

**UNIVERZITA KARLOVA**

**FAKULTA SOCIÁLNÍCH VĚD**

Institut sociologických studií

**Bc. Lucie Horáková**

**Analýza mediace v sociologii**

*Diplomová práce*

Praha 2017

Autor práce: **Lucie Horáková**

Vedoucí práce: **prof. RNDr. Jan Hendl, CSc.**

Rok obhajoby: 2017

## **Bibliografický záznam**

HORÁKOVÁ, Lucie. *Analýza mediace v sociologii*. Praha, 2017. 75 s. Diplomová práce (Mgr.) Univerzita Karlova, Fakulta sociálních věd, Institut sociologických studií. Katedra sociologie. Vedoucí diplomové práce prof. RNDr. Jan Hendl, CSc.

## **Abstrakt**

Diplomová práce „Analýza mediace v sociologii“ se zabývá analýzou mediace a možnostmi její aplikace v oblasti sociologie v závislosti na typu závislé proměnné, která do analýzy vstupuje. První možností je případ, kdy závislá proměnná je spojitá – v tomto případě je použit program SPSS a jeho doplněk PROCESS, který slouží přímo k analýze mediace. Druhou možností je případ, kdy závislá proměnná vstupující do analýzy je binární – doplněk PROCESS tuto variantu neumožňuje, tudíž je analýza provedena taktéž v programu SPSS sadou lineárních a logistických regresí dle metody Baron & Kenny. V práci jsou využity dvě případové studie z oblasti sociologie, jednak GSS (General Social Survey) a ISSP (International Social Survey Programme) a pomocí sekundární analýzy těchto dat jsou zkoumány důsledky přechodu ze spojitě závislé proměnné na binární.

## **Abstract**

Diploma thesis „Mediation Analysis in Sociology“ deals with mediation analysis and possibilities of its application in sociology, depending on the type of the dependent variable that enters the analysis. In the first case the dependent variable is continuous – in this case the SPSS software and its PROCESS add-on are used to directly analyse the mediation. In the second case the dependent variable that enters the analysis is binary – the PROCESS add-on doesn't allow this option; therefore, the analysis is performed in SPSS software by the set of linear and logistic regressions according to the Baron & Kenny method. Two case studies from the field of sociology, GSS (General Social

Survey) and ISSP (International Social Survey Programme), are used in the thesis and the consequences of the transition from continuous dependent variable to binary are examined using the secondary analysis of these data.

## **Klíčová slova**

Analýza mediace, mediátor, lineární regrese, logistická regrese, SPSS, PROCESS

## **Keywords**

Mediation analysis, mediator, linear regression, logistic regression, SPSS, PROCESS

**Rozsah práce:** 111 515 znaků

## **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.

## **Poděkování**

Na tomto místě bych ráda poděkovala vedoucímu práce prof. RNDr. Janu Hendlovi, CSc. za trpělivost, cenné rady, připomínky a příjemně strávený čas při konzultacích.

Dále bych ráda poděkovala svému partnerovi za podporu a za to, že nikdy neztrácel víru v mé schopnosti i v těžkých chvílích.

# Institut sociologických studií

## Teze diplomové práce

### **Projekt diplomové práce**

**Předpokládaný název:** Analýza mediace ve statistice

**Předpokládaný název v angličtině:** Mediation analysis in statistics

**Autor:** Lucie Horáková

**Konzultant:** prof. RNDr. Jan Hendl CSc.

### **Námět práce zahrnující formulaci a vstupní diskusi poznávaného problému**

Mediace ve statistice je procesem, který nám umožňuje zkoumat kauzální vztah mezi dvěma proměnnými. Jedná se o zapojení třetí proměnné do vztahu mezi dvěma, díky čemuž jsme schopni tento vztah přesněji modelovat. Pokud zkoumáme vztah mezi dvěma proměnnými, zpravidla jednu označujeme jako příčinu (tedy nezávislou či vysvětlující proměnnou) a druhou jako následek (tedy vysvětlovanou či závislou proměnnou). Pomocí mediace nicméně nezjišťujeme pouhý prostý vztah mezi těmito dvěma, ale o hlubší analýzu procesů, které se v tomto vztahu odehrávají, a co může ovlivnit jeho směr či sílu. Mediace je klíčovým konceptem pro spoustu teoretických i aplikovaných vědních oborů, jako je například psychologie (sociální, klinická, vývojová i kognitivní), či sociální a biomedicínské vědy. V praxi se používá zejména pro analýzy různých intervenčních, léčebných či jiných programů a měření jejich efektů.

### **Cíl práce**

Zahraniční literatura se tomuto tématu věnuje spíše omezeně, českých publikací na toto téma je spíše poskrovnu. Rozhodla jsem se tedy pro zkoumání tohoto tématu, abych přispěla k jeho rozšíření v tuzemské literatuře, a více přiblížila statistické a metodologické aspekty analýzy mediace i českým čtenářům. Cílem práce je také představit konkrétní aplikace právě ve společenskovědním výzkumu.

### **Předpokládané metody zpracování a předběžná struktura práce**

První částí práce bude představení teoretického rámce tématu, jeho výhod a kritiky, kde budu zcela primárně vycházet ze zahraničních pramenů. Díky omezenému rozsahu odborné literatury na toto téma a skepsi vědců v jeho spojení se statistikou budu za hlavní

zdroje této části práce považovat hlavně metodologické články. Autoři těchto článků jsou zejména Lawrence R. James, Jeanne M. Brett, nebo David A. Kenny. Důležitou odbornou publikací je pak práce Introduction to Statistical Mediation Analysis od Davida P. MacKinnona, která poskytuje jak teoretická východiska, tak mnoho příkladů aplikací od biologie po sociologii.

V teoretické části práce bude představen rozdíl mezi mediací a moderací, a také popsány jednotlivé modely mediace. Jako první bude představen jednoduchý model mediace. Hodnotou mediačního efektu v tomto modelu je velikost efektu nezávislé proměnné na závislou proměnnou, který předpokládáme, že probíhá nepřímo, právě skrz mediátorovou proměnnou (mediátorová proměnná je ovlivněna nezávislou proměnnou a jejím prostřednictvím ovlivňuje závislou). Jedná se tedy o intervenující efekt. U složitějších modelů, které budou popsány v dalších kapitolách, je potřeba použít přístupy víceúrovňového modelování dat. Jedná se o úsekovou analýzu, využití modelování pomocí latentních proměnných, nebo longitudinální modely mediace.

V hlavní analytické části se pak již budu věnovat samotné konkrétní aplikaci mediace v sociologii. V této části práce budou použita konkrétní kvantitativní data, na kterých budu mediaci demonstrovat – budou provedeny vlastní výpočty, na kterých ukážu vhodnost daných přístupů v těchto konkrétních situacích. Poslední částí pak bude diskuse, ve které budou shrnuta klíčová zjištění práce.

### **Orientační seznam literatury**

Kenny D. A. (2008) Reflections on Mediation, *Organizational Research Methods*, Vol. 11, 353–358.

James L. R., Brett J. M. (1984) Mediators, moderators, and test for mediation. *Journal of Applied Psychology*, Vol. 69, 307–321. 14

Baron R. M., Kenny D. A. (1986) The moderator-mediator variable distinction in socialpsychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, Vol. 51, 1173–1182.



MacKinnon D. P., Krull J. L., Lockwood C. M. (2000) Equivalence of the mediation, confounding, and suppression effect. *Prevention Science*, Vol. 1, 173–181.

Shrout P. E., Bolger N. (2002) Mediation in experimental and nonexperimental studies: New Procedures and recommendations. *Psychological Methods*, Vol. 7, 422–445.

Hendl J. (2009) Přehled statistických metod. Analýza a metaanalýza dat. Portál, Praha.

Gelman, Andrew, Hill, Jennifer. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Analytical Methods for Social Research. New York: Cambridge University Press.

Bryk A. S., Raudenbush S. W. (1992) *Hierarchical linear models*. Thousand Oaks, CA, Sage.

MacKinnon, David P. *Introduction to Statistical Mediation Analysis*. (2008). London: Routledge

MacKinnon, D. P., Wurpts, I. C., & Valente, M. J. (2014). Imagery and memory theory as known effect validation for mediation analysis.

Bolger, N. & Laurenceau, J-P. (2013). *Intensive longitudinal methods: An introduction to diary and experience sampling research*. Guilford: New York

Fairchild, A. J., MacKinnon, D. P., Taborga, M. P., & Taylor, A. B. (2009). R<sup>2</sup> effect-size measures for mediation analysis. *Behavior Research Methods*, 41(2), 486-498.

V Praze dne 31.5.2016

Podpis konzultanta práce:

# Obsah

<b>1</b>	<b>ÚVOD</b> .....	<b>2</b>
<b>2</b>	<b>TEORETICKÁ ČÁST</b> .....	<b>5</b>
2.1	ÚVOD DO PROBLEMATIKY MEDIACE .....	5
2.2	MEDIACE VS. MODERACE .....	5
2.3	METODY HLEDÁNÍ MEDIÁTORŮ .....	6
2.4	PODOBY ANALÝZY MEDIACE .....	8
2.4.1	<i>Mediace pro spojitá data</i> .....	8
2.4.2	<i>Mediace pro kategoriální data</i> .....	16
2.5	MEDIACE V PRAXI .....	18
2.5.1	<i>Předchozí aplikace</i> .....	18
2.6	KRITIKA .....	20
2.7	ODLIŠNOST PŘÍSTUPŮ .....	21
2.7.1	<i>Lineární regrese</i> .....	21
2.7.2	<i>Logistická regrese</i> .....	22
<b>3</b>	<b>METODOLOGICKÁ KAPITOLA</b> .....	<b>25</b>
3.1	CÍLE PRÁCE, VÝZKUMNÉ OTÁZKY A HYPOTÉZY .....	25
3.2	SEKUNDÁRNÍ ANALÝZA DAT .....	26
3.2.1	<i>Výhody a nevýhody sekundární analýzy dat</i> .....	27
3.2.2	<i>Požadovaná povaha dat</i> .....	27
3.3	ANALYTICKÝ POSTUP .....	28
3.4	VYHODNOCENÍ KVALITY REGRESNÍCH MODELŮ .....	29
3.4.1	<i>Lineární modely</i> .....	29
3.4.2	<i>Binární logistické modely</i> .....	30
3.5	LIMITACE MODELOVÁNÍ .....	30
3.6	POUŽITÝ SOFTWARE .....	31
3.6.1	<i>SPSS PROCESS</i> .....	31
3.6.2	<i>Alternativní software – R</i> .....	32
3.7	ETIKA VÝZKUMU .....	34
<b>4</b>	<b>ANALYTICKÁ ČÁST</b> .....	<b>35</b>
4.1	PŘÍPADOVÁ STUDIE Č. 1 – GENERAL SOCIAL SURVEY 2016 .....	35
4.1.1	<i>Vybrané proměnné</i> .....	35
4.1.2	<i>Analýza pro spojitou závislou proměnnou</i> .....	36
4.1.3	<i>Analýza pro binární závislou proměnnou</i> .....	41
4.2	PŘÍPADOVÁ STUDIE Č. 2 - INTERNATIONAL SOCIAL SURVEY PROGRAMME .....	46
4.2.1	<i>Vybrané proměnné</i> .....	47
4.2.2	<i>Analýza pro spojitou závislou proměnnou</i> .....	47
4.2.3	<i>Analýza pro binární závislou proměnnou</i> .....	50
4.3	ZHODNOCENÍ VÝZKUMNÝCH OTÁZEK A HYPOTÉZ .....	54
4.4	DISKUSE S LITERATUROU .....	55
4.4.1	<i>Limitace modelů</i> .....	55
4.4.2	<i>Doporučení pro další zkoumání</i> .....	57
<b>5</b>	<b>ZÁVĚR</b> .....	<b>59</b>
<b>6</b>	<b>SUMMARY</b> .....	<b>61</b>
<b>7</b>	<b>ZDROJE</b> .....	<b>63</b>

# 1 Úvod

Předkládaná diplomová práce se zabývá analýzou mediace ve statistice a jejím využití v sociálních vědách. Analýza mediace je velmi klíčovou analytickou metodou pro mnoho odvětví (zejména psychologii, dále ale také sociální a biomedicínské vědy), která je čím dál častěji aplikována v rámci výzkumů a má stále větší a větší možnosti. Co se týče aplikace v sociologii, zatím není tato metoda příliš rozšířená. Navíc co se týká české literatury, jedná se o poměrně zanedbané téma, a to i přesto, že v zahraniční literatuře se objevuje již celá desetiletí.

Hlavním záměrem této práce je popis této metody, její aplikace na vybraná sociologická data a tím pádem prozkoumání možností, které tato metoda má v oblasti sociologie. V práci jsou porovnány dva přístupy, jakými se dá analýza mediace aplikovat (oba dva jsou založeny na příslušné regresní analýze). V prvním případě se jedná o jednoduchou analýzu mediace u spojité závislé proměnné (tedy použitím lineární regrese), v druhém případě se jedná o jednoduchou analýzu mediace u binární závislé proměnné (tedy použitím logistické regrese). Aby bylo porovnání co nejvíce objektivní, v jeho rámci se jedná o stejná data, tedy v příkladech je vybrána taková závislá proměnná, kterou lze posléze binarizovat. Zahraniční literatura, která se věnuje analýze mediace je doposud spíše zaměřena na případ první, tedy spojitou proměnnou.

Z tohoto důvodu si práce klade za cíl tyto dva přístupy porovnat, s čímž se pojí několik hlavních výzkumných otázek. První hlavní otázka je, zda se významně liší výsledky analýzy, pokud tatáž závislá proměnná je spojitá či binární. Výsledky obou analýz, které jsou provedeny, jsou porovnány a je sledováno, zda došlo k výrazným změnám v číselných výsledcích, případně v celé interpretaci modelu. Nejde ale pouze o srovnání číselných výsledků, ale také o porovnání postupu, s jakým byly modely vytvářeny. Analýza dat je provedena v programu SPSS, kde je nutné k těmto dvěma typům závislých proměnných u analýzy mediace přistupovat rozdílně. Pokud se jedná o spojitou proměnnou, je možné použít SPSS doplněk PROCESS, který byl specificky vytvořen nad rámec programu pro analýzu mediace a moderace. V případě, že se jedná o binární proměnnou, v tomto případě nelze pracovat s PROCESsem, ale je potřeba mediaci provést jako sérii několika regresních analýz. Práce tedy kromě podoby modelů zkoumá, jaké jsou další důsledky přechodu ze spojité proměnné na proměnnou binární, a to zejména co se týče samotného průběhu analýzy (například jak se liší výstupy, jaké

informace dané přístupy neobsahují, či jaké informace naopak obsahují navíc). Díky této části je práce také částečně návodem, jak využívat některé funkce doplňku PROCESS, který v českém jazyce též neexistuje. V tomto smyslu se práce snaží doplnit mezery v české literatuře, která se věnuje problematice statistických metod.

Práce je rozdělena do tří hlavních bloků, a to na teoretickou, metodologickou a praktickou (analytickou) část. Teoretická část se zabývá zejména uvedením čtenáře do problematiky analýzy mediace. První částí je popis, co je mediace a jakým způsobem se liší od moderace (jelikož tato dvě témata jsou v literatuře často spojována a mohlo by dojít k jejich záměně). Následně tato část popisuje, jakými různými způsoby můžeme mediátorové proměnné v našich výzkumech hledat. Druhá část teoretické části již poté popisuje konkrétně, jakým způsobem analýzu mediace aplikujeme, a to opět pro spojitá i kategoriální data. V případě spojitých dat je zde mnoho možností – práce se nejpodrobněji zabývá jednoduchou analýzou mediace, a to z toho důvodu, že tento přístup je později v práci aplikován. V teoretické části jsou ale popsány i alternativní přístupy, které můžeme pro spojitá data použít – složitější (několikanásobné) modely mediace, víceúrovňové modely, longitudinální modely atd. Další částí teoretického bloku je představení aplikací z předchozích výzkumů, které použily mediaci jako svůj analytický nástroj. Jedna představená případová studie je ze zahraničního prostředí z oblasti behaviorální psychologie, [Marcus, Forsyth: 2010] a druhá z českého prostředí z oblasti sportovních věd [Psotta et al.:2012].

V metodologické kapitole je popsána metodologie práce – z tohoto pohledu se jedná o sekundární analýzu dat. V této části je rozebráno, jaké konkrétní požadavky byly kladeny na data, která byla následně do analýzy vybrána (z hlediska cílové populace, metody sběru dat a požadované povahy analyzovaných proměnných). Sekundární analýza dat je zde jako metoda rozebrána obecně, jsou popsány její výhody a nevýhody. Součástí kapitoly je také popis, jakým způsobem se vyhodnocují regresní modely, tedy rozdíly mezi lineární a binární logistickou regresí. Na závěr kapitoly je popsán použitý software (SPSS PROCESS), plus popis alternativního softwaru, který je možné pro analýzu mediace využít – R mediation package (který ale v práci použit následně není, je popsán pouze teoreticky).

V analytické části pak dochází k samotné aplikaci analýzy mediace na naše konkrétní vybraná data. V první části jsou naše data popsána – v první případové studii se jedná o průzkum General Social Survey (GSS) a v druhé případové studii o data z ISSP (International Social Survey Programme). Data jsou analyzována nejprve ve své původní podobě, poté je závislá proměnná převedena do binární podoby. V závěru práce je poté provedena diskuse s literaturou, která má dvě části. První částí je diskuse limitů těchto modelů na základě dostupné literatury v porovnání s aplikací v práci. V této části jsou zmíněny přímo i limity práce jako takové. V druhé části diskuse jsou pak provedena doporučení k dalšímu zkoumání těchto metod. Na závěr práce jsou veškeré výsledky ještě jednou shrnuty a porovnány s hypotézami a výzkumnými otázkami.

## 2 Teoretická část

### 2.1 Úvod do problematiky mediace

Mediací se zabýváme z toho důvodu, že díky ní můžeme prozkoumat kauzální vztahy mezi proměnnými. Jedná se o zapojení třetí proměnné do vztahu mezi dvěma proměnnými, což umožňuje lepší zkoumání vztahů. Pokud zkoumáme vztah mezi dvěma proměnnými, zpravidla jednu označujeme jako příčinu (čili nezávislou neboli vysvětlující proměnnou) a druhou jako následek, tedy vysvětlovanou či závislou proměnnou). Pomocí mediace nicméně nezjišťujeme pouhý prostý vztah mezi těmito dvěma, ale snažíme se o hlubší analýzu procesů, které se v tomto vztahu odehrávají, a co může ovlivnit jejich směr a sílu.

Často se využívá v teoretických a aplikovaných oborech, jako je psychologie (sociální, klinická, vývojová, kognitivní), v sociálních a biomedicínských vědách. Klíčovým cílem těchto společenskovedních i ostatních oborů je identifikování kauzálních mechanismů, které ovlivňují a určují lidské chování a formování názorů. Přímé testování například psychologických teorií ve smyslu vztahu testování přímého efektu stimulu na následek není dostatečné – často se právě pracuje s dalšími koncepty (právě v psychologii například poznání nebo emoce) jakožto mediátorem, který tento vztah modifikuje [Imai, Keele, Yamamoto 2010: 52]. Jednou z nejčastějších aplikací jsou pak výzkumy intervencí a efektů programů.

Mediátorové proměnné jsou takové, které jsou nám schopny ukázat, jakým způsobem jsou propojeny jevy. Respektive mediátorová proměnná je jakýmsi přechodovým můstkem mezi závisle a nezávisle proměnnou. Nezávislá proměnná je příčinnou mediátorové proměnné, která je potom příčinnou závislé proměnné [MacKinnon 2008: 2-3].

