

1

Jak se dělá věda

přírodní a sociální vědy
paradigma
stochastický charakter
přirozený systém
uzavřený systém
redukce informací
typy zkreslení
nepravá korelace
vývojová sekvence

chybějící střední člen
dvojitá příčina
redukce počtu proměnných
redukce analyzovaných vztahů
redukce populace
vzorek
redukce času
kauzalita

Kapitola 1
Jak se dělá věda

1.1 Co je vlastně věda?

Matematika, fyzika? Určitě. Sociologie? Doufejme. Filozofie, teologie? Zeptejte se kolegů z příslušných fakult a dostanete určitě kladnou odpověď. Ale co mají tyto všechny obory společného? A hlavně, jak poznáme, že něco je skutečně věda? Odpověď není snadná. Definice v encyklopediích a slovnících publikovaných v několika posledních stoletích nám nepomohou odpovědět na náš základní problém: jak poznat, co je a co není věda.

Teprve Thomas S. Kuhn (1962) analyzoval instituci vědy skutečně v sociologických termínech. Jeho přístup nám umožní vytvořit *operační definici* vědy, to jest popis operací, které musíme použít, abychom poznali, zda něco věda je. Kuhnovo pojetí – a to je pro naši diskusi velice důležité – nám pomůže zejména vyjasnit pozici sociologie v systému jiných vědních oborů.

Kuhn uvedl do teorie vědy dva důležité koncepty: *koncept paradigmatu a koncept „normální“ vědy*. **Normální věda** znamená pro Kuhna „výzkum pevně zakotvený v jednom či více minulých vědeckých výdobytcích, které určitá vědecká komunita přijímá jako základ pro budoucí praxi“. (Kuhn 1962, s. 10) Jinými slovy, normální věda je takový vědní obor, ve kterém komunita vědců přijímá shodné *paradigma*.

Slovníková definice termínu paradigma (v originále „paradigm“), „vzor (skloňovací apod.)“, nám příliš nepomůže. Kuhn používá tento termín v novém a velmi specifickém smyslu:

Paradigmata jsou přijímané příklady aktuální vědecké praxe, příklady, které zahrnují zákony, teorii, aplikace a instrumentaci. To vše poskytuje model, ze kterého vyvěrá určitá koherentní tradice vědeckého výzkumu.
(Kuhn 1962, s. 10)

Jenom to, co odpovídá paradigmatu, je vědou. O tom, co je přijatelné paradigma, rozhoduje komunita vědců. Definice vědy má tedy sociální charakter: **věda je to, co za vědu považují vědci v daném oboru**. Pátrat po nějaké objektivní, univerzální definici vědy je pak z hlediska naší diskuse zbytečné.

Kuhnova analýza „normální vědy“ nabízí velmi zajímavý vhled do sociální praxe produkce vědy. Paradigma má evidentní praktický význam: „Když vědec může přijmout paradigma jako zaručené, nepotřebuje zpravidla usilovat o znovuvybudování oboru začínaje prvními principy a nepotřebuje ospravedlňovat každý koncept, který uvedl.“ (Kuhn 1962, s. 19–20)

Naprostá většina aktivit probíhá v normální vědě uvnitř paradigmatu. „Hlavní operace, kterou se zabývá většina vědců v průběhu své kariéry, je oprásování.“ (Kuhn 1962, s. 24) Oprásováním míní Kuhn zpřesňování paradigmatu, jeho aplikaci na širší pole, a případně i odstraňování rozporů, které v paradigmatu zbývají.

Paradigma je nesporně důležitým nástrojem ekonomizace vědy. Nicméně tato výhoda není zadarmo: paradigma totiž zároveň omezuje množinu řešení a postupů, které jsou ve vědě dovoleny. Omezuje i soubor problémů, které normální věda smí řešit. Kuhn porovnává tuto situaci s řešením skládačky. Skládačka má jen jediné správné řešení. Stovky malých kousků musí být složeny tak, aby daly obraz Hradčan nebo půvabné mladé dámy. Přitom je snadné si představit, že umělec nebo dítě by elementy skládačky složili zcela jiným způsobem. Výsledný obraz by mohl být mnohem krásnější, mnohem významnější než očekávaný výsledek. Nicméně bylo by to **chybné řešení skládačky**.

„Kritériem dobré kvality skládačky není to, že výsledek je velice zajímavý nebo důležitý. Naopak, některé důležité problémy, jako kupř. léčení rakoviny nebo koncept trvalého míru, nejsou často skládačkou vůbec, protože nemají (uvnitř paradigmatu – M. D.) žádné řešení.“ (Kuhn 1962, s. 36–37)

Pracovat mimo rámec paradigmatu může být v normální vědě velice riskantní, v minulosti mnohý narušitel paradigmatu zemřel na hranici. V naší osvícenější době je pravděpodobnost, že takový narušitel bude publikovat ve vědeckých časopisech nebo že jeho habilitační práce bude přijata, nulová.

