

Jak studovat armádu v laboratoři

| | |
|--|---------------------------------|
| experiment | zrání |
| srovnávání statických skupin | prostředí |
| předběžné a následné sledování jedné skupiny | instrumentace |
| klasický experiment | regrese k průměru |
| následné pozorování na 2 skupinách | úmrtnost |
| Solomonův experiment | testování |
| typy zkreslení | limity aplikace |
| historie | pozice sociologie v systému věd |

Kapitola 2.
Jak studovat armádu v laboratoři

2.1. Vojáci v kině

Na následujícím fiktivním příkladu si budeme ilustrovat, co je a co není skutečný experiment a jaká je logika tohoto postupu. Mnoho Kanadčanů má z historických i ekonomických důvodů dosti smíšené postoje k USA. Řekněme, že by tento fakt v době války na Středním východě vzbuzoval obavy v kanadských vojenských autoritách a že se tyto autority rozhodly zlepšit postoje vojáků k Velkému Bratru. Mimo jiné byl připraven film o kanadsko-amerických vztazích. Účinnost tohoto filmu mohla být studována různými přístupy. Podívejme se na některé z nich:

Vzorec 1.:

Film byl promítán v posádkovém kině. Pak všichni vojáci zodpověděli dotazník. (Dotazník mohl být použit proto, že vojáci musí poslouchat, a tak dostatečná návratnost dotazníku byla zaručena.) Dotazník zjišťoval jen dvě věci. Zda voják viděl film, a sérii otázek měřících postoje k USA. Analýza výsledků ukázala, že postoje těch, kteří viděli film, jsou mnohem častěji pozitivní než postoje ostatních vojáků. Rozdíly mezi oběma skupinami byly tak výrazné, že bylo velmi nepravděpodobné, že vznikly náhodou.

Cvičení 2.1.

Můžeme si být jisti, že zjištěné rozdíly byly skutečně vyvolány filmem? Můžeme navrhnout nějakou alternativní hypotézu, která by vysvětlovala výsledky právě tak dobře?

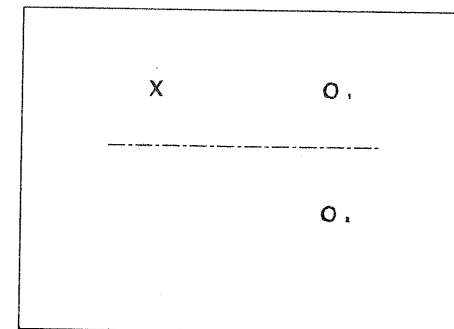
V tomto uspořádání jsme pozorovali, že existují souběžné změny v obou proměnných. Nevíme ale, co tu bylo dříve, zda postoje, nebo shlédnutí filmu. Jak uvidíme dále, ani vliv vnější příčiny nelze vyloučit. Byla zajištěna jenom jedna ze tří podmínek nutných pro ustavení existence kauzálního vztahu.

Campbell a Stanley, kteří napsali snad nejdůležitější knihu o logice experimentu v sociálních vědách, zařazují toto uspořádání mezi **před-experimentální vzorce** (pre-experimental designs) a používají pro něj název rovnávání statických skupin.

Zmínění autoři také zavedli grafický popis výzkumného uspořádání, který je od té doby téměř univerzálně používán. X v tomto grafu označuje expozici experimentální proměnné a O se subskriptem označuje pozorování nebo měření. Vzdálenost od leva do prava představuje tok času. Jedině v našem případě srovnání statických skupin používají také symbol čerchované čáry. K významu tohoto symbolu se vrátíme o něco později.

Graf 2.1.

Srovnávání statických skupin:



I když tento vzorec není příliš dobrým nástrojem, brzy uvidíme, že je pro sociologický výzkum velice důležitý.

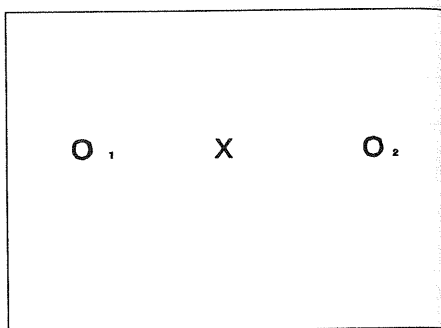
Vzorec 2.:

Tentokrát byl výzkum uspořádán jinak. Nejdříve byly měřeny postoje všech vojáků, po nějakém čase pak všichni vojáci viděli film, a pak byly postoje k U.S.A. měřeny znovu. Toto uspořádání, které bývá označováno jako předběžné a následné pozorování jedné skupiny (one group pretest post-test approach), je popsáno v našem grafu 1.6.

Řekněme, že po shlédnutí filmu se postoje vojáků vůči USA významně zlepšily. Můžeme si být teď jisti, že pozorované změny byly opravdu vyvolány shlédnutím filmu?

Graf 2.2.

Předběžné a následné pozorování jedné skupiny:



Předchozí model srovnávání statických skupin, nás ubezpečil jen o jedné ze tří nezbytných podmínek pro ustavení existence kauzálního spojení, o existenci souběžných změn. v našem přítomném vzorci k tomu ještě přistupuje informace o časovém pořadí. Nemusíme tedy pochybovat o směru kauzality. Je očividné, že postoje nemohly ovlivnit návštěvu filmu. Ale přesto se pokuste splnit úkol v následujícím cvičení.

Cvičení 2.2.

Navrhněte hypotézu, která by vysvětlila, že zmíněné výsledky bychom mohli dostat i kdyby film byl naprosto neefektivní.

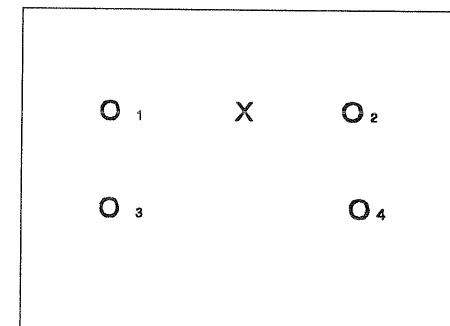
Když se nám podařilo vyřešit předchozí úkol, nebo když jsme si přečetli řešení na konci této kapitoly, je nám jasné, co je slabostí diskutovaného uspořádání. Předběžné a následné pozorování jedné skupiny není s to zajistit třetí podmínku pro ustavení existence kauzální souvislosti. Nemůže vyloučit zkrslující vliv nějaké vnější příčiny.