### 2.2 Mediace vs. moderace

Primárním záměrem této práce je popsat a prozkoumat metody zkoumání mediátorových proměnných – je tedy potřeba si ujasnit rozdíl mezi mediací a moderací. Obě jsou takové proměnné, které umožňují lépe zkoumat vztah mezi závisle a nezávisle proměnnou. Mediátor je taková proměnná, která je přímo dosazena do vztahu mezi dvěma

proměnnými, vložena mezi ně, a ovlivňující celý model vztahu jako takový. Za moderátorovou proměnnou považujeme takovou, která pouze ovlivňuje sílu vztahu, případně směr mezi závislou a nezávisle proměnnou. Působením mediátoru například můžeme regulovat sílu působení nezávislé proměnné na závisle proměnnou, v případě že je tento účinek pro nás žádoucí (například v případě zavádění intervenčních programů). Moderátor tedy můžeme považovat za kvalitativní proměnnou [Baron, Kenny 1986: 1174]. Například při známé analýze rozptylu (ANOVA) ji můžeme považovat za interakci mezi ohniskovou nezávislou proměnnou a faktorem, který specifikuje přesné podmínky vztahu.

### ***2.3 Metody hledání mediátorů***

David MacKinnon definuje několik způsobů, jak můžeme hledat v datech potenciální mediátorové proměnné. Uvádí také, že v rámci jedné studie nebo analýzy dat je obvyklá kombinace několika z těchto metod. Vždy je potřeba si uvědomit, jakým způsobem by mohl mediátor ovlivnit závisle proměnnou (výstup) a podle toho se je snažit nalézt a evaluovat. Zdůrazňuje také, že v případě, že hledáme mediátory z toho důvodu, abychom zvyšovali účinnost programů, vždy je třeba jejich působení analyzovat a dále s nimi pracovat při vylepšování programů.

Jako příklady mediátorových proměnných v oblasti výzkumu prevencí uvádí například následující: tématem je prevence psychických nemocí, správné zacházení se stresem je určeno jakožto mediátor a terapie a vyrovnání se s nemocí jakožto výstup. Jako další příklady můžeme zmínit například prevenci těhotenství mladistvých, kde jako mediátorovou proměnnou určujeme správnou komunikaci mezi rodičem a dítětem během dospívání, případně prevenci užívání drog, kde jsou jako mediátory určeny společenské normy a nácvik odolávání pokušením [MacKinnon 2008: 36].

První způsob, jak najít mediátorovou proměnnou je jednoduše pokusit se jí identifikovat pomocí intuice. I přesto, že ne vždy se tato metoda může zdát správná, historicky se osvědčila například v medicínských výzkumech HIV/AIDS v dobách, kdy o nemoci byly velmi omezené znalosti.

Druhou metodou je metoda využití focus groups, které diskutují dané téma, případně problém a způsoby, jak mu předejít. Jako příklad je uvedena focus group deseti

mladistvých, kteří byli usvědčeni z řízení pod vlivem alkoholu. V rámci diskuse se snaží přijít na to, jakým způsobem by se dalo konzumaci alkoholu a tím pádem i řízení pod vlivem zabránit.

Třetím způsobem, jak najít mediátory, je důkladná rešerše literatury a předchozích studií na zkoumané téma. Ideálně buď k tématu existují již recenzované stati, v opačném případě je řešením studium jakýchkoli jiných článků a posudků, které se tématu týkají.

Čtvrtý způsob je založen na teorii. Někteří výzkumníci v dnešní době svou evaluaci podkládají teoretickými základy – jejich východiskem je, že pokud teorie v předchozích případech již fungovala, je třeba ji využít opakovaně, jelikož bude tím pádem fungovat stále.

Pátý způsob je samostatná studie proměnných, které by mohly potenciálně souviset s výsledkem programu, zaměřená na jejich korelace. Pokud tímto způsobem jsou prozkoumány klíčové proměnné, výzkumník nemusí mít přesnou představu, z jakého důvodu je právě daná proměnná mediátorem, nicméně mediátorem i přesto může efektivně být. Výstupem z takové studie jsou pak kvantitativní vztahy mezi výslednou proměnnou a všemi jejími mediátory.

Posledním způsobem je aplikace již předchozí analýzy mediace ve výzkumu na podobné téma. Ideálním případem je, pokud již byly mediátory tohoto typu programu jednou identifikovány, což nám napomůže k výběru dalších mediátorů a nastavení designu, který bude nejvíce efektivní [MacKinnon 2008: 40-41].

V této práci budou, stejně jako ve většině výzkumů, mediátory hledány kombinací metod. Ideální kombinace vhodná pro potřeby této práce je kombinace hledání pomocí důkladné rešerše literatury, dostupné teorie a intuitivního úsudku, jelikož tak můžeme spoléhat nejen na vlastní úsudek, ale také na teoretické základy.



## 2.4 Podoby analýzy mediace

Tato práce se zabývá rozdílem v přístupech k analýze mediace, pokud je závislá proměnná spojitá či kategoriální (v našem případě binární). V následující kapitole bude popsána odlišnost těchto přístupů z teoretického pohledu.

### 2.4.1 Mediace pro spojitá data

Je mnoho možností, jak přistupovat k analýze mediace, pokud jsou naše data spojitá. V této části bude popsáno několik těchto základních přístupů – jednoduchá mediace (tedy model, který bude následně použit pro aplikaci v analytické části), vícenásobná mediace, longitudinální mediační model, mediace s latentní proměnnou, víceúrovňový mediační model a path analysis.

#### Jednoduchá mediace

Tato kapitola se bude zabývat tím, na jakém principu stojí jednoduchá analýza mediace jako taková – tento postup bude poté aplikovaný na konkrétní data v analytické kapitole. Popsány budou čtyři regresní rovnice, které zprostředkovávají informace potřebné pro analýzu mediace, následně budou popsány testy, které ověřují vhodnost modelu a výpočet intervalu spolehlivosti. V těchto ukázkách uvažujeme proměnnou  $Y$ , která je závislá proměnná, a kterou můžeme považovat za výstup – například již zmíněná úroveň zvládnutí psychických nemocí.  $X$  je nezávislá proměnná, pod kterou si zde můžeme představit intervenční program.

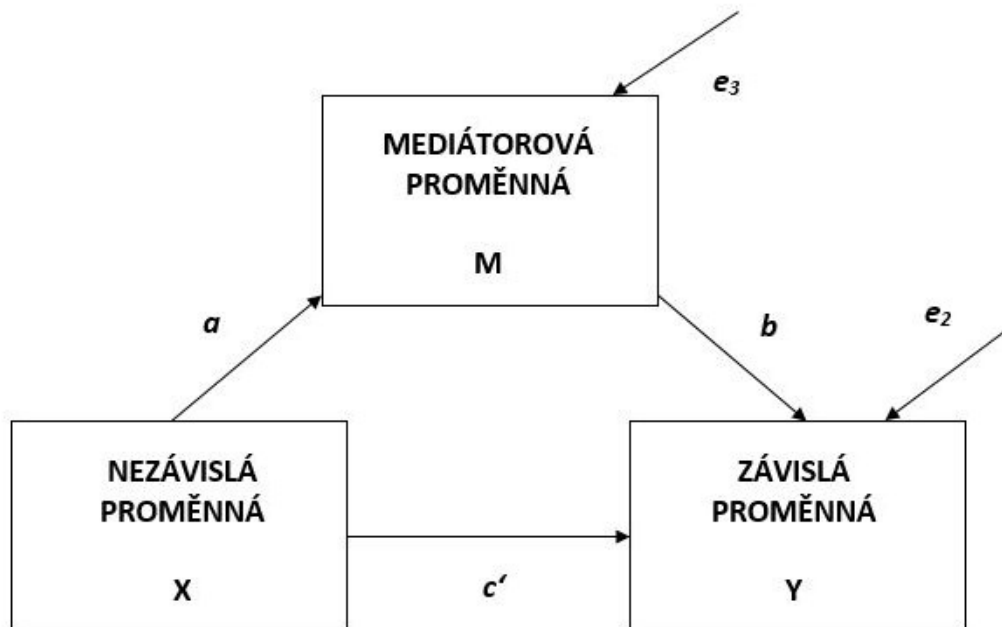
Schéma číslo 1 vyjadřuje model tří proměnných  $Y$ ,  $X$  a  $M$  (mediátor), kde probíhá mediace mezi těmito třemi. Kromě toho, že nezávislá proměnná ovlivňuje mediátor, který má následný vliv na závislou proměnnou, je zde také přímý vztah mezi závislou a nezávislou proměnnou, který existuje i nezávisle na působení mediátoru. Každá z šipek, která vyjadřuje směr vztahu, má u sebe popisek  $a$ ,  $b$  a  $c$ <sup>1</sup>

---

<sup>1</sup> Apostrof u vztahu  $c$  je zde z toho důvodu, že se jedná o vztah, který je upravený mediátorem. Pokud se v rovnicích uvedených níže objevuje pouze  $c$ , jedná se pouze o vztah mezi nezávislou a závislou proměnnou, který není upravený působením mediátoru.

Dále jsou zde uvedeny parametry  $e_2$ , který vyjadřuje část závislé proměnné, která není vysvětlena vztahem s nezávislou proměnnou a mediátorem, a  $e_3$ , který vyjadřuje část mediátoru, která není vysvětlena nezávislou proměnnou [MacKinnon 2008: 48-49].<sup>2</sup>

Obrázek č. 1 – Jednoduchá analýza mediace pro spojitá data



Zdroj: MacKinnon. *Introduction to Statistical Mediation Analysis*. 2008.

Při výpočtu jako takovém je pak potřeba provést několik regresních analýz.

- 1.1.  $Y = i_2 + cX + e_1$
- 1.2.  $M = i_3 + aX + e_3$
- 1.3.  $Y = bM + i_3 + e_2$
- 1.4.  $Y = i_2 + c'X + bM + e_2$

V první rovnici (1.1) provádíme regresi Y na X. Tento krok provádíme z toho důvodu, abychom prozkoumali vliv nezávisle proměnné na závisle proměnnou samostatně, bez působení mediátoru (z tohoto důvodu je zde uveden parametr  $c$  a ne  $c'$ , který je uveden v

<sup>2</sup> V modelu, kde se nevyskytuje mediátor, je část závisle proměnné, která není vyjádřena nezávisle proměnnou, označena jako  $e_1$ .

grafickém modelu výše). Tímto krokem zjišťujeme, zda je vůbec vhodný model, do kterého bychom potenciálně zapojili mediátor. V případě, že by mezi proměnnými žádný vztah modelovatelný pomocí regrese nebyl, nemělo by vůbec smysl o mediaci uvažovat. V případě, že mezi závisle a nezávisle proměnnou existuje významný vztah, tedy parametr  $c$  je nenulový, má smysl o zapojení mediátoru uvažovat [MacKinnon 2008: 49].

Druhým krokem (1.2) je provedení regrese  $M$  na  $X$ , tedy mediátoru  $M$  na nezávislou proměnnou. V tomto případě nás zajímá hodnota parametru  $a$ . V případě, že by hodnota  $a$  byla nulová a mezi těmito proměnnými nebyl vztah, nemohli bychom o  $M$  uvažovat jako o mediátoru. Obdobně, ve třetí rovnici (1.3) provádíme regresi  $Y$  na  $M$ , tedy závislou proměnnou (výsledek programu – v tomto případě míru vyrovnání se s duševními chorobami) ve vztahu k mediátoru (který v regresi vystupuje jako nezávislá proměnná). V tomto případě, pokud by hodnota parametru  $b_M$  byla nulová, není vztah mezi mediátorem a závislou proměnnou, respektive  $M$  nemůže být mediátorem. Tento třetí krok nicméně provádíme pouze v případě, pokud bychom předpokládali “úplnou mediaci”.

Posledním krokem je provedení mnohonásobné regrese  $Y$  na  $M$  a  $X$ . Tedy v našem příkladu regrese míry vyrovnání se s psychickými nemocemi na mediátoru (jakožto libovolného konstrukt, který vztah ovlivňuje) a programu jako takovém. V případě, že by parametr  $c'$  vycházel jako nulový, a naopak parametr  $b$  by vycházel nenulově, mohli bychom tvrdit, že celý efekt programu a vliv na vyrovnání se s psychickými nemocemi není způsobem programem, ale pouze mediátorem. Takový jev bychom nazývali “úplnou mediací”. V praxi se nicméně častěji setkáme s tzv. “parciální mediací”, tedy případem, kdy je výstup z programu ovlivněn jak programem jako takovým, tak působením mediátoru. Tato situace nastává v případě, že parametr  $c'$  (respektive jeho absolutní hodnota) je větší než nula, ale zároveň menší, než hodnota  $c$  v kroku 1 (1.1) [Hendl 2010: 3-4]

### **Vyhodnocení efektu mediace**

Celkový efekt, jakým působí nezávisle proměnná na závisle proměnnou je efekt, který jsme si výše definovali jako  $c$ . V tradiční studii by to byl celkový efekt, který by nás jako výzkumníky zajímal a který bychom pozorovali. Pokud ale do studie zapojíme analýzu

mediátorů, není tento celkový efekt až tak důležitý, ale zaměřujeme se na jiné, dodatečné informace, které nám poskytuje právě mediátor. K působení mediátoru může docházet i v situaci, kdy je nulový celkový efekt nezávisle proměnné na závisle proměnnou.

Jsou dvě možnosti, jakým způsobem vyhodnotíme velikost efektu. První možnost je taková, že se díváme na parametry  $a$  a  $b$  – jejich součin považujeme za mediační efekt – definujeme jako samostatný parametr  $ab$ . Je tomu tak z toho důvodu, že nezávislá proměnná  $X$  ovlivňuje nezávislou proměnnou  $Y$  nepřímo skrze mediátor  $M$ , efekt mediátoru tedy nazýváme nepřímým efektem (primární efekt, o který se jako výzkumníci v analýze mediace zajímáme) [Preacher, 2011: 95]. Efekt nezávislé proměnné na závisle proměnnou rozšířený o efekt mediátoru pak nazýváme přímým efektem. Mediační efekt je také roven rozdílu parametrů  $c$  a  $c'$ , tedy  $c - c'$ . Celkový efekt se tedy skládá z přímého efektu  $c'$  a nepřímého efektu  $ab = c - c'$  [MacKinnon 2008: 50].

### **Následné testy statistických hypotéz**

Autoři metody, která je popsána výše nicméně sami uznávají, že se nejedná o test mediace jako takový. Tato analýza nám ukazuje, zda a do jaké míry je přítomen mediační efekt, který můžeme zkoumat. V praxi je metoda odlišná v tom, že se používají i testy významnosti. Je tomu tak z toho důvodu, že pokud bychom zkoumali pouze koeficienty, může to být zavádějící, jelikož malé hodnoty koeficientů by mohly být významné v případech, kdy se jedná o velký výběr, naopak vysoké koeficienty by nemusely být statisticky významné v případě velkého výběru [Hendl 2010: 5].

Dalším důvodem k testování také může být, že v případě, že by byl nepřímý efekt při zavedení mediátorů ovlivněn jen nepatrně, mohlo by to i přesto znamenat, že absolutní hodnota působení mediátoru byla mnohem vyšší a tento postup to neodhalil. Argumentem pro testování hypotéz je také to, že pokud budeme přímo testovat hypotézu o rozdílu mezi celkovým efektem  $c$  a přímým efektem  $c'$ , jsme schopni mediační efekt odhalit příměji, než sérií regresních analýz [Preacher 2004: 719].

Testování významnosti mediačních efektů se provádí standardně tak, že je měřeno, zda se mediační efekt statisticky signifikantně liší od nuly. Jedním ze způsobů, jak toto otestovat je posouzení, zda je nula součástí intervalu spolehlivosti. Pokud se nula nachází vně intervalu spolehlivosti, mediační efekt je statisticky významný. Dále je možné

testování na základě standardní chyby (z-test), jehož hodnota je porovnána s hodnotou příslušného kvantilu normálního rozdělení. Pokud je absolutní hodnota testu vyšší, než 1,96, mediační efekt je statisticky odlišný od nuly na hladině významnosti 0,05 [MacKinnon 2008: 53].

### **Intervaly spolehlivosti**

Vzhledem k tomu, že velikost mediačního efektu je jen určitým odhadem, můžeme vytvářet jeho intervaly spolehlivosti. Díky intervalům spolehlivosti eliminujeme pravděpodobnost chyby odhadu a vytváříme možnost rozpětí hodnot efektu, místo prostého uvedení jediné hodnoty [MacKinnon 2008: 52]. Intervaly spolehlivosti standardně vytváříme symetrické, se stejnou hodnotou na horní hranici i dolní hranici mediačního efektu. V případě, že bychom chtěli vytvořit asymetrické intervaly spolehlivosti, odhad by byl podstatně přesnější, jelikož mediační efekt ve většině případů nedosahuje normálního rozdělení.

### **Mnohonásobná mediace**

V kapitole 2.4 byl popsán jednoduchý model mediace v případě, že do analýzy vstupuje pouze jeden mediátor. V praxi se ale mnohem častěji stane, že se nejedná o jednoduchý vztah mezi nezávislou a závislou proměnnou, do které vstupuje pouze jeden mediátor, ale často se jich objevuje více, které je potřeba do modelu zakomponovat. MacKinnon jako příklad z praxe uvádí model čtyř mediátorů, kde jako nezávislou proměnnou máme očekávání učitelů od svých studentů a jako závisle proměnnou studentovy výkony. Byla provedena metaanalýza dat 135 studentů a jako mediátory byly odhaleny: rozdílné sociální klima pro studenty, od kterých se více očekávalo, mnohem podrobnější feedback pro studenty, od kterých se více očekávalo, tendence učit více a na vyšší úrovni ty, od kterých se více očekávalo a tendence dávat studentům, od kterých se více očekávalo více příležitostí k odpovědi.

Pro naši analýzu je nicméně nejdůležitější, naznačit, jak se v případě více mediátorů změnil model popsáný v kapitole 2.4. Zůstává podobný, jako v předchozí kapitole, kde byla popsána jednoduchá mediace, ale v tomto případě se přidává ještě druhý mediátor. Kromě toho, že zde máme mediátor, který stavíme mezi nezávislou a závislou proměnnou, máme zde mediátor číslo dva, který stojí opět mezi nimi (nezávisle na prvním mediátoru) a vztah tak ovlivňuje také. Opět zde pracujeme s parametrem  $c$ , který popisuje

vztah mezi závislou a nezávislou proměnnou bez působení mediátorů a parametr  $c'$ , který popisuje vztah mezi nimi beroucí v potaz působení mediátorů [MacKinnon. 2008: 103-106].

### **Path analysis mediation models**

Modely mediace mohou být i více komplexní. Jedná se zejména o takové, které mají více než jednu nezávislou, případně závislou či mediátorovou proměnnou. Takovým modelům říkáme path analysis models [Statistics solutions: on-line]. S více než jednou závislou proměnnou nejsou již zpravidla vhodné metody, které se zakládají na lineární regresi – z toho důvodu, že korelace mezi závislými proměnnými nemohou být simultánně odhadnuty a musí tak být použit jiný přístup [MacKinnon 2008: 127].

Základem používané metody je tzv. analýza kovariančních struktur, tedy způsob, kdy je model evaluován na základě podobnosti kovarianční matice [Yuan, Bentler 1997: 767]. Možnosti v tomto případě jsou jednak „structural model“, který obsahuje informace o vztazích mezi konstrukty, případně „measurement model“, který popisuje, jakým způsobem dané konstrukty korelují, případně jaké mají vztahy k latentním či nevypozorovaným konstruktům (tedy ty, které nejsou přímo změřeny, ale je potřeba, aby byly odvozeny z konstruktů/proměnných, které změřeny jsou). Většinou obsahují dvě či více manifestních proměnných, které jsou „podezřelé“ z toho, že měří latentní konstrukt, s tím, že každá z nich pokrývá pouze jeho část, která může být oddělena od té části, která s latentním konstruktem nesouvisí. Speciální modely, kde pouze jedna manifestní proměnná měří jeden latentní konstrukt, se nazývají „manifest variable models“ [MacKinnon 2008: 128].

