

Cvičení 2.7.

Konstrukci hypotetických výsledků můžeme provést na základě následujícího souboru otázek:

- Ve kterých ze čtyř skupin se projevívá vliv filmu?
- Ve kterých vliv "historie"?
- Ve kterých vliv "testování"?

Odpověď na tyto otázky je snadná. "Historie" zasáhla všechny skupiny; předpokládáme, že vojáci ve všech skupinách se dozvěděli o útoku na Tel Aviv. "Testování" ovlivnilo jenom skupiny ve kterých bylo použito předběžné měření, a ovšem, film mohl mít vliv jen na obě experimentální skupiny.

Nejvyšší musí tedy být výsledky následného měření v první skupině. Je zde kumulován vliv všech faktorů: filmu, "testování" i "historie". Nejnižší bude skóre v druhé kontrolní skupině, ale i to bude vyšší, než bylo skóre prvních dvou skupin v předběžném měření. I tato skupina byla ovlivněna "historií". Výsledky ostatních skupin budou někde mezi těmito oběma extrémami. Jejich pořadí podle velikosti bude záviset na relativní síle obou zkreslujících faktorů. Tabulka odpovídající těmto podmínkám by mohla vypadat třeba takto:

| | Čas 1 | Film | Čas 2 |
|---------------------------|-------|------|-------|
| Experimentální skupina 1. | 50 | ANO | 85 |
| Kontrolní skupina 1. | 50 | NE | 65 |
| Experimentální skupina 2. | -- | ANO | 80 |
| Kontrolní skupina 2. | -- | NE | 60 |

A jistě si dovedete sami vypočítat, že "historie" odpovídá za 10 bodů přírůstku ve všech skupinách (rozdíl mezi "následným" měřením ve druhé kontrolní skupině a předběžným měřením v kterékoliv z prvních dvou skupin). Rozdíl mezi experimentální skupinou jedna a dvě v následném měření (stejně jako rozdíl mezi oběma kontrolními skupinami) nám dá sílu efektu "testování": 5 bodů. A konečně, čistý efekt filmu, 20 bodů, dostaneme porovnáním výsledků kterékoliv experimentální skupiny s odpovídající kontrolní skupinou a ještě několika dalšími způsoby. Víte jakými?

Pozorování nepozorovatelného

transformace informací
zkrácení v interview
validita
konzulita
kontrola validity
členství ve známé skupině
prediktivní validita

konstruovaná validita
mínění skupiny soudců
výčet obsahu
zjevná validita

Kapitola 3.

Pozorování nepozorovatelného

Od doby, co Američané vynalezli Richarda Nixona, neměl nikdy nikdo tolik trablů.

Richard Condon, BANDICOOT

Tohle bude krátká kapitola, ale obtíže, na něž tu narazíme, by stačily na celou knihu. V předchozí kapitole jsme hovořili o našem hlavním problému, o tom, že díky velikosti "přirozených systémů" v sociálních vědách není možné, docílit úplného popisu a úplné analýzy reality. Redukce je naším nejdůležitějším, ne však jediným epistemologickým problémem. Tím druhým je problém transformace informací.

V tomhle typu nepřijemnosti nejsme sami. Naprostá většina věd nedělá závěry přímo z analýzy reality, ale na základě záznamu, popisu této reality. To platí i pro přírodní vědy. Představa vědce, který zatřepe zkumavkou, obsah zkumavky změně barvu a vydá dramatický kout a vědec začne tančit radostí, že jeho teorie byla potvrzena, patří dnes spíše do světa televizní fikce. I zde jsou konečné závěry navrženy na základě popisu, na základě protokolů o mnoha experimentech či pozorováních. A popis nutně znamená transformaci, překlad reality do symbolů jazyka.

Cvičení 3.1.

Některé vědy se obejdou bez tohoto kroku, bez transformace reality do jejího popisu. Zkuste jmenovat nějakou takovou šťastnou vědu a vysvětlíte, proč nemusí analyzovat pouhý popis.

Problém je v tom, že transformace informací je v různých vědách a pro různé pozorované problémy různě obtížná a různě spolehlivá.

