

UNIVERZITA KARLOVA V PRAZE

FAKULTA SOCIÁLNÍCH VĚD

Institut sociologických studií, katedra sociologie

Bc. Petra Pospíšilová

**Možnosti analýzy latentních tříd:
případová studie diváků České televize**

Diplomová práce

Praha 2017

Autor práce: **Bc. Petra Pospíšilová**

Vedoucí práce: **PhDr. Ing. Petr Soukup**

Rok obhajoby: **2017**

Bibliografický záznam

POSPÍŠILOVÁ, Petra. *Možnosti analýzy latentních tříd: případová studie diváků České televize*. Praha, 2017. 65 s. Diplomová práce (Mgr.) Univerzita Karlova, Fakulta sociálních věd, Institut sociologických studií. Katedra sociologie. Vedoucí diplomové práce PhDr. Ing. Petr Soukup

Abstrakt

Diplomová práce „Možnosti analýzy latentních tříd: případová studie diváků České televize“ se zabývá analýzou latentních tříd a jejími možnostmi jako segmentační metody. Zkoumány jsou tři odlišné varianty – analýza latentních tříd pro spojená data z šesti vln výzkumu, simultánní analýza latentních tříd pro každou vlnu jednotlivě a analýza latentních tříd pro spojená data převedená do dichotomické podoby. Prostřednictvím sekundární analýzy dat jsou v případové studii zaměřené na postoj diváků k České televizi srovnány výsledky všech tří metod, popsány jejich silné a slabé stránky a je vybráno nejvhodnější řešení. Na základě toho jsou poté interpretovány čtyři nalezené segmenty diváků z hlediska jejich charakteristických vlastností a sociodemografických charakteristik. Oporou práce jsou texty klasických autorů v oblasti analýzy latentních tříd, na jejichž poznatky práce odkazuje.

Abstract

The thesis “The Potential of Latent Class Analysis: the Czech Television Audience Case Study” deals with latent class analysis and its potential as a segmentation method. Three different approaches are examined – latent class analysis for joint data from six research waves, simultaneous latent class analysis for each wave separately and latent class analysis for joint data transferred into dichotomous form. The results of all three methods are compared through secondary analysis of data from case study focused on audience’s perception of the Czech Television; their strong and weak points are described and the most suitable solution is chosen. Based on these findings, four identified segments of the audience are then interpreted from the point of their characteristic features and sociodemographic parameters. This paper is supported by works of classic authors in the area of latent class analysis; the thesis refers to their findings that also served as basis for composition of the research hypotheses.

Klíčová slova

Segmentace, analýza latentních tříd, faktorová analýza, veřejnoprávní funkce televize

Keywords

Segmentation, latent class analysis, factor analysis, public role of television

Rozsah práce: 134 850 znaků s mezerami

Prohlášení

1. Prohlašuji, že jsem předkládanou práci zpracovala samostatně a použila 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 ...

Bc. Petra Pospíšilová ...

Poděkování

Na tomto místě bych ráda poděkovala PhDr. Ing. Petru Soukupovi za cenné rady a vstřícný přístup při vedení této práce.

Děkuji také svým rodičům a prarodičům, kteří mi celých osmnáct let mého vzdělávání poskytovali útočiště plné lásky a podpory. V neposlední řadě děkuji svému partnerovi, který mi byl velkou oporou ve chvílích nejistoty.

TEZE DIPLOMOVÉ PRÁCE

Univerzita Karlova v Praze

Fakulta sociálních věd

Institut sociologických studií

Katedra Sociologie

Předpokládaný název práce: Možnosti analýzy latentních tříd: případová studie diváků České televize

Autor: Bc. Petra Pospíšilová

Konzultant: PhDr. Ing. Petr Soukup

Klíčová slova: Segmentace, analýza latentních tříd, faktorová analýza, Česká televize, veřejnoprávní funkce televize

Klíčová slova v Aj: Segmentation, latent class analysis, factor analysis, Czech television, public role of television

1. Teoretický kontext a námět práce

Typologie a jejich sestavování hrají v sociálně vědních výzkumech přirozeně zcela zásadní roli. Za touhou po jejich vytvoření stojí prostá myšlenka, že pozorované chování jedince může být v mnoha případech ovlivněno jeho příslušností k určitému typu. Klasifikace jsou proto jedním ze stěžejních konceptuálních úkolů a bez klasifikace by pravděpodobně žádné sociálně vědní výzkumy nemohly fungovat. [Bailey 1994: 1] Mnoho statických metod proto na potřebu vytváření typologií reagovalo a snažilo se přijít s řešením toho, jak různé typy jednání ve vzorcích lidského chování odhalovat a popisovat.

Segmentace jako pojem, pod nímž jsou tyto metody spojovány, vychází z principu rozčlenění zkoumaného vzorku populace na separované skupiny, které se vzájemně odlišují svými vlastnostmi. Cílem této analýzy je dosáhnout situace, kdy jsou jednotky spadající do jedné skupiny neboli segmentu, co nejvíce homogenní, ale ve vztahu k ostatním skupinám co nejvíce odlišné, tedy heterogenní. [Bailey 1994: 1] Při správném použití by poté segmentační analýza měla odpovědět na otázku, jak vypadají zkoumané jednotky a jaká jsou specifika, na jejichž základě mohou být rozřazeny do vzájemně odlišných skupin. Použity přitom mohou být proměnné od těch nejzákladnějších, jakými jsou například sociodemografické charakteristiky, až po komplikované postojové výroky.

V této diplomové práci bude jako jedna z možných metod segmentace představena analýza latentních tříd. Princip této techniky spočívá ve velmi obecné myšlence, že na vzorek dat by mělo být nahlíženo jako na data, která pocházejí ze dvou či více různě zkombinovaných populací. [Grover, Vriens 2006: 533] Pozorované proměnné nám však poskytují jen částečnou informaci o tom, do které populace (neboli třídy) má být který objekt zařazen. [Hendl 2015: 577]

Základní princip analýzy latentních tříd je proto postavena na rozdílu mezi tzv. manifestními a latentními proměnnými. Tyto pojmy odkazují k myšlence, že mnoho konceptů v sociálních vědách nemůže být pozorováno přímo. Existuje nespočet zajímavých teoretických konceptů, pro něž v rámci dostupných měření nalézáme jen částečně vyhovující indikátory. Každý z těchto měřitelných indikátorů však může být ovlivněn nepozorovanou, latentní, proměnnou, která jej definuje. [McCutcheon 1987: 5] Pokud chceme vztahy mezi pozorovatelnými proměnnými důkladně interpretovat, musíme analyzovat tyto skryté faktory. Latentní proměnné umožňují výzkumníkovi překročit hranice samotných dat, objevovat v datech skryté souvislosti a své závěry přenést na více teoretickou úroveň. [Henry 1999: 587] Z tohoto důvodu má smysl je blíže zkoumat a snaha o nacházení vztahů mezi proměnnými na latentní úrovni je jedním z pilířů této diplomové práce.

Analýza latentních tříd je v principech svého fungování velmi blízká faktorové analýze, proto s ní bývá často srovnávána. Klíčovým prvkem, který analýzu latentních tříd liší od faktorové analýzy, stejně jako od mnoha jiných segmentačních technik, je typ proměnných se kterými analýza pracuje. Zatímco faktory ve faktorové analýze jsou spojitě proměnné s nekonečným počtem kategorií, latentní proměnné v analýze latentních tříd jsou nominálního charakteru. Faktorová analýza tak vytváří kontinuální faktory, analýza latentních tříd tvoří diskrétní latentní třídy, díky čemuž vyniká oproti dalším segmentačním metodám. [Hagenaars, Halman 1989: 82]

2. Cíle práce a výzkumné otázky

Práce se zabývá analýzou latentních tříd jako segmentační technikou. K tomuto účelu jsem si určila dva základní cíle práce – metodologický a věcný. Prvním cílem je ukázat různá segmentační řešení postavená na variantách této statistické metody, poukázat na její silné a slabé stránky, diskutovat případné rozdíly mezi statistickým řešením a věcnou interpretací a v neposlední řadě vybrat nejvhodnější řešení. Dané segmenty budou poté do hloubky interpretovány.

Pro dosažení tohoto cíle práce jsem si definovala základní výzkumnou otázku, od které budou odvozeny hypotézy a související výzkumné otázky pro analytickou část práce. Výzkumná otázka zní takto:

Jaké jsou možnosti analýzy latentních tříd jako segmentační metody?

Aby mohly být jednotlivé varianty analýzy latentních tříd zhodnoceny z hlediska jejich praktického užití, bude významnou část práce tvořit případová studie. Ta se ve své tematické rovině zabývá obrazem České televize jako média veřejné služby. Cílem studie je segmentace televizních diváků do několika skupin na základě jich reflexe zkoumaných oblastí. Nalezené segmenty budou poté interpretovány a detailně charakterizovány. K tomuto cíli práce se pojí druhá, věcná, výzkumná otázka:

Jaké skupiny respondentů lze na základě jejich vnímání veřejnoprávní funkce České televize ve zkoumaném vzorku identifikovat?

3. Analyzovaná data

Data, která jsem se rozhodla v této práci využít, pocházejí z trackingové studie, kterou pro Českou televizi realizuje TNS Aisa. Cílem tohoto výzkumu je poskytnout vedení České televize a Radě pro rozhlasové a televizní vysílání ucelenou zpětnou vazbu na vysílání od televizních diváků. Klíčovými tématy výzkumu jsou především názory diváků na jednotlivé televizní stanice, jejich divácké preference či hodnocení veřejné služby České televize a jejího fungování jako média.

Výzkum je realizován od roku 2012 vždy dvakrát ročně, a to v jarní a podzimní vlně. Každá z vln zahrnuje přibližně 1 000 respondentů z cílové skupiny reprezentativní televizní populace ve věku 18+. Respondenti jsou v rámci každé vlny vybíráni kvótním výběrem, jehož kritérii jsou pohlaví, věk, vzdělání, velikost místa bydliště a kraj, ve kterém respondent žije. Takto zvolené kvóty odpovídají rozložení obyvatelstva v populaci na základě těchto kritérií. Zdrojem pro určování kvót jsou pravidelně aktualizovaná data Českého statistického úřadu.

Data byla sbírána nejprve metodou CATI (Computer Aided Personal Interviewing), v pozdějších vlnách metodou TAPI (Tablet-Assisted Personal Interviewing) dotazování.

Společnost TNS Aisa mi dala souhlas k využití dat ze všech osmi dosud realizovaných vln pro účely mé diplomové práce. K dispozici mám jak samotné datové soubory, tak veškeré výstupy, které z těchto dat TNS Aisa zpracovala. Výsledky mé případové studie poskytnu po dokončení práce oddělení výzkumu programu a auditoria České televize, které s nimi může naložit podle vlastního uvážení.

Dotazník používaný v této trackingové studii je poměrně obsáhlý a zkoumá celou řadu oblastí. Pro účely své diplomové práce jsem se rozhodla v případové studii analyzovat baterii výroků týkajících se názoru respondentů na Českou televizi jako celek. Konkrétní výroky, které respondenti hodnotili stupněm souhlasu na čtyř-položkové škále, jsou vypsány v následující tabulce:

Nyní vám postupně přečtu několik výroků týkajících se České televize jako celku. U každého výroku mi prosím řekněte, do jaké míry s ním osobně souhlasíte. Vyberte vždy jednu z variant odpovědí - rozhodně souhlasím, spíše souhlasím, spíše nesouhlasím nebo rozhodně nesouhlasím.

- B6_1 B6. ČT předkládá divákům pravdivý a nezkreslený obraz skutečnosti kolem nás.
- B6_2 B6. ČT mi pomáhá porozumět světu, ve kterém žijeme.
- B6_3 B6. Ve vysílání ČT dostávají v dostatečné míře prostor různé názory a úhly pohledu na jednotlivá témata.
- B6_4 B6. Informace předkládané ČT divákům jsou objektivní a vyvážené.
- B6_5 B6. ČT významně přispívá k rozvoji vzdělanosti svých diváků.
- B6_6 B6. ČT nabízí řadu pořadů, ze kterých se něčemu přiučím.
- B6_7 B6. ČT významně přispívá k tomu, aby se občané dokázali orientovat v oblasti práva, platných zákonů a norem.
- B6_8 B6. ČT dokáže přimět diváky k tomu, aby se aktivně podíleli na různých kulturně-spoločenských aktivitách.
- B6_9 B6. ČT má mezi televizemi vedoucí postavení v oblasti kulturních pořadů.
- B6_10 B6. ČT nabízí dostatek pořadů, které odpovídají mému vkusu a mým zájmům.
- B6_11 B6. ČT dává ve svém vysílání dostatečný prostor pro prezentaci kultury a života národnostních a etnických menšin v ČR.
- B6_12 B6. Při sledování ČT si dokážu dobře odpočinout.
- B6_13 B6. Při sledování ČT se dokážu dobře pobavit.
- B6_14 B6. Špičkové domácí a zahraniční sportovní události dostávají ve vysílání ČT patřičný prostor.
- B6_15 B6. Vysílání ČT jako celek přispívá k budování a posilování dobrých vztahů mezi lidmi a v celé společnosti.
- B6_16 B6. Vysílání ČT jako celek napomáhá k národnostní a etnické toleranci.
- B6_17 B6. Vysílání ČT jako celek napomáhá k náboženské toleranci.
- B6_18 B6. ČT je nezávislá na ekonomických a politických vlivech.
- B6_19 B6. ČT přináší divákům i pořady, které komerční TV nevysílají.
- B6_20 B6. Česká televize je důvěryhodná.
- B6_21 B6. Česká televize je inovativní.
- B6_22 B6. ČT navenek dostatečně prezentuje, jaké pořady a pro jaké skupiny diváků na svých kanálech nabízí.
- B6_23 B6. ČT prezentuje ve svém vysílání pro děti hodnoty slušnosti, vzdělanosti, úcty k národnostním menšinám a k životnímu prostředí.

4. Předpokládané metody zpracování

Práce bude vycházet jak z dostupných teoretických zdrojů, tak z vlastní analýzy dat. Analytická část práce má kvantitativní podobu a bude soustředěna na analýzu latentních tříd a její varianty. Technicky budou data zpracovávána v programech SPSS a MPlus.

5. Předpokládaná struktura práce

1. Úvod
2. Teoretická část práce
3. Metodologická část práce
4. Analytická část práce
5. Závěr

6. Orientační seznam literatury

BAILEY, K.D. *Typologies and taxonomies: an introduction to classification techniques*. Thousand Oaks, Calif.: Sage Publications. Sage university papers series, no. 07-102. ISBN 0-8039-5259-7. 1994

COOPER, J.C.B. *Factor Analysis: An Overview*. Taylor & Francis: The American Statistician, Vol. 37, No. 2 (1983), pp. 141-147. 1983

GOODMAN, L.A. *Exploratory Latent Structure Analysis Using Both Identifiable and Unidentifiable Models*, *Biometrika*, Vol. 61, No. 2., pp. 215-231. 1974

GROVER, J., VRIENS, M. *The Handbook of Marketing Research: Uses, Misuses, and Future Advances*. Thousand Oaks: Sage Publications, ISBN 141290997X. 2006

HAGENAARS, J.A, HALMAN, L.C. Searching for Ideal Types: The Potentialities of Latent Class Analysis. *Oxford University Press: European Sociological Review*, Vol. 5, No. 1 (May, 1989), pp. 81-96. 1989

HENDL, J. *Přehled statistických metod: analýza a metaanalýza dat*. Páté, rozšířené vydání. Praha: Portál. ISBN 9788026209812. 2015

HENRY N.W. *Latent Structure Analysis at Fifty*. Virginia Commonwealth University, Richmond, VA 23284-2014. pp. 587-592. 1999

JEŘÁBEK, H. *Paul Lazarsfeld's research methodology: biography, methods, famous projects*. Prague: Karolinum, ISBN 8024610981. 2006

JEŘÁBEK, H. SOUKUP, P. (eds.). *Advanced Lazarsfeldian methodology*. Prague: Karolinum, ISBN 9788024615219. 2008

LAZARSELD, P.F., HENRY, N.W. *Latent Structure Analysis*, Boston, Houghton Mifflin Company. 1968.

McCUTCHEON, A.L. *Latent Class Analysis*. Newbury Park: Sage Publications. ISBN 0803927525. 1987

WEDEL, M., WAGNER, K. A. *Market Segmentation: Conceptual a Methodological Foundations*. 2nd ed. Boston: Kluwer Academic Publishers. ISBN 0-7923-8635-3. 2000

OBSAH

ÚVOD	1
1 TEORETICKÁ ČÁST	2
1.1 SEGMENTAČNÍ ANALÝZA	2
1.2 METODY SEGMENTAČNÍ ANALÝZY	3
1.3 ANALÝZY S LATENTNÍ PROMĚNNOU	6
1.4 ANALÝZA LATENTNÍCH TŘÍD	7
1.4.1 Základní pojmy analýzy latentních tříd	10
1.4.2 Explorační analýza latentních tříd	18
1.4.3 Konfirmační analýza latentních tříd	18
1.4.4 Simultánní analýza latentních tříd	20
1.4.5 Srovnání faktorové analýzy a analýzy latentních tříd	21
1.4.6 Dosavadní případové studie	22
2 METODOLOGICKÁ ČÁST	24
2.1 VÝZKUMNÉ CÍLE, OTÁZKY A HYPOTÉZY	24
2.2 POPIS ANALYZOVANÝCH DAT	26
2.3 BATERIE OTÁZEK VYBRANÁ K ANALÝZE	27
2.4 ÉTIKA VÝZKUMU	29
2.5 VOLBA ANALYTICKÉHO POSTUPU	30
2.6 POUŽITÝ SOFTWARE	31
2.6.1 Analýza latentních tříd v programu Mplus	31
3 ANALYTICKÁ ČÁST	33
3.1 ZÁKLADNÍ POPIS SOUBORU	33
3.2 VÝSLEDKY ANALÝZY LATENTNÍCH TŘÍD PRO SPOJENÉ VLNY	33
3.3 VÝSLEDKY SIMULTÁNNÍ ANALÝZY LATENTNÍCH TŘÍD PRO VLNY JEDNOTLIVĚ	36
3.4 VÝSLEDKY ANALÝZY LATENTNÍCH TŘÍD PRO DICHOTOMICKÉ PROMĚNNÉ	39
3.5 SROVNÁNÍ TŘÍ POUŽITÝCH METOD	41
3.5.1 Analýza pro spojená data vs. simultánní analýza	41
3.5.2 Analýza pro spojená data vs. analýza pro dichotomické proměnné	44
3.5.3 Zhodnocení hypotéz výzkumu I.	46
3.6 CHARAKTERISTIKA NALEZENÝCH SEGMENTŮ	47
3.6.1 1. segment – Příznivci	47
3.6.2 2. segment – Podporovatelé	47
3.6.3 3. segment – Skeptici	48
3.6.4 4. segment – Nevyhranění	48
3.7 VLIV SOCIODEMOGRAFICKÝCH PROMĚNNÝCH NA SEGMENTY	49
3.7.1 Pohlaví	49

3.7.2	<i>Věk</i>	50
3.7.3	<i>Vzdělání</i>	51
3.7.4	<i>Velikost místa bydliště</i>	52
3.7.5	<i>Četnost sledování ČT</i>	53
3.7.6	<i>Zhodnocení hypotéz výzkumu II</i>	54
3.7.7	<i>Shrnutí charakteristik segmentů</i>	55
4	DISKUZE VÝSLEDKŮ, LIMITŮ A NÁVRHŮ PRO DALŠÍ VÝZKUM	56
	ZÁVĚR	57
	SUMMARY	59
	LITERATURA	62
	PŘÍLOHY	65

Úvod

Předkládaná diplomová práce se zabývá analýzou latentních tříd jako jednou z možných metod segmentace, které zatím především v českém prostředí nebylo dáno mnoho prostoru. Východiskem textu je myšlenka, že mnoho konceptů v sociálních vědách nemůže být měřeno přímo. Pozorovatelné neboli manifestní proměnné, které tvoří základ empirických studií, však mohou pomoci odhalit proměnné skryté neboli latentní, které kvůli jejich obecnosti či abstraktnosti nelze zachytit v samotném výzkumu. Nahlížíme-li skrze manifestní proměnné na proměnné latentní, můžeme v datech objevovat skryté vazby a souvislosti. Z tohoto důvodu má smysl je blíže zkoumat a snaha o nacházení vztahů mezi proměnnými na latentní úrovni je jedním z pilířů této diplomové práce.

Práce je členěna do tří hlavních celků, a to na teoretickou, metodologickou a analytickou část. V teoretické části práce je analýza latentních tříd zařazena do kontextu segmentačních metod a jsou vymezeny základní pojmy, se kterými analýza latentních tříd pracuje. Dále je představen základní princip jejího fungování, nastíněny odlišnosti mezi jejími jednotlivými podobami a stručně jsou popsány tři vybrané případové studie, ve kterých byla analýza latentních tříd aplikována.

Metodologická část práce vymezuje výzkumné otázky, hypotézy a analytické postupy. Cílem práce je ukázat a zhodnotit různá segmentační řešení postavená na analýze latentních tříd a jejich třech variantách – analýze latentních tříd pro spojená data z šesti vln výzkumu, simultánní analýze latentních tříd pro každou vln jednotlivě a analýze latentních tříd pro spojená data převedená do dichotomické podoby. Aby mohly být jednotlivé varianty analýzy představeny a posouzeny z hlediska jejich praktického užití, byla provedena případové studie. Případová studie se ve své tematické rovině zabývá obrazem České televize jako média veřejné služby. Cílem studie je segmentace televizních diváků do několika skupin na základě jejich reflexe zkoumaných oblastí. Data, která jsou pro tyto účely použita, pocházejí z šesti vln výzkumu, který pro Českou televizi pravidelně realizuje TNS Aisa.

Po provedení všech tří analýz v analytické části práce je vybráno nejvhodnější řešení, na kterém je postavena výsledná segmentace. Jednotlivé typy diváků jsou poté popsány z hlediska jejich charakteristických vlastností a sociodemografických charakteristik.

