**7**

**Korišćenje brojeva kao dokaza**

Političari imaju pomiješana osjećanja po pitanju zavisnosti njihove profesije od brojeva. S jedne strane, numerički dokazi čine se zadivljujuće transparentnim. Naučnici često svoje skupove podataka čine dostupnim široj javnosti, a neki naučni časopisi zahtijevaju od autora da to učine prilikom objavljivanja članka. Naučnici pokušavaju da jasno opišu svoje statističke tehnike i saopšte svoje glavne nalaze sa velikom preciznošću (npr. „Među zemljama korišćenim u ovoj studiji, povećanje stopa završetka srednjeg školovanja od deset procenata, povezana je sa povećanjem stope izlaznosti birača od procenata“). Oni koji analiziraju politiku upotrebom formalnih modela izvode svoje matematičke dokaze red po red. Svi ovi koraci bi trebalo da olakšaju drugima da imaju uvid u obavljeni posao, da ga ospore, ili da se na isti nadovežu. Ukratko, brojevi omogućavaju kumulativno znanje. Oni pomažu da političke nauke s pravom nose naziv nauke. Ovo je jedan od razloga zbog kojih se na nekim programima osnovnih studija i na većina postdiplomskih programa iz političkih nauka insistiraj da studenti uzmu najmanje jedan predmet iz oblasti statistike.

 Istraživanja zasnovana uglavnom na pisanim dokumentima čine se manje transparentnim. Napomene i bibliografije ukazuju samo na to gdje je moguće pronaći dokaze. Čitaoci bi i dalje morali provesti dosta vremena da pronađu izvorni materijal i prođu kroz njega. Autori rijetko objašnjavaju zašto su odlučili da uključe nekoliko dijelova dokaza iz određenog dokumenta, a isključe neke druge. Neki izvori, ukoliko su objavljeni na stranom jeziku ili pohranjeni u dalekoj arhivi, praktično su nedostupni za većinu ljudi. Zaključci su često manje precizni, a obično se navodi se da je neki faktor povezan sa nekim ishodom, ili da je faktor A važniji od faktora B. Ovo nije baš brza traka ako govorimo o intelektualnom napretku. Ne čudi što se većina članaka objavljenih u *American Political and Science Review*, vodećem časopisu u ovoj disciplini, oslanja na kvantitativnu analizu numeričkih dokaza.[[1]](#endnote-1)

 S druge strane, politikolozi mogu biti vrlo skeptični prema brojevima i onima koji ih analiziraju. Često ponavljamo dosjetku Mark Tvena o tri vrste laži – postoje laži, proklete laži i statistika[[2]](#footnote-1). Mi se samo donekle šalimo kada kažemo da tučemo ili mučimo podatke dok ne priznaju. Shvatamo da su mnogo nego će skraćenice BFF i YOLO ući u naš vokabular, kompjuterski naučnici imali su pojam GIGO – koji je značio *Garbage in, Garbage out* (Otpaci ulaze, otpaci izlaze, prim. prev.). GIGO znači da kad god su naše numeričke vrijednosti defektne, naša analiza tih brojeva će vjerovatno biti manjkava. Većinu vremena, naše mjere su nesavršene. Ponekad nam brojevi koji su namneophodni jednostavno nisu dostupni.[[3]](#endnote-2) S obzirom na ova ograničenja, možda bi bilo iskrenije zaključiti da je faktor A važniji od B, a ne da je A za 21% važnije. Politikolozi mogu ukazati na brojne situacije u kojima se statistička analiza pokazala pogrešnom. Vinjeta iz poglavlja 3 koja se odnosi na distribuciju stimulativnih finasijskih sredstava tokom Velike recesije, samo je jedan od primjera, a neke statističke tehnike su tako napredne, ili ezoterične, da malo ko od nas može stvarno analitički sagledati rad. Toliko o transparentnosti.

 Na svoj način, rad sa brojevima može biti jednako komplikovan kao rad sa riječima. Ako se sa njima ne barata na pravi način, obje vrste dokaza mogu ugroziti sve kvalitete koje naš projekat ima u pogledu dizajna istraživanja ili odabira teme. Mi treba da vodimo računa o svim našim dokazima (kao dobar advokat na suđenju). Naredni dio ovog poglavlja govori o tome šta kvalitet, raznovrsnost i osećaj za proporciju znače kada radite sa brojevima kao dokazima. Pošto smo vse eć dotakli važnih koncepata kao što su pristrasnost i triangulacija u poglavlju 6, njima se neću dodatno baviti. Ostatak ovog poglavlja će upoznati čitaoce sa nekoliko tehnika za analizu numeričkih podataka. Naglasak će biti na najčešće korišćenim tehnikama koje nam pomažu da odgovorimo na fundamentalna pitanja o tome šta se desilo, zašto i koga je briga. Kako možemo odrediti da li su dvije varijable direktno ili inverzno povezane? Kako da proverimo lažnu povezanost? Koji su rezultati značajni i koji i nisu toliko važni? Možemo li koristiti iste statističke tehnike ako su naše mjere nominalne, ordinalne, intervalne ili pak racio-mjere? Iako preporučujem da uzmete barem jedan predmet iz oblasti statistike kako biste saznali odgovore na ovakva pitanja, ovo poglavlje će vam pružiti osnovni uvid u opšte strategije i ključne koncepte.

**Kvalitet, raznovrsnost, osjećaj za proporciju**

Prije nego što napravimo bilo kakve proračune, potrebno je procijeniti kvalitet naših numeričkih podataka. Naši brojevi bi trebali biti autoritativni, i da dolaze iz zvaničnih ili poštovanih izvora. Studenti američke politike često koriste državnu statistiku iz Biroa za popis stanovništva ili Kancelarije za upravljanje i budžet. Studenti komparativne politike i međunarodnih odnosa mogu koristiti statistiku inostranih vlada ili međudržavnih entiteta kao što su Ujedinjene nacije i Svjetska banka. Često su to i najbolji brojevi. Ipak, vrijedi provjeriti postoje li i nedostaci. Na primjer, vlasti Argentine je pratila reputacija da potcjenjuje stopu inflaciju u ovoj zemlji. I to poprilično. Situacija je toliko loša da je u 2011. godini Međunarodni monetarni fond odlučio da potraži vjerodostojniji izvor za informacije o inflaciji u Argentini. Ubrzo nakon toga, Časopis *Economist* je učinio isto, objavivši da je „umoran od toga da bude nevoljni saučesnik u onome što izgleda kao namjeran pokušaj varanja birača i investitora.“ U istom članaku se ističe da su zvanični ekonomski podaci za Kinu „možda nepouzdani“.[[4]](#endnote-3) Neke akademske studije utvrdile su da su autoritarni režimi skloni da objavljuju previše optimistične brojeve, što vodi pristrasnosti veze između vrste režima i ekonomskog rasta, u korist režima.[[5]](#endnote-4)

 Vlade nisu jedini izvor autoritativnih brojeva. Neke vlade, posebno u siromašnim zemljama ili zemljama u kojima se odvijaju sukobi, nemaju dovoljno resursa za prikupljanje statističkih podataka u većoj mjeri. Čak i ako imaju kapaciteta, vjerovatno nećemo vjerovati ni jednoj vladi koja se mjeri nekim specifičnim dimenzijama („Duvajte u rogove! Predsjednik Vuvuzele s ponosom objavljuje da u njegovoj vladi nema korupcije“). Takvi brojevi bi mogli biti pristrasni. Stoga se, u želji da izmjerimo nivo korupcije javnog sektora u odrđenim zemljama, oslanjamo na nevladine organizacije kao što je *Transparency International*. Računamo na Freedom House i nekoliko drugih sličnih institucija kada je u pitanju saopštavanje informacija kako i koliko vlade širom svijeta štite politička prava i građanske slobode. Ispitivanja javnog mnjenja koja sponzorišu vlade takođe bi mogla biti sklona pristranosti („Dobrodošli u Sjevernu Koreju, gdje povjerenje u vladu uvijek iznosi 99%“). Srećom, možemo se okrenuti *Istraživanju svjetskih vrijednosti*, Međunarodnom istraživanju socijalnih programa, američkim nacionalnim izbornim studijama, glavnim medijima i ka nekoliko drugih organizacija ako nas interesuju mjerenja javnog mnjenja. Mnogi, mnogi naučnici se oslanjaju na ove izvore za autoritativne podatke prilikom istraživanja. Oni koji proučavaju međunarodne sukobe mogu se okrenuti skupu podataka *Korelati rata*, koji je potekao na Univerzitetu Mičigen ili Skupu podataka o oružanim sukobima, koji je povezan sa Univerzitetom Upsala u Švedskoj. Timovi istraživača razvili su brojne druge skupove podataka vezane za politiku (npr. AidData, Policy Agendas).