Určit pohlaví respondenta není obvykle obtížný problém. Ale mnohé důležité sociologické koncepty nejsou dostupné přímému pozorování, například koncept odcizení. Jiné jsou dostupné pozorování, a přece je sledujeme obvykle jenom nepřímě. Jistě by bylo možné zjistit

zřejmě pozorováním bydliště respondenta, počet dětí, povolání, vlastnictví auta, pohlaví atd., ale ani se vztahem na tyto věci respondenta jenom ptáme, prostě proto, že přímé sledování by bylo příliš časově náročné a nákladné.

3.1. A proč se neptat?

Většinou nám nezbude nic jiného, než se zeptat. Jenom nesmíme zapomenout, že každé nepřímé měření zvyšuje riziko zkreslení, a to tím více, čím více je kroků mezi zkoumaným fenoménem a naším protokolem, záznamem o tomto fenoménu.

I v něčem zdánlivě tak jednoduchém, jako je proces odpovědi na otázku je těch transformací přinejmenším dost. Podívejme se silným zvětšovací sklem na následující transformaci:

otázka --> odpověď

Mnoho kroků mezi těmito dvěma body má pro nás dosti důležitý význam. Měli bychom je mít stále na mysli, když kupř. připravujeme otázky pro interview nebo (nedej bože) pro dotazník.

(1) Především, zkoumaná osoba musí naši otázku správně porozumět. To není vždycky tak jednoduché. "Správně porozumět" znamená, že respondent pochopí otázku přesně tak, jak to měl výzkumník na mysli, a dále to znamená, že všichni dotazovaní musí pochopit otázku nuprosto stejným způsobem. Jde nám přece - alespoň v kvantitativním výzkumu - o srovnatelnost dat.

Cvičení 2.2.

Je s následující otázkou všechno v pořádku ?

- Jaký je nyní Váš příjem?

Zamyslete se nad tím na chvíli.

Častým hříchem při konstrukci otázek je, že mylně generalizujeme náš specifický okruh znalostí na celou populaci. Můžeme si být jisti, že opravdu každý ví co je HZDS? Můžeme si být jisti, že opravdu každý používá určitý termín ve stejném smyslu, jako my? Někdy můžeme být opravdu překvapeni:

Pohádka pro odrostlejší děti 2

Egri Bikaver aneb Jak to nevyšlo.

Před mnoha a mnoha lety jsme dělali výzkum v jedné malé horské obci. Velice chudá obec to byla, s tak špatným JZD, že v něm nebylo ani co ukradnout. Byl to kvalitativní výzkum. Mimo jiné nás zajímal životní styl obyvatel. Ptali jsme se také: "Co nejčastěji pijete?"

Odpověď prvního družstevníka nás lehce překvapila. Odpověď druhého, třetího, desátého nás prostě šokovala. Všechny odpovědi byly stejné. Všichni dotazovaní odpověděli, že nejčastěji pijí Egri Bikaver. To byl pro nás šok: tohle víno, maďarská Býčf krev, bylo v té době pěkně drahé. Rozhodně nad poměry lidí v téhle obci.

Ovšem jsem si hned vykonstruoval teorii. Krásnou a chytrou, kompenzační. Velmi nízká životní úroveň deprimuje družstevníky. Pít je důležitou součástí místní kultury. a tak se lidé snaží udržet vysoký status alespoň v této oblasti, když všude jinde je jejich status zoufale nízký.

Tak jsem se šel hned pochlubit se svojí teorií osobě, která rozuměla obci nejlépe, místnímu panu faráři. Když se pan farář přestal smát - a smál se moc dlouho - vysvětlil mi jednu zvláštnost místního slovníku: Egri Bikaver se tu říká hořčák, v té době nejlacinějšímu, přišernému ovocnému vínu. Říká se tomu tak proto, že "když se toho napijete, bolí vás hlava tak, že řvete jako bejka". A to byl zase jednou konec mé geniální teorie.

(2) V dalším kroku se musí zkoumaná osoba rozhodnout, zda je vůbec ochotna odpovědět na naši otázku. Některé otázky mohou být pro respondenta nepřijemné, může se cítit jimi ohrožen, může chtít chránit své soukromí. Ne vždy jsme schopni takovou situaci předvídat.