Vzorec 3.:

Tentokrát bylo použito formy klasického experimentu. Jeho logiku můžeme sledovat v grafu 1.7. Vojáci byli **náhodně** rozděleni do dvou skupin. V obou skupinách byly měřeny postoje k USA (O₁ a O₃ v našem grafu). Film byl však předveden jen jedné skupině, reprezentované horním řádkem v grafu. Pro skupinu, která byla vystavena experimentální proměnné, budeme užívat jméno **experimentální skupina**. Druhá skupina je nazývána **kontrolní**. Pak bylo na obou skupinách provedeno druhé měření (O₂ a O₄ v grafu).

Graf 2.3.

Klasický experiment



Následující tabulka obsahuje možné hypotetické výsledky. Čísla ve sloupci ČAS 1 jsou průměrné výsledky předběžného měření na obou skupinách, čísla ve sloupci ČAS 2 z následného. Čím vyšší číslo, tím kladnější jsou postoje.

Tabulka 2.1.

| SKUPINA | ČAS 1. | FILM | ČAS 2. |
|----------------|--------|------|--------|
| experimentální | 50 | ano | 70 |
| kontrolní | 50 | ne | 50 |

Jasně vidíme, že se postoje zlepšily v experimentální skupině, která viděla film. v kontrolní skupině výsledky zůstaly nezměněny. Zjistíme-li, že rozdíl mezi průměrným skóre 50 a 70 jsou statisticky významné (co to znamená, uvidíme v kapitole o statistice), můžeme prohlásit, že to byla experimentální proměnná - film, co změnilo postoje vojáků. Ale můžeme si být skutečně jisti, že vztah mezi našimi proměnnými (shlédnutí filmu a změny v postojích) má opravdu kauzální charakter?

Podívejme se, jak jsou tři podmínky pro ustavení kauzality respektovány v klasickém experimentu:

- **Souběžné změny** jsou zde zřejmé: přítomnost filmu je doprovázena změnou ve výsledcích, nepřítomnost filmu nevyvolává žádné změny v postojích.
- Informace o **časové posloupnosti** je zajištěna: známe postoje před i po shlédnutí filmu.
- Na rozdíl od všech až dosud diskutovaných uspořádání **můžeme vyloučit i zkreslující vliv nějaké vnější proměnné.**

Ilustrujme si tento fakt opět na fiktivních datech. Kdyby film byl neúčinný a změny byly vyvolány nějakým vnějším faktorem, pak bychom dostali výsledky podobné těm v tabulce 2.2.

Tabulka 2.2.

| SKUPINA | ČAS 1. | FILM | ČAS 2. |
|----------------|--------|------|--------|
| experimentální | 50 | ano | 70 |
| kontrolní | 50 | ne | 70 |

Každý vnější, s experimentem nesouvisející, vliv ovlivní nejen experimentální, ale i kontrolní skupinu. Kupř. zpráva o útoku na Tel Aviv by ovlivnila postoje v obou skupinách; třeba tak, jak to popisuje tabulka 1.3. Klasický experiment takto **upozorní na přítomnost zkreslení.**

Cvičení 2.3.

Jak bychom mohli interpretovat data obsažená v tabulce 2.3.?

Tabulka 2.3.

| SKUPINA | ČAS 1. | FILM | ČAS 2. |
|----------------|--------|------|--------|
| experimentální | 50 | ano | 80 |
| kontrolní | 50 | ne | 60 |

Klasický experiment je tedy velmi účinným nástrojem k ověření existence kauzální závislosti. Je jednou z forem **skutečného experimentálního přístupu** (true experimental design). Není však jeho jedinou formou. I v této knize se seznámíme alespoň s dvěma z nich. Všechny tyto postupy mají něco společného: každý skutečný experiment musí respektovat určité podmínky. Není-li některá podmínka dodržena, není tento postup skutečným experimentem, a tak nemůžeme na jeho základě učinit platné závěry o existenci kauzální závislosti.

Skutečný experiment musí respektovat všechny z následujících podmínek:

1. Výzkumník musí kontrolovat experimentální proměnnou.
2. Experiment musí použít alespoň jednu experimentální a jednu kontrolní skupinu.
3. Jednotlivci musí být zařazeni do těchto skupin náhodně.

První podmínka, **kontrola výzkumníka nad experimentální proměnnou**, znamená, že to musí být výzkumník (a ne zkoumané osoby), který rozhodne o tom, kdo bude vystaven zkoumanému podnětu. To byl problém, proč byl náš první vzorec (srovnávání statických skupin - graf 2.1.) tak neúčinný. V našem případě vojáci sami rozhodovali, zda půjdou, nebo

nepůjdou do kina. Proto jsme nebyli schopni určit, zda jejich postoje vedly k rozhodnutí shlédnout film nebo zda je platná hypotéza, že film změnil jejich postoje.

Druhé podmínce už také vlastně rozumíme. Provádíme-li měření jen na jedné skupině, tak jako v **předběžném a následném pozorování jedné skupiny** (graf 2.3.), vystavujeme se nebezpečí, že nějaký vnější vliv (bomby na Tel Aviv) budeme mylně interpretovat jako vliv experimentální proměnné.

Třetí podmínka je, že **pokusné osoby musí být zařazeny do experimentální a kontrolní skupiny náhodně**. Nad touto podmínkou se musíme na chvíli zamyslet.

Cvičení 2.4.

Představte si, že by někdo z nadřazených rozhodl, že třetí rota bude experimentální skupinou, a první rota bude reprezentovat kontrolní skupinu. Takové rozhodnutí může mít racionální jádro; nebyl by tak narušen průběh výcviku v ostatních jednotkách. Ale nevystavujeme se tím nějakému zkreslení?

Náhodné zařazení osob do experimentální a kontrolní skupiny sleduje velice náročný cíl. V ideálním případě by se měla experimentální skupina lišit od kontrolní jen v jediném bodu: experimentální skupina byla vystavena experimentální proměnné. Jinak by obě skupiny měly být totožné.