 Iako se ovi izvori čine vjerodostojnm i često koriste, oni i dalje imaju svoja ograničenja. Kao što smo saznali u poglavlju 2, naučnici se ne slažu o tome kako mjeriti velike koncepte poput demokratije i tolerancije, pa se bilo koji numerički indikator može smatrati manjkavim. Određeni indeks bi mogao isključiti ključnu dimenziju, uključiti irelevantne mjere ili agregirati komponente bez jasnog opravdanja. Čak i ako se slože konceptualnoj definiciji i ključnim dimenzijama, politikolozi se možda neće složiti oko odgovarajućim nivoom mjerenja, kao što je redni nasuprot intervalnom. Štaviše, svako ko blisko sarađuje sa ovim izvorima podataka shvata da su neki od brojeva, blago rečeno, nepouzdani. Američka nacionalna izborna studija već decenijama postavlja ljudima jednostavno pitanje glasanja. Prijavljena namjeravana izlaznost za predsjedničke izbore uvijek je iznosila između 70 i 80 posto od 60-ih godina prolog vijeka na ovamo. Prave cifre su, međutim, obično manje za 15 do 20 procentnih poena.[[6]](#endnote-5) Otkuda tako veliki jaz? Pa, mnogi ljudi misle da dobri građani glasaju, pa daju društveno poželjan odgovor čak i ako ne izlaze na birališta. Ako bi citirali brojeve dobijene putem ANIS da pokažemo da je izlaznost birača u Sjedinjenim Državama veća u poređenju sa izlaznošću u većem dijelu Evrope, ispali bismo glupi.

 Kako je greška u mjerenju sveprisutna prijetnja, političari često analiziraju odnose na više načina. Pretpostavimo da smo zainteresovani za odnos između obrazovanja i političke participacije u mnogim zemljama. Iz našeg pregleda postojeće literature otkrivamo da je obrazovanje mjereno oridnalno ili koristeći racio-mjere, zavisno od autora i zemalja koje se proučavaju. Racio-mjere uključuju procenat svršenih srednjoškolaca, procenat diplomiranih visokoškolaca i procenat pismenih ljudi. Ordinalna mjera bi mogla biti jednostavna i data u vidu kategorije koja odražava najviši stepen postignuća - manje od srednje stručne spreme, srednja stručna sprema, više od srednje stručne spreme. Takođe smo u literaturi našli i nekoliko mjera političkog učešća. Zatim možemo koristiti statističke procedure (npr. korelaciju, unakrsnu tabulaciju) da izračunamo odnose između svake mjere svake varijable.Tako ćemo triangulirati naše dokaze. Ako su svi rezultati prilično slični, bićemo uvjereniji u karakterisanje odnosa između obrazovanja i političkog učešća.[[7]](#endnote-6) Ako se rezultati razlikuju, onda bismo otkrili gdje i koliko su mjerenja bila bitna. I jeda i drugi rezultat može biti suštinski interesantan.

 Politikolozi stalno koriste ovaj pristup. Pomenuću dva primjera sa kojima bi trebalo da ste već upoznati. Kada je Elen Imergut testirala vezu između „moći doktora“ i učešća vlade u zdravstvenoj zaštiti, ona je prvo mjerila na tri različita načina: kroz broj doktora kao udijela ukupne populacije; procenat doktora koji su pripadali glavnom medicinskom udruženju u toj zemlji; i procenat poslanika koji su bili doktori. Sve tri numeričke mjere obuhvatile su različite aspekte apstraktnog koncepta, moć doktora. Međutim, nijedna od mjera nije bila usklađena sa očekivanjima sa zdravstvenom politikom što je omogućilo Imergutovoj da odbaci ovo objašnjenje.[[8]](#endnote-7) Kada je Majkl Ros testirao različite uzročne mehanizme koji povezuju prirodne resurse sa demokratijom u mnogim zemljama, jedna od mogućnosti bila je i represija. Više prihoda od nafte moglo bi generisati resurse potrebne za vlade da se obruše na bilo koga ko bi zagovarao demokratiju. Izmjerio je ovu varijablu dva načina, kao budžet na vojsku/ bruto društveni proizvod i kao broj vojnika/radna snaga. Prva mjera nije bila statistički značajna, ali druga jeste. Sam po sebi, veliki vojni budžet nije bio toliko važan, novac je morao da se troši na vojnike. Više nafte, više vojnika, manje demokratije.[[9]](#endnote-8) Kako ova dva primjera ukazuju, ovaj pristup numeričkim dokazima može se koristiti ako je studija mala ili velika. U svojoj ulozi inspektora, trebali bi primijetiti da li autori mjere svoje ključne varijable na više načina kako bi dokazali postojanost njihovih nalaza. Kada god je moguće, treba da prihvatimo ovaj isti pristup u sopstvenom istraživanju. Ako, pak, nemamo veliko povjerenje u naše numeričke mjere, triangulacija ima dobar smisao.

 Kao što je pomenuto u poglavlju 4, poređenja sa velikim *n* su obično adekvatnija od studija slučajeva sa malim *n* u pogledu otkrivanja i postupanja u slučaju lažne veze. Više slučajeva omogućava nam da analiziramo više kombinacija varijabli. Neko ko analizira odgovore nakon anketiranja 1.000 i više ljudi, čak i karakteristike 100 i više zemalja plus ima dovoljno slučajeva da testira razne potencijalno konfundirajuće varijable. Naravno, ne bi trebali uključivati ​​svaku numeričku varijablu do koje bi mogli doći. Moramo se fokusirati na teorijski relevantne varijable. Za odnos između obrazovanja i izlaznosti birača, mogli bismo da kontrolišemo prihode, ideologiju i starost... ali ne i koliko puta godišnje jedemo u Čipotlu (restoranu iz istoimenog lanca restorana u SAD, prim. prev.). Provjera uticaja drugih varijabli mogla bi dovesti do kreiranjaviše statističkih regresionih modela. Ros je upravo to uradio prilikom analize odnosa između prirodnih resursa i demokratije: kreirao je tri različita modela za kontrolu geografije, a svaki od tih modela uključivao je kontrole prihoda po glavi stanovnika i procenta muslimana u svakoj od zemalja. Prilikom testiranja mogućih uzročnih mehanizama, stvorio je najmanje tri statistička modela za svaki mehanizam.[[10]](#endnote-9)

 Triangulacija numeričkih dokaza, onda, može uključiti višestruke mjere pojedinačnog koncepta, ili više kombinacija različitih mjera. Moraćete sa mi povjerujete na riječ, ali moći će da uključuje različite statističke tehnike koje se primjenjuju na isti skup podataka.[[11]](#endnote-10) Kada upoređujete ukupne performanse linearnih regresionih modela (učinak koji pokazuje kako se model poklapa sa podacima), neki političari preferiraju izračunavanje prilagođene R-kvadrat statistike, dok su drugi skloniji računanju standardne greške u procjeni (engleska skraćenica je SEE). Da bi zadovoljili ove dvije publike, mogli bismo da prijavimo obje brojke.[[12]](#endnote-11) Kada rade s panelnim podacima, koji pokrivaju mnoge slučajeve tokom nekoliko godina, naučnici moraju da računaju na nasumične efekte ili fiksne efekte za relevantne varijable koje su isključene iz njihovih modela. Oni mogu testirati obje pretpostavke i prijaviti rezultate. Ako i kada saznate više o statistikama, naići ćete na mnoge druge slučajeve u kojima ne postoji pravi način da analizirate odnose između numeričkih varijabli. Statistička analiza često nalaže odluke koje su stvar lične procjene i suda, što je čini dijelom umjetnošću a dijelom naukom. Što je više prostora za procjenu, veća je potreba za triangulacijom.

