdílem mezi

první a druhou hodnotou parametru. Pro větší přehlednost shrnuji nastavení hodnot parametru jistoty záměru volby zvažované strany také v tabulce níže (*Tabulka 11*).

| Jistota záměru volby preferované strany | Parametr jistoty volby zvažované strany | | |
|---|---|---|---|
| | Uvedl preferovanou stranu | | Neuvedl preferovanou stranu |
| | Preferovaná strana | Ostatní strany | Všechny strany |
| Zcela jistý(á) | 1 | N.A. | 1/počet zvažovaných stran |
| Skoro jistý(á) | 0,65 | 0,35/ počet ostatních stran | |
| Nepříliš jistý(á) | 0,55 | 0,45/ počet ostatních stran | |
| Vůbec si není jistý(á) | 0,5 | 0,5/ počet ostatních stran | |

*Tabulka 11: Parametr jistoty volby zvažované strany v navrhovaném volebním modelu*

### 4.2.4. Minule volená strana a vliv času uplynulého od minulých voleb

Poslední důležitou složkou výpočtu navrhovaného volebního modelu je zohlednění vlivu času od minulých voleb v teoretických četnostech proměnné minulého volebního chování (dále minulé volby). U té tedy nepotřebuji stanovit hodnoty parametru, ale upravit na základě poznatků ze sekundární analýzy dat z omnibusu Median její upravené teoretické četnosti pro vážení. Ze srovnávaných modelů modelu MEDIAN používá vážení podle minulého volebního chování ještě model KANTAR, ty nicméně vliv času uplynulého od minulých voleb u jednotlivých stran nezohledňují.

Já tyto teoretické četnosti pro vážení upravuji podle toho, jaký v sekundární analýze dat z omnibusu Median u jednotlivých politických stran vyšel podíl vlivu zapomínání a vlivu výběru respondentů na odchylku od výsledku minulých voleb. V sekundární analýze jsem u každé politické strany získal údaj indikativní tomu, jak velkou část odchylky a kterým směrem u ní tvoří vliv času uplynulého od minulých voleb. Každá z těchto odchylek má velikost (počet procentních bodů) a směr (buďto kladný nebo záporný vzhledem ke skutečným výsledkům voleb). Pro úpravu teoretických četností pro konstrukci navrhovaného volebního modelu potřebuji stanovit, jaký podíl odchylky pozorované v roce 2017 **nebyl** obsažený již odchylce pozorované v roce 2013 – nyní popíšu, jak jsem k tomuto údaji u jednotlivých stran došel.

Pokud měly odchylky v roce 2013 a v roce 2017 stejný směr (buďto obě záporný, nebo obě kladný), a odchylka z roku 2017 z nich byla ta menší, pak byla celá obsažená už v odchylce pozorované v roce 2013 a její podíl vzniklý uplynutím času od voleb je tedy

nulový. Pokud měly odchylky stejný směr, ale odchylka v roce 2017 z nich byla větší, je tímto vzniklým podílem rozdíl obou odchylek. Pokud mají odchylky různý směr, počítá se celá odchylka z roku 2017 (protože se vyvinula novým směrem). Tímto postupem u každé strany získám hodnotu (v procentních bodech), kterou nazývám „nová odchylka“. Právě součtem nové odchylky a původních teoretických četností minulé volby vznikají upravené teoretické četnosti minulé volby. Ty dopočítám do 100 % podle účasti u voleb do PSP ČR v roce 2013 a dalších neproduktivních hodnot, čímž získám finální hodnoty teoretických četností pro výpočet předrestukturační váhové proměnné v navrhovaném volebním modelu. Pro přehlednost uvádím hodnoty u jednotlivých stran dle postupu výše také v tabulce[49] (*Tabulka 12*).

| | Volby 2013 | Odchylka 2013 (p.b.) | Odchylka 2017 (p.b.) | Nová odchylka (p.b.) | Upravené T. Č. | Finální T. Č. |
|---|---|---|---|---|---|---|
| ČSSD | 20,46 % | 6,16 | 4,82 | 0,00 | 20,46 % | 10,93 % |
| ANO 2011 | 18,66 % | 3,20 | 4,66 | 1,45 | 20,11 % | 10,74 % |
| KSČM | 14,91 % | -0,92 | -6,55 | -5,63 | 9,28 % | 4,96 % |
| TOP 09 | 12,00 % | 0,10 | 0,88 | 0,78 | 12,78 % | 6,83 % |
| ODS | 7,73 % | 0,18 | 3,50 | 3,32 | 11,05 % | 5,90 % |
| Úsvit | 6,89 % | -3,06 | -3,09 | -0,03 | 6,86 % | 3,67 % |
| KDU-ČSL | 6,78 % | -0,97 | 0,01 | 0,01 | 6,79 % | 3,63 % |
| Strana zelených | 3,20 % | -0,78 | -2,31 | -1,53 | 1,67 % | 0,89 % |
| Česká pirátská strana | 2,66 % | -0,50 | 1,03 | 1,03 | 3,69 % | 1,97 % |
| Svobodní | 2,47 % | -1,29 | -0,92 | 0,00 | 2,47 % | 1,32 % |
| jiná strana | 4,22 % | -2,13 | -2,03 | 0,00 | 4,22 % | 2,26 % |
| Už si nepamatuje | | | | | | 4,73 % |
| Neúčast | | | | | | 36,38 % |
| Voleb se nemohl(a) zúčastnit kvůli věku | | | | | | 5,48 % |

*Tabulka 12: Úprava teoretických četností vážící proměnné minulé volby*

[49] Osobně považuji tento parametr navrhovaného modelu za jeho nejslabší část, respektive nejméně přesvědčivě argumentovanou. Nikoliv ale proto, že by mi tento postup přišel konceptuálně neospravedlnitelný (naopak), ale zejména proto, že nemám momentálně k dispozici větší datové soubory pro sekundární analýzu dat (například ze všech vln omnibusu Median mezi lety 2013 a 2017), abych mohl vliv uplynulého času posoudit detailněji. V současné situaci totiž mnoho výše popsaných odchylek spadá vzhledem k velikosti výběrového souboru do tolerance statistické chyby. Přesto jsem se rozhodl v upravených teoretických četnostech zohlednit všechny rozdíly, které splnily výše popsané podmínky, bez ohledu na jejich velikost. Protože minulá volba je pouze jednou z proměnných používaných při výpočtu předrestrukturační váhové proměnné, jsou rozdíly vytvořené takto malými úpravami teoretických četností spíše zanedbatelné. Tím pádem jsou větší rozdíly u jednotlivých stran zohledněné (ODS, KSČM) s většími rozdíly, zatímco ty menší se výrazně neprojeví – tento postup považuji za čistší (respektive minimálně stejně tak oprávněný), jako kdybych nastavil arbitrární cut-off stran, u kterých budeme rozdíl brát v potaz, dle velikosti odchylky.

## 4.3. Kritéria srovnávání volebních modelů

Než přistoupím k samotnému srovnání výsledků jednotlivých volebních modelů, je potřeba nejdříve stanovit a popsat kritéria, jejichž pomocí modely srovnávám. Jedná se o šest statistik, jejichž výběr vychází z vlastní diplomové práce konzultanta (Prokop 2012), jejímž výstupem byl mimo jiné způsob konstrukce volebního modelu Median. Vzhledem k tomu, že v této práci navrhuji způsob optimalizace daného modelu, považuji za vhodné použít k posouzení této optimalizace ta samá kritéria, která byla původně použita při představení původního postupu. Z původní konzultantovy práce přebírám jak výčet kritérií, tak odůvodnění jejich použití.