Jak se můžeme přiblížit tomuto cíli? V podstatě to není příliš těžké. Pracují zde za nás zákony pravděpodobnosti. Budeme o tom hovořit podrobně a uvidíme, že je obvykle těžší pochopit logiku náhodného výběru a zejména uvěřit jí, než tento výběr skutečně provést.

Náhodné začlenění vojáků do experimentální a kontrolní skupiny je technicky velice jednoduché. Obvykle použijeme velice rafinovanou techniku: hodíme si korunou. Padne-li řečně hlava, voják bude v kontrolní skupině, když lev, pak bude v experimentální. Každý voják má tak **stejnou pravděpodobnost**, že bude vybrán do experimentální skupiny a jsou-li skupiny dostatečně veliké, budeme mít v obou přibližně stejné procento vojáků s vyšším vzděláním, těch, kteří se právě vrátili z namáhavého cvičení, kteří měli zaražené vycházky, těch, které právě bolí zuby, kteří měli šťastnou schůzku s děvčetem atd. Obě skupiny si budou **podobné ve všech známých i neznámých vlastnostech** vojáků. Budou se lišit významně jen

v jediném: jenom experimentální skupina uvidí film. To však platí jen tehdy, kdy ekvivalence obou skupin byla vytvořena opravdu náhodným zařazováním subjektů do obou skupin. Jiné postupy pro docílení ekvivalence jako "matching" jsou z tohoto hlediska nedostatečné.

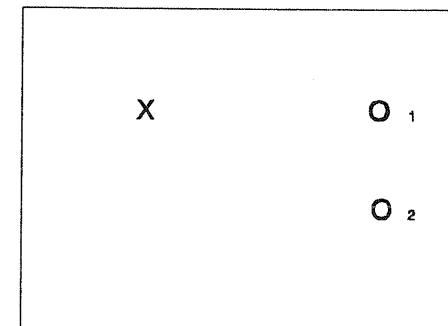
Vzorec 4.:

Následné pozorování na dvou skupinách je podobné klasickému experimentu s jednou důležitou výjimkou: předběžné pozorování bylo vypuštěno. Vojáci byli opět náhodně rozděleni do dvou skupin, experimentální skupině byl předveden film, a teprve potom byly v obou skupinách měřeny postoje. Vidíme to jasně v grafu 2.8.

Na první pohled se zdá takové uspořádání experimentu nemožné. Jak můžeme poznat, že nastaly nebo nenastaly nějaké změny, když neznáme stav mínění před tím než byla exponována experimentální proměnná. Ale je tomu opravdu tak? Zamysleme se znovu nad hypotetickými výsledky klasického experimentu:

Graf 2.4.

Následné pozorování
na dvou skupinách



Tabulka 2.4.

| SKUPINA | ČAS 1. | FILM | ČAS 2. |
|----------------|--------|------|--------|
| experimentální | 50 | ano | 70 |
| kontrolní | 50 | ne | 50 |

Vidíme, že obě předběžná pozorování nám dala totožné výsledky a stejně jsou i výsledky následného pozorování v kontrolní skupině. Když použijeme symboly z grafu 1.7., můžeme to vyjádřit takto:

$$O_1 = O_3 = O_4$$

Známe-li tedy výsledky následného pozorování v kontrolní skupině, dostáváme automaticky i informaci o výsledcích předběžného pozorování v obou skupinách. Neměli bychom být tímto faktem překvapeni. Víme přece, že byly-li pokusné osoby zařazeny do obou skupin skutečně náhodně, experimentální skupina se bude lišit od kontrolní jenom tím, že její členové byli vystaveni působení experimentálního podnětu.

Campbell a Stanley v citované knize argumentují, že jediným ospravedlněním pro používání předběžného měření je tradice. Výzkumníkům v průběhu jejich profesionální socializace bylo předběžné testování předloženo jako daná, nediskutovaná součást experimentu. Nyní se cítí mnohem bezpečněji, je-li předběžné měření součástí jejich experimentálního uspořádání, i když tento krok je v řadě experimentů zbytečný. Zdůrazněme však ještě jednou, že toto vše platí jen tehdy, když výběr do obou skupin byl proveden opravdu náhodně a jsou tedy ekvivalentní. To také předpokládá, jak o tom budeme mluvit později, že náš vzorek je dostatečně veliký. Nicméně v tomto výzkumném uspořádání jsou, ve srovnání s klasickým experimentem, určité rozdíly v tom, **jak** jsme informováni o přítomnosti zkreslení.

Cvičení 2.4.

Navrhněte, prosím, jaké výsledky bychom mohli obdržet, použijeme-li experiment na dvou skupinách pouze s následným měřením, když film by byl zcela neúčinný a změny byly vyvolány útokem na Tel Aviv.

Odpověď na předchozí cvičení nás přesvědčila, že i tento experiment odhalí přítomnost tohoto typu zkreslení. Nicméně obdobné výsledky (shodná čísla pro obě skupiny) bychom dostali i tehdy, když film byl zcela neúčinný. V situaci, kdy výsledný efekt je kombinací jak nějakého vnějšího činitele a účinnosti filmu (tedy něco takového, co je pro klasický experiment popsáno v tabulce 1.3.), viděli bychom z výsledků 80 a 60, že film byl účinný, ale nevěděli bychom, že nastala nějaká obecná změna v postojích, vyvolaná vnější příčinou. Je-li ovšem našim cílem měřit účinnost filmu, pak tato situace není vůbec na závadu.

Podívejme se však ještě na další, dosti složité uspořádání skutečného experimentu.

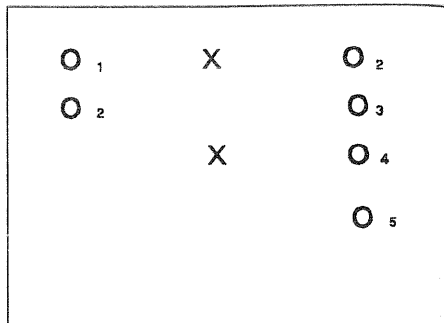
Vzorec 5.:

Tento typ, nazývaný Salomonův experiment na čtyřech skupinách, používá dvě experimentální a dvě kontrolní skupiny.

