

## Výmluvy ve formě hypotéz

H 1:

Sociologie je ještě mladá věda, neměla dosud čas vyvinout své vlastní paradigma.

ALE:

Kybernetice není ještě ani padesát let a je nesporně "normální" vědou. Moderní sociologie byla pokřtěna Comtem v roce 1839. Nicméně to, co bychom dnes označili za sociologii, nalezneme již v Platónovi a jinde.

H 2:

Sociální jevy jsou rychle proměnné a proto je velmi obtížné je analyzovat.

Naši kolegové v chemii a fyzice byli schopni definovat všechny důležité vlastnosti těžkých prvků, i když poločas života těchto prvků je jen nepatrný zlomek vteřiny.

H 3:

Mnohé sociální jevy nejsou dostupné přímému pozorování.

To je pravda, ale i kolegové v přírodních vědách jsou odkázáni na nepřímé pozorování. Nikdo nikdy neviděl atom, jenom jeho reprezentaci.

H 4:

Chytří studenti pochopí, že studium sociálních věd není cestou k finančně úspěšné kariéře. Jenom ti méně chytří volí náš obor.

Tuto hypotézu necháme raději bez komentáře.

H 5:

Sociální jevy jsou vzájemně propojeny. Často se zdá, že všechno souvisí se vším ostatním.

Tahle hypotéza stojí za vážné zamyšlení.

Ze všech hypotéz si jen ta poslední zaslouží podrobnou analýzu.

### 1.3. Co nezkoumat

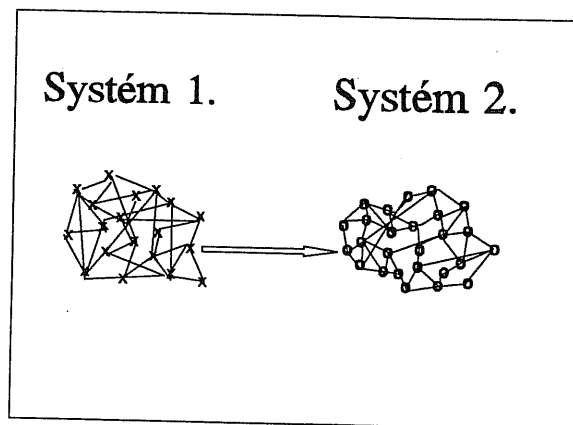
Ashby (1965) se kdysi zamyslel nad elementárními kroky v empirickém výzkumu. Každý vědec přirozeně vybírá pro své pozorování jenom tu část reality, kterou považuje za relevantní. Nicméně výběr této části reality nemůže být arbitrární. Musí respektovat to, co Ashby označuje za "přirozený systém."

Přirozený systém je definován jako soubor proměnných, které jsou navzájem spojeny mnoha vztahy. Naproti tomu počet vztahů spojujících jeden přirozený systém s jiným přirozeným systémem je veličinou nižšího řádu.

Můžeme si to představit třeba tak, jak to ukazuje náš graf: jako skupinu mnoha křížků, představujících jednoho přirozeného systému. Jednotlivé křížky jsou spojeny mnoha vazbami. Můžeme si třeba představit, že každý křížek je spojen se všemi zbývajícími křížky. Druhý přirozený systém je představován souborem mnoha malých kroužků, které jsou opět vzájemně spojeny mnoha a mnoha spoji. Ale existuje jenom jediný spoj mezi oběma systémy.

Graf 1.1.

"Přirozené" systémy



Proč je koncept "přirozeného systému" tak důležitý? Protože není možné vysvětlit chování systému, který není **informačně uzavřený**. Informačně uzavřený je takový systém, který nemůže být ovlivněn ničím zvenku **bez vědomí výzkumníka**. Interpretace chování informačně otevřeného systému může vést k vážnému zkreslení:

#### Pohádka pro odrostlejší děti 1.

#### O zlém profesorovi a informačně otevřeném systému.

Byl jednou jeden profesor, ne moc dobrý profesor a ne moc dobrý člověk. Zvláště zlý byl na svého asistenta. a tak se jednoho dne ten asistent vzbouřil. Začal tajně přidávat do profesorových zkumavek jakousi substanci. Zlý profesor dostával náhle neočekávané výsledky, které jeho věda nemohla vysvětlit. Byl z toho opravdu zoufalý. Opustil univerzitu a začal v Kotcích prodávat květiny. Tak se informačně neuzavřený systém zasloužil o proměnu špatného profesora v užitečnou lidskou bytost.

A právě zde je místo pro koncept přirozeného systému. Když výzkumník správně vybere úplný přirozený systém, musí kontrolovat jenom několik málo vstupů, přicházejících z jiných přirozených systémů.

Náš příklad s křížky a kolečky nám pomůže ilustrovat situaci, kdy výzkumník vybral neúplný přirozený systém, třeba proto, že není známo, co patří do systému, nebo protože systém obsahuje proměnné, které nejsme s to pozorovat. Řekněme, že vybral jenom polovinu křížků. Nyní musí kontrolovat nejen těch několik málo linek spojujících křížky s kolečky, ale i mnoho a mnoho kanálů, spojujících vybrané křížky s těmi nevybranými. Jeho úkol je teď mnohem obtížnější.

#### **Cvičení 1.1.**

*Co se stane, když výzkumník vybere více než jeden přirozený systém?  
(Odpověď naleznete na konci této kapitoly.)*

Interpretace neúplně popsaného systému může vést k velice vážnému zkreslení. A zde jsme u klíčového bodu naší diskuse. Existuje nějaký rozdíl mezi přirozenými systémy přírodních a sociálních věd? Představme si jednoduchý fyzikální experiment. Řekněme, že bychom chtěli experimentálně ustavit, jaký je bod varu vody. Které proměnné musíme sledovat, abychom dostali uspokojivě spolehlivé výsledky? Přirozeně musíme měřit teplotu vody. Ze středoškolské fyziky si ještě pamatujeme, že bod varu závisí na tlaku vzduchu a že bychom také měli kontrolovat čistotu vody. Pro jistotu bychom se měli také přesvědčit, zda je teploměr správně kalibrován. Několik málo elementů nám zde stačí k uspokojivě úplnému popisu přirozeného systému.

Vezměme si naproti tomu jednoduchý příklad vztahů v oblasti sociální. Chceme třeba studovat, co může ovlivnit příjem jedince. Jistě: povolání, pozice v zaměstnání, místo práce, délka vzdělání, typ vzdělání, pracovní zkušenost, věk, pohlaví, zdravotní stav, rodinné poměry, členství v určitých organizacích, vzdělání rodičů, povolání rodičů, vzdělání a povolání manžela nebo manželky, vzdělání a povolání dětí nebo jiných členů domácnosti, členství v různých neformálních organizacích, příbuzní v zahraničí, přístup k různým nedostatkovým produktům atd. Mohli bychom pokračovat s tímto výčtem ještě velmi dlouho. Povšimněme si také, že mnoho z bodů, které jsme právě zmínili, představuje ve skutečnosti celý soubor proměnných. Kupř. takové body jako pracovní zkušenost, členství v organizacích,

příslušnost k neformálním skupinám musí být měřeny celou sérií otázek v dotazníku nebo interview.