1 Teoretická část

V této části práce bude představen teoretický základ, ze kterého předložený text vychází. Popsány budou základní myšlenky segmentační analýzy, jednotlivé metody a jejich specifika. Pozornost bude poté zaměřena na analýzu latentních tříd, ze které vychází analytická část této práce.

1.1 Segmentační analýza

Typologie a jejich sestavování hrají v sociálně vědních výzkumech přirozeně zcela zásadní roli. Za touhou po jejich vytvoření stojí prostá myšlenka, že pozorované chování jedince může být v mnoha případech ovlivněno jeho příslušností k určitému typu. Klasifikace jsou proto jedním ze stěžejních konceptuálních úkolů. Bez klasifikace by pravděpodobně žádné sociálně vědní výzkumy nemohly fungovat. Význam klasifikace dobře vystihuje ve svém textu K.D. Bailey, který metodologickou potřebu klasifikace přirovnává k elektrině – každý z nás ji denně využívá, často však bez znalosti toho, jak funguje. [Bailey 1994: 1] Mnoho statistických metod proto na potřebu vytváření typologií reagovalo a snažilo se přijít s řešením toho, jak tyto fundamentální typy jednání ve vzorcích lidského chování odhalovat a popisovat.

Segmentační analýza vychází z myšlenky rozčlenění zkoumaného vzorku populace na separované skupiny, které se vzájemně odlišují svými vlastnostmi. Cílem této analýzy je dosáhnout situace, kdy jsou jednotky spadající do jedné skupiny neboli segmentu, co nejvíce homogenní, ale ve vztahu k ostatním skupinám co nejvíce odlišné, tedy heterogenní. Statisticky řečeno se analýza snaží minimalizovat rozptyl uvnitř skupin, a naopak maximalizovat rozptyl mezi jednotlivými skupinami. [Bailey 1994: 1] Při správném použití by poté segmentační analýza měla odpovědět na otázku, jak vypadají zkoumané jednotky a jaká jsou jejich specifika, na jejichž základě mohou být rozřazeny do vzájemně odlišných skupin.

Podíváme-li se na tento problém z matematického pohledu, typologie nejsou ničím jiným než lokalizací jednotlivých objektů v multidimenzionálních tabulkách. [Hagenaars, Halman 1989: 81] Analýza typologií za použití techniky latentních tříd představuje nejobecnější podobu modelu latentních tříd, který bude představen níže. [McCutcheon 1987: 13] V následující kapitole budou stručně představeny základní přístupy k segmentační analýze dat.

1.2 Metody segmentační analýzy

Segmentační analýza jako obecná snaha o seskupování vzájemně si podobných jednotek zahrnuje řadu statistických metod a postupů. Metody používané v segmentačních výzkumech bývají děleny podle dvou základních kritérií. První členění rozlišuje *a-priori* a *post-hoc* přístup. Základním rozdílem mezi těmito dvěma přístupy je fakt, že zatímco v *a-priori* postupu rozhodne o počtu vytvářených segmentů výzkumník předem na základě svého uvážení, u *post-hoc* metod je počet a typ segmentů ovlivněn až výsledky analýzy dat. [Wedel, Wagner 2000: 16]

Druhou možností klasifikace segmentačních metod je poté členění na *deskriptivní* a *prediktivní* statistické metody. Deskriptivní metody analyzují souvislosti v rámci jednoho setu segmentačního vzorku bez ohledu na závislost či nezávislost proměnných. Oproti tomu prediktivní metody analyzují souvislosti mezi dvěma množinami proměnných, kdy jedna sestává z nezávisle proměnných a druhou tvoří naopak pouze proměnné závislé. [Wedel, Wagner 2000: 16]

Tato dvě kritéria dohromady vytvářejí celkem čtyři možné kombinace klasifikace segmentačních metod, které jsou spolu s příklady konkrétních technik zobrazeny v následující tabulce (tabulka č. 1) a budou stručně popsány v podkapitolách této části práce.

Tabulka č. 1 – Klasifikace segmentačních metod

	A-priori	Post-hoc
Deskriptivní	Kontingenční tabulky Log-lineární modely	Metody shlukové analýzy Fuzzy techniky Smíšené modely
Prediktivní	Kontingenční tabulky Regresní, logitová a diskriminační analýza	AID, ANN Smíšené regresní modely

Zdroj: [Wedel, Wagner 2000: 16, tabulka 3.1]

1.2.1.1 A-priori deskriptivní metody

Na a-priori deskriptivním přístupu k segmentaci jsou založeny ty metody, u nichž je typ a počet segmentů definován již před samotným sběrem dat. [Wedel, Wagner 2000: 18] V praxi se tedy tyto metody využívají v situaci, kdy výzkumník předem ví, do jakých skupin chce zkoumanou populaci rozdělit, a analýzy slouží k pouhému rozdělení vzorku do těchto nadefinovaných skupin.

Nejvýraznějším zástupcem této kategorie metod jsou kontingenční tabulky. Ty byly velmi populárním nástrojem především v počátcích segmentačních výzkumů.

Kontingenční tabulky jsou stejně jako ostatní metody z této kategorie vhodné pro získání rychlého pohledu na segmenty a souvislosti mezi segmentačními bázemi. Ačkoli nejsou kontingenční tabulky ani například log-lineární modely ve svých zjištěních zcela vyčerpávající, jsou často používanými technikami, a to především v hybridních segmentačních technikách kombinujících a-priori a post-hoc přístup. Jedná se o dvoufázový proces, kdy je nejprve pomocí a-priori metody vzorek rozčleněn na dvě či více předem definovaných skupin vzešlých z jedné konkrétní proměnné zjišťované přímou otázkou v dotazníku (například. kuřáci vs. nekuřáci). Následně jsou v každé z těchto skupin provedeny post-hoc analýzy, často shlukového charakteru, které obě skupiny dále rozdělí do segmentů. Za výhodu těchto metod je považován fakt, že využívají silných stránek obou odlišných přístupů, které kombinují. Efektivita tohoto postupu však závisí především na post-hoc technikách použitých ve druhém kroku. [Wedel, Wagner 2000: 18]

1.2.1.2 Post-hoc deskriptivní metody

V post-hoc deskriptivním přístupu jsou segmenty identifikovány podle homogenity v rámci skupiny měřených proměnných. Obvykle se nejprve berou v úvahu sociodemografické charakteristiky, poté jsou data zpracována za použití shlukovacích metod, díky kterým jsou vytvořeny jednotlivé skupiny a popsány jejich charakteristiky. Počet segmentů je tak v tomto přístupu dán samotnými daty a použitou metodologií. [Wedel, Wagner 2000: 19]

1.2.1.3 A-priori prediktivní metody

A-priori prediktivní přístup vychází z definice segmentů vytvořených a-priori deskriptivním přístupem a následného použití prediktivních modelů pro popsání vztahu mezi příslušností k danému segmentu a skupinou nezávislých proměnných. V praxi se tedy jedná o dva kroky, kdy jsou nejprve a-priori definovány segmenty a poté jsou tyto segmenty popsány pomocí nezávislých proměnných. [Wedel, Wagner 2000: 22-23]

Tento přístup k segmentaci je možné rozdělit na dvě odlišné metody. Tzv. *forward* metoda je postavena na myšlence, že základní charakteristiky, tedy zejména např. sociodemografické charakteristiky, jsou nejprve použity pro vytvoření předem daných (a-priori) segmentů, které jsou poté vztaženy k dalším měřeným proměnným.

[Wedel, Wagner 2000: 22] Použijeme-li opět příklad s kouřením, v rámci této metody mohou být nejprve na základě sociodemografických charakteristik vytvořeny skupiny lidí ve zkoumané populaci, a až poté testováno, zda takto vytvořené segmenty předpovídají jedinci, zda bude kuřákem a případně jak silným, a to na základě otázek pojících se k životnímu stylu. Tzv. *backward* metoda je postavená na opačném postupu. V prvním kroku jsou tedy nejprve definovány segmenty na základě otázek týkajících se životního stylu. Ty jsou poté popsány na základě sociodemografických charakteristik. Uvedeme-li stejný příklad, populace by byla nejprve rozdělena na silné a slabé kuřáky, a až poté by bylo ověřováno, zda lze tyto segmenty popsat na základě obecných sociodemografických charakteristik. [Wedel, Wagner 2000: 22]

1.2.1.4 Post-hoc prediktivní metody

Post-hoc prediktivní metody identifikují segmenty na základě odhadu vztahu mezi závisle proměnnými a skupinou prediktorů. Segmenty vytvořené tímto přístupem jsou homogenní ve vztahu mezi závisle a nezávisle proměnnými. Populace je tak rozdělena segmenty, které jsou co do závisle proměnných maximálně odlišné, a to na základě nezávisle proměnných, kterými jsou nejčastěji sociodemografické charakteristiky. [Wedel, Wagner 2000: 23]

Jak vyplývá z popisu těchto čtyř skupin segmentačních metod, a-priori segmentační techniky jsou v porovnání s ostatními nejjednoduššími, současně však i nejméně efektivními metodami. Jejich síla spočívá především v hybridních technikách, kde jsou kombinovány s post-hoc technikami. [Wedel, Wagner 2000: 34] Podíváme-li se na základní segmentační techniky z hlediska tohoto rozdělení, v praxi se naopak jako velmi užitečné a často používané jeví shlukovací metody, které spadají mezi post-hoc deskriptivní přístupy. Diskriminační analýza je jedním z představitelů a-priori prediktivních metod. Vhodná je spíše pro popis segmentů než pro jejich identifikaci. Conjoint analýza je důležitým zástupcem skupiny post-hoc prediktivních metod, a to především díky svému dvoufázovému postupu – v první fázi jsou získány důležité hodnoty na základě hodnocení respondentů či statistických odhadů. Ve druhém kroku jsou poté subjekty shlukovány na základě podobností těchto klíčových hodnot a je určena prediktivní rovnice pro každý segment. Analýza latentních tříd, o které

pojednává tato práce, se řadí mezi post-hoc deskriptivní přístupy. [Wedel, Wagner 2000: 17-28]

1.3 Analýzy s latentní proměnnou

Dříve než bude přistoupeno k popisu analýzy latentních tříd, je důležité definovat dva základní pojmy, které prostupují celou práci. Těmito pojmy jsou *manifestní* a *latentní proměnné*. Znamějším pojmem než proměnné jsou pravděpodobně manifestní a latentní funkce sociologa R. K. Mertona, který rozdílem v těchto termínech popisuje na jedné straně funkce, které jsou v souladu se záměry jednajících osob, mají tedy manifestní charakter, a na straně druhé skryté, přímo nezamýšlené, tedy latentní, funkce jednání. Podstata těchto termínů zůstává stejná i u proměnných, které tyto pojmy popisují. Mnoho konceptů v sociálních vědách nemůže být pozorováno přímo. Existuje nespočet zajímavých teoretických konceptů, pro něž v rámci dostupných měření nalézáme jen částečně vyhovující indikátory. Vezmeme-li si například koncept náboženské oddanosti, zjistíme, že jej nemůžeme měřit přímo. Můžeme však předpokládat, že nábožensky oddanější jsou lidé, kteří častěji chodí kostela, častěji se modlí či přikládají náboženství větší význam ve svém osobním životě. Poté předpokládáme, že každý z těchto měřitelných indikátorů je ovlivněn nepozorovanou, latentní, proměnnou a můžeme studovat vzorce vzájemných vztahů mezi pozorovatelnými proměnnými, abychom pochopili a mohli charakterizovat latentní proměnnou stojící v pozadí těchto vztahů. [McCutcheon 1987: 5]

Manifestní proměnné jsou tak pozorovatelnými proměnnými, které tvoří základ empirických studií. Latentní proměnné jsou ty, které kvůli jejich obecnosti či abstraktnosti nelze měřit přímo, a proto je na ně nahlíženo skrze proměnné manifestní. Předpoklad, že existují i nepozorovatelné latentní proměnné, poté umožňuje výzkumníkovi překročit hranice samotných dat, objevovat v datech skryté souvislosti a své závěry přenést na více teoretickou úroveň. [Henry 1999: 587] Snaha o nacházení vztahů mezi proměnnými na latentní úrovni je jedním z pilířů této diplomové práce.

Možnost analyzovat vztah manifestních latentních proměnných pod sebou zahrnuje pojem *strukturální modelování* či *modelování pomocí strukturálních rovnic*. Analýzy pracující s latentními proměnnými jsou schopny podchytit celou řadu statistických konceptů, jako jsou například chybějící data, efekt náhody, smíšené modely, shluky či právě latentní třídy. Statisticky se dělí na přístupy, které pracují se

spojitou latentní proměnnou, a na ty, které na latentní proměnnou pohlížejí jako na kategoriální, jak bude popsáno dále.

1.4 Analýza latentních tříd

Ačkoli byly základní myšlenky analýzy latentních tříd popsány již v polovině minulého století, dosud jí bylo v porovnání s obecně známou a často užívanou shlukovou, faktorovou či diskriminační analýzou věnováno velmi málo prostoru. V současnosti však je jako smíšený model segmentační analýzy stále více používanou technikou. Její princip spočívá ve velmi obecné myšlence, že na vzorek dat by mělo být nahlíženo jako na data, která pocházejí ze dvou či více různě zkombinovaných populací. [Grover, Vriens 2006: 533] Pozorované proměnné nám však poskytují jen částečnou informaci o tom, do které populace (neboli třídy) má být který objekt zařazen. [Hendl 2015: 577] Analýza latentních tříd tak stejně jako předchozí zmíněné metody reaguje na potřebu vytváření typologií v sociálně vědních výzkumech a odhalování skrytých vzorců lidského jednání. Je tedy jednou z důležitých technik pro vytváření typologií. A to jako metoda pro empirické popsání několika latentních typů v rámci skupiny měřitelných indikátorů, či jako technika testování, zda typologie předpokládaná na základě teorie adekvátně reprezentuje data. [McCutcheon 1987: 8]

Se základní myšlenkou analýzy kategorických dat přišel v 50. letech minulého století P. F. Lazarsfeld, který představil model *analýzy latentních struktur*. Lazarsfeld představil tuto metodu jako nástroj pro vytváření typologií neboli také klastrů, postavených na dichotomických pozorovatelných proměnných. Cílem této metody bylo vyřešit problém měření tzv. *povahových konceptů*. Koncepty považoval Lazarsfeld obecně za teoretické konstrukty, které je možno definovat pomocí pozorovatelných znaků. Povahové koncepty, jsou poté ty koncepty, které nevycházejí z přímo měřitelných charakteristik, ale spíše ze schopnosti některých objektů vykazovat za specifických podmínek specifické reakce. To nás přivádí k myšlence, že lidské chování je ovlivněno skrytou strukturou. A právě tuto strukturu nazývá Lazarsfeld *latentní strukturou*.

Za pomoci analýzy latentních struktur se tak výzkumníci snaží skrze manifestní data pozorovat latentní struktury. [Lazarsfeld 1969: 351-353 in Jeřábek, Soukup 2008: 90] Vztahy mezi pozorovatelnými proměnnými a koncepty jsou poté vyjádřeny pravděpodobností, že se manifestní proměnná bude u objektu vyskytovat spolu se

skrytým latentním konceptem. [Lazarsfeld, Henry 1968: 3] Čím větší počet proměnných máme, tím vyšší je šance, že správně odhadneme počet latentních tříd a následně je efektivně charakterizujeme. Tuto pravděpodobnost dobře vystihuje známý příklad Paula Lazarsfelda: „*Pokud chceme odhadnout proporci bílých míčků v krabici, zlepšuje se náš odhad s každým dalším nezávislým tahem míčku z krabice. Sociolog ovšem nemůže položit tutéž otázku pětkrát za sebou. Proto klade několik různých (souvisejících) otázek, aby mohl vyslovit přesnější soudy*“ [Lazarsfeld, Henry 1968: 5]

Z analýzy latentních struktur se později vyvinula *analýza latentních tříd*. Pod pojmem analýza latentních tříd se skrývá skupina metod sloužících k analýze více směrných kontingenčních tabulek. [Wedel, Wagner 2000: 21] V rámci té mohou být analyzována kategoriální data, tedy data nominálního či ordinálního charakteru stejně, jako je tomu v analýze latentních struktur. Analýza latentních tříd také stejně jako analýza latentních struktur předpokládá, že pozorované souvislosti mezi kategoriálními proměnnými vycházejí z nepozorovatelných latentních proměnných, které mají specifický počet tříd. [Jeřábek, Soukup 2008: 93-94]

Podíváme-li se na typ dat, se kterými analýza latentních tříd pracuje, zjistíme, že manifestní proměnné, které vstupují do analýzy, mohou být jak dichotomického, tak polytomického charakteru. Skrze kategoriální manifestní proměnné jsou nalézány opět kategoriální latentní proměnné. Více známou faktorovou analýzu dat považoval Lazarsfeld za jednu z metod analýzy latentních struktur, která pro něj představovala variantu pro práci s kontinuálními latentními proměnnými, tedy faktory, které vycházejí z kontinuálních proměnných manifestních. Analýzu latentních tříd poté považoval za obdobu faktorové analýzy pro data kvalitativní povahy, která vedou výzkumníka k empirické identifikaci diskrétních latentních proměnných z nominálních či ordinálních dat. [McCutcheon 1987: 7]

Základní myšlenka analýzy latentních tříd je vhodně shrnuta v následujícím prostém obrázku (obrázek č.1):

Obrázek č. 1 – Model analýzy latentních tříd

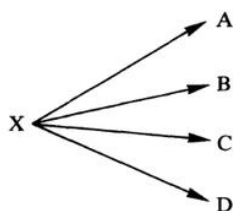


FIGURE 1 *The basic latent class model*

Zdroj: [Hagenaars, Halman 1989: 84]

X zde představuje latentní, tedy skrytou proměnnou. Kategorie latentní proměnné X jsou nazývány latentními třídami. Tyto latentní třídy reprezentují vniklou typologii. Latentní proměnná X zde ovlivňuje manifestní proměnné A, B, C a D, které jsou na rozdíl od ní samotné měřitelné. Stejně jako ve faktorové analýze jsou i zde manifestní proměnné vzájemně provázané, avšak jen díky latentní proměnné X. Žádné přímé vztahy mezi proměnnými zde neexistují. Tedy například A nemá přímý vliv na B. Vztahy mezi manifestními proměnnými tak mohou být popsány pouze na základě vztahu každé z nich s latentní proměnnou. [Hagenaars, Halman 1989: 84] Jinými slovy je základním východiskem práce s latentními proměnnými předpoklad, že latentní proměnná vysvětluje vztahy mezi manifestními proměnnými. Latentní proměnná je tak považována za „skutečný“ zdroj naměřených asociací. [McCutcheon 1987: 5-6] Tento klíčový předpoklad je označován jako *local independence*, neboli *lokální nezávislost proměnných* a je popsán v samostatné podkapitole v části práce zaměřené na základní pojmy analýzy latentních tříd.

Analýza latentních tříd spadá mezi tzv. *mixture models*, tedy *smíšené modely*. Tento pojem označuje celou řadu specifických modelů, které mají dva základní společné prvky. Všechny jsou tzv. *finite mixture models*, tedy *konečnými smíšenými modely*. Slovo konečný zde vyjadřuje, že tyto modely pojímají celkové rozložení jedné či více proměnných jako směsici konečného počtu rozložení komponentů, která je obvykle jednodušší a přehlednější než původní celkové rozložení. Druhým společným rysem smíšených modelů je, že příslušnost komponentů je pro některé či všechny jedince v populaci skrytá, tedy latentní. Více než vyjádření rozložení populace jako směsice známých skupin popisují smíšené modely celkové rozložení populace jako směsici určitého počtu skupin neznámých. Také díky neznámému počtu, formě a

proporci jednotlivých komponentů se smíšené modely řadí k modelům latentním. [Little 2013: 551-552]

U každého latentního smíšeného modelu můžeme identifikovat dvě základní části. První částí je model měření, který popisuje vztah mezi skrytou latentní proměnnou a pozorovatelnými manifestními proměnnými, které stojí v jejím popředí. Tento model tak specifikuje počet latentních tříd a distribuci jednotlivých indikátorů této latentní proměnné. Strukturální model poté specifikuje distribuci latentní proměnné v rámci populace a vztah mezi latentní proměnnou, prediktory a závěry. [Little 2013: 552]

Postup analýzy latentních tříd je takový, že na počátku stojí mnohorozměrná kontingenční tabulka, v níž ve většině případů existují závislosti mezi proměnnými. Ty mohou být ovlivněny právě tím, že za nimi stojí nepozorovaná latentní proměnná. Kontingenční tabulka je poté postupně rozkládána na více menších tabulek. Výzkumník musí před tímto procesem odhadnout, kolik tříd by se za proměnnými mohlo skrývat, stejně jako se například u faktorové analýzy rozhoduje o počtu faktorů. V praxi je však často vyzkoušeno několik modelů pro různý počet latentních tříd, a výzkumník se až na základě výsledků rozhodne, který model je nejvhodnější. [Green, Carmone, Wachspress 1976: 171]

V sociologických výzkumech se často setkáváme s daty kategoriální povahy. I z tohoto důvodu je analýza latentních tříd velmi užitečnou a přínosnou metodou a je s podivem, že v praxi je jen zřídka kdy využívána.

Každá analýza v sobě nese nedostatky, kterým musí výzkumník při sestavování typologií čelit. Analýza latentních tříd přirozeně nemůže vyřešit všechny problémy segmentačních analýz, ale troufám si tvrdit, že v mnoha případech je vhodnější metodou než shluková, faktorová či diskriminační analýza, neboť sestavuje typologie postavené na latentní úrovni (na rozdíl od shlukové a diskriminační analýzy, které se pohybují pouze na manifestní úrovni), všechny zahrnuté proměnné jsou nahlíženy v nominálním měřítku a ve formě vztahů existujících mezi proměnnými nejsou žádná výrazná omezení. [Hagenaars, Halman 1989: 83]

1.4.1 Základní pojmy analýzy latentních tříd

Předchozí kapitola byla zaměřena na obecnou ideu a cíle analýzy latentních tříd. Následující kapitola se zabývá jednotlivými pojmy, které jsou pro analýzu latentních tříd klíčové a jež je třeba definovat. Vysvětlena bude lokální nezávislost proměnných,

pravděpodobnost latentních tříd, latentní podmíněná pravděpodobnost či iterace, aby na tyto koncepty mohlo být dále navázáno v metodologické a analytické části práce.