 Osjećaj za proporciju, bar po mom iskustvu, funkcioniše malo drugačije za brojeve nego riječi. Oni koji rade sa brojevima manje se oslanjaju na opšte znanje. Većina ljudi ne poznaje jednostavne činjenice, kao što su prosječna izlaznost birača ili nivo potrošnje za vojsku u bilo kojoj zemlji osim svoje (a možda čak ni za svoju). Istraživači obično moraju identifikovati izvore za sve svoje brojeve. Međutim, oni koji rade sa brojevima vjerovatno će se u velikoj mjeri oslanjati na neki jedinstveni autoritativni izvor za razliku od svojih kolega koji prvenstveno rade sa dokumentima. Statistička analiza stavova prema imigrantima, koja je ranije pomenuta u ovoj knjizi, zasnivala se uglavnom na osnovu dokaza Evropskog društvenog istraživanja iz 2002-3.[[13]](#endnote-12) Projekat koji je osmišljen da objasni razliku u izlaznosti na izbore među američkim državama može u potpunosti biti zasnova na brojevima iz Biroa za popis stanovništva SAD. Kod ovakvih projekata, osjećaj za proporciju, može podrazumijevati odbranu određene teze višestrukim dokazima iz jednog izvora podataka, umjesto pojedinačnih dokaza iz više izvora (tj. na način na koji politikolozi uglavnom rade sa dokumentima). U nekim drugim slučajevima, međutim, osjećaj za proporciju, nas navodi da numeričke i tekstualne dokaza tretiramo jednako. I dalje nam je potrebno više dokaza kada su činjenice ili njihovo tumačenje sporni. I dalje nam je potrebno još dokaza da bi dokazali da je A prouzrokovalo B nego kada dokazujemo kovarijaciju između A i B.

 Ostatak ovog poglavlja predstavlja veoma kratak uvod u statistiku, koji je ekvivalent jednosatnoj turi po Britanskom muzeju. Moja namjera je da se pred čitaoce iznesu na neki jednostavne, ali potencijalno moćni načini analize brojeva, uz nadu da će nakon toga moći da efikasnije proučavaju dalju literaturu i poželjeti da saznaju mnogo više. Ukoliko se pravilno koristi, Statistika nam može reći mnogo o tome šta se i zašto dogodilo u politici. Između ostalog, moramo da usaglasimo naše statističke procedurame sa nivoom mjerenja u našim podacima. Počećemo sa jednom varijablom podataka, a nastaviti sa dvije, i konačno tri ili više varijabli.

**Analiziranje jedne varijable**

Koko politikologe zanimaju veze, većina naših statističkih tehnika uključuje dve ili više varijabli. Ipak, postoje i slučajevi kada analiza samo jedne varijable može biti od koristi. Na primjer, ona može biti ključna informacija za naš izbor predmeta: možemo identifikovati tipične slučajeve i ekstremne slučajeve za dalju, intenzivniju studiju.[[14]](#endnote-13) Takođe nam može reći da li su naši podaci distribuirani normalno ili nabnormalno, što bi moglo uticati na bilo koju naknadnu statističku analizu koju namjeravamo da obavljamo. Dobar deskriptivni rad često nalaže da generalizujemo u vezi sa jednom varijablom. Kada su Džon Sajds i Džek Sitrin analizirali stavove Evropljana pram imigrantima, njihova prve tabela sa brojevima sadržala je procentualni prikaz stanovnika koji su se izjasnili pozitivno, neutralno ili negativno o mogućim posljedicama imigracije.[[15]](#endnote-14) Onaj ko proučava prednost prethodnog mandata Kongresu bi možda trebalo da izračuna prosječnu razliku u pobjedi na izborima kada se predtavnici kandiduju za novi mandat. Onaj ko proučava pogranične sukobe u dvadesetom veku možda bi želio da sazna koliko ih je riješeno za manje od godinu dana, a koliko sukoba je trajalo duže od deset godina.

 Prvi podatak koji nam je potreban tiče se preciznosti naših brojeva. Neki brojevi su izraženi savršeno precizno, dok su bliže tome da se smatraju dobrim približnim pretpostavkama. Često stepen preciznosti zavisi od toga da li su naši brojevi nastali iz uzorka ili čitave populacije. Uzorci vode ka brojevima sa plus ili minus marginom greške, čineći nas uvjerenim da pravi broj leži negde unutar te margine greške. Procjene nivoa podrške predsjedniku često se zasnivaju na uzorku iz nacionalne populacije, između 1.000 i 1.500 ljudi. Iako bi mediji mogli da objave da je podrška na 44 odsto, stvarni broj u populaciji je vjerovatno negde između 41 i 47 procenata (44 ± margina greške od 3 procenta); malo je vjerovatno da je pravi broj u populaciji tačno 44,00 procenata. Ako podrška za dva mjeseca bude 42 posto, mediji mogu objaviti da je došlo do pada, i vjerovatno bi bili u pravu. Takođe je moguće, s obzirom na marginu greške, da nije došlo do promjee, ili čak da je došlo do povećanja podrške. Brojevi dobijeni iz većih uzoraka, kao što je *US Current Population Survey*, imaju manju marginu greške. Brojevi dobijeni iz cjelokupne populacije još su precizniji. Kada zemlje prijavljuju stopu smrtnosti odojčadi ili ukupan broj birača na poslednjim izborima, ti brojevi obično odražavaju svako pojedinačno rođenje i svakog birača. Suština je da ne treba prtjerivati sa preciznošću naših brojeva, pogotovo kada daju ključni dokaz za naše argumente.

 Ako radimo sa grupom brojeva, sledeći podatak koji nam je potreban je tipična vrijednost, ili ono što je poznato kao **mjera centralne tendencije**. Kako ćemo je računati, zavisi od nivoa mjerenja varijable (vidi poglavlje 2). Za nominalnu varijablu kao što je regija ili religija, koristimo **modus**, što najčešće znači vrijednost. Prema informacija iz CIA World Factbook, glavne religije u Singapuru su budizam (34%), hrišćanstvo (18%), bez religije (16%), islam (14%), taoizam (11%) i hinduizam (5%). Modusni odgovor je budizam; tipičan stanovnik Singapura je budista. U Španiji je 94% stanovništva rimokatoličke vjeroispovijesti, što je čini modusnom religijom.[[16]](#endnote-15) Dakle, modus može ali i ne mora opisivati većinu slučajeva za određenu varijablu. Kada je u pitanju ordinalna varijabla obično izračunavamo medijanu, koja odražava srednju vrijednost. Ordinalna varijabla je ona u kojoj se vrijednosti mogu poredati (npr. od manjih ka višim), ali rastojanja između vrijednosti su nepravilna ili nepoznata. Povjerenje pojedinaca u vladu je dobar primjer. Godine 1964, 22 procenta Amerikanaca izjavilo je da vjeruje nacionalnoj vladi da ponekad radi ono što je ispravno, 62 procenta je izjavilo da vlada to čini većinu vremena, a 14 procenata da vlada ispravno postupa skoro uvijek. Srednji odgovor, sa pola odgovora ispod i pola iznad, bio bi onaj iz grupe odgovora „većinu vremena“. Do 2008. godine, medijana se promijenila: 68 procenata, odgovoralo je sa „ponekad“, u poređenju sa samo 25, odnosno 5 procenata za druge dvije kategorije. Novi srednji odgovor bio je „ponekad“, što ukazuje na pad povjerenja.[[17]](#endnote-16)

 Kada su varijable intervalnog ili racio tipa, a rastojanja između svake vrijednosti ista, računamo **aritmetičku sredinu** (ono što većina ljudi naziva prosjekom). Dodamo sve vrijednosti i podijelimo ukupan broj sa brojem slučajeva. Na kraju poglavlja 5, objavio sam podatke za dvadeset pet fiktivnih zemalja. Tri od tih ponuđenih varijabli – prihod po glavi stanovnika, procenat diplomiranih srednjoškolaca i izlaznost glasača - izraženi su kao racio mjere. U tipičnoj zemlji, odziv birača bio je na nivo pod 70 posto. U prosjeku, 57 posto odraslih u ovim zemljama završilo je srednju školu. Ako bolje pogledamo tu tabelu, primijetićemo da nijedna zemlja nije imala odziv birača tačno 70 posto, a nijedna nije imala upravo 57 posto svršenih srednjoškolaca. Aritmetička redina ne mora predstavljati stvarni rezultat iz našeg skupa podataka.