Již tento první pohled důrazně naznačuje, že přirozené systémy v sociálních vědách jsou mnohem rozsáhlejší než systémy v exaktní vědě. Nezapomeňme, že pokud si umíme představit situaci, ve které "něco" může ovlivnit to, co studujeme, také ono "něco" patří do přirozeného systému. Kdybychom tedy snili o úplném popisu systému faktorů ovlivňujících příjem, museli bychom do našeho výčtu zahrnout i takové věci jako tělesnou váhu, barvu očí a vlasů, vlastnictví určitých předmětů a stovky, pravděpodobně i tisíce podobných proměnných.

#### Cvičení 1.2:

Vyjmenujte několik proměnných, o kterých můžete předpokládat, že nemohou mít žádný vliv na výši příjmu.

Je tedy zřejmé, že počet proměnných v "přirozených systémech" v oblasti zájmu sociálních věd je nesmírný. Již jenom z hlediska počtu proměnných není valná naděje, že bychom mohli popsat přirozený systém perfektně.

Analýza neúplně popsaného systému je spojena s vysokým rizikem zkreslení. Podívejme se nyní na logiku tohoto zkreslení.

#### 1.4. V tom šílenství je systém

Představme si, že systém, který máme zkoumat, sestává jenom ze tří proměnných. Jenom dvě z těchto proměnných (X a Y) jsme zahrnuli do našeho výzkumu, ne však třetí proměnnou (Z). Třeba proto, že o ní nevíme, nebo ji prostě neumíme měřit. Podívejme se na některé typy zkreslení, které tak mohou nastat. Můžeme je - značně zjednodušeně - klasifikovat do následujících kategorií:

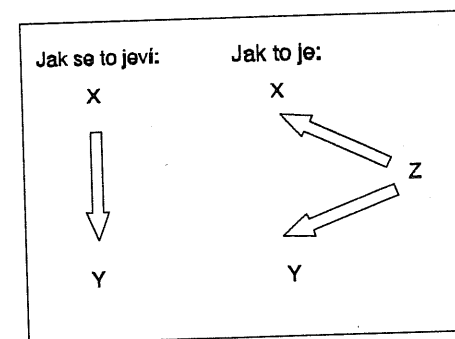
- nepravá korelace
- vývojová sekvence
- chybějící střední člen
- dvojí příčina

#### Nepravá korelace

V řadě evropských regionů bylo zjištěno, že čím více čápů žije v určité krajině, tím vyšší je tam porodnost. Korelační koeficienty byly tak významné, že je velice nepravděpodobné, že zjištěná souvislost je náhodná. Jsme tedy ochotni přijmout hypotézu, že čápi přece jen nosí děti? Asi sotva. Ale pak je naší povinností navrhnout hypotézu, která by uspokojivě vysvětlovala naměřenou souvislost.

Graf 1.2.

#### Nepravá korelace



Toto je klasický příklad nepravé korelace ("spurious correlation"). Zkreslení vzniká tehdy, když třetí nepozorovaná nebo neanalyzovaná proměnná ovlivňuje nějak obě proměnné X a Y, které studujeme.

#### Cvičení 1.3.

Podívejte se pečlivě na graf 1.2., popisující nepravou korelaci. Navrhněte, co může být to tajemné Z.

Jistě nám nehrozí nebezpečí, že bychom přijali hypotézu, že čápi nosí děti. Ale představme si, že nepravá korelace se zdá potvrzovat naši oblíbenou hypotézu. Potom výzkumník musí mít objektivnost anděla a trpělivost nerostného krystalu, aby pracně zabil to, co se po měsíce pokoušel dokázat.

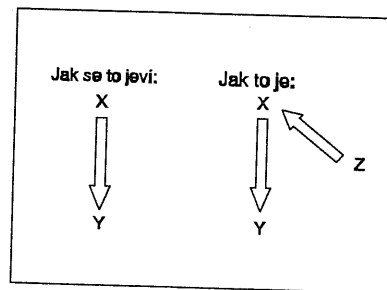
Nepravá korelace je skutečným nebezpečím ve výzkumu. Není to ani tak technický problém analýzy, ale spíše problém lidské kvality výzkumníka.

#### Vývojová sekvence

Tak nazýváme zkrslení, způsobené faktem, že proměnná X, která ovlivňuje Y, je určována předcházející, ale nepozorovanou proměnnou Z.

Graf 1.3.

Vývojová sekvence



Taková situace je skutečně naprosto nevyhnutelná. Každá příčina má totiž jinou příčinu, ta zase jinou příčinu, která má opět svoji příčinu, a tak bychom mohli pokračovat až k aktu stvoření, nebo k tomu, co astronomové nazývají Big Bang. To je problém velmi dobře známý filozofům, kteří ho obvykle nazývají "regresus ad infinitum"

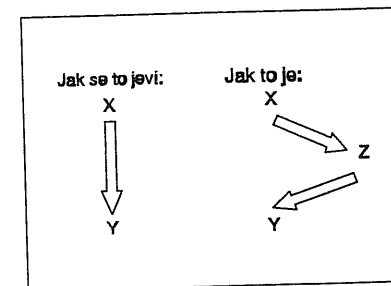
Nicméně, někdy může předčasné přerušlení kauzálního řetězce vést k mylné interpretaci. Některé studie tvrdí, že četba pornografické literatury vyvolává násilné chování mužů k ženám. Nelze však vyloučit, že je zde nějaký předcházející činitel, jako kupř. autoritativní metoda socializace respondenta v dětství, který vyvolal silný zájem jedince o pornografii. Takový omyl je závažný zejména tehdy, když cílem výzkumu je sociální intervence.

#### Chybějící střední člen

Tak je označována situace, kde mezi nezávisle proměnnou X a závislou Y je ještě proměnná Z, kterou jsme nezahrnuli do analýzy. Graf 1.3. tuto situaci jasně popisuje. Je to opět konfigurace, která je téměř všudypřítomná. Kdybychom se jakousi sociologickou lupou podívali, co se děje mezi nějakou příčinou a jejím následkem, existuje ještě řada mezikroků. Často můžeme tyto elementy ignorovat bez rizika zkrslení. Ne však vždycky.

Graf 1.4.

Chybějící střední člen



Řekněme, že X reprezentuje pohlaví respondentů a Y jejich skóre v testu inteligence. Je možné, že výsledky žen, a to zejména žen příslušejících k nižším sociálním třídám, by byly signifikantně nižší, než výsledky mužů.

#### **Cvičení 1.4.**

*Zamyslete se, prosím, nad předchozím odstavcem a navrhněte alternativní hypotézu, ukazující mužským šovinistům, že takové výsledky nepotvrzují superioritu nás, pánů tvorstva.*

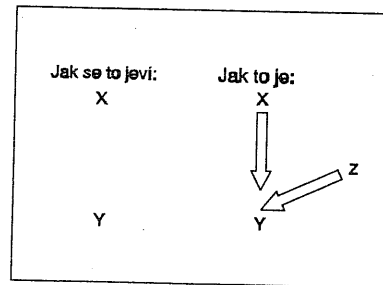
Zkrslení tohoto typu může být nebezpečné. Můžeme je často najít v quasivědeckých pracích, podporujících rasismus, v některých politických pamfletech atd. Mnohé noviny se dopouštějí tohoto hříchu z nevědomosti, když publikují výsledky statistických šetření.

### Dvojitá příčina

Takto můžeme označit situaci, kdy závislá proměnná Y má dvě příčiny, ale jenom jedna z nich, X, byla zahrnuta do výzkumu. Toto je asi nejčastější problém výzkumu v sociálních vědách. Pravděpodobně neexistuje žádný sociální jev, který by měl jedinou příčinu. I v našem nesmírně zmenšeném vesmíru, složeném jenom ze tří proměnných, si můžeme představit, jaké zkrácení může vyvolat, není-li tato další příčina zahrnuta do analýzy.

