VALIDITA A PROBLÉMY USUZOVÁNÍ V EXPERIMENTÁLNÍCH A OBSERVAČNÍCH STUDIÍCH

POL 494, 4.5. 2016

Validita jako „aproximace k pravdě“ (Shadish et. al. 2002)

- **Validita (v emprickém výzkumu): naše vědění (získané inferencí) je přibližně pravdivé.**

Komponenty validity: **externí a interní validita**

- **Interní validita**: přibližná pravda inference nebo získaného vědění v rámci populace, kterou jsme studovali („jak dobře jsme zkoumali, co nás zajímá“)

- **Externí validita**: přibližná pravda inference nebo získaného vědění mimo studovanou populaci. („jestli to, co jsme zjistili, lze automaticky vztáhnout i mimo zkoumanou populaci“).
- Stačí k porozumění konceptu validity? Ne

Dekonstrukce interní validity

- Interní validita má několik komponent:

- **Statistickou**
- **Kauzální**
- **Konstruktovou**

Statistická validita

- Obecně 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.
- Při studiu kauzality zjišťuje, zda existuje statisticky významný vztah (kovariance) mezi proměnnými, které výzkumníka zajímají a zda lze určit jeho velikost
- Manski: SV je o tom, jak velikost a variabilita vzorku ovlivňuje závěry, které můžeme udělat o populaci.

Diskuse o statistické významnosti

- Statistická významnost= pravděpodobnost, že náš vybraný vzorek pochází ze souboru, ve kterém platí nulová hypotéza.
- Konvence většinou je 0.05, chceme nižší, abychom řekli, že náš výsledek je statisticky významný
- „Vyšší statistická významnost neznamená větší důležitost výsledku“
- Důležitá i **velikost** efektu

Př.

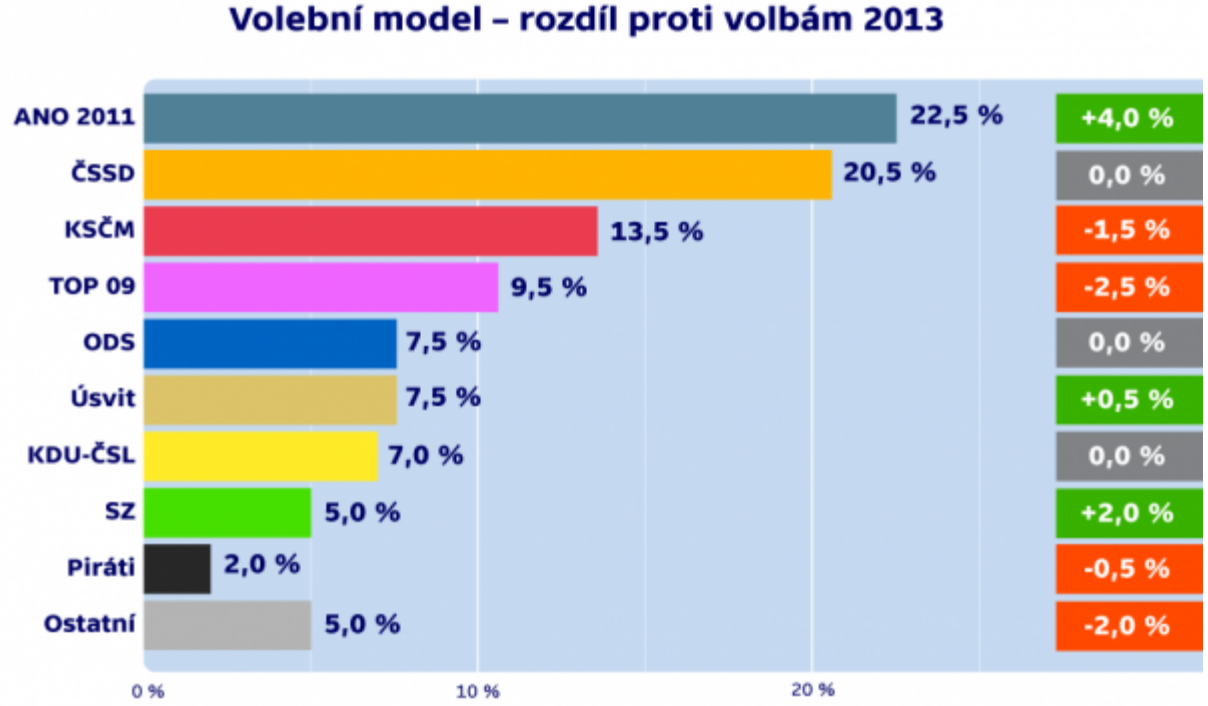
- 500 kandidátů si v komunálních volbách si dělá personalizovanou kampaň, 500 ne
- Chceme zjistit, zda personalizace kampaně má vliv na jejich volební výsledek
- Zjistíme, že mezi skupinami jsou **statisticky významné** rozdíly
- Skupina s personalizovanou kampaní získala průměrně o 0,4% hlasů více.
- Věcná významnost **sporná**.