Jsou jimi **suma čtverců odchylek**, která se standardně používá pro srovnávání očekávaných a pozorovaných hodnot. Následující čtyři charakteristiky se v párech věnují vždy **maximální** a **průměrné** odchylce každého modelu, jednou v **absolutních** číslech, podruhé v **relativních** (procentech odchylky od kýžené hodnoty). Maximální odchylka je důležitá vlastnost modelu, protože vysoká maximální odchylka může podstatně ovlivnit jeho kredibilitu v očích veřejnosti. Průměrná odchylka, je podobně jako suma čtverců výše, indikativní vlastnostem modelu jako celku. Zařazení relativních odchylek nám umožňuje zhodnotit lépe odchylky například u menších stran, kde je odchylka v absolutních číslech sice malá, ale může se jednat o velikost rovnou podstatné části zisku celé strany. Poslední srovnávanou statistikou je **počet chybných kvalifikací do PSP ČR**, který může opět mít podstatný vliv na kredibilitu výsledků modelu v očích veřejnosti – Zatímco u stran na prvních místech v zisku hlasů není odchylka v řádu jednotek procent tak kritickým pochybením, u stran, které se pohybují kolem pětiprocentní hranice volební klauzule, může taková odchylka v modelu mít po zveřejnění závažné důsledky – způsobit spekulace (jedním či druhým směrem, zda se daná strana „dostane do sněmovny“, atd. (Hrbková and Chytilek 2013)

Jedinou statistikou srovnávanou v (Prokop 2012), kterou nesrovnávám ve své diplomové práci, je **absolutní odchylka v rozdílu mezi první a druhou stranou**. Tu bylo smysluplné sledovat v době, kdy první dvě místa ve výsledcích voleb patřila tradičně dvou největším stranám, typicky ODS a ČSSD (viz např. Český Statistický Úřad 2010). Nicméně od voleb do PSP ČR v roce 2013 se tato situace změnila, druhou nejsilnější stranou po ČSSD se stalo ANO 2011 (Český Statistický Úřad 2013), které se pak ve volbách roku 2017 stalo s náskokem nejsilněji zastoupenou stranou v PSP ČR, zatímco ČSSD se propadla na šesté

místo. (Český Statistický Úřad 2017) V situaci, kdy je v PSP ČR jeden silný lídr (hnutí ANO 2011) a souboje mezi ostatními stranami se vedou spíše o druhé a třetí místo, jsou výsledky takové statistky pro srovnávání volebních modelů zavádějící a tvrdím, že je vhodné ji ze srovnání vyřadit.

### 4.4. Posouzení navrhovaného volebního modelu (výsledky)

Přestože srovnání jednotlivých modelů provádím zejména prostřednictvím výše popsaných statistik, některé jejich vlastnosti jsou evidentní už za samotné tabulky jejich výsledků. (*Tabulka 13*) Ve všech modelech je podstatně nadhodnocena ČSSD, která je ve všech modelech nejhůře predikovanou stranou (podobnost napříč modely může indikovat, že se jedná o vlastnost datového souboru omnibusu Median, spíše než všech srovnávaných modelů zároveň – podobně mám tendenci interpretovat i konzistentní podhodnocení položky „Ostatní strany"), druhou konzistentně nadhodnocovanou stranou je TOP 09. Naopak obecně podhodnocovanými stranami jsou Piráti a SPD (které se v těchto volbách do PSP dostaly poprvé, v případě SPD se zároveň rovnalo o první volby do PSP ČR od založení subjektu). Podstatně podhodnoceny byly i strany ANO 2011 (kromě modelu CVVM2017 – to může být spojené se silným zastoupením voličů ANO 2011 mezi respondenty se silným záměrem účasti u voleb) a STAN (kromě modelu Median – potenciálně naopak díky většímu zohlednění nerozhodnutých voličů), u kterého se může jednat o vliv pozdějšího nástupu do volební kampaně.

Další informace o srovnávaných modelech získáme z tabulky jejich statistik. (*Tabulka 14*) Ve všech modelech byla přítomná jedna strana, u které nesouhlasila predikce kvalifikace do PSP, a ve všech případech to byla shodně strana STAN (která byla ve všech modelech podhodnocená, zároveň se jednalo o novou stranu, která před volbami navíc opustila koalici s KDU-ČSL).

Pokud seřadím všechny srovnávané modely v každé ze sledovaných statistiky vzestupně (dle odchylek), a pak z těchto pořadí vypočítám aritmetický průměr, získám tím jednoduchý index indikativní tomu, jak dobře si ve sledovaných kritériích daný model stojí oproti ostatním srovnávaným modelům. Hodnotu tohoto indexu prezentuji v tabulce. (*Tabulka 15*) Statistiku počtu chybných predikcí kvalifikace do PSP vzhledem ke shodným výsledkům zanedbávám.

V indexu si navrhovaný volební model vede velmi dobře – na základě průměrného pořadí těchto kritérií si v rámci srovnávaných volebních modelů navrhovaný model stojí nejlépe s průměrným pořadím 1,4. Tento výsledek je přinejmenším indikativní tomu, že volební model naplňuje cíle, které byly v této diplomové práci stanoveny pro jeho konstrukci, a že má cenu model do budoucna dále testovat jako potenciální nový způsob konstrukce volebního modelu Median.

| | VOLBY 2017 | Návrh | odch. (p.b.) | MEDIAN | odch. (p.b.) | CVVM 2017 | odch. (p.b.) | CVVM 2018 | odch. (p.b.) | KANTAR | odch. (p.b.) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| ANO 2011 | 29,64 % | 27,12 % | -2,52 | 27,08 % | -2,56 | 30,06 % | 0,42 | 29,10 % | -0,54 | 27,74 % | -1,90 |
| ODS | 11,32 % | 11,00 % | -0,32 | 9,78 % | -1,54 | 10,68 % | -0,64 | 10,76 % | -0,56 | 10,30 % | -1,02 |
| Piráti | 10,79 % | 8,70 % | -2,09 | 8,51 % | -2,28 | 9,07 % | -1,72 | 8,60 % | -2,19 | 8,06 % | -2,73 |
| SPD | 10,64 % | 9,30 % | -1,34 | 9,23 % | -1,41 | 7,75 % | -2,89 | 7,81 % | -2,83 | 9,54 % | -1,10 |
| KSČM | 7,76 % | 8,15 % | 0,39 | 9,49 % | 1,73 | 7,33 % | -0,43 | 7,95 % | 0,19 | 10,55 % | 2,79 |
| ČSSD | 7,27 % | 10,44 % | 3,17 | 10,24 % | 2,97 | 12,79 % | 5,52 | 12,76 % | 5,49 | 11,38 % | 4,11 |
| KDU-ČSL | 5,80 % | 6,57 % | 0,77 | 6,29 % | 0,49 | 6,43 % | 0,63 | 6,45 % | 0,65 | 6,70 % | 0,90 |
| TOP 09 | 5,32 % | 7,91 % | 2,59 | 6,73 % | 1,41 | 6,96 % | 1,64 | 7,13 % | 1,81 | 6,61 % | 1,29 |
| STAN | 5,18 % | 3,78 % | -1,40 | 4,76 % | -0,42 | 3,02 % | -2,16 | 3,78 % | -1,40 | 3,51 % | -1,67 |
| Svobodní | 1,56 % | 1,38 % | -0,18 | 1,22 % | -0,34 | 0,99 % | -0,57 | 0,98 % | -0,58 | 0,97 % | -0,59 |
| Zelení | 1,46 % | 1,88 % | 0,42 | 2,38 % | 0,92 | 2,06 % | 0,60 | 1,80 % | 0,34 | 1,73 % | 0,27 |
| Ostatní | 6,28 % | 3,77 % | -2,51 | 4,29 % | -1,99 | 2,85 % | -3,43 | 3,86 % | -2,42 | 2,90 % | -3,38 |

*Tabulka 13: Srovnání výstupu navrhovaného volebního modelu dle výsledků*

| HODNOTICÍ KRITÉRIUM | NÁVRH | MEDIAN | CVVM 2017 | CVVM 2018 | KANTAR |
|---|---|---|---|---|---|
| Suma čtverců odchylek | 38,63 | 35,30 | 62,78 | 55,57 | 55,05 |
| Průměrná absoutní odchylka u strany | 1,48 | 1,51 | 1,72 | 1,58 | 1,81 |
| Maximální absolutní odchylka u strany | 3,17 | 2,97 | 5,52 | 5,49 | 4,11 |
| Průměrná relativní odchylka u strany | 22 % | 23 % | 29 % | 25 % | 27 % |
| Maximální relativní odchylka u strany | 49 % | 63 % | 76 % | 76 % | 56 % |
| Počet chybných predikcí kvalifikace do PSP | 1 | 1 | 1 | 1 | 1 |

*Tabulka 14: Srovnání navrhovaného volebního modelu dle hodnoticích kritérií*

| PRŮMĚRNÉ POŘADÍ | |
|---|---|
| NÁVRH | 1,40 |
| MEDIAN | 1,80 |
| CVVM2017 | 4,80 |
| CVVM2018 | 3,60 |
| KANTAR | 3,40 |