Jak je potom zásadní pokrok vědy vůbec možný? V historii vědy se stane poměrně zřídka, že více a více pozorování se zdá být v rozporu s existujícím paradigmatem. Tato rozporná pozorování najdou posléze cestu do prestižních vědeckých žurnálů, jsou ostře napadena, a z bouře krvavých vědeckých diskusí se zrodí nové přijaté paradigma. Kuhn označuje tento proces za **vědeckou revoluci**. Takové revoluce nejsou časté. Příkladem vědecké revoluce je třeba opuštění zeměměřného světového názoru nebo přijetí relativistické fyziky.

Z Kuhnovy koncepce vědy vyplývá ještě další důležitý závěr: obsah paradigmatu je v různých vědních oborech různý, každá vědecká komunita v určitém oboru vytvořila vlastní paradigma. To znamená, že neexistuje univerzální definice vědy; jsou jen definice přijímané v rámci určitého vědního oboru.

Rozdíly v paradigmatu mohou být dramatické i u oborů velice příbuzných. Kuhn uvádí příklad fyziky a chemie: „Pro chemika je atom helia molekulou, protože se chová jako molekula z hlediska kinetické teorie plynů. Na druhé straně, pro fyzika atom helia molekulou není, protože nevykazuje molekulární spektrum.“ (Kuhn 1962, s. 50)

Důležitější je, že ne všechny vědní obory patří do kategorie „normální vědy“. Jenom některé tradiční obory, jako matematika a astronomie, měly své první paradigma již v časně historii vědy. Jiné dozrávaly do paradigmatálního stadia mnohem později. V dalších vědách, jako kupř. v některých součástech biologie zabývajících se problémy dědičnosti, je paradigma ještě novinkou. A citujme Kuhna ještě jednou: „**Zůstává otevřenou otázkou, zda sociální vědy mají vůbec paradigma.**“ (Kuhn 1962, s. 15)

Mezi sociology je jistě nesrovnatelně méně jednoty v teoriích a metodách, než je tomu kupř. ve fyzice. Někdy se zdá, že se sociologové všude na světě shodují pouze v jediném bodu, totiž že nejsou dostatečně placeni. I když sociální vědy jsou v předparadigmatálním stadiu, element procesu vývoje vědy se vztahuje i na naše pole. Mnohem důležitější však je, že Kuhn nás staví před otázku, **proč není sociologie paradigmatální vědou.**

1.2 Kam se podělo paradigma?

Snad bychom mohli Kuhnovi namítnout, že alespoň v některých oblastech našeho oboru máme elementy paradigmatu. (Později v této knize budeme hovořit o kvantitativním a kvalitativním paradigmatu v metodologii sociologického výzkumu.) Nicméně univerzální paradigma, které by bylo univerzálně přijímáno alespoň pro nějaký podobor sociologie, skutečně neexistuje.

Neměli bychom být vlastně šťastní, že nejsme spoutáni svěřací kazajkou paradigmatu? Bohužel to není tak jednoduché. Obory patřící do domény normální vědy jsou často označovány jako vědy exaktní. Jejich nálezy se signifikantně liší od nálezů společenských věd:

- Nálezy exaktních věd jsou mnohem přesnější a spolehlivější než nálezy společenských věd. Exaktní vědy jsou často (i když ne vždy) schopny produkovat nálezy deterministického charakteru: „Když X, a jenom když X, pak vždy Y.“ Naproti tomu naše závěry mají vždycky **stochastický, pravděpodobnostní charakter**.
- Závěry v exaktních vědách mají mnohem univerzálnější platnost. Naše nálezy mohou být obvykle platné jen pro prostředí, ze kterého jsme sebrali naše data.
- Exaktní vědy používají experiment mnohem častěji nežli my a jsou mnohem častěji schopny nabídnout spolehlivou výpověď o kauzálním charakteru vztahu mezi proměnnými. Ustavit důkaz o kauzalitě je ve společenských vědách mnohem obtížnější, a často nemožné.

Pravděpodobně bychom mohli v tomto ne právě radostném výčtu ještě chvíli pokračovat. Ale již nyní nám tento výčet vnucuje otázku, proč je tomu tak. Tady je několik možných vysvětlení:

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 Platonovi 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.

H5:

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.