- VS.

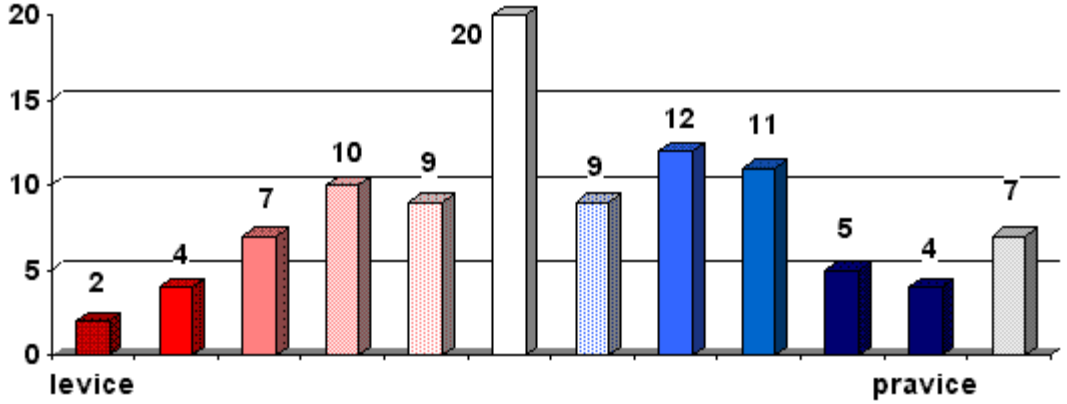
- 15 kandidátů si v komunálních volbách si dělá personalizovanou kampaň, 15 ne
- Chceme zjistit, zda personalizace kampaně má vliv na jejich volební výsledek
- Zjistíme, že mezi skupinami nejsou **statisticky významné** rozdíly
- Skupina s personalizovanou kampaní získala průměrně o 3,5% hlasů více.
- Věcná významnost potenciálně vysoká, statistickou validitu **narušuje malé**

Příklad statistické validity: průzkumy veřejného mínění

- TNS Aisa 2/2014 (vzhledem ke statistické chybě lze z dat- srovnání 2/2014 a voleb- usuzovat velmi málo).