*Tabulka 15: Průměrná hodnota indexu pořadí srovnávaných volebních modelů*

# 5. Diskuse navrhovaného modelu

Jak na druhém, tak na prvním místě se v indexu umístily mnou navrhovaný volební model a volební model společnosti Median. Takový výsledek může být problematický – navrhovaný model sdílí většinu konceptuálních charakteristik s modelem Median, na kterém je založen, a ten byl navíc rovněž navržen mimo jiné na základě analýzy dat ze sběrů společnosti Median. Pokud se tedy v datech skrývá pravidelnost, která není specifická modelům, ale datům agentury Median, nejsem schopný to odlišit. Nicméně, parametry navrhovaného modelu byly navrženy na základě analýzy dat CHPS, která nebyla sebrána za tímto účelem, což opět mluví ve prospěch navrženého modelu. Kdybychom přeci jen chtěli ověřit, že navrhovaný model projevuje dobré vlastnosti i při aplikaci na datových souborech pocházejících z jiných sběrů, bylo by vhodné provést podobnou komparaci jako v (Kreidl 2005) a následně porovnat výstupy z jednotlivých datových souborů. Pokud bychom ale přistoupili na závěry této 14 let staré studie, můžeme předpokládat, že rozdíl vytvořený volbou jiné společnosti pro sběr dat by neměl zásadní vliv, a že tato vysoká umístění můžeme přičítat skutečně vlastnostem obou volebních modelů.

Oproti volebnímu modelu MEDIAN jsou ale evidentní i některé rozdíly. Metodika navrhovaného modelu má například oproti modelu MEDIAN přes velkou míru shody v metodice obou modelů podstatně menší odchylku oproti výsledkům voleb u ODS a KSČM (která je zároveň v absolutních číslech v rámci obou modelů patří mezi nejmenší s hodnotami -0,32 a 0,39 dle pořadí). Tento rozdíl nevznikl náhodou – jedná se o strany, u kterých v navrhovaném modelu došlo k největší úpravě teoretických četností minulé volby na základě analýzy vlivu času uplynulého od voleb, a tím pádem nebylo jejich zastoupení v modelu ovlivněno vážením (protože vyšlo najevo, že se nejedná o chybu způsobenou výběrem).

Druhou zajímavou a potenciálně pozitivní (pro používání navrhovaného modelu a jemu podobných modelů) informací vycházející z výše srovnávaných statistik je, že minimálně v kontextu sběru dat několik týdnů před volbami si více komplexní modely (které berou v úvahu nerozhodnuté voliče a jejich preference) vedou lépe než méně komplexní volební modely. V obou verzích indexu výše se modely se model CVVM2017 (od jehož používání už Centrum pro výzkum veřejného mínění upustilo) umístil na poslední pozici s tím. že model CVVM2017 měl v obou případech více než o jedno místo horší průměrné postavení v žebříčku oproti modelu CVVM2018 nebo KANTAR. Tento rozdíl u modelu CVVM2017 může být způsoben vyřazením respondentů, kteří uvádějí, že by se případných voleb spíše nezúčastnili[50]. Model KANTAR se umístil na třetím místě (před modelem CVVM2018). Přestože je model KANTAR méně komplexní kvůli nezohledňováním nerozhodnutých voličů, oproti modelu CVVM2018 nicméně váží data podle minulého volebního chování – to model CVVM2018 nezohledňuje, používá ale parametr záměru účasti a jistoty záměru volby strany u nerozhodnutých voličů. Volební modely (navrhovaný a MEDIAN), které zohledňují všechny výše zmíněné faktory zároveň, nicméně ve stanovených kritériích vycházely nejlépe.

Kromě silných stránek navrhovaného modelu (i ve srovnání s modelem MEDIAN) nám výše popsaná sekundární analýza v kombinaci se srovnáním ostatních modelů přinesly i některé další poznatky o modelech ostatních agentur. Volební model agentury Kantar předpokládá, že voliči, kteří nepovažují za pravděpodobné změnu preferované strany, ji budou volt se 100% pravděpodobností – to je, minimálně podle výsledků sekundární analýzy dat v této práci, nadhodnocený předpoklad.

V sekundární analýze dat se také povedlo ověřit, že záměr účasti nelze (vzhledem k reálné účasti u následných voleb) vnímat pouze jako binární indikátor – tak s ním nicméně nakládají například model KANTAR a starší volební model CVVM2017, které ho používají de facto pouze jako cut-off kritérium pro vstup respondentů do modelu. Novější volební model CVVM2018 a model MEDIAN, už zohledňují záměr volební účasti jako parametr váhy odpovědí jednotlivých respondentů ve výsledném modelu, ale, jak se povedlo znázornit ve výsledcích sekundární analýzy dat této práce, vztah záměru účasti

[50] Podobně by mohly i ostatní modely získat dotazováním zvažovaných volených stran u respondentů, kteří deklarují, že by se případných voleb rozhodně nezúčastnili, viz *3.3.1 Volební účast, záměr volební účasti a minulá volební* účast.

u voleb a reálné volební účasti je zároveň ovlivněn také účastí respondenta u minulých voleb.

Možným zdrojem kritiky výše prezentovaných postupů[51] je, že fakt, že takto konstruovaný model (na datech z těsně předvolebního období) je vhodné používat při výzkumu právě v době těsně před volbami, nemusí znamenat, že bude tím správným nástrojem pro období dejme tomu dva a půl roku před volbami. S takovou připomínkou spíše souhlasím, a právě vzhledem k tomu, že bezprostředně před volbami se děje mnoho událostí, které mají ovlivnit rozhodování voličů, si možná dokonce myslím, že může být vhodné používat různě konstruované modely v různých kontextech – zde vytvořený model byl navržen podle toho, jak se chová na datech bezprostředně před volbami, a byl navržen právě pro tento kontext, jinak konstruovaný model by určitě mohl být vhodnější pro jiný kontext. Na výzkumníkovi je potom volba mezi přímou porovnatelností výstupů vytvořených na základě shodné metodiky v časové řadě, a volebním modelem, u kterého v dané době předpokládá lepší zachycení rozložení situace ve společnosti.

# 6. Závěry

V práci se mi podařilo postupně ověřit, a na základě stanovených kritérií potvrdit, všechny vytyčené hypotézy. Výsledky této konfirmační analýzy jsem následně využil pro nastavení konkrétního způsobu konstrukce volebního modelu. Výstupy tohoto navrhovaného volebního modelu jsem porovnal s výstupy dalších způsobů konstrukce volebních modelů. Tím se mi podařilo splnit hlavní cíl této diplomové práce, stanovený v její úvodní kapitole. Ve zvolených hodnoticích kritériích vykazuje navrhovaný model lepší hodnoty než ostatní srovnávané modely, a to včetně volebního modelu agentury Median. V tomto smyslu se tedy skutečně jedná o jeho optimalizaci.

Konkrétním a dobře viditelným projevem jedné ze silných stránek navrhovaného modelu byl fakt, že strany, u kterých došlo na základě výsledků sekundární analýzy dat z vln omnibusu Median XI/XII 2013 a IX/X 2017 (konkrétně ODS a KSČM) vykazovaly ve výsledném srovnání volebních modelů minimální absolutní odchylky od výsledků voleb do PSP ČR v roce 2017. Tento výsledek považuji za indikativní vhodnosti postupu úpravy

[51] Dalšími legitimními kritikami postupů během dosahování cílů této práce jsou kromě srovnávání výsledků volebního modelu s výsledky voleb i volba konkrétních proměnných a nejsilnější pozice parametru záměru účasti a minulé volby ze všech parametrů modelu. Ty jsem nicméně už komentoval v příslušných kapitolách této práce, a zde je pouze zmiňuji pro větší míru úplnosti výčtu diskutabilních aspektů.

teoretických četností váhové proměnné minulé volby, a další analytické práce v této oblasti (viz dále).

Vzhledem k tomu, že zadání této práce vzešlo původně od autora v ní optimalizovaného modelu, a že se podařilo optimalizovaný způsob konstrukce volebního modelu navrhnout, považuji za vhodné se v závěru práce zamyslet, jak by s výše prezentovanými poznatky mohla agentura Median dále pracovat. Nemám tím samozřejmě na mysli, že by agentura Median návrh volebního modelu rovnou přejala a začala je publikovat.