Zamyslíme-li se na chvíli nad tímto typem experimentu, uvidíme, že je to něco, co už vlastně dobře známe. Salomonův experiment je vlastně kombinací klasického experimentu (prvé dva řádky grafu) a experimentu používajícího jen následné měření na dvou skupinách (třetí a čtvrtý řádek grafu). Automaticky nám vytane otázka, k čemu je dobré takové plýtvání. Každý z obou skutečných experimentů je schopen ustavit existenci kauzálního spojení mezi proměnnými sám o sobě.

Graf 2.5.

Salomonův experiment na čtyřech skupinách



Důvod pro použití Salomonova experimentu je jiný. I když každý skutečný experiment nám bude signalizovat přítomnost velké většiny zkreslení, jenom Salomonův experiment nám umožní rozlišit, které ze zkreslení (nebo která podskupina zkreslení) je přítomno.

Stanley a Campbell navrhli velice moudrou klasifikaci možných zkreslení. My zde probereme alespoň ta nejdůležitější a budeme sledovat, jak jsou různá výzkumná uspořádání těmito zkresleními zranitelná.

Podíváme se blíže na následující typy zkreslení:

- historie
- zrání
- prostředí
- instrumentace
- regrese k průměru
- experimentální úmrtnost
- testování

První typ zkreslení, zvaný **historie**, už známe. Nastane tehdy, když nějaký vnější činitel (bomby na Tel Aviv), koinciduje v čase s uvedením experimentální proměnné a je mylně

interpretován jako vliv experimentální proměnné (filmu). Viděli jsme také v tabulkách 2.2. a 2.3., že ve skutečném experimentu nás výsledky na přítomnost tohoto zkreslení upozorní.

Druhý typ zkreslení, **zrání** (maturation), nastane tehdy, kdy změny vyvolané prostě tokem času jsou zaměněny za efekt experimentální proměnné. V dlouhodobých výzkumech to může být opravdu vliv dospívání. Kdybychom kupř. testovali novou metodu výuky čtení v průběhu prvního roku ve škole, zjistíme určitě nějaké zlepšení, i kdyby tato metoda byla velice špatná. Sedmileté dítě je mnohem zralejší než šestileté. Ve skutečném experimentu nás existence kontrolní skupiny opět upozorní na přítomnost tohoto zkreslení.

Třetí typ zkreslení, **prostředí**, není uvedeno v Campbellově a Duncanově výčtu. Oba autoři (a mnoho jiných po nich) zařazují pod záhlaví "zrání" i vliv toho, že zkoumané osoby jsou v době následného měření unavenější, zrudnější, nervóznější, hladovější atd., než byly v době předběžného měření. To se pak může projevit v tom, že postoje zjištěné v následném pozorování jsou mnohem častěji negativní, než by odpovídalo skutečnému vlivu experimentální proměnné. Zkreslení tohoto typu může ovšem nastat i v krátkodobých výzkumech. My se však domníváme, že tato situace většinou nepatří pod toto záhlaví. Zkreslení nebylo vyvoláno jenom tokem času, ale něčím, co je spojeno s organizací experimentu. Třeba to může být nutnost čekat mezi jednotlivými výzkumnými kroky, špatně větraná místnost, zákaz kouření atd. Ve většině případů bude těmto vlivům vystavena jenom experimentální skupina. Obě skupiny nejsou již ekvivalentní. Proto ani skutečný **experiment není schopen kontrolovat tento typ zkreslení**. Tak je možné, že ohodnotíme efekt experimentální proměnné mylně. Toto zkreslení je kvalitativně odlišné od ostatních zde zmíněných. Experiment je proti němu bezbranný a neexistuje žádný univerzální a snadný recept, jak tomuto zkreslení zabránit.

Instrumentace je zkreslení, vyvolané změnami v nástrojích měření mezi předběžným a následným pozorováním. Na prvý pohled se zdá existence takového zkreslení absurdní: nikdo přece nebude tak pošetilý, aby měnil nástroj uprostřed hry. Nezapomeňme však, že nástrojem měření může být i lidský pozorovatel. Ten může být v době následného pozorování unavenější, než byl při pozorování v čase 1, nebo díky zkušenosti z předběžného měření, se stal lepším pozorovatelem, atd. Někdy je však opravdu nutné, změnit náš nástroj technicky. Když kupř. časový odstup mezi předběžným a následným měřením je krátký, vojáci si mohou pamatovat, jak odpověděli na otázky měřícího postroje poprvé. Je dost pravděpodobné, že mnozí

budou prostě opakovat své původní odpovědi, i když se jejich původní postoje změnily. Stane se to třeba proto, že si pokusné osoby nechťejí zadat. Proto je pro následné měření používána jiná sada otázek. Jak uvidíme později, v pojednání o problémech transformací informace, připravit ekvivalentní sadu indikátorů je značně obtížný úkol. Snadno se může stát, že jeden soubor bude senzitivnější než druhý, že stejná úroveň postojů bude v jedné sadě vyjádřena vyšším skóre, než ve druhé. To pak může vést k mylným závěrům o efektu experimentální proměnné.

Každý skutečný experiment odhalí existenci tohoto zkreslení. V experimentu na dvou skupinách, používajícím jenom následné měření, toto zkreslení prakticky vůbec neexistuje. Zde měříme každou skupinu jenom jednou. Není tedy důvod, proč měnit sadu otázek. Protože obě měření jsou v tomto výzkumném uspořádání prakticky souběžná, i změny v lidském pozorovateli jsou méně pravděpodobné.

Regrese k průměru je nepříliš časté zkreslení v sociologickém výzkumu. Toto zkreslení nastane jenom tehdy, když experimentální skupina byla vybrána na základě nějakých extrémních výsledků. Chtěli bychom třeba testovat novou metodu výuky statistice, určenou zvláště pro studenty, kteří mají se statistikou velké potíže. Dali bychom studentům statistický test, a pro výzkum bychom vybrali jenom ty, jejichž známka by byla pod určitým limitem. Na těchto studentech bychom zkoušeli naši novou metodu. Při tomto uspořádání bychom na konci pokusu zjistili zlepšení i tehdy, když by naše metoda výuky byla neúčinná. Jak je to možné?