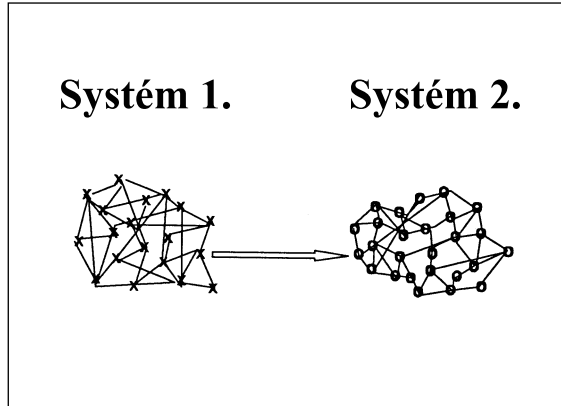
1.3 Co nezkoumat

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 elementy 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 protože 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

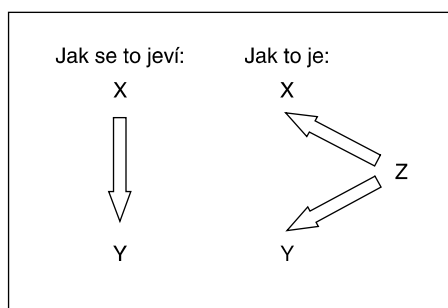
Představme si, že systém, který máme zkoumat, sestává jenom ze tří proměnných. Jenom dvě z těchto proměnných (X a Y) jsme zahrnuli do našeho výzkumu, ne však třetí proměnnou (Z). Třeba proto, že o ní nevíme, nebo ji prostě neumíme měřit. Podívejme se na některé typy zkreslení, které tak mohou nastat.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

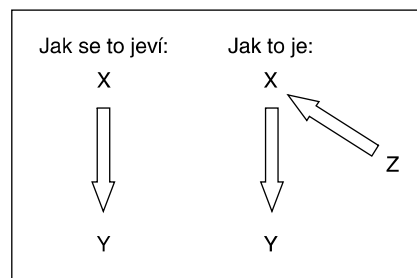
Jistě nám nehrozí nebezpečí, že bychom přijali hypotézu, že čápi nosí děti. Ale představme si, že nepravá korelace se zdá potvrzovat naši oblíbenou hypotézu. Potom výzkumník musí mít objektivnost anděla a trpělivost nerostného krystalu, aby pracně zabil to, co se po měsíce pokoušel dokázat.

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

Vývojová sekvence

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

Graf 1.3
Vývojová sekvence



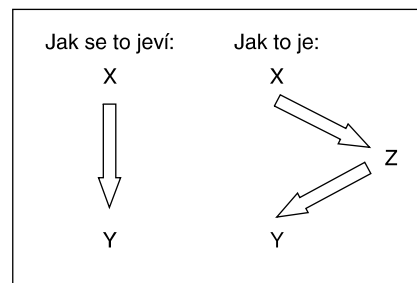
Taková situace je skutečně naprosto nevyhnutelná. Každá příčina má totiž jinou příčinu, ta zase jinou příčinu, která má opět svoji příčinu, a tak bychom mohli pokračovat až k aktu stvoření, nebo k tomu, co astronomové nazývají Big Bang. To je problém velmi dobře známý filozofům, kteří ho obvykle nazývají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

Nicméně někdy může předčasné přerušení 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 zkreslení. 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

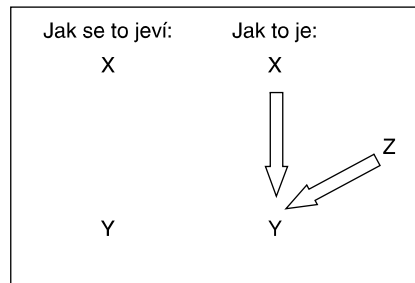
Zamyslete se, prosím, nad předchozím odstavcem a navrhnete alternativní hypotézu ukazující mužským šovinistům, že takové výsledky nepotvrzují superioritu nás, pánů tvorstva.

Zkreslení tohoto typu může být nebezpečné. Můžeme je často najít v kvazivědeckých 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 zkreslení 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 zkreslení?

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 zkreslení, 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 popsaným přirozeným systémem**, a tudíž budeme vždy vystaveni zkreslení, která jsme diskutovali v předchozím textu. Tato zkreslení 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.

Redukci počtu pozorovaných proměnných

a její efekt na vznik zkreslení 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. Nieméně 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 stu 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 30 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 jmény 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 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 závalu, 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, závislé proměnné. Zjistit tento fakt v sociálních vědách může být obtížné a velmi často i nemožné. S výjimkou experimentálního postupu (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, v 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.