1.4.1.1 Lokální nezávislost proměnných

Klíčovým předpokladem analýzy latentních tříd je *lokální nezávislost proměnných*. Tento předpoklad je postaven na myšlence, že indikátory objektů, které byly klasifikovány v rámci analýzy latentních tříd jako objekty spadající do jedné latentní třídy, jsou jeden na druhém statisticky nezávislé. [Jeřábek 2006: 55] Jinými slovy lze říci, že jediné, co mají manifestní proměnné společné, je fakt, že jsou indikátory stejné latentní proměnné. Manifestní proměnné tak spolu sice asociují, ale tato souvislost by se vytratila, pokud by neexistovala daná latentní proměnná.

Myšlenku lokální nezávislosti můžeme ilustrovat na případu, kdy máme dvě manifestní kategoriální proměnné A a B. Vazba mezi nimi je ovlivněna třetí proměnnou C, také kategoriálního charakteru. Souvislosti, které dokážeme identifikovat mezi manifestními proměnnými vstupujícími do analýzy, jsou tedy způsobené latentní proměnnou, kterou hledáme. Lokální nezávislost poté znamená, že proměnné A a B jsou na sobě nezávislé v jednotlivých kategoriích proměnné C. Kritérium lokální nezávislosti tak obecně řečeno poskytuje metodu pro rozhodování, zda vztahy mezi měřenými proměnnými jsou odvozeny od skrytých vysvětlujících proměnných. Pokud je několik vzájemně propojených proměnných označeno za lokálně nezávislé v rámci kategorií třetí proměnné, vycházíme z toho, že tato proměnná reprezentuje onen „skutečný“ předmět našeho zájmu. [McCutcheon 1987: 17] Jinými slovy lze říci, že latentní třídu definuje fakt, že pokud bychom odstranili efekt příslušnosti k latentní třídě na data, všechna zbylá měření by byla náhodná.

McCutcheon [McCutcheon 1987: 11-13] uvádí dva typy interpretace vztahů mezi dvěma či více proměnnými, které nejsou navzájem nezávislé. První typ interpretace je postaven na myšlence kauzality, tedy předpokladu, že jedna z proměnných může být označena za nezávislou a druhá za závislou proměnnou. Druhý typ interpretace toto kauzální pořadí nahrazuje myšlenkou symetrických vztahů mezi proměnnými. Tyto symetrické vztahy mohou být podle Rosenberga [Rosenberg 1968: 3-7] rozděleny na pět typů:

1. Proměnné jsou alternativními indikátory stejného konceptu, který vysvětlují. Typickým příkladem tohoto typu vztahu mohou být položky v baterii otázek dotazníkového šetření, které jsou vysoce vzájemně provázané, neboť popisují stejný koncept.
2. Proměnné jsou různými částmi jednoho komplexu. McCutcheon uvádí v této souvislosti příklad souvislosti mezi návštěvou opery a výběrem Martini jako drinku. Oba tyto faktory odkazují ke stejnému komplexu, kterým je životní styl jednice.
3. Třetím typem je funkcionální souvislost mezi proměnnými. Tento typ interpretace je postaven na myšlence, že některé elementy jednání se obvykle vyskytují, nebo naopak nevyskytují, spolu s jinými. Důvodem je propojení funkcí těchto proměnných, které splňují v rámci jednoho systému, který je předmětem studie. McCutcheon zde uvádí příklad organizace s formálními a neosobními vztahy, kde mohou být současně identifikována například administrativní pravidla procesů. Pokud jeden z těchto konceptů chybí, je pravděpodobné, že se v organizaci nebude vyskytovat ani koncept druhý.
4. Čtvrtým typem interpretace je efekt společné příčiny. Tento typ popisuje situaci, kdy proměnným předchází společná nadřazená proměnná, která obě dvě sledované proměnné ovlivňuje. Příkladem může být průměrná spotřeba elektřiny a porodnost konkrétní země. Obě tyto proměnné mohou být vysvětlovány skrze úroveň rozvinutosti země, ale nelze říci, že porodnost ovlivňuje spotřebu elektřiny a naopak.
5. Posledním typem interpretace jsou náhodné vztahy mezi proměnnými. Korelace mezi proměnnými nejsou postaveny na žádném logicky vysvětlitelném vztahu a jedná se pouze o náhodný vztah.

Analýza latentních tříd je technikou vhodnou především pro zkoumání a interpretaci symetrických vztahů mezi diskretními kategoriálními proměnnými. Ačkoli může být užita i ve své kauzální podobě, nejedná se o příliš rozšířenou podobu. Rozhodnutí o použití jednoho z pěti nastíněných přístupů k interpretaci poté závisí na

výzkumníkovi, teorii a logickém pohledu na zkoumaný koncept. Interpretace symetrických vztahů jako částí společného komplexu či jako efektu společné příčiny je často používána především při studiu typologií. [McCutcheon 1987: 13]

1.4.1.2 Latentní podmíněná pravděpodobnost

Pro výklad v kontextu analýzy latentních tříd je zcela zásadním pojmem pravděpodobnost. Vrátime-li se na chvíli zpět k analýze latentních struktur, je důležité podotknout, že Lazarsfeld uvažoval vstupní data pro analýzu latentních struktur jako naměřené hodnoty vážené pravděpodobnostmi. Každá odpověď tak může být považována za pravděpodobnost, tedy hodnotu od 0 do 1, příklonu k odpovědi „ano“ či „ne“ za hypotetického předpokladu, že by jedinec byl dotazován opakovaně a nepamatoval si svoje předchozí odpovědi. Jak již bylo popsáno výše v kapitole věnované základním principům fungování analýzy latentních tříd, analýza latentních tříd rozšiřuje analýzu latentních struktur kromě jiného i o možnost pracovat s ne-dichotomickými proměnnými. Pro daného respondenta tak může být určena pravděpodobnost volby každé z nabízených možností odpovědi. [Jeřábek 2006: 54]

Pojmem *latentní podmíněné pravděpodobnosti* jsou v analýze latentních tříd označovány skupiny odhadnutých pravděpodobností určitých odpovědí při přináležení do určité latentní třídy. [Eaton a col. 1989: 106 in Jeřábek, Soukup 2008: 97-98] Měření je tedy vztah mezi každou hodnotou manifestní proměnné a latentní třídou, což může být přirovnáno například k faktorovým zátěžím, které vyjadřují korelaci mezi proměnnou a faktorem. Latentní podmíněné pravděpodobnosti obdobně vyjadřují pravděpodobnost, s jakou bude zkoumaný objekt dosahovat určitých hodnot měřených proměnných. [McCutcheon 1987: 33]

Takto spočítané latentní podmíněné pravděpodobnosti vzešlé z dat jsou jako celek označovány jako tzv. *response pattern* neboli *vzorec odpovědí*. Představit si tento vzorec můžeme jako matici, kde ve sloupcích jsou umístěny zkoumané položky a v řádcích poté jednotliví respondenti. Máme-li tedy m dichotomických položek, dostáváme 2^m vzorců odpovědí. Sociologie jako obor se poté zabývá frekvencí výskytu individuálních vzorců odpovědí. [Jeřábek 2006: 53] Latentní podmíněné pravděpodobnosti jsou klíčovým výstupem pro interpretaci jednotlivých latentních tříd, jak bude ukázáno dále v rámci provedené případové studie.

1.4.1.3 Pravděpodobnost latentních tříd

Třetím důležitým pojmem pro interpretaci výsledků analýzy latentních tříd je pojem *pravděpodobnosti latentních tříd*. Pravděpodobnosti latentních tříd popisují rozložení tříd latentní proměnné. Důležitou roli zde hrají dva základní aspekty – počet tříd a jejich relativní velikost. Počet tříd vyjadřuje počet latentních typů definovaných modelem latentních tříd pro námi zkoumané měřitelné proměnné. Pokud má tedy například latentní proměnná tři třídy, znamená to, že zkoumaná populace může být rozdělena na tři základní typy respondentů. [McCutcheon 1987: 18-19] Minimální počet tříd je přirozeně dvě, neboť jedna latentní třída by značila úplnou nezávislost měřených proměnných. Velikost jednotlivých tříd poté také představuje důležitý aspekt pro interpretaci výsledků analýzy. Říká nám, nakolik je zkoumaná populace rovnoměrně rozložena do jednotlivých typů, či zda naopak jedna třída nevysvětluje oproti ostatním třídám velkou část populace. [Jeřábek, Soukup 2008: 97]

1.4.1.4 Metoda maximální věrohodnosti

V praxi se výzkumník často dostává do situace, kdy nemá věrohodnou teorii, která by mu na základě pouhého úsudku předpovídala ideální počet latentních tříd, které mají být v rámci analýzy vytvořeny. V takovém případě musí být počet tříd odhadnut nepřímo prostřednictvím testování jednotlivých modelů a jejich srovnání. U jednotlivých modelů se porovnává jejich absolutní a relativní přesnost a je vybrán model, který nejlépe pasuje na zkoumaná data. [Geiser 2013: 237]

Přesnost modelu je stejně jako parametry odhadována na základě tzv. *maximum likelihood estimation (MLE)*, tedy metody maximální věrohodnosti, sloužící pro konstrukci odhadu parametru. Jedná se o proceduru, která poskytuje odhady parametrů statistického modelu tím, že hledá hodnotu parametru, která maximalizuje věrohodnost modelu.

Od kritéria maximální věrohodnosti je odvozeno tzv. *loglikelihood estimation*, které slouží jako východisko pro výpočet dalších statistik.

1.4.1.5 Počet iterací

Analýza latentních tříd vyžaduje ve svém postupu určení počátečních neboli také startovacích odhadů podmíněných pravděpodobností a pravděpodobností latentních tříd.

Po určení počátečních hodnot dojde k přepočtu těchto pravděpodobností a jejich srovnání s předchozími odhady. Jeden tento krok se nazývá iterace.

Proces iterací tedy začíná se zkušebními hodnotami pro podmíněné pravděpodobnosti a pravděpodobnosti latentních tříd modelu, které jsou počítány stále znovu, dokud není proces zastaven. Při rozhodování o počtu provedených iterací může být postupováno dvěma odlišnými metodami. V prvním případě může výzkumník na základě svého odhadu určit přesný počet iterací, která mají být provedeny. Nevýhodou takového přístupu je riziko, že další provedené odhady po zastavení iterací by mohly přinést podstatné změny, což výzkumník nezjistí. Druhou metodou pro určení počtu iterací může být takový postup, kdy je předem určena hodnota tolerance, s jakou se mohou lišit odhady na začátku a konci každé iterace. Jinými slovy se tedy jedná o rozdíl mezi zadanými vstupními hodnotami a hodnotami vypočítanými. Pokud se rozdíl nachází v toleranci, mohou být iterace zastaveny. [McCutcheon 1987: 23-25]

Při nastavování počtu iterací je třeba mít na paměti, že maximální věrohodnost má kromě svého globálního maxima také lokální maximum. Pozor je třeba dát na to, že výpočet může skončit právě u tohoto lokálního maxima, namísto globálního. Abychom se tomuto vyhnuli a nedošlo k nesprávnému výpočtu dalších statistik, je potřeba počítat opakovaně s různými počátečními odhady a kontrolovat, zda bylo dosaženo vícekrát stejných výsledků. Ty značí, že bylo nalezeno globální, nikoli lokální maximum. [Geiser 2013: 237-239]

1.4.1.6 Pravděpodobnostní a loglineární parametrizace

Přibližně od devadesátých let minulého století se analýza latentních tříd stávala důležitou metodou pro analýzu kategoriálních dat. Stejně jako se užití této metody stávalo stále častějším, rozšiřovala se i škála témat a oblastí, v rámci níž byla analýza latentních tříd jako analytická metoda využívána. Jak uvádí Alan McCutcheon, hlavním důvodem užitečnosti analýzy latentních tříd je fakt, že je možné použít ve dvou odlišných, ale zcela rovnocenných parametrizacích pro tvorbu modelu – pravděpodobnostní parametrizaci a loglineární parametrizaci. [Hagenaars, McCutcheon 2002: 56] Možnost volby typu parametrizace je nesporně jednou z velkých předností analýzy latentních tříd, neboť tím činí vznikající model velmi flexibilním a použitelným napříč širokou škálou výzkumných případů. [Hagenaars, McCutcheon 2002: 57]

Pravděpodobnostní parametrizace je pravděpodobně častěji užívanou variantou parametrizace, neboť bývá označována za intuitivnější volbu. Pravděpodobnostní parametrizace neomezeného modelu latentních tříd je charakteristická dvěma typy kategoriálních proměnných, kterými jsou manifestní a latentní proměnné, a stejně tak dvěma typy parametrů – podmíněnou pravděpodobností a pravděpodobností latentních tříd, které byly popsány v předchozích kapitolách 1.4.1.2 a 1.4.1.3. Hypotézy jsou v rámci analýzy latentních tříd testovány tak, že jsou nejprve nastaveny restriktce, a pak je testováno, jak tyto restriktce ovlivňují to, nakolik vytvořený model vystihuje data. Pravděpodobnostní parametrizace požaduje, aby se součet pravděpodobností všech latentních tříd rovnal 1. To jinými slovy znamená, že v daném případě existuje latentní třída pro každou z možných odpovědí vyskytujících se v datech. Tato parametrizace má i další praktickou výhodu – pokud by byla v analýze zpracovávána dichotomická data, nebylo by nutné v interpretaci latentních tříd uvádět pravděpodobnost přináležení případu do obou dotazovaných kategorií, neboť tento součet je roven 1. Pokud je tedy známa pravděpodobnost pro jednu variantu odpovědí, druhá přirozeně tvoří dopočet do 1. [Hagenaars, McCutcheon 2002: 58-59]

Model latentních tříd může být v některých případech prezentován také jako loglineární model. Loglineární parametrizace se však liší od běžných loglineárních modelů, a to ve dvou základních bodech. Zaprvé loglineární model pro analýzu latentních tříd pracuje s latentními, tedy skrytými, proměnnými. Zadruhé jsou poté zahrnuty pouze parametry mezi latentní proměnnou a každým indikátorem. Všechny ostatní případy zahrnující kombinace mezi indikátory jsou nastaveny na 0. [Hagenaars, McCutcheon 2002: 61]

1.4.1.7 Informační kritéria

Vzdálíme-li se na chvíli od analýzy latentních tříd ke statistice v jejím nejširším pojetí, jedním ze základních pojmů tohoto oboru je statistická významnost. Statistická významnost neboli jinými slovy testování nulové hypotézy vychází z formulace dvou hypotéz o situaci v základním souboru – nulové hypotézy a hypotézy alternativní. [Soukup 2010: 77-79] Po definování hypotéz je testováno, zda empiricky získaná data odpovídají předem určenému teoretickému předpokladu. Cílem je zjistit, zda je možné nulovou hypotézu na dané hladině významnosti zamítnout, či nikoliv. Při interpretaci statistické významnosti je poté pozornost soustředěna na konkrétní obecně přijímané

hodnoty pro testování statistických hypotéz, a na jejich základě je vyneseno verдик. [Soukup 2010: 79]

Soukup však ve svém článku *Nesprávná užívání statistické významnosti a jejich možná řešení* polemizuje na tím, že testování hypotéz na tomto principu má několik významných nedostatků. Za všechny jmenujme nereálnost nulových hypotéz, až téměř mechanické používání 5% hladiny významnosti či myšlenku, že pojem statisticky významné nemusí nutně znamenat i věcně důležité. [např. Cohen 1994 in Soukup 2010: 83] Těmto problémům se snaží předejít přístup založený na porovnání několika modelů na základě informačních kritérií, ze kterého vychází analýza latentních tříd a jenž bude pro hodnocení statistické vhodnosti použit v rámci analytické části této práce.

Klíčovým principem porovnávání na základě informačních kritérií je fakt, že namísto generování jedné hypotézy o jednom modelu, jako je tomu u výše popsaného běžného testování statistické významnosti, je srovnáváno několik vzájemně i konkurujících modelů současně. Na základě věrohodností funkce je poté vypočtena hodnota informačního kritéria pro každý z modelů. Tato funkce zohledňuje navíc složitost modelu, která tak ovlivňuje hodnotu informačního kritéria. [Soukup 2010: 95] Dvěma nejvíce užívanými kritérii jsou Akaikeho informační kritérium (AIC) a bayesovské Schwarzovo informační kritérium (BIC). AIC pracuje s věrohodností daného modelu a jeho složitostí. Čím je model složitější, tím je hodnota AIC vyšší, neboť s každým dalším parametrem se hodnota AIC zvyšuje o dvě jednotky. Jak již sám název napovídá, BIC vychází z bayesovské statistiky a je postaveno na poměrovém porovnávání zkoumaného modelu a modelu saturevaného, tedy modelu obsahujícího všechny parametry. Kritérium BIC je považováno za konzervativnější, preferuje tedy jednodušší modely, AIC naopak upřednostňuje modely složitější. Co se volby mezi těmito dvěma kritérii týče, v sociologii dominuje užívání kritéria BIC. [Soukup 2010: 96-97] Z obecného pohledu by však kritéria měla jít ruku v ruce.

Informační kritéria nemají z hlediska vyhodnocování provedených analýz sama o sobě žádný smysl, cílem je jejich vzájemné srovnání mezi několika modely. Při hodnocení je za statisticky vhodnější považován model s nižšími hodnotami AIC a BIC. Zde je však třeba mít na paměti, že se jedná o čistě statistickou vhodnost modelu, která nemusí nutně označovat i věcně nejvhodnější model. Důraz by tak kromě statistických ukazatelů měl být u analýzy latentních tříd kladen i na logickou a smysluplnou interpretaci výsledků.

1.4.2 Explorační analýza latentních tříd

Analýza latentních tříd bývá nejčastěji užívána právě ve své explorační podobě, kdy nejsou předem specifikovány žádné parametry. Typicky je tak využívána ve chvíli, kdy neexistují žádné silné předem definované hypotézy napovídající z dat počet a povahu latentních tříd. Ačkoli je tento přístup dobře odůvodnitelný, může být i velmi limitující, pokud je již známo více informací o populaci a proměnných, které jsou součástí dané studie. Typickou otázkou, která značí vhodnost užití explorační analýzy latentních tříd, můžeme shrnout takto: *Kolik latentních tříd se skrývá za skupinou kategorických pozorovatelných proměnných?* [Finch, Bronk 2011: 132-133]

Explorační analýza latentních tříd postupuje obvykle tak, že je ze zkoumaných dat nejprve vytvořen model s jednou latentní třídou, který je nazýván *nezávislým modelem*. Poté následuje vytvoření modelu se dvěma třídami. Počet tříd se dále navyšuje až do chvíle, kdy vzniklý model vhodně odpovídá datům. [Jeřábek, Soukup 2008: 98] Problém explorační analýzy latentních tříd však spočívá v tom, že právě vyhovující počet tříd nelze nikdy odhadnout s naprostou přesností. Vždy se jedná o pouze přibližný odhad. [Hojitink 2001: 563 in Jeřábek, Soukup 2008: 99] Druhým nezbytným krokem analýzy je zvolení počáteční sady parametrů. Ty slouží jako základ iterativního algoritmu maximálně věrohodného odhadu. „*Zkoušením různých počátečních hodnot odhadů parametrů můžeme porovnat jednotlivá řešení dosažená iterativním postupem, abychom zjistili, které z nich minimalizuje chí-kvadrát statistiku pravděpodobnostního poměru*“ [Goodman 1974: 218] Dále jsou jednotlivé sestavené modely testovány pomocí chí-kvadrát testů. Na základě nich výzkumník přijímá, nebo zamítá tyto modely. Poslední fází explorativní analýzy latentních tříd je interpretace modelu, který výzkumník zvolil jako nejvhodnější, a popis jednotlivých latentních tříd. Zde hrají velkou roli zkušenosti výzkumníka a jeho vlastní odhad, neboť je v mnoha případech těžké porovnat, kolik latentních tříd nejlépe vystihuje zkoumaná data.

1.4.3 Konfirmační analýza latentních tříd

V rámci konfirmační analýzy latentních tříd jsou testovány hypotézy o pravděpodobnostních hodnotách latentních tříd. [McCutcheon 1987: 1987] Konfirmační analýza latentních tříd tak výzkumníkovi nabízí vhodný nástroj pro modelování a testování specifických hypotéz o skrytých vzorcích v rámci pozorovatelných proměnných. Tyto hypotézy o povaze a počtu latentních tříd jsou poté promítnuty do

vznikajícího modelu jako parametry. Otázka, která předchází konfirmační analýze latentních tříd, může znít například takto: *Existují tři latentní třídy skryté za skupinou kategoriálních proměnných, kde skupina A dosahuje vyšší pravděpodobnosti odpovědi než skupina B a C, jak nám napovídá teorie?* [Finch, Bronk 2011: 132-133]

V případě konfirmačního přístupu k analýze latentních tříd si může výzkumník stanovit specifické hodnoty pravděpodobností modelu. Stejně tak může být předem určena i pravděpodobnost, že případ spadá do té které latentní třídy. McCutcheon [McCutcheon 2002] popisuje tři typy omezení parametrů, které mohou být v rámci konfirmačního modelování latentních tříd použity. První možností je situace, kdy je předem definováno, že jsou jedna nebo více hodnot parametru stejné napříč všemi latentními třídami. V takovém případě mluvíme o tzv. *restrikci rovnosti*. Druhým typem omezení je tzv. *deterministická restrikce*, která se soustředí na testování toho, zda podmíněné pravděpodobnosti odpovědi dosahují určité stejné hodnoty pro jednu či více latentních tříd. Tato hodnota je často 1 nebo 0. Tento přístup je využíván v případech, kdy je výzkumník na základě teorie schopen usoudit, že např. určitý druh chování se zcela jistě nebude v konkrétních latentních třídách vyskytovat. Pro tuto položku proto a-priori nastaví hodnotu na 0. Třetí případ zahrnuje použití nerovných restrikcí pro testování hypotéz s ohledem na relativní pravděpodobnost, zda položka do latentní třídy spadá, či nikoli. Tuto *restrikci nerovnosti* je vhodné použít v případě, kdy výzkumník předpokládá, že jednotlivé latentní třídy je možno seřadit podle jejich ovlivnitelnosti konkrétní položkou. [Finch, Bronk 2011: 135-136]

Svým praktickým provedením se konfirmační analýza latentních tříd nijak zásadně neliší od analýzy explorační. Specifikum je zde především v množství parametrů, které jsou v rámci konfirmační verze odhadovány. Celkově je jich méně, a to kvůli restrikcím, které tvorbu tohoto modelu ovlivňují. Tím se současně zvyšuje i počet stupňů volnosti tohoto modelu. Druhá odlišnost vyplývá již ze samotné podstaty rozdílu mezi exploračním a konfirmačním přístupem k analýze. Zatímco u exploračního modelu je analytický proces zakončen interpretací vzniklých latentních tříd, u konfirmačního modelu má výzkumník o způsobu interpretace modelu jasnou představu již před samotnou analýzou. Tu poté zakončuje pouze zhodnocením, nakolik jeho počáteční představa odpovídala datům.