Za oprávněný další postup nicméně požaduji provést ještě několik dalších srovnání návrhu se stávajícím modelem Medianu. Ideálním začátkem těchto srovnání by bylo využít faktu, že aplikace zde prezentovaného návrhu nevyžaduje na straně Medianu žádné změny v dotazníku. Tím pádem bude naprosto proveditelné přepočítat současné časové řady volebního modelu Medianu na výše navržený model, a následně porovnat, jaký měl tento přepočet na tyto časové řady vliv.

Dalším možným krokem by bylo v nově vznikajících volebních modelech Medianu (v budoucích vlnách omnibusu Median) vypočítávat po určitou dobu kromě současného volebního modelu i zde navrhovaný model, a použit jeho výsledky pro další interní srovnání obou postupů. Pokud by v těchto srovnáních navrhovaný model vycházel oproti současnému volebnímu modelu konzistentně lépe, jednalo by se podle mého názoru o dostatečně silný argument začít zvažovat zařazení zde navrhovaného volebního modelu coby stálého hlavního postupu výpočtu volebního modelu agentury. Navrhovaný způsob výpočtu stejně tak nemusí být přejímán jako celek, využity mohou být i jenom některé jeho parametry.

Jednou ze změn v metodice sběru dat jejich omnibusu, kterou bych agentuře Median rád navrhl, je odstranit z dotazníku filtr, který vyřazuje respondenty, kteří uvedou, že by se případných voleb rozhodně nezúčastnili, z dalšího dotazování jejich preferovaných stran. Ze sekundární analýzy vyplynulo, že podstatná část těchto respondentů se reálných voleb přeci jen zúčastní, a tím pádem by podle mého názoru bylo vhodné je z volebního modelu nevyřazovat, ale přistupovat k nim pouze jako k nerozhodnutým respondentům (s pravděpodobně specifickými hodnotami ostatních parametrů modelu). Takováto změna by pravděpodobně vyžadovala kromě změny filtru i změnu konkrétní formulace otázky na

volební preference pro tyto respondenty – takovým způsobem, aby zohledňovala jejich předchozí vyloučení účasti v případných volbách[52].

Vliv takové změny na model by bylo nicméně možné posoudit až po jejím zavedení, a vyžadovala by další analytickou práci. Tento závěr sekundární analýzy (respektive doporučení) nicméně považuji za vhodný podnět ke zvážení i pro jiné agentury (kromě Medianu), které v současné době filtrují respondenty vylučující účast u případných voleb z volebního modelu, nebo z dalšího dotazování volebních preferencí[53]. Kromě výše jmenované změny bych agentuře Median na základě rešerše literatury rád doporučil i zařazení otázky ohledně zájmu respondenta o politiku, která by mohla umožnit další podstatnou optimalizaci některých parametrů volebního modelu, konkrétně zejména volební účasti. (srov. s Tabery and Šamanová 2012)

Na závěr své diplomové práce bych rád ještě jmenoval dva možné směry pro další výzkum a analýzy, které podle mého názoru z poznatků v této práci vyplývají. Prvním je možný vztah misreportingu v proměnné minulé volby a vývoji rozložení volebních preferencích od doby minulých voleb do období sběru dat. V této práci se v tomto ohledu podařilo získat pouze dílčí poznatky, které tomuto nicméně nasvědčují, a proto by bylo vhodné se jim věnovat detailněji. Tato teoretická perspektiva k problému misreportingu by zároveň byla konzistentní s vlivem teoretického konceptu kognitivní disonance (Waldahl and Aardal 1982), a mohla by rovněž přispět k další optimalizaci vážení dat ve volebním modelu dle minulého volebního chování.

Za druhé vhodné směřování dalších výzkumů a analýz považuji podrobnější prozkoumání vztahu mezi proměnnou jistoty záměru volby preferované politické strany a záměrem účasti, podobně jako se jim věnuje např. (Linek 2018, 11), ale se zapojením pokročilejší mnohorozměrové analýzy v kombinaci s dalšími proměnnými, například s minulou volební účastí (Denny and Doyle 2009) nebo konkrétní preferovanou stranou, případně skupin stran podle vhodné typologie[54]. Výsledkem takové analýzy by potenciálně mohl být volební model, který zároveň dokáže nerozhodnuté respondenty klasifikovat ještě jemněji,

[52] Příkladem může být otázka, kterou v tomto kontextu používá výzkumná agentura STEM: *„Pokud byste (přesto) šel(šla) volit, kterou stranu byste příští týden volil(a) do Poslanecké sněmovny Parlamentu ČR?“* (STEM 2019)

[53] Podobné tvrzení můžeme udělat i o ostatních poznatcích v této práci – přestože byl model záměrně testován na datech společnosti Median, považuji jejich testování za podnětný směr analytické práce i pro další výzkumné agentury, které je momentálně nezohledňují.

[54] Předpokládám zde možné rozdíly mezi stranami např. podle míry disciplinovanosti jejich voličů. (Linek 2012)

ale ve stále přiměřené (doufám) míře komplexity, a založený na teoretických i metodologických poznatcích z odborné literatury.

# 7. Summary

This graduate thesis utilizes a secondary data analysis of a Czech Household Panel Study Wave 3 dataset gathered before the 2017 Czech parliamentary election and a follow-up after said election to confirm the relationship between the following variable groups: reported voter turnout and the intent to vote; reported turnout and reported turnout in the prior election, controlling for the intention to vote; and declared pre-election voting preference of a particular party and reported voting for said party. A secondary analysis of the XI/XII 2013 wave and the IX/X 2017 wave of the Median omnibus survey data is also performed in order to confirm the effect on time elapsed since previous election on previous voting behavior recall. All the hypothesized relationships are confirmed. The findings from the secondary data analyses are utilized in constructing an optimized likely voter model based on the model published by a Czech research agency called Median. This proposed model is compared to four other likely voter models that are or have been published in the Czech Republic. When compared through a set of statistical criteria outlined in (Prokop 2012) upon the construction of the Median voter model, the proposed optimized model shows the highest ranking out of all the compared models. The proposed model is then recommended for further testing and possible adaptation by Median. Exploring further research topics regarding the model´s construction such as cognitive dissonance in prior voting recall and the effect of prior voting behavior on the strength of voting preferences are suggested.

# 8. Použitá literatura


Adamany, David, and Philip Dubois. 1975. "The 'Forgetful' Voter and an Underreported Vote." *The Public Opinion Quarterly* 39 (2): 227–31. http://www.jstor.org/stable/2748149.

Adamany, David, and Mack C. II. Shelley. 1980. "Encore! The Forgetful Voter." *The Public Opinion Quarterly* 44 (2): 234–40.

Agresti, Alan, and Dennis Wackerly. 1977. "Some Exact Conditional Tests of Independence for R X C Cross-Classification Tables." *Psychometrika* 42 (1): 111–25. https://link.springer.com/article/10.1007%2FBF02293748.

Aldrich, John H., Jacob M. Montgomery, and Wendy Wood. 2011. "Turnout as a Habit." *Political Behavior* 33 (4): 535–63. https://doi.org/10.1007/s11109-010-9148-3.

Ansolabehere, Stephen, and Eitan Hersh. 2012. "Validation: What Big Data Reveal About Survey Misreporting and the Real Electorate." *Political Analysis* 20 (04): 437–59. https://doi.org/10.1093/pan/mps023.

Belli, Robert F., Michael W. Traugott, Margaret Young, and Katherine A. McGonagle. 1999. "Reducing Vote Overreporting in Surveys: Social Desirability, Memory Failure, and Source Monitoring." *Public Opinion Quarterly* 63: 90–108. https://doi.org/10.1086/297704.

Benoit, Kenneth. 1998. "Forecasting Parliamentary Outcomes in Multiparty Elections: Hungary 1998." *Paper Prepared for Annual Meeting of the American Political Science Association*. Boston. https://www.researchgate.net/publication/2315383_Forecasting_Parliamentary_Outcomes_in_Multiparty_Elections_Hungary_1998.

Bernstein, Robert, Anita Chadha, and Robert Montjoy. 2001. "Overreporting Voting: Why It Happens and Why It Matters." *The Public Opinion Quarterly* 65 (1): 22–44.

Blalock, H. M. 1963. "Correlated Independent Variables: The Problem of Multicollinearity." *Social Forces* 42 (2): 233–37. https://doi.org/10.1093/sf/42.2.233.