[[Pak se náš respondent musí rozhodnout, jak odpoví. Odpoví pravdivě? Zná vůbec pravdivou odpověď na naši otázku? Institut für Demoskopie zjistil, že hlavní příčinou nepravdivých odpovědí v Německé spolkové republice v šedesátých letech byl prostý fakt, že respondent neznal pravdivou odpověď (Noelle 1967). Neznalost respondenta je velice častým zdrojem zkreslení; zkoumaná osoba nerada přiznává, že nezná odpověď, že nerozumí některému výrazu v otázce a tak stejně odpoví. Přirozeně taková odpověď je pro náš výzkum neplatná, ale my o tom nevíme.

Není příliš těžké empiricky dokázat, že riziko takového zkreslení je dost vysoké. Před několika lety moji studenti provedli následující test. Do krátkého dotazníku, zaměřeného na některé etické aspekty politického rozhodování, jsme zařadili vysloveně nesmyslnou otázku. Její český ekvivalent by zněl asi takto:

- Domníváte se, že koncept kufanditní pumpřidntnosti je:

- 1.-mravně akceptovatelný
- 2.-akceptovatelný jen někdy
- 3.-neakceptovatelný
- 4.-nevím

Dotazník byl pak aplikován na malý a nereprezentativní vzorek posluchačů prvního ročníku a výsledky byly docela zajímavé. Jen menšina respondentů zvolila jedinou "pravdivou" alternativu čtyři. Nestálo by za to, vyzkoušet, zda výsledky u nás by byly odlišné od výsledků z Kanady?

Ovšem někdy je nepravdivá odpověď záměrná. Obvykle proto, že určitá alternativa odpovědi je pro respondenta spojena s představou nějaké sankce. V Severní Americe může být obtížné získat pravdivé informace o příjmu a finančních záležitostech vůbec. "Co kdyby tazatel měl nějaké spojení s daňovými úřady nebo s konkurencí?" U nás, v ne tak dávné minulosti, hrála silnou zkreslující úlohu obava z politického postihu. Předpovědět možnost takového zkreslení je často nesnadné. v jednom z našich prvních výzkumů, asi v roce 1966, jsme dostali v malých vesnicích na moravskoslovenském pomezí řadu "radioaktivních" odpovědí. Ale v jedné z obcí naprostá většina respondentů odmítla říci, co měli v neděli k obědu. Proč, to nevím. Dobře to vystihuje výrok jednoho mého kanadského kolegy: "Typický respondent se spíše přizná k manželské nevěře, než by připustil, že si nečistí pravidelně zuby."

Zkreslení může vyvolat i představa pozitivní sankce. Respondent volí nepravdivou odpověď, aby se ukázal v nejlepším světle, nebo chce prostě tazatele potěšit. To dobře ilustruje naše pravdivá Pohádka 3.

Pohádka pro odrostlejší děti 3.

O vzorném tazateli a vzorné zkoumané osobě.

Byl jednou jeden rozsáhlý výzkum o využití volného času ve velkém průmyslovém městě. Bylo to velmi náročné pro respondenty. Žádali jsme je, aby po řadu dnů zaznamenávali v 15 minutových intervalech, co dělali od probuzení do usnutí. Tazatelé byli studenti naší fakulty. Mezi nimi Jirka. Na Jirku připadla také jedna velice stará paní. Když uslyšela, o co jde, posílala Jirku za sousedkou. "Ta je mladá, pořád je někde, chodí na schůze a tak." Náš Jirka jí laskavě a moudře vysvětlil teorii konstrukce vzorku, význam sociologického výzkumu a nakonec babička podlehla a když Jirka po několika dnech přišel, aby odnesl záznam, byl moc překvapen. Babička byla v divadle, na výstavě, dokonce i na schůzích. Jirka jí moc pochválil a stará paní řekla: "Ale jsem moc ráda, že to mám za sebou. Já bych už víc toho chození po těch candrbálech nevydržela."

Tak to je vzácný případ toho, jak výzkum může zkreslit ne odpověď, ale chování zkoumané osoby. Všimněte si, že jsem známého sociologa průmyslu, dr. Jiřího Orta vůbec nejmenoval.