1.4.4 Simultánní analýza latentních tříd

Analýza latentních tříd v podobě, v jaké byla doposud v této práci nahlížena, byla zaměřena na model latentních tříd pro data sesbíraná v rámci jednorázového výzkumu. Simultánní analýza latentních tříd rozšiřuje tyto poznatky o možnost srovnávat latentní struktury v datech z několika různých výzkumů. Pokud tedy máme k dispozici identická měření provedená ve dvou či více skupinách a cílem je jejich vzájemná komparace, simultánní analýza latentních tříd představuje pro tuto úlohu vhodný nástroj. Stejně tak užitečným nástrojem může tato metoda být, pokud se jedná o měření stejné populace v různých časových bodech, tedy o longitudinální studii s cílem sledovat určité trendy. [McCutcheon 1987: 61]

Model latentních tříd pro více skupin obsahuje tři typy kategoriálních proměnných – sadu pozorovatelných a měřených indikátorů, jednu či více nepozorovatelných latentních proměnných a skupinovou proměnnou (v angličtině tzv. *grouping variable*), která je kategoriální povahy a může být asociována jak s měřitelnými indikátory, tak s latentní proměnnou. Model dále vychází z faktu, že pro skupinovou latentní proměnnou platí předpoklad lokální nezávislosti proměnných stejně jako pro proměnnou latentní. Simultánní analýza latentních tříd poté studuje ekvivalenci měření napříč skupinami. Jinými slovy se tedy snaží odpovědět na otázku, do jaké míry jsou položky odpovědí stejné napříč skupinami v závislosti na příslušnosti jednice k latentní třídě. K tomuto účelu může být zvolena pravděpodobnostní nebo loglineární parametrizace, které byly popsány v samostatné kapitole 1.4.1.6. [Kankaraš, Vermunt 2014: 1-2] Častěji užívanou volbou je pravděpodobnostní parametrizace, kdy je požadováno, aby se součet pravděpodobností všech latentních tříd rovnal 1.

Simultánní analýza latentních tříd může vyústit v několik různých výsledků. Prvním, extrémním, výsledkem je situace, kdy se skupiny vzájemně liší jak počtem, tak i povahou latentních tříd. V takové chvíli mluvíme o zcela heterogenních skupinách, které z podstaty nemohou mít stejnou latentní strukturu. Druhou možností výsledku jsou tzv. *neomezené heterogenní modely s t-třídami*. V latentní struktuře odlišných skupin byl nalezen stejný počet tříd, aniž by byly omezeny skupinové parametry (např. podmíněné pravděpodobnosti a pravděpodobnosti latentních tříd). V takové chvíli je třeba zvážit, zda struktura tříd identifikovaných v jedné skupině odpovídá struktuře tříd skupiny druhé. Jinými slovy se ptáme, zda ty stejné podmíněné pravděpodobnosti platí pro všechny srovnávané skupiny. Třetím případem jsou tzv. *částečně homogenní*

skupiny, které se vyznačují tím, že restrikce rovnosti může být použita na některé z podmíněných pravděpodobností u všech zkoumaných skupin. V závislosti na počtu uložených restrikcí poté můžeme nahlížet latentní strukturu těchto skupin jako vzájemně podobnou. Posledním zjištěním, ve které může simultánní analýza latentních tříd vyústit, je pochopitelně případ *homogenních skupin*. Tato situace nastává ve chvíli, kdy je možné použití restrikce rovnosti u všech parametrů, což jinými slovy znamená, že latentní struktura všech skupin je stejná. [McCutcheon 1987: 61-62]

1.4.5 Srovnání faktorové analýzy a analýzy latentních tříd

Jak již bylo naznačeno výše, analýza latentních struktur, potažmo tedy i analýza latentních tříd, je v principech svého fungování velmi blízká faktorové analýze, proto bude jejich komparaci věnována tato kapitola.

Cíl faktorové analýzy spočívá stejně jako cíl analýzy latentních tříd ve snaze aproximovat datovou matici, která do analýzy vstupuje, maticí nižšího řádu. Zatímco u faktorové analýzy se jedná o matici korelační, u analýzy latentních tříd se nám jedná o matici relativních čteností, tedy pravděpodobností. [Green 1952]

Jak faktorová analýza, tak i analýza latentních tříd mohou mít explorační i konfirmační podobu. Při exploračním modelu vysvětlují latentní struktury v rámci sady pozorovatelných proměnných, zatímco ve své konfirmační podobě testují hypotézy o latentních strukturách u těchto proměnných. [McCutcheon 1987:8 in Jeřábek, Soukup 2008: 96]

Co obě analýzy naopak odlišuje, je povaha faktorů či tříd, které z analýzy vycházejí. Faktory ve faktorové analýze jsou spojitě proměnné s nekonečným počtem kategorií. Všechny skryté faktory jsou měřeny v intervalech a všechny vztahy mezi faktory samotnými a pozorovanými proměnnými musí mít lineární charakter. Analýza latentních tříd oproti tomu neklade žádné předpoklady o podobě vztahu mezi manifestními a latentními proměnnými. Nepozorovatelné latentní proměnné jsou poté kategoriální povahy. Jinými slovy, zatímco faktorová analýza vytváří kontinuální faktory, analýza latentních tříd tvoří diskrétní latentní třídy. [Hagenaars, Halman 1989: 82]

Ve srovnání s klasickým shlukováním přináší analýza latentních tříd také jeden velmi podstatný rozdíl. Konvenční shlukovací metody jsou ze své podstaty tzv. alokačními technikami. To znamená, že každému případu je přiřazena hodnota 1 nebo 0,

podle toho, zda do shluku spadá či nikoli. Analýza latentních tříd oproti tomu staví třídy na základě pravděpodobnosti. Třídy tak mají fuzzy podobu, kdy každému případu je přiřazena procentuální šance, s jakou do dané třídy spadá. Objekty tak nejsou přiřazeny pouze do jedné třídy, ale vyjadřujeme u nich pravděpodobnost, s jakou spadají do každé z vytvářených tříd. Tento rozdíl ve výsledku znamená, že alokační metody budou vždy podceňovat variabilitu uvnitř jednoho shluku, a naopak přeceňovat odlišnost a vzdálenost mezi jednotlivými klastry. [Grover, Vriens 2006: 534-535]

1.4.6 Dosavadní případové studie

Krátká kapitola o dosud publikovaných případových studiích si klade za cíl ukázat, že analýza latentních tříd bývá jako segmentační metoda používána v celé řadě oborů jak humanitních, tak i přírodovědných.

Britští sociologové Savage, Devine, Cunningham a další spoluautoři aplikovali analýzu latentních tříd jako výzkumnou metodu ve studii zkoumající sociální třídy, jejich postavení a vzájemné odlišnosti. Studie „*A New Model of Social Class? Findings from the BBC's Great British Class Survey Experiment*“ [Savage, Devine, Cunningham a kol. 2013] zkoumá různá kritéria pro definování sociálních tříd. Výsledkem studie je sedmitřídní multidimenzionální model sociálního rozdělení a polarizace v Británii. Z pohledu této diplomové práce je však přednější metodologická stránka studie než její věcné závěry. Autoři ve studii hledají způsob, jak šetrně odlišit mezi ekonomickými ukazateli a kulturním a sociálním kapitálem jedince, aby bylo správně odhadnuto, kde leží hranice jednotlivých sociálních tříd. Svůj model latentních tříd stavějí na myšlence, že některé parametry statistických modelů odlišují nepozorovatelné podskupiny, které vytvářejí jednotlivé kategorie kategoriální latentní proměnné. [Savage, Devine, Cunningham a kol. 2013: 228-229]

Ve studii je standardizováno šest proměnných, které podle autorů specifikují sociální postavení jedince (status sociálních kontaktů, celkový počet kontaktů, dva druhy kulturního kapitálu, příjem jedince a výše majetku) a na jejich základě je vytvářen model latentních tříd. Aby byl překonán problém šikmosti dat, kombinují autoři data z několika zdrojů a používají váhy k dorovnání některých kategorií. Beyesovské informační kritérium (BIC) v explorační analýze latentních tříd ukázalo jako nejvhodnější model se sedmi latentními třídami, který byl i věcně vyhodnocen jako vyhovující. Vzniklé proporce jednotlivých tříd jsou relativně vyrovnané, žádná třída

není oproti ostatním výrazně málo zastoupena, což autoři považují za další znak správnosti modelu. Třídní příslušnost jedince je následně korelována se sociodemografickými charakteristikami ve snaze hlouběji popsat vzniklé třídy. [Savage, Devine, Cunningham a kol. 2013: 229-234]

Zajímavou studií v medicínsko-psychologické oblasti je také studie s názvem *The „Structure of Posttraumatic Stress Disorder“* [Breslau a kol. 2005], která se zabývá tím, jak mohou být pacienti s posttraumatickou stresovou poruchou rozřazeni do skupin podle jejich symptomového profilu. Cílem výzkumu tak byla segmentace pacientů za účelem vytvoření klinicky homogenních skupin, jejichž znalost by mohla usnadnit další vědecké výzkumy v této oblasti. Zajímavé na této studii je, že na rozdíl od ostatních zmíněných případových studií pracuje s daty sesbíranými ve dvou komunitních skupinách a pracuje tedy se dvěma odlišnými populacemi. Každý ze dvou vzorků byl však v tomto případě testován samostatně procesem systematické replikace. Na základě sedmnácti zkoumaných symptomů bylo dokázáno, že souvislosti mezi měřitelnými symptomy jsou ovlivněny latentní strukturou třídní klasifikace na pozadí. Informace o této struktuře byly vyhodnoceny na základě výsledků o pravděpodobnosti přináležení do latentní třídy a velikosti jednotlivých tříd. Výsledky byly interpretovány pro každou ze skupin zvlášť, ačkoli u obou dvou skupin vyšel nejpřesvědčivěji model se třemi latentními třídami, které si co do věcného charakteru byly mezi skupinami velmi podobné.

Třetí zajímavou případovou studií je studie z oblasti marketingu s názvem *A latent class segmentation analysis of e-shoppers* [Bhatnagar, Ghose 2004], jejímž cílem je segmentace zákazníků nakupujících na internetu. Segmentace je v tomto případě postavena na datech o nákupním chování jedinců napříč mnoha kategoriemi zboží. Autoři se snaží ukázat, že analýza latentních tříd může v tomto případě přinést mnoho benefitů oproti klasickému profilování postavenému na sociodemografických charakteristikách. Kromě informací o nákupním chování uskutečnili autoři v rámci studie dotazníkové šetření, které bylo zaměřeno na hodnocení jednotlivých nákupních webů. Tímto způsobem byla analýza latentních tříd obohacena o baterii otázek se škálou souhlasu Likertova typu. Respondentům byl také předložen seznam kategorií a byli tázáni, který druh zboží nakupují na internetu a zda je pro ně tento druh zboží důležitý. [Bhatnagar, Ghose 2004: 760-764]

Analýza latentních tříd pro vzorek 1330 respondentů bez nastavení jakýchkoli parametrů a omezení ukázala podle Bayesovského informačního kritéria jako

nejvhodnější model se třemi latentními třídami, čehož se autoři drželi. Na základě informací o tom, jaké druhy zboží preferují u nákupu přes internet členové jednotlivých segmentů, byly poté segmenty charakterizovány. V posledním kroku byly identifikované segmenty evaluovány z pohledu sociodemografických charakteristik. Vyhodnoceny byly i vybrané otázky jednotlivě, což poskytlo autorům dobrý podklad pro další interpretace segmentů.

Ze tří uvedených případových studií vyplývá, že základními výstupy výzkumu jsou tři klíčové informace – počet, velikost a charakter latentních tříd. Jakkoli může tato informace znít triviálně, považují ji za důležitou pro vlastní případovou studii, neboť žádné z těchto kritérií nesmí být opomenuto.

2 Metodologická část

Tato část textu je zaměřena na vymezení případové studie, která byla v rámci této diplomové práce realizována. Postupně budou definovány výzkumné otázky a hypotézy, popsána analyzovaná data a konkrétní položky vybrané k analýze, zvolen analytický postup či okomentován software, ve kterém byla data zpracována.

2.1 Výzkumné cíle, otázky a hypotézy

Práce se zabývá analýzou latentních tříd jako jednou z možných metod segmentace. K tomuto účelu jsem si určila dva základní cíle práce – metodologický a věcný. Prvním cílem je ukázat různá segmentační řešení postavená na variantách této statistické metody, poukázat na její silné a slabé stránky, diskutovat případné rozdíly mezi statistickým řešením a věcnou interpretací a v neposlední řadě vybrat nejvhodnější řešení. Dané segmenty budou poté do hloubky interpretovány.

Srovnány budou celkem tři možné přístupy k práci s daty vzešlymi z několika vln výzkumu v rámci analýzy latentních tříd – analýza pro spojená data ze všech vln, simultánní analýza pro jednotlivé vlny a analýza pro spojená data s dichotomickými proměnnými.

Pro dosažení tohoto cíle práce jsem si definovala základní výzkumnou otázku, od které budou odvozeny hypotézy a související výzkumné otázky pro analytickou část práce. Výzkumná otázka zní takto:

Jaké jsou možnosti analýzy latentních tříd jako segmentační metody?

Aby mohly být jednotlivé varianty analýzy latentních tříd zhodnoceny z hlediska jejich praktického užití, bude významnou část práce tvořit případová studie. Ta se ve své tematické rovině zabývá obrazem České televize jako média veřejné služby. Vybrané otázky směřují k tomu, jak dotazovaní respondenti vnímají Českou televizi, nakolik v jejich očích naplňuje svoji veřejnoprávní funkci či jakými atributy disponuje. Cílem studie je segmentace televizních diváků do několika skupin na základě jejich reflexe zkoumaných oblastí. Nalezené segmenty budou poté interpretovány, detailně charakterizovány a budou vytyčeny stěžejní rozdíly v diváckých profilech těchto skupin. K tomuto cíli práce se pojí druhá, věcná, výzkumná otázka:

Jaké skupiny respondentů lze na základě jejich vnímání veřejnoprávní funkce České televize ve zkoumaném vzorku identifikovat?

Na základě teoretické části práce a prozkoumání dat, která jsou detailně popsána v další kapitole, jsou definovány následující hypotézy a s nimi spojené dílčí výzkumné otázky.

H1: Latentní struktura nalezená v rámci analýzy latentních tříd pro spojená data ze všech šesti vln dohromady se ve svých výsledcích výrazně neliší od latentní struktury každé vlny samostatně.

Související výzkumné otázky: Která metoda segmentace poskytuje přesvědčivější výsledky pro data analyzovaná v rámci této případové studie? Model s kolika latentními třídami je nejvhodnější? Odchyluje se některá ze zkoumaných vln ve svých výsledcích od vln ostatních? Do kolika segmentů lze respondenty rozdělit? Jak jsou jednotlivé segmenty proporčně zastoupeny? Jaké tyto segmenty jsou z hlediska svého charakteru?

H2: Použití dichotomických proměnných oproti zachování čtyř-položkové škály souhlasu ubírá hodnotnou část popisu jednotlivých tříd, v důsledku tak mají modely odlišnou strukturu.

Související výzkumné otázky: Mají modely stejnou strukturu co do počtu, velikosti a charakteru tříd?

H3: Jednotlivé segmenty lze odlišit na základě sociodemografických charakteristik jejich členů.

Související výzkumné otázky: Lze jednotlivé segmenty charakterizovat na základě sociodemografických charakteristik jejich příslušníků? Jaké charakteristiky jsou klíčové? Které naopak nehrají žádnou roli?

2.2 Popis analyzovaných dat

Data, která jsem se rozhodla v této práci využít, pocházejí z trackingové studie, kterou pro Českou televizi realizuje TNS Aisa. Cílem tohoto výzkumu je poskytnout vedení České televize a Radě pro rozhlasové a televizní vysílání ucelenou zpětnou vazbu na vysílání od televizních diváků. Klíčovými tématy výzkumu jsou především názory diváků na jednotlivé televizní stanice, jejich divácké preference, postoje výroky týkající se imageového vnímání českých televizních kanálů a v neposlední řadě hodnocení veřejné služby České televize a jejího fungování jako média.

Výzkum probíhá od roku 2012 vždy dvakrát ročně, a to v jarní a podzimní vlně. Každá z vln je realizována na přibližně 1 000 respondentech. Následující tabulka (tabulka č.2) zahrnuje konkrétní údaje o každé z dosud realizovaných vln výzkumu:

Tabulka č. 2 – Popis vln výzkumu

Rok	Vlna	Číslo vlny	Počet respondentů	Termín realizace
2012	Podzim	1	1106	15. 11. – 2. 12. 2012
2013	Jaro	2	1000	2. 5. – 13. 5. 2013
	Podzim	3	1111	22. 11. – 15. 12. 2013
2014	Jaro	4	1003	15. 5. – 8. 6. 2014
	Podzim	5	1003	14. 11. – 3. 12. 2014
2015	Jaro	6	1008	15. 5. – 24. 6. 2015
	Podzim	7	1002	10. 11. – 14. 12. 2015
2016	Jaro	8	1003	11. 5. – 10. 6. 2016

Do analýzy dat v rámci této případové studie jsem se rozhodla zařadit pouze 3. až 8. vlnu, tedy celkem 6 vln, které proběhly mezi podzimem roku 2013 a jarem 2016. K tomuto kroku jsem se rozhodla z toho důvodu, že do prvních dvou vln nebylo zařazeno 5 výroků, které považuji za zajímavé a přínosné pro tuto případovou studii. Celkový počet respondentů v rámci analýzy je tedy 6 130.

Cílovou skupinou výzkumu je reprezentativní televizní populace ve věku 18+. Respondenti jsou v rámci každé vlny vybíráni kvótním výběrem, jehož kritéria jsou pohlaví, věk, vzdělání, velikost místa bydliště a kraj, ve kterém respondent žije. Takto zvolené kvóty odpovídají rozložení obyvatelstva v populaci na základě těchto kritérií. Zdrojem pro určování kvót jsou data Českého statistického úřadu.

Data byla poté sbírána nejprve metodou CATI (Computer Aided Personal Interviewing), v pozdějších vlnách metodou TAPI (Tablet-Assisted Personal Interviewing) dotazování. V rámci dotazovacích situací je v těchto metodách přítomen tazatel, který respondentovi předčítá předem připravené otázky a zaznamenává jeho odpovědi. Rozdíl mezi těmito metodami je pouze v použití počítače či tabletu jako záznamového prostředku, z hlediska metodologie výzkumu je tak považují za totožné.

Průměrná délka dotazování se v jednotlivých vlnách pohybovala mezi 25-35 minutami v závislosti na délce konkrétního dotazníku. V pozdějších vlnách přibylo v dotazníku oproti vlnám dřívějším několik otázek či byly měněny otázky v ad-hoc části výzkumu sestavovaného na základě aktuálního dění ve společnosti.