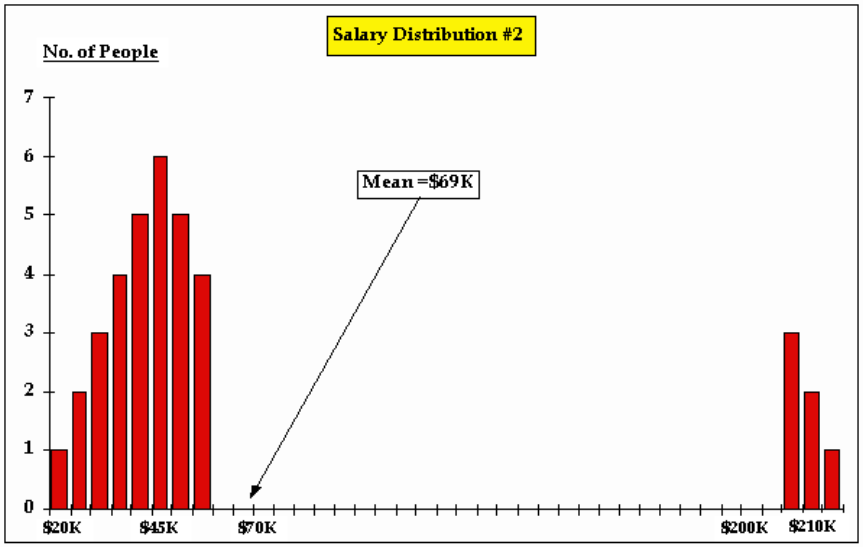


Př. Statistické validity: distribuce hodnot (vhodná statistika)



Viceméně normální rozložení

U dat obvykle předpokládáme normální distribuci, pokud není, musíme tomu přizpůsobit statistické nástroje, pomocí kterých je analyzujeme



Např. zde není průměr vhodnou statistikou, zajímá-li nás centrální tendence v populaci

Statistická replikace

- Prostředek zvyšování statistické validity: buďto se použije jiný vzorek ze stejné populace nebo jiná statistická metoda u stejného vzorku (v DP a observačních studiích to neděláme- časově a finančně nákladné, v experimentech často ano).

- Souvisí s tzv. **replikační krizí** v sociálních vědách

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 je vztah (to je statistická validita), ale o **identifikaci proměnných**, které se na něm podílí.

„Čtyři kauzální překážky“ (viz předchozí přednášky**).**

Příklady problematické kauzální validity

- **„Budu zkoumat, jak míra profesionalizace kampaně ovlivňuje podporu strany“** (čtvrtá překážka).

- **„Zjistila jsem v datech, že pohlaví ovlivňuje ideologickou orientaci“ (**první překážka, pokud nedoplníme o přesvědčivý mechanismus a konstatujeme jen na základě dat).

- **„Data sice navrhovaný vztah zatím nepotvrdila, ale mechanismus mezi závislou a nezávislou proměnnou je**

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).

Příklady otázek, které si klademe u konstruktové validity (MW 2010:261)

- Jsou možné volby/úkoly subjektů (např. to, na co odpovídají v dotazníku, jak se rozhodují v experimentu) stejné jako v teorii? Získáme pak data, která bezpečně můžeme vztáhnout k teoretickým konceptům?
- Jsou proměnné, které považuje teorie za konstantní, i konstantní ve výzkumu?
- Je rozsah informace dostupný aktérům obdobný, jak to předpokládá teorie?
- Jsou subjekty ze stejné populace jako předpokládá teorie?

- **Př: Stouffer (minulá hodina, špatná**

Hrozby interní validitě (Campbell-Stanley 1966)

- Selekce
- Historie
- Zrání
- Opakované testování
- Instrumentace (nedokonalé instrumentální proměnné)
- Regrese k průměru
- Úmrtnost (subjektů)
- Zkreslení vyvolané pozorovatelem (obtrusivní techniky)

Externí validita

- Úzce souvisí s koncepty „robustnosti“ výsledku a „**vědecké replikace**“.

- 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 **stejné populaci**.

JAK UVAŽOVAT O VALIDITĚ: VÝZKUM POLITICKÝCH KAMPANÍ

Kampaň jako „sebeposilující se proces

- **Vícenásobná ekvilibria** (procesy s mnoha možnými stabilními výsledky)
- **Nahodilost** (i malé události, pokud se přihodí se ve správný moment, mohou mít dalekosáhlé následky)
- **Načasování a pořadí** (záleží na načasování a na pořadí: Tilly: „Pokud analyzujeme sérií událostí, pak to, kdy se stanou, ovlivňuje jak se – další – stanou“).
- **Rezistence:** v momentě, kdy je nějaký proces zřetelně nastartován („kritický zlom“), začíná být odolný ke změně

Jak uvažovat o politických událostech: kdy a jak rychle se projeví příčina a následek?

