

# Výzkum v sociálních vědách III.

POL 181 a BSS 104, 26.10.2017

# „Čtyři kauzální překážky“

**Abychom mohli mezi dvěma proměnnými konstatovat kauzální vztah** (nezávislá proměnná X ovlivňuje závislou Y):

1. Musí existovat věrohodný mechanismus, který spojuje X a Y.
2. Musíme si být jisti, že to není naopak a Y neovlivňuje X
3. Mění se Y s tím, jak se mění X (kovariance).
4. Neexistuje nějaká proměnná Z, která zároveň ovlivňuje X a Y.

# PŘÍKLAD

Jsme HRC a vyrobili jsme následující video:

<https://www.youtube.com/watch?v=RaxNEzA3jRs>

chceme zkoumat, jestli „funguje“

- **Zkuste navrhnout, co znamená „funguje“, tak, aby to byl vztah mezi proměnnými**
- **Zkuste „funguje“ operacionalizovat**

# Cesty, jak zjišťovat, zda video funguje

- Máme hypotézu, že ti, kdo viděli video, budou Donalda Trumpa hodnotit hůře, než ti, kdo ho neviděli („Sledování videa s negativní reklamou ovlivňuje hodnocení kandidáta, kterého se týká“)
- Abychom ji ověřili, potřebujeme **data**
- V rámci jakého výzkumného postupu se pohybujeme?
- Kolik případů bychom asi tak měli zkoumat a jakých?

# Od designu výzkumu k datům: Jak si vybírat případy

- Kromě **jednopřípadových studií** (další přednáška) sbíráme často data o větším množství případů, které jsou buďto z hlediska výběru případů:
  - **Cenzus** (vybraný vzorek rovná se celá zkoumaná populace)
  - **Výběr** (vybíráme jenom určité jednotky populace).
  - **Kdy zvolit cenzus, kdy vzorek a jak případy vybrat** (pokud zkoumáme politické systémy, strany atd.) **viz následující přednáška**

# Druhy výběrů („pokud zkoumáme lidi“)

- **Kvótní:** předem víme, že ve vzorku chceme určitá % jednotek s určitými charakteristikami, nemusí se shodovat se zastoupením v celé populaci.

- **Náhodné** (nekontrolujeme charakteristiky populace):

**Nevědecké:** ankety, „samovýběry“ (zde nemá každý stejnou šanci být ve vzorku, lidé vybírají sami sebe, že odpoví nebo neodpoví- ankety na internetu, ale (bohužel) i některé diplomové práce, distribuující dotazník metodou „sněhové koule“.

**Vědecké:** mechanismy zajištění „náhodnosti“ v kvótním výběru: systematicčnost (každý má stejnou šanci), stratifikovaný náhodný výběr (rozdělíme do skupin, z těch náhodně vybíráme), vícestupňový náhodný výběr (na začátku např. populace ČR, rozdělí se na vzájemně zastupitelné shluky)- každý subjekt v populaci má stejnou šanci, že se dostane do našeho vzorku (vs. např. výzkumy SANEPu).

**Co bychom si vybrali v našem příkladu?**

# Jak to můžeme zkoumat

**observační studií nebo experimentem (alias, pokud zkoumáme živé lidi, výběrová šetření)**

Výzkumný plán	Podstata	Typický postup
Experiment	Měří se efekty, které vznikají, když se manipuluje s jednou nebo více nezávislými proměnnými	Vybere se vzorek populace, jedinci nebo objekty se přiřadí k různým experimentálním podmínkám, uskuteční se plánované změny nezávislých proměnných. Měří se malý počet cílových proměnných. Analýza dat
Výběrové šetření	Sběr informací ve standardizované podobě od většího množství jednotek	Vybere se vzorek ze známé populace, od každého prvku se získá určité množství informací standardizovaným způsobem (dotazník nebo strukturovaný rozhovor). Analýza dat