Následující tabulka (tabulka č.3) obsahuje shrnutí research designu případové studie:

Tabulka č. 3 – Research design

Zadavatel	Česká televize
Agentura realizující výzkum	TNS Aisa
Období sběru data	Podzim 2012 – jaro 2016 2 vlny ročně, celkem tedy 8 vln
Velikost výběrového souboru	N = 8 236 (každá vlna cca 1000 respondentů)
Specifikace výběrového souboru	Reprezentativní televizní populace ve věku 18+
Druh výběru	Kvótní výběr (kritéria - pohlaví, věk, vzdělání, velikost místa bydliště a kraj)
Metoda sběru data	CATI + TAPI

2.3 Baterie otázek vybraná k analýze

Volba následující baterie otázek nebyla náhodná. Její výběr byl proveden s ohledem na to, aby jednotlivé výroky pokryly dostatečné množství tematických okruhů, které se vztahují k pozici České televize jako veřejnoprávního média, a na jejich základě mohly být identifikovány jednotlivé typy diváků. V této kapitole jsou

představeny a blíže popsány položky, které byly z dotazníku vybrány jako vhodné pro tuto případovou studii. Jedná se o baterii výroků týkajících se názoru respondentů na Českou televizi jako celek. Konkrétní výroky, které respondenti hodnotili stupněm souhlasu na 4 položkové škále, jsou vypsány v následující tabulce (tabulka č.4):

Tabulka č. 4 – Baterie otázek

<p><i>Nyní vám postupně přečtu několik výroků týkajících se České televize jako celku. U každého výroku mi prosím řekněte, do jaké míry s ním osobně souhlasíte. Vyberte vždy jednu z variant odpovědi - rozhodně souhlasím, spíše souhlasím, spíše nesouhlasím nebo rozhodně nesouhlasím.</i></p>
<p>B6_1 B6. ČT předkládá divákům pravdivý a nezkrácený obraz skutečnosti kolem nás.</p> <p>B6_2 B6. ČT mi pomáhá porozumět světu, ve kterém žijeme.</p> <p>B6_3 B6. Ve vysílání ČT dostávají v dostatečné míře prostor různé názory a úhly pohledu na jednotlivá témata.</p> <p>B6_4 B6. Informace předkládané ČT divákům jsou objektivní a vyvážené.</p> <p>B6_5 B6. ČT významně přispívá k rozvoji vzdělanosti svých diváků.</p> <p>B6_6 B6. ČT nabízí řadu pořadů, ze kterých se něčemu přiučím.</p> <p>B6_7 B6. ČT významně přispívá k tomu, aby se občané dokázali orientovat v oblasti práva, platných zákonů a norem.</p> <p>B6_8 B6. ČT dokáže přimět diváky k tomu, aby se aktivně podíleli na různých kulturně-spoločenských aktivitách.</p> <p>B6_9 B6. ČT má mezi televizemi vedoucí postavení v oblasti kulturních pořadů.</p> <p>B6_10 B6. ČT nabízí dostatek pořadů, které odpovídají mému vkusu a mým zájmům.</p> <p>B6_11 B6. ČT dává ve svém vysílání dostatečný prostor pro prezentaci kultury a života národnostních a etnických menšin v ČR.</p> <p>B6_12 B6. Při sledování ČT si dokážu dobře odpočinout.</p> <p>B6_13 B6. Při sledování ČT se dokážu dobře pobavit.</p> <p>B6_14 B6. Špičkové domácí a zahraniční sportovní události dostávají ve vysílání ČT patřičný prostor.</p> <p>B6_15 B6. Vysílání ČT jako celek přispívá k budování a posilování dobrých vztahů mezi lidmi a v celé společnosti.</p> <p>B6_16 B6. Vysílání ČT jako celek napomáhá k národnosti a etnické toleranci.</p> <p>B6_17 B6. Vysílání ČT jako celek napomáhá k náboženské toleranci.</p> <p>B6_18 B6. ČT je nezávislá na ekonomických a politických vlivech.</p> <p>B6_19 B6. ČT přináší divákům i pořady, které komerční TV nevysílají.</p> <p>B6_20 B6. Česká televize je důvěryhodná.</p> <p>B6_21 B6. Česká televize je inovativní.</p> <p>B6_22 B6. ČT navenek dostatečně prezentuje, jaké pořady a pro jaké skupiny diváků na svých kanálech nabízí.</p> <p>B6_23 B6. ČT prezentuje ve svém vysílání pro děti hodnoty slušnosti, vzdělanosti, úcty k národnostním menšinám a k životnímu prostředí.</p>

Pro lepší pochopení výroků bych výroky pomyslně rozdělila do tří tematických kategorií. První kategorii označuji jako „způsob prezentace skutečnosti“. Spadají sem výroky týkající se objektivity a nezkrácenosti informací uveřejňovaných Českou televizí či názory na její nestrannost a ovlivnitelnost. Druhou kategorii bych definovala jako „porozumění světu“. Tato skupina výroků se zabývá otázkou, zda Česká televize

prezentuje názory vyváženou formou, poskytuje informace z různých oblastí či zda prostřednictvím svého vysílání vzdělává diváky. Do třetí kategorie „kultura, tolerance a vztahy mezi lidmi“ poté spadají výroky zaměřené na národnostní a etnickou toleranci, rozšiřování obecného kulturního povědomí či posilování dobrých vztahů mezi lidmi. Předpokládám, že typologie diváků vytvořená v analytické části práce bude tyto oblasti odrážet, a to tím způsobem, že se jednotlivé typy diváků od sebe budou lišit v kategorii, kterou jejich příslušníci nejvíce nebo naopak nejméně kritizují.

2.4 Etika výzkumu

Povaha výzkumného tématu, tedy postoj k České televizi jako k médiu veřejné služby, není nikterak osobního či citlivého charakteru. Baterie otázek mapující tuto oblast proto nemusí být zvláště uzpůsobována podle zásad a doporučení pro práci s citlivými tématy. Odhlédneme-li však od této konkrétní baterie a zaměříme-li se na dotazník jako celek, zjistíme, že poslední přibližně čtvrtina dotazníku zjišťuje politické preference respondenta, které slouží České televizi jako podklady pro tvorbu volebního modelu. Otázky zabývající se politickými názory a postoji již mohou zcela jistě představovat výraznou bariéru pro respondenta a ovlivnit nejen jeho ochotu odpovídat, ale i pravdivost jeho výroků. Z tohoto důvodu je politicky zaměřená část výzkumu zařazena až na konec dotazníku, kdy je počítáno s tzv. „zeptej se a uteč“ přístupem. Vzhledem k tomu, že tato část dotazníku není předmětem této případové studie, nebudou zde dále rozebírány problémy sociální desirability či přístupy k práci s citlivými tématy. Co by však na tomto místě mělo být stručně zhodnoceno, jsou obecné etické zásady výzkumu a míra jejich naplnění.

Jak již bylo uvedeno výše, analyzovaná data pocházejí z výzkumu, který pro Českou televizi realizuje agentura TSN Aisa. Agentura TNS Aisa je členem asociace SIMAR, tedy sdružení agentur pro výzkum trhu, které dle svých základních ustanovení dohlíží na kvalitu dat a dodržování obecně uznávaných etických zásad při jejich sběru a práci s nimi. Důraz je tak kladen na důvěrné zacházení s osobními údaji, anonymitu respondentů, umožnění adekvátní kontroly výzkumů, zodpovědnou interpretaci dat či zveřejnění informací o zadavateli výzkumu. Členství ve sdružení SIMAR dále dokládá, že TNS Aisa dodržuje ve své práci zásady vydané mezinárodní asociací ESOMAR. [Simar 2016]

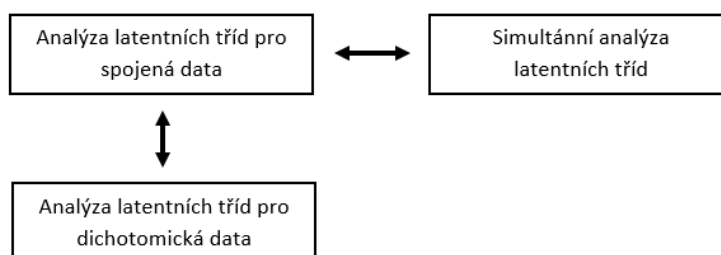
2.5 Volba analytického postupu

Jak již bylo uvedeno v kapitole věnované metodologickému popisu dat, pro účely své práce jsem se rozhodla nevyužít nasbíraných dat ze všech osmi realizovaných vln, ale omezit se pouze na šest vln, které proběhly mezi podzimem roku 2013 a jarem 2016. S daty z šesti různých vln lze nakládat dvěma možnými způsoby – analyzovat každou vlnu zvlášť prostřednictvím simultánní analýzy latentních tříd, nebo data ze všech vln spojit do jednoho souboru.

Cílem této práce je srovnat výsledky obou těchto variant, data tedy budou analyzována jak ve spojené formě, tak ve variantě pro každou vlnu zvlášť. Důležité je zde však upozornit na rozdíl mezi srovnáváním výsledků z jednotlivých vln a longitudinálních studií. Pokud by bylo cílem v práci sledovat trendy, tedy vývoj a případné změny v charakteru latentních tříd napříč vlnami, musela by populace zůstat u všech vln stejná. V rámci tohoto výzkumu však respondenti nebyli dotazováni opakovaně, ale pro každou vlnu byli vybíráni respondenti noví. V rámci této sekundární analýzy dat proto není možné realizovat longitudinální výzkum stejné skupiny jedinců po delší časové období.

Třetím typem analýzy, která bude v rámci této práce provedena, je analýza pro spojená data přepočítaná do dichotomického charakteru proměnných. Jedná se tedy o období první analýzy pouze s tím rozdílem, že u jednotlivých otázek bude čtyř-položková škála stupně souhlasu nahrazena svoji dvou-položkovou variantou. Srovnávání analýzy pro spojená data a simultánní analýzy latentních tříd sleduje přirozeně jiné cíle, než srovnání práce s více-položkovou škálou odpovědí či dichotomickými proměnnými. Analýzy tak nebudou srovnávány každá s každou, ale srovnání bude provedeno dle následujícího schématu (obrázek č.2):

Obrázek č.2 – Schéma srovnávání analýz



Jak by mělo být patrné z obrázku, sledovány jsou dva odlišné koncepty – analýza latentních tříd pro spojená data versus simultánní analýza latentních tříd a analýza

latentních tříd proměnné se čtyř-položkovou škálou odpovědí versus obdoba této analýzy pro přepočítaná dichotomická data.

2.6 Použitý software

Data jsou analyzována ve dvou statistických softwarech – SPSS a Mplus. Zatímco SPSS je poměrně tradiční a hojně využívaný statistický program, Mplus je, využíváno spíše okrajově. Jedná se o program pro statistické modelování, který nabízí širokou škálu modelů, odhadů a algoritmů. Program původně vzniknul pro strukturní modelování, postupně byl však rozšířen o celou řadu dalších procedur. Velké výhody užití programu Mplus oproti SPSS přichází ve chvíli, kdy je cílem analýzy pracovat s daty na jejich latentní úrovni, neboť program Mplus je pro práci s latentními proměnnými přímo určen a nabízí širší škálu možností než SPSS. Vytvářet statistické modely v Mplus lze za skrze kontinuální latentní proměnné, kategorické latentní proměnné či kombinaci obou dvou. [Soukup 2015]

V rámci této práce jsem se rozhodla program Mplus využít pro vytváření modelů analýzy latentních tříd. Protože Mplus vyžaduje velmi specifický formát vstupních dat a nabízí jen omezené možnosti jejich úpravy, i v těchto analýzách bylo pro základní procedury využito SPSS, ze kterého byl až následně generován soubor vstupních dat pro Mplus.

Za jednu z klíčových výhod Mplus oproti SPSS považuji fakt, že Mplus je schopné pracovat se všemi typy proměnných a neklade příliš vysoké nároky na použitou techniku. Z praktického pohledu je poté zásadní vysoká i rychlost, s kterou je Mplus schopné analýzy zpracovat. Zásadní slabinou Mplus jsou popisky proměnných a kategorií dat, které nelze definovat ve vstupním datovém souboru, ale musí být nastaveny až v rámci použité syntaxe, což pro mě osobně znesnadňuje počáteční orientaci v datech a přidělová práci. [Soukup 2015]

2.6.1 Analýza latentních tříd v programu Mplus

Základní podmínkou analýzy latentních tříd v programu Mplus je specifikace smíšeného modelu, který je v Mplus označován jako *type=mixture*. [Geiser 2013: 235] Smíšené modelování zde odkazuje na model s kategoriální latentní proměnnou, kde latentní proměnná reprezentuje subpopulaci, přičemž příslušnost případů ke konkrétním

třídám není známa, ale je usuzována z dat. [Muthén, Muthén 2010: 47] Cílem analýzy latentních tříd je „odmísit“ tyto subpopulace a vytvořit z nich homogenní skupiny. [Geiser 2013: 237]

Jak již bylo uvedeno výše, program Mplus vzniknul primárně pro účely strukturního modelování, na které jej považují za vhodný, nicméně pro další pomocné analýzy, jakými jsou například závislosti příslušnosti k latentní třídě na sociodemografických charakteristikách, je nutné vrátit data zpět do SPSS. Z tohoto důvodu je dalším důležitým příkazem, který musí být v syntaxi nastaven, příkaz *variable: auxiliary = Id*. Díky tomuto pokynu vytvoří Mplus nový soubor dat, kde bude pro každého respondenta uložena jako proměnná příslušnost k latentní třídě. Ta je počítána na základě odhadů parametrů modelu, tedy pravděpodobnosti, s jakou jedinec spadá do konkrétní latentní třídy. Proměnná *id* je zde v roli pomocné proměnné, která sama nevstupuje do analýzy, ale identifikuje respondenta. V syntaxi je dále třeba specifikovat typ jednotlivých proměnných, počet latentních tříd či nastavit proměnnou, podle které budou data převážena. [Geiser 2013: 235]

Jak již bylo rozebráno v teoretické části práce, dalším kritériem, které je třeba v rámci analýzy latentních tříd vzít v úvahu, je počet provedených iterací. V Mplus je defaultně nastaven počet náhodných počátečních hodnot generovaných programem parametrů na 10. To je poměrně malé číslo, které je potřeba navýšit, a to především pro komplexnější modely s přibližně třemi a více třídami. Jako relativně bezpečné nastavení se podle Geisera ukazuje alespoň 500 počátečních hodnot. Příkaz *starts = 500 50* naprogramuje Mplus, aby v prvním kroku optimalizace použilo 500 počátečních hodnot místo defaultně nastavených 10 a aby vybralo 50 z nich, které dosahují nejvyšší loglikelihood hodnoty. Ve druhém kroku optimalizačního procesu je poté znovu použito pouze těch 50 hodnot. Za výsledné řešení je poté považováno to s nejvyšší loglikelihood hodnotou. [Geiser 2013: 237-239] Stejně tak je dobré zvýšit počet iterací pro prvotní fázi alespoň na 50 z defaultně nastavených 10.

Pro usnadnění výběru vhodného modelu existuje celá řada informačních kritérií. Mplus rozlišuje mezi kritérii odhadujícími absolutní a relativní přesnost. Absolutní přesnost (*absolute model fit*) vyjadřuje nakolik model vystihuje zkoumaná data, relativní přesnost se poté ptá, jak dobře si model stojí ve srovnání s ostatními modely. Pro odhad absolutní přesnosti používá Mplus dvou ukazatelů – *likelihood ratio* (LR) a *Pearsonův X^2 test*. Problém je, pokud je v datech příliš mnoho případů a relativně málo

položek, či naopak příliš mnoho položek na malý vzorek. V takovém případě mluvíme o *řidkosti dat*. [Geiser 2013: 227-229]

3 Analytická část

V první části práce byl představen teoretický rámec, ze kterého tato diplomová práce vychází. Metodologická část byla zaměřena na vymezení případové studie, definování základních výzkumných otázek a hypotéz. Třetí, analytická, část textu z výše popsané teorie a metodologie vychází a klade si za cíl představit výsledky provedených analýz.

3.1 Základní popis souboru

Prvním krokem analytické části práce je prozkoumání základních deskriptivních statistik. Všechny základní statistické testy jsem provedla v programu SPSS. Tabulka č. 5 ukazuje důležitou informaci, že z celkovým 6 130 případů jich může být v některých případech do analýzy zařazena pouze přibližně polovina, konkrétně 3 354 případů. Tento úbytek je způsoben častým výskytem odpovědí „nevím“ či „bez odpovědi“ u jednotlivých položek baterie otázek, na které je SPSS velmi citlivé a mohou být z některých analýz vyřazeny. Další základní průzkumné statistiky ukázaly, že celkový průměr hodnot odpovědí u jednotlivých položek je 2,09, z čehož lze vyčíst, že respondenti se jako celek přirozeně klonili k názorovému středu.

Tabulka č. 5 – Deskriptivní statistiky

Deskriptivní statistiky			
		N	%
Počet případů	Validní	3354,5	54,7
	Vyřazené	2775,5	45,3
	Celkem	6130,0	100,0

3.2 Výsledky analýzy latentních tříd pro spojené vlny

Analýzu latentních tříd jsem provedla ve statistickém programu Mplus. Vzhledem k tomu, že v Mplus nelze upravovat a čistit datový soubor, základní úkony jsem provedla v SPSS, kde jsem si data připravila a až poté z nich vytvořila datovou matici určenou pro Mplus. Abych program zbytečně nezatěžovala a pracoval rychleji, ve výsledné matici jsem ponechala pouze zkoumanou baterii otázek a váhu.

Ačkoli jsem již měla několik předpokladů o podobě latentních tříd, vyzkoušela jsem pro úplnost postupně modely s dvěma až osmi latentními třídami. Následující tabulka (tabulka č.6) zobrazuje hodnoty kritérií AIC a BIC pro jednotlivé modely. Samy o sobě nemají hodnoty těchto kritérií žádnou vypovídající hodnotu, klíčové jsou však pro statistické srovnání vhodnosti jednotlivých modelů. Obecně platí, že čím nižší je hodnota AIC a BIC daného modelu oproti modelům ostatním, tím je model vhodnější. Tato kritéria je však nutné brát pouze jako orientační statistický ukazatel a vždy je důležité i věcně zvážit logiku a smysluplnost utvořených latentních tříd.

Tabulka č. 6 – informační kritéria

Informační kritéria – srovnání modelů		
	AIC	BIC
2 třídy	268 819	269 753
3 třídy	255 884	257 289
4 třídy	252 112	253 998
5 tříd	249 224	251 570
6 tříd	247 430	250 246
7 tříd	246 191	249 478
8 tříd	249 997	249 754

Rozhodnutí o vyhovujícím počtu latentních tříd je zcela nezbytné pro shrnutí a smysluplné vysvětlení rozdílů mezi jednotlivými skupinami. Obecná tendence by měla být hledat model s co nejnižším počtem latentních tříd smysluplně vysvětlujících data. [Geiser 2013: 240] Tomu odpovídá i požadavek na velikost jednotlivých tříd, kdy jedním z cílů analýzy latentních tříd je nalézt nejmenší možnou velikost latentní třídy, která adekvátně popisuje data. [Breslau a kol. 2005: 1345] Shodně dle AIC a BIC kritérií je nejvhodnější model se sedmi latentními třídami. Po věcném prozkoumání vzniklých segmentů jsem však došla k závěru, že nejvhodnější je model se čtyřmi třídami, který dle mého názoru dostatečně pokrývá názorové spektrum, ale zbytečně neštěpí respondenty do velkého počtu skupin. Tři třídy, které jsou v modelu se sedmi latentními třídami navíc podle mě nepřinášejí žádné výrazně nové charakteristiky, a i co do počtu jedinců jsou poměrně málo zastoupené. Dle informačních kritérií AIC a BIC můžeme také vidět, že zatímco mezi třemi a čtyřmi třídami je hodnota snížena relativně výrazně, u dalších modelů jde již o zlepšení přibližně poloviční. Řídíme-li se doporučením z dosavadní literatury [např. Burnham, Anderson 2004: 271 in Soukup 2010: 97], rozdíl na úrovni dvou jednotek je zanedbatelný, rozdíl mezi přibližně čtyřmi a sedmi jednotkami již značí významnější posun a rozdíly nad deset jednotek ukazují na

jednoznačnou statistickou preferenci modelu s nižší AIC a BIC hodnotou. Dále bude tedy pracováno s modelem čtyř latentních tříd.

Následující tabulka (tabulka č. 7) zobrazuje hodnoty jednotlivých informačních kritérií modelu se čtyřmi latentními třídami. Log-likelihood hodnota -125 777 je stejná pro všech 50 nejlepších počátečních hodnot, což značí statisticky dobrý model a ukazuje na replikovatelnost řešení napříč různými počátečními hodnotami. Pro model se čtyřmi třídami byl tedy použit dostatečný počet replikací a riziko problému lokálního maxima bylo překonáno. Druhým důležitým ukazatelem je dosažená hodnota entropie, která značí kvalitu klasifikace na škále od nuly do jedné. Čím blíže je hodnota jedné, tím větší přesnosti klasifikace dosahuje. Dosažená hodnota 0,878 tedy může být považována za dobrý výsledek. Výsledná klasifikace je relativně kvalitní a přesná. Třetím klíčovým ukazatelem je průměrná hodnota pravděpodobnosti, s jakou je respondent přiřazen do dané latentní třídy. Hodnota 1 značí vysokou reliabilitu vytvořené klasifikace a jedinci jsou tedy do příslušné latentní třídy přiřazeni s průměrně vysokou mírou pravděpodobnosti. Doporučená minimální hodnota je podle Geisera [Geiser 2013: 248] 0,8. Dosažené výsledky tedy mohou být považovány za dobré.

Tabulka č. 7 – Shrnutí informačních kritérií I.

Informační kritéria modelu se 4 latentními třídami		
Log-likelihood – hodnota H0	-125 777,47	
AIC	252 112,94	
BIC	253 988,08	
Entropie	0,878	
Průměrná pravděpodobnost příslušnosti k latentní třídě	Třída č. 1	0,961
	Třída č. 2	0,915
	Třída č. 3	0,928
	Třída č. 4	0,919

Následující tabulka (tabulka č.8) zobrazuje procentuální zastoupení respondentů v jednotlivých latentních třídách. Vidíme, že třetí segment je výrazně menší než zbylé tři segmenty, avšak zahrnuje pod sebe 10 % respondentů, což není nikterak zanedbatelný podíl.

Tabulka č. 8 – Podíl respondentů v latentních třídách

Procentuální rozložení respondentů v jednotlivých třídách	
Latentní třída	Podíl respondentů v %
1	28
2	32
3	10
4	30

Klíčovým výstupem analýzy latentních tříd pro segmentaci respondentů je tabulka odhadnutých pravděpodobností daných odpovědí při přináležení do určité latentní třídy. Pro ilustraci výsledků všech tří srovnávacích metod a jejich možnou komparaci byla vytvořena tabulka, resp. čtyři totožné tabulky – každá pro jednu latentní třídu, která obsahuje výsledky všech tří analýz. Tabulka je pro svoji velikost umístěna v příloze (příloha č. 1). Ve sloupcích jsou uvedeny výsledky pro jednotlivé typy analýzy. Řádky tabulky zobrazují odhadnuté pravděpodobnosti daných odpovědí při přináležení do určité latentní třídy. Pro usnadnění orientace v tabulce je u každé odpovědi umístěna šipka značící četnost dané odpovědi pro každý segment. Šipka směřující nahoru značí, že daná pravděpodobnost má hodnotu vyšší nebo rovno 67 % a nachází se tedy v horní třetině. Vodorovná šipka značí pravděpodobnost spadající do rozmezí 33–67 %, tedy střední hodnoty a šipka směřující dolů je znakem pro spodní třetinu pravděpodobností odpovědí, tedy hodnoty nižší než 33 %. Tabulka je k možnému nahlédnutí v příloze a zde budou prezentovány pouze z ní vzešlé závěry.

3.3 Výsledky simultánní analýzy latentních tříd pro vlny jednotlivě

Prvním krokem analýzy je volba počtu latentních tříd, se kterými má model počítat. Vzhledem k tomu, že se v tomto případě jedná o modelování s totožnými daty jako v předchozí analýze, a to pouze s tím rozdílem, že v předešlé variantě byla data za jednotlivé vlny spojena dohromady, zatímco u této varianty jsou zachovány informace o zařazení každého případu do jedné z šesti vln, nepředpokládám, že by se struktura vytvořeného modelu lišila co do počtu nalezených latentních tříd. Z tohoto důvodu byl hned na počátku analýzy vybrán model se čtyřmi latentními třídami.