Bolstein, Richard. 1991. "Comparison of the Likelihood to Vote Among Preelection Poll Respondents and Nonrespondents." *The Public Opinion Quarterly* 55 (4): 648–50. http://www.jstor.org/stable/2749412.

Camilli, Gregory, and Kenneth D. Hopkins. 1978. "Applicability of Chi-Square to 2 × 2 Contingency Tables with Small Expected Cell Frequencies." *Psychological Bulletin* 85 (1): 163–67. https://doi.org/10.1037/0033-2909.85.1.163.

Český Statistický Úřad. 2010. "Celkové Výsledky Hlasování." 2010. https://www.volby.cz/pls/ps2010/ps2?xjazyk=CZ.

Český statistický úřad. 2013. “Celkové Výsledky Hlasování.” Volby Do Poslanecké Sněmovny Parlamentu České Republiky Konané ve Dnech 25.10. – 26.10.2013. 2013. https://volby.cz/pls/ps2013/ps2?xjazyk=CZ.

Český statistický úřad. 2017. “Celkové Výsledky Hlasování.” Volby Do Poslanecké Sněmovny Parlamentu České Republiky Konané ve Dnech 20.10. – 21.10.2017 (Promítnuto Usnesení NSS). 2017. https://www.volby.cz/pls/ps2017nss/ps2?xjazyk=CZ.

Chábová, Kristýna. 2013. “Komparace Metodických Postupů Předvolebních Průzkumů Agentur CVVM , Ppm Factum , Median a STEM.” *Naše Společnost* 1: 25–35.

CHPS. 2018. “CAPI Individuální Dotazník.”

Cohen, Jacob. 1988. *Statistical Power Analysis for the Behavioral Sciences*. New York: Lawrence Erlbaum Associates, Publishers.

Cumming, Robert Graham. 1990. “Is Probability Sampling Always Better? A Comparison of Results from a Quota and a Probability Sample Survey.” *Community Health Studies* 14 (2): 132–37. https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1753-6405.1990.tb00033.x.

Cutts, David, Edward Fieldhouse, and Peter John. 2009. “Is Voting Habit Forming? The Longitudinal Impact of a GOTV Campaign in the UK.” *Journal of Elections, Public Opinion and Parties* 19 (3): 251–63. https://doi.org/10.1080/17457280903073914.

CVVM. 2017. “Stranické Preference a Volební Model v Září 2017.” https://cvvm.soc.cas.cz/media/com_form2content/documents/c2/a4415/f9/pv170925.pdf .

CVVM. 2018. “Volební Model v Listopadu 2018,”. https://cvvm.soc.cas.cz/cz/tiskove-zpravy/politicke/volby-a-strany/4761-volebni-model-v-listopadu-2018.

Dassonneville, Ruth, and Marc Hooghe. 2016. “The Noise of the Vote Recall Question : The Validity of the Vote Recall Question in Panel Studies in Belgium, Germany, and the Netherlands.” *International Journal of Public Opinion Research* 29 (2).

Delavande, Adeline, and Charles F. Manski. 2010. “Probabilistic Polling And Voting In The 2008 Presidential Election.” *Public Opinion Quarterly* 74 (3): 433–59. https://doi.org/10.1093/poq/nfq019.

Denny, Kevin, and Orla Doyle. 2009. “Does Voting History Matter? Analysing Persistence in Turnout.” *American Journal of Political Science* 53 (1): 17–35. https://doi.org/10.1111/j.1540-5907.2008.00355.x.

Dillman, Don A, Glenn Phelps, Robert Tortora, Karen Swift, Julie Kohrell, Jodi Berck, and Benjamin L Messer. 2009. "Response Rate and Measurement Differences in Mixed Mode Surveys Using Mail, Telephone, Interactive Voice Response and the Internet." *Social Science Research* March (1): 1–18. https://www.sciencedirect.com/science/article/pii/S0049089X08000306.

Durand, Claire, Andre Blais, and Sebastien Vachon. 2002. "Accounting for Biases in Election Surveys : The Case of the 1998 Quebec Election." *Journal of Official Statistics* 18 (1): 25–44. https://www.scb.se/contentassets/ca21efb41fee47d293bbee5bf7be7fb3/accounting-for-biases-in-election-surveys-the-case-of-the-1998-quebec-election.pdf.

Erikson, Robert S., Costas Panagopoulos, and Christopher Wlezien. 2004. "Likely (and Unlikely) Voters and the Assessment of Campaign Dynamics." *Public Opinion Quarterly* 68 (4): 588–601. https://doi.org/10.1093/poq/nfh041.

ESOMAR. 2011. "Esomar Guideline for Online Research." https://www.esomar.org/uploads/public/knowledge-and-standards/codes-and-guidelines/ESOMAR_Guideline-for-online-research.pdf.

Granberg, Donald, and Sören Holmberg. 1990. "The Intention-Behavior Relationship Among U. S. and Swedish Voters." *Social Psychology Quarterly* 53 (1): 44–54.

Hrbková, Lenka, and Roman Chytilek. 2013. "Předvolební Průzkumy Veřejného Mínění a Předpojatost Tisku v České Republice." *Politologický Časopis* 20 (1): 32–48. https://doi.org/10.5817/pc2013-1-32.

IBM Corp. 2014. "IBM SPSS Statistics 23 Core System User´s Guide." ftp://public.dhe.ibm.com/software/analytics/spss/documentation/statistics/23.0/en/client/Manuals/IBM_SPSS_Statistics_Core_System_User_Guide.pdf.

IBM Corp. 2015. "IBM SPSS Regression 23."

Kreidl, Martin. 2005. "Zhodnocení Vlivu Práce Výzkumných Agentur Na Konstruktovou Validitu Škál." *Sociologický Časopis/Czech Sociological Review* 41 (1): 103–23.

Krejčí, Jindřich. 2008. "Co Je to Návratnost u Výběrových Dotazníkových Šetření." *SOCIOweb* 12: 1–12. http://www.socioweb.cz/upl/editorial/download/161_socioweb_12_08.pdf.

Kudrnáčová, Michaela. 2019. "České Panelové Šetření Domácností – Dokumentace k Datům." Vol. Verze 4.0.

Lebeda, Tomáš. 2003. "Průzkumy Volebních Preferencí Před Volbami Do Sněmovny 2002." *Politologický Časopis* 1: 22–37. http://www.politologickycasopis.cz/userfiles/file/2003/1/2003-1-2-Tom†Á Lebeda - PrÖzkumy volebn°ch preferenc° pžed volbami do Snōmovny 2002.pdf.

Lebeda, Tomáš, Jindřich Krejčí, and Yana Leontiyeva. 2004. "Volební Preference, Jak Jim Správně Porozumět." https://cvvm.soc.cas.cz/media/com_form2content/documents/c3/a542/f28/100049s_lebeda-vyzkumy.pdf.

Lebeda, Tomáš, Lukáš Linek, Pat Lyons, and Klára Vlachová. 2007. *Voliči a Volby 2006*. Praha: Sociologický ústav AV ČR. http://www.politologickycasopis.cz/userfiles/file/2003/1/2003-1-2-Tom†Á Lebeda - PrÖzkumy volebn°ch preferenc° pžed volbami do Snōmovny 2002.pdf.

Linek, Lukáš. 2009. "Jak Měřit Stranickou Identifikaci?" *Data a Výzkum - SDA Info* 3 (2): 187–210. http://dav.soc.cas.cz/uploads/21ba4ab0ec55c45e08f0737c645bbbf714256fdb_DaV09_2_pp187_210.pdf.

Linek, Lukáš. 2010. "Zhodnocení Valenční Teorie Volebního Chování Na Základě Oteřvených Otázek Ohledně Důvodů Volební Neúčasti." *Naše Společnost* 2. https://cvvm.soc.cas.cz/media/com_form2content/documents/c3/a136/f28/Linek,%2520Luk%25c3%25a1%25c5%25a1.%2520Zhodnocen%25c3%25ad%2520valen%25c4%258dn%25c3%25ad%2520teorie%2520volebn%25c3%25adho%2520chov%25c3%25a1n%25c3%25ad%2520na%2520z%25c3%25a1klad%25c4.

Linek, Lukáš. 2011. "Proč Se Měnila Úroveň Účasti ve Volbách Do Polaneké Sněmovny v Letech 1996–2010?" *Sociologický Časopis/Czech Sociological Review* 47 (1): 9–32.