- (4) I verbalizace odpovědi, zejména v rozhovoru, může být spojena s některými problémy. Pro respondenta může být nepřijatelné vyslovit určitou nepřijemnou alternativu, použít neobvyklé slovo. Obvykle jsme však schopni těmto situacím předejít: i v řízeném rozhovoru můžeme u kritických otázek předložit dotazované osobě tišší listek se seznamem alternativních odpovědí a zeptat se, která alternativa nejlépe vystihuje jeho situaci. Pro respondenta může být příjemnější odpovědět třeba číslem 5, než vyslovit přímo, že jeho manželství je velice nešťastné a že již často uvažoval o rozvodu.
- (5) Záznam odpovědi může také vyvolat zkreslení. U otevřených otázek a zejména v kvalitativním výzkumu proces záznamu odpovědí - a někdy potřebujeme velice přesný záznam - může narušit přirozený charakter interakce mezi tazatelem a zkoumanou osobou. V standardizovaném rozhovoru je hlavním zdrojem zkreslení seznam nabídnutých kategorií. Neúplný soubor alternativních kategorií a kategorie, které se překrývají, patří mezi naše

způsobů řízení. K tomu ještě přistupuje méně častý omyl, výběr chybné kategorie, ať již tazatelem nebo respondentem.

Vidíme tedy, že i zdánlivě jednoduchý proces, jako je odpovídat na otázku, představuje celou řadu transformací, z nichž každá může být spojena s povážlivým zkreslením. Co si pak můžeme se zkoumáním konceptů, které nemají přímou empirickou reprezentaci? A v sociologii máme takových konceptů opravdu mnoho. Odcizení, prestiž povolání, třídní vědomí, autorita a stovky jiných konceptů, se kterými pracuje sociologická teorie, je téměř nemožné pozorovat přímo.

3.1 Jak nám respondent řekne to, co neví

Ve většině věd jsou pozorování a měření převážně nepřímá. Studovaný fenomén je reprezentován nějakým indikátorem. Kupř. tělesná teplota je indikována pozicí rtuť v stupnici teploměru, věk zkoumané osoby je reprezentován její odpovědí na příslušnou otázku. Čím větší je vzdálenost mezi tím, co chceme studovat, a indikátorem, který jsme schopni pozorovat, tím větší je riziko zkreslení. V sociologii je tato vzdálenost značná a většinou zahrnuje dlouhou řadu transformací.

Nalezení spolehlivého ukazatele je jednou z nejkritičtějších operací ve výzkumu. Dopustíme-li se omylu v této fázi, všechny následující operace budou neplatné. Měli bychom zde aplikovat řízení, populární mezi lidmi kolem počítačů od dob kdy počítače byly velké jako dům a byly jenom dva. Dnes se vlastně nepoužívá řízení, ale jen zkratka: GIGO. Znamená to "garbage in, garbage out". Volně přeloženo to znamená asi "svinstvo na vstupu dá svinstvo na výstupu". Výzkum můžeme pokazit kdykoliv, ale nezačtyčené omyly při konstrukci indikátoru jsou fatální.

Abychom se mohli produktivně bavit o problémech s tím spojených, musíme zavést dva nové pojmy: pojem validity a pojem reliability.

Validní měření

je takové měření, které měří skutečně to, co jsme zamýšleli měřit.

Reliabilní měření

je takové měření, které nám při opakované aplikaci dává shodné výsledky, pokud se ovšem stav pozorovaného objektu nezměnil.

Obě definice jsou jednoduché a nezní příliš vědecky. Ale uvidíme, že jsou velmi užitečným a pro výzkum nezbytným nástrojem.

Podívejme se však nejdříve na některé příklady narušení validity a reliability. Naše pohádka č.4 ilustruje takový příklad.

Pohádka pro odrostlejší děti 4.

O muzeu a validitě

Bylo jednou v jednom krásném městě krásné Muzeum objevů vědy a techniky. Jeho ředitelem byl podnikavý a sečtělý mladý muž, a ten se jednou dočetl o studii podniknuté v jednom muzeu v Chicagu, která použila opotřebených koberec před jednotlivými exponáty jako ukazatele jejich popularity. Dočetl se také, že tato technika byla použita i jinde, a tak se rozhodl, že ji použije i ve svém muzeu.