# Experiment

- 2 klíčové charakteristiky:



výzkumník **manipuluje** s nezávislou proměnnou

výzkumník si **náhodně vybírá** a kontroluje, koho vystaví jakému podnětu nezávislé proměnné (např. kdo uvidí film)

Výhody: možnost odfiltrovat vliv třetích proměnných (protože skupiny jsou úplně ve všem kromě sledování videa stejné)

# Co bychom dělali v našem úkolu

- Nezávislá proměnná- sledování videa
- Lidi z našeho vzorku bychom pozvali do laboratoře, náhodně je rozdělili skupin, jedna skupina by viděla video, druhá ne
- Následně bychom měřili závislou proměnnou
- Je rozdíl mezi hodnotou závislé proměnné efekt našeho klipu?
- Musíme závislou proměnnou měřit i předtím?

# Observační studie

- Výzkumník sleduje jednotky výzkumu a měří hodnoty proměnných

**Nesnaží se žádnou z (nezávislých) proměnných manipulovat**

# Dva druhy observačních studií: průřezové a longitudinální

**Průřezové** (*cross - sectional*): zaměřují se na zkoumání více případů v jednom čase

**Longitudinální** (*time - series*): srovnávají jeden případ ve více časech

## **Kombinace obou**

Příklad: existuje hypotéza, že s vyšší nezaměstnaností se zvyšuje deficit státního rozpočtu.

V prvním případě by se testovalo např. na zemích EU v roce 2010, ve druhém na konkrétní zemi v dlouhé časové řadě.

# Co bychom dělali v našem úkolu

- Video bychom zveřejnili
- Následně bychom hledali lidi, kteří ho viděli a kteří ho neviděli a měřili u nich závislou proměnnou.
- Je rozdíl mezi hodnotou závislé proměnné efekt našeho klipu?

- **Jaká jsou slabší místa našeho poznání o efektu videa, získaného pomocí experimentu a pomocí observační studie?**
- **Jak experiment a observační studie ob stojí na čtyřech kauzálních překážkách?**

# Experiment

- První překážka (vztah mezi proměnnými musí dávat smysl)- překonali bychom spolehlivě
- Druhá překážka (nejdřív působí nezávislá proměnná, pak se mění závislá)- překonali bychom spolehlivě
- Třetí překážka (s tím, jak se v datech mění hodnota nezávislé proměnné, mění se i závislá)- pokud bychom něco naměřili, překonali bychom, problém by ale byl s velikostí efektu kvůli nepřírozenosti laboratorního prostředí, jsou cesty, jak to vylepšit-„experimentální realismus“
- Čtvrtá překážka (závislou i nezávislou proměnnou neovlivňuje jiná proměnná)- překonali bychom

# Observační studie

- První překážka (vztah mezi proměnnými musí dávat smysl)- překonali bychom spolehlivě
- Druhá překážka (nejdřív působí nezávislá proměnná, pak se mění závislá)- problematické, co když si klip vybrali hlavně naši mobilizovaní voliči a naopak Trumpovi voliči se mu vyhli?
- Třetí překážka ((s tím, jak se v datech mění hodnota nezávislé proměnné, mění se i závislá)- pokud bychom něco naměřili, překonali bychom
- Čtvrtá překážka (závislou i nezávislou proměnnou neovlivňuje jiná proměnná)- obojí by mohl například ovlivňovat zájem o politiku, vzdělání, stranická identifikace, nepřekonali bychom



# Závěr

- U tohoto konkrétního problému máme větší šanci redukovat omyl pomocí experimentu
- Není to tak ale **vždy, experimenty i observační studie mají svá silná a slabá místa**, která buďto povzbuzují nebo limitují jejich použití v konkrétním případě.