Linek, Lukáš. 2012. "Vliv Volební Účasti Na Zisky Politických Stran ve Volbách Do Poslanecké Sněmovny." *Acta Politologica* 4 (1): 1–17.

Linek, Lukáš. 2018. "Účast ve Sněmovních Volbách : Poznatky , Měření a Obtíže s Předpovědí." http://simar.cz/assets/media/files/assets/uploads/01_Linek_Ucast_ve_snemovnich_volbach.pdf.

Lipps, Oliver, Nicolas Pekari, and Caroline Roberts. 2015. "Undercoverage and Nonresponse in a List-Sampled Telephone Election Survey" 9 (2): 71–82.

McHugh, Mary L. 2013. "The Chi-Square Test of Independence Lessons in Biostatistics." *Biochemia Medica* 23 (2): 143–49. https://doi.org/10.11613/BM.2013.018.

MEDIAN. 2017. "Sněmovní Volební Model MEDIAN Září 2017." http://www.median.eu/cs/wp-content/uploads/2017/09/MEDIAN_Snemovni_volebni_model_2017_09-1.pdf.

MEDIAN. 2018. "Sněmovní Volební Model Říjen 2018,". http://www.median.eu/cs/wp-content/uploads/2018/11/MEDIAN_Snemovni_volebni_model_2018_10.pdf.

Millar, Morgan M., and Don A. Dillman. 2011. “Improving Response to Web and Mixed-Mode Surveys.” *Public Opinion Quarterly* 75 (2): 249–69. https://doi.org/10.1093/poq/nfr003.

Mood, Carina. 2010. “Logistic Regression: Why We Cannot Do What We Think We Can Do, and What We Can Do about It.” *European Sociological Review* 26 (1): 67–82. https://doi.org/10.1093/esr/jcp006.

Murphy, John W., and Christian A. Schlaerth. 2010. “Where Are Your Data? A Critique of Secondary Data Analysis in Sociological Research.” *Humanity & Society* 34 (November): 379–90. https://journals.sagepub.com/doi/abs/10.1177/016059761003400405.

Olsen, Randall J. 2005. “The Problem of Respondent Attrition: Survey Methodology Is Key.” *Monthly Labor Review* 128 (2): 63. http://www.bls.gov/opub/mlr/2005/02/art9full.pdf.

Paul, Prabasaj, Michael L. Pennell, and Stanley Lemeshow. 2012. “Standardizing the Power of the Hosmer-Lemeshow Goodness of Fit Test in Large Data Sets.” *Statistics in Medicine*, 67–80. https://doi.org/10.1002/sim.5525.

Prokop, Daniel. 2012. “Výzkum Volebních Preferencí v ČR: Návrh Metodologické Optimalizace.” Univerzita Karlova, Fakulta sociálních věd. https://is.cuni.cz/webapps/zzp/detail/106481?lang=en.

Prokop, Daniel. 2019. “Ústní Konzultace s Konzultantem D. Prokopem 16. 4. 2019, ZŠ Boleslavova, Praha.”

Radler, Barry T., and Carol D. Ryff. 2010. “Who Participates? Accounting for Longitudinal Retention in the MIDUS National Study of Health and Well-Being.” *Journal of Aging and Health* 22 (3): 307–31. https://doi.org/10.1177/0898264309358617.

Sciarini, Pascal, Goldberg, Andreas C. 2015. “Lost on the Way: Nonresponse and Its Influence on Turnout Bias in Postelection Surveys.” *International Journal of Public Opinion Research*, December, edv049. https://doi.org/10.1093/ijpor/edv049.

Sciarini, Pascal, Goldberg, Andreas C.. 2016. “Turnout Bias in Postelection Surveys: Political Involvement, Survey Participation, and Vote Overreporting.” *Journal of Survey Statistics and Methodology* 4 (1): 110–37. https://doi.org/10.1093/jssam/smv039.

Sigelman, Lee. 1982. “The Nonvoting Voter in Voting Research.” *American Journal of Political Science* 26 (1): 47–56. http://www.jstor.org/stable/2110838.

Silver, Brian D., Barbara A. Anderson, and Paul R. Abramson. 1986. “Who Overreports Voting?” *American Political Science Review* 80 (2): 613–24. https://doi.org/10.2307/1958277.

Soukup, Petr. 2010. “Nesprávná Užívání Statistické Významnosti a Jejich Možná Řešení.” *Data a Výzkum - SDA Info* 4 (2): 77–104. http://dav.soc.cas.cz/uploads/27e65d18f9df9bee6df1af9649f82b267f9cccda_DaV10_2_s77_104.pdf.

STEM. 2019. “Volební Preference – Duben 2019.” 2019. https://www.stem.cz/stem-volebni-preference-duben-2019/.

Stocké, Volker. 2007. “Response Privacy and Elapsed Time Since Election Date as Determinants for Vote Overreporting.” *Internation Journal of Public Opinion Research* 19 (2).

Sudman, Seymour, and Norman M. Bradburn. 1973. “Effects of Time and Memory Factors on Response in Surveys.” *Journal of the American Statistical Association* 68 (344): 805–15. https://doi.org/10.1080/01621459.1973.10481428.

Tabery, Paulína, and Gabriela Šamanová. 2012. “Volební Kampaň a Deklarovaná Volební Účast ve Výzkumu Veřejného Mínění.” *Naše Společnost* 1: 19–28. https://cvvm.soc.cas.cz/media/com_form2content/documents/c3/a1523/f28/Volebni kampan a deklarovana volebni ucast.pdf.

TNS Kantar. 2019. “Trendy Česka Leden 2019.” https://ct24.ceskatelevize.cz/sites/default/files/2247140-trendy_ceska_2019_vlna_46_volebni_model.pdf.

Waldahl, Ragnar, and Bernt Olav Aardal. 1982. “Can We Trust Recall-Data ?” *Scandinavian Political Studies* 5 (2): 101–16.

Wuensch, Karl L. 2019. “Correlation and Regression Analysis: SPSS,” 1–13. http://core.ecu.edu/psyc/wuenschk/SPSS/CorrRegr-SPSS.pdf.

# Projekt diplomové práce

Klíčová slova: volební chování, volební model, volební preference, volební účast, konfirmační analýza, metodologie, metodika, vážení

## Úvod

Během posledních dvou let zabraly v České republice podstatnou část veřejného a společenského života témata a události spojené s volbami. V roce 2018 proběhly jak prezidentské, tak senátní a komunální volby, a v roce 2017 volby do poslanecké sněmovny parlamentu ČR. Tyto události byly v médiích a na sociálních sítích provázeny kromě jiného i množstvím průběžně publikovaných výzkumů volebních preferencí a volebních modelů. Přestože se volebním modelům ve výše zmíněné komunikaci dává poměrně velký prostor, neexistuje v současné době v České republice ustálený a obecně používaný způsob jejich výpočtu. I když by se to podle podobně vypadajících publikovaných výstupů nemuselo zdát, téměř každá výzkumná agentura publikující volební modely používá při přípravě těchto výstupů vlastní metodiku, které se od sebe následně liší jak samotným postupem, tak mírou komplexity. Tato situace, kde neexistuje rozšířený konsensus na tom, jak přesně tyto modely vytvářet, otevírá, naštěstí pro mě, poměrně velký prostor pro diskusi a inovace v této oblasti.

V navrhované diplomové práci bych rád využil několika otázek položených v rámci výzkumu Czech Household Panel Study týkajících se volebního chování, a použil je k ověření některých předpokladů, na kterých v současné době staví volební model agentura Median. Pokud se tyto předpoklady podaří potvrdit (a případně pokud se tak učinit nepodaří), využiji samotných poznatků získaných v rámci této sekundární analýzy dat k navržení optimalizace některých postupů při tvorbě volebního modelu agentury Median, které následně i otestuji na reálných datech, a porovnám s výstupy některých stávajících způsobů tvorby volebních modelů. Ve svých doporučeních budu vycházet z volebního modelu společnosti Median, která mi pro diplomovou práci poskytne přístup k některým zkoumaným datům, stejně jako technické zázemí pro jednotlivé analýzy. Od tohoto návrhu diplomové práce si slibuji zejména možnost pracovat s reálnými a profesionálně sbíranými daty a potenciální využití mých poznatků v aplikovaném výzkumu.

