

# VALIDITA A REPLIKACE

POL 565 11. 10. 2016

# Co je validita výzkumu?

- K čemu to je?
- Jak ji hodnotíme?
- Kdo ji posuzuje?

# Validita

- Co jsme se dozvěděli na základě našich dat a jak moc tomu můžeme věřit?
  - Výsledky, které platí (tedy jsou pravdivé) pro danou populaci a které jsou natolik robustní, že je lze zobecnit i mimo studovanou populaci.
- **Validita** je přibližná **pravdivost** vědecké inference nebo tvrzení (Morton a Williams 2010).
- K čemu ale vztahujeme tu pravdivost?
- Dělení na **INTERNÍ** a **EXTERNÍ** validitu (Campbell 1957)

# Jak chápat validitu?

- Interní validita:
  - Přibližná pravdivost inference nebo tvrzení **v rámci cílové populace**
- Externí validita:
  - Přibližná pravdivost inference nebo tvrzení pro pozorování **mimo cílovou populaci**, která je předmětem výzkumu.
- Jaká je logika validity experimentálního výzkumu v tomto tradičním dělení validity?
  - Pomůcka: Jaké jsou silné a slabé stránky experimentu?

# Validita (Cronbach 198)

- Kauzální validita
- Statistická validita
- Konstruktová validita
- Externí validita

# Statistická validita

- *Udává, zda existuje statisticky významná kovariance mezi proměnnými a zda je tento vztah dostatečně silný.*
- Otázka statistické analýzy dat.
- Je vztah signifikantní? Co když je na hranici významnosti 5 %?
- Jsou splněny základní statistické předpoklady (distribuce proměnných, homogenita rozptylu atd.), je vhodné použít neparametrické testy?

# Statistická validita

- Statistická replikace jako metoda ověřování statistické validity
- Jiný vzorek dané populace (při zachování konstruktové validity)
- Nebo stejný vzorek a jiné statistické nástroje (při zachování konstruktové validity)
- Problémy se statistickou validitou objeveny v psychologickém výzkumu
- Program StatCheck replikuje analýzu, řada nesrovnalostí a chyb ve většině z testovaných 50 tis. článků (Nijuten a Hartgerink. 2015)

# Statistická validita

- V některých případech statistické replikace odhaleny podvody
- Kauza LaCour a umělé generování dat ve field experimentu v Californii (LaCour a Green 2014)



# Kauzální validita

- Typicky považována za silnou stránku experimentu
- Srovnání s observačním výzkumem?
- Udává, zda vztahy, které výzkumník identifikuje v dané populaci, jsou kauzální.
- Udává, zda změna v nezávislé proměnné přináší změnu v závislé proměnné.
- Zda lze vyloučit kauzální vztah mezi nezávislou proměnnou a ostatními faktory

# Konstruktová validita

- Udává, zda jsou inference odvozené z dat platné pro danou teorii (teoretické konstrukty), která je testována.
- Typicky zaměňována za “interní validitu”
- Jsou proměnné sledované designem dobrou reprezentací proměnných implikovaných v teorii?
- Jsou ostatní proměnné konstantní?
- Mají subjekty informace, které předpokládá teorie?
- Mají subjekty stejnou možnost behaviorální reakce, jakou předpokládá teorie?
- Jsou měřící nástroje vhodné?

# Konstruktová validita ve vztahu k externí validitě

- Pro zobecnění je konstruktová validita klíčová
- Pokud máme pozitivní výsledky ve studii s vysokou konstruktovou validitou, je to dobrý základ pro teoretické implikace pro JINÉ POPULACI
- Pokud máme negativní výsledky ve studii s vysokou konstruktovou validitou na jedné populaci, může vzniknout nový teoretický předpoklad (vztahující se opět k jiné populaci než v originální studii)

# Hrozby interní validity

- Treatment neovlivňuje subjekty podle teoretického předpokladu
  - Např. nízká pozornost, nedůvěra
  - Neschopnost vyvolat zkoumaný fenomén (př. Sociální vyloučení, strach atd.)
  - Ohrožena konstruktová i kauzální validita
- Úmrtnost v experimentu – pokud k ní dojde po náhodném přiřazení k treatmentu
- Nedodání treatmentu hrozí především u field experimentů (kauzální validita)

# Hrozby IV

- Správná operacionalizace proměnných
- Nahrazování klíčových proměnných lépe měřitelnými koncepty
- Experimentátor nemanipuluje treatment (kvaziexperimenty, přírodní experimenty)
- Non-compliance, nedostatečné testování
- U psychologických experimentů vadí, když subjekty uhodnou cíl experimentu
- Sociální desirabilita

# Externí validita

- Zobecnitelnost mimo zkoumanou populaci
- Teorie jsou obecně platné. Jsou obecně platné i výsledky experimentu?
- Odpovídají exp. subjekty skutečnému světu?  
Odpovídá treatment stimulům v běžném světě?  
Odpovídá experimentální kontext kontextu v reálném světě?
- Achillova pata experimentálního výzkumu?
- Politické vědy jsou posedlé externí validitou (Rose McDermott 2010).
- Často zaměňována za EKOLOGICKOU VALIDITU: otázka prostředí, odpovídá zkoumanému prostředí IRL?