Problém je spojen s nedokonalostí měření v předběžném testu. Některé studenty jsme vybrali omylem. Vybrali jsme nesprávně několik studentů, kteří jsou dobří ve statistice, ale v době testu byli třeba nachlazení, stalo se jim něco nepříjemného nebo měli prostě smůlu. Ze stejných důvodů několik špatných studentů bylo omylem klasifikováno jako dobří. Ti však jsou vyloučeni z experimentu a nemohou tak kompenzovat zkreslení opačného směru. Při následném měření se tento omyl již nemůže opakovat a tak zjistíme nějaké zlepšení průměru i tehdy, když experimentální proměnná byla zcela neefektivní. Tomuto zkreslení je snadné předejít, když z populace "špatných" studentů vybrané tak nedokonale, jak jsme popsali výše, vytvoříme technikou náhodného výběru kontrolní a experimentální skupinu a aplikujeme na ně kterýkoliv z typů skutečného výzkumu. V následném měření sice bude stále přítomno zkreslení vyvolané regresí v obou skupinách, ale rozdíl v průměrném skóre skupiny, na kterou

byla aplikována nová technika a skóre kontrolní skupiny vyjádří skutečný efekt experimentální proměnné.

Experimentální úmrtnost je zkreslení, vyvolané tím, že některé osoby v průběhu výzkumu prostě odpadnou. Zkreslením se to stane teprve tehdy, když ztráta případů je systematická, když ztrácíme případy signifikantně více jenom v jedné z obou experimentálních skupin. Následující cvičení nám ilustruje tento problém:

Cvičení 2.5.

Profesor P.I. Toma nebyl spokojen s výukou metodám sociologického výzkumu. Navrhl proto novou metodu učení, zahrnující týdenní krátké známkové testy, důraz na statistickou analýzu a soustavnou kontrolu práce studentů. Chtěl si ověřit účinnost této metody a jako odborník připravil k tomu účelu perfektní klasický experiment. Před začátkem školního roku rozdělil zapsané studenty náhodně do experimentální a kontrolní třídy. Před začátkem přednášek obě skupiny psaly předběžný test. Nová metoda výuky byla uvedena jen v experimentální třídě, kontrolní třída byla vyučována tradičním způsobem. Na konci semestru obě třídy psaly opět test. Aby bylo vyloučeno subjektivní zkreslení, byla použita metoda slepého experimentu: všechny testy byly známkovány profesorem, který neznal studenty a nevěděl, kdo je v kontrolní a kdo v experimentální skupině. Výsledky experimentu jsou shrnuty v tabulce 1.5.

Tabulka 2.5.

| | Experimentální třída | Kontrolní třída |
|----------------|----------------------|-----------------|
| Předběžný test | 52.84 N=61 | 53.05 N=60 |
| Závěrečný test | 83.10 N=30 | 68.00 N=57 |

Čísla v polích tabulky reprezentují průměrná skóre (vyšší číslo = lepší výsledek), N reprezentuje počet studentů. Můžeme si být jisti, že nová metoda přináší lepší výsledky?

Žádná z experimentálních metod nám neukáže přítomnost tohoto typu zkreslení. Na druhou stranu, kontrolovat přítomnost tohoto zkreslení není vůbec obtížné. Stačí pouze sledovat počty pokusných osob na začátku a konci experimentu.

Testování je snad nejdůležitější typ zkreslení. Nastává tehdy, když předběžné měření samo ovlivní závisle proměnnou a my se mylně domníváme, že tyto změny byly vyvolány experimentálním stimulem. Můžeme si snadno představit, že otázky, kterými jsme měřili postoje vojáků v předběžném testu, mohou spustit u některých osob myšlenkový proces, který sám o sobě povede ke změně mínění o USA. Tak obdržíme zlepšení v postojích i tehdy když film byl totálně neúčinný.

Nebezpečí, že bychom skutečně zaměnili efekt prvního testu za efekt experimentálního proměnné, existuje jedině v předexperimentálních uspořádáních v "předběžném a následném pozorování na jedné skupině". Ve skutečném experimentu toto nebezpečí nehrozí. V uspořádání používajícím jen následné pozorování na dvou skupinách, zkreslení vlivem předběžného testování nemůže vůbec nastat, protože předběžné měření zde vůbec neexistuje. V klasickém experimentu je toto zkreslení dosti pravděpodobné, ale změny v následných skóre kontrolní skupiny nás upozorní, že část přírůstku ve skóre experimentální skupiny byla vyvolána nějakým zkreslujícím vlivem. Nevíme však jakým. Stejnou distribuci výsledků vyvolá i historie, instrumentace a poměrně zřídka se vyskytující zranění.

Někdy může být důležité, odlišit vliv testování od ostatních typů zkreslení, zejména velice často se vyskytujícího zkreslení typu historie. Zde je právě síla **Solomonova experimentu na čtyřech skupinách**. Tato časově, organizačně i finančně náročná technika nám pomůže diferencovat mezi vlivem testování a některými jinými typy zkreslení.

Nyní by bylo asi vhodné si osvěžit, jak vlastně tento experiment vypadá. Graf 2.5. nám připomene, jak je tento experiment organizován. Je to kombinace klasického experimentu (první dva řádky) a experimentu na dvou skupinách, používajícím jen "následné" měření (poslední dva řádky). Pokusné osoby byly ovšem zařazeny do všech čtyř skupin technikou náhodného výběru. Následující tabulka obsahuje fiktivní výsledky ze Solomonova typu experimentu. Pokusme se ji interpretovat.

byla aplikována nová technika a skóre kontrolní skupiny vyjádří skutečný efekt experimentálního proměnné.

Experimentální úmrtnost je zkreslení, vyvolané tím, že některé osoby v průběhu výzkumu prostě odpadnou. Zkreslením se to stane teprve tehdy, když ztráta případů je systematická, když ztrácíme případy signifikantně více jenom v jedné z obou experimentálních skupin. Následující cvičení nám ilustruje tento problém:

Cvičení 2.5.