Výsledky výzkumu byly poněkud překvapivé. Jako nejúspěšnější byl vyhodnocen exponát věnovaný úspěchům čtyřletého plánu hospodářské výstavby, sklář s uprášenými diagramy a plakáty, věcní, o které v té době předstíraly zdvořilý zájem jen reprezentanti nejvyšších mocenských orgánů.

Výzkum byl publikován, ředitel dostal vyznamenání a muzeum podstatně zvýšený rozpočet. Jenom několik set zasvěcených vědělo, že kolem nejúspěšnější expozice vede jediný přístup k jediným toaletám v muzeu.

Druhé velké nebezpečí je v narušení reliability. Je to podobná situace jako bychom měřili délku měřítkem, které se tu zkracuje, tu prodlužuje, a my o tom nevíme. Uveďme si docela realistický příklad nereliabilní otázky, jakou čas od času potkáváme ve výzkumech:

Chodíte do kina

-často
-ne tak často
-občas
-zřídka
-vůbec ne

Odpověď "často" může znamenat pro respondenta A "jednou za měsíc", pro respondenta B "dvakrát týdně", pro mě "dvakrát za rok" a znám jednoho televizního dramaturga, u kterého si význam odpovědi "často" vůbec netroufám odhadnout. V každém případě je tato otázka velice nereliabilní.

Zkusme v následujícím cvičení, jak jsme porozuměli pojmu validity a reliability. Otázka v tomto cvičení je jednoduchá, ale není právě nejjednodušší.

Cvičení 3.3.

Může nereliabilní měření být validní?

A co opačná situace: Může být měření, která je nevalidní, být přesto reliabilní? Asi ano. Je to takové měření, které nám dává konstantní, ale chybné výsledky. Je to podobná situace ve fyzickém měření, kdy by naše měřítko bylo kalibrováno v jiných jednotkách, než se domníváme. Následující příklad je hodně přehnaný, ale vystihuje dobře náš problém: domníváme se, že měříme v centimetrech a měřítko bylo kalibrováno v palcích. Při opakovaném měření stejného objektu dostaneme stejné výsledky, které však interpretujeme v centimetrech, a tak jsou naše závěry nevalidní.

Zajistit validitu pozorování je pro sociologický výzkum nesmírně důležité. My opravdu

musíme často od zkoumané osoby získat informaci, kterou tato osoba vůbec nezná. Tedy alespoň ne v té formě, v jaké ji hodláme použít pro náš výzkum. Platí se průměrného respondenta, zda se cítí alienován, bylo by zřejmě absurdní. Nejhorší na tom je, že mnoho respondentů by nám na tuto otázku stejně odpovědělo, třeba by nevěděli, co slovo alienace znamená; nechť si prostě přiznají svoji neznalost. I když mnozí z nás mají spíše skeptický názor na měření inteligence, IQ je stále užívaným konceptem. IQ měříme i u osob, které nikdy o tomto konceptu neslyšely. Na některé otázky dnes pravděpodobně nedostanete pravdivou odpověď. Jen málokterý respondent je ochoten přiznat, že je rasista, nebo že bije manželku. Proto se snažíme obdržet takovou informaci různými prostředky nepřímého měření. Povíme si o nich víc v příští kapitole. Ale už teď můžeme snadno odhadnout, že takové měření bude vystaveno značnému riziku zkreslení.

V takových případech je nezbytné kontrolovat, zda navrhané měření je skutečně validní. Většina z následujících výčetů technik pro kontrolu validity není snadná. Zdaleka ne všechny techniky, zejména ne ty nejužitečnější, je možno aplikovat ve všech situacích. Testování validity přidává další krok do výzkumného procesu. Je to krok časově náročný, a tedy i nákladný. To by nás však nemělo odradit od aplikace těchto technik, kdykoliv to je jen trochu možné. Nezapomeňme na GIGO!

První, a pravděpodobně nejmocnější, skupina technik pro kontrolu validity je založena na srovnání s nějakým vnějším kritériem.

1. Validita založená na členství ve známé skupině

Logika této techniky je velice jednoduchá: zkusíme náš nástroj na skupině, o níž víme, že má vlastnost, kterou má náš nástroj měřit. Tak např. otázky měřící rasismus by mohly být testovány na členech rasistických organizací. Jejich skóre by mělo být významně vyšší, než skóre vzorku z obecné populace. Nepřímé otázky, které mají zjistit, zda respondent bije ženu by mohly být vyzkoušeny na lidech odsouzených opakovaně pro tento delikt. Hlavní problém je zde existence a dostupnost takové skupiny.