# Externí validita

- Externí validita musí být ustavena na základě empirických důkazů
- EV je otázkou **VĚDECKÉ REPLIKACE** k testování stejných teoretických konstruktů
  - Na nové populaci
  - Na novém vzorku
  - Za použití jiného nastavení
  - Za specifikace environmentálních podmínek
  - Stress test: změna v experimentální protokolu, aniž bychom měli teoretickou predikci, co bude znamenat. Pokud předpoklad původní teorie neplatí, není teorie robustní.

# Příklad: Druckman, McDermott 2008. Replikace Kahnemana a Tverskyho

*Představte si, že USA se připravuje na epidemie neobvyklé asijské nemoci, na kterou dle očekávání zemře 600 lidí. Zvažují se dva alternativní programy pro boj s nemocí. Který program zvolíte:*

A: Pokud bude přijat, bude zachráněno 200 lidí

B: Pokud bude přijat, s 1/3 pravděpodobností bude 600 lidí zachráněno a se 2/3 pravděpodobností nikdo nebude zachráněn.

Vs.

A: Pokud bude přijat, zemře 400 lidí.

B: Pokud bude přijat, s 1/3 pravděpodobností nikdo nezemře a se 2/3 pravděpodobností 600 lidí zemře.



# Příklad: Druckman, McDermott 2008

- Představte si, že komunita, ve které bydlíte, získala 3000 \$ z vládního grantu na rozvoj. Peníze budou muset být okamžitě investovány do jednoho ze dvou programů.
- Y: Když bude přijat, komunita získá 1000\$.
- Z: Když bude přijat, s 50% šancí komunita získá 2000\$ a s 50% nezíská nic.

Vs.

- Y: Když bude přijat, komunita ztratí 2000\$.
- Z: Když bude přijat, s 50% šancí komunita ztratí 2000\$ a s 50% neztratí nic.

# Externí validita – hrozby a slabiny

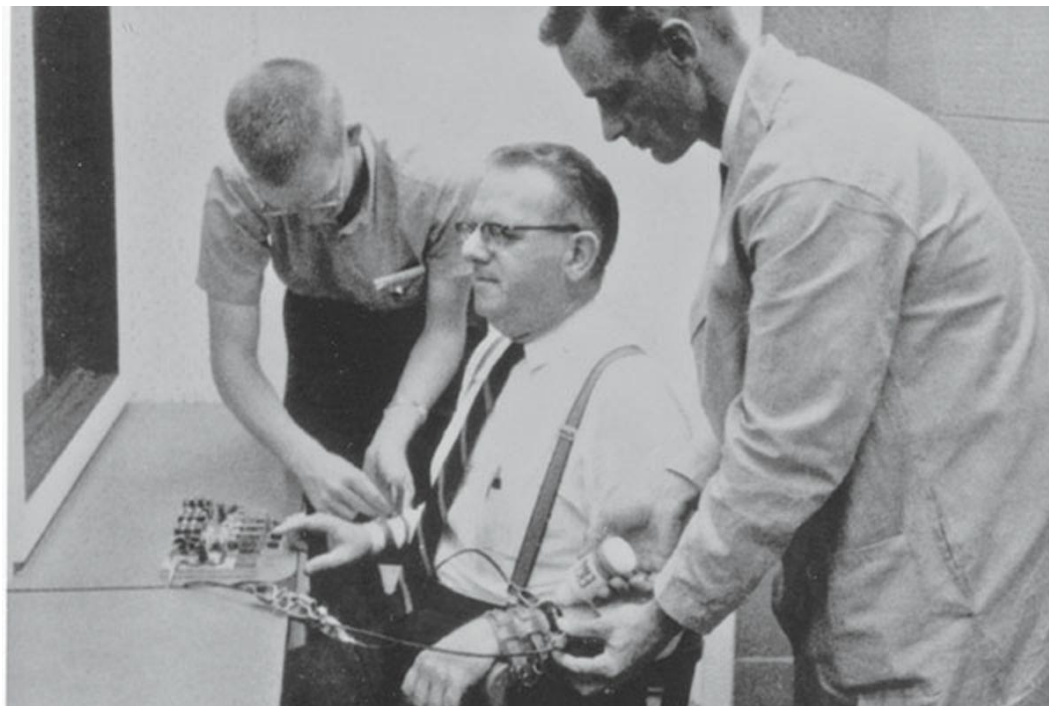
- Ekologická validita
  - Odpovídá experimentální prostředí skutečnému prostředí?
  - Vysoký stupeň umělosti! Především v laboratoři.
  - Vysoká míra kontroly prostředí implikuje jeho umělost.
  - Jak je na tom fMRI?
  - Důraz na **věcný realismus**