		Efekt	
		Rychlý	Pomalý
Příčina	Rychlá	tornádo	meteorit
	Pomalá	zemětřesení	globální oteplování

Zdroj: Pierson 2006:81

Abbot: „Podobně důležitým kauzálním událostem netrvá stejnou dobu, aby se staly“

- Tornádo: rychlá (okamžitě se demonstrující) příčina i následek (politický skandál).
- Zemětřesení: pomalá příčina (demografická změna) rychlý efekt (revoluce), tzv. **threshold effect nebo** řada navazujících příčin ústících v dramatický efekt – **causal chain**
- Meteorit: rychlá příčina (přijatá politika), pomalý efekt (příklad: penzijní reforma). – **cummulative effect**
- Globální oteplování: pomalá příčina (gramotnost, deurbanizace) a pomalý efekt (stranická identifikace), tzv. **cummulative causes**

Nejčastější analýza politiky: „Tornádo“ (podporují média):

příčina → následek (efekt)

U sebeposilujících se procesů („Zemětřesení“, „Meteorit“, „Globální oteplování“):

příčina → následek (efekt) → následek

Politické (volební) kampaně: jak je pojímat (jsme si jisti, jak rychle se dějí)?

1. Jako tornádo
2. Jako sebeposilující se událost: *Campaigns can be graduated from those that barely exist to those that consume parties, voters and the media(Brady- Johnston 2006:3).*

Z hlediska časových sekvencí jsou nejčastěji pojímány jako „threshold effect“ nebo „causal chain“, nikdy ale nebylo prokázáno, že i efekty se projevují rychle (a ne pomalu).

Liší se v různých typech voleb (PR vs. většinové systémy, jeden kandidát vs. vícenásobná volba)

Politické kampaně: co na nich zkoumat: Příčiny efektů vs. efekty příčin

- Pokud se např. ptáme, co (všechno) ovlivňuje volební účast, zkoumáme **příčiny efektů**

- Pokud se např. ptáme, jaký vliv má informovanost na volební účast, zkoumáme **efekt** (jedné) **příčiny**

Techniky zkoumání efektů v kampaních

- **Ideál**: lze uplatnit na **různě dlouhé** kauzální procesy

- 1. EXPERIMENT
- 2. OBSERVAČNÍ STUDIE
- 2A Panelové šetření
- 2B Průřezová studie (rolling cross section)

Experiment (Ansolabehere-Iyengar 1995)

- Hlavní charakteristika: intervence do „DGP“.
- Příčiny/Stimuly: Informace
- Efekty: Deklarované/skutečné chování
- Hlavní výhoda: **kontrola je síla** (volba podnětu/ů, vystavení osob podnětu, laboratorní c polní experimenty).

Velikost efektu1 Velikost efektu2

Skupina1 Skupina2

..

Podnět 1 Podnět 2

Význam využití experimentu v kampaních

- Zjistit, co má potenciál způsobit „efekt“ (kritický zlom)
- Poskytuje návody k redukci komplexity problému, dosavadních orientací recipienta
- Testuje (možnou) selektivitu percepce vystavení podnětu
- Campaigns and Elections (2014): ***Experiments have been at the core of the progressive revolution in modeling, targeting, persuasion and turnout over the past decade.***

http://www.campaignsandelections.com/magazine/us-edition/428067/vaccines-vs-leeches.thtml

Problém: vychýlenost směrem k výzkumu typu „tornádo“ (nevíme, jak rychle se IRL projeví)

Příklad experimentu: (Gerber, Kaplan, Bergan 2007: *Newspaper bias* a identifikace se stranou/kandidátem)

Volby guvernéra ve Virginii

Populace: Osoby, které neodebíraly žádné noviny (kontrolní proměnné věk, pohlaví, stranická identifikace), 3 skupiny (celkem 3000+ zúčastněných)

Intervence: jedna skupina dostávala zadarmo Washington Post (liberální), druhá Washington Times (konzervativní), třetí žádné po dobu 2 měsíců před volbami

Zjištění: Preference pro Rep a Dem kandidáta se mezi skupinou 1 a 2 lišila o 8%, lišil se i postoj k tématům kampaně, méně významně. Tón novin má efekt na výsledek kampaně.