## Téma

V navrhované diplomové práci se budu zaměřovat na tři parametry současného volebního modelu agentury Median: první zvažovanou volenou stranu, záměr volební účasti a deklarovanou volební účast u předchozích voleb do PSP ČR. Výběr této konkrétní trojice je daný výčtem otázek z Czech Household Panel Study, které se týkají volebního chování. V následujících odstavcích stručně popíšu, v jaké podobě se tyto parametry v současné době měří ve volebních modelech třech českých výzkumných institucí: Centra pro výzkum veřejného mínění, STEM a společnosti Median.

První zvažovanou volenou stranu měří všechny tři výše zmíněné agentury relativně podobným způsobem, otevřenou otázkou bez pomoci karet se seznamem stran. Konkrétní znění těchto otázek se liší. V případě CVVM jde o otevřenou otázku *„Pokud byste k volbám šel, kterou stranu byste volil?"*, (CVVM 2018, 2) u instituce STEM *„Pokud byste (přesto) šel(šla) volit, kterou stranu byste příští týden volil(a)do Poslanecké sněmovny Parlamentu ČR?"* (STEM 2018) a u společnosti Median *"Kterou stranu nebo hnutí byste nejspíše volil(a), pokud byste v současnosti šel / šla k volbám do Poslanecké sněmovny?"*. (Median 2018, 12) V případě, že si respondent na základě této otevřené otázky žádnou první zvažovanou volenou stranu nevybere, může být u CVVM a Medianu tento parametr dotázán ještě jednou otázkou. V případě CVVM jde o otevřenou otázku *„Vzhledem k tomu, že si nejste svou volbou zcela jistý, jaké (další) strany byste zvažoval?"* (CVVM 2018, 2) na níž může jako odpověď uvést až dvě strany.

U společnosti Median potom *"Můžete prosím uvést všechny strany, o jejichž volbě byste uvažoval(a)?"* (Median 2018, 12), výčtovou otázku s rotujícím seznamem možných odpovědí, která zaznamenává pořadí zvolených stran dle respondentovy preference.

Z doplňujících otázek CVVM a Median je patrné, že některé volební modely (včetně modelu Median, který pro účely navrhované práce slouží jako výchozí), pracují s více než jednou zvažovanou volenou stranou pro jednoho respondenta. Vzhledem k tomu, že v rámci CHPS je ale dotazována pouze jedna, budu se v navrhované práci zaměřovat pouze na proměnnou první volené strany, tedy té, kterou v rámci zamýšlené sekundární analýzy dat mohu ověřovat. Ze stejného důvodu (absence odpovídajících dat z CHPS) se nebudu věnovat míře jistoty volby, kterou první zvažované volené straně respondent přiřkl.

Záměr volební účasti měří tyto tři agentury poměrně podobnými otázkami. U CVVM otázkou *„Představte si, že by příští týden byly volby do Poslanecké sněmovny. Šel byste*

*volit? Rozhodně ano, spíše ano, spíše ne, rozhodně ne, nebo nevíte?"* (CVVM 2018, 2) v případě STEM *„Pokud by se příští týden konaly volby do Poslanecké sněmovny, šel byste Vy osobně volit?"* (STEM 2018) a ve výzkumech společnosti Median otázkou *„Pokud by se nyní konaly volby do Poslanecké sněmovny, šel/šla byste volit?"*. (Median 2018, 12) Nejedná se tedy doslova o otázku týkající se záměru účastnit se samotných voleb v zákonem nastaveném čase, ale hypotetických voleb konajících se u všech agentur shodně týden ode dne rozhovoru. Vzhledem k datu sběru testovaných dat velmi blízko samotných voleb do PSP ČR v roce 2017 budu v diplomové práci s touto proměnnou nakládat právě ve významu záměru účastnit se reálných voleb.

Více specifická je situace u proměnných účasti a volené strany u minulých voleb do PSP ČR. Na základě rešerše v současné době mám za to, že s těmito proměnnými pracuje pouze volební model společnosti Median, který obě z nich měří následující otázkou: *„Kterou stranu nebo hnutí jste volil(a) při volbách do Poslanecké sněmovny, které proběhly v říjnu 2017?"* (Median 2018, 12) Její součástí je možnost odpovědi indikující neúčast, zvolená strana indikuje účast.

Za jedno z hlavních rizik navrhované diplomové práce považuji nutnost pečlivého nakládání s možnými zdroji chyby a posunů v analyzovaných datech. U všech zkoumaných proměnných se nejedná o postojová data, ale o deklarovaný záměr jednání, případně o reportované jednání v minulosti. Ani v jednom z těchto případů tedy nemůžeme považovat samotné datum za zaměnitelné s jednáním samotným. Deklarované záměry se do času samotných voleb mohou změnit, ať už například vlivem politických kampaní spojených s již popsanými jevy jako underdog efekt (Donsbach 2001, 22) nebo bandwagon efekt, (McAllister and Studlar 1991, 721) případně strategického načasování zásadních mediálních událostí. U reportovaného jednání v minulosti vstupuje do hry ne vždy spolehlivá paměť, (Gutek 1978) a případně i záměrný misreporting, například z důvodů sociální desirability (Fisher 1993) příslušnosti k určitým politickým stranám na základě výsledku voleb, a naopak. U obojího je třeba mít na paměti ještě fakt, že volební chování patří mezi citlivá osobní data, (ESOMAR 2011, 14) a tedy na ně pravděpodobně respondentů ze zkoumaného souboru odmítla odpovědět. Tato rizika se budu snažit mitigovat zejména vhodným zvolením analyzovaných datových souborů co možná nejblíže příslušným volbám do PSP ČR, případě jejich srovnávání s datovými soubory sebranými v čase od těchto voleb co možná nejvzdálenější. Samotné výše zmíněné zdroje chyb

v měření se pak budu snažit při analýze nebrat jako samozřejmost, ale naopak je na základě této analýzy zohlednit v nastavení parametrů výstupního navrhovaného volebního modelu.

## Metody

Praktická část mojí práce bude sestávat ze dvou hlavních sekcí. První sekce se bude věnovat konfirmační sekundární analýze dat z Czech Household Panel Study, konkrétně otázkám týkajících se deklarované volební účasti a volebních preferencí, případně v minulosti volené strany. Druhou sekcí potom bude aplikace poznatků z první části v návrhu konkrétního způsobu výpočtu volebního modelu a následně srovnání výstupů tohoto způsobu výpočtu s některými stávajícími způsoby.

V první sekci budu ověřovat několik hypotéz, které v současné době navrhuji v následujícím znění (finální formulace se pravděpodobně změní na základě dalšího studia literatury):

- H1: Respondenti, kteří před volbami deklarovali záměr se voleb do PSP ČR účastnit, po volbách deklarovali, že se jich účastnili, častěji než respondenti, kteří před volbami záměr se voleb účastnit nedeklarovali.
- H2: Respondenti, kteří deklarovali účast u voleb do PSP ČR v roce 2013, po volbách do PSP ČR v roce 2017 deklarovali, že se jich účastnili, častěji než respondenti, kteří účast u voleb do PSP ČR v roce 2013 nedeklarovali.
- H3 (kombinovaná H1; H2): Respondenti, kteří před volbami do PSP ČR v roce 2017 deklarovali záměr se voleb účastnit, a zároveň deklarovali účast u voleb do PSP ČR 2013, po volbách do PSP ČR v roce 2017 deklarovali, že se jich zúčastnili, častěji než respondenti, kteří nedeklarovali výše zmíněn záměr a předchozí účast.
- H4: Respondenti, kteří deklarují, že ve volbách do PSP ČR v roce 2013 volili určitou politickou stranu, deklarují, že tuto stranu budou volit ve volbách do PSP ČR v roce 2017, častěji než respondenti, kteří to nedeklarují.
- H5: Respondenti, kteří před volbami do PSP ČR v roce 2017 deklarovali, že by v nich volili určitou politickou stranu, po těchto volbách deklarovali, že tuto stranu volili, častěji než respondenti, kteří tuto preferenci před volbami nedeklarovali.