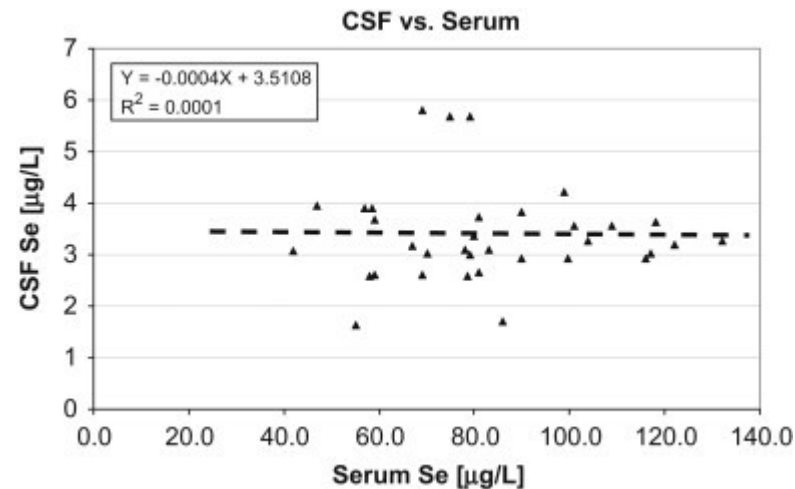
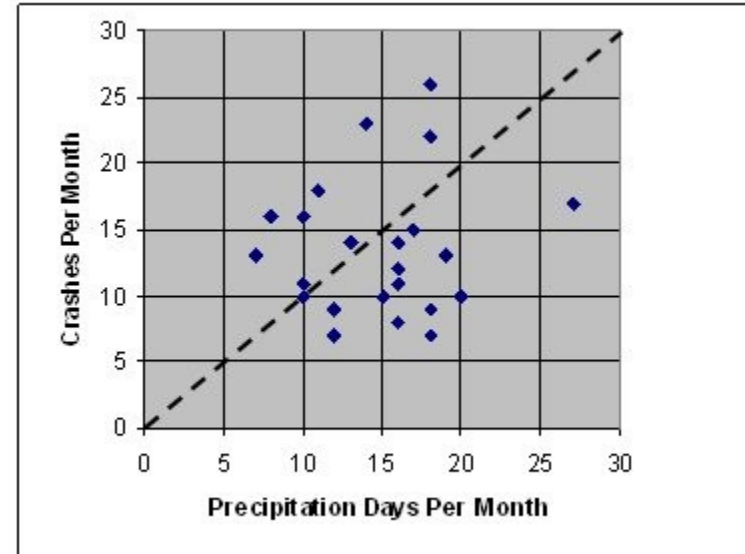
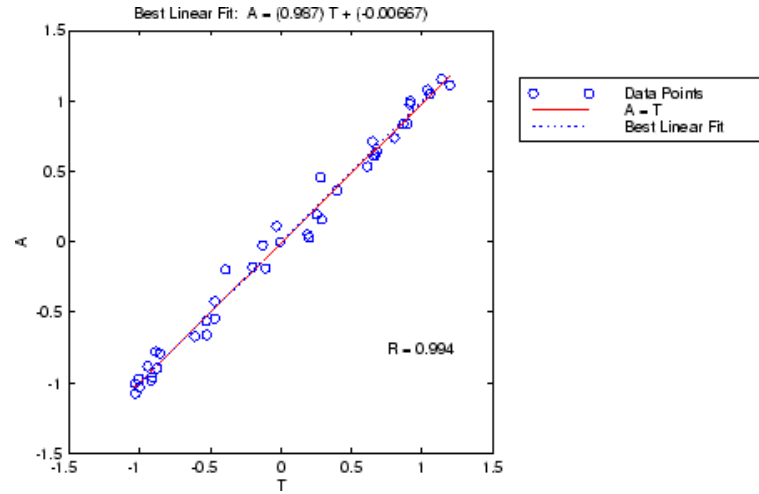
# Nevýhody experimentu a observačních studií

- **Experiment:**
- obtížná proveditelnost v některých případech/nemožnost provést. Některými nezávislými proměnnými totiž nejde manipulovat (nezávisle je přiřazovat).
- problematická externí validita (v reálném světě nemusí být kovariance stejná jako v laboratoři, tj. problémy na třetí překážce).
  