 Konačno, potrebna nam je određena **mjera disperzije**, neke indikacija kako se podaci distribuiraju oko tipične vrijednosti. *Distribucija* podrazumijeva jasan red, pa se nominalne varijable tretiraju drugačije od ostalih.[[18]](#endnote-17) Nas zanimaju disperzija vrijednosti za intervalne, ordinalne i racio mjere. Normalna distribucija ima jedan jedini vrh u sredini i simetrična je sa obje strane - klasična krivulja u obliku zvona. Modus, medijana i aritmetička sredina su identični. Ovo odgovara prvoj ilustraciji na slici 7.1. Kada podaci imaju ovaj oblik, većina vrijednosti se koncetriše oko sredine. U stvarnom svijetu politike, podaci se ne ponašaju uvijek normalno. Političari se bave **zakrivljenom** (tj. asimetričnom) distribucijom. Neke su **negativno asimetrične**, što znači da na lijevoj strani distribucije postoji veći kraj (sl. 7.1, ilustracija b). Taj kraj obično smanjuje vrijednost aritmetičke sredine tako da je ona manja od medijane ili modusa. Pretpostavimo da su vam ocjene na testovima na nekom predmetu bile 0, 90, 90, 90 i 100. Vaša prosječna ocjena bi bila 74, što je dosta niže od medijane koja je 90. Može se desiti i suprotno. **Pozitivno asimetrična distribucija** ima veći kraj sa desne strane, što obično povlači aritmetičku sredinu iznad proseka (slika 7.1)[[19]](#endnote-18) Distribucija prihoda u zemljama često ima ovaj oblik, pošto mali broj veoma bogatih ljudi pripada desnom kraju distribucije.Mogući su takođe i drugi oblici. Ponekad nailazimo dvomodusnu distribuciju u kojoj je većina vrijednosti grupisana oko dva vrha. Osjećanja prema određenoj ličnosti iz svijeta politike mogu podijeliti javnost u dva različita tabora, pri čemu će malo njih biti u sredini. Podaci bi mogli biti vrlo ravnomjerno raspoređeni, bez jasno izraženog vrha. Kada god je to moguće, savjetujem studente da generišu slike svojih podataka tako da mogu vidjeti oblik distribucije. Stupčani grafikoni (*bar charts*), histogrami i kutijasti dijagrami mogu biti vrlo korisni u ovom pogledu („pita“ grafikoni (*pie charts*), ne toliko).

Slika 7.1. Klasične distribucije varijable. Napomena: Različite vrijednosti za određenu varijablu se pojavljuju duž horizontalne ose i njihove frekvencije duž vertikalne.

Postoji nekoliko načina mjerenja distribucije ordinalnih, intervalnih i racio varijabli. Najjednostavniji i verovatno najmanje koristan je pun **opseg** vrijednosti. Takozvana skala termometra osjećanja obično ispitanicima omogućava da ponude vrijednost od 0 do 100. Ako bi najniži prijavljeni rezultat bio 10 a najviši 95, onda bi opseg bio 10-95. Jedna mana opsega ješto su dovoljne samo dvije opservacije da se definišu krajnje tačke. Ogromna većina ljudi bi mogla dati Džordžu V. Buša rejting od 30 do 70, ali ako jedna osoba da ocjenu 0 i a druga 100, onda je pokriven cijeli opseg. Često je bolja opcija je **interkvartilni opseg**, koji opisuje srednjih 50 procenata distribucije, između dvadesetpetog i sedamdesetpetog percentila. Nekoliko ekstremnih slučajeva odstupanja neće iskriviti ovu statistiku. Većina koledža, na primjer, izvještava o interkvartilnom opseg rezultata standardizovanih testova kod studenata. Umjesto da kažu samo da prosječan kandidat koji polaže prijemni ispit ima rezultat na od, recimo, 24, navode da je srednja polovina kandidata postizala rezultate od 20 do 27 (što znači da je jedna četvrtina ostavarila rezlultat ispod 20 bodova a jedna četvrtinu iznad 27). Kada su u pitanju prihodi domaćinstava u Sjedinjenim Državama, interkvartilni raspon u 2013. godini bio je između 25.000 i 90.000 dolara. Iako bi Amerikanci koji prihoduju više od 100.000 dolara mogli da smatraju sebe srednjom klasom oni, tehnički govoreći, ne pripadaju populaciji sa srednjom visinom prihoda.[[20]](#endnote-19)

 Većina statističkih softverskih paketa, kao što su SPSS i Stata, olakšavaju izračunavanje **standardne devijacije** i **mjere asimetričnosti** jedne varijable. Obje mjere funkcionišu sve dok su varijable intervalne ili racio mjere. Obje mjere zasnivaju se na rastojanjima između svake opservacije i aritmetičke sredine, mada se tačne matematičke formule razlikuju. Standardna devijacija najbolje funkcioniše kada je distribucija vrijednosti blizu normale. U tom slučaju, oko dvije trećine podataka nalaziće se u okviru jednestandardne devijacije iznad i ispod sredine, a gotovo svi podaci (95%) mogu se naći u okviru dvije standardne devijacije aritmetičke sredine. Ako su ocjene na testu za cijelurazred normalno distribuirane, sa srednjom vrijednošću od 83 i standardnom devijacijom od 6,5, onda su svi imali skor između 70 i 96. Kada distribucija varijable nije toliko normalna, mjera asimetrije funkcioniše malo bolje (iako stvarno čudne distribucije mogu da je zavaraju). Za razliku od standardne devijacije, čija je vrijednost uvijek pozitivna, mjere asimetrije može biti pozitivna ili negativna. Ne iznenađuje što pozitivan broj ukazuje na pozitivnu (desnu) simetriju, dok negativan broj ukazuje na negativnu (lijevu) simetriju. Velika vrijednost asimetričnosti, sama po sebi, nije znak nevolje. Neki statističari savjetuju da kada je apsolutna vrijednost mjere asimetričnosti veća od dvostruke vrijednosti njene standardne greške, podatke treba smatrati pretjerano asimetričnim.

 Hajde da se vratimo na izmišljeni skup podataka iz poglavlja 5. Kada sam analizirao broj izašlih birača koristeći SPSS, saznao sam da je srednja vrijednost iznosila 69,96 posto a medijana 68 posto. Interkvartilni raspon povećao se sa 59 na 83,5 procenata. Standardno odstupanje iznosilo je 13,5, što znači da je većina zemalja bila u nivou između 53 i 96 posto kada je u pitanju izlaznost. (Džojrajd, Nju Trenton i Vuvuzela bile su jedine zemlje koje su izvan ovog opsega.) Mjera simetrije je bila 0.03, a standardna greška simetrije je bila .464. Prema gore navedenom pravilu, varijabilna izlaznost birača nije asimetrična, i trebalo bi da navedem aritmetičku sredinu kao tipičnu vrijednost. Nasuprot tome, pet ocjena na testu koji su ranije pomenuti su loše asimetrične, čime medijana od 90 odražava tipičniji skor od aritmetički prosječnih 74.

 Zašto bismo vodili računa da li su naši podaci distribuirani normalno ili ne? Pa, velika mjera asimetričnosti može nas dovesti do preispitivanja naše mjere centralne tendencije. Konkretno, možda ćemo poželjeti da pređemo sa aritmetičke sredine na medijanu ili da prijavimo obje. Prihodi u Sjedinjenim Državama su očigledno asimetrični na desno, zahvaljujući relativno malom broju multimilionera i milijardera. Prema američkom Birou za popis stanovništva, aritmetička sredina prihoda domaćinstava u 2013. godini iznosila je 72.641 dolara, dok je medijana bila 51.939 dolara.[[21]](#endnote-20) U pitanju je prilično velika razlika. Ako želimo da razgovaramo o resursima dostupnim tipičnom američkom domaćinstvu, trebalo bi da se pozovemo na aritmetičku sredinu. Više ljudi će biti grupisano oko te cifre nego oko medijane. Međutim, ako upoređujemo prihode širom zemalja i želimo da procijenimo koliko je prihoda dostupno u svakoj zemlji, onda aritmetička sredina može biti sasvim prikladna. Izbor između aritmetičke sredine i medijane zavisi djelimično od distribucije podataka, a djelimično i od većih pitanja koja postavljamo. Drugi razlog zašto je distribucija važna je što mnoge statističke tehnike za detekciju odnosa među varijablama, kao što su različiti oblici regresione analize, pretpostavljaju da se svaka varijabla normalno distribuira. Kada je varijabla loše asimetrična, politikolog je obično transformiše logaritamski kako bi došao do normalne distribucije. Rosovi statistički modeli prokletstva resursa djelimično su iz ovog razloga koristili prirodnu mjeru prihoda po glavi stanovnika.[[22]](#endnote-21)

**Analizaranje dvije varijable**

Jedan od preduslova za bilo kakvu deskriptivnu ili uzročnu vezu jeste da vrijednosti mjera variraju zajedno, bilo direktno ili inverzno. Kod studija slučajeva možemo dobro pretresti podatke i vidjeti da li se pojavljuju neki šabloni. To je ono što je Imergutova uradila u odbacivanju hipoteze o moći doktora. Međutim, osim u nekoliko slučajeva, oči nas mogu prevariti, ili naši skromni mozgovi možda neće moći da obrade sve brojeve koji su pred nama. Hvala bogu na računarima i statističkom softveru.