Graf 1.5.

### Dvojitá příčina



### Cvičení 1.5.

*Představme si třeba, že X je vzdělání jedince a že Y je jeho příjem. Pokud předpokládáme, že se vzděláním příjem poroste, mohli bychom zjistit, že souvislost je velice nízká, nebo dokonce nulová. Co mohlo vyvolat toto zkrácení?*

Teoreticky by bylo možné namítnout, že v některých situacích neměřená proměnná Z může posilovat vliv příčiny X. Ale to je krajně nepravděpodobné. Jak vidíme z grafu 1.4., mezi X a Y není žádný příčinný vztah. I naše cvičení 1.5. je platné jenom uvnitř našeho nerealisticky miniaturizovaného systému tří proměnných. Je jasné, že realisticky by bylo třeba zahrnout další proměnnou "věk", která ovlivňuje vzdělání, a prostřednictvím "zkušenosti", "seniority" i příjem.

Ve skutečnosti "dvojitá příčina" (double causation) má efekt náhodné disturbance, náhodného rušení. Tedy téměř vždy vede k tomu, že naměřená souvislost mezi X a Y je nižší, než skutečná. Tento vzorec bychom mohli právem označit za "nepravou nezávislost". Ta je pak důležitou příčinou stochastického, pravděpodobnostního charakteru poznatků sociálních věd.

### 1.5. Máme toho moc

Představme si, že bychom opravdu snili o vytvoření modelu, který by nám pomohl odhadnout výši jednotlivcovu platu tak přesně, jak je to jen možné. V předchozích odstavcích jsme ukázali, že bychom k tomu potřebovali informaci ne o stovkách, ale pravděpodobně o tisících proměnných. I letmá úvaha nás brzy přesvědčí, že takový projekt nemá naději na úspěch: trvání výzkumného rozhovoru i délka dotazníku jsou objektivně omezeny. Rozhovor delší než jednu hodinu si můžeme dovolit jen ve velmi výjimečných případech, třeba tehdy, pokud je předmět našeho výzkumu zajímavý pro respondenta. Návratnost dotazníku, beztak všude na světě nízká, se vzrůstající délkou dotazníku rychle klesá. Nízká návratnost je pak zdrojem závažného zkrácení, jak o tom budeme hovořit v naší kapitole o populaci a vzorku. Nadto sběr dat je aktivitou hodně nákladnou. Praxe sociologického výzkumu ukazuje, že standardní výzkumy pracují s několika desítkami proměnných a jen u mála skutečně rozsáhlých výzkumů počet proměnných překročí hranici dvou set.

Je tedy zřejmé, že v sociálních vědách **prakticky nikdy nebudeme pracovat s úplně popsáním přirozeným systémem** a tudíž budeme vždy vystaveni zkrácením, která jsme diskutovali v předchozím textu. Tato zkrácení nebudou mít tak jasný charakter; doména výzkumu nesestává jen ze tří proměnných.

Nutnost pracovat s **redukovaným popisem reality** je ještě zdůrazněna následujícím faktem, že redukce informace v sociologickém výzkumu probíhá na několika úrovních.

- redukce počtu pozorovaných proměnných;
- redukce počtu analyzovaných vztahů mezi nimi;
- redukce populace na vzorek a
- redukce časového kontinua na popis jednoho, nebo několika málo časových bodů.

### Redukce počtu pozorovaných proměnných

a její efekt na vznik zkrácení jsme popsali v paragrafu 1.4.

### Redukce počtu analyzovaných vztahů mezi proměnnými

způsobuje zkreslení totožná s redukcí počtu pozorovaných proměnných. Je to logické a docela banální: když nevypočítáme, jak třeba délka pracovní zkušenosti souvisí s příjmem, efekt je stejný, jako bychom informaci o zkušenosti nezískali vůbec. Nicméně ve srovnání s předchozí situací existuje zde zásadní rozdíl: redukce v počtu vztahů je vždycky mnohem rozsáhlejší. Ze všech vztahů, které existují v přirozeném systému budeme schopni studovat jenom zlomek, který je menší, než procento proměnných, které jsme ze všech relevantních proměnných zahrnuli do popisu. Nadto, my zde ještě více redukuje systém, který byl již redukován co do počtu proměnných.

Ilustrujme si to opět na našem příkladu s příjmem, jako závisle proměnnou. Pro úplné pochopení tohoto systému např. nestačí vědět, jaký vliv na příjem má vzdělání. Je třeba také vědět, zda efekt vzdělání je totožný pro jedince s různým zaměstnáním. Je stejný pro muže jako pro ženy? Jak je vliv vzdělání modifikován pracovní zkušeností v oboru, ve kterém respondent nyní pracuje, případně zkušeností v jiných oborech? Jaký je vliv vzdělání u žen, náležejících do věkové skupiny 30-40, s dvěma dětmi ve věku 10 až 14 let, které žijí v obcích s populací 10 až 20 tisíc, které pracují v administrativě v textilním průmyslu, které mají více než deset let praxe v oboru, jejichž manžel má vysokoškolské vzdělání, ale je v invalidním důchodu atd.? Zkrátka, pro úplnou analýzu bychom potřebovali skutečně vyčíslit nejen vztahy mezi všemi proměnnými, ale i mezi jejich všemi možnými kombinacemi. Již tato úloha by vyžadovala vypočítat, analyzovat a interpretovat skutečně astronomické množství koeficientů.

Řekněme, že bychom analyzovali vliv souboru o sto proměnných na příjem. Tento soubor nezávisle proměnných může být kombinován do  $2^{100}$  různých vzorců. Tedy pro úplný popis vlivů těchto proměnných přímo na příjem by byl počet potřebných koeficientů vyjádřen číslem 1 následovaným asi 30ti nulami. To však nezahrnuje analýzu nepřímých vlivů. Z teoretických i praktických důvodů může být pro nás důležité vědět nejen to, jaký je přímý vliv vzdělání na příjem, ale i to, jak vzdělání ovlivňuje zaměstnání, které pak ovlivní příjem.

V kapitole 5. uvidíme, že situace je ještě mnohem komplikovanější, že z hlediska statistiky mají proměnné různý charakter. Některé, *nominální*, jsou představovány jenom jmeny kategorií (kupř. kraj, ve kterém respondent žije). O jiných, *ordinálních*, jsme schopni říci, že určitá kategorie je vyšší (lepší, silnější atd.) než jiná, ale nejsme schopni říci, kolikrát je lepší (kupř. zlatá a stříbrná medaile ve sportovní soutěži). Konečně existují *intervalové* proměnné, které mají skutečně kvantitativní charakter. U nich je možné říci, kolikrát je určitá kategorie vyšší. (Je zřejmé, že příjem čtyři tisíce je dvakrát vyšší, než příjem dva tisíce.) Pro každý z těchto typů proměnných musíme použít *jiný soubor statistických operací*. V téže kapitole uvidíme, že souvislost mezi proměnnými může mít různý tvar, od přímky ke křivkám různých stupňů. Každý typ souvislosti je sledován jiným typem operací.