Výhody x nevýhody zvoleného postupu?

Hrozby externí validitě (McDermott 2011)

- Umělost experimentu
- Úmrtnost, Nenáhodné výběry (zde vztažená ke generalizování)
- Hawthornův efekt (efekt morčete)
- Vliv proměnných, které nejsou v laboratoři přítomné
- Různá „síla“ stimulu IRL a v laboratoři

Panelová šetření

Stejní respondenti

Postoje čas 1 → Postoje čas 2

↑

„Efekt“

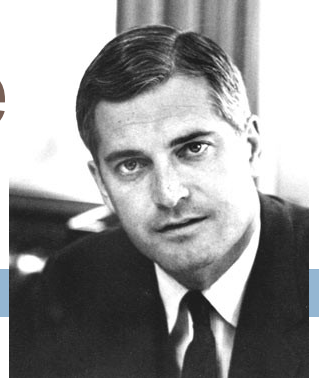
Výhody panelu

- Přináší do výzkumu čas, „prodlužuje“ zkoumaný proces, umožňuje zavádět do výzkumu „předem“ naměřené vysvětlující proměnné
- Postoje v Čase 1 (vyvolané např. příčinami, předcházejícími kampani) se stávají jednou z vysvětlujících proměnných pro postoje v Čase 2
- Snadný odhad chyby měření
- Používán v prvních amerických studiích prezidentských kampaní

Nevýhody panelu

- **Nástroj**: Změny v postoji v čase 2 jsou vyvolány tázáním v čase 1
- **Nákladnost, „úmrtnost“**
- **Hrubost informace** (kampaně jsou události s řadou –potenciálních- efektů, je obtížné připsat změnu postoje **jednomu** konkrétnímu)
- Nevíme, **kdy** efekt měřit.