Profesor P.I. Toma nebyl spokojen s výukou metodám sociologického výzkumu. Navrhl proto novou metodu učení, zahrnující týdenní krátké známkové testy, důraz na statistickou analýzu a soustavnou kontrolu práce studentů. Chtěl si ověřit účinnost této metody a jako odborník připravil k tomu účelu perfektní klasický experiment. Před začátkem školního roku rozdělil zapsané studenty náhodně do experimentální a kontrolní třídy. Před začátkem přednášek obě skupiny psaly předběžný test. Nová metoda výuky byla uvedena jen v experimentální třídě, kontrolní třída byla vyučována tradičním způsobem. Na konci semestru obě třídy psaly opět test. Aby bylo vyloučeno subjektivní zkreslení, byla použita metoda slepého experimentu: všechny testy byly známkovány profesorem, který neznal studenty a nevěděl, kdo je v kontrolní a kdo v experimentální skupině. Výsledky experimentu jsou shrnuty v tabulce 1.5.

Tabulka 2.5.

| | Experimentální třída | Kontrolní třída |
|----------------|----------------------|-----------------|
| Předběžný test | 52.84 N=61 | 53.05 N=60 |
| Závěrečný test | 83.10 N=30 | 68.00 N=57 |

Čísla v polích tabulky reprezentují průměrná skóre (vyšší číslo = lepší výsledek), N reprezentuje počet studentů. Můžeme si být jisti, že nová metoda přináší lepší výsledky?

Žádná z experimentálních metod nám neukáže přítomnost tohoto typu zkreslení. Na druhou stranu, kontrolovat přítomnost tohoto zkreslení není vůbec obtížné. Stačí pouze sledovat počty pokusných osob na začátku a konci experimentu.

Testování je snad nejdůležitější typ zkreslení. Nastává tehdy, když předběžné měření samo ovlivní závisle proměnnou a my se mylně domníváme, že tyto změny byly vyvolány experimentálním stimulem. Můžeme si snadno představit, že otázky, kterými jsme měřili postoje vojáků v předběžném testu, mohou spustit u některých osob myšlenkový proces, který sám o sobě povede ke změně mínění o USA. Tak obdržíme zlepšení v postojích i tehdy když film byl totálně neúčinný.

Nebezpečí, že bychom skutečně zaměnili efekt prvního testu za efekt experimentální proměnné, existuje jedině v předexperimentálních uspořádáních v "předběžném a následném pozorování na jedné skupině". Ve skutečném experimentu toto nebezpečí nehrozí. V uspořádání používajícím jen následné pozorování na dvou skupinách, zkreslení vlivem předběžného testování nemůže vůbec nastat, protože předběžné měření zde vůbec neexistuje. V klasickém experimentu je toto zkreslení dosti pravděpodobné, ale změny v následných skóre kontrolní skupiny nás upozorní, že část přírůstku ve skóre experimentální skupiny byla vyvolána nějakým zkreslujícím vlivem. Nevíme však jakým. Stejnou distribuci výsledků vyvolá i historie, instrumentace a poměrně zřídka se vyskytující zrání.

Někdy může být důležité, odlišit vliv testování od ostatních typů zkreslení, zejména velice často se vyskytujícího zkreslení typu historie. Zde je právě síla **Solomonova experimentu na čtyřech skupinách**. Tato časově, organizačně i finančně náročná technika nám pomůže diferencovat mezi vlivem testování a některými jinými typy zkreslení.

Nyní by bylo asi vhodné si osvěžit, jak vlastně tento experiment vypadá. Graf 2.5. nám připomene, jak je tento experiment organizován. Je to kombinace klasického experimentu (první dva řádky) a experimentu na dvou skupinách, používajícím jen "následné" měření (poslední dva řádky). Pokusné osoby byly ovšem zařazeny do všech čtyř skupin technikou náhodného výběru. Následující tabulka obsahuje fiktivní výsledky ze Solomonova typu experimentu. Pokusme se ji interpretovat.

Tabulka 2.6.

Průměrná skóre v Solomonově experimentu

| | Čas 1 | Film | Čas 2 |
|---------------------------|-------|------|-------|
| Experimentální skupina 1. | 50 | ANO | 80 |
| Kontrolní skupina 1. | 50 | NE | 60 |
| Experimentální skupina 2. | -- | ANO | 70 |
| Kontrolní skupina 2. | -- | NE | 50 |

Původní skóre v čase 1. jsou shodná ve všech čtyřech skupinách. Pro první dvě jsme to zjistili měřeními, pro druhé dvě to můžeme snadno odvodit. Známe výsledky pro následné pozorování pro kontrolní skupinu 2. Jak jsme poznali v souvislosti s experimentem používajícím jen "následné" měření na dvou skupinách, neměřený stav mínění obou skupin před experimentem se rovná výsledkům kontrolní skupiny z následného měření.

Když sledujeme výsledky následného měření, vidíme zřetelné zlepšení postojů v obou experimentálních skupinách, v první (která je součástí klasického experimentu) je tento přírůstek silnější. Ale vidíme také, že skóre vzrostlo rovněž v první kontrolní skupině, zatímco "následné" měření ve druhé kontrolní skupině zůstalo nezměněno.

To nám signalizuje přítomnost nějakého zkreslení, které se však objevuje jen v prvních dvou skupinách. Nemůže to být "historie", zpráva o útoku na Tel Aviv by ovlivnila postoje všech vojáků a skóre v obou kontrolních skupinách by bylo stejné. (V našem případě by skóre v kontrolní skupině bylo také 60.)

Takovou konfiguraci výsledků bychom mohli získat také vlivem "zrání", tehdy, kdyby osoby v prvních dvou skupinách byly vystaveny třeba dlouhému čekání mezi prvním měřením, filmem a následným pozorováním. To je v našem případě velice nepravděpodobné.

Uspořádání pokusu zcela vylučuje výskyt "regrese" (skupiny nebyly vybrány na základě nějakých extrémních výsledků) a "mortality". Vojáci by pravděpodobně neměli svobodu vypadnout z výzkumu a nadto by takový výpadek byl lehce pozorovatelný.

Zbývají tedy jen dva možné typy zkreslení, které jsou srovnatelné s distribucí výsledků v tabulce 1.6. Je to "instrumentace" a "testování". Pokud jsme pro následné měření v první experimentální skupině použili stejnou sadu otázek jako v prvním pozorování, je vliv "instrumentace" vyloučen. Ale i kdyby byla pro následné měření použita odlišná sada, nemůžeme zlepšení v kontrolní skupině 1. vysvětlit větší senzitivností druhého dotazníku. Stejný dotazník byl přeci použit i pro druhou kontrolní skupinu a její průměrné skóre by muselo být také 60.