Je tedy zřejmé, že úplná analýza vztahů mezi proměnnými by přesáhla kapacitu největších superpočítačů a a z praktických, ekonomických a časových důvodů je neuskutečnitelná. Představa, že by takový objem informace mohl být smysluplně interpretován, je prostě absurdní. **Nutnost pracovat s redukováným počtem analyzovaných vztahů není dána současným stavem vědy a technologie, ale dá se říci, že má objektivní, absolutní charakter.**

### Redukce populace na vzorek

Diskuse tohoto typu redukce bude mnohem optimističtější, než předchozí výklad. Problémy jsou zde spíše technického a zejména ekonomického, než epistemologického charakteru. Teorie i praxe konstrukce vzorku je dobře propracována. Budeme se jí zabývat v kapitole 4. Zde jen poznamenejme, že za určitých okolností jsme schopni definovat velikost chyby v měření, které se dopustíme, když pozorujeme vzorek namísto celé populace.

### Redukce časového kontinua na jeden časový bod

Naprostá většina sociálních jevů se mění s průběhem času. Některé jevy zvolna, jiné velice rychle. Naproti tomu ve výzkumné praxi získáváme nejčastěji jenom obraz jediného časového bodu. Filmový záznam nahrazujeme statickou fotografií. To nemusí být v mnoha případech na závadu, pokud neusilujeme o **kauzální vysvětlení**, o vysvětlení **příčinné souvislosti**. Na tomto místě se nebudeme zabývat zajímavými epistemologickými a teoretickými aspekty

kauzální analýzy. Spíše se podíváme jak poznat, zda změny v proměnné X jsou opravdu příčinou změn v proměnné Y.

Abychom mohli prohlásit, že vztah mezi dvěma proměnnými má kauzální charakter, musí být splněny všechny následující podmínky:

- (1) Musí existovat souběžné změny v obou proměnných,
- (2) musíme vyloučit existenci nějaké další, vnější příčiny a
- (3) změny v obou proměnných se musí objevit v logickém časovém pořadí.

S **prvním** bodem nebudeme mít velké problémy. Souběžné změny mohou být relativně snadno měřeny různými statistickými koeficienty souvislosti. O některých z nich budeme hovořit v naší statistické kapitole 8. Ovšem zjištění, že existuje významná souvislost mezi proměnnými, samo o sobě nestačí. Jinak bychom mohli v našem příkladu v odstavci 1.4. akceptovat hypotézu, že čápi nosí děti.

Respektovat **druhý bod** je v sociálních vědách, s výjimkou experimentu, skutečně nemožné. My už vlastně víme, proč tomu tak je. Ukázali jsme, že jsme schopni popsat a analyzovat "přirozený systém" jen velmi neúplně. Nelze tedy naprosto vyloučit, že studované proměnné jsou vystaveny vlivu nějaké vnější proměnné, a my o tomto vlivu nevíme. Jedině skutečný experiment - pokud je vůbec uskutečnitelný - se může s touto situací vypořádat. Jak - o tom bude řeč v příští kapitole.

**Třetí podmínka** znamená, že změny v nezávisle proměnné, tedy v té proměnné, kterou považujeme za příčinu, musí nastat dříve než v druhé, závisle proměnné. Zjistit tento fakt v sociálních vědách může být obtížné a velmi často i nemožné. S výjimkou experimentálního přístupu (a s výjimkou některých případů přímého pozorování) je časové kontinuum redukováno na jediný časový bod. Kupř. odpovědi zaznamenané v dotazníku nezrcadlí nic víc než, řekněme, názory respondenta minulý čtvrtek odpoledne. Nevíme, co se stalo dříve a co později. Můžeme třeba zjistit, že lidé, kteří čtou určitý časopis, obhajují určitý názor mnohem častěji, než ostatní respondenti. Nejsme však schopni říci, zda čtení časopisu přispělo k vytvoření tohoto názoru, nebo zda lidé čtou časopis proto, že odpovídá jejich názorům. V neexperimentální situaci je většinou nemožné zjistit časovou posloupnost přímo ze

získaných dat. Výjimkou jsou takové situace, kdy víme, že určitá proměnná prostě **musí** být příčinou. Když zjistíme, že ženy používají rtěnku mnohem častěji než muži, nebudeme pochybovat o směru kauzality. Prostě víme, že i kdybychom začali užívat rtěnku denně, nezmění nás to v krásné ženy. Ale tato naše znalost je importována z vnějšku, nepochází z výzkumných dat samotných.

Někdy, ale poměrně zřídka, můžeme tento problém obejít tím, že výzkum na stejné populaci opakujeme po určitém časovém odstupu. To je nejen nákladné, ale také můžeme tento postup aplikovat jenom tehdy, když můžeme s dostatečnou jistotou očekávat, že populace bude vystavena vlivu nezávisle proměnné v období mezi prvním a druhým výzkumem. Někdy to je snadné, když kupř. studujeme vliv zahájení vysokoškolského studia na názory mladých lidí. Studujeme-li však, jaký vliv na postoje má třeba náhodný dramatický zážitek, pak je tento postup neaplikovatelný.

Dalo by se namítnout, že se můžeme zeptat respondenta, co se stalo dříve a co později. Ale otázky o minulosti jsou notoricky nespolehlivé. Pokud se nejedná o něco, co bylo pro zkoumanou osobu subjektivně skutečně důležité, respondent prostě nemusí znát tuto informaci. Velice často nám sice odpoví, ale tato odpověď bude bohužel neplatná.

Rozhodně špatnou strategií by bylo přesunout na respondenta odpověď na otázku, co bylo příčinou a co následkem. To je zejména důležité v oblasti motivace. Respondent prostě nezná nebo nedovede vyjádřit skutečné důvody svého určitého rozhodnutí. Profesor Hyhlík ve svých přednáškách z psychologie zdůrazňoval, že otázka "proč" do interview nebo do dotazníku prostě nepatří. Odpověď na toto "proč" musí hledat výzkumník a ne respondent.

Je tedy zřejmé, že redukce časového kontinua a redukce počtu sledovaných proměnných a vztahů mezi nimi doslova **vylučuje kauzální analýzu dat v neexperimentální situaci**. Jak jsou tato omezení překonávána v experimentu a jaké jsou meze aplikace experimentu v sociálních vědách, o tom bude řeč v následující kapitole.

## Řešení úkolů z kapitoly 1.

### Cvičení 1.1.

Nestane se příliš mnoho. Při analýze se brzy zjistí, že skupina proměnných nesouvisí s ostatními. Výzkumník ztratil jenom nějaký čas sbíráním dat, která skutečně nepotřeboval. Nebezpečí chybné interpretace zde opravdu nehrozí. Mimoto brzy uvidíme, že nebezpečí tohoto typu omylu je v sociologii nulové.

### Cvičení 1.2.

To je velice těžká otázka. Sám jsem s jejím řešením nikdy příliš neuspěl. Dlouho jsem v přednáškách používal příklad respondentovy krevní skupiny, až mi někdo správně namítl, že v některých zemích, kde se platí za odběr krve, dárci s řídce se vyskytující krevní skupinou dostávají vyšší odměnu než ostatní. Jiný můj oblíbený příklad, číslo bot, byl rovněž právem napaden. Velikost chodidla může být asociována se silou lýtkových svalů a tento fakt může ovlivnit výkon profesionálního atleta, a tedy i jeho příjem. Pak nám zbývají takové příklady, jako den v týdnu, na který se respondent narodil (pokud ovšem nevěříme na nějaký zvláštní druh astrologie). Ve střední Evropě by asi nemělo vliv číslo domu. (Naproti tomu v severoamerických městech tento předpoklad neplatí. Městské bloky mají velice často shodnou velikost a rohové parcely jsou téměř vždy největší a mohou tak produkovat větší příjem z nájmu.)