Hodnoty informačních kritérií pro tento model jsou uvedeny ve shrnující tabulce. Log-likelihood hodnota je v tomto případě -134 331. Je to tedy nejlepší hodnota, která

se objevila ve výpisu log-likelihood hodnot pro různé sady počátečních hodnot. V teoretické části práce byla jedna podkapitola věnovaná počátečním hodnotám parametrů, iteracím a jejich doporučenému počtu. Obecně platí, že je dobré zvýšit počet iterací pro prvotní fázi alespoň na 50 z defaultně nastavených 10. U předchozí analýzy pro spojená data byl tento počet iterací dostatečný a 50 replikací stačilo k tomu, aby bylo překonáno riziko lokálního maxima a log-likelihood hodnota byla stejná pro všech 50 nejlepších počátečních hodnot. Model, který je počítán v rámci této analýzy je o poznání složitější, neboť v sobě u každého respondenta kombinuje jak příslušnost k latentní třídě, tak i příslušnost k vlně, v rámci které byl respondent dotazován. Z tohoto důvodu se počet iterací 50 ukázal jako příliš malý a hrozilo, že vypočítaná hodnota log-likelihood je pouhým lokálním, nikoliv celkovým maximem. Analýza tak byla znovu přepočítána, a to s počtem iterací nastavených na 100. Tento počet je již dostatečný a výsledná hodnota log-likelihood -134 331 byla replikována.

Vzhledem k tomu, že v analýze latentních tříd pro spojené vlny i v její simultánní době je pracováno se stejnými daty, stejným počtem případů a stejným počtem modelovaných latentních tříd, mohou být vzájemně srovnávána informační kritéria AIC a BIC. Jak hodnoty AIC, tak i BIC, jsou o přibližně 20 jednotek nižší u modelu latentních tříd počítaném pro spojená data. Řídíme-li se obecným míněním, že pokles hodnoty o 10 jednotek již značí významný posun [např. Burnham, Anderson 2004: 271 *in* Soukup 2010: 97], vedou tato kritéria k jasné preferenci modelu vycházejícího ze spojených dat. Hodnota entropie je posuzována opačným pohledem, tedy čím vyšší je hodnota (na škále 0-1), tím vyšší přesnosti klasifikace model dosahuje. Srovnáme-li hodnotu entropie pro předešlý model 0,878 s hodnotou tohoto modelu 0,962, vidíme, že model počítaný simultánní cestou podstatně lépe sedí co do vzniklé klasifikace na zkoumaná data. Tento výsledek je však z pohledu charakteru obou metod přirozený – u simultánní analýzy vzniká klasifikace pro každou vlnu zvlášť. Výsledky jsou tedy interpretovány jednotlivě pro každou vlnu a až následně je z nich možné činit závěry o jakési „celkové podobě“ vzniklých tříd. Je tedy logické, že tento přístup lépe vystihuje daná data, než „zprůměrovaná“ klasifikace vytvořená ze spojených dat. Hodnoty informačních kritérií shrnuje následující tabulka (tabulka č.9):

Tabulka č. 9 – Shrnutí informačních kritérií II.

Informační kritéria modelu se 4 latentními třídami	
Log-likelihood – hodnota H0	-134 331,716
AIC	271 991,43
BIC	283 175,09
Entropie	0,962

V tabulce oproti předešlé analýze chybí údaje o průměrné pravděpodobnosti příslušnosti k latentní třídě. Tato statistika není v simultánní analýze latentních tříd počítána, neboť kvůli tomu, že respondenti nejsou rozděleni pouze do latentních tříd, ale také podle jednotlivých vln, by napočítaných pravděpodobností muselo být mnoho. Následující tabulka (tabulka č. 10) shrnuje procentuální zastoupení respondentů v jednotlivých třídách:

Tabulka č. 10 – Podíl respondentů v latentních třídách

Procentuální rozložení respondentů v jednotlivých třídách	
Latentní třída	Podíl respondentů v %
1	24
2	28
3	16
4	32

Procentuální rozložení respondentů v jednotlivých třídách je mírně vyrovnanější, než tomu bylo v předešlé analýze, kde nejmenší třída dosahovala 10 %. Nejméně zastoupená třída zde má 16 %, což lze považovat za dobrý výsledek a znak, že počet tříd byl zvolen správně, neboť každá třída zahrnuje dostatečný počet případů. Podrobnějšímu srovnání tří zde předkládaných analýz s ohledem na procentuální zastoupení respondentů v jednotlivých třídách bude věnován prostor v kapitole 3.5.

Nejdůležitějším výstupem analýzy latentních tříd pro segmentaci respondentů je tabulka odhadnutých pravděpodobností daných odpovědí při přináležení do určité latentní třídy, je spolu s výsledky zbylých analýz umístěna v příloze (příloha č. 1).

3.4 Výsledky analýzy latentních tříd pro dichotomické proměnné

Jako třetí možný přístup k analýze latentních tříd byla vybrána její obdoba pro dichotomická data. Data byla nejprve upravena v programu SPSS, kde byly překódovány jednotlivé odpovědi, a to tak, že možnosti „rozhodně souhlasím“ a „spíše souhlasím“ byly spojeny a nově označeny jako „ano“ a možnosti „spíše nesouhlasím“ a „rozhodně nesouhlasím“ byly sdruženy pod nově vzniklou kategorií „ne“. Z původní čtyř-položkové škály souhlasu s daným výrokiem tak vznikla dichotomická proměnná. Dále byl zvolen stejný postup analýzy latentních tříd jako u první zde předložené varianty, tedy analýzy latentních tříd pro spojená data.

U analýzy latentních tříd pro dichotomické proměnné mám na rozdíl od simultánní analýzy důvod předpokládat, že struktura nalezených tříd se může lišit. Z tohoto důvodu bylo stejně jako v případě první analýzy pro spojené vlny vyzkoušeno postupně celkem sedm odlišných řešení, každé pro jiný počet latentních tříd. Z přehledové tabulky hodnot informačních kritérií pro jednotlivé modely (tabulka č. 11) můžeme vyčíst, že pro dichotomické proměnné by statistiky nejvhodnější byl model s osmi latentními třídami. Dle dosavadní literatury, jak již bylo popsáno výše, jsou však rozdíly informačních kritérií AIC a BIC menší než dvě jednotky zanedbatelné. K výraznějšímu skoku tedy dochází pouze v rozdílu mezi dvěma či třemi latentními třídami, s každou další přidanou třídou se již model příliš nezlepšuje. Aby mohly být modely z jednotlivých typů analýzy porovnány mezi sebou, byl vybrán jako výsledný model se čtyřmi latentními třídami.

Tabulka č. 11 – Informační kritéria

Informační kritéria – srovnání modelů		
	AIC	BIC
2 třídy	122 046	122 362
3 třídy	118 239	118 716
4 třídy	117 044	117 683
5 tříd	115 951	116 751
6 tříd	115 176	116 138
7 tříd	114 736	115 858
8 tříd	114 390	115 674

Hodnoty klíčových informačních kritérií pro model s dichotomickými proměnnými a čtyřmi latentními třídami shrnuje tabulka č.12. Log-likelihood hodnota -

58 427 je stejná pro všech 50 nejlepších počátečních hodnot, což značí, stejně jako v případě modelu pro spojené vlny, statisticky dobrý model a ukazuje na replikovatelnost řešení napříč různými počátečními hodnotami. Tento model se nepotýká s problémem lokálního maxima, neboť byl použit dostatečný počet replikací. Hodnota Entropie 0,819 je i v tomto modelu relativně vysoká. Vzniklou klasifikaci tak lze považovat za kvalitní, a to i přesto, že u zde popisované analýzy pro dichotomické proměnné je hodnota Entropie nejnižší ze tří nabízených řešení. Posledním důležitým informačním kritériem pro vzniklý model je průměrná hodnota pravděpodobnosti, s jakou je respondent přiřazen do dané latentní třídy. Jak již bylo řečeno výše, doporučená minimální hodnota je dle Geisera [Geiser 2013: 248] 0,8, což hodnoty pro všechny čtyři třídy v tomto případě splňují. Obecně je tedy vysoká pravděpodobnost, že každý z respondentů byl přiřazen do správné latentní třídy, což je důležité pro zkoumání vlivu sociodemografických charakteristik na příslušnost k latentní třídě popsaného v kapitole 3.7.

Tabulka č. 12 – Shrnutí informačních kritérií III.

Informační kritéria modelu se 4 latentními třídami		
Log-likelihood – hodnota H0	-58 427,34	
AIC	117 044,69	
BIC	117 683,18	
Entropie	0,819	
Průměrná pravděpodobnost příslušnosti k latentní třídě	Třída č. 1	0,916
	Třída č. 2	0,849
	Třída č. 3	0,878
	Třída č. 4	0,926

Následující tabulka (tabulka č.13) zobrazuje procentuální zastoupení respondentů v jednotlivých latentních třídách. V modelu latentních tříd pro dichotomické proměnné jsou první a druhá třída zastoupeny o více než větším podílem respondentů než třetí a čtvrtá třída. Srovnání poměrného zastoupení jednotlivých tříd pro modely vzešlé z jednotlivých typů analýz bude dále popsáno v kapitole 3.5.

Tabulka č. 13 – Podíl respondentů v latentních třídách

Procentuální rozložení respondentů v jednotlivých třídách	
Latentní třída	Podíl respondentů v %
1	38
2	35
3	15
4	12

Klíčová tabulka odhadnutých pravděpodobností daných odpovědí při přiřazení do určité latentní třídy, je umístěna v příloze (příloha č. 1).

3.5 Srovnání tří použitých metod

V analytické části práce byla představena tři možná pojetí analýzy latentních tříd – analýza pro spojená data z šesti vln, simultánní analýza pro každou z šesti vln zvlášť a analýza pro spojená dichotomická data. V předchozích třech kapitolách byly popsány výsledky každé z analýz. Cílem této kapitoly je shrnout závěry vzešlé z jednotlivých vln a srovnat vzájemně jednotlivé přístupy. Srovnání zde bude uvedeno dvojí – analýza latentních tříd pro spojená data versus simultánní analýza latentních tříd, a poté analýza latentních tříd pro spojená data versus stejná analýza pro data dichotomická. Tyto dva pohledy vycházejí z definovaných cílů práce popsaných v metodologické části práce.

3.5.1 Analýza pro spojená data vs. simultánní analýza

První srovnání se zabývá rozdíly mezi analýzou pro spojená data a simultánní analýzou. Na základě posouzení informačních kritérií AIC a BIC vychází z dvojice zmíněných analýz jako statisticky výrazně vhodnější model pracující se spojenými daty. Z pohledu entropie vychází naopak jako lepší model vypočítaný simultánní analýzou. Jak již však bylo popsáno v kapitole 3.3, toto zjištění vychází z faktu, že model pracující se spojenými daty vytváří jednu sadu odhadnutých pravděpodobností daných odpovědí při přiřazení do určité latentní třídy, zatímco simultánní analýza vytváří sadu pravděpodobností pro každou vln zvlášť (viz. příloha č. 1). Dle hodnoty entropie se proto model vzešlý ze simultánní analýzy jeví jako vhodnější, ačkoli svou podobou znesnadňuje věcnou interpretaci výsledků.

Srovnání výsledků obou analýz z pohledu interpretace, tedy věcného popisu vzniklých segmentů, je obecně velmi obtížné, neboť neexistuje žádný přesně definovaný klíč, podle něhož by bylo možné při srovnávání postupovat. Jako hlavní metoda srovnávání proto bylo zvoleno grafické posouzení, které se jeví jako vhodné a názorné. Z bližšího pohledu na grafické vyjádření jednotlivých segmentů za pomoci šipek v tabulce latentních podmíněných pravděpodobností (příloha č. 1) je zřejmé, že po věcné stránce vykazují obě analýzy stejný charakter nalezených segmentů. Všechny analyzované vlny přirozeně nejsou co do latentních podmíněných pravděpodobností

totožné, současně však z dat není patrný žádný výraznější rozdíl jedné konkrétní vlny oproti ostatním. Ukazuje se tedy, že postoj k České televizi byl minimálně v rozmezí tří zkoumaných let relativně stabilním konceptem, kdy pro každou vlnu lze identifikovat dané typy diváků. Z tabulky je možno vyčíst, že se v rámci jednotlivých vln měnil především poměr odpovědí „rozhodně souhlasím“ a „spíše souhlasím“, tedy síla přesvědčení, avšak zařazení většiny respondentů dané třídy na pozitivní či negativní stranu škály zůstalo stejné. Graficky je toto možno posoudit na základě počtu jednotlivých typů šipek v rádcích, kdy převládající šipka ve většině případů odpovídá šipkám pro spojená data. Na tomto místě je důležité podotknout, že analýza pro spojená data není pouhým průměrem vypočítaných hodnot šesti zkoumaných vln. Proto je třeba se i ve výsledcích soustředit na celkový charakter segmentů, tedy směr na čtyř-položkové škále souhlasu, jakým jsou odpovědi orientovány, spíše než na přesná čísla.

Z celkového pohledu je charakter tříd velmi podobný a čtyři třídy nalezené v analýze pro spojená data je proto možné identifikovat i v simultánním modelu. Tento poznatek považuji za jeden z důkazů toho, že obě varianty analýzy jsou srovnatelné a obě dvě mohou být použity pro úlohy s podobnými daty, jaká jsou použita v této práci.

Odlišně v každé analýze vychází procentuální zastoupení respondentů v jednotlivých třídách. Z tabulky č. 14 je patrné, že analýza pro spojené vlny i simultánní analýza určila shodně jako nejmenší latentní třídu č.3. V poměrném zastoupení zbylých tříd se však analýzy vzájemně odlišují. Analýza pro spojená data určila jako největší latentní třídu č.2, u simultánní analýzy je největší latentní třída č. 4. Tyto dvě třídy jsou, jak bude popsáno v kapitole 3.6, názorově středovými a jsou si do velké míry podobné, proto je jejich odlišná preference pochopitelná.

Tabulka č. 14 – Srovnání podílů respondentů v latentních třídách

Procentuální zastoupení respondentů (v %)			
	Spojené vlny	Simultánní analýza	Dichotomické proměnné
1. třída	28	24	38
2. třída	32	28	35
3. třída	10	16	15
4. třída	30	32	12

Zajímavé informace pro srovnání struktury a velikosti latentní tříd v jednotlivých analýzách přináší pohled na predikci latentní třídy na úrovni jedince. Tabulka č. 15 zobrazuje počty respondentů pro každou kombinaci určení latentní třídy ve

srovnávaných analýzách. Modře označena jsou pole tabulky vyjadřující situaci, kdy byl respondent v obou analýzách zařazen do stejné latentní třídy. Zbylá pole popisují „přesun“ jedince do jiné latentní třídy. Na úrovni jedince predikovaly obě analýzy stejnou latentní třídu pro 2 804 respondentů, změna latentní třídy naopak nastala u 3 326 respondentů z celkové počtu 6 130. Predikce latentní příslušnosti jedince se tedy shodují na přibližně 46%. Z bližšího pohledu na tabulku je však patrné, že rozdílné zařazení respondenta do tříd neodporuje logice věcné interpretace. Z hlediska věcné interpretace jsou krajními třídami typologie třídy č. 1 a 3, jak bude popsáno v kapitole 3.6. Pouze ve 28 případech nastal přesun z jedné této třídy do druhé. Největší část přesunů, konkrétně 830 respondentů, proběhla z 1. třídy do 2. třídy. Interpretační těžiště obou těchto tříd se nachází na pozitivní straně škály souhlasu a respondenti tak byly zařazeni pouze do více či naopak méně vyhraněné skupiny. Druhý nejvyšší počet přesunů proběhl z 2. třídy do třídy 4., celkem se jednalo o 822 jedinců, což je pohyb mezi dvěma co do postojů středově zaměřenými skupinami.

Tabulka č. 15 – Predikce na úrovni jedince

Kontingenční tabulka – Predikce na úrovni jedince						
		Simultánní analýza				Počet případů v latentní třídě
		1. třída	2. třída	3.třída	4. třída	
Analýza pro spojená data	1. třída	624	830	11	309	1 774
	2. třída	337	799	3	822	1 961
	3.třída	17	1	585	20	623
	4. třída	460	132	384	796	1 772
Počet případů v latentní třídě		1 438	1 762	938	1 947	6 130

Ze srovnání analýzy latentních tříd pro spojená data a simultánní analýzy latentních tříd vyplývá, že pro účely této konkrétní případové studie jsou obě metody funkční a plně použitelné. Nejdůležitějším výstupem z analýzy latentních tříd pro účely segmentace zkoumané populace je informace o počtu, velikosti a věcném charakteru nalezených tříd. Obě analýzy vyústily v nalezení čtyř latentních tříd, které jsou z hlediska věcné interpretace mezi oběma analýzami srovnatelné. V čem se analýzy liší je procentuální zastoupení respondentů v jednotlivých třídách.

Vzhledem k tomu, že na základě analýzy latentních tříd je vytvářen popis jednotlivých segmentů zkoumané populace, jeví se jako praktičtější a v praxi využitelnější metoda práce se spojenými vlnami, neboť výstupem je pouze jedna sada

pravděpodobností určitých odpovědí při přiřazení do určité latentní třídy. Jednu sadu pravděpodobností lze ve výsledcích práce interpretovat důkladněji a přesněji než šest odlišných sad, které při snaze o vytvoření jedné spojené typologie nechávají velký prostor výzkumníkovi pro jeho vlastní interpretaci a pohled na problematiku. Simultánní analýza by byla vhodnější variantou, pokud by bylo cílem studie sledovat rozdíly mezi jednotlivými vlnami. Pokud je však vytvářena typologie, jako je tomu v tomto případě, není tolik důležité, zda jedna konkrétní vlna oproti ostatním vybočuje. Důležitý je celek a ten je u analýzy pro spojená data možné interpretovat přesněji.

3.5.2 Analýza pro spojená data vs. analýza pro dichotomické proměnné

Podíváme-li se na informační kritéria, mohou být přímo srovnávány pouze první dvě analýzy, tedy analýza pro spojená data a analýza simultánní. U analýzy pro dichotomické proměnné již dochází ke změně dat jejich překódováním. Data vstupující do modelu se tak mění a nemohou být srovnávána informační kritéria pro tento model a modely se čtyřpoložkovou škálou souhlasu pro každý výrok.

Hodnota entropie je v případě práce s dichotomickými proměnnými nižší než u analýzy se čtyř-položkovou škálou. Tento rozdíl je opodstatnitelný, neboť menší rozpětí škály vede k nižší diferenciaci dat a výsledná klasifikace poté nemůže dosahovat takové přesnosti, jako klasifikace postavená na delší škále zachycující charakter odpovědí respondentů do většího detailu. Podíváme-li se na sadu odhadnutých pravděpodobností daných odpovědí při přiřazení do určité latentní třídy pro tuto analýzu (příloha č. 1), můžeme vidět, že charakter jednotlivých tříd zůstal opět zachován. Poměr vypočítaných pravděpodobností na pozitivní versus negativní straně škály u analýzy pro spojená data přibližně odpovídá poměru hodnot „ano“ a „ne“ u analýzy pro data dichotomická. Pokud je v analýze pro spojená data u nějaké z otázek vyšší podíl jedné strany škály, než je tomu u ostatních otázek v dané třídě, můžeme tento rozdíl pozorovat i u dichotomických proměnných. Příkladem mohou být otázky B6_17 a B6_18 u latentní třídy č.1. Oproti ostatním otázkám je součet vypočítaných procent na negativní straně škály výrazně vyšší (22 a 23%) než je průměr zbylých otázek (7%) a výrazně vyšší je i pravděpodobnost odpovědi „ne“ u dichotomické analýzy. Obecně lze říci, že u všech zkoumaných otázek odpovídá součet odhadnutých latentních pravděpodobností jedné strany škály s přesností na 10% (ve výjimečných případech 15%) zastoupení odpovědi

„ano“ a „ne“ u analýzy pro dichotomická data. Tento poznatek mě vede k závěru, že struktura latentních tříd zůstává u obou analýz stejná. Důležité je však upozornit, že tento závěr vychází z analýzy pouze jedné sady dat a nemusí proto platit obecně.

Tabulka č. 16 zobrazuje stejně jako v předchozím případě informace o predikci latentní třídy na úrovni jedince. Na úrovni jedince predikovaly obě analýzy stejnou latentní třídu pro 3 141 respondentů, změna latentní třídy naopak nastala u 2 989 respondentů z celkové počtu 6 130. Predikce dvou srovnávaných analýz se tedy shodují na 52%. Podíváme-li se však na tabulku blíže, zjistíme, že rozdílné zařazení respondenta do tříd stejně jako v předchozím případě neodporuje logice věcné interpretace. V žádném ze zkoumaných případů nenastal přesun z 1. třídy do 3. třídy a naopak, což značí srovnatelnost modelů. Tyto dvě třídy stojí ve své interpretaci zcela proti sobě a přesun mezi nimi by ukazoval na zásadní nesoulad mezi srovnávanými analýzami. Největší část přesunů se odehrála mezi 1. a 2. třídou a 2. a čtvrtou třídou. Stejného výsledku dosáhlo i předchozí srovnání analýzy pro spojená data a simultánní analýzy, kde bylo popsáno, že tyto přesuny jsou logicky opodstatněné.