 Kada analiziramo odnose između varijabli, tražimo tri dijela informacije. Prvi i najosnovniji je da li je odnos **statistički značajan**. Laički objašnjeno otkrivanje statističke značajnosti svodi se na „hej, mislim da ovaj odnos stvarno postoji“. Preciznija verzija je da moramo biti 95 posto sigurni da postoji veza u cijeloj populaciji prije nego što je nazovemo statistički značajnom. Moramo biti vrlo, vrlo samouvjereni prije nego što zaključimo tako nešto iz naših podataka. Drugim riječima, vjerovatnoća da je nulta hipoteza tačna (tj. da ne postoji veza) mora biti 5% ili manja. Neki politikolozi tvrde da značaj postoji ako je manje od 10 procenata vjerovatnoće da je nulta hipoteza tačna, ali najvažnija stvar je to što smo mi politikolozi, po obrazovanja, prilično oprezna grupa ljudi. Biti 75 posto siguran da veza postoji, za nas jednostavno nije dovoljno. Radije bismo zaključili da ne postoji veza, čak i ako postoji, nego da svijetu kažemo da postoji veza a da ona ne postoji.[[23]](#endnote-22)

 Kada odnos nije statistički značajan, to je sve što treba da znamo. Nema smisla reći da je veza između dvije varijable statistički beznačajna, direktna i skromna po svojoj snazi. Ako veze nema to znači je uopšte nema.[[24]](#footnote-2) No, kada je odnos statistički značajan, onda moramo procijeniti njegovu snagu. Prečesto politikolozi kažu da su njihovi rezultati „značajni“ bez dodatnog objašnjenja. Iako će mnogi čitaoci izjednačiti značaj sa materijalnom važnošću, to često samo znači da statistički značajan odnos postoji. Postoji velika razlika između statističkog i materijalnog značaja. Statistički značajan odnos mogao bi biti slab ili jak, trivijalan ili značajan. Ako vam kažem, u najvećem povjerenju, da je porast broja pismenosti od 50 posto bio povezan sa povećanjem izlaznosti birača od 0,5 do 1,0 posto, možda i ne biste bili impresionirani. Država bi mogla da uloži značajan napor u unapređenje obrazovnog sistema, a izlaznost bi se tek neznatno pomjerila sa postojeće vrijednosti. Jupi. Za razliku od statističke značajnosti, politikolozi nemaju opštevažeći standard koji se koristi da se veza kvalifikuje kao kao snažna, skromna, slaba, važna, ili trivijalna. Govoriti o materijalnom značaju je komplikovano, donekle i subjektivno, i, po mom mišljenju, od ključne važnosti, jer se direktno tiče pitanja „Koga je briga?“ koje bi trebalo biti ključno za naš rad.

 Na kraju, i pravac odnosa je važan. Želimo da znamo da li su dvije varijable u direktnoj ili inverznoj vezi. Ovo je malo nezgodno utvrditi za nominalne varijable, jer nemaju pravac, samo različite kategorije. Nema nikakvog smisla u tome da u državi sa više regija mora automatski vladati i veća podrška kontroli vlasništva nad oružjem. Ali mogli bismo *koristiti* statistike da shvatimo da li, na primjer, ljudi koji žive na jugu manje podržavaju kontrolu vlasništva nad oružjem nego ljudi koji žive u drugim regijama SAD.

 U odabiru određene statističke tehnike, još jednom moramo biti svjesni nivoa mjerenja za naše varijable. Kada su obje varijable **kategoričke**, što znači nominalne ili ordinalne, vršimo **ukrštanje**. Ova tehnika je vrlo česta u studijama javnog mnjenja i ponašanja tokom glasanja. Postoji li veza između rase (nominalna) i podrške smrtnoj kazni (ordinalna)? Rod i stranačka pripadnost su povezani? Da li je opšti nivo nečijeg obrazovanja (ordinalna) u vezi sa njegovom podrškom regulativi u oblasti zaštite životne sredine (nominalna)? Da li je veća vjerovatnoća da će članovi sindikata češće glasati za ljevičarsku stranku na izborima u odnosu na one koji to nisu?

 Ono što mi pokušavamo da generišemo je tabela, koja se često naziva tabelom kontingencije, sa vrijednostima jedne varijable, odozgo, koje se pojjavljuju u vidu kolona, i vrijednostima druge varijable, sa strane, koje se pojavljuju u vidu redova. Tradicionalno, nezavisna varijabla daje se u kolonama, a zavisna varijabla u redovima. Da biste odgovorili na prvo postavljeno pitanje, rasa bi bila naša varijabla u kolonama, a stavovi prema smrtnoj kazni bi bili varijabla koja se pojavljuje u redovima. Ne možemo očekivati da stavovi prema smrtnoj kazni promijene nekog bijelca u crnca, ali možemo očekivati da pripadnost različitoj rasi utiče na njihov stav po ovom pitanju. Za ovaj odnos, vrlo jednostavna tabela može imati dvije kolone (crnci i bijelci) i četiri reda (Podržava snažno, Podržava - ali ne snažno, Protivi se – ali ne snažno, Protivi se snažno). Svaka od osam ćelija prikazaće ukupan broj ispitanika, vjerovatno iz nacionalne ankete, koji se uklapaju u date kolone i redove. Pošto se brojevi u redovima mogu teško interpretirati, treba da izračunamo procente po kolonama – koji procenat crnaca kaže da snažno podržava smrtnu kaznu, koji procenat podržava, ali ne snažno, i tako dalje. Mi bismo onda uporedili kolone, u potrazi za smislenim razlikama između bijelaca i crnaca.

 Iako razlike mogu biti tako vidljive odmah da znamo da se nešto događa, pomaže i to zvedemo nekoliko statističkih proračuna kako bi bili sigurni.[[25]](#footnote-3) **Pirsonov hi-kvadrat test** će nam pokazati statističku značajnost odnosa između dvije kategoričke varijable.[[26]](#endnote-23) Tačna vrijednost hi-kvadraat statistički je manje važna od njegove *p-* vrijednosti koja „nam pokazuje vjerovatnoću da ćemo vidjeti posmatrani odnos između dvije varijable u našem uzorku podataka, ako tu nije bilo stvarne veze između njih u populaciji koja nije posmatrana (tj. nulta hipotheza).[[27]](#endnote-24) Po pravilu, želimo da vidimo da li je vrijednost *p*- manja od 0,05, što znači da je vjerovatnoća manja od 5 posto. Kada je u pitanju snaga veze, možemo birati između nekoliko različitih **mjera povezanosti**. Ako je barem jedna od varijabli nominalna možemo izračunati Lambdu, Kramorov V, ili Somersov D.[[28]](#footnote-4) Moguće vrijednosti ovog statističkog opsega su od 0 do 1; što bliže nuli, to je veza slabija. Ako su obje varijable ordinalne, kao nivo obrazovanja i podrška zaštiti životne sredine, obično računamo snagu veze sa Kendalovim Tau, gamom, ili Spirmanovim Rho.[[29]](#endnote-25) Ovaj statistički raspon je od -1 do +1, što nam govori i snagu i pravac odnosa. Negativna vrijednost za ovaj statistički podatak ukazuje na obrnut odnos, pa što je bliža broju -1, jača je veza. Pozitivna vrijednost znači direktnu vezu, a veće znači jače. Što je broj bliži nuli, iz oba pravca, veza je slabija. Nigdje nisam vidio autoritativan prevod ovih brojeva u jednostavan jezik pridjeva. Kramerovo V ili Kendalov Tau od 0.31 mogu ukazivati na skromnu vezu po nekim istraživačima i prilično snažnu po drugima.