### Cvičení 1.3.

Jistě, máte pravdu. Čím více průmyslu, tím více exhalací a tím horší životní podmínky pro čápy. Zároveň je známo, že průmyslové oblasti mají často mnohem nižší porodnost, než oblasti zemědělské.

### Cvičení 1.4.

To bylo lehké, že? Většina testů byla vyvinuta mužskými profesory, náležejícími alespoň do vyšší střední třídy. Jejich socializace, která je ve většině společností odlišná od socializace žen, se odráží v obsahu testů. Tím jsou úkoly testů pro ženy, a zejména pro ženy z nižších tříd, těžší.

### Cvičení 1.5.

Téměř ve všech průmyslových zemích mají starší generace nižší vzdělání, než ty mladší. Zároveň ale, s přibývajícím věkem, jedinec získává větší zkušenost, senioritu, a jeho plat roste. Tento faktor může naměřený efekt vzdělání snížit, nebo dokonce anulovat.

## Jak studovat armádu v laboratoři

experiment  
srovnávání statických skupin  
předběžné a následné sledování jedné skupiny  
klasický experiment  
následné pozorování na 2 skupinách  
Solomonův experiment  
typy zkreslení  
historie

zrání  
prostředí  
instrumentace  
regrese k průměru  
úmrtnost  
testování  
limity aplikace  
pozice sociologie v systému věd



Kapitola 4.  
Jak si nakreslit plán  
aneb  
Na co jsou hypotézy

I'm short enough and ugly enough  
to succeed on my own.

Woody Allen, *Play it again Sam*

Až dosud jsme si jenom stěžovali. Stěžovali jsme si na nezvládnutelnou velikost přirozených systémů v sociálních vědách, na trampoty, které z toho vyplývají: na nutnost pracovat s neúplně popsaným a ještě neúplněji analyzovatelným systémem, na to, že jsme v důsledku této **redukce** vystaveni mnoha typům zkreslení. Stěžovali jsme si na to, že jsme schopni potvrdit existenci příčinné závislosti téměř výhradně jen experimentem.

A stěžovali jsme si také na náš druhý veliký problém: na nutnost nepřímého měření, na to, že **transformace** studovaného jevu do indikátorů, ve kterých tento jev pozorujeme, měříme a analyzujeme, je mnohem delší a mnohem riskantnější než ve většině jiných věd. Nařkali jsme nad nejistou validitou našeho měření.

V téhle kapitole se podíváme na to, co se s tím dá dělat. Neočekávejte tu žádné zázračné recepty. Udělat slušný empirický výzkum, to dá fušku. Mnoho a mnoho hodin přemýšlení a psaní ještě před tím, než máme vůbec možnost ušpinit si ruce fakty. Dlouhé dny v terénu, dlouhé dny strávené analýzou dat, a to se nakonec může stát, že naše hypotéza bude fakty vyvrácena. Alespoň jsme zjistili, že tudy cesta nevede. Horší je, když ji naše výsledky ani nepotvrdí, ani nevyvrátí. I to se někdy stane. Ale přesto přese všechno, je to velká krása dělat výzkum. Vždyť studovat vztahy mezi lidmi je velké dobrodružství. Pevnina sociálních vztahů a struktur je ještě jen málo prozkoumaná a zaplňovat prázdná místa na mapě je veliké dobrodružství. Zkuste to, a uvidíte, že se vám to bude líbit. Ale pozor, je to návykové.

V téhle kapitole se budeme zabývat spíše logickými pravidly výzkumu než podávat recepty, jak technicky připravit výzkum. Tím se zabývá celá řada jiných knih. Věříme však, že tato

pravidla jsou důležitou pomůckou i pro optimalizaci rozhodnutí o mnohých technických krocích ve výzkumu.

Ještě jednu důležitou poznámku než opravdu začneme: v této kapitole, stejně jako v předchozích, se budeme zabývat hlavně kvantitativním výzkumem, a to ještě většinou v jeho silně standardizované verzi. Není to proto, že bychom tento typ výzkumu považovali za nadřazený. Jediný důvod pro tento popis je fakt, že v kvantitativním výzkumu jsou všechna naše zranitelná místa jasně viditelná.

#### 4.1. Výzkum jako idiot

Můžeme začít třeba pravdivou pohádkou z počítačového pravěku:

##### Pohádka pro odrostlejší děti 6.

##### O tom, co je počítač

Před dávnými a dávnými časy byl představen tisku jeden z prvních počítačů. Byl obrovský, nesmírně drahý, pomalý a neuměl o moc víc, než trochu lepší kalkulačka. Ale protože to bylo dávno a dávno, novináři byli okouzlení. Hovořili s jedním z tvůrců toho počítače:

Novinář: "Pane profesore, ten počítač je génius!"

Profesor: "Ale, pánové, to se mýlíte. To je úplný idiot. Umí je jednu věc: rozeznat nulu od jedničky. Ale umí to velmi rychle a nedělá chyby."

Empirický kvantitativní výzkum v sociologii je také takový idiot. Umí jenom rozeznat, zda existuje souvislost mezi dvěma nebo více proměnnými. Bohužel, neumí to rychle a dělá mnoho chyb.

Kvantitativní výzkum není nic jiného než testování hypotéz. Podívejme se blíže, co to vlastně znamená v termínech epistemologie vědy. Při produkci poznání zde užíváme v podstatě dva odlišné postupy: indukci a dedukci.

**Deduktivní metoda** vychází z teorie nebo z obecně formulovaného problému. Teoretický nebo praktický problém je přeložen do jazyka hypotéz. Hypotézy navrhují, jaké spojení mezi proměnnými bychom měli najít, je-li naše hypotéza pravdivá. Pak následuje sběr dat. Odpovídají-li závislosti mezi sebranými daty vzorci předpovězenému v hypotézách, přijmeme hypotézy jako platné. Jinak musíme hypotézy odmítnout.

**Induktivní metoda** je značně odlišná. Začíná pozorováním, ve kterém pátráme po pravidelnostech, vzorcích, které snad existují v objektivní realitě. Objevené pravidelnosti popíšeme ve formě předběžných závěrů. Ty pak ověřujeme dalším pozorováním. Konečným produktem je nová teorie.

A teď si zkusme, zda jsme rozdílům mezi dedukcí a indukcí dobře porozuměli:

**Cvičení 4.1.**

*Kterou z těchto metod používal velmistr dedukce, slavný detektiv Sherlock Holmes?*

Tabulka 4.1.

DEDUKCE	INDUKCE
Teorie	Pozorování
▼	▼
Hypotézy	Nalezené pravidelnosti
▼	▼
Pozorování	Předběžné závěry
▼	▼
Přijaté/zamítnuté hypotézy	Teorie

Kvantitativní výzkum používá deduktivní metodu. To znamená, že je schopen řešit jen určitou kategorii problémů:

Kvantitativní výzkum může nalézt řešení jen pro takové problémy, které je možno popsat v termínech vztahů mezi pozorovatelnými proměnnými.

Přeformulovat problém tak, aby vyhovoval těmto podmínkám, přeložit teoretický problém do jazyka výzkumu, nemusí být snadné. Naštěstí máme nástroj, který nám tuto operaci podstatně ulehčí. Jsou to **pracovní hypotézy**. Je to velice užitečný nástroj, který nám pomůže minimalizovat problémy vyplývající z redukce a transformace informací, nevyhnutelné v sociologickém výzkumu.