# Věcný vs. experimentální realismus

- Experimentální realismus:
  - Dokáže experiment vyvolat v subjektech ty procesy, které teorie předpokládá?
  - Jsou subjekty dostatečně zaujaté?
  - Dokáže zamýšlená manipulace vyvolat to, co chceme?
  - Jde o psychologickou zkušenost subjektů
  - Pokud není, co se stane? Co to znamená pro validitu?

# Věcný vs. experimentální realismus

- Jaká je validita Milgramových experimentů?



# Věcný vs. experimentální realismus

- Wasler, Aronson a Abrahams 1966:
  - subjekty četly noviny, ve kterých byl článek o trestním stíhání v Portugalsku
  - Framing experiment
  - Vysoký věcný realismus, čtení novin odráží skutečný život
  - Nízký experimentální realismus: studenti v experimentu nebyli ovlivněni manipulací

# Vzorky

- Lze zobecnit výsledek na základě dat nenáhodného vzorku?
- Nejčastější nenáhodný vzorek jsou studenti.
- Vidíte nějaký potenciální problém?

# Studentské vzorky

- Sears 1986: studentské vzorky jsou příliš úzká datová základna
- Systematicky se liší od celkové populace
- Slabě vyvinuté identity a sebe-pochopení
- Slabé a nekonsistentní postoje
- Snadno ovlivnitelní vnějšími faktory
- Spory o to, jak se studentské vzorky liší od celkové populace.

# Studentské vzorky

- Miller a Krosnick 2000: studenti mají menší zkušenosti s politickými kampaněmi, liší se efekty (priming experimenty)
- Druckman a Kam (2011): studenti se neliší v klíčových proměnných
  - Stranická identita, ideologie, politické postoje na témata imigrace či homosexuality, zájem o politiku, konzumace médií



# Studentské vzorky

- Je pro validitu experimentálního výzkumu důležité, jak moc studenti odpovídají parametrům celé populace?
- Ne nezbytně!
- Cílem experimentu **není** vyvození obecných závěrů.
- Cílem experimentu je **testování teorie**.
- Externí validita je otázkou REPLIKACE

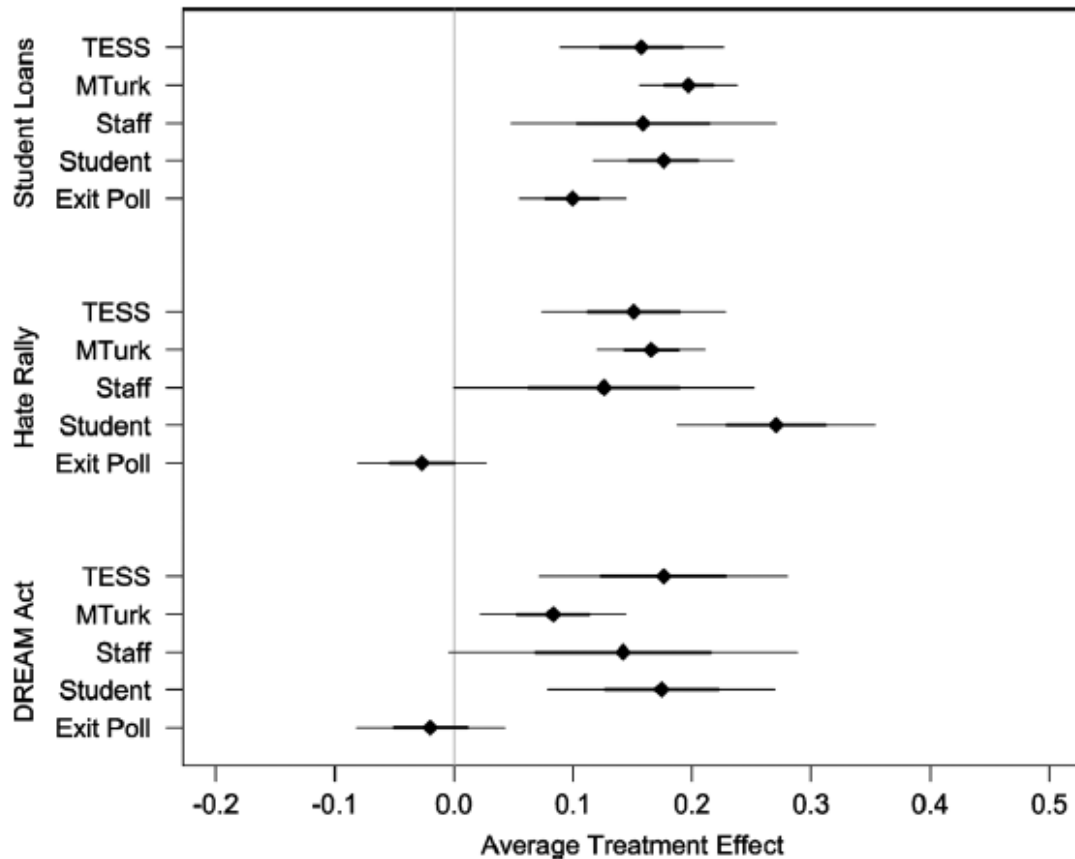
# Online “pracovníci”