Můžeme tedy uzavřít, že oním zkreslením, které je přítomné v naší tabulce je "testování". Zlepšení postojů v kontrolní skupině (z 50 na 60) bylo vyvoláno prvním měřením. Ale vliv testování ovlivnil také prvou experimentální skupinu. Jenom část přírůstku 30 bodů (z 50 na 80) byla vyvolána filmem. Deset bodů je způsobeno instrumentací. Čistý vliv filmu je jen 20 bodů. To je potvrzeno i čísly v posledních dvou řádcích tabulky. Protože zde neexistuje předběžné měření, zkreslení testováním zde neexistuje. Rozdíl mezi skóry experimentální a kontrolní skupiny reprezentuje tak čistý efekt filmu a potvrzuje náš předchozí závěr. Je to opět 20 bodů ($70 - 50 = 20$).

Tato výklad nebyl příliš snadný. Možná, že by stálo za to projít předchozí stránky ještě jednou. Salomonův experiment je velice efektivní nástroj. Nejen, že nám často umožní rozlišovat mezi různými typy zkreslení, ale hlavně sdružuje poznávací sílu klasického experimentu a experimentu na dvou skupinách používajícího jenom následné měření. Nadto nám pochopení tohoto typu experimentu poskytne dobrý vhled do logiky experimentu a kauzálního zdůvodňování obecně.

Když jsme si tedy znovu přečetli předchozí stránky, ověřme si na následujícím cvičení, jak dobře jsme porozuměli této problematice.

Cvičení 2.7.

Připravte, prosím, tabulku fiktivních výsledků Solomonova experimentu, ve kterém byl film účinný, ale kde bylo také přítomno zkreslení vyvolané "historií" a "testováním".

2.2. Proč to neděláme...

Všechno co mám rád, je buď drahé, nebo nezákonné, nebo nemravné, nebo se po tom tloustne.

Americká lidová moudrost

Je zřejmé, že skutečný experiment je nesmírně účinným nástrojem výzkumu. Ve většině případů je to jediný prostředek, jak dokázat existenci kauzální závislosti. Když však listujeme stránkami vědeckých sociologických žurnálů, brzy zjistíme, že jen poměrně malá menšina publikovaných empirických studií používá experimentální metodu. Proč jsme, my sociologové, tak nezodpovědní, že se vzdáváme jednoho z neúčinnějších nástrojů vědecké analýzy?

Bohužel to nezávisí na nás. Víme přece, že jednou z nezbytných podmínek skutečného experimentu je, že to musí být experimentátor, kdo manipuluje experimentální proměnnou. a háček je právě v tom, že dlouhou řadou proměnných, zcela základních pro většinu sociálních problémů, nejsme schopni manipulovat. U některých proměnných (jako věk, pohlaví) je to prostě fyzicky nemožné. U mnoha a mnoha jiných je to prakticky nemožné, nebo manipulace by byla eticky nepřijatelná. Nemůžeme z pokusných důvodů změnit vzdělání pokusných osob, jejich povolání, příjem, rodinný stav atd.

Dalo by se namítnout, že nemožnost manipulovat kupř. takovou proměnnou jako věk, by nemělo představovat velký problém. Víme přece, jak jsme se již zmínili, že věk nemůže být závisle proměnnou. Nicméně informace o časové následnosti sama o sobě je nedostatečná k tomu, abychom vyloučili vliv nějaké vnější příčiny. To je problém dobře známý třeba v gerontologickém výzkumu.

Můžeme sice z náhodného vzorku populace vybrat jedince, náležející k různým věkovým skupinám a porovnávat je ve vztahu k nějaké závisle proměnné. Řekněme, že by nás zajímalo, jak se s věkem mění obecná spokojenost se životem. Jistě nebudeme pochybovat, co je nezávisle proměnná. Problém je jinde: nemůžeme zjistit, jaký vliv na zjištěné změny v postojích mělo samo stárnutí a jaký je podíl jiných faktorů sdružených s kalendářním věkem. Lidé různých ročníků vyrůstali v různých sociálních podmínkách, vzdělání bylo pro ně různě dostupné, různé významné celospolečenské události (hospodářské krize, války, dramatické změny v sociálním a politickém uspořádání) je zasáhly v různém věku, nebo je

nezasáhly vůbec. Mnoho takovýchto vlivů může mít silný vliv na současnou spokojenost se životem a my nejsme s to, protože nejsme schopni manipulovat nezávisle proměnnou (věk), odlišit tyto vlivy od vlivu stárnutí.

Takováto situace je bohužel typická pro většinu empirických sociologických šetření. Většinou jsme schopni použít jen předexperimentální uspořádání. Téměř vždy to bude **srovnávání statických skupin**, které je znázorněno v grafu 2.1. Tento graf je podobný grafu popisujícímu skutečný experiment typu "srovnávání dvou skupin v následném měření" (graf 2.4.). Je tu však jeden, pro nás tragický, rozdíl: tato čára znamená, že my jsme nerozhodli, kdo uvidí film a kdo ne, že jsme nezařadili náhodně vojáky do experimentální a do kontrolní skupiny. Manipulace experimentální proměnné prostě není do tohoto postupu zahrnuta. Jenom jsme se ptali: "Viděl jsi film? Máš rád USA?"

Srovnávání statických skupin, prakticky jediný postup aplikovatelný v dotazníkovém šetření a téměř jediný použitelný v neexperimentálním výzkumu vůbec,

- je schopen nalézt existenci souběžných změn,
- nemůže poskytnout informaci o časovém pořadí vyvozenou přímo z výzkumných dat, a zejména
- není s to kontrolovat přímo zkreslující vliv nějakých vnějších vlivů.

To znamená, že v neexperimentálním výzkumu jsme ohrožováni všemi druhy zkreslení, které jsme probírali v této kapitole.

Je to tedy opět velikost přirozených systémů ve společnosti, souvislost všeho se vším, co způsobuje tyto naše problémy.

2.3. Co je jádrem pudla?

Doufejme, že tato kapitola zdůvodnila, že naše nemalé výzkumné problémy nejsou produktem inferiority sociologů, ale že mají svůj objektivní kořen v nesmírné velikosti přirozených systémů, se kterými musíme pracovat. Ale proč vůbec potřebujeme tak veliké systémy?