#### 4.2. Aby hypotézy nebyly jen hypotetické

Cesta od teorie k výzkumu je dlouhá a musíme se snažit, aby nebyla příliš klikatá. První krok od teorie k výzkumu je obvykle formulace základních, ještě dost obecných hypotéz. Taková hypotéza může být definována třeba takto: "*Hypotéza je očekávání o charakteru věcí, vyvozené z teorie*" (Babbie 1979). Uveďme si alespoň jeden příklad takové hypotézy:

##### Hypotéza 1.

Čím více podnětů týkajících se politiky osoba dostává, tím vyšší je pravděpodobnost, že bude participovat v politice.

Lester W. Milbrath, Political Participation, Chicago 1965, str.39

Tato hypotéza je mnohem konkrétnější, než většina hypotéz na této úrovni. Přesto nemůže být ještě bezprostředním východiskem pro technickou konstrukci výzkumu a výzkumných nástrojů. Je třeba ji rozložit do souboru **pracovních hypotéz**.

My definujeme pracovní hypotézu takto:

Tabulka 4.2.

1. Pracovní hypotéza je tvrzení, předpovídající existenci souvislosti mezi dvěma nebo více proměnnými.
2. Všechny proměnné zmíněné v hypotéze musí mít validní operační definici.
3. Soubor pracovních hypotéz musí zahrnovat nejen proměnné reprezentující zkoumané koncepty, ale i ty proměnné, které mohou významně zkruslit interpretaci testovaných vztahů.

Především si musíme ujasnit, co znamená termín "operační definice". Začneme jednoduchým příkladem. Co je vlastně kyselina? Chemik by nám nabídl definici, popisující "kyselinu jako sloučeninu s vodíkovým atomem, který je schopen odštěpit se ve vodním roztoku jako vodíkový kation  $H^+$  a udělit roztoku kyselou reakci.." (Příruční slovník naučný, 1963. II.díl, str.741).

Operační definice kyseliny by mohla znít asi takto: *Kyselina je taková tekutina, která změni barvu lakmusového papírku z modré na červenou*. Vidíme, že operační definice je mnohem méně precizní, daleko volněji spojená s teoretickými koncepty. **Nevysvětluje nám vůbec, co kyselina je**. Má však jednu velice užitečnou vlastnost: dává nám bezprostřední **návod, jak poznat, co je kyselina**.

V operační definici je koncept vyjádřen popisem operací, kterými bude měřen.

Ovšem i operační definice musí být zakotvena v teorii. Problém je, že v sociálních vědách

je vzdálenost mezi teoretickým konceptem a jeho operační definicí značná, a to může ohrozit validitu našich operací. Ale to všechno už známe: diskuse o operačních definicích je v podstatě opakováním naší diskuse o indikátorech z předchozí kapitoly.

Podívejme se zde alespoň na jeden klasický příklad operacionalizace. Emile Durkheim uvedl ve své klasické knize Sebevražda (1897) koncept anomie. Anomie je obvykle definována jako nepřítomnost norem. V dobách dramatických sociálních změn staré sociální normy ztrácí platnost a nové normy nejsou ještě obecně přijímány. Jednotlivec si již není jist, co bude spojeno s negativní sankcí a co s odměnou. Durkheim zvolil sebevraždu, "výsostně individualistický a osobní akt", za ukazatel anomie. Nesmírně komplexní teoretický koncept anomie byl silně **redukován**: anomie určité oblasti je operačně definována jako počet sebevražd za rok připadajících na každých sto tisíc obyvatel. I stručná úvaha nám ukáže, jak mnoho zajímavých a důležitých dimenzí původního konceptu je v této operacionalizaci ztraceno.

Redukce informací je nevyhnutelné břemeno, které nás bude provázet na každém kroku výzkumných operací. To však nemusí být fatální. Zhoubné by to bylo tehdy, kdybychom uvěřili, že operační definice je perfektní reprezentací teoretického konceptu, že můžeme skutečně položit rovnítko mezi anomii a frekvencí sebevražd.

Naštěstí praxe empirického sociologického výzkumu se obvykle zabývá koncepty značně jednoduššími, než je Durkheimova anomie. Uvidíme také brzy, že empirický výzkum není s to testovat platnost "grand theories" (a většinou ani teorií středního rozsahu) jako celku. Tyto operace jsou jednoznačně doménou teorie.

A teď se už můžeme vrátit k mnohem skromnějším konceptům a zamyslet se nad jejich překladem do jazyka pracovních hypotéz. Vraťme se k Milbrathově hypotéze o participaci v politice:

#### Hypotéza 1.

Čím více podnětů týkajících se politiky osoba dostává, tím vyšší je pravděpodobnost, že bude participovat v politice.

Podívejme se, jak tato hypotéza odpovídá naší definici pracovní hypotézy, uvedené v tabulce

3.2. Hypotéza předpovídá existenci spojení mezi dvěma proměnnými, ale pro žádnou z nich nemáme operační definici a ještě nevíme, co znamená třetí podmínka, zmíněná v tabulce. Je však zřejmé, že hypotéza 1. není pracovní hypotézou, že musí být do ní teprve přeložena.

Nejdříve můžeme začít s pátráním po měřitelných ukazatelích politické participace. Není těžké vyjmenovat jich celou řadu. Za krátkou dobu bychom mohli skončit se seznamem obsahujícím několik desítek přijatelných indikátorů. Ale které z nich vybrat? Takový seznam může obsahovat všechno, od faktu, že respondent je třeba předsedou vlády, kandiduje na politickou pozici, přes účast na demonstracích, až po prosté hlasování ve volbách a všichni víte z nedávné minulosti, že stávka, účast na mši, na určitém divadelním představení nebo koncertě může být navýsost politickým aktem. Stejně dobře si umíte představit, že záměrná neúčast v oficiálních kanálech politické participace, může být velice silným a významným vyjádřením politického postoje.

Co tedy vybrat? Záleží především na cíli výzkumu, na teoretickém rámci, ve kterém chceme politickou participaci studovat. Milbrathova hypotéza by mohla být použita právě tak ve studii zabývající se psychologickými a filosofickými dimenzemi politické participace, jako ve výzkumu snažícím se porovnat participaci ve standardních politických kanálech u různých skupin obyvatelstva. Operační definice politické participace by ovšem musela být v obou případech různá.

A ovšem, definice závisí na konkrétním politickém a historickém kontextu. Operační definice politické participace obyvatel hlavního města Prahy 1991 by se podstatně lišila od definice participace na téže populaci o pět let dříve. Jiné indikátory musí být použity pro měření politické participace v různých politických systémech. Indikátory použitelné v demokratických zemích by sotva daly smysluplné výsledky, kdyby byly aplikovány v zemi, kde je demokracie silně omezena. I v zemích se shodným politickým systémem mohou být ukazatele platné v jedné zemi být neplatné v jiné, stejně demokratické. Prostě se tyto země liší v politické kultuře, různé politické akty mohou mít různý význam a nebo nemusí mít v určité zemi vůbec žádný význam.

Participace v běžných formách politiky je v severní Americe měřena často ukazateli podobnými těm uvedeným v naší tabulce 4.3. Tyto ukazatele bývají často exportovány do jiných zemí, bohužel často jen mechanicky, a to může vyvolat povážlivé zkreslení,

nebezpečné zejména ve srovnávacím výzkumu. K tomu se vrátíme v jedné z našich budoucích pohádek. Bude to smutná pohádka, protože je pravdivá.