Příklad: Vnímání Johna Turne kanadských volbách 1988



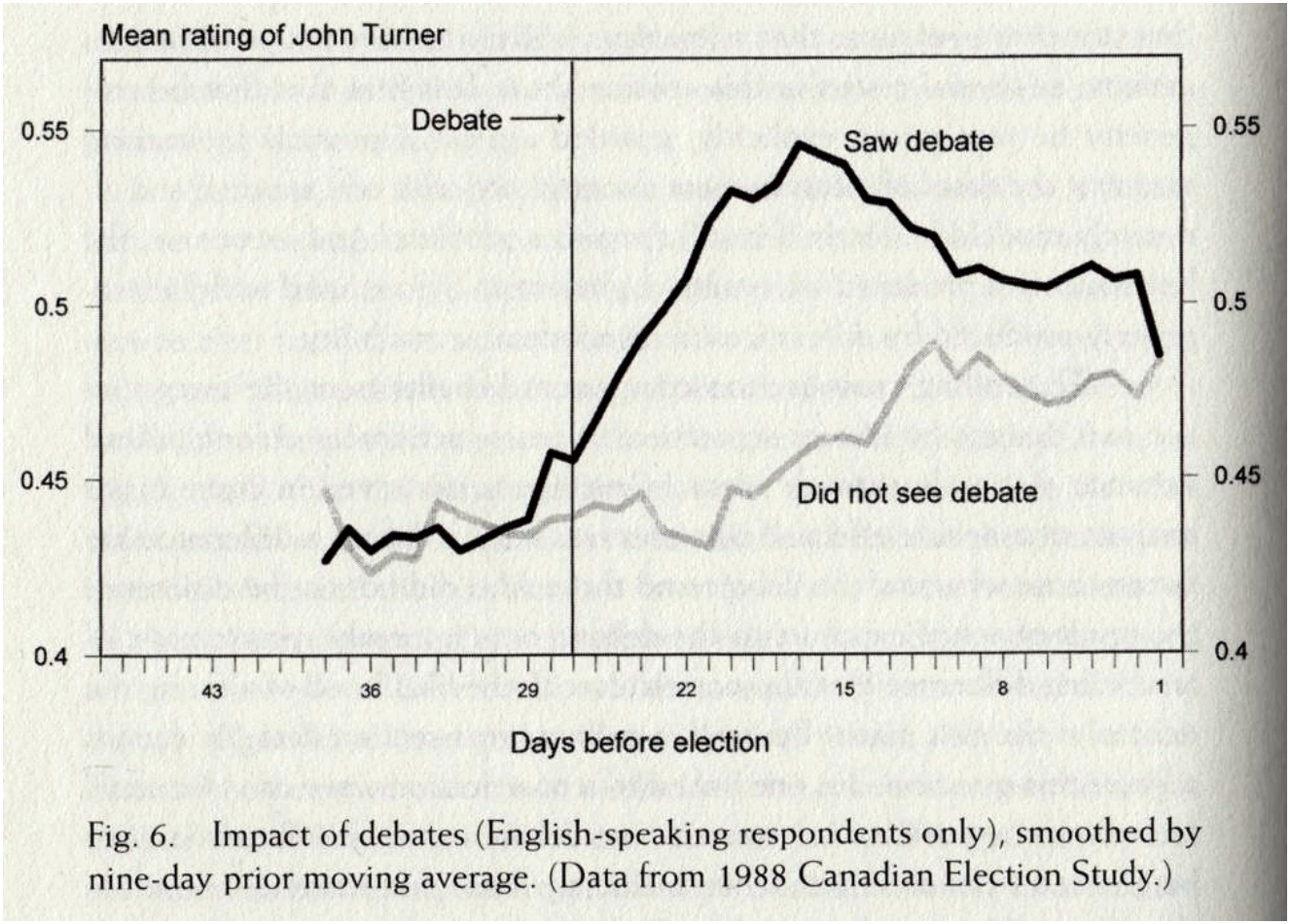
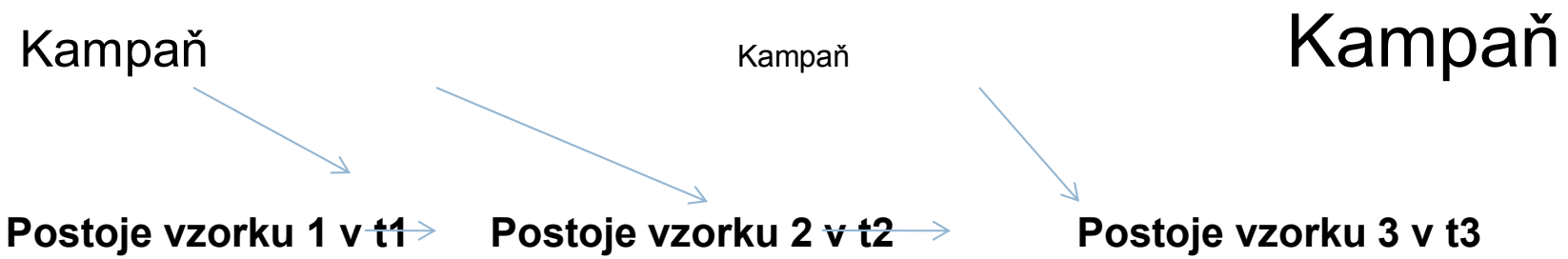


Fig. 6. Impact of debates (English-speaking respondents only), smoothed by nine-day prior moving average. (Data from 1988 Canadian Election Study.)

Graf zachycuje hodnocení J. Turnera se speciálním důrazem na hodnocení těch, kdo viděli a neviděli debatu. Jeho interpretace je složitá, hodnocení JT u těch, kdo ji sledovali, se zvyšovalo již před ní. Po ní nejprve ti, kdo debatu nesledovali, hodnotili JT hůře, ale následně i oni lépe. Bylo to proto, že se dozvěděli o jeho dobrém výkonu později? Nebo se stala nějaká jiná, do výzkumu nezahrnutá událost, jež způsobila, že JT začaly obě původní skupiny hodnotit lépe? Naznačené otázky ukazují problémy se studiem kauzality v kampaních, říci že

Tento příklad demonstruje zásadní problémy se 4. kauzální překážkou, typické pro studium politických procesů

Postupující průřezová studie *(rolling cross-section study)*



Design:

Existuje cílová velikost vzorku, např. N=10000, ta je doplňována v pravidelných intervalech o určitý počet rozhovorů, v daném intervalu se vždy začíná rozhovory, které se nepovedlo dokončit v minulých intervalech

Používáno např. v US prezidentské kampani 2008.