 Konkretan primjer može poslužiti kao potvrda ovih opštih zaključaka. Hajde da se vratimo na pitanje o rasi i pogledu na smrtnu kaznu. Slučajno sam sačuvao dio studije Američke nacionalne izborne studije (eng. skr. ANES) iz 2000. na svom kompjuteru, dijelom zbog istorijskih predsjedničkih izbora te godine. To istraživanje uključuje pitanja o rasi i smrtnoj kazni, i omogućilo mi je da izvršim ukrštanje (vidi tabelu 7.1). Uobičajen pregled kolona pokazuje neke razlike između crnaca i bijelaca. Izgleda da su crnci bili više nego dvostruko snažno protiv smrtne kazne, a bijelci su češće nego crnci bili za to da je snažno podržavaju. Međutim, kada su u pitanju srednja dva odgovora, crnci i bijelci su bili vrlo slični. Dakle, nije baš jasno kako možemo okarakterišemo cjelokupan odnos. Sada kada je zbirna statistika postala od pomoći p-vrijednost (.000) u hi-kvadrat statistici potvrđuje da je odnos bio statistički značajan.[[30]](#footnote-5) Kramerovo V, mjera asocijacije iznosila je 0,211. Lično, ja bih ovo nazvao umjerenim odnosom, ali se mišljenje drugih može razlikovati. Pozitivan znak Kramerovog V ne govori nam ništa, jer je rasa nominalna varijabla. Moramo da pogledamo tabelu procenata kako bi tumačili pravac. Da sumiram moje nalaze, mogu da navedem da je „došlo do umjerene (Kramerovo V = .211) i statistički značajne veza između rase i stavova prema smrtnoj kazni u 2000. Crnci su naginjali ka tome da je manje podržavaju nego bijelci. Više od 35 posto crnaca kaže da su snažno protiv smrtne kazne, u poređenju sa samo 14 posto bijelaca."

|  |
| --- |
| **Tabela 7.1 Ukrštanje rase i stavova prema smrtnoj kazni** |
|  | Bijelci | Crnci |
| Snažno podržava (%) | 57,7 | 33,5 |
| Podržava (%) | 18,3 | 18,3 |
| Protivi se (%) | 10.5 | 12,6 |
| Snažno se protivi (%) | 13,6 | 35,6 |
| *N* | 1320 | 191 |

Izvor: Američka nacionalna izborna studija, 2000.

*Bilješka.* **Pirsonov hi-kvadrat = 67,559 ; p = D00; Kramerovo V = 211. Kolona procenti ne može iznositi tačno 100,0 zbog zaokruživanja.**

 Kada su obje varijable intervalne ili racio, koristimo različitu tehniku - analize korelacije. **Pirsonov koeficijent** korelacije jedna je od najčešćih statistika u svim političkim naukama.[[31]](#endnote-26) Uobičajena skraćenica za koeficijent korelacije je *r*, a on može da varira od -1 do +1. Što se više udaljava od 0, bilo u pozitivnom ili negativnom smjeru, odnos je jači. U kontekstu korelacije, snaga se odnosi na stepen linearnog uklapanja naše dvije varijable. Zamislite grafikon s prhodima po glavi stanovnika zemlje na x osi i izlaznošću birača na y osi. Svaka zemlja u našem istraživanju biće predstavljena jednom tačkom (npr. x = $14,047; y = 56%). Ako sve ove tačke formiraju jednu liniju, koja ide nakoso gore tako da su veća primanja povezana s većom izlaznošću, onda je r jednak +1. [[32]](#endnote-27) U stvarnom životu, podaci su rijetko posloženi tako uredno. Što više naši podaci odstupaju od prave linije i naizgled razbacani nasumično, manja je vrijednost *r*. Kada je Robert Patnam analizirao korelaciju između ekonomske modernosti i rada institucija u regijama Italije, većina opservacija bila je u neposrednoj blizini linije, ali ne i direktno na njoj (*r* = 0,77). Pipa Noris je pronašla slabiju korelaciju (r=.20) između toga koliko često Amerikanci čitaju novine i koliko često su kontaktirali vladine zvaničnike. U principu, više jednog ponašanja vodilo je ka više ovog drugog. Bez obzira na to, može se naći veliki broj ljudi koji svaki dan čitaju novine, ali nikada nisu kontaktirali nekog službenika, i mnogi drugi koji rijetko čitaju

novine a češće ostvaruju kontakte.[[33]](#endnote-28)

 Ipak, nisu svi odnosi u politici su linearni. Neki mogu biti zakrivljeni, na razne načine, a koeficijent korelacije neće otkriti te odnose. R bi bio varljivo nizak. Ovo je još jedan razlog za prikazivanje naših podataka kroz slike. Radeći sa dvije varijable, intervalne ili racio, možemo napraviti dijagram rasipanja i vizuelno pregledati podatke za bilo koje nelinearne obrasce.

 Intervalne i racio mjere se često nazivaju kontinuirane varijable[[34]](#endnote-29) , a ANES 2000 uključivala je veliki broj njih. Nekoliko ih je predstavljalo klasična pitanja karakteristična za termometar osjećanja, gdje se pojedinci pitaju šta misle o određenom političaur, grupu ljudi, ili dijelu vlasti. Ovi rezultati mogu varirati od 0 (vrlo hladno) do 100 (vrlo toplo), gdje 50 ukazuje na neutralno osjećanje. Tabela 7.2 prikazuje matricu korelacija za par pitanja karakterističnih za termometar osjećanja.[[35]](#endnote-30) Nije iznenađujuće, osjećanja prema dva glavna predsjednička kandidata, Alu Goru i Džordžu Bušu, su bila inverzno povezana (r = -.414) i statistički zanačajna.[[36]](#footnote-6) Generalno govoreći, topliji pojedinci osjetio naginjali su ka Bušu, a hladniji ka Goru, i obrnuto. Osjećanja prema Goru i prema feministkinjama su u direktnoj vezi (r = 0,331), što je bio i slučaj sa Bušom i vojskom (r = 0,270); u oba slučaja veze su bile statistički značajne. Ništa novo. Ipak, možda najviše iznenađuje nedostatak korelacije između osjećanja ispitanika prema feministkinjama i vojsci. Takozvana p- vrijednost za statističku značajnost ukazuje na 22 posto šanse da je nulta hipoteza bila u pravu. U pitanju je vrijednost koja je dobrano iznad naših praga od 5 posto. Kao dobri politikolozi, morali bismo zaključiti da je postojala povezanost između ove dvije varijable. Neki Amerikanci verovatno imaju hladna osjećanja prema feministkinjama i vojsci, neki topla prema oboje, ali mnogi su imali pomiješana osjećanaja. Prije analize ovih brojeva, ja bih se dobro uvjerio da su ove dvije varijable inverzno povezane. Nešto sam saznao i to je u redu.

|  |
| --- |
| **Tabela 7.2 Korelacijska matrica skorova na termometru osjećanja** |
|  | **Gor** | **Buš** | **Feministkinje** | **Armija** |
| Al Gor |  |  |  |  |
| *r* | 1 | -.414\*\* | .331\*\* | -.072\*\* |
| Stat. znač. |  | .000 | .000 | .005 |
| *N* | 1774 | 1747 | 1410 | 1494 |
| Džordž V. Buš |  |  |  |  |
| *r* | -.414\*\* | 1 | -.203\*\* | .270\*\* |
| Stat. znač. | .000 |  | .000 | .000 |
| *N* | 1774 | 1761 | 1403 | 1487 |
| Feministikinje |  |  |  |  |
| *r* | .331\*\* | -.203\*\* | 1 | .033 |
| Stat. znač. | .000 | .000 |  | .219 |
| *N* | 1410 | 1403 | 1427 | 1413 |
| Armija |  |  |  |  |
| *r* | -.072\*\* | .270\*\* | .033 | 1 |
| Stat. znač. | .005 | .000 | .219 |  |
| *N* | 1494 | 1487 | 1413 | 1517 |

 Pretpostavimo da smo tražili korelacije među dvadeset pet izmišljenih zemalja navedenih u poglavlju 5. Mogli bismo isto učiniti i kada je u pitanju izlaznost birača, dohodak po glavi stanovnika, procenat svršenih srednjoškolaca, kako su u svim slučajevima u pitanju kontinuirane varijable. U svjetlu analize smo upravo izvršili putem skorova na termometru osjećanja, rezultati su pomalo zbunjujući. Korelacija između izlaznost birača i broja osoba koje su završile srednju školu je jača (. 473) od svega što smo vidjeli u tabeli 7.2. Ipak, vrijednost *p* je 0,017, što se kvalifikuje kao statistički značajna vrijednost, ali je malo veća od mnogih uporedivih brojki u tabeli 7.2. Kako možemo biti manje sigurni u vezu koja se čini jačom? Osim toga, korelacija između izlaznost birača i dohotka po stanovniku čini se prilično jakom (*r*= 0,372), ali pada na našem testu statističke značajnosti jer je *p*- vrijednost 0,067. Kako je to moguće? Odgovor na oba pitanja je isti - zbog veličine uzorka. Za pitanja vezana za termometar osjećanja, imali smo negdje između 1.400 i 1.800 odgovora. Bilo je mnogo lakše biti siguran da neka veza postoi nego kad smo imali samo dvadeset pet zemalja. U principu, što je manji broj slučajeva, veće mora biti r prije nego se veza može nazvati statistički značajnom. Loša stvar je, opet, to što neko sa velikim uzorkom može ponosno izjaviti da je određena veza statistički značajna, čak i ako je u pitanju veoma slaba veza.[[37]](#endnote-31)