Naše tabulka je adaptací ukazatelů hierarchie politické participace, navržených Milbrathem (Milbrath 1982, str. 18) Indikátory jsou řazeny sestupně. První skupina zahrnuje aktivity, které jsou v angloamerické literatuře označovány jako "gladiátorské" aktivity, poslední skupina "divácké" aktivity. Prostřední reprezentuje přechod mezi oběma vyhraněnými typy.

Tabulka 4.3.

Výkon veřejné funkce nebo funkce v politické straně Kandidování na takovou pozici Vybírání peněz pro politické účely Participace na politických schůzích plánujících strategii Aktivní členství v politické straně Aktivní účast v předvolební kampani
Účast na politických schůzích a shromážděních Peněžní dary politické straně nebo kandidátovi Psaní dopisů veřejným nebo politickým funkcionářům
Politické plakáty na nárazníku auta, na trávníku před domem Přemlouvání druhých, aby hlasovali určitým způsobem Iniciování politické diskuse Hlasování

Ponechme stranou otázku platnosti hierarchie jednotlivých ukazatelů a podívejme se na problém validity těchto ukazatelů v našich poměrech.

#### Cvičení 4.2.

Zkuste odhadnout, které z těchto ukazatelů by mohly vyvolat zkeslení, kdyby byly použity pro výzkum politické participace v našich poměrech.

I když bychom pravděpodobně pro výzkum u nás některé z těchto ukazatelů vypustili, je zřejmé, že naše závisle proměnná, politická participace, musela být měřena v celé řadě proměnných. Protože si nemůžeme být jisti, že každý z indikátorů odráží politickou participaci ve stejné míře, každý z indikátorů musí být použit ve zvláštní pracovní hypotéze.

Podívejme se teď na naši nezávisle proměnnou, expozici politickým stimulům. Ta by mohla být sledována kupř. následujícími ukazateli:

- frekvence sledování politických pořadů v TV
- frekvence sledování politických pořadů v rozhlasu
- frekvence čtení politických úvodníků
- politické debaty v rodině
- politické debaty s přáteli
- odebírání nebo pravidelná četba politicky orientovaných periodik

Jistě by bylo možno v tomto výčtu pokračovat. Tradice politické participace v rodině může být velmi silným politickým stimulem. Tento jev by ovšem musel být operacionalizován do několika proměnných (kupř. kdo z rodiny participoval v politice; jak; jak intenzivně; kdy atd.) Jiným, pravděpodobně silným politickým stimulem by mohla být četba knih s politickou tematikou, a tak bychom mohli ještě hodnou chvíli pokračovat v našem výčtu.

Pro jednoduchost předpokládejme, že pro popis politických stimulů použijeme jenom prvních šest proměnných. Každá z nich ovšem musí být použita v samostatné hypotéze; každý z těchto politických podnětů může mít různě silný vliv na různé typy politické participace. Řekněme, že bychom pro popis politické participace použili třeba jen šest proměnných. Některé z nich jsou u nás sotva validní, jak jste sami zjistili ve cvičení 4.2. U jiných je frekvence výskytu tak nízká, že je v řadě výzkumů můžeme ignorovat (kupř. v USA v prvních třech typech "gladiátorské" participace působilo méně než 2% populace).

Máme tedy 6 proměnných popisujících politické stimuly a 6 dalších, popisujících politickou participaci. Naše pracovní hypotézy by mohly vypadat třeba takto:

- Osoby, které čtou pravidelně politické úvodníky, hlasují ve volbách častěji, než osoby, které je nečtou.
- Osoby, které čtou pravidelně politické úvodníky, jsou častěji aktivními členy politické strany než osoby ostatní.

- Čím častěji respondent sleduje televizní politické pořady, tím vyšší je pravděpodobnost, že se bude aktivně podílet na předvolební agitaci.

Abychom popsali všechny vlivy stimulů na všechny typy participace, museli bychom sformulovat 36 podobných pracovních hypotéz. Vidíme, že nám počet hypotéz pozehaně roste a to ještě nejsme zdaleka u konce s naším úkolem. Zatím jsme se ještě ani nestačili vypořádávat s třetí podmínkou z naší definice pracovní hypotézy: s nutností zahrnout do souboru hypotézy o proměnných, které by mohly podstatně zkreslit naši interpretaci vztahů mezi diskutovanými proměnnými.

A opět je tu naše stará bolest: komplexnost přirozených systémů v sociálních vědách. Podívejme se alespoň na nejmarkantnější příklady:

I kdyby čtení úvodníků nemělo vůbec žádný vliv na aktivní členství v politické straně, je vysoce pravděpodobné, že souvislost mezi čtením a členstvím stejně naměříme. Je to projev naší staré známé z prvé kapitoly, nepravé korelace. Ve většině demokratických společností vzdělanější osoby participují v politice více, než lidé méně vzdělaní. Ale více vzdělaní respondenti budou také s daleko větší pravděpodobností vystaveni politickým stimulům. Je tedy možné, že souvislost mezi politickými stimuly a participací je zcela nebo částečně vyvolána vlivem vzdělání na obě diskutované proměnné. Podobnou roli může hrát kupř. i sociálně ekonomický status. Osoby v určitém ekonomickém postavení mohou být angažovány v politice, protože je to dobré pro jejich obchod. Z téhož důvodu pátrají po informaci o politických problémech a záměrně se tedy vystavují politickým stimulům.

Takových situací, kde nějaká vnější proměnná ovlivňuje vztah mezi studovanými jevy, může být mnoho. Naštěstí, statistická analýza je schopna odizolovat vliv těchto vnějších proměnných. Můžeme toho však docílit jen tehdy, víme-li o každém respondentovi, jaké je jeho vzdělání, socioekonomický status atd. I kdyby náš výzkumný cíl byl velice úzký, kdybychom se zajímali jenom o vztah mezi politickými stimuly a participací, musíme do našeho souboru pracovních hypotéz zahrnout třeba tyto:

- Čím vyšší je jednotlivcovu vzdělání, tím je vyšší pravděpodobnost, že bude navštěvovat politické schůze a shromáždění.
- Osoby s vyšším vzděláním čtou politické úvodníky častěji než osoby méně vzdělané.

- Osoby, které vlastní v obci obchod, budou participovat v předvolební kampani pro místní volby častěji, než osoby ostatní.

Takových hypotéz budeme muset formulovat mnoho: víme z literatury i z denní zkušenosti, že účast v politických aktivitách je silně ovlivněna proměnnými jako pohlaví, věk, sociabilita jedince, jeho sociální sebedůvěra a mnoho dalšími a většina z těchto proměnných může být asociována s tím, jak silně je jedinec vystaven politickým stimulům.

Soubor pracovních hypotéz musí zahrnout hypotézy o všech proměnných, které mohou mít významný vliv na závisle proměnnou a jsou přitom asociovány také s nezávisle proměnnou.

A ještě jedna důležitá terminologická poznámka: **termín "závislá" a "nezávislá" proměnná neznamenají existenci kauzální závislosti** a často ani ne očekávaný směr vlivu. Většinou je to jen terminologická konvence, jejíž užitečnost vyplyne, až se budeme zabývat statistickou analýzou. Kupříkladu v našem případě, kdyby naše obecná hypotéza byla potvrzena daty, tento výsledek může být interpretován v termínech expozice stimulům ovlivňujících participaci, stejně dobře jako vliv participace na expozici stimulům. Ale to všechno už dobře známe z diskuse kauzality a experimentu v první kapitole.