Se vzrůstající komplexitou modelů s větším počtem mediátorů je mnohem více složitější udržet přehled o informacích o všech parametrech v regresních rovnicích – z toho důvodu se v těchto metodách využívají kovarianční matice, které slouží k utřídění těchto informací. Když jsou matice vytvořeny, jsou dále použity maticové rovnice, které slouží k odhadu klíčových efektů, včetně efektů způsobených mediátorovými proměnnými a jejich standardních chyb. Maticové rovnice slouží také k lepšímu porozumění pomocným výpočtům [MacKinnon 2008: 130].

### **Mediace s latentní proměnnou**

Tento způsob rozšiřuje strukturní modely popsané v předchozí kapitole o možnost explicitně namodelovat chybu měření, díky čemuž je možné překonat problémy, které může způsobit při odhadu mediačního efektu. Chyba měření je běžnou záležitostí, se kterou se můžeme setkat ve všech odvětvích výzkumu včetně sociálních věd. Abychom porozuměli efektu, který má chyba měření na naše data, je nutné porozumět zkreslení, které je chybou způsobeno. Chyba může být náhodná, způsobená nesystematickými faktory, případně nenáhodná, způsobená systematickými faktory [Chescher 1991: 451]. Jsou dva klíčové aspekty měření, které mohou být ovlivněny; validita (tedy zda je měřeno to co má být měřeno) a reliabilita (tedy zda můžeme námi naměřené výstupy zobecnit).

Jeden ze způsobů, jak zredukovat chybu měření je specifikovat v modelu, jaký vztah mají individuální měření k hypotetickým, nebo latentním konstruktům, které nás zajímají. Tyto modely používají sady indikátorů latentních konstruktů pro modelování vztahů mezi indikátory a latentními konstrukty ve dvou rovinách – tedy skutečný vztah a chybu [MacKinnon 2008: 176].

### **Longitudinální mediační model**

Všechny modely, které jsou popsány výše, neberou nicméně v potaz měření stejných proměnných při opakovaných příležitostech. Z různých důvodů nicméně opakované měření zlepšuje možnosti interpretace procesů mediace, protože změna mezi jednotlivými případy může být porovnána se změnou u jednotlivých případů v čase [MacKinnon 2008: 193].

Některé aspekty longitudinálních dat jsou schopny proces mediace lépe osvětlit. Jsou schopny vypovídat více o aktuálních rolích závislé, nezávislé proměnné a mediátoru. Na rozdíl od jednoduchého mediačního modelu, kde jsou tyto vztahy měřeny pouze při jedné příležitosti a na základě této příležitosti evaluovány, je v této variantě možné tyto proměnné zkoumat pod jinými teoretickými nebo empirickými podmínkami. Longitudinální data také umožňují zkoumat, zda změny v mediátoru způsobují obdobné změny u závislé proměnné. Dalo by se říci, že je výhodou, pokud sběr dat probíhá alespoň ve třech vlnách – to již velmi znásobuje pravděpodobnost správné analýzy, evaluace a následných závěrů ohledně mediačního efektu [MacKinnon 2008: 193].

Další výhodou longitudinálních dat je pak to, že umožňují zkoumání změn jak u jednotlivých případů, tak mezi proměnnými („cross-sectional“). Je to klíčové zejména z toho důvodu, že změny jednotlivého případu mohou být zcela jiné než změny mezi různými případy. Například prediktory toho, proč jedna osoba dosáhla mnohem vyššího skóre u závislé proměnné mohou být zcela jiné než prediktory toho, že se v čase skóre jedné osoby výrazně zvýšilo oproti skóre ostatních – toto nám longitudinální data umožňují zachytit.

Třetí výhodou pak je, že longitudinální data umožňují alternativní vysvětlení „cross-sectional“ mediačních efektů. Díky změnám uvnitř případů jsou tak z vysvětlení vypouštěny efekty, které jsou dány statickými diferencemi mezi případy (tudíž každý jednotlivý případ se tak stává svou vlastní kontrolou). Například biologické faktory, jako je genetika, nejsou pravděpodobně vysvětlení longitudinálních vztahů, jelikož tyto proměnné se pravděpodobně nemění napříč vlnami měření [MacKinnon 2008: 195].

### **Víceúrovňový mediační model**

Multilevel mediation models jsou příkladem toho, jak pracovat s modely, pokud jsou data sbírána na více než jedné úrovni (tedy například školy, nemocnice či rodiny – v těchto případech jsou jednotlivé případy součástí jednotlivých skupin). Za skupiny také můžeme považovat zeměpisné skupiny, tedy města, státy, země. Tyto případy samozřejmě analýzu mediace sice poměrně výrazně komplikují, na druhou stranu zvyšují informační hodnotu, kterou analýza poskytuje.

Jelikož jedinci, kteří jsou součástí jedné skupiny sdílí některé společné charakteristiky, je také velmi pravděpodobné, že by mohli odpovídat stejným způsobem. Může to být způsobeno například komunikací mezi členy, podobným backgroundem, či podobným response biasem<sup>3</sup>. Důsledek toho pak je, že data jednotlivých případů na sebe nejsou

---

<sup>3</sup> Rozumíme tím jakékoli faktory, které brání respondentovi v přesné nebo pravdivé odpovědi. Nejčastěji k nim dochází ve studiích, které jsou založeny na respondentově sebe-evaluaci, u strukturovaných rozhovorů či dotazníků. Může pak docházet ke zkreslením, která mají výrazný vliv na validitu dat [Bradburn et al. 1978: 221].



navzájem nezávislá, čímž porušujeme předpoklady nezávislého pozorování, které jsou vyžadovány pro metody analýzy mediace popsané v kapitole 2.4.

## 2.4.2 Mediace pro kategoriální data

V kapitolách výše byly popsány přístupy, jakým způsobem se dá analýza mediace provádět, pokud data, na které analýzu aplikujeme, jsou spojitá. Záměrem této práce nicméně je i aplikovat analýzu na kategoriální (v tomto případě konkrétně binární) data, z toho důvodu bude v následující kapitole popsán i tento přístup.

V praxi se často setkáváme se případy, kde zkoumáme vztah mezi nezávislou proměnnou (proměnnými) a závislou proměnnou, která je binární. Často se s tím můžeme setkat zejména v medicínských výzkumech, které slouží například k prevencím různých nemocí – závislou proměnnou tedy v těchto případech je, zda je sledovaná nemoc přítomna či ne (proměnná může nabývat pouze těchto dvou hodnot, jelikož zde není možnost žádné částečně kladné či záporné odpovědi). Metody, které se v tomto případě nejčastěji používají jsou logistická regrese, případně probitová analýza<sup>4</sup> [MacKinnon 2008: 300].

### Postup při analýze

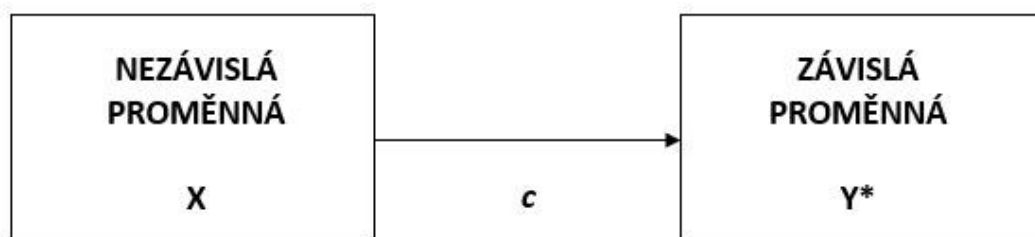
Stejně jako u spojitě proměnné je možné počítat velikost efektu mediace pomocí tří rovnic, které byly již popsány v kapitole výše. Průběh výpočtu, postup i podoba vztahu mezi závislou, nezávislou a mediátorovou proměnnou jsou v zásadě stejné, nicméně rovnice jsou pouze upraveny do podoby, která umožňuje výpočet s kategoriální závislou proměnnou.

$$2.1. \quad Y^* = i_1 + cX + e_1$$

---

<sup>4</sup> Probitová analýza, nebo také probitová regrese je velmi podobná logistické regresi. V tomto případě používáme normální rozdělení (místo logitu používáme z-skór). Výstupy jsou velmi podobné jako u logistické regrese, ale ne zcela stejné. Respektive koeficienty probitové regrese se rovnají zhruba 0,625násobku koeficientů logistické regrese. Existují aspekty, díky kterým může být v analýze mediace tato varianta preferována před logistickou regresí, respektive při pokročilem modelování (například pomocí programu MPlus) se v praxi využívá častěji [MacKinnon 2008: 302].

Obrázek č. 2 – Závislá a nezávislá proměnná

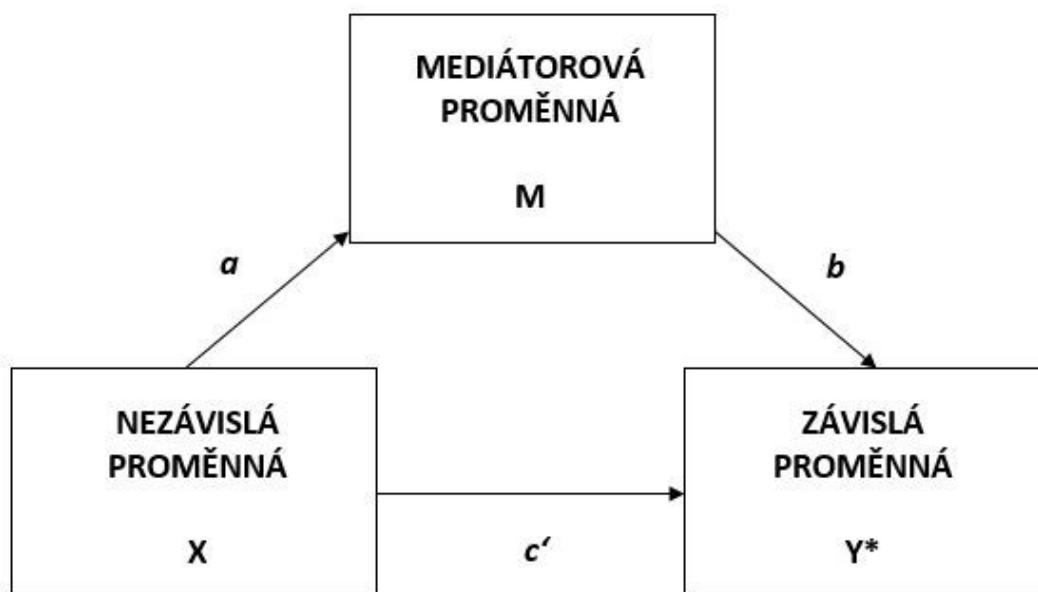


Zdroj: MacKinnon. *Introduction to Statistical Mediation Analysis*. 2008.

$$2.2. \quad Y^* = i_2 + c'X + bM + e_2$$

$$2.3. \quad M = i_3 + aX + e_3$$

Obrázek č. 3 – Jednoduchá analýza mediace pro kategoriální data



Zdroj: MacKinnon. *Introduction to Statistical Mediation Analysis*. 2008.

$Y^*$  je dichotomizovaná závislá proměnná (s hodnotami 0 a 1),  $X$  je nezávislá proměnná a  $M$  je mediátor. Konstanty a rezidua v příslušných rovnicích jsou popsány písmeny  $i_i$  a  $e_i$ .

Stejně jako je popsáno v kapitole 2.4, velikost mediačního efektu lze odhadnout pomocí difference mezi regresními koeficienty nezávislé proměnné, tj.  $c - c'$ , případně jako součin parametrů  $ab$  [MacKinnon 2008: 303-304]

## ***2.5 Mediace v praxi***

Pro ilustraci, jakým způsobem může aplikace mediace pomoci v praxi a zároveň jako příklad předchozích studií byly vybrány výzkumy, na kterých je postup v praxi popsán.

### **2.5.1 Předchozí aplikace**

V této kapitole bude popsáno, jakým způsobem byly metody analýzy mediace aplikovány v minulosti. Do této kapitoly byly vybrány dvě případové studie, které jsou zde popsány. První je z českého prostředí – výzkum Rudolfa Psotty z oblasti sportovních věd [Psotta et al.: 2012] a druhá ze zahraničního prostředí, tedy výzkumy, které provedly Bess H. Marcus a LeighAnn H. Forsyth [Marcus, Forsyth: 2010] v oblasti psychologie a behaviorálních věd.

#### **Případová studie č. 1 – Psotta et al.**

Záměr této studie [Psotta et al.: 2012] byl odhalit a popsat úroveň tělesné aktivity dětí v pozdějším školním věku s pohybovými obtížemi, porovnat je s dětmi bez pohybových obtíží a určit tak možné mediátory vztahu mezi úrovní tělesné aktivity a pohybovými obtížemi. Úroveň pohybové aktivity byla považována za závislou proměnnou, úroveň pohybových schopností za nezávislou proměnnou. Jako mediátor byla sledována úroveň pozornosti dětí.

Studie se zúčastnilo 212 dětí ze 6 základních škol ve 4 regionech v České republice. Nejprve prošli testováním na úroveň jejich pohybových (motorických schopností) za pomoci škály Movement Assessment Battery for Children-2 (MABC-2; Henderson, Sugden, & Barnett, 2007). Děti, které byly identifikovány jako případy s lehkými, nebo signifikantními motorickými obtížemi ( $n=15$ ), byly zahrnuty do studie ( $MD^+$ ). Druhá skupina bez motorických obtíží ( $MD^-$ ) byla vybrána náhodným výběrem z těch, které tyto obtíže nevykazovaly ( $n=27$ ). Úroveň pohybové aktivity byla pak zkoumána týmem výzkumných asistentů, pomocí akcelerometru v kombinaci s vyplňováním self-reportu po dobu sedmi dní.

Mediátor, tedy pozornost dětí, byl měřen pomocí dvou psychodiagnostických metod. Takzvaným „d2 testem“ selektivní pozornosti, který měří jak úroveň pozornosti, tak úroveň koncentrace. Úkol pro subjekt je za určitý časový limit vybrat pouze žádoucí vizuální objekt, v tomto případě písmena „d“ se dvěma čárkami nad nebo pod písmenem, během toho, co se dívá na čtrnáct řádků, kde se vyskytují i jiné matoucí objekty (písmena „d“ s méně nebo více čárkami, případně písmena „p“). Druhá metoda pak byla Numerical square test [Jirásek: 1975], který spočívá v tom, že respondent vidí čísla od 1 do 25 a snaží se je co nejrychleji seřadit do správného pořadí.

Jako statistická metoda sloužící k počítání výsledků byla použita metoda mediace dle Baron & Kenny modelu. Výsledky ukázaly na to, že jsou signifikantní rozdíly v pohybové aktivitě mezi dětmi skupin MD<sup>+</sup> a MD<sup>-</sup> (tedy děti spadající do kategorie MD<sup>-</sup> dosáhly povětšinou nižšího skóru v pohybové aktivitě, než děti kategorie MD<sup>+</sup>). K významným rozdílům došlo zejména v počtu kroků za týden, v průběhu pracovních dnů. Dle výsledků došlo také k tomu, že byla objevena statistická významnost mezi výsledky pohybových předpokladů a mediátorovou proměnnou, a to u vztahu mezi hrubou motorikou a energetickým výdejem za týden. Velikost mediačního efektu koncentrace dětí se pohybovala v rozmezí 12–22 % [Psotta 2012: 30-37].

### **Případová studie č. 2 – Marcus et al.**

Tato konkrétní studie [Marcus, Forsyth: 2010] se zaměřila na zkoumání efektivnosti dvou programů, které byly recipientům doručovány v tištěné podobě. Převážně se jednalo o zdravé dospělé jedince se sedavým typem zaměstnání. První program obsahoval „na míru šité“ motivační manuály, díky kterým bylo možné se informovat o individuálním pokroku (program byl založen spíše teoreticky). Druhá verze pak byl kontrolní program, který byl založen na manuálech Americké kardiologické asociace (AHA), ale v tomto případě se nejednalo o individualizované materiály (nicméně rozsah byl podobný jako v případě prvního, individualizovaného programu). I přes jejich univerzalitu nicméně skvěle doporučovaly, jakým způsobem může dojít k behaviorální změně, s níž má pak souviset nárůst pohybové aktivity. Oba druhy materiálů byly rozdány ve čtyřech vlnách – na začátku programu, po prvním, třetím a šestém měsíci.

Výsledky byly takové, že v obou případech programy vedly ke zvýšení pohybové aktivity. V první skupině, která obdržela individualizované materiály, byl pokrok nicméně větší než v kontrolní skupině. Z toho důvodu výzkumníci zkoumali odlišnosti v teoretických modelech, které byly použity, aby zjistili, co je skutečně pravou příčinou, která způsobila větší efektivitu individualizovaných materiálů. Prokázali, že behaviorální procesy, které byly součástí individualizovaných plánů vedly ke změně velice efektivně – tedy připomínání úkolů (jako uložení cvičebního oblečení na viditelné místo), přijímání osobních závazků, sebeodměňování a náhradní alternativy pomohly k lepším výsledkům. Pouze dva z těchto behaviorálních procesů dosáhly signifikantního výsledku i u kontrolní skupiny. Výzkumníci tedy na základě analýzy působících mediátorů vyvodili závěry, že ve chvíli, kdy mají jedinci možnost individualizované zpětné vazby o využívání behaviorálních procesů, začnou být povzbuzeni k užívání těchto procesů více, než by je povzbudila obecná standardizovaná informace o nich. Větší používání procesů zároveň zvyšuje šanci jedince dosáhnout doporučené úrovně pohybové aktivity [Marcus, Forsyth 2010: 52-53].

## **2.6 Kritika**

I přes veliké množství výhod a možností nových pohledů, které analýza mediace nabízí, v literatuře také čelí kritice. Green, Ha a Bullock například považují analýzu mediace za „Black Box experiment“ (tedy že v sociálních vědách je často deklarováno, že se skryté vztahy neboli „black box“ podařilo otevřít, ale velmi často tomu tak není). Považují ji za smysluplnou, zejména pokud se jedná o exaktní, technické vědy (jako je například medicína či biologie) – vědci, kteří zkoumají například účinek nějaké látky (jako příklad uvádějí mediátor prevence kurdějí – vitamín C), dovedou porozumět všem biochemickým procesům, na základě, kterých jsou schopni určit, jakým způsobem daná látka jakožto mediátor působí [Green, Ha, Bullock 2010: 200].

Sociální vědci se dle autorů této kritiky snaží pracovat obdobným způsobem, nicméně jen malá část z nich dovede vztah mediátoru k ostatním proměnným určit a popsat naprosto přesně (čímž jsou myšleny případy, které jsou podloženy solidními vědeckými důkazy, a ne pouhými předpoklady). Zejména kritizují právě regresní modely, které se v současnosti pro analýzu mediace nejhojněji využívají – tvrdí, že tyto modely stojí na naivních předpokladech. Jejich kritika spočívá zejména v následujících třech bodech.

Zprvė to, ťe konvenční regresní pŕístupy k analýze mediace stojí často na silných a nesplnitelných předpokladech. Zadruhé kritizují, ťe experimentální prostředí, ve kterém jsou sociálněvědní výzkumy často prováděny, je velmi nepraktické pro zkoumání kauzalit, jelikoť mohou být pouze zdánlivé, a tudíť je prakticky nemoťné zkoumat mediátory, pokud nejsou kauzální vztahy pevně stanoveny. Zatřetí, i pokud kauzální vztahy stanoveny jsou, je zpravidla konceptuálně mnohem sloťitější popsat vztahy s mediátory, než se na první pohled zdá [Green, Ha, Bullock 2010: 202].