Hlavním výstupem této konfirmační analýzy by nemělo být potvrzení výčtu hypotéz výše, ale, pokud se podaří tyto hypotézy potvrdit, na základě konkrétních výsledků této analýzy nastavit hodnoty, kterými by se výše zkoumané parametry měly podílet na finálním

výstupu volebního modelu, tady druhá sekce diplomové práce. V ní budu vycházet ze způsobu, v níž v současné době agentura Median počítá svůj volební model, kde jsou tzv. široká data všech zvažovaných stran jednotlivých respondentů restrukturována na dlouhá data, kdy každá jedna zvažovaná volená strana má na základě socio-demografických charakteristik a dalších parametrů vlastní speciální váhu. Právě pro parametry minulé volební účasti a volby, avizované účasti a zvažované volené strany nastavím hodnoty na základě poznatků z první části diplomové práce. Ostatní parametry budou vycházet ze stávajícího způsobu výpočtu agentury Median.

Výsledný volební model (případně modely) následně srovnám jak se samotnými výsledky voleb, tak s výstupy stávajících volebních modelů agentury Median a CVVM. Jak můj navrhovaný model, tak srovnávané modely budou vypočítány na stejném datovém souboru (Omnibus Median, září 2017). Toto srovnání by mi mělo umožnit ověřit hlavní tezi celé diplomové práce, a tedy že výsledky volebního modelu vytvořeného výše popsaným způsobem na základě poznatků z dat sebraných v rámci Czech Household Panel Study budou výsledkům voleb podobat více než volební modely vytvořené jinými způsoby.

## Předpokládaná struktura práce

- Úvod
- Rešerše literatury
    - Zdroje chyb ve výzkumech volebního chování
    - Volební účast
    - Minulá volba
    - Volební preference
- Metody
- Data
    - CHPS 2013, 2015, 2017
    - Omnibus Median 2017
    - Panelový výzkum a možná rizika
- Analýza
    - Konfirmační analýza
    - Zapracování poznatků do volebního modelu
    - Výpočet navrhovaných volebních modelů
- Výsledky
    - Poznatky z CHPS
    - Srovnání navrhovaného modelu a stávajících volebních modelů
- Diskuse
- Závěry

## Literatura (projektu)


CVVM. 2018. “Volební Model v Listopadu 2018.” https://cvvm.soc.cas.cz/cz/tiskove-zpravy/politicke/volby-a-strany/4761-volebni-model-v-listopadu-2018.

Donsbach, Wolfgang. 2001. “Who’s Afraid of Election Polls? Normative and Empirical Arguments for the Freedom of Pre-Election Surveys.” https://wapor.org/wp-content/uploads/who-is-afraid-of-opinion-polls.pdf.

ESOMAR. 2011. “Esomar Guideline for Online Research.” https://www.esomar.org/uploads/public/knowledge-and-standards/codes-and-guidelines/ESOMAR_Guideline-for-online-research.pdf.

Fisher, Robert J. 1993. “Social Desirability Bias and the Validity of Indirect Questioning.” *Journal of Consumer Research* 20. https://www.jstor.org/stable/2489277.

Gutek, Barbara A. 1978. “On the Accuracy of Retrospective Attitudinal Data.” *The Public Opinion Quarterly* 42 (3): 390–401. https://www.jstor.org/stable/pdf/2748302.pdf.

McAllister, Ian, and Donley T. Studlar. 1991. “Bandwagon, Underdog, or Projection? Opinion Polls and Electoral Choice in Critain, 1979-1987.” *The Journal of Politics* 53 (3): 720–41. http://www.jstor.org/stable/2131577.

Median. 2018. “Sněmovní Volební Model Říjen 2018.” http://www.median.eu/cs/wp-content/uploads/2018/11/MEDIAN_Snemovni_volebni_model_2018_10.pdf.

STEM. 2018. “Volební Preference – Listopad 2018.” 2018. https://www.stem.cz/stem-volebni-preference-listopad-2018/.

# Přílohy

Poznámka: Na základě pokynů konzultanta nepřikládám přímo k práci vzhledem k proprietárnímu charakteru datového souboru omnibusu Median k této práci datové soubory a použité syntaxe pro software SPSS. Proto jsem se v jednotlivých kapitolách snažil jak analýzy, tak způsob konstrukce modelu popsat velmi podrobně. V případě potřeby mohu poskytnout použité datové soubory a syntaxe na vyžádání. Kontaktní e-mail: 80477199@fsv.cuni.cz

**Příloha 1: Deskriptivní statistiky v práci používaných vah (kde uvedeno)**

| Váha | Datový soubor | Minimum | Maximum | 10. percentil | 90. percentil |
|---|---|---|---|---|---|
| socio-demografie +minulá volba(2013) | CHPS (po imputaci) | 0,300001 | 2,005654 | 0,504913 | 2,988840 |
| socio-demografické charakteristiky | omnibus Median 2013 | 0,300001 | 3,000012 | 0,533428 | 2,660224 |
| socio-demografické charakteristiky | omnibus Median 2017 | 0,499931 | 1,999725 | 0,499931 | 1,922846 |
| socio-demografie +minulá volba(2013) | omnibus Median 2017 | 0,300206 | 3,002057 | 0,314147 | 2,004237 |
| Upravená před-restrukturační váha | omnibus Median 2017 | 0,501414 | 2,005654 | 0,501414 | 1,937867 |

## Příloha 2: Výstup logistické regrese používané při imputaci dat v 3.2.1.1

**Omnibus Tests of Model Coefficients**

| | | Chi-square | df | Sig. |
|---|---|---|---|---|
| Step 1 | Step | 506,764 | 19 | ,000 |
| | Block | 506,764 | 19 | ,000 |
| | Model | 506,764 | 19 | ,000 |

**Model Summary**

| Step | -2 Log likelihood | Cox & Snell R Square | Nagelkerke R Square |
|---|---|---|---|
| 1 | 908,185[a] | ,316 | ,483 |

a. Estimation terminated at iteration number 20 because maximum iterations has been reached.

**Variables in the Equation**

| | | B | S.E. | Wald | df | Sig. | Exp(B) |
|---|---|---|---|---|---|---|---|
| Step 1[a] | pracer | | | 21,651 | 5 | ,001 | |
| | pracer(1) | -,656 | ,395 | 2,751 | 1 | ,097 | ,519 |
| | pracer(2) | -1,698 | ,455 | 13,894 | 1 | ,000 | ,183 |
| | pracer(3) | -,782 | ,622 | 1,583 | 1 | ,208 | ,457 |
| | pracer(4) | -,730 | ,420 | 3,021 | 1 | ,082 | ,482 |
| | pracer(5) | ,270 | ,746 | ,131 | 1 | ,717 | 1,310 |
| | edu4 | | | 3,663 | 3 | ,300 | |
| | edu4(1) | ,092 | ,361 | ,065 | 1 | ,798 | 1,097 |
| | edu4(2) | -,327 | ,272 | 1,452 | 1 | ,228 | ,721 |
| | edu4(3) | -,014 | ,264 | ,003 | 1 | ,957 | ,986 |
| | polint | | | 3,269 | 3 | ,352 | |
| | polint(1) | ,404 | ,455 | ,789 | 1 | ,375 | 1,498 |
| | polint(2) | ,212 | ,257 | ,679 | 1 | ,410 | 1,236 |
| | polint(3) | ,367 | ,209 | 3,082 | 1 | ,079 | 1,443 |
| | wvote_r | | | 80,969 | 5 | ,000 | |
| | wvote_r(1) | ,570 | ,571 | ,994 | 1 | ,319 | 1,767 |
| | wvote_r(2) | -,208 | ,548 | ,144 | 1 | ,704 | ,812 |
| | wvote_r(3) | -1,386 | ,539 | 6,626 | 1 | ,010 | ,250 |
| | wvote_r(4) | -1,856 | ,527 | 12,397 | 1 | ,000 | ,156 |
| | wvote_r(5) | -22,486 | 15068,000 | ,000 | 1 | ,999 | ,000 |
| | ucast2013 | | | 47,405 | 3 | ,000 | |
| | ucast2013(1) | 1,936 | ,729 | 7,060 | 1 | ,008 | 6,934 |
| | ucast2013(2) | ,336 | ,719 | ,218 | 1 | ,640 | 1,399 |
| | ucast2013(3) | 1,227 | ,890 | 1,900 | 1 | ,168 | 3,411 |
| | Constant | 1,484 | ,997 | 2,218 | 1 | ,136 | 4,412 |

a. Variable(s) entered on step 1: pracer, edu4, polint, wvote_r, ucast2013.