#### 4.3. Je to k něčemu?

Teď už jasně vidíme, že konstrukce pracovních hypotéz je nesnadný úkol, již jen vzhledem k pouhému počtu nezbytných hypotéz. Je to úkol náročný, většinou spíše nudný, avšak bohužel nezbytný. Pracovní hypotézy hrají ve výzkumu několik velice důležitých rolí:

##### 1. Formulace pracovních hypotéz je testem, zda je výzkum vůbec možný.

V kvantitativním výzkumu můžeme zkoumat jen takové problémy, které je možné přeložit do jazyka pracovních hypotéz, tj. jen takové problémy, jež je možno vyjádřit jako vztahy mezi proměnnými, pro které máme validní operační definici.

## 2. Pracovní hypotézy jsou důležitým nástrojem pro optimalizaci redukce informací.

Lék na naši nejnebezpečnější chorobu, nutnost pracovat s neúplným popisem reality, nebude asi nikdy vynalezen. Musíme se tedy naučit, jak optimálně s touhle nemocí žít. Musíme si být rozumně jisti, že jsme vybrali alespoň ty nejdůležitější proměnné, které ovlivňují zkoumaný problém, a zároveň se vyhnout sbírání informací, které nutně nepotřebujeme. Víme už sice, že zavedení proměnných, které nepatří do zkoumaného systému, nevyvolá zkreslení, ale každá proměnná navíc znamená investici času, peněz a energie, a to je něco, co si prostě nemůžeme dovolit. Většinou jsou naše limity překročeny ještě dříve, než jsme zahrnuli do výzkumu proměnné, které považujeme za nezbytné. Omyly tohoto typu jsou bohužel dosti časté. Náš výzkum má totiž ještě jednu populární dětskou nemoc: začíná zpravidla formulací otázek pro interview nebo dotazník. To je zajímavá, vzrušující a tvořivá činnost; zdaleka ne tak nudná, jako formulace pracovních hypotéz. Skončíme se seznamem chytrých a zajímavých otázek, které s našim problémem nějak souvisejí. Bohužel, při analýze dat zjistíme, že pro řadu proměnných nemáme žádné použití. To není moc dobré. Ale horší je, když při analýze zjistíme, že kdybychom položili jednu jedinou, hloupou otázku navíc, mohli jsme opravdu něco objevit. To je opravdu nepříjemné, já vím. Sám jsem si to taky vyzkoušel.

Ovšem existence pracovních hypotéz, jak svědčí naše pravdivá pohádka č. 7, nemusí být automatickou pojistkou proti katastrofám tohoto typu.

### Pohádka pro odrostlejší děti 7.

#### Pohádka v prvé osobě jednotného čísla.

Bylo, nebylo, byl jsem kdysi zodpovědný za metodologii mezinárodního srovnávacího výzkumu. V průběhu se k nám připojili kolegové z jedné půvabné země. Převzali většinu našich pracovních hypotéz, přidali několik svých, změnili několik operačních definic tam, kde by naše původní indikátory byly v jejich zemi neplatné, a pečlivě testovali validitu nových indikátorů. Připravili takřka perfektní výzkumný nástroj.

Ale když přišla data na zpracování do Toronta, nevěřil jsem svým očím: Hledal jsem a hledal, ale nikde jsem nemohl najít údaj o pohlaví respondenta. Při přepisování dotazového archu vypadla rubrika pro záznam pohlaví. Nikdo si toho nevšiml, až bylo pozdě. U některých osob bylo možno pohlaví odhadnout. Uvedl-li respondent povolání "žena v domácnosti", pravděpodobně to nebyl muž. Ale ve většině případů museli tazatelé znovu do terénu. Nechtěl bych být v jejich kůži. Nemusí být právě příjemné zaklepat znovu na téměř tři sta dveří s pošetilou otázkou: Jste muž, nebo žena?

A tak, máme-li smůlu, ani pracovní hypotézy nepomohou.

## 3. Pracovní hypotézy obsahují základní informaci pro optimální rozhodnutí o technikách výzkumu

Tak jsme zase nazpátek u naší dětské nemoci "začneme s dotazníkem". Nejenže dotazník se v podmínkách postindustriální společnosti stává pomalu ale jistě nepoužitelným monstrem, ale na samém začátku výzkumu nemáme prostě dostatek informací k rozhodnutí, kterou z dost bohaté pokladnice technik sběru informací vybrat. Teprve pracovní hypotézy nám umožní se optimálně rozhodnout. Tomuhle problému však věnujeme později celou kapitolu. Pracovní hypotézy, protože popisují strukturu spojení mezi proměnnými, ukáží nám, alespoň v hrubých obrysech, jaký typ statistických operací budeme potřebovat. I v tomto časném stadiu výzkumu tak budeme schopni odhadnout, zda pro analýzu dat použijeme kalkulačku, osobní počítač, nebo zda budeme muset - je-li náš vzorek opravdu obrovský - bojovat o čas na střediskovém počítači.

## 4. Pracovní hypotézy jsou základem pro odhad rozsahu výzkumu.

Většinou teprve potom, když jsme připravili soubor pracovních hypotéz, jsme schopni alespoň přibližně odhadnout, jak náročný bude náš výzkum. Obvykle budeme překvapeni; teprve proces formulace pracovních hypotéz nám ukáže, jak mnoho proměnných, a ještě více analyzovaných vztahů mezi nimi potřebujeme pro řešení zdánlivě jednoduchého problému. Pracovní hypotézy jsou důležitým výchozím bodem pro časové, a tedy i finanční plánování výzkumu. A na závěr ještě jedno velice nevědecké, ale užitečné doporučení pro odhad času, který budeme potřebovat: udělejme nejdříve spíše pesimistický odhad, násobme odhadnutý čas třemi, a když všechno dobře půjde, snad tento nový termín splníme. Tenhle spíše prostoduchý recept se nám už často dobře osvědčil.

Vidíme tedy, že ty nepřjemné pracovní hypotézy mohou být velice užitečné. Neměli bychom se jim vyhýbat. Alespoň ne ze začátku, alespoň ne po prvních deset, dvacet let výzkumné praxe. Ale pak už je z toho návyk; dost zdravý a hlavně bezpečný.

#### Řešení úkolů z kapitoly 4.

##### Cvičení 4.1.

Bohužel, tento mistr dedukce používal induktivní metodu. Snažil se najít vzorec, vysvětlení, které existuje ve fyzických stopách chování. Podívejte se pro potvrzení na kteroukoliv z Doyleových povídek, ve které tento hrdina vystupuje, a pokud by Vám tento důkaz nestačil, přečtěte si "Skandál v Čechách", kde Holmes explicitně popisuje svoji gnozeologii. Ale s tím se v téhle knížce ještě setkáme.

##### Cvičení 4.2.

Některé z ukazatelů prostě neodpovídají naší politické kultuře a tradicím, nicméně jejich použití by nevyvolalo vážné zkreslení. Jen frekvence pozitivních odpovědí by byla nízká. Psaní dopisů veřejným činitelům, finanční příspěvky na politické účely jsou u nás mnohem řidší než v USA, kde byla tato stupnice vytvořena. Jedna z nejběžnějších forem amerického politického výrazu, "bumperstickers" a volební tabule umístěné na trávníku před rodinnými domky u nás téměř neexistují. Když však jsou takové indikátory použity beze změn ve srovnávacím výzkumu, vyvolají nebezpečné etnocentrické zkreslení. S jedním takovým případem se později seznámíme.