I přes tyto argumenty samozřejmě autoři zdůrazňují, ťe svou kritikou rozhodně nechtějí naznačit, ťe by hledání mediátorových proměnných bylo zbytečné, nebo dokonce nemoťné. Naopak stanovování cest, pomocí kterých mapujeme vztahy mezi proměnnými, může mít obrovský teoretický a praktický význam. Problém, se kterým se setkáváme je spíše ten, ťe sociální vědci jsou často netrpěliví a ukvapení ve svých závěrech, a tudíť nedokáťí objektivně popsat, jak vlastně efekt vzniká.

Vzhledem ke komplexnosti analýzy mediace je třeba vždy počítat s tím, ťe závěry nelze vyvozovat jednoduše, ale ťe tento experimentální pŕístup vyžaduje podstatně dlouhodobější práci, zkoumání z různých úhlů pohledu atd. Je například potřeba si uvědomit, ťe existují různé podmínky, za kterých efekty působí silněji či slaběji a s těmito experimentálními situacemi odpovídajícím způsobem pracovat. V případě, ťe si výzkumník sám uvědomuje tyto limitace, je možné dosáhnout relevantních závěrů [Green, Ha, Bullock 2010: 203-208].

## ***2.7 Odlišnost pŕístupů***

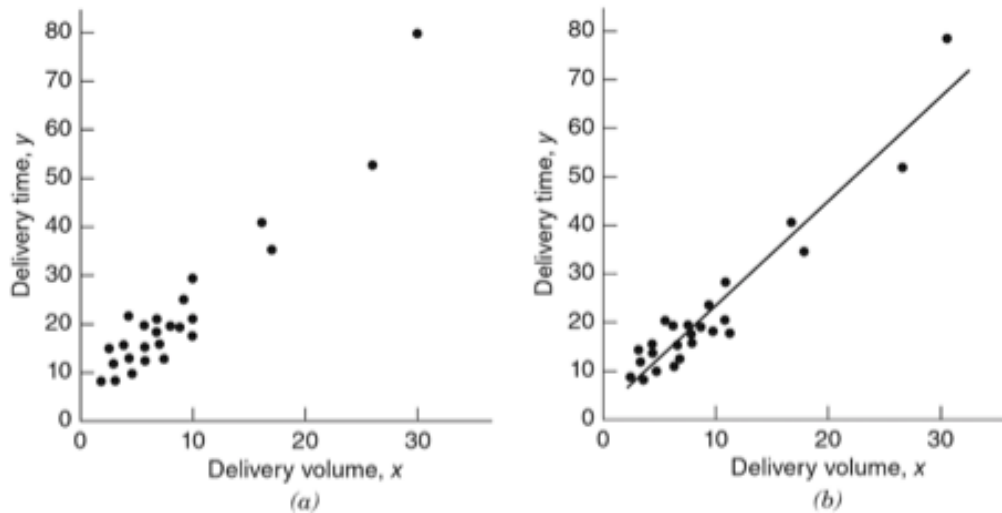
### **2.7.1 Lineární regrese**

Pro ilustraci, jaký je rozdíl mezi lineární a logistickou regresí a v čem tak spočívají specifika modelu, zde uvedu základní charakteristiky obou pŕístupů.

Lineární regrese je statistická technika, díky které můžeme vyjadřovat vztah mezi nezávislou a závislou proměnnou (za předpokladu, ťe obě z těchto proměnných jsou spojité) a predikovat možný průběh závislé proměnné. Lineární závislost vyjádříme tak,

že data proložíme regresní přímkou, která vztah definuje a zároveň predikuje [Montgomery et al. 2012: 2].

Obrázek č. 4 – Lineární regrese<sup>5</sup>



Zdroj: Montgomery, Peck, Vining. *Introduction to Linear Regression Analysis*. 2012

Pokud předpokládáme, že  $X$  je nezávislá proměnná a  $Y$  závislá proměnná, můžeme regresní křivku definovat jako následující vztah (vzhledem k tomu, že je zde pouze jedna závislá proměnná, jedná se o jednoduchou lineární regresi):

$$3.1. \quad y = \beta_0 + \beta_1 x + E$$

$\beta_0$  je konstantou,  $\beta_1$  je směrnici regresní přímky a  $E$  jsou chyby v predikci (odhadu) neboli rezidua (náhodné proměnné, které jsou do modelu zavedeny pro případ odchylek mezi skutečnou hodnotou  $y$  a očekávanou hodnotou  $y$ ) [Montgomery et al. 2012: 3].

## 2.7.2 Logistická regrese

Hlavní odlišnost, kterou má logistická regrese od klasické lineární regrese je, že závislá proměnná je binární nebo dichotomická. Je zde tedy rozdíl ve výběru parametrického

<sup>5</sup> V tomto případě modelujeme, jak množství zboží, které je potřeba doplnit do automatu na občerstvení ovlivňuje dobu, za kterou je naplněn. Na prvním grafu vidíme bodový graf jednotlivých případů, na druhém pak regresní přímkou, která je modelována na základě těchto případů, metodou nejmenších čtverců.

modelu a v předpokladech. Pokud je toto bráno v potaz, logistická regrese dále funguje na stejném principu, jako regrese lineární [Hosmer, Lemeshow 2000: 2]

Závislá proměnná  $Y$  může nabývat hodnot 0 a 1. Na základě hodnot nezávislých proměnných hodnotu závislé proměnné klasifikujeme, stejně jako u klasické lineární regrese, nicméně v tomto případě predikujeme pravděpodobnost, s jakou při dané hodnotě nezávislé proměnné nastane jedna nebo druhá hodnota závisle proměnné. Pojem pravděpodobnost zde nahrazujeme pojmem šance (že jev nastane, tj.  $Y=1$ ), kterou definujeme jako poměr pravděpodobnosti, že jev (který popisuje závisle proměnná) nastane a pravděpodobnosti, že jev nenastane ( $Y=0$ ) [Řeháková 2000: 476]. Tedy matematicky zapíšeme jako:

$$4.1. \quad P(Y = 1) = P(Y = 1) / [1 - P(Y = 1)]$$

Vzhledem k tomu, že šance má minimální hodnotu nula, provádíme její transformaci na její přirozený logaritmus. Tuto proměnnou nazýváme logit, její hodnoty se již pohybují od mínus do plus nekonečna a definujeme jí následujícím vztahem:

$$4.2. \quad \text{logit}(Y) = \ln \{P(Y = 1) / [1 - P(Y = 1)]\}$$

Co se týče následné interpretace, vždy je potřeba rozlišovat tyto tři pojmy, jejichž rozdíl byl vysvětlen výše. Častěji se pro interpretaci používá šance než logit.

Proměnné, které do logistické regrese vstupují jako nezávislé, jsou jak spojité, tak kategoriální. Je ale nutné, aby kategoriální proměnná byla ordinální – v případě, že by se jednalo o nominální proměnnou, u které by nebylo možné seřadit jednotlivé kategorie v pořadí, regresní model by nebylo možné vytvořit (což je případ i lineární regrese popsané v předchozí kapitole).<sup>6</sup> Rovnice modelu je následující:

---

<sup>6</sup> V tomto případě, pokud jsou proměnné nominální, řeší se to vytvořením indikátorových proměnných. Pokud má proměnná  $n$  kategorií, vytvoří se  $n-1$  nových indikátorových proměnných s tím, že každá z nich je jednou z kategorií původní proměnné. Jedna kategorie je vynechána (referenční kategorie), což je na našem uvážení, která. [Řeháková 2000: 477].



$$4.3. \quad \text{logit}(Y) = \alpha + \beta_1 X_1 + \dots + \beta_K X_K - 1 + \sum \beta_{Ki} D_{Ki}$$

Způsob, jakým interpretujeme regresní koeficienty je buď takový, že náš výstup jsou částečně standardizované regresní koeficienty, které nám dovedou seřadit nezávisle proměnné podle síly jejich vztahu k výstupu, druhým způsobem je pak vytvoření plně standardizovaných regresních koeficientů. Ty dokáží udat pořadí nezávislých proměnných podle síly vztahu, ale mohou být také interpretovány stejným způsobem jako standardizované koeficienty ve vícenásobné regresi. [Menard 2011: 1416].

### 3 Metodologická kapitola

V této části bude popsána metodologie diplomové práce. Jelikož se jednalo o sekundární analýzu dat, první fází bylo vyhledání dat, která byla vhodná pro účely této práce. Další fází bylo nalezení dvojice závislých a nezávislých proměnných, které do analýzy vstupovaly, a které splnily předpoklady, které jsou popsány v kapitole 2.4 (jednoduchá mediace s jednou závislou proměnnou, jednou nezávislou proměnnou a jedním mediátorem). V případě, že tato dvojice, potažmo trojice proměnných byla na základě splnění daných podmínek zařazena do analýzy, dalším krokem bylo nalezení a otestování příslušného mediátoru. Mimoto je v této kapitole popsáno, jakým způsobem vyhodnocujeme kvalitu regresních modelů, jaký byl použitý software (SPSS PROCESS) a jaký je alternativní software (R mediation package). Závěrem je v kapitole popsána etika výzkumu.

#### 3.1 Cíle práce, výzkumné otázky a hypotézy

Cílem této práce je zkoumání analýzy mediace – v literatuře, která se doposud tématu věnuje je mediace aplikována na situaci, kdy je závislá proměnná spojitá – je tedy možné analýzu provést s využitím klasické lineární regrese. V práci byla použita data, pomocí kterých bylo možné mediaci aplikovat a popsat.

Kromě provedení tohoto přístupu se nicméně práce věnovala i jinému problému – ve stejných datech byly vybrané proměnné převedeny na binární. Z toho důvodu bylo potřeba provést binární logistickou regresi. Výsledky těchto analýz byly následně porovnány a diskutovány, zkoumány byly také důsledky přechodu od lineární k logistické regresi. Z této části plyne výzkumná otázka, zda se liší výsledky analýzy mediace, pokud tatáž závislá proměnná je spojitá či binární, z čehož plyne hypotéza:

**H1:** Výsledky a kvalita modelů u analýzy mediace se neliší, ať je závislá proměnná spojitá či dichotomická<sup>7</sup>.

---

<sup>7</sup> V případě našich konkrétních případových studií.

Dalším cílem, který práce měla, bylo pak zkoumat rozdíly práce v SPSS, pokud používáme klasický modul regresních analýz, případně doplněk PROCESS. Doplněk PROCESS používáme v případě spojité proměnné, k druhému, klasickému přístupu se vracíme v případě dichotomické proměnné (se kterou PROCESS nedovede pracovat). I přesto, že předpokládáme stejné výsledky, co se týká modelů, proces jejich vytvoření se bude pravděpodobně lišit. Výzkumná otázka tedy zní, jakým způsobem se liší aplikace analýzy mediace, pokud používáme doplněk PROCESS oproti použití sady regresních analýz.

**H2:** V programu SPSS jsou výrazné rozdíly v aplikaci analýzy mediace, pokud použijeme doplněk PROCESS nebo aplikujeme mediaci jako sadu klasických regresních analýz (zejména ve výstupech, které jsou v případě použití doplňku PROCESS omezeny pouze na parametry, které se týkají pouze mediace jako takové).

*Související výzkumné otázky:* V čem konkrétně se liší výstupy, pokud při analýze mediace použijeme nebo nepoužijeme doplněk PROCESS? O co jsou tyto výstupy rozšířeny a co jim naopak chybí? Je možné modely interpretovat stejně? Jaké jsou výhody a nevýhody jednotlivých přístupů? Jak se liší syntax u jednotlivých přístupů?

### ***3.2 Sekundární analýza dat***

Jelikož se v této práci jedná primárně o zkoumání statistické metody mediace, z metodologického pohledu byla využita sekundární analýza již sesbíraných dat, tedy data, která byla již sebrána nějakým jiným výzkumným subjektem (výzkumník, výzkumná instituce, výzkumná agentura atd.). Data, která byla použita v této diplomové práci, jsou data ze studií GSS (General Social Survey) a ISSP (International Social Survey Programme), které jsou podrobněji popsány v analytické kapitole.

Sekundární analýza dat se od primární liší v tom, že při primární analýze analyzuje data přímo ten, kdo je sesbíral. Primární data bývají nejčastěji sbírána pomocí rozhovorů, dotazníků, pozorování, focus groups, případně dalšími metodami či kombinací těchto metod. Sekundární analýza je tedy alternativou, která zaprvé umožňuje výzkumníkovi přístup k většímu množství informací, než pokud by se omezoval pouze na data, která

sám sesbíral, a také odpovídat pomocí dat na výzkumné otázky, ke kterým data ve svém původním účelu nesloužila. [Vartanian 2011: 3-4].

### **3.2.1 Výhody a nevýhody sekundární analýzy dat**

Hlavní výhodou sekundární analýzy dat je zejména to, že umožňuje výzkumníkovi práci s mnohem větším objemem dat, než by mohlo být v jeho možnostech sesbírat. Nemusí se navíc omezovat na data, která sesbíral sám, ale má možnost kombinovat i s jinými, která mohou jeho výzkumný záměr doplnit i přesto, že jejich původní záměr byl jiný.

I přesto, že sekundární analýza dat má mnoho výhod, je potřeba si dávat pozor na její možné limitace a nevýhody. Jedním z argumentů, proč používat primární data by mohlo být to, že nemáme plnou kontrolu nad přesným zněním položených otázek, díky čemuž může být misinterpretován jejich původní význam. Dalším problémem sekundárních dat obecně je to, že ne vždy je nalezena odpověď na otázku, kterou bychom jako výzkumníci chtěli přesně zodpovědět z toho důvodu, že v původním výzkumu tato otázka zazněla trochu jinak.

Zjednodušeně by se tedy dalo říct, že při sekundární analýze dat výzkumník na úkor kontroly nad daty, která analyzuje, získává úsporu času, peněz a komfortu. Ztrátou kontroly se rozumí zejména kontrola nad způsobem, jakým byla data sbírána – i v případě, že přesně známe a rozumíme znění otázky, nikdy nevíme, jakým způsobem byla položena (zda zde například byl významný rozdíl mezi jednotlivými tazateli, do jaké míry osoba tazatele dokázala ovlivnit respondenta atd.). Nejsme schopni určit, zda respondenti přesně porozuměli otázkám či ne. Neznáme response rate ve výzkumu, ani nevíme, jakým způsobem data byla čištěna, jak byly imputovány chybějící hodnoty atd. [Vartanian 2011: 14-17].

### **3.2.2 Požadovaná povaha dat**

Do diplomové práce byly vybrány dva příklady dat, na kterých byla ilustrována mediace za použití modelů lineární a logistické regrese.

V této diplomové práci nebylo směřodonné, jaká byla cílová populace v jednotlivých výzkumech, jejichž data do analýzy vstoupila. Bylo potřeba se pouze ujistit, zda byl výběr respondentů dostatečně reprezentativní, nicméně nebyly kladeny žádné požadavky na charakteristiky respondentů, jako je věk, vzdělání a jiné sociodemografické faktory. Nutná nebyla ani konzistence v jejich charakteristikách napříč jednotlivými příklady.

Stejně jako nezáleželo na cílové populaci, nebylo také určeno, jakou metodou by měla být data sesbírána. Vzhledem k tomu, že byly důležité pouze jednotlivé proměnné a jejich charakteristiky, na metodě sběru dat nezáleželo.

V datech byly vyhledány příklady dvojic závislých a nezávislých proměnných, které vykazují určitý vztah. Prvotní ověření vztahu proběhlo použitím korelačních koeficientů (Pearsonův koeficient), které určily, zda má s daty smysl pracovat, zda mezi nimi existuje nějaká závislost. Z tohoto důvodu bylo potřeba, aby ve své původní podobě data byla spojitá [Bolboaca, Jantschi: on-line].

### ***3.3 Analytický postup***

U proměnných, které splnily tyto předpoklady pro provedení analýzy, byl pomocí metod popsaných v kapitole 2.3 nalezen příslušný mediátor, který by tento vztah mohl potenciálně ovlivnit, a následně bude tento vztah otestován metodou Baron & Kenny (pomocí sady několika vzájemně souvisejících regresních analýz).

Dalším krokem pak bylo závislou proměnnou, která v původním vztahu byla spojitá, převést na proměnnou binární a celý proces provést znovu, tentokrát za použití sady logistických regresí.

Na základě tohoto postupu byla v závěru práce provedena diskuse a srovnání použitých metod na základě dvou zmíněných příkladů. Dále byla vyhodnocena specifika, silné a slabé stránky jednotlivých přístupů a diskutována jejich vhodnost použití pro tento typ dat.

### 3.4 *Vyhodnocení kvality regresních modelů*

V analytické části byly mezi sebou porovnávány lineární a regresní modely stejných dat v jejich původní a binarizované podobě. Sledovány jsou tedy dva odlišné koncepty modelů, jejichž kvalita musí být vyhodnocena. Abychom mohli posoudit, který z modelů je pro zkoumání vhodnější, musíme sledovat jednotlivé parametry regresí.

#### 3.4.1 Lineární modely

V případě, že testujeme regresní model, vždy je jako první potřeba testovat významnost jednotlivých parametrů (t-test). Pokud testujeme jednotlivé parametry, stavíme nulovou hypotézu, že regresní parametr, který testujeme, je roven nule, proti alternativní hypotéze, že je různý od nuly. Testovací statistika má Studentovo rozdělení s  $(n-k)$  stupni volnosti.

Dále se používá F-test (omnibus test globální hypotézy), kdy se při modelu s konstantou  $\beta_1$  testuje nulová hypotéza  $H: \beta_2 = \beta_3 = \dots = \beta_k = 0$ , oproti které stojí alternativní hypotéza, že  $\beta_j$  není roven nule pro alespoň jedno  $j = 2, 3, \dots, k$ . Testové kritérium je statistika, která má při platnosti nulové hypotézy Fisher-Snedecorovo rozdělení s  $(k-1)$  a  $(n-k)$  stupni volnosti.

Co se týče výsledků těchto dvou testů – v případě, že F-test i všechny t-testy vychází jako statisticky významné, model je vhodný, jelikož vystihuje variabilitu proměnné  $Y$ , pokud jsou všechny testy statisticky nevýznamné, model vhodný není. Pokud je významný pouze F-test a nevýznamné některé jednotlivé dílčí t-testy, model vhodný být může, nicméně je potřeba zvážit vypuštění nevýznamných parametrů z modelu, pokud by byly nevýznamné všechny, jedná se o paradox, který je způsobem multikolinearitou (lineární závislostí mezi jednotlivými regresory) – model je tedy celkově vhodný, ale sám o sobě není významný žádný jeho člen [Neubauer: on-line].

Dalším z mnoha kritérií na jejichž základě se určuje kvalita regresního modelu je koeficient determinace ( $R^2$ ). Koeficient determinace určuje, kolik procent celkové validity je vysvětleno regresním modelem. Tento koeficient značíme jako  $R^2$  a používá se nejčastěji pro ohodnocení vhodnosti modelu. Koeficient determinace nabývá hodnot v intervalu mezi  $\langle 0, 1 \rangle$ . Čím je jeho hodnota vyšší (čím více se blíží k horní hranici 1), tím je model vhodnější, jelikož tím lépe popisuje daná data. Vyjadřujeme ho vztahem  $R^2 = S_T/S_Y$ , kde  $S_Y$  nazýváme celkovým součtem čtverců a  $S_T$  teoretickým součtem čtverců

[Mielcová, Stoklasová, Ramík: on-line]. Čím vyšší tedy bude hodnota koeficientu determinace (čím více se bude blížit k 1), tím můžeme regresní model považovat za kvalitnější a vhodnější.