Pro odpověď se můžeme vrátit až k samým začátkům moderní sociologie. August Comte se ve svém grandiosním "Course de la philosophie positive" (1830 - 1841) zabýval také tříděním věd. I když filosofická podstata tohoto systému byla mnohokrát napadena, je kupř. principem, o který se opírá Deweyovo desetinné třídění, používané v řadě knihoven celého světa.

Comte navrhl tuto hierarchii věd:

| | |
|---|------------|
| 6 | sociologie |
| 5 | biologie |
| 4 | chemie |
| 3 | fyzika |
| 2 | astronomie |
| 1 | matematika |

Emanuel Chalupný ve své *Logice věd* (Chalupný 1945, p. 56-57) komentuje, že "každá věda číselně pozdější má *složitější* předmět než věda předchozí a opírá se o tuto vědu. ... Zároveň s tímto *přirostem* složitosti však nastává *úbytek povšečnosti* předmětu věd".

Jistě, naše dnešní universum věd je daleko širší, ale základní Comtov princip je stále aplikovatelný. Prostě matematik - uvnitř své profesionální činnosti matematika - se nepotřebuje zabývat výsledky chemie. Chemik však pro svoji profesionální činnost potřebuje výsledky práce matematika. Biolog, jakožto biolog (doufejme, že ne jako občan), může bezpečně ignorovat výsledky práce sociálních věd. Naopak my v sociálních vědách musíme respektovat relevantní nálezy všech věd, které mají v Comtově stupnici nižší číslo. Člověk je nejen sociální bytostí, ale je i bytostí biologickou, žijící ve fyzickém světě, ovládaném objektivními zákony přírodních věd.

Toto je pravděpodobně důvod, proč jsou přirozené systémy sociálních věd k nezvládnutelnosti obrovské. Tato rozměrnost je opět příčinou klesající obecnosti nálezů s přirůstající pozicí vědy v Comteově stupnici. Zákony matematiky mají univerzální platnost. Mohou. Jsou čistou abstrakcí. Nálezy biologa jsou aplikovatelné přinejmenším třeba na živočišný druh na kterém experimentoval. Nálezy sociologa můžeme přijmout jako platné jen pro zkoumanou populaci (a častěji jen pro použitý vzorek) v dané historické a sociální situaci. Každá generalizace nad

tuto úroveň má spekulativní charakter. Nemusí být mylná, ale nemá dostatečnou empirickou podporu.

Nutnost pracovat s neúplně popsanou a neúplně analyzovanou realitou není poslední z našich bolestí. Na další problémy spojené s transformací potřebujeme novou kapitolu.

Řešení úkolů z kapitoly 1.

Cvičení 2.1.

Nelze vyloučit, že směr kauzality byl právě opačný: vojáci šli na film proto, že měli pozitivnější vztah k U.S.A. Vojáci s negativním postojem na film spíše nešli.

Cvičení 2.2.

Je zcela možné, že změny v postojích byly vyvolány nějakou třetí vnější příčinou, která koincidovala v čase s uvedením filmu. Kupř. v období mezi prvním a druhým měřením byla uveřejněna zpráva o řízených střelách, které dopadly do obytných čtvrtí Tel Avivu. Byla to tato zpráva, a ne film, co ovlivnilo postoje vojáků k válce a k USA.

Cvičení 2.3.

Data v tabulce 2.3. ukazují, že zlepšení v postojích nastalo nejen v experimentální, ale i kontrolní skupině. Nicméně zlepšení je zřetelně silnější ve skupině, která viděla film. To nás vede k následujícímu závěru: film měl efekt na postoje, ale byl zde přítomen i vliv nějakého vnějšího faktoru.

Cvičení 2.4.

Jistěže ano. Můžeme si snadno představit, že jedna z rot je specializována a je v ní daleko vyšší procento více vzdělaných vojáků. Nebo jedna rota se právě vrátila z namáhavého cvičení, nebo její velitel zarazil vojákům vycházky. To všechno, a mnoho dalších vlivů, kterým byla vystavena **jen jedna** ze skupin, může ovlivnit postoje členů této skupiny. My pak můžeme mylně usoudit, že tyto rozdíly jsou funkcí expozice experimentálnímu podnětu (filmu).

Cvičení 2.5.

Odpověď je zde snadná. Obě následná pozorování by byla shodná; dostali bychom řekněme průměrné skóre 70 v experimentální i kontrolní skupině.

Cvičení 2.6.

Předpokládejme, že rozdíl mezi 83.10 a 68.00 je skutečně významný; studenti v experimentální třídě mají mnohem lepší známky, než v kontrolní. Ale podívejme se na počty studentů; na všech amerických a kanadských univerzitách je po několik týdnů po začátku pro studenty velice snadné nechat se vyškrtnout z vybraného kursu, který se jim nelíbí. Zatímco kontrolní třída ztratila jen 5% studentů, z experimentální třídy odpadla polovina. Je tedy pravděpodobné, že lepší výsledky nejsou dány tím, že by nová metoda produkovala lepší znalost, ale tím, že rigorózní postup ponechal v experimentální třídě jen vysoce motivované studenty, ty ochotné soustavně a tvrdě pracovat, ty kteří nemají potíže se statistikou. Ostatní, kteří by svými výkony snížili známkový průměr skupiny, byli rigorózní metodou učení zastrašení a vystoupili z kursu.

Cvičení 2.7.

Konstrukci hypotetických výsledků můžeme provést na základě následujícího souboru otázek:

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

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

Nejvyšší musí tedy být výsledky následného měření v první skupině. Je zde kumulován všech faktorů: filmu, "testování" i "historie". Nejnižší bude skóre v druhé kontrolní skupině, i to bude vyšší, než bylo skóre prvních dvou skupin v předběžném měření. I tato skupina je ovlivněna "historií". Výsledky ostatních skupin budou někde mezi těmito oběma extrémy. Je pořadí podle velikosti bude záviset na relativní síle obou zkreslujících faktorů. Tabu 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 přírůstkem 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, 2 bodů, dostaneme porovnáním výsledků kterékoliv experimentální skupiny s odpovídající kontrolní skupinou a ještě několika dalšími způsoby. Víte jakými?