Tabulka č. 16 – Predikce na úrovni jedince

Kontingenční tabulka – predikce na úrovni jedince						
		Analýza pro dichotomická data				Počet případů v latentní třídě
		1. třída	2. třída	3.třída	4. třída	
Analýza pro spojená data	1. třída	1 325	395	0	54	1 774
	2. třída	1 102	821	0	38	1 961
	3.třída	0	32	489	102	623
	4. třída	1	846	419	506	1 772
Počet případů v latentní třídě		2 428	2 094	908	700	6 130

Mají-li však být jednotlivé segmenty diváků do hloubky interpretovány, v analýze pro dichotomické proměnné se ztrácí část informace, která slouží k detailnějšímu pochopení rozdílů mezi skupinami. V předchozí kapitole bylo popsáno, že jednotlivé vlny se od sebe liší zastoupením jednotlivých stupňů souhlasu, tedy odpovědí „rozhodně souhlasím/nesouhlasím“ a „spíše souhlasím/nesouhlasím“. Rozdíl v poměru těchto odpovědí je důležitý i pro toto srovnání. Zatímco analýza pro čtyřpoložkovou škálu tento rozdíl zachová a je následně možné jej interpretovat, dichotomická varianta jej smaže shrnutím obou odpovědí do jedné. Pozorován tak může být celkový charakter třídy, nikoli však jednotlivé nuance.

Jak již bylo popsáno v teoretické a metodologické části práce, významnější případové studie zpracovávané analýzou latentních tříd pracují často s dichotomickým typem otázek. Poznatek o hodnotě entropie, a především také věcná interpretace jednotlivých tříd na základě sady odhadnutých pravděpodobností daných odpovědí při přiřazení do určité latentní třídy mě však vede k důležitému argumentu pro preferenci vícepoložkových škál odpovědí.

3.5.3 Zhodnocení hypotéz výzkumu I.

V metodologické části práce byly definovány tři výzkumné hypotézy. Cílem této podkapitoly je stručné shrnutí závěrů pro hypotézy č. 1 a 2, které se pojí k této části výzkumu.

H1: Latentní struktura nalezená v rámci analýzy latentních tříd pro spojená data ze všech šesti vln dohromady se ve svých výsledcích výrazně neliší od latentní struktury každé vlny samostatně.

Ano, tato hypotéza může být na základě výše popsaných výsledků potvrzena. Latentní struktura obou analýz je velmi srovnatelná. Přesvědčivější, a především vhodnější formu výsledků pro data analyzovaná v rámci této případové studie však poskytuje analýza pro spojená data. Preferenci modelu pro spojená data potvrzují i hodnoty informačních kritérií, které jednoznačně upřednostňují model pro spojená data. Hodnoty AIC i BIC, jsou u tohoto modelu přibližně o 20 jednotek nižší, což může být považováno za významný rozdíl.

H2: Použití dichotomických proměnných oproti zachování čtyř-položkové škály souhlasu ubírá hodnotnou část popisu jednotlivých tříd, v důsledku tak mají modely odlišnou strukturu.

Ne, tato hypotéza nemůže být potvrzena. Analýza ukázala, že oba modely mají stejnou strukturu latentních tříd. Charakter tříd je stejný, interpretace u dichotomických proměnných však ztrácí na košatosti. Dichotomizace je výhodná při velkém počtu položek, a to jak z pohledu interpretace, tak i prezentace závěrů, a proto je v mnoha případech upřednostňovanou volbou. V kontextu této případové studie však považuji za vhodnější zachování čtyřpoložkové škály, která poskytuje více informací pro věcnou interpretaci latentních tříd.

Každá z tří předkládaných možností analýzy má klady a zápory. Na základě výše popsaných argumentů a cíle práce popsat jednotlivé typy diváků České televize se jako nejvhodnější varianta analýzy jeví analýza pro spojená data. Vzniklá segmentace diváků, která bude představena v následující kapitole, proto vychází z této analýzy.

3.6 Charakteristika nalezených segmentů

Analýza latentních tříd a její věcné zvážení ukázalo jako nejvhodnější řešení model se čtyřmi latentními třídami. Jinými slovy lze tedy zkoumaný vzorek diváků České televize rozdělit na základě jejich postoje k televizi do čtyř segmentů. Segmenty byly pojmenovány jako Příznivci, Podporovatelé, Skeptici a Nevyhranění. Tato kapitola si klade za cíl nalezené segmenty diváků České televize popsat.

3.6.1 1. segment – Příznivci

Prvním segmentem jsou Příznivci České televize, kteří tvoří 28 % zkoumané populace. Obecně se jedná o respondenty, kteří mají velmi pozitivní vztah k České televizi a k většině výroků se staví kladným souhlasem. Domnívají se, že Česká televize prezentuje objektivní, vyvážený a nezkreslený obraz světa kolem nás. Zaměříme-li se na důvěryhodnost České televize, diváci z tohoto segmentu jsou o ní přesvědčeni ze všech nejvíce, neboť s více jak sedmdesátiprocentní pravděpodobností odpovídají na výrok, že Česká televize je důvěryhodná, možností „rozhodně souhlasím“. Stejně tak si tito jedinci myslí, že Česká televize pomáhá rozvíjet všeobecný přehled u svých diváků, vzdělává a poskytuje rady z nejrůznějších oblastí života ve společnosti. Oblastí, kterou si lidé v tomto segmentu jsou oproti jiným sférám veřejnoprávnosti České televize nejméně jistí, je podpora kultury a národnostní a etnické tolerance. Přesto však i s výroky pojíci se k těmto tématům souhlasí.

3.6.2 2. segment – Podporovatelé

Druhý segment byl nazván jako Podporovatelé České televize a je se svým diváckým zastoupením 32 % nejpočetnějším segmentem. U všech výroků se diváci této skupiny kloní k pozitivní straně škály. Obecně je tento segment podobný čtvrtému

segmentu Nevyhraněných. U charakteristiky je významná především oblast kultury, národnostní a etnické tolerance a obecně vztahů mezi lidmi, neboť tyto výroky hodnotí diváci tohoto segmentu méně pozitivně než výroky ostatní. Nejméně přesvědčeni jsou příslušníci tohoto segmentu o nezávislosti ČT na politických a ekonomických vlivech. Pro celkovou názorovou charakteristiku segmentu je také důležité, že u většiny výroků má možnost „rozhodně nesouhlasím“ nulové zastoupení, což podporuje celkově pozitivní naladění těchto diváků vůči České televizi.

3.6.3 3. segment – Skeptici

V pořadí třetí segment tvoří Skeptici, kterých je ve zkoumané populaci 10 %. Diváci, kteří spadají do této skupiny, se ke všem výroky staví převážně negativně. Skepticky se stavějí především ke způsobu, jakým Česká televize prezentuje skutečnost a napomáhá divákům porozumět světu, ve kterém žijeme. Programová skladba České televize nenabízí dostatek pořadů, které by odpovídaly jejich vkusu a zájmům a nepovažují ji ani za inovativní. Zatímco diváci ze segmentu Příznivců České televizi důvěřují, jedinci z tohoto segmentu ji za příliš důvěryhodnou nepovažují. Diváci tohoto typu si myslí, že Česká televize je pod politickými a ekonomickými vlivy. Otázky, které se jako jedny z mála dostaly u těchto diváků na pozitivní stranu škály, se týkají spíše okrajových témat co do veřejnoprávní funkce České televize, jakými jsou například dostatečný prostor ve vysílání vyhrazený sportovním událostem či nabídka pořadů, které komerční televize divákům nenabízejí.

3.6.4 4. segment – Nevyhranění

Diváci tvořící čtvrtý segment jsou v rámci této případové studie označeni jako Nevyhranění a tvoří 30 % zkoumaných diváků. Odpovědi těchto diváků na zkoumané výroky oscilují mezi možnostmi spíše souhlasím“ a „spíše nesouhlasím“, proto je lze považovat za určitý názorový střed. U většiny výroků se však kloní spíše k pozitivní straně škály. Výjimkou je oblast výroků zaměřená na prezentaci skutečnosti, kde Nevyhranění tíhnou spíše ke kritice České televize. Programy, které Česká televize ve svém vysílání nabízí, příliš neodpovídají vkusu těchto diváků, kteří kritizují také fakt, že Česká televize v jejich očích není ekonomicky a politicky nezávislá.

Čtyři vzniklé segmenty byly na základě svého charakteru pojmenovány jako Příznivci, Podporovatelé, Skeptici a Nevyhranění. Analýza latentních tříd má obecně tendenci vytvářet krajní skupiny doplněné o skupiny střední, proto výsledná segmentace odpovídá očekávání. Pro tyto segmenty byla dále testována závislost na sociodemografických charakteristikách, které se ukázaly jako vlivný ukazatel pro interpretaci.

3.7 Vliv sociodemografických proměnných na segmenty

Pohled na jednotlivé segmenty z hlediska sociodemografických charakteristik může jistě přinést zajímavé poznatky. Statisticky se v tomto případě jedná o testování závislosti nominální závislé proměnné „příslušnost k latentní třídě“ na nominálních či ordinálních sociodemografických proměnných „pohlaví“, „věková kategorie“, „nejvyšší ukončené vzdělání“, „velikost místa bydliště“ a „četnost sledování ČT“. Z důvodu charakteru proměnných, u kterých nemá smysl počítat průměr, byly pro zjišťování závislosti zvoleny kontingenční tabulky, tedy dvourozměrné analýzy vztahu mezi dvěma proměnnými.

3.7.1 Pohlaví

První základní sociodemografickou charakteristikou, u které byl zkoumán její vliv na příslušnost k latentní třídě, je pohlaví. Následující kontingenční tabulka (tabulka č. 17) zobrazuje rozložení hodnot proměnných. U kontingenčních tabulek nedává smysl pracovat s absolutními četnostmi, proto jsou v této i v dalších kontingenčních tabulkách zobrazena sloupcová procenta.

Tabulka č. 17 – Kontingenční tabulka

Příslušnost k latentní třídě	Kontingenční tabulka – Pohlaví	
	Muž	Žena
Skeptici	11 %	10 %
Nevyhranění	30 %	27 %
Podporovatelé	32 %	32 %
Příznivci	27 %	31 %
Celkem	100 % 2976	100 % 3153

Z tabulky můžeme vyčíst, že například u čtvrtého segmentu Příznivců lehce převažují ženy nad muži. Aby bylo ověřeno, zda je tato závislost statisticky významná, či zda se jedná pouze o náhodu či výběrovou chybu, byl proveden chí-kvadrát test. Hodnota testového kritéria Pearsonova chí-kvadrát testu je v tomto případě 15,522 a její hladina významnosti je 0,001. Nulová hypotéza o nezávislosti vztahu mezi pohlavím a příslušností k latentní třídě byla proto zamítnuta a vliv pohlaví na charakteristiku jednotlivých segmentů může být považován za statisticky opodstatněný. Dále byl spočítán kontingenční koeficient Cramerovo V, který vyjadřuje těsnost vzájemné souvislosti dvou proměnných. Hodnota koeficientu je v tomto případě 0,050, což značí velmi slabou závislost.

Aby však výsledky mohly být dále specifikovány, byla pro každou hodnotu vypočítána adjustovaná rezidua, která zobrazuje následující tabulka (tabulka č. 18).

Tabulka č. 18 – Adjustovaná rezidua

Příslušnost k latentní třídě	Adjustovaná rezidua – Pohlaví	
	Muž	Žena
Skeptici	1,8	-1,8
Nevyhranění	2,6	-2,6
Podporovatelé	-0,5	0,5
Příznivci	-3,3	3,3

Adjustované reziduum vychází z rozdílu mezi napozorovanou a očekávanou četností, tedy tzv. *deltou*, kterou standardizuje podělením její směrodatnou odchylkou. Adjustovaná rezidua proto mohou být testována z hlediska jejich statistické významnosti. Pokud je hodnota vyšší v absolutní hodnotě než 2,00, můžeme si být s 95 % pravděpodobností jistí, že daný rozdíl mezi napozorovanou a očekávanou četností je statisticky významný. [Mareš, Rabušic, Soukup 2015: 257]

Podíváme-li se na vztah mezi pohlavím a příslušností k latentní třídy, vidíme, že do skupiny Nevyhraněných spadají častěji muži než ženy, a naopak v segmentu Příznivců je se dají očekávat spíše ženy než muži. U skupin Skeptiků a Podporovatelů nehraje pohlaví respondenta významnější roli.

3.7.2 Věk

V pořadí druhou zkoumanou sociodemografickou charakteristikou je věk. Respondenti byli rozděleni do tří věkových skupin a následující tabulka (tabulka č. 19)

zobrazuje zastoupení respondentů pro jednotlivé kombinace věku a příslušného segmentu. Vidíme, že výrazný rozdíl v zastoupení je především u druhého segmentu Nevyhraněných, kde převažují mladí lidé nad staršími 55 let.

Tabulka č. 19 – Kontingenční tabulka

Příslušnost k latentní třídě	Kontingenční tabulka – Věk		
	18-35	36-54	55+
Skeptici	12 %	10 %	10 %
Nevyhranění	33 %	29 %	25 %
Podporovatelé	30 %	33 %	32 %
Příznivci	25 %	29 %	33 %
Celkem	100 %	100 %	100 %
	1858	2081	2192

Hodnota testového kritéria Pearsonova chí-kvadrát testu pro věkovou proměnou je 55,539 a její pravděpodobnost chyby prvního druhu je menší než 0,0005. Opět je tedy možné zamítnout nulovou hypotézu o nezávislosti vztahu mezi věkem a příslušností k latentní třídě a lze mluvit o vlivu této proměnné na zařazení respondenta do jednoho ze čtyř segmentů. Kontingenční koeficient Cramerovo V dosahuje hodnoty 0,067, což značí, že tato závislost je relativně malá, ale jistý vliv přece jen má.

Adjustovaná rezidua dále ukazují, že mladí lidé budou oproti lidem nad 55 let výrazně častěji ztotožňovat své názory se skupinou Nevyhraněných. Opačné zastoupení charakterizuje skupinu Příznivců, pro které je typický spíše starší věk. Z tabulky níže (tabulka č. 20) je také vidět, že ve skupině Skeptiků jsou více zastoupeni mladí lidé pod 35 let.

Tabulka č. 20 – Adjustovaná rezidua

Příslušnost k latentní třídě	Adjustovaná rezidua – Věk		
	18-35	36-54	55+
Skeptici	2,6	-1,3	-1,2
Nevyhranění	5,0	-0,2	-4,6
Podporovatelé	-1,8	1,6	0,2
Příznivci	-4,9	-0,5	5,2

3.7.3 Vzdělání

Třetí sociodemografickou proměnnou, u které byl zkoumán její vliv na příslušnost k latentní třídě je nejvyšší ukončené vzdělání. Již samotná kontingenční tabulka (tabulka

č. 21) nenaznačuje žádné významnější rozdíly mezi vzdělanostními skupinami, jak lze vidět níže.

Tabulka č. 21 – Kontingenční tabulka

Příslušnost k latentní třídě	Kontingenční tabulka – Vzdělání			
	základní	střední bez maturity	střední s maturitou	vysokoškolské
Skeptici	12 %	11 %	9 %	11 %
Nevyhranění	30 %	28 %	29 %	28 %
Podporovatelé	31 %	32 %	33 %	33 %
Příznivci	27 %	29 %	30 %	28 %
Celkem	100 % 1025	100 % 2187	100 % 2104	100 % 815

Hodnota testového kritéria Pearsonova chí-kvadrát testu pro vzdělání je 11,897 a hladina významnosti 0,219 překročila obecně přijímanou hodnotu signifikance 0,05. V případě vzdělání tedy nelze zamítnout nulovou hypotézu o nezávislosti vztahu mezi vzděláním a příslušností k latentní třídě. Hodnota kontingenčního koeficientu Cramerova V je v tomto případě 0,025, což značí velmi slabý vztah proměnné vzdělání a příslušnosti k latentní třídě. Nejvyšší ukončené vzdělání respondenta nemá výraznější vliv na jeho postoj k České televizi, a s tím spojené zařazení do jednoho ze čtyř segmentů diváků.

3.7.4 Velikost místa bydliště

Velikost místa bydliště se na rozdíl od vzdělání projevila jako proměnná, která příslušnost respondenta k latentní třídě ovlivňuje, neboť v následující kontingenční tabulce (tabulka č.22) můžeme pozorovat rozdíly mezi zastoupením obyvatel z různě velkých obcí a měst v jednotlivých latentních třídách.

Tabulka č. 22 – Kontingenční tabulka

Příslušnost k latentní třídě	Kontingenční tabulka – Velikost místa bydliště				
	do 999 obyvatel	1 000 - 4 999 obyvatel	5 000 - 19 999 obyvatel	20 000 - 99 999 obyvatel	100 000 a více obyvatel
Skeptici	8 %	9 %	8 %	11 %	14 %
Nevyhranění	28 %	28 %	30 %	27 %	31 %
Podporovatelé	33 %	36 %	30 %	29 %	32 %
Příznivci	31 %	26 %	32 %	33 %	23 %
Celkem	100 % 1029	100 % 1271	100 % 1116	100 % 1326	100 % 1388

Hodnota testového kritéria Pearsonova chí-kvadrát testu pro věkovou proměnou je 82,195 a její pravděpodobnost chyby prvního druhu je 0,000. V tomto případě je tedy možné zamítnout nulovou hypotézu o nezávislosti vztahu mezi věkem a příslušností k latentní třídě a lze mluvit o vlivu této proměnné na zařazení respondenta do jednoho ze čtyř segmentů. Hodnota kontingenčního koeficientu Cramerova V je 0,067, což poukazuje na malý vliv.

Řídíme-li se kritériem, že hodnota adjustovaného rezidua vyšší v absolutní hodnotě než 2,00 značí statistický významný rozdíl mezi skupinami, vyčteme z následující tabulky, že pro skupinu Skeptiků jsou charakterističtí lidé z velkých měst nad 100 000 obyvatel. Segment Podporovatelů má naopak velké zastoupení diváků z menších obcí velikosti 1 000 až 4 999 obyvatel. Pro příznivce je poté typické středně velké město od velikosti 5 000 do 99 999 obyvatel. Konkrétní hodnoty reziduí zobrazuje i s vyznačením významných rozdílů následující tabulka (tabulka č. 23):

Tabulka č. 23 – Adjustovaná rezidua

Příslušnost k latentní třídě	Adjustovaná rezidua – Velikost místa bydliště				
	do 999 obyvatel	1 000 - 4 999 obyvatel	5 000 - 19 999 obyvatel	20 000 - 99 999 obyvatel	100 000 a více obyvatel
Skeptici	-2,8	-1,3	-3,0	1,4	5,2
Nevyhranění	-0,5	-0,5	0,6	-1,9	2,3
Podporovatelé	0,8	3,5	-1,4	-2,4	-0,5
Příznivci	2	-2	3	4	-5

3.7.5 Četnost sledování ČT

Z dalších proměnných, které by mohly ovlivňovat příslušnost respondenta k latentní třídě přinesla poměrně zajímavé výsledky otázka zjišťující, jak často respondent sleduje Českou televizi. Kontingenční tabulka (tabulka č. 24) v tomto případě zobrazuje velké procentuální rozdíly v zastoupení jednotlivých kategorií.

Tabulka č. 24 – Kontingenční tabulka

Příslušnost k latentní třídě	Kontingenční tabulka – Jak často sledujete ČT?				
	každý den	3-6x v týdnu	1-2x v týdnu	méně často	nikdy
Skeptici	5 %	8 %	11 %	24 %	48 %
Nevyhranění	23 %	28 %	34 %	40 %	35 %
Podporovatelé	34 %	36 %	32 %	24 %	7 %
Příznivci	39 %	29 %	24 %	12 %	11 %
Celkem	100 % 2239	100 % 1673	100 % 1372	100 % 679	100 % 164

Hodnota testového kritéria Pearsonova chí-kvadrát testu je v tomto případě 735,248 a její hladina významnosti je 0,000. Nulová hypotéza o nezávislosti vztahu mezi pohlavím a příslušností k latentní třídy byla proto jednoznačně zamítnuta a můžeme říci, že čestnost sledování České televize významně ovlivňuje příslušnost respondenta k latentní třídě. Kontingenční koeficient Cramerovo V dosáhl hodnoty 0,20 a je tedy v porovnání s ostatními sociodemografickými charakteristikami proměnnou nejvíce ovlivňující příslušnost k latentní třídě.

Z tabulky adjustovaných reziduí (tabulka č. 25) vidíme, že rozdíly mezi jednotlivými skupinami jsou zde nejvýznamnější ze všech testovaných proměnných. Obecný vztah je takový, že Skeptici Českou televizi nesledují, zatímco Příznivci ano. Tato charakteristika bude více rozebrána u popisu jednotlivých segmentů.

Tabulka č. 25 – Adjustovaná rezidua

Příslušnost k latentní třídě	Adjustovaná rezidua – Jak často sledujete ČT?				
	každý den	3-6x v týdnu	1-2x v týdnu	méně často	nikdy
Skeptici	-11	-4,2	1,5	12,5	16
Nevyhranění	-7,5	-1,4	4,4	6,9	1,7
Podporovatelé	2,2	4	-0,5	-4,6	-7
Příznivci	12,5	0,1	-4,9	-10,5	-5,1

3.7.6 Zhodnocení hypotéz výzkumu II.

Pro tuto část práce zaměřenou na vliv sociodemografických charakteristik na příslušnost k latentní třídě byla definována následující hypotéza:

H3: Jednotlivé segmenty lze odlišit na základě sociodemografických charakteristik jejich členů.

Tato hypotéza může být potvrzena pouze částečně – některé sociodemografické charakteristiky mají vliv, ten však není nikterak významný. Nejvýznamnější charakteristikou je četnost sledování České televize, vzdělání naopak nemá žádný vliv na příslušnost k latentní třídě.

3.7.7 Shrnutí charakteristik segmentů

Jak již bylo řečeno dříve, individuální rozdíly pozorovatelné ve struktuře odpovědí respondentů jsou podle myšlenky analýzy latentních tříd způsobeny příslušností k odlišným latentním třídám. [Geiser 2013: 233] Fakt, že každá třída má své specifické charakteristiky, je v rámci této analýzy zcela klíčový, a proto jsou hlavními posuzovanými kritérii počet, velikost a charakter latentních tříd.

Ve zkoumané populaci diváků České televize byly identifikovány čtyři segmenty diváků. Z obecného pohledu se jedná o dvě krajní skupiny, tedy segment diváků, kteří k České televizi mají kladný vztah, a naopak těch, jejichž vztah je negativní. Druhé dva segmenty se vyznačují středovým a středově pozitivním postojem. Většinový charakter vztahu všech diváků zahrnutých do výzkumu k České televizi je pozitivní. Toto tvrzení vychází z úvahy, že sečteme-li procentuální zastoupení Příznivců a Podporovatelů, jejichž vztah k České televizi se nachází na pozitivní straně škály, dostaneme 60 % všech respondentů. Oproti tomu Skeptici, kteří se jako jediní u většiny výroku kloní k negativní straně škály, představují pouhých 10 % respondentů. Nevyhranění poté jako středová skupina zastupují neutrální postoj k České televizi a tvoří zbylých 30 %.