### 3.4.2 Binární logistické modely

Stejně jako v lineární regresi je v logistických regresních modelech potřeba testovat významnost regresních koeficientů. V případě logistické regrese zcela stejně ověřujeme hypotézu, že regresní koeficient se liší od nuly. V logistické regresi můžeme použít tzv. Waldovo kritérium  $(b_i / s(b_i))^2$  - tedy příslušný koeficient dělený vlastní čtvercovou odchylkou. Test nicméně nemusí být vždy spolehlivý, pokud je hodnota koeficientu  $b$ , a tedy i jeho směrodatná odchylka, vysoká.

Další možnost, jak posoudit vhodnost logistického regresního modelu je chí-kvadrát test. Nízké hodnoty chí-kvadrátu ukazují na shodu (mezi pozorovanými a očekávanými četnostmi). Stejně jako u lineární regrese také v logistické regresi vystupuje parametr  $R^2$ , který nám vhodnost modelu popisuje, a můžeme tak mezi sebou porovnávat jejich kvalitu. Ve vyhodnocení modelu logistické regrese používáme také Hosmer-Lemeshow test, který ověřuje celkovou shodu modelu s pozorováním (overall goodness of fit) [Allison, Paul: on-line].

### 3.5 Limitace modelování

Co se týče přístupu k modelování jako takovému – v případě, že se rozhodneme pro vytvoření určitého modelu, vždy je potřeba si nejprve uvědomit, že data nemusela být vždy sesbírána za účelem vytvoření tohoto konkrétního modelu a z toho důvodu myslet na několik možných problémů. Tyto limitace se opět týkají zejména případu, kdy provádíme sekundární analýzu. Například je možné, že chybí nějaké klíčové proměnné či parametry, které výsledný fenomén reálně způsobují více než nezávislá proměnná, se kterou v modelu pracujeme, nicméně tato proměnná nebyla do šetření zahrnuta. Může se stát, že klíčové proměnné nebyly správně operacionalizovány. K chybě může také dojít při samotném výběru základního souboru, který je do šetření zahrnut – nemusí být vždy reprezentativní vzhledem k cílové populaci. Vždy je také dávat pozor na chybějící

hodnoty které nemusí být missing at random,<sup>8</sup> ale může se například jednat opakovaně o jednotky, které z výzkumu vypadly z nějakého důvodu již dříve, případně (zejména ve výzkumech zdraví nebo panelových výzkumech) jednotky již nejsou neschopny odpovídat z důvodu stáří nebo nemoci [Harrel 2001: 6-12].

### ***3.6 Použitý software***

V této kapitole je popsán software, který byl pro tuto sekundární analýzu dat použit. Jednalo se o doplněk PROCESS macro, který byl nainstalován nad rámec programu IBM SPSS 24 pro rozšíření jeho funkcí. Dále je zde popsán alternativní program, který lze pro analýzu mediace použít, a to R – mediation package (tento program nebyl v práci použit pro aplikaci, je zde představen pouze teoreticky).

#### **3.6.1 SPSS PROCESS**

Sekundární analýza dat byla v této práci provedena pomocí programu IBM SPSS 24 a to dvěma různými způsoby.

První možností byla instalace doplňku PROCESS (verze 2.16), jehož autorem je Andrew F. Hayes. Tento způsob analýzy byl vybrán z toho důvodu, že díky tomu se práce neomezila pouze na popis analýzy mediace, ale také popis funkcí tohoto doplňku.

Doplňek PROCESS lze kromě SPSS používat také ve statistickém programu SAS. Slouží zejména k analýze mediace a moderace, ale má i spoustu jiných možností (modely, které si přejeme počítat jsou zde označeny čísly, celkově v této verzi je 76 možností). Pokud chceme počítat jednoduchou mediaci, najdeme ji v dialogovém okně pod číslem 4. Jelikož PROCESS pracuje pomocí regresního přístupu k odhadům přímých a nepřímých efektů mediace, najdeme ho v záložce „Analyze -> Regression -> PROCESS“. PROCESS je možné stejně jako u klasických regresních analýz používat tím způsobem,

---

<sup>8</sup> Jedná se o jeden z typů chybějících dat. V případě, že jsou chybějící data upořádána missing at random, je chybějící odpověď spjata s pozorovanými daty, ale ne s hodnotou odpovědi jako takovou. Jako příklad můžeme považovat respondenta z dlouholetého panelu, který je ale již natolik nemocný, že nemůže odpovídat na události z minulosti, protože si je nepamatuje. Souvisí to tedy se stavem respondenta, ale ne s přímo událostí jako takovou. Jedná se o alternativu k MCAR (missing completely at random, chybějící hodnota nezávisí na hodnotách proměnné ani hodnotách odpovědí u jiných otázek) a NMAR (not missing at random, missing je spjat s hodnotou chybějící odpovědi jako takovou [Rubin 1976: 582].



že navolené příkazy nakopírujeme do syntaxového okna, nicméně i pokud se jedná pouze o jednoduchou mediaci, i v tom případě má syntax cca 5000 řádků (z toho důvodu není použitý syntax k práci přiložen).

Zřejmou limitací PROCESSu je fakt, že jej lze použít pouze v první části práce, a to pro případy spojitých proměnných. V případě, kdy je proměnná dichotomická, je potřeba přistoupit k jinému způsobu, tedy k sadě regresí, kterou popisují Baron & Kenny, tedy za využití SPSS v klasickém módu provést nejprve logistickou regresi X na Y, poté lineární regresi X na M (jelikož M zůstává spojitou proměnnou, není převedena na dichotomickou) a poté několikanásobnou logistickou regresi X a M na Y. Jedná se v podstatě o totéž, co je provedeno pomocí PROCESSu, pouze klasickou cestou, přes několik oddělených analýz.

### 3.6.2 Alternativní software – R

V následující kapitole bude zhruba popsáno, jakým způsobem je možné aplikovat analýzu mediace v prostředí programu R (jakožto možná alternativa pro SPSS PROCESS) a jakým způsobem můžeme interpretovat mediační efekty, včetně jednoho příkladu. Při použití analýzy mediace v R je potřeba mít nainstalovaný balíček Mediation package, který uživateli umožňuje zkoumat roli kauzálních mechanismů při použití různých typů dat a pozorovat, jakým způsobem se mění výstupy při změně předpokladů [Tingley et al. 2014: 2].

Mediation package v R dokáže vyhodnotit jednak koeficient “average direct effects” (ADE), tak “average causal mediation effects” (ACME). Vytváření modelu mediace probíhá ve dvou krocích – je potřeba specifikovat dva statistické modely. Model pravděpodobnostního rozdělení mediátoru  $M_i$  za předpokladu (nezávislé proměnné)  $T_i$ , a sady dalších ovlivňujících proměnných  $X_i$ , plus pravděpodobnostní rozdělení výstupu (závislé proměnné)  $Y_i$  za předpokladu daných  $T_i$ ,  $M_i$  a  $X_i$ . Tyto modely jsou konstruovány zvláště a umožňují výpočet mediační funkce, která odhaduje parametr ACME a další klíčové hodnoty.

Pro odhad těchto parametrů musíme, jak je již řečeno výše, specifikovat konkrétní proměnné, které do něj vstupují (pomocí příkazů *med.fit* a *out.fit*). V průběhu analýzy je možno dále specifikovat, kolik simulací bude odhad parametrů, nicméně Tingley et al.

doporučují využít co nejvyšší možný počet. Standardně používaný typ simulace se nazývá quasi-Bayesian Monte Carlo method a je založený na normálním rozdělení. Dále může být mediační funkce numericky shrnuta pomocí summary function, která počítá bodové odhady, intervaly spolehlivosti, p-hodnoty pro ACME, ADE a celkové mediační efekty [Tingley et al. 2014: 6-7].

Příklad takové tabulky výstupů je uveden níže.

Obrázek č. 5 – Analýza mediace v R

```

Quasi-Bayesian Confidence Intervals

              Estimate 95% CI Lower 95% CI Upper p-value
ACME (control)      0.0791      0.0351      0.1501      0.00
ACME (treated)      0.0804      0.0367      0.1557      0.00
ADE (control)       0.0206     -0.0976      0.1158      0.70
ADE (treated)       0.0218     -0.1053      0.1226      0.70

Total Effect        0.1009     -0.0497      0.2339      0.14
Prop. Mediated (control) 0.6946     -6.3109      3.6793      0.14
Prop. Mediated (treated) 0.7118     -5.7936      3.4965      0.14
ACME (average)      0.0798      0.0359      0.1537      0.00
ADE (average)       0.0212     -0.1014      0.1192      0.70
Prop. Mediated (average) 0.7032     -6.0523      3.5879      0.14

Sample Size Used: 265

Simulations: 100

```

Zdroj: Tingley et al. R Package for Casual Mediation Analysis. 2014

V tomto případě se jednalo o experiment, kde účastníci byli vystavováni různým mediálním příběhům a autoři zkoumali, jakým způsobem tyto vlivy ovlivnily jejich postoje a politické chování ve vztahu k imigrační politice.<sup>9</sup> Jako mediátorová proměnná zde byl určen strach.

V tomto konkrétním příkladu na první pohled vidíme, že parametr ACME je statisticky významně odlišný od nuly (můžeme vyčíst z p-hodnoty), ale ADE a celkový efekt ne. O

---

<sup>9</sup> Brader et. al, 2008

výsledku analýzy nám tabulka napovídá to, že mediátor, tedy strach, zvyšuje emocionální citlivost účastníků experimentu.

Mediační analýza v R pak umožňuje i další možnosti, jako je například využití alternativních simulací (např. nonparametric bootstrap), vytváření grafického znázornění parametrů, nebo další testování vztahu nezávislé proměnné a mediátoru pomocí testu (funkce TMint) [Tingley et al. 2014: 5-10].

### ***3.7 Etika výzkumu***

V diplomové práci je také potřeba se zabývat etickými souvislostmi, které jsou s výzkumem spojeny. Vzhledem k tomu, že veškerá data, která byla sekundárně analyzována jsou veřejně dostupná pro další zpracování a jejich zdrojová databáze řádně odcitována, z tohoto důvodu nedošlo k žádnému porušení autorských práv původního realizátora výzkumu.

Obecně, pokud provádíme sekundární analýzu dat, je potřeba dbát zejména na to, aby původní výzkum proběhl v souladu se základními etickými pravidly, tedy: účast subjektů na výzkumu je dobrovolná, jelikož informace, které výzkumný subjekt získá, mohou být velice osobní a pro okolí respondenta neznámé. Z tohoto důvodu je také třeba se ujistit, že byla dodržena anonymita a diskrétnost. Anonymity je dosaženo v případě, že výzkumník sice dokáže identifikovat respondenta ve spojení s jeho odpovědí, ale tyto údaje nikdy nezveřejňuje (v datech bývají jména a jiné osobní údaje respondentů anonymizována, případně nahrazena čísly) [Babbie 2014].

## 4 Analytická část

### 4.1 Případová studie č. 1 – General Social Survey 2016

Data, která byla vybrána pro první případovou studii, jsou ze studie General Social Survey [General Social Survey: on-line]. GSS je studie, která sbírá data o americké společnosti za účelem monitorovat a vysvětlovat trendy, případně konstanty v postojích, v chování a jiných vlastnostech. GSS v podobě, v jaké probíhá v současné době vznikla v roce 1972, nicméně předtím byla velká část otázek převzata v z jiných studií, tudíž můžeme určité trendy mapovat již více než 70 let [General Social Survey: on-line].

GSS obsahuje obrovský set různých otázek, a to jak demografických, tak behaviorálních či postojových. Co se týče témat, pokrývají civilní práva, zločin, násilí, meziskupinovou toleranci, morálku, psychické zdraví, sociální mobilitu, traumatické události, stres.

Data byla vybrána zejména z toho důvodu, že se jedná o nejlepší zdroj sociologických a postojových dat v americké populaci. Program během let svého působení prodělal spoustu metodologických změn a optimalizací, které vedly k tomu, aby použitím dat bylo co nejlépe možné provádět analýzu americké společnosti jako takové, ale i její porovnání s jinými státy. Díky tomu se jejich tvůrci ujistí, že jsou data zveřejňována v nejlepší kvalitě. Během procesu bylo napsáno více než 124 metodologických reportů, které sloužily k optimalizaci kvality dat [General Social Survey: on-line].

#### 4.1.1 Vybrané proměnné

Z důvodu, aby data byla co nejvíce aktuální, byla pro případovou studii vybrána nejnovější vlna, tedy data za rok 2016. Data byla volně dostupná ke stažení na webu General Social Survey. Co se týče konkrétní závislé proměnné, bylo potřeba, aby byla vybrána taková zkoumaná proměnná, která je v datech spojitá (tedy respondenti odpovídali konkrétním číslem, nesmělo se jednat o ordinální proměnnou, např. Likertovu škálu) a zároveň bylo možné pro účely porovnání metod proměnnou převést na dichotomickou. Z tohoto důvodu byla jako závislá proměnná vybrána proměnná „socioekonomický status respondentova partnera – cosei 10“, která nabývá libovolných hodnot od 0 do 100. Jako nezávislá proměnná, která by měla tuto proměnnou predikovat, byla vybrána proměnná „počet dokončených let vzdělání – educ“ a jako potenciální mediátor, který by měl do tohoto vztahu vstupovat, byla vybrána proměnná

„socioekonomický status respondenta – sei10“. Předpokládali jsme tedy, že čím vyšší bude vzdělání respondenta (tedy konkrétně kolik let vzdělání absolvoval, z důvodu zachování spojitosti proměnné), tím vyšší bude socioekonomický status partnera, kterého si vybere. Dále by do tohoto vztahu měl vstupovat vlastní socioekonomický status respondenta, jakožto mediátor.

#### 4.1.2 Analýza pro spojitou závislou proměnnou

Při výpočtu modelu jednoduché mediace pomocí PROCESS doplnku můžeme poměrně jednoduše a přehledně analyzovat části analýzy a jednotlivé dílčí procesy, které v modelu probíhají a které byly popsány v teoretické části výše. V zásadě se jedná o sadu několika lineárních regresí, tedy:

- 1) proměnná X predikuje Y – path c
- 2) proměnná X predikuje M – path a
- 3) proměnné X a M predikují Y
  - a) proměnná M predikuje Y – path b
  - b) proměnná X již nepredikuje Y, nebo je tento vztah menší – path c'

Tyto jednotlivé části jsou popsány v uvedených výstupech. Procesy (tedy jednotlivé body 1, 2 a 3) jsou v tomto pořadí seřazeny záměrně, jelikož na sebe logicky navazují. Pokud není splněn předchozí předpoklad, další již nemá smysl atd. V SPSS jsou nicméně výstupy seřazeny v abecedním pořadí, tudíž je potřeba si správně tyto dílčí procesy seřadit a interpretovat.

Tabulka č. 1 – Shrnutí modelu

Model = 4 $Y = cosei10$ $X = educ$ $M = sei10$
---

Zdroj: General Social Survey 2016. Vlastní výpočty.

První část výstupu, který doplněk PROCESS vypočítá, je shrnutí, co bylo počítáno. Tabulka je shrnutí proměnných a jejich předpokládaných vztahů, tedy v tomto případě Y

(cosei10 – socioekonomický status respondentova partnera, závislá proměnná), X (educ – počet dokončených let vzdělání respondenta, nezávislá proměnná), M (sei10 – socioekonomický status respondenta – mediátor).

Tabulka č. 2 - Model Summary

Outcome: sei10 (M)

R	R-sq	MSE	F	df1	df2	p
0.5051	0.2551	317.3630	79.1067	1.0000	231.0000	0.0000

Zdroj: General Social Survey 2016. Vlastní výpočty.

Tabulka č. 3 - Model

	coeff	se	t	p	LLCI	ULCI
constant	-11.0980	6.1325	-1.8097	0.0716	-23.1808	0.9848
Educ	3.9683	0.4462	8.8942	0.0000	3.0892	4.8474

Zdroj: General Social Survey 2016. Vlastní výpočty.

V prvním modelu figuruje nezávislá proměnná (educ) a mediátor (sei10). Jedná se tedy o proces číslo 2 – proměnná X predikuje M. Otázkou, kterou je zde, i v ostatních procesech, třeba si zodpovědět je, zda je tento model statisticky významný, což můžeme posoudit několika způsoby. V první tabulce tedy můžeme vidět globální F-test, který nám určuje statistickou významnost (nenulovost) všech regresních koeficientů. Můžeme si jeho hodnotu porovnat s tabulkami F-rozdělení, případně stačí porovnat dopočítané p-hodnoty s hladinou statistické významnosti  $\alpha$ . Pokud je p-hodnota nižší, než hladina významnosti  $\alpha$  (pro účely této práce budeme předpokládat její hodnotu 0,05), můžeme zamítnout nulovou hypotézu, že jsou regresní koeficienty nulové – což v tomto případě tak je. Vidíme, že i hodnota  $R^2$  je poměrně vysoká (0,2551).

V druhé tabulce se můžeme podívat na testy jednotlivých dílčích regresních koeficientů, které jsou také důležité. V tomto případě se jedná o výsledky t-testů nulové hypotézy, že jsou koeficienty nulové, tudíž statisticky nevýznamné. Hodnotu t-testů můžeme opět porovnat s tabulkou Studentova rozdělení, případně nám opět stačí pracovat pouze s p-hodnotou. V tomto případě je dopočítaná p-hodnota také nulová, tudíž koeficienty jsou statisticky významné (nenulové). Můžeme také vidět intervaly spolehlivosti koeficientů, které nám také potvrzují, že koeficienty nejspíše nulové nejsou. Vzhledem k tomu, že

došlo ke statisticky významnému výsledku v případě F-testu i dílčích t-testů, můžeme tento model považovat za vhodný. Bod číslo 2 jsme tedy splnili (proměnná X predikuje M). Z tohoto modelu vyplývá, že vzdělání respondenta predikuje socioekonomický status respondenta.

Tabulka č. 4 - Model Summary

Outcome cosei10

R	R-sq	MSE	F	df1	df2	p
0.3943	0.1555	450.4285	21.1749	2.0000	230.0000	0.0000

Zdroj: General Social Survey 2016. Vlastní výpočty.

Tabulka č. 5 - Model

	coeff	se	t	p	LLCI	ULCI
constant	9.3935	7.3575	1.2767	0.2030	-5.1032	23.8902
sei10	0.3151	0.0784	4.0204	0.0001	0.1607	0.4696
educ	1.4694	0.6159	2.3860	0.0178	0.2560	2.6829

Zdroj: General Social Survey 2016. Vlastní výpočty.

V této části výstupu je popsán třetí proces, tedy zda nezávislá proměnná a mediátor společně predikují závislou proměnnou (tedy vícenásobný regresní model). V první tabulce opět na první pohled vidíme, zda je model statisticky významný či ne – v tomto případě se na první pohled již zdá, že model statisticky významný je – dopočítaná p-hodnota globálního F-testu je 0,000, tedy stále nižší, než hladina významnosti (což by nám mělo naznačit, že dle výsledku testu jsme zamítli nulovou hypotézu, že jsou všechny koeficienty nulové a přijmuli alternativní, že nulové nejsou). Tento výsledek je poměrně pozitivní, jelikož jsme tím splnili další podmínku, díky které můžeme identifikovat a interpretovat mediaci. Budeme tedy v analýze nadále pokračovat.