Způsob konstrukce vzorku (opakuje se v pravidelné periodě).

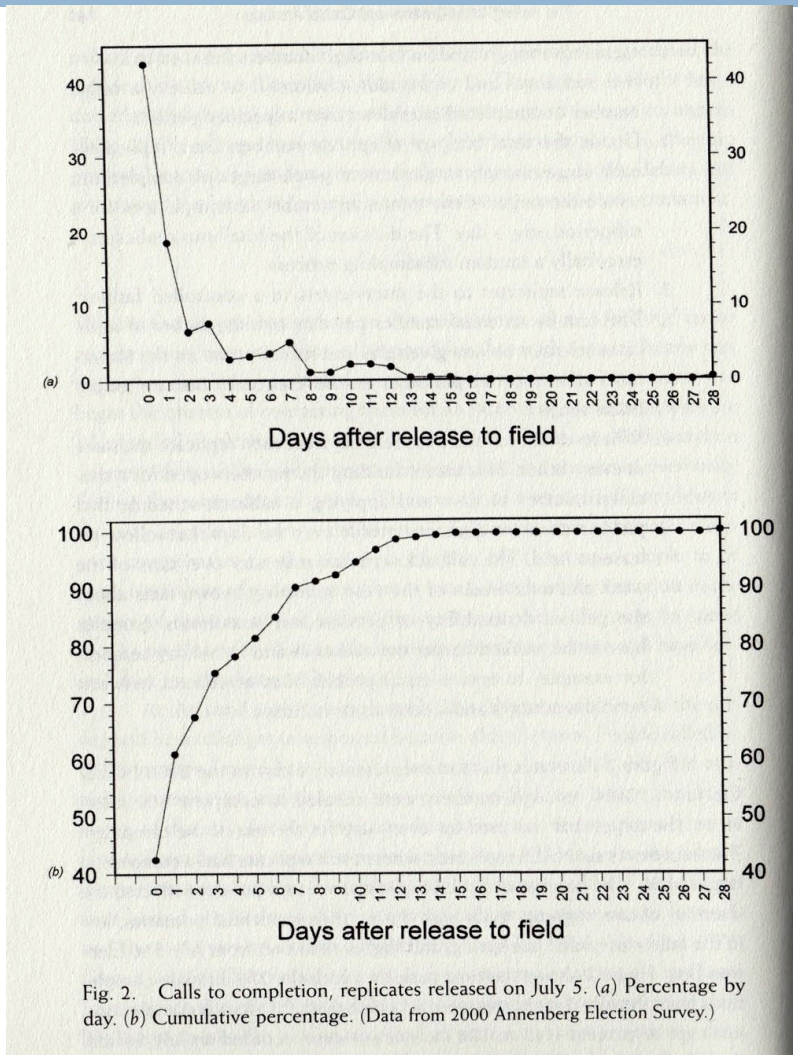


Fig. 2. Calls to completion, replicates released on July 5. (*a*) Percentage by day. (*b*) Cumulative percentage. (Data from 2000 Annenberg Election Survey.)

Srovnání/výhody RCS: panel vs. cross-section: Gore v kampani 2000

Zdroj: Annenbergská studie 2000

Zatímco na základě panelového šetření by se zdálo, že debaty přispěly zásadním způsobem ke snížení hodnocení upřímnosti AG v kampani, RCS ukazuje, že věc je složitější, např. před první debatou již toto hodnocení prudce klesalo a po ní se dokonce mírně zvýšilo. Z hlediska šancí na izolaci/měření některých proměnných v kampani tak lze RCS považovat za asi lepší nástroj než panel.