- **Observační studie**
- Velmi obtížná kontrola „čtvrté kauzální překážky“
- Nemůžeme si být skoro nikdy zcela jisti, že ve hře nejsou ještě nezávislé proměnné.
- Jednoduše proto, že lidé, kteří podobně jako v experimentu „viděli“ a „neviděli“ film, nejsou stejní, sami si to vybrali, jestli se budou dívat a třetí proměnné ovlivnily právě i rozhodnutí, zda se nechat vystavit efektu třetí proměnné.

# Testování hypotéz: co se děje s hypotézami, když už máme data

- Probíhá jinak, než v každodenním životě (tam si spíše vybíráme argumenty, podporující naše tvrzení)
- Ve vědě je testování hypotézy založeno na tom, že vědec vyvine nezanedbatelné úsilí, aby svou hypotézu prostřednictvím dat zpochybnil a zamítl (falzifikoval)
- Proti **hypotéze** (očekávání o stavu pozorování vztahu mezi proměnnými) stojí **nulová hypotéza** („to, co předpokládáme, neplatí“ „to, co budeme pozorovat, pokud neplatí to, co předpokládáme“).
- Př: **Hypotéza** říká: Se zvýšením A (vzděláním) se zvyšuje B (pravděpodobnost volební účasti). **Nulová hypotéza**: Se zvýšením A se B nezvyšuje (= nemění se ve zkoumaném souboru nebo snižuje).
- Používáme, tzv. testovou statistiku, říká nám, jaká je pravděpodobnost toho, co jsme naměřili, pokud by platila nulová hypotéza. Na základě toho přijímáme, resp. odmítáme nulovou hypotézu (jde o konvenci, kdy přijímáme a kdy odmítáme).
- Pokud není v průběhu výzkumu hypotéza zamítnuta, je na konci výzkumného procesu **kauzální teorie**, vysvětlující stav závislé proměnné prostřednictvím působení nezávislé proměnné (nebo proměnných).
- Proces vědeckého ověřování hypotéz** je velmi **přísný, aby nebyla** hypotéza zamítnuta, musíme např. ve velkém souboru mít jasnou („statisticky významnou“) evidenci o tom, že vztah existuje.

# Jak se statisticky prověřuje: Příklady lineárních závislostí



# Potřebujeme vždy hypotézy?

- Ne, někdy stačí i výzkumná otázka
- Př: H01: Pravděpodobnost, že jedinec půjde volit, roste se vzděláním
- H02: Pravděpodobnost, že jedinec půjde volit, klesá s tím, čím více svítí slunce...

**vs.**

- Čím (Jakými faktory) lze vysvětlit volební účast?

Jaké jsou techniky sběru  
dat?

# Techniky sběru dat

- Techniky sběru dat představují prostředky, pomocí kterých jsou získávána data. Obvykle k nim saháme v momentě, kdy už víme koho (jakou populaci a vzorek z ní) zkoumáme.
- Mezi základní techniky sběru dat patří **pozorování, dotazování, obsahová analýza a sekundární analýza**
- Techniky sběru dat se dále mohou dělit dělí na **obtrusivní** (vtíravé)- dochází při nich k interferenci se zkoumaným systémem- řadí se sem zejména dotazník, rozhovor a otevřené nezúčastněné pozorování a **neobtrusivní**, při nichž výzkumník neinterferuje se zkoumaným systémem (studium dokumentů, sekundární analýza, skryté nezúčastněné pozorování).
- **Induktivní a deduktivní strategie** používá **jiné techniky sběru dat**- pro induktivní jsou typické nestandardizovaný rozhovor, zúčastněné pozorování a analýza osobních dokumentů, zatímco pro deduktivní výzkum jsou to dotazník, standardizovaný rozhovor a nezúčastněné pozorování.