Když se podíváme do druhé tabulky, nalezneme v ní informace o jednotlivých koeficientech. Další informace (proces 3a), kterou získáváme je, zda mediátor predikuje závislou proměnnou (řádek sei10), tedy zda je příslušný regresní koeficient statisticky významný. V tomto případě p-hodnota není zcela nulová, nicméně zůstává na hodnotě 0,01, což je stále méně, než 0,05, což je pro nás příznivý výsledek. Nyní se podívejme na krok 3b, tedy zda již X nepredikuje M, případně je tento vztah menší – tudíž bychom

předpokládali, že hodnota koeficientu zde nebude statisticky významná (což by byl ideální výsledek), případně že dopočítaná p-hodnota bude vyšší, než v případě, že predikujeme vztah pouze mezi závislou a nezávislou proměnnou – což v tomto případě vyšlo přesně dle předpokladu, jelikož je p-hodnota (v řádku educ) rovna hodnotě 0,0178 (což v další části modelu uvidíme, že je vyšší než při popisu vztahu mezi X a Y). Nedošlo tedy k ideálnímu výsledku, regresní koeficient je stále statisticky významný, nicméně je slabší než v následujícím kroku (je zřejmé i z intervalu spolehlivosti, který se pohybuje mezi hodnotami 0,2560 a 2,6829). Dochází tedy zřejmě k parciální mediaci.

#### Total Effect Model

Tabulka č. 6 - Model Summary

Outcome: cosei10

R	R-sq	MSE	F	df1	df2	p
0.3101	0.0961	479.9967	24.5729	1.0000	231.0000	0.0000

Zdroj: General Social Survey 2016. Vlastní výpočty.

Tabulka č. 7 - Model

	coeff	se	t	p	LLCI	ULCI
constant	5.8961	7.5419	0.7818	0.4351	-8.9636	20.7558
educ	2.7200	0.5487	4.9571	0.000	1.6389	3.8011

Zdroj: General Social Survey 2016. Vlastní výpočty.

V této části výstupu se díváme na celkové efekty v našem modelu, což nás pomyslně vrací k prvnímu procesu (tedy zda nezávislá proměnná predikuje závislou proměnnou). U tohoto modelu vidíme opět příznivý výsledek. Z celkového F-testu o modelu je zřejmé, že celkový model je statisticky významný (jelikož p-hodnota vychází 0,000, tedy nižší než hraniční hodnota 0,05). Hodnota  $R^2$  je pro nás také příznivá, 0,0961, tedy téměř 10 %.

Podívejme se samozřejmě i na hodnotu jednotlivých regresních koeficientů. V tomto případě vidíme, že tento model může být vhodný, jelikož je zde určitý vztah mezi závislou a nezávislou proměnnou. Srovnajme si to ještě s následující částí výstupu.



## Total, direct and indirect effects

Tabulka č. 8 - Total effect of X on Y

Effect	SE	t	p	LLCI	ULCI
2.7200	0.5487	4.9571	0.0000	1.6389	3.8011

Zdroj: General Social Survey 2016. Vlastní výpočty.

Tabulka č. 9 - Direct effect of X on Y

Effect	SE	t	p	LLCI	ULCI
1.4694	0.6159	2.3860	0.0178	0.2560	2.6829

Zdroj: General Social Survey 2016. Vlastní výpočty.

Tabulka č. 10 - Indirect effect of X on Y

Effect	Boot SE	BootLLCI	BootULCI
1.2506	0.3782	0.5703	2.0583

Zdroj: General Social Survey 2016. Vlastní výpočty.

V této části vidíme v podstatě zopakované to, co už jsme zjistili výše, nicméně výhodou doplňku PROCESS oproti sadě několika regresí je v tom, že nám mediační efekty sám vypočte a seřadí do těchto přehledných tabulek. Vidíme na první pohled, že celkový efekt závislé proměnné na nezávislou proměnnou zde je (první tabulka), tedy 2,72. Zajímavé je, že pokud se podíváme na přímý efekt, neovlivněný žádnou jinou proměnnou (druhá tabulka), je podstatně menší, dokonce 1,4694, a méně statisticky významný ( $\text{sig} = 0,0178$ ). Zbytek podílu na celkovém efektu pak vytváří efekt nepřímý, jehož velikost je 1,2506. Toto si můžeme ověřit ještě s intervaly spolehlivosti, u kterých se díváme, zda v sobě obsahují nulu či ne. U přímého efektu tomu tak není (interval spolehlivosti je ohraničen hodnotami 1,6389 a 3,8011), stejně tak je tomu u přímého efektu (kde je interval spolehlivosti ohraničen hodnotami 0,2560 a 2,6829). Interval spolehlivosti nepřímého efektu se pohybuje mezi hodnotami 0,5703 a 2,0583. Je tedy velmi pravděpodobné, že se velikost efektu okolo nuly nepohybuje, tudíž zde k mediačnímu efektu dochází.

Z tohoto plyne několik závěrů, které zde ještě jednou shrnu – vzhledem k tomu, že celkově je model statisticky významný, což plyne zejména z procesu číslo 1, můžeme říct, že nezávislá proměnná (tedy počet let vzdělání respondenta) predikuje závislou proměnnou (socioekonomický status respondentova partnera). Statisticky významný vztah je také mezi nezávislou proměnnou a mediátorem (socioekonomickým statusem respondenta) – zde je již dopočítaná p-hodnota celého modelu nulová. Z třetího procesu nám vyplývá, že pokud do analýzy zařadíme mediátor, vztah mezi závislou a nezávislou proměnnou je rázem slabší (t-test o nulovosti tohoto parametru vykazuje vyšší hodnotu sig), jelikož je ovlivněn mediátorem, což vidíme jak z této části procesu jako takové, ale potvrzuje se nám to i v tabulce, která popisuje jednotlivé efekty proměnných na sebe navzájem. V případě, že hodnotíme velikost pouze přímého efektu nezávislé proměnné na závislou proměnnou, tento efekt je menší, než když bereme v úvahu efekt celkový, tedy do vztahu započítáme i mediátorovou proměnnou „socioekonomický status respondenta“.

Jako závěr této případové studie tedy můžeme říci, že čím je vyšší počet let respondentova studia, tím vyšší socioekonomický status má partner, kterého si vybere. Tento vztah ale částečně ovlivňuje i vlastní socioekonomický status respondenta, jakožto mediátor.

#### 4.1.3 Analýza pro binární závislou proměnnou

V této části je model popsán v případě, že se jedná o stejná data, pouze závislá proměnná tentokrát není kardinální, ale kategoriální (binární). Naši závislou proměnnou Y je socioekonomický status respondentova partnera, popsáný číselnou hodnotou. Pro účely této analýzy jsme tedy překódovali v SPSS pomocí příkazu RECODE proměnnou tak, aby měla již pouze dvě hodnoty, a to hodnotu 0 („Nižší status“), pod kterou byly zakódovány hodnoty 0-50, tedy spodní polovina a 1 („Vyšší status“), pod kterou byly zakódovány všechny ostatní hodnoty, tedy hodnoty v horní polovině (51-100). Frekvenční tabulka této proměnné je následující:

Tabulka č. 11 – Partners socioeconomic status binary

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Lower	170	5,9	69,4	69,4
	Higher	75	2,6	30,6	100,0
	Total	245	8,5	100,0	

Zdroj: General Social Survey 2016. Vlastní výpočty.

Jak jsem již psala výše, pokud máme dichotomickou (nebo jakoukoli jinou kategoriální proměnnou) v modelu, nelze použít jednoduchý postup, který umožňuje doplněk PROCESS; nicméně i tak lze analýzu provést – je pouze potřeba ji provést postupně, pomocí sad několika (v tomto případě binárních logistických) regresí. Nevýhodou tohoto přístupu je, že mediační efekt musí být dopočten ručně odvozením z regresních koeficientů. Výhodou je, že můžeme postup jednotlivými procesy provést tak, jak máme nadefinováno v kapitole 2.4.2 v pořadí, které se hodí, a ne podle abecedy, jako to bylo v automatickém výpočtu. Výstupy jsou v tomto případě podrobnější, jelikož zde používáme postup, který je určen pro regresní analýzu jako takovou (tedy provedení regresní analýzy SPSS bez použití doplňků týkající se analýzy mediace). Popsány jsou zde tedy pouze nejrelevantnější výstupy a nejsou zde ukázány všechny tabulky.

Tabulka č. 12 - Omnibus Tests of Model Coefficients

		Chi-square	Df	Sig.
Step 1	Step	25,242	1	0,000
	Block	25,242	1	0,000
	Model	25,242	1	0,000

Zdroj: General Social Survey 2016. Vlastní výpočty.

Jako první provádíme logistickou regresi nezávislé proměnné na závisle proměnnou, tedy ptáme se, zda počet let vzdělání respondenta predikuje socioekonomický status, který bude mít jím vybraný partner. Tato tabulka je prvním indikátorem toho, že tento model by mohl být vhodný, stejně jako v případě předchozí kapitoly. V této tabulce můžeme najít srovnání našeho modelu s nulovým modelem, tedy takovým, kde figuruje pouze konstanta a všechny ostatní koeficienty (v tomto případě koeficient proměnné X) jsou nulové. Vidíme, že z tabulky zatím vypadá, že se zde setkáváme se statistickou odlišností našeho modelu od modelu nulového – dopočítaná p-hodnota je všude nulová, můžeme tudíž zamítnout nulovou hypotézu, že by v našem modelu byly všechny koeficienty nulové (tudíž že by náš model byl shodný s modelem nulovým).

Tabulka č. 13 - Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	272,128	0,100	0,141

Zdroj: General Social Survey 2016. Vlastní výpočty.

Dalším kritériem, pomocí kterého můžeme určovat, zda je model kvalitní či ne, jsou hodnoty pseudo koeficientu determinace. Na rozdíl od lineární regrese zde máme koeficienty dva, musíme také dbát na to, že tyto koeficienty nemohou být interpretovány stejně jako klasický koeficient determinace, navíc nemají stejné rozpětí hodnot. Koeficient Cox & Snell R Square nabývá maximální hodnoty, která je o něco nižší než 1, kdežto Nagelkerke R Square nabývá maximální hodnotu 1. Je tedy potřeba mít na paměti, že vzhledem k odlišné stupnici bude hodnota Cox & Snell R square vždy o něco vyšší. Platí ale, že čím je hodnota vyšší, tím je model vhodnější, tudíž v tomto případě jsou hodnoty koeficientů opět poměrně příznivé, což by nám měly potvrdit ještě poslední dvě tabulky.

Tabulka č. 14 - Hosmer and Lemeshow Test

Step	Chi-square	Df	Sig.
1	15,251	6	0,018

Zdroj: General Social Survey 2016. Vlastní výpočty.

Tabulka č. 15 - Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1	Highest year of school completed	0,274	0,058	22,294	1	0,000	1,316
	Constant	-4,549	0,824	30,508	1	0,000	0,011

Zdroj: General Social Survey 2016. Vlastní výpočty.

Tato tabulka nám v podstatě potvrzuje závěry z předchozích dvou. Vidíme, že model skutečně statisticky významný je. Posledním potvrzením je také Waldovo kritérium, což je koeficient dělený vlastní směrodatnou odchylkou  $(b_i / s(b_i))^2$ . Pokud je hodnota Waldova testu statisticky významná, můžeme říct, že proměnná patří do modelu a že je koeficient nenulový, což v tomto případě říct nedokážeme (jelikož je zde tato hodnota 0,240). Hosmer & Lemeshow test testuje shodu skutečných a odhadovaných hodnot (overall goodness of fit) – vidíme že dopočítaná p-hodnota je o něco vyšší než 0 (0,018), ale stále nižší než 0,05. Nulovou hypotézu o shodnosti tedy zamítáme.

Na rozdíl od předchozího postupu počítání tedy na první pohled vidíme, že tento model vhodný je, respektive že minimálně splňuje první z podmínek pro zkoumání analýzy mediace (zjevnou výhodou tohoto přístupu je, že tento výsledek vidíme hned, a ne až v několikátém kroku, kvůli abecednímu pořadí). Nezávislá proměnná tedy predikuje závislou proměnnou. Budeme tedy pokračovat v postupu dále abychom zjistili, zda jsou vztahy mezi proměnnými stejné, jako když jsme model popsali v kapitole 2.4.1, kde jsme pracovali se spojitou závislou proměnnou.

Následující regrese se týká nezávislé proměnné a mediátoru. Tedy pokročili jsme k procesu číslo 2 a zkoumáme, zda proměnná, kterou jsme si označili jako nezávislou, predikuje proměnnou, kterou jsme si označili jako mediátor. V tomto případě jsme museli použít klasickou lineární regresi, jelikož jsme zpátky u vztahu mezi dvěma spojitými proměnnými.

Tabulka č. 16 - Model Summary

Model	R	R square	Adjusted R Square	Std. Error of the Estimate
1	0,282	0,079	0,079	22,0951

Zdroj: General Social Survey 2016. Vlastní výpočty.

Vzhledem k tomu, že se jedná opět o lineární regresi, je to v podstatě stejný krok, jako v kapitole 4.1.2 v procesu číslo 2, jelikož jsme prováděli změny pouze v proměnné Y, ale zbylé dvě proměnné by měly zůstat stejné, a tudíž i jejich vztah. Vidíme, že se na první pohled výsledek příliš nezměnil ve vztahu k celému modelu, nicméně celkem překvapivé je, že výsledek není zcela stejný i přesto, že by se mělo jednat o totožný statistický test – v předchozí kapitole byl výsledek koeficientu determinace 0,2551, v tomto případě je nižší, tedy 0,079. Je tedy zřejmé, že výsledky při použití těchto dvou rozdílných metod nejsou zcela stejné. Tento výsledek koeficientu determinace je nicméně pro tuto část analýzy i přesto poměrně příznivý.

Tabulka č. 17 - ANOVA

ANOVA						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	114291,210	1	114291,210	234,111	0,000
	Residual	1324465,741	2713	488,192		
	Total	1438756,952	2714			

Zdroj: General Social Survey 2016. Vlastní výpočty.

Tabulka č. 18 - Coefficients

Coefficients						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	30,874	1,167		26,451	0,000
	Highest year of school completed	1,166	0,076	0,282	15,301	0,000

Zdroj: General Social Survey 2016. Vlastní výpočty.

V následujících tabulkách vidíme nejprve celkový F-test o regresním modelu. V tomto případě opět není nutné si jeho výsledek porovnávat s tabulkami, ale usuzujeme pouze z dopočítané p-hodnoty, která je nulová, tedy stejně jako v předchozí kapitole můžeme říct, že tato část modelu je statisticky významná. To samé můžeme říct o koeficientech, výsledek je tedy stejný – na první pohled vidíme, že vzdělání respondenta predikuje jeho vlastní socioekonomický status, tudíž tomu můžeme rozumět tak, že čím vyšší má respondent počet let ukončeného vzdělání, tím vyšší má svůj vlastní socioekonomický status. Tato část analýzy má tedy také výsledky, které potřebujeme, abychom mohli dále pokračovat.

Tabulka č. 19 - Variables in the Equation

Variables in the Equation							
		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1	Highest year of school completed	0,159	0,066	5,865	1	0,015	1,172
	R's socioeconomic index (2010)	0,029	0,008	12,046	1	0,001	1,029
	Constant	-4,262	0,847	25,300	1	0,000	0,014

Zdroj: General Social Survey 2016. Vlastní výpočty.

Posledním krokem je proces číslo 3, kde už uvádím pouze výslednou tabulku koeficientů, jelikož výsledek je přesně takový, jaký se dal očekávat. Vidíme, že proměnná, kterou jsme si označili jako nezávislou (počet dokončených let vzdělání respondenta) ovlivňuje proměnnou, kterou jsme si označili jako závislou (socioekonomický status respondentova partnera). Tento vztah je ale menší v případě, že do analýzy zařadíme i mediátorovou proměnnou, tedy respondentův vlastní socioekonomický status a tím pádem na základě toho můžeme říct, stejně jako v kapitole 4.1.2, že námi určený vztah mezi závislou, nezávislou proměnnou a mediátorem, nám byl na základě našich výpočtů potvrzen, a to i v tomto alternativním případě, kdy je závislá proměnná binární.

## ***4.2 Případová studie č. 2 - International Social Survey Programme***

Druhým datovým souborem, který byl použit pro účely aplikace postupů popsaných v této práci, byl dataset z výzkumu ISSP (International Social Survey Programme) [ISSP: online].

ISSP je mezinárodní program, který řídí každoroční výzkumy relevantní pro témata sociálních věd. ISSP byl založen v roce 1984 (zakládajícími členy byly Rakousko, Německo, UK a USA) a od té doby se výzkumů pod jeho záštitou zúčastnilo více než milion respondentů. Dotazníky jsou postaveny tak, že jsou mezi sebou porovnatelné na úrovni geografické (jednotlivé země) i časové (jednotlivé roky). ISSP také do svých výzkumů pravidelně přidává nová témata, podle toho, jak se vyvíjejí trendy sociálních věd (každá země má pro tyto změny jeden hlas). Metodologie výzkumů se stále vyvíjí také, na základě rozhodování metodologické komise, která sídlí ve Švýcarsku [ISSP: online].

Pro podobnost s tematikou předchozích dat byl vybrán datový soubor z roku 2012 s názvem International Social Survey Programme: Family and Changing Gender Roles IV – ISSP 2012. Datový soubor byl nalezen přímo skrze stránky ISSP, kde byl poté odkaz na volné stažení prostřednictvím datového archivu Gesis. Konkrétní proměnné, které se nicméně zdály nejvhodnější pro naši analýzu, se tentokrát netýkaly rodiny, ale úspěšnosti v zaměstnání respondenta.

### 4.2.1 Vybrané proměnné

Jako závislá proměnná byla zvolena proměnná „Počet zaměstnanců, na které respondent dohlíží/supervizuje“, tedy volně by se to dalo přeložit jako počet podřízených. Tuto proměnnou jsme mohli lehce analyzovat, jelikož je kardinální a zároveň z ní lze jednoduše udělat dichotomická proměnná (tedy hodnota „Ano“ pro ty, co podřízené mají a hodnota „Ne“ pro ty, co podřízené nemají). Jako nezávislou proměnnou, která by tuto proměnnou měla predikovat, jsme zvolili „Počet let vzdělání respondenta“ – předpokládali jsme, že tyto dvě proměnné budou souviset, jelikož čím vyšší má respondent vzdělání, tím vyšší by měl mít v zaměstnání postavení, tudíž i vyšší počet podřízených. Jako mediátor, který budeme zkoumat, byl zvolen respondentův věk, jelikož jsme předpokládali, že i ten do toho bude vstupovat (počet podřízených se zvyšuje i díky zásluhám).

### 4.2.2 Analýza pro spojitou závislou proměnnou

V případě spojitě proměnné byla opět analýza provedena pomocí doplňku PROCESS. Analýza proběhla sadou stejných regresí, které byly následně vyhodnoceny.

Tabulka č. 20 – Shrnutí modelu

Model = 4
Y = NSUP
X = EDUCYRS
M = AGE

Zdroj: International Social Survey Programme 2012. Vlastní výpočty.

První část výstupu shrnuje námi navržený model – „Počet podřízených“ jako závislá proměnná Y, „Počet let vzdělání“ jako nezávislá proměnná X a „Věk“ jako mediátor.

Tabulka č. 21 - Model Summary

Outcome: AGE (M)

R	R-sq	MSE	F	df1	df2	p
0.957	0.092	3120.8985	570.3168	1.0000	61752.0000	0.0000

Zdroj: International Social Survey Programme 2012. Vlastní výpočty.



Tabulka č. 22 - Model

	coeff	se	t	p	LLCI	ULCI
constant	45.6872	0.3129	146.0040	0.0000	45.0739	46.3006
EDUCYRS (X)	0.3754	0.0157	23.8813	0.0000	0.3446	0.4062

Zdroj: International Social Survey Programme 2012. Vlastní výpočty.

První model nám shrnuje druhý proces, kde se ptáme, zda proměnná X predikuje proměnnou M (opět začínáme druhým procesem, jelikož zde probíhá abecední řazení). Dle výsledků je tento první model statisticky významný. Vidíme to na první pohled z výsledku celkového F-testu o modelu, který vykazuje nulovou dopočítanou p-hodnotu. Když se podíváme na t-testy jednotlivých regresních koeficientů, vidíme znovu, že tento model nulový není, jelikož regresní koeficient se statisticky významně liší od nuly, což potvrzuje nejen nulová dopočítaná p-hodnota, ale také intervaly spolehlivosti, které ukazují, že regresní koeficienty nejsou nulové.