2. Prediktivní validita

porovnává předpověď založenou na testovaném měření se skutečnými výsledky. Např. validita škály, která má předpovídat pravděpodobnost úspěchu středoškolských studentů na vysoké škole, je testována porovnáním výsledků testu se známkami, které student na univerzitě

skutečně docílil. Jen tehdy, je-li shoda obou souborů dat dostatečná, je test prohlášen za validní a bude i nadále používán. Je zřejmé, že tento postup může být smysluplně použit pro testování měření, která mají být používána opakovaně, a nelze jej použít pro jednorázový výzkum.

3. Souběžná validita

je testována tak, že měření téže vlastnosti je prováděno dvěma nebo více různými postupy. Můžeme třeba použít různé formulace nepřímých otázek, kombinovat je s projekčními otázkami (o těch hovoříme v příští kapitole) a můžeme v tomto souboru použít i přímou otázku. Čím větší je shoda mezi těmito různými přístupy k měření, tím je pravděpodobnější, že náš postup je validní. Velikou výhodou souběžné validity je, že tato technika je aplikovatelná takřka univerzálně. Určitou nevýhodou je však to, že plný test validity obdržíme teprve tehdy, když data byla sebrána a analyzována.

V další skupině máme jen jednu techniku:

4. Konstruovaná validita

V této technice je konstruován test hypotéz logicky spojících zkoumanou vlastnost s technikou měření, kterou používáme. Ilustrujme si to na tomto příkladu:

Profesor P.I. Toma vyvinul škálu pro měření frustrace. Aby dokázal, že je tato škála validní, zkonstruoval soubor hypotéz předpovídajících, za jakých okolností můžeme očekávat vysokou úroveň frustrace:

- Hypotéza 1: Očekávání těžkého úkolu v blízké budoucnosti zvyšuje hladinu frustrace.
- Hypotéza 2: Čím negativněji je tento úkol vnímán, tím bude frustrace silnější.
- Hypotéza 3: Čím méně zajímavý je očekávaný úkol, tím vyšší bude frustrace.

Studenti sociologie, očekávající zkoušku ze statistiky byli testováni. Jejich frustrace byla významně vyšší než frustrace kontrolního vzorku studentů, kteří neočekávali žádnou zkoušku. P.I. Tomova škála byla přijata jako validní ukazatel frustrace.

(Adaptováno z Therese L. Baker, *Doing Social Research*, 1988)

Poslední skupinu testů validity představují testy založené na obsahu měřeného jevu:

5. Validita založená na mříňní skupiny soudců

Skupina zkušených knihovniců, učitelů literatury, nám může pomoci posoudit, zda naše posouzení čtenářské vyspělosti zkoumaných osob, založené na výčtu přečtených knih, je přijatelné. Důležité je, aby každý z členů poroty pracoval nezávisle na druhých. Shoda mezi výroky různých soudců je pak dobrým ukazatelem validity.

6. Validita testovaná výčtem obsahu (Domain of Meaning)

To je v podstatě test, zda naše měření dostatečně kryje doménu jevu, který zkoumáme. Tento test může být soiva použit pro validizaci abstraktnějších konceptů. Výčet obsahu je užitečný zejména tehdy, kdy samotná definice zkoumaného konceptu není dostatečným nástrojem pro výzkumné operace. Kupř. srovnávat volný čas zkoumaných osob není snadný úkol. Prostý výčet aktivit, odpovídající definici tohoto testu validity není dostatečným řešením: jedna a tatáž aktivita je pro jednoho respondenta milovanou rekreací a pro jiného velice nepřijemnou povinností. Proto jsme pro výzkum o využití volného času navrhli tento postup: Po určité dobu budou zaznamenávány všechny aktivity zkoumaných osob a jejich trvání. Pak budou respondenti dotazováni na každou aktivitu, zda ji považují za volný čas. Takový postup může poskytnout validní, srovnatelnou definici rozsahu volného času.