 Usko vezana za analizu korelacije je jednostavna linearna regresija. Umjesto utvrđivanja da li su dvije varijable u vezi na bilo koji način, jednostavna regresija od nas zahtijeva da navedemo zavisnu varijable i nezavisnu varijablu. Polako se odmičemo od korelacije i krećemo sprema uzročnosti (iako je razdaljina između njih prilično velika). U studiji koja se tiče obrazovanja i izlazanosti birača, istraživač će, sasvim razumno pretpostaviti, da je je prvo nezavisna varijabla, a drugo zavisna varijabla. Među dvadeset pet izmišljenih zemalja smo analizirali, jednostavna linearna regresija nam govori da je u prosjeku, što je porast od 10 procenata u broju svršenih srednjoškolaca povezan sa povećanjem od 3,6 procenata u izlaznosti birača. To nam takođe govori da su varijacije u procentu svršenih srednjoškolaca u svim zemljama odgovorne za 19 posto varijacija u pogledu izlasnosti birača. Pojasniću odakle mi ove brojke za koji tren, kada dođemo do hijerarhijske regresione analize, koja se koristi mnogo češće nego jednostavna.

 Ukrštanja, korelacije, i jednostavne regresije su najčešći oblici bivarijatne analize podataka. Povremeno, želimo da analiziramo bivarijatni odnos između kategoričke i kontinuirane varijable, i u tom slučaju treba nam drugačija tehnika. Na primjer, pretpostavimo da želimo procijeniti vezu između pola i skora na termometru osjećanja kada je u pitanju Hilari Klinton. Ili smo možda imali kontinuiranu mjeru političke tolerancije kod pojedinaca u bilo kojoj zemlji, a željeli smo da vidimo da li je ista bila

vezana za vjersku pripadnost. Jedna opcija je **testiranje razlika aritmetičkih sredina**, koje najbolje funkcioniše kada je nezavisna varijabla kategorična i zavisna kontinuirana.[[38]](#endnote-32) Osnovna ideja je da se uporede aritmetičke sredine (i ukupne distribucije) za svaku vrijednost kategorične varijable. Testiranjem razlika aritmetičkih sredina bi se poredila, na primjer, osjećanja muškaraci i žena prema Hilari Klinton, ili kakove rezultate hrišćani, muslimani ili hindusi postižu na skali tolerancije.[[39]](#endnote-33) Relevantna mjera asocijacije zove se eta-kvadrat, i vrijednosti mogu varirati od 0 do 1. Statistička značajnost može utvrditi t-testom ako poredimo dvije grupe, ili f-testom za dvije ili više grupa.

 Šta se dešava ako koristimo pogrešnu statističku proceduru? Pa, naš odgovor može da se kreće u rasponu od toga da smo „gotovo u pravu“ do toga da „potpuno griješimo“. Pretpostavimo da analiziramo dvije nominalne varijable putem ukrštanja, a mjerimo snagu njihove veze s Kendalovim Tau, koji je pogodan za dvije ordinalne varijable, umjesto sa Kramerovim V. Šanse su dobre (ali ne i garantovane) da će odgovor koji dobijemo biti veome blizu tačnom, dijelom zbog toga što su i nominalne i redne varijable kategorične. Naša statistička analiza može se činiti pomalo neopreznom, ali ne i duboko pogrešnom. Međutim, ako bi koristili ukrštanje u situaciji kada bi trebalo napraviiti matricu korelacija, ili obrnuto, onda su šanse da napravimo veliku grešku značajno izvjesnije. Možda ćemo „pronaći“ vezu koja zapravo ne postoji, ili odbaciti odnos koji zaista postoji. Miješanje kategoričkih varijabli sa kontinuiranim varijablama može ozbiljno ugroziti našu analizu.

**Analiziranje tri ili više varijabli**

 Za većinu politikologa, statistička analiza daje prave rezultate tek kada se radi sa tri ili više varijabli odjednom. Tada možemo provjeriti lažnu povezanost i početi da odbacujemo određene hipoteze. Tada možemo odrediti koje su od naših nezavisnih varijabli manje ili više vezane za naše zavisne varijable. Možemo čak biti u stanju da detektujemo interakcije između naših nezavisnih varijabli. Više se ne moramo pretvarati da je politika oblikovana diskretnim parovima varijabli .

 U najopštijem smislu, politikolozi koristiti multivarijatne analize podataka kako bi u određenoj mjeri uveli kontrolu nad složenim i često misterioznim političkim pojavama. Kontrola je ključna. U poglavlju 4, upoznali smo se sa jednim od načina na koji politikolozi uspostavljaju kontrolu nad nekim djelićem svjetske politike – kroz eksperimente. Eksperimentalni dizajn istraživanja, u svom najčistijem obliku, omogućava istraživačima da manipulišu jednom nezavisnom varijablom, držeći svaki drugi mogući uticaj nepromjenjivim. Sve primijećene promjene u zavisnoj varijabli moraju biti povezane sa promjenom u toj jednoj nezavisnoj varijabli. Može biti naznačeno čak da nezavisne varijable izazivaju promjene (iako je precizan uzročni mehanizam često nepoznat). Mnoge manifestacije svijeta politike, međutim, nisu prikladne za eksperimente, iz praktičnih i etičkih razloga. Politikolozi koji se oslanjaju na posmatrački pristup, a posebno oni koji sprovode poređenja sa velikim *n*, koriste statističke tehnike za aproksimiranje one vrste kontrole koja se dobija eksperimentima.

 Multivarijatne tehnike se često baziraju na pristupima analiziranja dvije varijable numeričkih podataka. Sa kategoričkim podacima, na primjer, možemo raditi uporednu analizu kontrolišući treću varijablu. U suštini, izvorno ukrštanje se dijeli na više ukrštanja za svaku vrijednost kontrolne varijable.[[40]](#endnote-34) Nakon toga međusobno upoređujemo ova nova ukrštanja i na taj način možemo ustanoviti koliko se razlikuju. Možda bismo željeli da znamo li posmatrani odnos između rase i stavova prema smrtnoj kazni važi i za muškarce i za žene. Drugim riječima, želimo znati da li ova relacija važi za različite polove. Nakon **ukrštanja sa kontrolom**, otkrivamo da pol ne znači mnogo, barem nije u 2000. godini. Odnos između rase i stavova prema smrtnoj kazni je statistički značajan, i za muškarce i za žene. Korelacija je malo jača kod muškaraca, nego što je to slučaj sa ženama (Kramerovo V = .247 u odnosu na .190), ali nije u pitanju ništa dramatično. Naš sljedeći korak je da odbacimo pol i pokušamo da kontrolišemo obrazovanje; naš skup podataka ima ordinalnu mjeru koja dijeli odrasle u tri kategorije na osnovu nivoa obrazovanja (manje od srednje škole, srednja škola, više od srednje škole). Rezultati su prilično isti kao i za pol. Sva tri nova ukrštanja su statistički značajna, a mjera snage među njima varira samo malo. Sada smo još više uvjereni da je rasa povezana sa stavom prema smrtnoj kazni, jer smo imali u vidu nekoliko drugih mogućih faktora, a ključna statistika se vrlo malo promijenila. Da je rezultat bio značajno drugačiji za neku od naših kontrolnih varijabli, bili bismo manje uvjereni da rasa igra važnu ulogu. Naša prvobitna otkrića bi bila djelimično ili potpuno lažna.

 Nažalost, **kontrola** treće varijable je praktično sve što možemo postići sa ukrštanjem. Ukoliko bismo pokušali kontrolisati dvije ili tri varijable istovremeno, neki od naših unosa bi mogli imati vrlo malo ili nimalo opservacija (slučajeva). U suštini, iste teškoće se mogu pojaviti i kada kontrolišemo samo jednu varijablu više sa nekoliko mogućih vrijednosti, kao što je skala sa sedam tačaka za političke ideologije. Kontrolisanje kontinuirane varijable kao što je prihod po glavi stanovnika samo bi pogoršalo ovaj problem za nekoliko stepeni. Osnovni problem sastoji se u tome što se naše statističke tehnike zasnivaju na poređenjima posmatranih distribucija podataka u svakom polju tabele i hipotetičkih distribucija vrijednosti koje nisu u vezi (npr. u odnosu na nultu hipotezu). Kada neka polja imaju malo ili nemaju upšte podataka, odnosi sa drugim poljima moraju biti prilično jaki prije nego što se označe kao statistički značajni. U ekstremnim slučajevima, naš statistički softver neće biti u mogućnosti da izračuna statistički značaj ili snagu asocijacije.