Tabulka č. 23 – Model Summary

Outcome: NSUP (Y)

R	R-sq	MSE	F	df1	df2	p
0.722	0.0052	4186683.38	161.5677	2.0000	61751.0000	0.0000

Zdroj: International Social Survey Programme 2012. Vlastní výpočty.

Tabulka č. 24 - Model

	coeff	se	t	p	LLCI	ULCI
constant	260.3514	13.2929	19.5858	0.0000	234.2973	286.4055
AGE (M)	1.9976	0.1474	13.5531	0.0000	1.7087	2.2865
EDUCYRS (X)	6.0491	0.5784	10.4582	0.0000	4.9154	7.1828

Zdroj: International Social Survey Programme 2012. Vlastní výpočty.

V tomto modelu vidíme proces číslo 3, tedy zda závislou proměnnou predikují nezávislá proměnná a mediátor. Vidíme, že i tato podmínka je splněna, jelikož celkový F-test má nulovou p-hodnotu, to samé i dílčí t-testy. Regresní koeficienty mají navíc poměrně vysoké hodnoty a z jejich intervalů spolehlivosti vidíme, že nejsou nulové. Zda probíhá mediace nebo ne zjistíme z následujícího regresního modelu.

## Total Effect Model

Tabulka č. 25 - Model Summary

Outcome: NSUP

R	R-sq	MSE	F	df1	df2	p
0.0474	0.0022	4199069.25	139.0372	1.0000	61752.0000	0.0000

Zdroj: International Social Survey Programme 2012. Vlastní výpočty.

Tabulka č. 26 - Model

	coeff	se	t	p	LLCI	ULCI
constant	351.6163	11.4780	30.6339	0.0000	329.1194	374.1132
EDUCYRS	6.7990	0.5766	11.7914	0.0000	5.6689	7.9292

Zdroj: International Social Survey Programme 2012. Vlastní výpočty.

Toto je abecedně poslední model, ale ve skutečnosti první proces, který se nás ptá na základ – tedy zda proměnná X predikuje proměnnou Y. Dupočítaná p-hodnota globálního F-testu je stejně jako u obou ostatních procesů nulová, můžeme tedy říct, že všechny tři procesy splnily své předpoklady, tudíž můžeme model vyhodnotit jako celek a určit, zda je proměnná „Věk“ skutečně mediátorem, případně velikost mediačního efektu. První hodnoty, které nás v tomto případě zajímají jsou hodnoty regresního koeficientu u prvního a třetího procesu. Pozitivním výsledkem je, že regresní koeficient proměnné X ve třetím procesu je menší než regresní koeficient v prvním procesu. To nám značí, že jeho hodnota poklesla ve chvíli, kdy do vztahu mezi proměnnými vstoupil mediátor, který také částečně závislou proměnnou Y ovlivňuje. Ideálním výsledkem v tomto případě by bylo, aby se regresní koeficient v této části snížil natolik, že by nebyl statisticky významný (což by znamenalo mediační efekt), nicméně i přesto to neznamena, že je výsledek špatný, mediace může být pouze parciální. O velikosti efektu se můžeme pak přesvědčit v posledních tabulkách.

Tabulka č. 27 - Total effect of X on Y

Effect	SE	t	p	LLCI	ULCI
6.7990	0.5766	11.7914	0.0000	5.6689	7.9292

Zdroj: International Social Survey Programme 2012. Vlastní výpočty.

Tabulka č. 28 - Direct effect of X on Y

Effect	SE	t	p	LLCI	ULCI
6.0491	0.5784	10.4582	0.0000	4.9154	7.1828

Zdroj: International Social Survey Programme 2012. Vlastní výpočty.

Tabulka č. 29 - Indirect effect of X on Y

	Effect	Boot SE	BootLLCI	BootULCI
AGE	0.7499	0.1464	0.5034	1.0976

Zdroj: International Social Survey Programme 2012. Vlastní výpočty.

Zde vidíme, že pokud bychom brali v úvahu pouze působení nezávisle proměnné na závisle proměnnou, velikost tohoto efektu by byla 6,0491. Celkový efekt je ale o něco větší a to 6,799. Co k velikosti efektu přispívá je právě působení mediátoru, jehož velikost je uvedena v poslední tabulce – tedy 0,7499 (o tolik se celkový efekt zvýší, pokud do vztahu zapojíme mediátor). Z intervalů spolehlivosti navíc vidíme, že tento efekt není nulový.

Z tohoto modelu tedy vyplývá, že se nám podařilo popsat vztah, který jsme původně předpokládali. Zdá se, že je pravda, že to, kolik má člověk podřízených je dáno tím, kolik let strávil ve škole, respektive se zvyšujícím se počtem let vzdělání se zvyšuje počet podřízených. Vztah mezi těmito dvěma proměnnými nicméně není takto jednoduchý, jelikož do modelu vstupuje věk respondenta, jakožto mediátor, který vztah ovlivňuje také – je tedy zřejmé, že na počet podřízených má vliv i věk respondenta (se kterým se pravděpodobně pojí i zkušenosti a zásluhy).

### 4.2.3 Analýza pro binární závislou proměnnou

V této části se opět budeme snažit porovnat výsledek v případě, že závislá proměnná bude binární. Proměnnou jsme překódovali pomocí příkazu RECODE do podoby, kde má již pouze dvě hodnoty, a to hodnotu 0 („Ne“), pod kterou byla zakódována pouze původní hodnota 0 (tedy že respondent uvedl, že nemá podřízené žádné), a 1 („Ano“), pod kterou byly zakódovány všechny ostatní hodnoty, tedy pokud respondent uvedl, že má libovolný počet podřízených. Frekvenční tabulka této proměnné je následující:

Tabulka č. 30 - Number of subordinates binary

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	No	47577	77,0	77,0	77,0
	Yes	14177	23,0	23,0	100,0
	Total	61754	100,0	100,0	

Zdroj: International Social Survey Programme 2012. Vlastní výpočty.

Tato analýza, jelikož se jedná o binární proměnnou, již také není prováděna pomocí doplňku PROCESS, ale sadou několika lineárních a binárních logistických regresí.

Tabulka č. 31 - Omnibus Tests of Model Coefficients

		Chi-square	Df	Sig.
Step 1	Step	259,044	1	0,000
	Block	259,044	1	0,000
	Model	259,044	1	0,000

Zdroj: International Social Survey Programme 2012. Vlastní výpočty.

Nejprve si porovnejme náš model s nulovým modelem. Na první pohled se zatím zdá, že náš model není nulový, jelikož v testu, který srovnává náš model s modelem nulovým s výsledky sig=0,000, tedy naše regresní koeficienty mají jiné hodnoty než nula.

Tabulka č. 32 - Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	66281,987	0,004	0,006

Zdroj: International Social Survey Programme 2012. Vlastní výpočty.

Méně pozitivní je tato tabulka, která nám ukazuje výsledek pseudo koeficientů determinace. Zde i v případě, že bereme v úvahu odlišnou interpretaci a odlišné rozpětí těchto koeficientů, není tento výsledek příliš příznivý.

Tabulka č. 33 - Hosmer and Lemeshow Test

Step	Chi-square	Df	Sig.
1	1358,969	8	0,000

Zdroj: International Social Survey Programme 2012. Vlastní výpočty.

Tabulka č. 34 - Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1	EDUCYRS	0,010	0,001	276,845	1	0,000	1,010
	Constant	-1,350	0,013	10947,774	1	0,000	0,259

Zdroj: International Social Survey Programme 2012. Vlastní výpočty.

U Hosmer & Lemenshow testu vzhledem k nulové hodnotě signifikance zamítáme nulovou hypotézu o shodě skutečných a odhadovaných hodnot. Pozitivní závěr nicméně můžeme vyvodit z dílčích t-testů, pomocí kterých můžeme vidět, že naše regresní koeficienty jsou statisticky významné (nenulové). Hodnota signifikance Waldova testu je příznivá (sig=0,000).

Závěr z tohoto procesu zatím navazuje na závěry modelu předchozí kapitoly, tedy první podmínka o vztahu nezávislé a závislé proměnné je splněna – nezávislá proměnná predikuje závislou proměnnou. Nyní bude testován druhý proces, tedy vztah nezávislé proměnné a mediátoru. Zde jsou obě proměnné spojitě, tudíž je využita lineární regrese.

Tabulka č. 35 - Model Summary

Model	R	R square	Adjusted R Square	Std. Error of the Estimate
1	0,096	0,009	0,009	55,86500

Zdroj: International Social Survey Programme 2012. Vlastní výpočty.

Tabulka č. 36 - ANOVA

ANOVA						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1779900,719	1	1779900,719	570,317	0,000
	Residual	192721726,200	61752	3120,899		
	Total	194501626,900	61753			

Zdroj: International Social Survey Programme 2012. Vlastní výpočty.

Tabulka č. 37 - Coefficients

Coefficients						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	45,687	0,313		146,004	0,000
	EDUCYRS	0,375	0,016	0,096	23,881	0,000

Zdroj: International Social Survey Programme 2012. Vlastní výpočty.

Když tyto výsledky srovnáme s předchozí případovou studií, vidíme, že zde se naopak výsledky druhého procesu (lineární regresní analýzy) neliší téměř vůbec. Koeficient determinace zůstává na nízké hodnotě 0,009, nicméně celkový F-test o modelu má příznivý výsledek (sig=0,000), stejně tak jednotlivé dílčí t-testy o regresních koeficientech. Zdá se tedy, že i druhá podmínka je splněna a závislá proměnná predikuje mediátor.

Tabulka č. 38 - Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	66137,761	0,007	0,010

Zdroj: International Social Survey Programme 2012. Vlastní výpočty.

Tabulka č. 39 - Hosmer and Lemeshow Test

Step	Chi-square	Df	Sig.
1	2559,085	8	0,000

Zdroj: International Social Survey Programme 2012. Vlastní výpočty.

Tabulka č. 40 - Variables in the Equation

Variables in the Equation							
		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1	EDUCYRS	0,009	0,001	231,899	1	0,000	1,009
	AGE	0,002	0,000	133,376	1	0,000	1,002
	Constant	-1,435	0,015	9164,593	1	0,000	0,238

Zdroj: International Social Survey Programme 2012. Vlastní výpočty.

V posledních výstupech je uveden výstup z procesu číslo 3, tedy několikanásobné binární logistické regrese, kde zkoumáme, zda závislou proměnnou predikují proměnná nezávislá a mediátor, a jakým způsobem se změní vztah X a Y po zapojení mediátoru. V tomto modelu je tato změna stejná jako v případě analýzy v PROCESSu. Vidíme, že regresní koeficient nezávislé proměnné se mírně zmenšil v případě, že do modelu byl zapojen mediátor (tedy jelikož je tato regrese logistická, snižuje se tím pravděpodobnost výskytu pozitivního výsledku u závislé proměnné). Nevýhodou tohoto postupu je, že zde nikde nenajdeme explicitní hodnotu mediačního efektu, ale musíme si ho dopočítat podle postupu v kapitole 2.4.2 (tedy jako rozdíl  $c-c'$ , nebo jako součin parametrů  $ab$ ).

Vidíme, že výsledek je stejný jako při použití předchozího způsobu analýzy. Opět se nám tedy podařilo popsat předpokládaný vztah mezi proměnnými pomocí tohoto modelu (interpretace modelu se nezměnila po změně závislé proměnné na binární).

### ***4.3 Zhodnocení výzkumných otázek a hypotéz***

První výzkumnou otázkou bylo, jakým způsobem se liší výsledky v případě, že je závislá proměnná vstupující do analýzy mediace spojitá, nebo kategoriální (v tomto případě konkrétně dichotomická). Z této části plynula hypotéza, že tyto výsledky se neliší. Před vyhodnocením této hypotézy je potřeba si uvědomit, že v této práci byly použity pouze dvě případové studie, výsledek tedy platí pouze pro ně a v tuto chvíli nedokážeme vyslovit tento závěr jako obecnou pravdu (díky tomu, že v praxi mohou působit i jiné nezávislé proměnné nebo mediátory, které daný vztah ovlivňují). V našem případě nicméně byly oba modely vyhodnoceny stejně, ať jsme již použili jakýkoli z typů proměnné. Podařilo se tedy vyřešit problém nespojitě proměnné a model byl vyhodnocen i přesto (stejně byl určen vztah mezi závislou, nezávislou proměnnou a mediátorem).

Druhým výzkumným záměrem byl samotný popis jednotlivých procesů (regresí) pojících se k analýze mediace, s čímž se pojila výzkumná otázka, jakým způsobem se liší analytický postup v případě, kdy používáme či nepoužíváme doplněk PROCESS sloužící k analýze mediace (s čímž se pojí řada výzkumných podotázek). PROCESS se liší

zejména v tom, že vzhledem k tomu, že tato jeho konkrétní funkce<sup>10</sup> slouží výhradně k analýze mediace, u jednotlivých regresních procesů jsou zobrazeny pouze informace, které slouží k vyhodnocení analýz mediace (zejména hodnoty regresních koeficientů,  $R^2$ , testy významnosti parametrů, globální F-test o modelu a intervaly spolehlivosti). U regresního přístupu lze sice (v případě, že si navolíme) zobrazit mnohem víc parametrů o modelu, ty ale nejsou relevantní pro analýzu mediace.

Co je naopak velikou výhodou PROCESSu je, že dovede sám spočítat mediační efekty, přímý a nepřímý efekt nezávislé proměnné na závislou proměnnou. Sada regresí toto v podstatě umožňuje také, ale efekty je potřeba si dopočítat.

Zjevná nevýhoda PROCESSu je, že nelze nastavit, v jakém pořadí chceme jednotlivé regrese zobrazit. V případě, že chceme vyhodnocovat mediační efekt a posuzovat, zda je celý model vhodný, musíme postupovat ve vyhodnocování jednotlivých regresí s určitou posloupností, nicméně PROCESS zobrazuje jednotlivé procesy abecedně. Může se tedy stát, že vyhodnocujeme model, který vlastně od začátku nemá smysl (protože nezávislá proměnná nepredikuje závislou proměnnou), ale odhalíme to až v průběhu interpretace. V PROCESSu by také bylo pro běžného uživatele velmi složité si upravovat syntax, jelikož je velmi dlouhý (cca 5000 řádků v případě pouze jednoduché analýzy mediace), což také přispívá ke zpomalení výkonnosti celého programu.

#### **4.4 Diskuse s literaturou**

V předchozí kapitole jsou popsány dvě empirické případové studie, na kterých jsme demonstrovali využití analýzy mediace v sociologii. V této části budou tyto výsledky porovnány a diskutovány s dostupnou literaturou. První část se týká limitace těchto modelů, což je pojí se slabými stránkami a limitací celé této práce. Druhá část diskuse se zabývá budoucími doporučeními, jakým způsobem k analýze mediace přistupovat.

##### **4.4.1 Limitace modelů**

Kromě limitů modelů, které jsou založeny na regresi je potřeba také brát v úvahu validitu celého modelu, který obsahuje mediátorovou proměnnou jako takového. První částí diskuse validity je diskuse validity závěrů, které plynou z modelu. Jako příklad uvádí ve

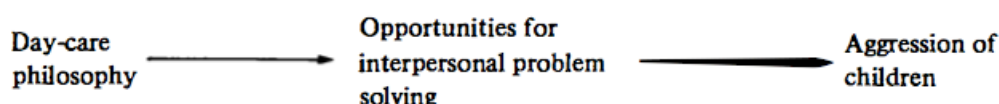
---

<sup>10</sup> Funkce číslo 4



svém díle *Estimating the effects of social interventions* Judd a Kenny. V tomto případě je v obrázku níže popsán mediační model, týkající se stacionářů, které mohou mít své specifické metody léčby, které ovlivňují množství způsobů, díky kterým je jedinec si schopný vyřešit svůj problém, což má dopady na redukci agresivního chování u dětí.

Obrázek č. 6 – Ilustrace vztahů mezi proměnnými



Zdroj: Judd, Kenny. *Estimating the effects of social interventions*. 1981.

Zprvė, stejně jako obecně u sociálních věd se zde setkáme s příliš malou mírou standardizace – model nedokáže vzít v potaz to, že ne na všechny jedince působí nezávislá proměnná stejnou mírou (tedy v tomto případě můžeme říci, že ne každý z jedinců dostává stejnou “dávku” terapie).

Když tedy tuto teorii aplikujeme na námi zkoumaná data, je nám jasné, že i v našem případě může toto být problém. Neznáme jednotlivé případy natolik, abychom mohli vyslovit zcela nevyvratitelný závěr. Judd a Kenny popisují, že jejich nezávislá proměnná nemusí mít stejný vliv na všechny zkoumané jednotky z toho důvodu, že může mít různé podoby a různým způsobem na případy působit.

V našem případě jsme tedy měli nezávislou proměnnou „vzdělání respondenta“, respektive počet let vzdělání respondenta. Odpověď na tuto otázku je pouze číslo, které nám uvádí, kolik let respondent navštěvoval nějakou vzdělávací instituci. Neříká nám ale naprosto nic o kvalitě jeho vzdělání. Může se také jednat o respondenty, kteří se například několikrát neúspěšně pokusili o vysokou školu, ale neuspěli, nicméně jako léta vzdělání to počítají. Dalo by se také předpokládat, že respondenti s určitým druhem vzdělání a konkrétními obory, budou úspěšnější než jiní. Nemůžeme tedy přesně určit, do jaké míry respondenta nezávislá proměnná ovlivňuje, tudíž nikdy tedy nemůžeme s naprostou jistotou říci, že i když jsme užitím příslušných metod prokázali určitý efekt mediace, tak je namodelovaný vztah bez výhrad správný.

Další oblast, na kterou se musíme zaměřit je, za jakých podmínek vůbec mediační efekt určujeme. Pokud bychom chtěli tvrdit, že zde k mediaci dochází, musíme vyslovit tři následující závěry: 1) Nezávislá proměnná způsobuje závisle proměnnou – bez toho by nemělo smysl žádný mediační efekt zkoumat. 2) Nezávislá proměnná způsobuje potenciální mediátor. Zároveň pokud o proměnné řekneme, že je mediátorem, musí být závislou proměnnou alespoň zčásti určena. 3) Mediátorová proměnná musí ovlivňovat výstup (závislou proměnnou). Pokud tento výstup přímo neovlivňuje, nemůžeme ji považovat za mediátor [Judd, Kenny 1981: 206-208]. To, že bez těchto závěrů nemůžeme prokázat mediaci je v praxi poměrně problematické – jelikož simultánní působení všech třech těchto podmínek se obtížně prokazuje. Zvláště pokud je řetěz mediací velmi dlouhý, může docházet k velmi malé korelaci mezi vstupem a výstupem.

U modelů mediace se také můžeme setkat s problematickou vnitřní validitou při testování mediačního modelu. Nezávislá proměnná, která do modelu vstupuje, je vždy regulovatelná, kdežto u mediátoru tomu často tak není. Můžeme tedy sice předpokládat vliv manipulovatelné nezávislé proměnné na výsledek (závisle proměnnou), ale nemůžeme se závěry počítat zcela stejně, pokud zapojíme mediátor jakožto nemanipulovatelnou proměnnou. Může zde docházet k chybě měření, která ovlivňuje validitu celého modelu [Judd, Kenny 1981: 208].