1. Zjevná validita (Face Validity)

Většina učebnic výzkumných metod uvádí tuto techniku mezi jinými technikami testování validity. My se domníváme, že je to spíše popis situace, kdy kontrola validity není vyžadována. Kupř. jistě není třeba testovat validitu otázky v dotazníku, zjišťující respondentovo pohlaví. Bohužel, spoléhání na zjevnou validitu je aplikováno i na složitější situace. Babbie (1989, str.124) uvádí tento příklad: Počet stížností podaných dělníky odborové organizaci by mohl být použit jako ukazatel pracovní morálky. Asi bychom se shodli, že to může mít něco společného s měřenou vlastností. Mohli bychom se shodnout také na tom, že je to lepší ukazatel, než, řekněme, počet knih, které si dělníci vypůjčili v veřejné knihovny. Potud Babbie. Zjevná validita je založena na tom, že intuitivně spoléháme na to, že existuje dostatečná spojitost mezi zkoumaným jevem a naším měřením. Vyjádřeno velice prostě: "Face validity" je eufemismus pro situaci, kdy kontrola validity nebyla provedena vůbec. Někdy, jako tomu bylo v prvním příkladu v tomto odstavci, není na tom nic špatného. Ve složitějších situacích to může být nebezpečné.

Kdyby se vám snad po těchto kapitole zdálo, že náš profesionální život je příliš jednoduchý, přidejme nakonec ještě jednu nepřijemnou připomínku: Neexistuje nic takového jako univerzálně validní měření určitého konceptu. Validita prokázaná v určitém kontextu a pro určitý účel je automaticky platná jen pro tento kontext a shodný účel. Indikátor určitého jevu, který je perfektně validní pro jeden účel, může být nevalidní pro měření téhož jevu v jiné souvislosti a naopak.

Cvičení 3.4.

V této kapitole jsme diskutovali otázku, měřící frekvenci chazení do kina v kategoriích "často", "na tak často", "občas" atd. Dospěli jsme k závěru, že tato otázka není reliabilní, a tedy není ani validní. Umíme si představit situaci, ve které by byla validním měřením?

To také znamená, že validní indikátory "necestuj" dobře: exportujeme-li validní indikátor ze země do země, z jedné kultury do jiné, snadno může ztratit svoji validitu. Kupř. jedna škála měřící validně odcizení v USA obsahuje mimo jiné otázky: Čtete pravidelně Reader's Digest? Chodíte pravidelně do kostela? Negativní odpověď na tyto otázky signalizuje přítomnost odcizení. Když je tato škála exportována třeba jenom do Kanady, výsledky měření jsou poněkud zkreslené: Reader's Digest není v Kanadě příliš populární, chazení do kostela se zdá mít v kanadském kulturním kontextu jiný význam než v USA. Importovat takovou škálu v její původní formě k nám, by bylo prostě absurdní.

Uveďme si alespoň ještě jeden příklad. v době, kdy se odehrála naše pohádka č.5, byla ve výzkumu vztahů na pracovišti v USA ještě hojně používána škála zvaná SRA Attitude Survey (Dalas 1958). Tato škála byla považována za spolehlivé měřítko pracovní morálky, za "morale audit".

Pohádka pro odrostlejší děti 5.

Co by se mohlo stát...

Náš oblíbený profesor P.J.Toma se rozhodl studovat, jaký vliv na dělníky má participace v dělnické samosprávě. Jeho základní hypotéza byla, že čím vyšší je tato participace, tím vyšší bude pracovní morálka dělníků. Pro měření závisle proměnné použil škálu z SRA Attitude Survey.

Analýza dat potvrdila, že existuje skutečně souvislost mezi participací v samosprávě a morálkou dělníků. Vypočítané koeficienty byly překvapivě silné a velice významné. Problém byl v tom, že koeficienty byly negativní: čím vyšší participace, tím nižší morálka.

Profesor P.J.Toma přijal hypotézu, že participace v dělnické samosprávě podvrací pracovní morálku, zvyšuje jejich alienaci, a je proto zavržená. Problém je v tom, že v SRA škále je považován za indikaci nízké pracovní morálky souhlas s tvrzením jako:

29. Špatné pracovní podmínky mi znemožňují, abych podal nejlepší pracovní výkon.
42. Vedení nám nedává jasné příkazy a instrukce.
44. Vedení nás neinformuje o věcech, které bychom měli vědět.
62. Náš šéf by měl být přátelštější k zaměstnancům.