- Amazon Mechanical Turk
  - Online crowdsourcing platforma
  - Profesionální subjekty
  - Nereprezentativní vzorky
  - Vysoká dostupnost
  - Jak moc jsou to kvalitní vzorky?

# Amazon Mturk

- Mullinix et al.: porovnání reprezentativního vzorku, studentů, zaměstnanců univerzity, lidí odchytnutých ve volební místnosti (volby 2012)
- Replikace 3 framing experimentů: studentské půjčky, hate rally, imigrace,
- Srovnatelný směr i velikost efektů
- Replikace 20 experimentů s původně reprezentativní vzorky v Mturk.
- Srovnatelný směr i velikost efektů

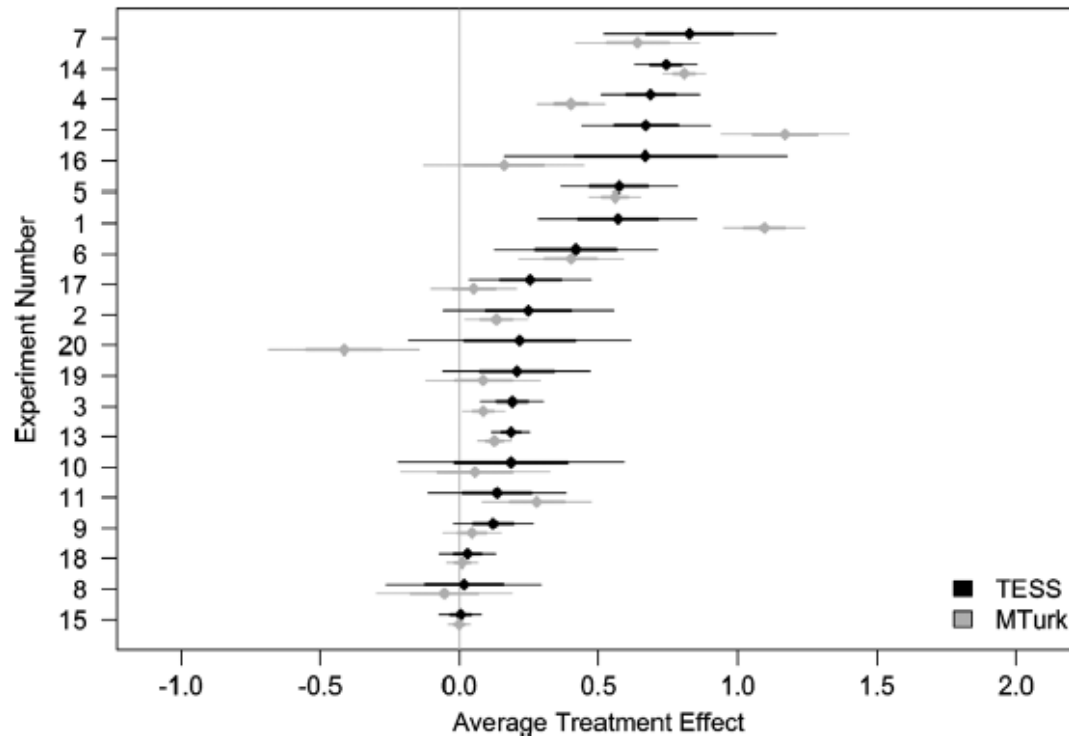
# Mullinix et al. 2016



*Figure 1*  
Study 1 Results.