Pokud se vrátíme obecně k metodě analýzy sekundárních dat, ze kterých jsou vytvářeny modely, narazíme na další limity a problémy, které se této práci mohou týkat. V tomto případě jsme analyzovali data, která měla původně zcela jiný účel, proč byla sesbírána, než je tato studie. Je možné, že na základě našeho úsudku nebyl vztah mezi proměnnými (zejména těmi nekorelujícími) určen zcela správně, ale vstupuje tam nějaký jiný fenomén, který do analýzy zahrnut nebyl. Neznáme také způsob, jakým byly zkoumané proměnné operacionalizovány (nevíme tedy, zda jejich význam nebyl nějak pozměněn), případně zda nedošlo k chybě při výběru použitého základního souboru.

#### **4.4.2 Doporučení pro další zkoumání**

V této práci dochází pouze k sekundární analýze dat, což se pojí s výše zmíněnými limitacemi. V případě, že bychom se chtěli analýze mediace věnovat více, lepším přístupem je pořízení nových dat s tím, že budou zcela primárně sloužit tomuto účelu. Ve většině výzkumů je velmi důležité, jakým způsobem je nastaven research design často

více, než užití statistické metody [MacKinnon 2008: 397]. V případě, že víme dopředu, že budeme chtít pro naši studii použít analýzu mediace, je dopředu potřeba zkoumaný vztah předpokládat a na základě toho sestavit naši metodologii. Toto můžeme předpokládat na základě metod popsaných v kapitole 2.4. Zejména je potřeba se dostatečně zaměřit na teorii, na základě které je potřeba potenciální vztahy předpokládat. I přesto, že v případě našich modelů byl tento parametr splněn, je potřeba také zmínit, že je vždy potřeba dbát na dostatečnou velikost vzorku, aby naše závěry byly zobecnitelné. Obzvláště v případě, kdy užíváme metodu asymetrického intervalu spolehlivosti, je potřeba aby vzorek byl alespoň  $n=500$ , aby model dokázal zachytit i drobné efekty.

V případě, že provádíme libovolnou studii, vždy je potřeba dbát na její reliabilitu. Mezi metody, které se používají k posouzení reliability patří například metoda štěpení (srovnání výsledků, kterých bylo u týchž respondentů dosaženo pomocí dvou různých, ale ekvivalentních forem dotazníků), což ale nebývá v reálných šetření příliš častá metoda z důvodu časové a ekonomické náročnosti. V případě kontinuálního výzkumu lze reliabilitu posuzovat pomocí míry shody mezi prvním a opakovaným měřením (v takové situaci je nutné určit optimální dobu odstupů, aby nebyla příliš krátká a respondenti si vše ještě nepamatovali, nebo příliš dlouhá, kde může dojít ke změně měřené vlastnosti). Další možností je zadání stejného dotazníku dvěma reprezentativním výběrům téhož základního souboru a poté opět vyhodnocení míry shody [Chrásková 2016: 166].

Další možností, jak tuto analýzu rozšířit by pak bylo zahrnout do studie více mediátorů (tedy vytvořit mnohonásobný model mediace). Jelikož v praxi není tak časté, aby byla závislá proměnná jenom jedním fenoménem, logickým dalším krokem by bylo prozkoumání jiného mediátoru, případně pak vytvoření jednoho souhrnného modelu, kde by byl popsán vztah všech proměnných najednou. Postup v tomto případě většinou je, že je nejprve vytvořen model jednoduché mediace (jako v našem modelu) a ostatní faktory ovlivňující vztahy v modelu (ostatní proměnné) až později po ověření tohoto vztahu [MacKinnon 2008: 400].

Dalším logickým krokem by také bylo namísto stávajícího jednoduchého modelu vytvořit model longitudinální, tedy sledovat námi určené trendy v námi určeném časovém úseku – což by v našem případě nebyl problém, jelikož se jedná o každoročně opakované studie.

Tímto bychom mohli ukázat, že výsledky, které nám vyšly nejsou jenom specifika daného časového období, ale i obecně platné.

## 5 Závěr

Předložená diplomová práce se zabývá analýzou mediace a příklady její aplikace v sociologii. Jako jeden z cílů si práce kladla popsat tuto metodu, představit její využití pro sociologii (kde doposud příliš často využívána není) a představit její aplikaci na konkrétních datech sociologického výzkumu. Tento cíl byl stanoven zejména z důvodu snahy o rozšíření tuzemské literatury o tuto problematiku, jelikož se jí doposud příliš nevěnuje. Dalším cílem bylo porovnat přístupy k analýze mediace v případě spojité či dichotomické závislé proměnné. Tyto cíle práce naplnila, jelikož došlo k představení a porovnání těchto metod, a to jak na teoretické, tak praktické rovině.

V teoretické části práce je popsán úvod do problematiky mediace a stručně charakterizovány přístupy k ní. Vymezeny jsou zejména různé přístupy k problematice v závislosti na typu proměnných (tyto kapitoly jsou rozděleny podle toho, zda se jedná o spojitou či dichotomickou závislou proměnnou). Z důvodu zasazení do teoretického rámce jsou v této části práce popsány jednoduché regresní modely, na kterých je analýza mediace založena, dále pak konkrétní aplikace v různých předchozích výzkumech.

V metodologické a analytické části je poté popsána aplikace na sociologických datech ze dvou studií: GSS (General Social Survey) a ISSP (International Social Survey Programme). V rámci analýzy se jednalo o jednoduchou mediaci, kde se model skládal z jedné závislé proměnné, jedné nezávislé proměnné a jednoho mediátoru. V obou těchto případových studiích byla jako nezávislá proměnná určena „počet let vzdělání respondenta“. V prvním případě byla porovnávána se závislou proměnnou „socioekonomický status respondentova partnera“ a jako mediátor byl určen „socioekonomický status respondenta“. V druhém případě byla porovnávána se závislou proměnnou „počet podřízených“ a jako mediátor byl určen „věk“. Oba tyto příklady byly aplikovány nejprve v původní podobě svých dat (se spojitou závislou proměnnou) užitím programu SPSS a jeho doplňku PROCESS. V obou těchto případech se užitím sady příslušných analýz podařilo prokázat mediační efekt, který je zde také popsán a interpretován na základě výstupů z programu.

Následně byly v obou případech závislé proměnné binarizovány a tyto dva přístupy porovnány (tento způsob analýzy musel být proveden pomocí sady lineárních a logistických regresí podle Baron & Kenny metody, jelikož PROCESS neumožňuje práci s binární závislou proměnnou). Nejprve byly porovnány výsledky a vliv na model jako takový. Na základě těchto dvou případových studií jsme vyslovili závěry, že v případě, když změním typ závislé proměnné, výstupy z analýzy jsou velmi podobné, co se týká koeficientů a mediačních efektů, celkový model, jakým je popsán vztah mezi proměnnými zůstává zcela stejný (tento závěr byl vysloven pouze pro tyto případové studie, jak je uvedeno v diskusi, v praxi by mohlo model ovlivňovat spousta dalších faktorů a výsledek by tak mohl být jiný). Limitace tohoto zjištění také spočívá v tom, že jsme v rámci této práce nebrali v úvahu další faktory, které by mohly na tento vztah působit (např. zařazením více nezávislých proměnných či mediátorů).

Dále práce odpovídala na otázky, jak se liší provedení analýzy dat v případě, že používáme nebo nepoužíváme PROCESS. Oba z těchto přístupů nám umožňují určit, zda je námi nadefinovaný model správný. Doplněk PROCESS má výhodu v tom, že vše probíhá v rámci jednoho nadefinovaného příkazu (tzn. je možné zadat jeden příkaz, který vypočítá celou sadu regresí). Jelikož je tento doplněk specializován právě na analýzu mediace, výstupy jsou omezeny zejména na informace, které nám pomohou tento vztah určit a odhalit. Ve výstupech nenajdeme příliš informací navíc o jednotlivých modelech, které nám jsou schopny poskytnout regresní analýzy jako takové (díky čemuž jsou výstupy velmi přehledné). PROCESS také dovede sám vypočítat mediační efekty, které tak není nutné dohledávat a dopočítávat z regresních koeficientů. Nevýhodou PROCESSu pak je, že nezobrazuje spousta informací o jednotlivých regresích, které bychom zjistili v případě, že bychom sadu regresí prováděli ručně. Nevýhodou také je, že je velmi složité zkontrolovat syntax díky jeho délce (cca 5000 řádků), jeho délka navíc výrazně ovlivňuje a zpomaluje výkonost programu.

Dle mého názoru je analýza mediace velmi zajímavou a přínosnou metodou, která by byla v sociologických výzkumech velmi užitečná. Pomocí ní bychom mohli odhalit spousta kauzálních vztahů, obohatit tak vztahy mezi závislými a nezávislými proměnnými, což by umožnilo vytváření nových teorií.

## 6 Summary

This diploma thesis deals with mediation analysis and examples of its application in sociology. One of the aims of this thesis was to describe this method, to introduce its use for sociology (where it is not used too often) and to present its application on specific sociological research data. This objective was determined mainly by the effort to extend the Czech literature on this issue, as it is not being pursued much. Another objective was to compare approaches to mediation analysis in the case of a continuous or dichotomous dependent variable. These goals were fulfilled, as these methods were presented and compared, both on theoretical and practical level.

In the theoretical part of the thesis an introduction to the problems of mediation is described and the approaches to it are briefly characterized. In particular, different approaches to the problem are defined depending on the type of variables (these chapters are divided according to whether it is a continuous or dichotomous dependent variable). As part of the theoretical framework, simple regression models on which mediation analysis is based, as well as specific applications in various previous studies, are described in this part of the thesis.

In the methodological and analytical section, application to sociological data from two studies: GSS (General Social Survey) and ISSP (International Social Survey Program) is described. In the analysis, it was a simple mediation where the model consisted of one dependent variable, one independent variable and one mediator. In both case studies, the "number of years of the respondent's education" was determined as an independent variable. In the first case, it was compared with the dependent variable "socio-economic status of the respondent's partner", and the "socio-economic status of the respondent" was determined as the mediator. In the second case, it was compared with the dependent variable "number of subordinates" and "age" was determined as the mediator. Both of these examples were first applied in the original form of their data (with a continuous dependent variable) using SPSS software and its PROCESS add-on. In both cases, using a set of relevant analyzes, it was able to prove the mediation effect, which is also described and interpreted based on program outputs.

Subsequently, in both cases the dependent variables were binarized and these two approaches were compared (this method of analysis had to be done using the Baron & Kenny method of linear and logistic regression as PROCESS does not allow working with a binary dependent variable). First, the results and the impact on the model as such were compared. Based on these two case studies, we concluded that if we change the type of dependent variable, the outputs from the analysis are very similar in terms of coefficients and mediation effects; the overall model, that describes the relationship between the variables, remains the same (this conclusion was only spoken for these case studies, as stated in the discussion, in practice the model could be influenced by a number of other factors and the result could be different). The limitation of this finding also lies in the fact that we did not take into account other factors that might influence this relationship (eg by including more independent variables or mediators).

In addition, the thesis responded to questions how the data analysis differs if we use or do not use PROCESS. Both of these approaches allow us to determine whether the model defined by us is right. The PROCESS add-on has the advantage that everything runs under one defined command (i.e., it is possible to enter one command that calculates the entire set of regressions). Since this add-on is specialized in mediation analysis, outputs are limited especially to information that will help us to determine this relationship and to reveal it. In the outputs, we do not find too much information about the individual models that regression analyzes as such are able to provide (which makes the outputs very clear). PROCESS can also calculate mediation effects by itself, so they do not have to be traced and calculated from regression coefficients. The disadvantage of PROCESS is that it does not show a lot of information about individual regressions that we would find out if we would manually set the regression set. The drawback is also that it is very difficult to check the syntax due to its length (about 5000 lines), its length also greatly affects and slows the performance of the program.

In my opinion, mediation analysis is a very interesting and beneficial method that would be very useful in sociological research. With it, we could discover many causal relationships, thus enriching the relationships between dependent and independent variables, which would allow the creation of new theories.

## 7 Zdroje

### Seznam literatury

Judd, Charles M., Kenny, David A. *Estimating the effects of social interventions*. 1981. Cambridge: Cambridge University Press

MacKinnon, David P. *Introduction to Statistical Mediation Analysis*. 2008. New York: Lawrence Erlbaum Associates

Harrell, Frank E. Jr. *Regression Modeling Strategies*. 2001. New York: Springer

Hosmer, David W., Lemeshow, Stanley. *Applied Logistic Regression*. 2000. Hoboken: John Wiley & Sons, Inc.

Vartanian, Thomas P. *Secondary data analysis*. 2011. New York: Oxford university press

Babbie, Earl. *The basics of social research*. 2014. Wadsworth: Cengage learning

Chráška, Miroslav. *Metody pedagogického výzkumu: Základy kvantitativního výzkumu*. 2016. Praha: Grada Publishing

Montgomery, Douglas C., Peck, Elizabeth A., Vining, Geoffrey G. *Introduction to Linear Regression Analysis*. 2012. Hoboken: John Wiley & Sons, Inc.

Marcus, Bess H., Forsyth LeighAnn H. *Psychologie aktivního způsobu života*. 2010. Praha: Portál

### Články

Hendl, Jan. Analýza působení mediátorových a moderátorových proměnných. Informační bulletin České statistické společnosti. Vol. 21, No. 3 (March 2010), pp. 1-7. Dostupné z: <<http://www.statspol.cz/bulletiny/ib-2010-1-web.pdf>>



Baron, Reuben M., Kenny, David A. The Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations. *Journal of Personality and Social Psychology*. Vol. 51, No. 6 (1986), pp. 1173-1182. Dostupné z: <<http://www.sesp.org/files/The%20Moderator-Baron.pdf>>

Řeháková, Blanka. Nebojte se logistické regrese. *Sociologický časopis*. Vol. 36, No. 4 (2000), pp. 475-492. Dostupné z: <[http://sreview.soc.cas.cz/uploads/5f6961faa17dd98a67cfb71a5205469d297369f5\\_372\\_475REHAK.pdf](http://sreview.soc.cas.cz/uploads/5f6961faa17dd98a67cfb71a5205469d297369f5_372_475REHAK.pdf)>

Menard, Scott. Standards for Standardized Logistic Regression Coefficients. *Social Forces*. Vol. 89, No. 4 (June 2011), pp. 1409-1428. Dostupné z: <[http://www.jstor.org/stable/pdf/41290135.pdf?seq=1#page\\_scan\\_tab\\_contents](http://www.jstor.org/stable/pdf/41290135.pdf?seq=1#page_scan_tab_contents)>

Tingley, Dustin et al. Mediation: R Package for Casual Mediation Analysis. *Journal of Statistical Software*. Vol. 59, No. 5 (August 2014), pp. 1-38. Dostupné z: <<https://www.jstatsoft.org/index.php/jss/article/view/v059i05/v059i05.pdf>>

Preacher, Kristopher J. Effect Size Measures for Mediation Models: Quantitative Strategies for Communicating Indirect Effects. *Psychological methods*. Vol. 16, No. 2 (2011), pp. 93-115. Dostupné z: <[http://quantpsy.org/pubs/preacher\\_kelley\\_2011.pdf](http://quantpsy.org/pubs/preacher_kelley_2011.pdf)>

Preacher, Kristopher J., Hayes, Andrew H. SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments & Computers*. Vol. 36, No. 4 (November 2004), pp. 717-731. Dostupné z: <<http://link.springer.com/article/10.3758/BF03206553>>

Imai, Kosuke, Keele, Luke, Yamamoto, Teppei. Identification, Inference and Sensitivity Analysis for Causal Mediation Effects. *Statistical Science*. Vol. 25, No. 1 (February 2010), pp. 51-71. Dostupné z: <[http://www.jstor.org/stable/pdf/41058997.pdf?seq=1#page\\_scan\\_tab\\_contents](http://www.jstor.org/stable/pdf/41058997.pdf?seq=1#page_scan_tab_contents)>

Rubin, Donald B. Inference and missing data. *Biometrika*. Vol. 63, No. 3 (1976) pp. 581-592. Dostupné z: <<http://www.jstor.org/stable/2335739>>

Eves, Howard W. A Very Brief History of Statistics. *The College Mathematics Journal*. Vol. 33, No. 4 (September 2002), pp. 306-308. Dostupné z: <<http://www.jstor.org/stable/1559052>>

Green, Donald P., Ha, Shang E., Bullock, John G. Enough Already about "Black Box" Experiments: Studying Mediation Is More Difficult than Most Scholars Suppose. *The Annals of the American Academy of Political and Social Science*. Vol. 628, Field Experiments in Comparative Politics and Policy (March 2010), pp. 200-208. Dostupné z: <<http://www.jstor.org/stable/40607522>>

Yuan, Ke-Hai, Bentler, Peter M. Mean and Covariance Structure Analysis: Theoretical and Practical Improvements. *Journal of the American Statistical Association*. Vol. 92, No. 438 (June 1997), pp. 767-774. Dostupné z: <<http://www.jstor.org/stable/2965725>>

Chesher, Andrew. The Effect of Measurement Error. *Biometrika*. Vol. 78, No. 3 (September 1991), pp. 451-462. Dostupné z: <<http://www.jstor.org/stable/2337015>>

Bradburn, Norman M., Sudman, Seymour, Blair, Ed, Stocking, Carol. Question Threat and Response Bias. *The Public Opinion Quarterly*. Vol. 42, No. 2 (1978), pp. 221-234. Dostupné z: <<http://www.jstor.org/stable/2748391>>

Psotta, Rudolf, Kokštejn, Jakub, Frýbort, Pavel. Relationships Between Levels of Motor Coordination, Attention and Physical Activity in Children: The Mediation Model. *Acta Univ. Palacki*. Vol. 42, No. 4 (November 2012), pp. 29-40. Dostupné z: <[https://www.researchgate.net/publication/277311574\\_Relationships\\_between\\_levels\\_of\\_motor\\_coordination\\_attention\\_and\\_physical\\_activity\\_in\\_children\\_The\\_mediation\\_model](https://www.researchgate.net/publication/277311574_Relationships_between_levels_of_motor_coordination_attention_and_physical_activity_in_children_The_mediation_model)>

## Internetové zdroje

Logistická regrese [online]. cit. 10.1.2017. Dostupné z:  
<<http://www.trilobyte.cz/downloadfree/qcemanual/logreg.pdf>>

Regresní a korelační analýza [online]. cit. 8.1.2017. Dostupné z:  
<<https://homen.vsb.cz/~oti73/cdpast1/KAP09/KAP09.HTM>>

Bolboaca, Sorana-Daniela, Jantschi, Lorentz. Pearson versus Spearman, Kendall's Tau Correlation Analysis on Structure-Activity Relationships of Biologic Active Compounds [online]. cit. 11.2.2016. Dostupné z:  
<[http://ljs.academicdirect.org/A09/179\\_200.htm](http://ljs.academicdirect.org/A09/179_200.htm)>

Lineární regrese [on-line]. cit. 10.2.2017. Dostupné z:  
<<http://mdg.vsb.cz/wiki/public/Excel4.pdf>>

Neubauer, Jiří. Testování hypotéz o parametrech regresního modelu [on-line]. Cit. 8.2.2017. Dostupné z: <<http://k101.unob.cz/~neubauer/pdf/regrese3.pdf>>

Path Analysis [on-line]. cit. 15.3.2017. Dostupné z:  
<<http://www.statisticssolutions.com/factor-analysis-sem-path-analysis/>>

Allison, Paul. Why I don't Trust the Hosmer-Lemeshow Test for Logistic Regression. cit. 10.3.2017. Dostupné z: <<https://statisticalhorizons.com/hosmer-lemeshow>>

Mielcová, Elena, Stoklasová, Radmila, Ramík, Jaroslav. Lineární regrese [on-line]. Cit. 12.4.2017. Dostupné z: <<http://mdg.vsb.cz/wiki/public/Excel4.pdf>>

## Zdroje dat

General Social Survey. Dostupné z: <<http://gss.norc.org/>>

International Social Survey Programme. Dostupné z:  
<<http://www.gesis.org/issp/home/>>