Náš P.J.Toma si neuvědomil, že tato tvrzení, stejně jako mnoho jiných ve škále, vyjadřují postoje, které by měl mít každý člen dělnické samosprávy. Zřejmě mu na jeho univerzitě utajili, že posláním samosprávy je zaujímat kritický postoj k formálnímu vedení závodu. Domníval se, že měří souvislost mezi členstvím v samosprávě a pracovní morálkou. Ve skutečnosti však měřil souvislost mezi členstvím a postoji, které jsou očekávány od členů samosprávy. Není tedy divu, že jeho koeficienty byly vysoké: koreloval jen dvě části jednoho jevu.

Výzkum popsáný v naší pohádce nám tedy neřekl nic o vlivu samosprávy na pracovní morálku. Ale není bezcenný. Ukázal nám zřetelně, že měření, jehož validita byla uspokojivě prokázána v jednom sociálním prostředí, může být tragicky (či komicky) nevalidní na jiné populaci.

Do stejné kategorie patří i problémy třídního, sexistického a rasového zkreslení testů

inteligence. Jejich výsledky snad mohou být validní pro bílé studenty z rodin vyšší střední třídy, tedy v prostředí, ve kterém byly vyvinuty a ve kterém byla jejich validita nejčastěji testována. Aplikujeme-li však tyto testy na černého kluka z gheta, jehož negramotní rodiče žijí na veřejné podpoře, nebo na rómskou ženu z východoslovenské cikánské vesnice, výsledky mohou být daleko méně validní.

Pokud si snad myslíte, že našich trampot s logikou empirického výzkumu v sociologii není ještě dost, máme tu ještě jednu obzvlášť pěknou Jobovu zvěst. Měli bychom si ji velice dobře pamatovat:

Všechny závěry z empirického výzkumu, statistické důkazy atd. se nevztahují na jevy, které chceme studovat, ale jen na jejich indikátory.

Řekněme, že sledujeme vliv vzdělání na politickou participaci. Všechny pozorované vztahy, souvislosti, statistické důkazy se skutečně nevztahují na tyto dva koncepty, ale jen na jejich ukazatele. Skutečně tedy měříme souvislost mezi léty formálního vzdělání a faktem, zda respondent hlasoval v posledních volbách, přispíval finančně na volební fond nějaké strany, chodil do průvodu a prapory vyvěšoval, nebo přilepil volební heslo na nárazník svého auta. Někdy je rozdíl mezi indikátorem a studovaným konceptem zanedbatelný. Častěji je však zdravé, pořádně se zamyslet nad vzdáleností mezi konceptem a jeho ukazateli. Ale k tomuto problému se v naší knize ještě několikrát vrátíme.

Formální úkolů z kapitoly 3.

Cvičení 3.1.
Do této kategorie patří formální vědy, jako matematika a formální logika. Jejich předmětem je čirá abstrakce, a tak kategorie překladu "reality" do jejího popisu tu není aplikovatelná.

Cvičení 3.2.
To je opravdu špatná otázka. Nemůžeme vědět, zda částka, kterou respondenti uvedli, představuje jeho čistý nebo hrubý příjem, zda zahrnuje také příjem z vedlejšího povolání. Ptáme se, zda je to týdenní, dvoutýdenní, nebo měsíční příjem. Někteří respondenti nám odpoví hodinovým příjmem. Rozhodně nám tato otázka neposkytne výsledky srovnatelné pro všechny osoby v našem vzorku.

Cvičení 3.3.
Měření, které není reliabilní, nám nedává stabilní výsledky. V každém pozorování měří něco jiného, nebo měří jinak. Taková operace tedy měří něco jiného, než má. Je tedy nevalidní.

Cvičení 3.4.
Tato otázka je nevalidní jakožto ukazatel kvantity chůze do kina. Mohla by však být validním ukazatelem toho, jak je nasycena respondentova potřeba chodit do kina. V tom případě by bylo užitečné měřit frekvenci jinou, objektivnější otázkou, třeba: "Kolikrát jste byl za poslední měsíc v kině?"