# Sekundární analýza

- Sekundární analýzou se rozumí využití dat, která byla již dříve získána k jiným (výzkumným) účelům.
- Bývá v sociálních vědách zmiňována na posledním místě, pro nás často hlavní technika sběru dat
- Disman: „Sociálněvědné výzkumy testují omezený soubor hypotéz a tyto testy představují jen omezenou množinu relevantních kombinací sebraných proměnných. V každém výzkumu je využita jen část užitečné informace, která byla v datech nashromážděna“.



# Příklady

- Pokud chceme vědět, jaká politická témata považují občané za nejpalčivější: data CVVM
- Sebezařazení ideologické: CVVM, Eurobarometr
- Vztah k autoritě: European Values Study
- Průzkumy veřejného mínění před volbami: řada agentur
- Výsledky voleb: volební komise (ČSÚ)

# Obsahová analýza

- Obsahovou analýzu děláme často, obvykle je to cesta, jak operacionalizovat a měřit nějaký vztah mezi koncepty, z nichž některý se týká obsahu.
- U analýzy dokumentů se jedná o empirickou metodu k systematickému, intersubjektivně prováděnému zkoumání obsahových a formálních znaků a sdělení, případně i autora a adresáta sdělení.
- Dokument je obecně jakýkoliv hmotný záznam lidské činnosti (úřední statistika, dopisy, osobní deníky, plakáty, letáky, články v odborných časopisech, hmotné stopy chování). Obsahová analýza může být použita i v kombinaci s jinými technikami, např. při zpracování dlouhých otevřených otázek v rozhovoru a obecně v kvalitativním výzkumu. I v obsahové analýze se obvykle pracuje s populací a vzorkem (populace = soubor sdělení)

# Obsahová analýza- příklady

- Srovnáváme MfD a HN v tom, jak píší o Hnutí ANO.
- Pokud spočteme celkový počet článků o politice a zkoumáme, v kolika z nich se objevují hesla jako „ANO “ nebo „Babiš“, děláme **kvantitativní obsahovou analýzu.**
- Pokud si vyberem články s heslem „Babiš“ a zkoumáme, zda je zmíněno v pozitivním, negativním nebo neutrálním módu, děláme **kvalitativní obsahovou analýzu.**

# Dotazování

- Patří sem nejen dotazníky, ale i rozhovory, fokusové skupiny...
- 1. Podle **stupně předstrukturovanosti** situace dotazování na málo strukturované, částečně strukturované a silně strukturované dotazování
- 2. Podle **stupně standardizace výzkumných nástrojů** a podmínek na nestandardizované dotazování, částečně standardizované dotazování a plně standardizované dotazování (dotazník bývá vysoce standardizovaný, rozhovor ne).
- 3. Podle formy získání dat **je ústní a písemné dotazování**, přičemž ústní lze dále dělit na přímý a telefonický (v současnosti typologii problematizují, resp. rozšiřují nová média, hlavně internet)
- 4. Podle počtu dotazovaných dotazování s jednotlivcem či se skupinou
- Podoba a „dramaturgie“ dotazníku podrobně- ***Disman: Jak se vyrábí sociologická znalost (nutnost, pokud budete mít dotazník v bakalářské práci!)***