Podíváme-li se na jednotlivé segmenty ze sociodemografického pohledu, v případě Příznivců České televize se jedná o mladé lidi ve věku 18–35 let, kteří žijí ve středně velkých městech. Většinou se jedná o ženy. Co se sledování České televize týče, jsou tito diváci nejvěrnějšími diváky. Téměř 40 % z nich se na Českou televizi dívá denně. Představují tak přesný opak Skeptiků, kteří Českou televizi typicky nesledují téměř vůbec.

Jedinou sociodemografickou charakteristikou, která vystihuje diváky segmentu Podporovatelů, je velikost místa bydliště. Podporovatelé pocházejí typicky z menší obce, konkrétně velikosti 1 000 – 4 999 obyvatel. Podíváme-li se na četnost sledování České televize těmito diváky, zjistíme, že ji sledují několikrát týdně. I z pohledu sociodemografie se tedy Podporovatelé ukazují jako více méně středová skupina bez tendence k extrémům.

Co se Skeptiků týče, jsou diváky z tohoto segmentu nejčastěji mladí lidé z velkých měst nad 100 000 obyvatel. Četnost sledování České televize ukazuje, že Skeptici, kteří se staví k České televizi převážně negativně, jsou těmi, kteří ji sledují ze všech skupin nejméně.

Mezi Nevyhraněné se na základě sociodemografických charakteristik řadí především mladí muži bydlící ve větších městech. Typický divák ze segmentu Nevyhraněných se na Českou televizi dívá spíše méně často, určitě ne každý den.

Následující tabulka (tabulka č. 26) shrnuje charakter jednotlivých segmentů:

Tabulka č. 26 – Shrnutí charakteristik segmentů

Shrnutí charakteristik segmentů					
	Tři tematické kategorie otázek			Sociodemografické charakteristiky	Velikost
	Způsob prezentace skutečnosti	Porozumění světu	Kultura, tolerance a vztahy mezi lidmi		
Příznivci	ČT velmi důvěřují a věří, že poskytuje pravdivý, objektivní, nezkreslený obraz skutečnosti	ČT určitě vzdělává a rozvíjí své diváky.	Touto kategorií jsou si jistí nejméně, přesto souhlasí, že ČT posiluje kulturu a toleranci.	Mladé ženy ze středně velkých měst. Na ČT se dívají nejčastěji, často denně.	28 %
Podporovatelé	Obecně spíše souhlasí s prezentovaným obrazem skutečnosti.	Spíše souhlasí se všemi body této kategorie.	S výroky z této kategorie souhlasí méně než se zbývajícími dvěma kategoriemi.	Obyvatelé menších obcí (1 000 – 4 999 obyvatel). ČT sledují několikrát týdně.	32 %
Skeptici	ČT není důvěryhodná a neprezentuje skutečnost věrohodně. ČT je pod politickými a ekonomickými vlivy.	ČT nepomáhá svým divákům v porozumění světu a spíše nevzdělává.	ČT nepomáhá kultuře ani toleranci ve společnosti.	Mladí lidé z velkých měst nad 100 000 obyvatel. Nesledují ČT téměř vůbec.	10 %
Nevyhranění	Spíše kritičtí ke způsobu prezentace skutečnosti.	Středový postoj.	Středový postoj.	Mladí muži z větších měst.	30 %

4 Diskuze výsledků, limitů a návrhů pro další výzkum

Jsem si vědoma, že předkládaná diplomová práce je případovou studií užití analýzy latentních tříd se zcela konkrétními daty, která jistě mají svá specifika, a nekladu si proto za cíl zobecnit závěry o použitých metodách. Domnívám se však, že výsledky realizované studie mohou sloužit jako nástroj dalšího srovnávání či poskytnout inspiraci pro další podobně orientované výzkumy.

Zajímavý předmět možného pokračování výzkumu spatřuji v otázce odlišného zastoupení respondentů v jednotlivých latentních třídách při srovnávání modelů počítaných třemi zkoumanými metodami. Rozsah a cíle této diplomové práce již neumožnily teoretický ani praktický výzkum toho, proč se modely se stejným počtem, strukturou a charakterem latentních tříd liší procentuálním zastoupením respondentů v každé třídě.

Možné pokračování výzkumu vidím také v problematice práce s více položkovou škálou oproti užití dichotomických proměnných. V současnosti bývá analýza latentních tříd v případových studiích často užitá pro práci s dichotomickými daty, domnívám se však, že jsou tím závěry ochuzeny o podstatnou část informací.

Závěr

Předložená diplomová práce se zabývá analýzou latentních tříd a možnostmi jejího využití. V teoretické části práce je popsána pozice analýzy latentních tříd mezi ostatními segmentačními metodami, koncept latentních proměnných a práce s nimi či stručný přehled historického vývoje této metody. Navazující kapitoly byly poté koncipovány s ohledem na praktickou neexistenci česky psané literatury na toto téma. Značný prostor je proto věnován teoretickému rámci, do kterého je analýza latentních tříd zasazena, seznámení s ideou stojící v pozadí této statistické techniky a vymezení základních pojmů, se kterými analýza latentních tříd pracuje. V rámci rešerše dosavadní literatury byly poté představeny významné případové studie využívající analýzu latentních tříd a shrnuta jejich hlavní zjištění.

Metodologická část práce je zaměřena na případovou studii, která byla v rámci této diplomové práce realizována. Ve své obsahové rovině se studie zabývá obrazem České televize jako média veřejné služby. Vybrána byla baterie otázek zaměřená na to, jak dotazovaní respondenti vnímají Českou televizi, nakolik v jejich očích naplňuje svoji veřejnoprávní funkci či jakými atributy disponuje. Cílem studie je segmentace televizních diváků do několika skupin na základě jejich reflexe zkoumaných oblastí.

Z metodologického pohledu je tato případová studie sekundární analýzou dat. Za pomoci těchto dat byly představeny tři možnosti uchopení analýzy latentních tříd v datech z šesti samostatně sbíraných vln – analýza latentních tříd pro spojená data, simultánní analýza latentních tříd a analýza latentních tříd pro dichotomické proměnné. Z předkládaného textu by mělo být patrné, že se jedná o kombinaci dvou různých

konceptů. V první řadě je srovnáván přístup postavený na spojených datech oproti simultánní analýze každé vlny zvlášť. Ve druhém případě jde o posouzení vhodnosti zachování vícepoložkové škály souhlasu u analyzovaných otázek, či jejich nahrazení dichotomickými proměnnými.

V analytické části práce byly vytvořeny tři modely latentních tříd pro klasifikaci diváků České televize, každý za využití jednoho ze zkoumaných typů analýzy latentních tříd. Výzkum skrze komparaci těchto metod ukázal, že nejvhodnější technikou je analýza latentních tříd pro spojená data. Hlavním argumentem pro tuto volbu jsou možnosti interpretace výsledků, na jejichž základě je postavena výsledná typologie. Na rozdíl od simultánní analýzy jsou v analýze pro spojená data věcné závěry o nalezených latentních třídách průkaznější, přehlednější a zřejmější, neboť je generována pouze jedna sada odhadnutých pravděpodobností daných odpovědí při přiřazení do určité latentní třídy. Simultánní analýza generuje sad šest, tedy jednu sadu pro každou vlnu, a jejich spojení do jedné ucelené klasifikace probíhající pouze v rovině vlastní interpretace může být velmi ovlivněno úsudkem výzkumníka. Zdůraznění konkrétních jevů, a naopak potlačení jiných tak může hrát významnější roli, než je tomu u analýzy pro spojená data. Preferenci analýzy pro spojená data podpořily i hodnoty informačních kritérií, které byly u této analýzy významně nižší než u simultánní analýzy.

Ze srovnání užití čtyřpoložkové škály stupně souhlasu či z ní přepočítaných dichotomických proměnných je v této práci argumentováno ve prospěch zachování škály v její delší podobě. Hlavním důvodem je ztráta velké části věcných informací u dichotomických proměnných. Interpretaci odhadnutých pravděpodobností daných odpovědí při přiřazení do určité latentní třídy činí dichotomické proměnné výrazně snazší, což může být v mnoha situacích výhodné, ale v případové studii realizované se záměrem o důkladné pochopení jednotlivých segmentů zkoumané populace a rozdílů mezi nimi smazává tento typ analýzy mnoho podstatných detailů.

Všechny tři provedené varianty analýzy jsou plně funkční a srovnatelné. Nalezeny byly charakterově stejné třídy, avšak liší se v poměrném zastoupení jedinců v jednotlivých třídách. Z tohoto důvodu je třeba mít na zřeteli klady i zápory každé z předkládaných analýz a nepodceňovat výběr vhodné metody pro daný výzkum.

Z věcného hlediska byly v rámci této diplomové práce nalezeny čtyři segmenty diváků na základě jejich vnímání České televize jako média veřejné služby. Tyto segmenty byly označeny jako Příznivci, Podporovatelé, Skeptici a Nevyhranění. Příznivci tvoří 28 % zkoumané populace a obecně se jedná o respondenty, kteří mají

velmi pozitivní vztah k České televizi a k většině výroků se staví kladným souhlasem. Podporovatelé jsou se svým podílem 32 % respondentů nejpočetnějším segmentem. Jedná se o diváky, kteří jsou vůči České televizi také naladěni pozitivně. Méně přesvědčeni jsou však o podpoře oblasti kultury a národnostní a etnické tolerance ze strany České televize, stejně jako o její ekonomické a politické nezávislosti. Třetí segment byl nazván Skeptici – těch je ve zkoumané populaci 10 %. Diváci, kteří spadají do této skupiny, se ke všem výrokům staví převážně negativně. Programová skladba České televize nenabízí dostatek pořadů, které by odpovídaly jejich vkusu a zájmům, proto Českou televizi nesledují téměř vůbec. Kritizují její nestrannost a způsob, jakým divákům předkládá skutečnost. Posledním segmentem jsou Nevyhranění tvořící 30 % dotazované populace, kteří představují jakýsi názorový střed. Výjimkou je oblast výroků zaměřená na prezentaci skutečnosti, kde Nevyhranění tíhnou spíše ke kritice České televize.

Analýzu latentních tříd osobně považuji za velmi užitečný nástroj segmentace a doufám, že předkládaná diplomová práce by mohla inspirovat další studie a k jejímu užítí.

Summary

The presented thesis deals with latent class analysis and potential of its application. The theoretical part of the paper describes the position of latent class analysis among other segmentation methods, the concept of latent variables and its applications and a brief historical overview of this method's evolvement. The following chapters were structured with regard to the fact that there are almost no publications on the topic written in the Czech language. Considerable focus is therefore given to the theoretical framework to which the latent class analysis is set into together with introduction of the idea standing behind this statistical technique and definition of basic terms which the latent class analysis uses. Subsequently, influential case studies using the latent class analysis are presented and their main outcomes are summarized.

Methodological part of the paper is focused on the case study realized within the thesis. The study's subject focuses on the image of the Czech Television as a public service media. The chosen set of questions concentrate on how the respondents perceive the Czech Television, how they value it's performance as a public service media and

what are the attributes that it disposes with. The aim of the study is to make a segmentation of the television audience into several groups according to their reflection towards the examined areas.

From methodological point of view, this case study represents a secondary data analysis. With data from six different research waves provided, three ways of conducting a latent class analysis are introduced - the latent class analysis for joint data, simultaneous latent class analysis and latent class analysis for dichotomous variables. It should be apparent from the provided text, that combination of two concepts is being used. Firstly, the approach based on joint data is compared with the one using simultaneous analysis of each research wave separately. Secondly, the suitability of preserving multi-item approval scale to the analysed questions is evaluated in comparison to their replacement with dichotomous variables.

Three models of latent class analyses for classification of the Czech Television audience were created in the analytical part of the paper; one for each type of latent class analysis concerned. The research comparing these methods has shown that the most suitable technique is the latent class analysis for joint data. The main argument supporting this conclusion is the possibility to interpret results which serve as basis for the final typology. In contrast to the simultaneous analysis, this approach provides more conclusive, transparent and apparent findings because only one set of latent class probabilities is generated. The simultaneous analysis generates six of these sets (one for each wave) and their merging into one complete classification is done only by interpretation. Therefore, the result may be significantly influenced by researcher's own judgement. Evaluation of importance of different phenomena can therefore play more significant role here than in case of the joint data analysis.

When choosing between usage of four-item approval scale and dichotomous variables recalculated out of it, this paper argues in favour of preserving the scale in its longer form. The main reason for this is that a large number of factual information is lost in case of dichotomous variables. The latter approach makes the interpretation of latent class probability much easier; this can be useful in some cases, but in case of a study made with the intention to obtain deep understanding of individual segments of examined population, many important details are effaced in this type of analysis.

From the factual point of view, four segments of the Czech Television viewers were identified based on their perception of the Czech Television as a public service media. These segments were labelled as Supporters, Moderate Supporters, Sceptics and Ambivalent Viewers. Supporters make 28 % of the respondents and they generally represent people with very positive approach towards the Czech Television and they agree with majority of the statements. With 32% share the Moderate Supporters represent the biggest segment. This category involves respondents that are also positive towards the Czech Television but are not completely convinced that it sufficiently supports culture and ethnic and multinational tolerance. They also have concerns about its economic and political independence. The third segment called Sceptics counts for 10% of the respondents. They have negative view on majority of the statements. The program schedule of the Czech Television does not offer enough titles that they would be interested in and they criticize it's impartiality and the way in which the reality is presented; so they hardly ever watch it. The last segment of Ambivalent Viewers consists of 30% share of the respondents. They stand approximately in the middle of the opinion scale concerning most of the topics with the exception of questions concerning presentation of reality where these respondents lean more towards criticism.

In my opinion, the latent class analysis is a very useful tool of segmentation and I hope that the presented thesis will inspire more researches towards its use.

Literatura

BAILEY, K.D. *Cluster Analysis*. Wiley: Sociological Methodology, Vol. 6 (1975), pp. 59-128. 1975

BAILEY, K.D. *Typologies and taxonomies: an introduction to classification techniques*. Thousand Oaks, Calif.: Sage Publications. Sage university papers series, no. 07-102. ISBN 0-8039-5259-7. 1994

BHATNAGAR, A. GHOSE, S. *A latent class segmentation analysis of e-shoppers*. Journal of Business Research [online]. 2004, **57**(7), 758-767 [cit. 2016-11-25]. DOI: 10.1016/S0148-2963(02)00357-0. ISSN 01482963. Dostupné z: <http://linkinghub.elsevier.com/retrieve/pii/S0148296302003570>

BRESLAU, N. REBOUSSIN, B. A., ANTHONY, J. C. STORR, C. L. *The Structure of Posttraumatic Stress Disorder*. Archives of General Psychiatry [online]. 2005, 62(12), 1343- [cit. 2016-11-19]. DOI: 10.1001/archpsyc.62.12.1343. ISSN 0003990x. Dostupné z: <http://archpsyc.jamanetwork.com/article.aspx?doi=10.1001/archpsyc.62.12.1343>

BURNHAM, K. P.; ANDERSON, D. R. *Multimodel Inference: Understanding AIC and BIC in Model Selection*. Sociological Methods & Research. Vol. 33: 261–304. 2004 in SOUKUP, P. *Nesprávná užívání statistické významnosti a jejich možná řešení*. Data a výzkum - SDA Info 2010, Vol. 4, No. 2: 77-104. (c) Sociologický ústav AV ČR, v.v.i., Praha. 2010

COHEN, J. *The earth is round ($p < .05$)*. American Psychologist. Vol. 49: 997–1003. 1994 in SOUKUP, P. *Nesprávná užívání statistické významnosti a jejich možná řešení*. Data a výzkum - SDA Info 2010, Vol. 4, No. 2: 77-104. (c) Sociologický ústav AV ČR, v.v.i., Praha. 2010

COOPER, J.C.B. *Factor Analysis: An Overview*. Taylor & Francis: The American Statistician, Vol. 37, No. 2 (1983), pp. 141-147. 1983

FINCH, W.H. BRONK, K.C. *Conducting Confirmatory Latent Class Analysis Using Mplus*, Structural Equation Modeling: A Multidisciplinary Journal, 18:1, 132-151. 2011

GEISER, Ch. *Data analysis with Mplus*. New York: The Guilford Press, 2013. ISBN 9781462507825.

GOODMAN, L.A. *Exploratory Latent Structure Analysis Using Both Identifiable and Unidentifiable Models*, Biometrika, Vol. 61, No. 2., pp. 215-231. 1974

GREEN, B.F. *Latent Structure Analysis and its Relation to Factor Analysis*, Journal of the American Statistical Association, Vol. 47, No. 257. pp. 71-76. 1952

GREEN, P.E. CARMONE, F.J. WACHSPRESS D.P. *Consumer Segmentation via Latent Class Analysis*. Oxford University Press: Journal of Consumer Research , Vol. 3, No. 3 , pp. 170-174. 1976

GROVER, J., VRIENS, M. *The Handbook of Marketing Research: Uses, Misuses, and Future Advances*. Thousand Oaks: Sage Publications, ISBN 141290997X. 2006

HAGENAARS, J. A., MCCUTCHEON A. L. *Applied latent class analysis*. New York: Cambridge University Press, 2002. ISBN 0521594510.

HAGENAARS, J.A, HALMAN, L.C. *Searching for Ideal Types: The Potentialities of Latent Class Analysis*. Oxford University Press: *European Sociological Review*, Vol. 5, No. 1 (May, 1989), pp. 81-96. 1989

HENDL, J. *Přehled statistických metod: analýza a metaanalýza dat*. Páté, rozšířené vydání. Praha: Portál. ISBN 9788026209812. 2015

HENRY, N.W. *Latent Structure Analysis at Fifty*. Virginia Commonwealth University, Richmond, VA 23284-2014. pp. 587-592. 1999

HOIJTIK, H *Confirmatory Latent Class Analysis: Model Selection Using Bayes Factors and (Pseudo) Likelihood Ratio Statistics*. *Multivariate Behavioral Research*, 36, no. 4, pp. 563-588. 2001

JEŘÁBEK, H. SOUKUP, P. (eds.). *Advanced Lazarsfeldian methodology*. Prague: Karolinum, ISBN 9788024615219. 2008

JEŘÁBEK, H. *Paul Lazarsfeld's research methodology: biography, methods, famous projects*. Prague: Karolinum, ISBN 8024610981. 2006

KANKARAŠ, M., VERMUNT J. K. *Simultaneous Latent-Class Analysis Across Groups*. *Encyclopedia of Quality of Life and Well-Being Research* [online]. Dordrecht: Springer Netherlands, 2014, 5969 [cit. 2016-12-30]. DOI: 10.1007/978-94-007-0753-5_2711. ISBN 978-94-007-0752-8. Dostupné z: http://link.springer.com/10.1007/978-94-007-0753-5_2711

KLECKA, W R. *Discriminant Analysis*, Beverly Hills: Sage Publications. 1980

LAZARSELD, P.F. *A Conceptual Introduction to Latent Structure Analysis*. 1969 In: LAZARSELD, P.F. *Mathematical Thinking in the Social Sciences*. New York, Russel and Russel (reprint, 1.st ed. 1954).

LAZARSELD, P.F., HENRY, N.W. *Latent Structure Analysis*, Boston, Houghton Mifflin Company. 1968.

LITTLE, T.D. (ed.). *The Oxford handbook of quantitative methods*. New York: Oxford University Press. Oxford library of psychology. ISBN 9780195342802. 2013

MAREŠ, P., RABUŠIC, L. a SOUKUP, P. *Analýza sociálněvědních dat (nejen) v SPSS*. Brno: Masarykova univerzita, 2015. ISBN 9788021063624.

MCCUTCHEON, A. L. *Basic concepts and procedures in single-and multiple-group latent class analysis*. In J. A. Hagenaars & A. L. McCutcheon (Eds.), *Applied latent class analysis* (pp. 57–88). Cambridge, UK: Cambridge University Press. 2002

McCUTCHEON, A.L. *Latent Class Analysis*. Newbury Park: Sage Publications. ISBN 0803927525. 1987

MUTHÉN, L, K, MUTHÉN, B,O *MPlus User Guide* [online]. Version 6, 2010. [cit. 2016-10-06]. Dostupné z: <https://www.statmodel.com/download/usersguide/Chapter12.pdf>

ROSENBERG, M. *The Logic of Survey Analysis*. New York: Basic Books. 1968 In: McCUTCHEON, A.L. *Latent Class Analysis*. Newbury Park: Sage Publications. ISBN 0803927525. 1987

SAVAGE, M., DEVINE F., CUNNINGHAM N., et al. *A New Model of Social Class? Findings from the BBC's Great British Class Survey Experiment*. *Sociology* [online]. 2013, 47(2), 219-250 [cit. 2016-11-18]. DOI: 10.1177/0038038513481128. ISSN 00380385. Dostupné z: <http://soc.sagepub.com/cgi/doi/10.1177/0038038513481128>

SIMAR [online]. [cit. 2016-11-17]. Dostupné z: <http://simar.cz/o-nas/eticke-zasady-oboru.html>

SOUKUP, P. *Nesprávná užívání statistické významnosti a jejich možná řešení*. *Data a výzkum - SDA Info* 2010, Vol. 4, No. 2: 77-104. (c) Sociologický ústav AV ČR, v.v.i., Praha. 2010

SOUKUP,P, *Advanced Data Analysis in MPlus* (přednáška) Praha: Univerzita Karlova v Praze, Fakulta sociálních věd. LS 2015/2016

WEDEL, M., WAGNER, K. A. *Market Segmentation: Conceptual a Methodological Foundations*. 2nd ed. Boston: Kluwer Academic Publishers. ISBN 0-7923-8635-3. 2000

Přílohy

Z grafických důvodů jsou přílohy umístěny na následujících stránkách.