*Note:* points are average treatment effects (difference between group means), and bars representing one and two standard errors for the mean-difference.

# Mullinix et al. 2016



*Figure 2*

## Control vs. Treatment Group 1.

*Note:* points are average treatment effects (difference between control and treatment group means), and bars representing one and two standard errors for the mean-difference. Many of the experiments have multiple treatment groups. This figure focuses on the first treatment group.

# Google Consumer Survey

- Kratší design (max 10 otázek)
- Krátké otázky
- Automatický sběr demografických dat přes Google
- Santoso, Stein a Stevenson 2016: replikace 3 (klasických) experimentů
  - Rámování sociálních dávek jako pomoci chudým
  - Asijská nemoc
  - List experiment
- Systematicky menší efekty: důsledek nízké pozornosti ze strany subjektů

# Nereprezentativní vzorky

- Nevíme, jak se vztahují k cílové populaci teorie
- Heterogenita treatmentu, ne na všechny působí stejně, můžeme mít vzorek, ve kterém absentují nebo naopak převažují zprostředkující proměnné
  - Někdy lze odhadnout, změřit a zahrnout do analýzy
  - Je nutné uvážit, zda je daný vzorek vhodný (např. u studentů a ideologií)
- Hrozba nízké pozornosti u online vzorků: ohrožen experimentální realismus, konstruktová validita
- Řešení? Lze ji měřit?
  - Indikátorem je latence odpovědí
- Změna subjektů v čase, souvisí s “profesionalizací” (především online vzorky)

Je některý typ validity důležitější?  
Máme něco upřednostnit?



# Interní validita vs. Externí validita?

- Větší Interní validita vede k nižší externí validitě (laboratoř)
- Větší externí validita vede k nižší interní validitě (field)
- Je možné interní validity brát jako nutnou podmínku externí validity.
  - *Můžeme zobecňovat mimo zkoumanou populaci, pokud naše výsledky nejsou dostatečně pravdivé v rámci zkoumané populace?*

# Replikace

- Nutný předpoklad ověřování validity a robustnosti v experimentálním výzkumu
- Znovuvytvoření podmínek pro reprodukci výsledků na nových datech,
- Známé i neznámé faktory mohou moderovat rozdíl ve velikosti efektů
- Neschopnost replikace mohou značit false positive originálního výzkumu nebo false negative replikace
- Důraz na transparentnost, zveřejnění experimentálního protokolu i datových matic (často podmínka publikace)

# Replikační krize: problémy experimentálního výzkumu

- Netransparentnost
- Publikační bias
- P-Hacking (researcher degree of freedom)
  - Ukončení sběru dat když  $p < .05$
  - Analýza mnoha vztahů, reportování jen těch, kde  $p < .05$
  - Existence mnoha podmínek, reportování je těch, kde  $p < .05$
  - Použití kovariátů k dosažení  $p < .05$
  - Vyřazení participantů k dosažení  $p < .05$
  - Transformace dat k dosažení  $p < .05$
- Nedostatečné specifikace podmínek

# Replikační krize?

- Open Science Collaboration 2015 (270 autorů):
  - Replikace 100 psychologických experimentů
  - Snaha reprodukovat originální efekty (effect size)
  - Pozitivní výsledek: 35 případů
  - Reprodukce velikost efektu (v 95% CI efektu replikovaného výsledku) – úspěšná v 47,4 %
  - V 82 případech byly původní velikosti efektu větší
  - Neznamená to neplatnost teorií, v originálním výzkumu nemuselo jít nutně o false positive (mohou působit neočekávané nepozorované faktory), publikační bias
  - Problém není omezen jen na psychologii

# Replikační krize?

- Gilbert et al.: původní replikační projekt nedokázal větně replikovat původní experimenty
- Popisují rozdíly mezi šesti původními experimenty a jejich replikacemi
- Zpochybnění replikačního přístupu
- Debata pokračuje, Susan Fiske obviňuje psychologickou obec z “metodologického terorismu”

# Důsledky?

- Některé teorie odvolány.
- D. Carney, spoluautorka vlivné a populární studie o POWER POSING (Carney, Cuddy, Yap 2010):

*“As evidence has come in over these past 2+ years, my views have updated to reflect the evidence. As such, I do not believe that “power pose” effects are real.”*