# Pozorování

- **V politologii nejméně časté**
- **1. Naivní/vědecké** (vědecké pozorování se vyznačuje plánovanými postupy, systematickostí a konkrétním výzkumným účelem, zatímco naivní pozorování slouží zpravidla ke získávání každodenních zkušeností)
- **2. Strukturované a nestrukturované pozorování.** Účelem nestrukturovaného pozorování je pokud možno zhuštěný popis politického jednání. Začíná s se spíše vágně formulovaným seznamem otázek, které připouští otevřenost k neočekávanému. Pozorování je doprovázeno analýzou: poznámky jsou strukturovány. Na jejich základě jsou formulovány kategorie, které dávají návod k dalšímu pozorování. U strukturovaného pozorování se naproti tomu předpokládá schéma kategorií ke klasifikaci způsobů jednání předem.
- **3. Otevřené a skryté pozorování.** Při otevřeném pozorování vystupuje pozorovatele otevřeně jako výzkumník, zatímco při skrytém pozorování svoji identitu skrývá (což při strukturovaném a systematickém pozorování lze ztěžít).
- **4. Zúčastněné a nezúčastněné pozorování.** Při zúčastněném pozorování je výzkumník sám elementem pozorovaného sociálního pole, zatímco v opačném případě zůstává vně. Zúčastněné pozorování lze dále dělit na aktivní a pasivní pozorování podle toho, zda se se zkoumaným politickým polem identifikuje či nikoliv.

# Jak hodnotit „kvalitu“ výzkumného designu: **Interní a externí validita**

Pokud můžeme hodně vsadit na to, že výzkum má dobrou schopnost v rámci zkoumaných případů bezpečně odhalit kauzalitu (že  $x$  ovlivňuje  $y$ ) má vysokou **interní validitu** (a naopak)

Pokud si můžeme být jisti, že výsledky našeho výzkumu jsou platné i mimo kontext (případně zkoumané případy), v němž jsme ho provedli, má vysokou **externí validitu** (a naopak)

# Dekonstrukce interní validity

- Interní validita má několik komponent:
  - **Statistickou**
  - **Kauzální**
  - **Konstruktovou**

# Statistická validita

- Zjišťuje, zda existuje statisticky významný vztah (kovariance) mezi proměnnými, které výzkumníka zajímají a zda lze určit jeho velikost
- Jde o efektivní a přesné použití statistických nástrojů (statistické usuzování- například se předpokládá něco o distribuci hodnot proměnných nebo o vztahu proměnných a podle toho se používá statistika), posouzení statistické významnosti a síly vztahu.
- Manski: SV je o tom, jak velikost a variabilita vzorku ovlivňuje závěry, které můžeme udělat o populaci.



# Kauzální validita

Jde o určení toho, zda variance, nalezená v datech, má kauzální charakter.

Prakticky to znamená určit (a být si jist), že změna v  $T$  způsobuje  $Y$ .

Nejde o to, jak velký/silný je vztah (to je statistická validita), ale o **identifikaci proměnných**, které se na něm podílí („4 kauzální překážky“).

„Skutečně je naše nezávislá proměnná nezávislá“?

# Konstruktová validita

- Hodnocení toho, jak validní je pozorování/data pro teorii, k níž je vztaženo.
- Širší než kauzální validita, nejde jen o to, zda mezi T a Y je kauzální vztah, ale o to, zda tento vztah jde dobře (validně) uplatnit i na T a Y v již existující teorii (z níž jsme vyšli).
- „Operacionalizovali jsme proměnné (případně převzali z nějakého předchozího výzkumu) tak, že teď naše zjištění dobře můžeme vztáhnout k teorii?“

# Externí validita

- Úzce souvisí s koncepty „**robustnosti**“ **výsledku** a „**vědecké replikace**“.
- Zajímá se o to, jestli výsledek, který jsme získali na nějaké populaci, je snadno přenositelný i na populaci jinou („robustnost výsledku“).
- Vědecká replikace odkazuje k situaci, kdy (např. v teoretickém vakuu) buďto **opakujeme** náš **výzkum** na **jiné populaci**, abychom ověřili naše původní zjištění nebo **rozšíříme teorii** o další předpoklady na **stejně populaci**.
- „Platí naše zjištění i v jiném kontextu, než jsme ho zjistili?“ „Pokud ne, proč“?