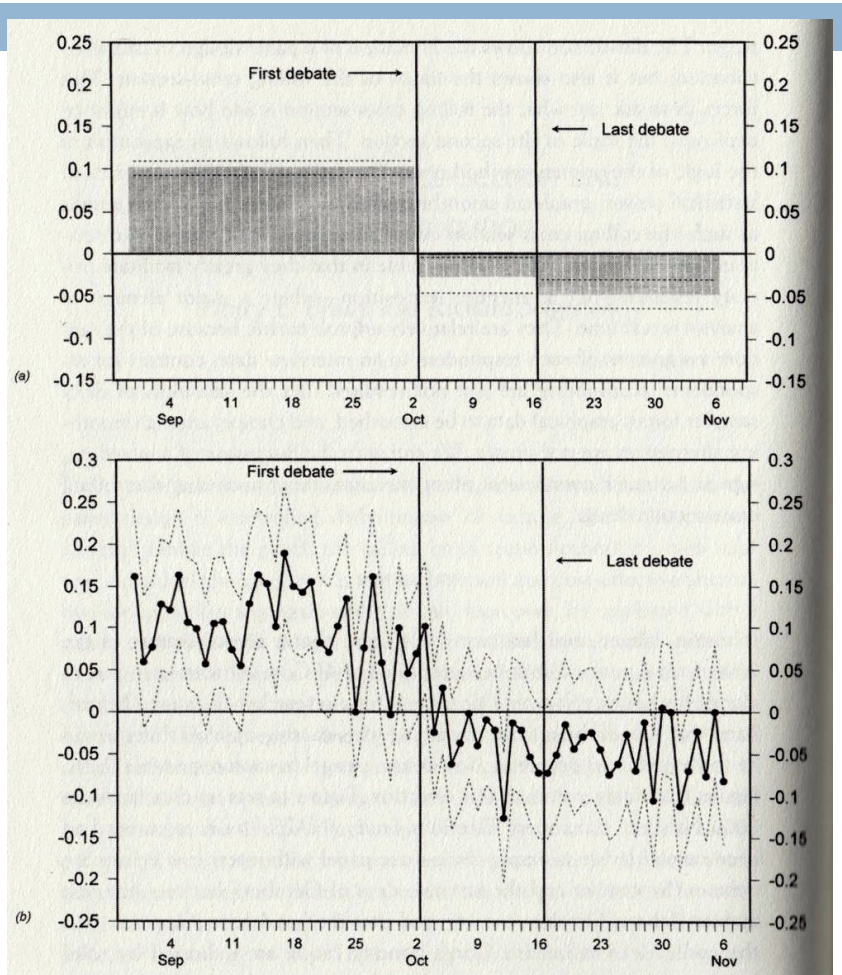


Fig. 1. Debates and perceptions of Al Gore's honesty. (*a*) Pre-post means; whole-period estimates; dashed lines are approximate 95 percent confidence interval. (*b*) Daily means; daily estimates; dashed lines are approximate 95 percent confidence interval. (Data from 2000 Annenberg Election Survey.)

Nevýhody/problémy

- Nejistota ohledně toho, jestli jsou efekty kampaně „quickly unfolding“ nebo „slowly moving“ a jak je měřit
- Malé velikosti denních vzorků (trendy vs. „sampling noise“)
- Nutnost sledovat kontext

Observační studie vs. experimenty a studium příčin a následků- shrnutí.

- Vysoká interní validita experimentů vs. sporná interní validita observačních studií („téma kontroly“).
- Vyšší externí validita observačních studií vs. nižší externí validita experimentů („téma generalizace“).

Literatura

- Pierson 2006: Politics in Time. History, Institutions and Social Analysis. Princeton University Press
- Brady&Johnston 2006: Capturing Campaign Effects. Uof Michigan Press
- Morton&Williams 2010: Experimental Political Science and the Study of Causality. Princeton:Princeton University Press.