 Radeći sa kontinuiranim varijablama, mogli bismo sa korelacije preći na **parcijalnu korelaciju**, što nam omogućava kontrolu treće varijable.U namjeri da objasni zašto neke regionalne vlade u Italiji postižu bolje rezultate od drugih, Patnam je testirao dvije opšte hipoteze, jednu ukorijenjenu u ekonomiji, a drugu u kulturi. Našao je jaku, pozitivnu korelaciju (r = 0,77) između njegovog indeksa ekonomske modernosti i indeksa rada institucija a još jaču korelacija (r = .92) između njegovog indeksa građanske zajednice i institucionalnog rada. Obje veze su bile statistički značajne, a čini se i veoma važne. Sljedeći Patnamov korak bio je računanje dvije parcijalne korelacije. Prva je predstavljala odnos modernosti ekonomije i rada institucija, kontrolišući građanske zajednice. Ta veza više nije bila statistički značajna. Drugi je bio između građanske zajednice i rada institucija, kontrolišući modernost ekonomije. Ta veza ostala je statistički značajna. Kao rezultat toga, Patnam je imao dokaz da je kultura važnija od ekonomije kada je u pitanju uticaj na rezultate rada vlasti.[[41]](#endnote-35)

 Oba ukrštanja i korelacije nam omogavaju da pređemo sa dvije na tri varijable, ali ne i mnogo više od toga.  **Višestruka regresiona analiza** se oslanja na jednostavne linearne regresije i više odgovara kontrolisanju varijabli. Umjesto samo jedne varijable, možemo simultano testirati dvije, tri, šest, trinaest – dakle, bilo koji broj varijabli .[[42]](#endnote-36) Pretpostavimo da imamo samo dvije nezavisne varijable (*A* i *B*) i jednu zavisnu varijablu (*Z*). Model višestruke regresije nam omogućava da procijenimo odnos između *A* i *Z* kontrolišući *B*, kao i odnos između *B* i *2* , kontrolišući *A*. U oba odnosa možemo identifikovati statističi značaj, materijalnu veličinu i pravac. Model nam takođe može statistički reći koliko je kombinacijom *A* i *B* nastalo varijacija u *Z*. To je mnogo korisnih informacija. Da smo imali nešto više razrađen model sa četiri nezavisne varijable ( *A*, *B*, *C*, *D*), onda smo mogli utvrditi odnos između *A* i *Z* kontrolišući *B*, *C* i *D*, i tako dalje. Tokom godina, razvijeno je nekoliko vrsta tehnika regresije za potrebe tumačenja različitih vrstapodataka i različitih vrsta istraživačkih pitanja. Za primjere date ispod, koristiću višestruku regresiju, i to metod najmanjih kvadratnih distanci, što je najosnovnija verzija.

 Na osnovu podataka iz mojih dvadeset pet fiktivnih zemalja, jednostavna linearna regresija je otkrila da je varijabla svršenih srednjoškolaca direktno povezana sa izlaznošću birača i da je statistički značajna. U prosjeku, porast od 10 procenata u varijabli obrazovanja povezan je sa povećanjem od 3.6 procenata u izlaznosti birača. Otkuda takvi brojevi? Kada sam provjeravao rezultate iz SPSS-a, primijetio sam da je *p*-vrijednost varijable srednjoškolaca (na osnovu statističkog *t* - testa) bila .017. Kao što je poznato, sve ispod .05 se računa kao statistički značajno. Nestandardizovani koeficijent (označen kao *B* ili *b*) bio je .364, što znači da je povećanje od jedne jedinice u nezavisnoj varijabli povezano sa povećanjem od .364 u zavisnoj varijabli. (Radi lakšeg izražavanja, ja sam brojeve množio sa 10) Za određivanje ukupne moći modela, oslonio sam se na prilagođenu R-kvadrat statistiku, koja je iznosila .190. Ona je blizak rođak rođak koeficijentu korelacije, *r*, i prilagođena je za broj nezavisnih varijabli u modelu.[[43]](#endnote-37) Vrijednost prilagođenog R-kvadrata može biti u rasponu od 0 do 1, a .190 znači da ova varijabla predstavlja 19 odsto varijacija u zavisnoj varijabli kada je izlaznost birača u pitanju.[[44]](#footnote-7)

 Kada sam dodao drugu nezavisnu varijablu, dohodak po stanovniku, rezultati su u početku bili zbunjujući. Nijedna nezavisna varijabla, kontrolišući druge, nije bila statistički značajna, a prilagođeni R-kvadrat jedva da je prešao (.192)[[45]](#footnote-8) . U nekim aspektima se ovaj model činio gorim od mog jednostavnog linearnog modela. Osnovni problem, ispostavilo se, bio je taj što su moje dvije nezavisne varijabe bile povezane međusobno (*r* .427). Sa toliko preklapanja u podacima, ne bih imao puno dodatne koristi od uključivanja obje varijable u model. Stoga sam izuzeo varijablu dohotka i dodao mjeru koja se tiče potrebe registracije birača prije izlaska na izbore, koja nije u korelaciji sa mojom varijablom obrazovanja.[[46]](#endnote-38) Kada su iskombinovane ove posljednje dvije varijable, vrijednost prilagođenog R-kvadrata skočila je na .335, što ukazuje da moj novi model može objasniti jednu trećinu varijacije izlaznosti birača. Statistički značaj varijable obrazovanja jedva da je promašio uobičajeni nivo od .05 (*p* = 0,055). Parcijalni koeficijent regresije ove varijable spustio se na .271. Varijabla za registraciju birača je, međutim, bila statistički značajna (*p* .023), kontrolišući obrazovanje, a pravac nestandardizovanog koeficijenta je bio negativan. To ima smisla: kako su zahtjevi za registraciju glasača veći, izlaznost birača opada. Tačna vrijednost *B* je bila -6,493, što znači da ako vrijednost nezavisne varijable poraste za jednu jedinicu (npr. od niže ka srednjoj vrijednosti), izlaznost birača pada u prosjeku za 6,5 procentnih poena, kontrolišući obrazovanje.

 Mogao bih lako povjerovati u to da se obje nezavisne varijable u ovom modelu čine važnima. Mi smo i dalje 94,5 procenata sigurni da je varijabla obrazovanja statistički značajna, kontrolišući zahtjeve registracije. Možda bi bilo korisno znati koja je od ove dvije varijable suštinski važnija, ali one su izražene u vrlo različitim jedinicama, što ih čini teškim za poređenje. Srećom, SPSS takođe stvara standardizovani koeficijent regresije (Beta) koji u suštini odbacuje određene jedinice varijabli i tako izračunava koliko je promjena u zavisnoj varijabli povezano sa promjenom jedne standardne devijacije u svakoj nezavisnoj varijabli. Što je veća apsolutna vrijednost Beta koeficijenta, veći je mogući uticaj. Shodno ovoj mjeri, moja varijabla registracije (-.426) je bila malo jača od moje varijable obrazovanje (0,352).[[47]](#endnote-39) Kada sam vratio varijablu prihoda po glavi stanovnika, stvarajući tako model sa tri nezavisne varijable, varijabla registracije birača ostala je statistički značajna, a njen nestandardizovani koeficijent je u suštini ostao nepromijenjen. Ni varijabla prihoda po glavi stanovnika, ni varijabla obrazovanja nisu bile statistički značajne (.273 i .178, respektivno). Stoga se čini da su, unutar granica ove male studije o izlaznosti birača, zahtjevi za registraciju bili najvažnija nezavisna varijabla, statistički i sadržajno.

 Iako jednostavan primjer kao što je ovaj može ilustrovati logiku višestruke regresione analize, on je atipičan, uzimajući u obzir mali broj slučajeva i nezavisnih varijabli (i potpuno lažne podatke). Tabela 7.3 analizira stav prema klimatskim promjenama i predstavlja nešto realniji model. Podaci potiču iz istraživanja sprovedenog na Internetu, među Amerikancima 2010. godine. Prethodne studije sugerisale su da su politički stavovi prilično tijesno povezani sa pogledom na klimatske promjene: konzervativci su za razliku od liberala promjene ove vrste rjeđe manje doživljavali kao problem ili nešto što iziskuje reakciju vlade. Autor studije istraživanja, Majkl Džons, smatra da je ta mjera političke ideologije bila preširoka. Želio je da ispita da li će određenije mjere sistema vjerovanja pratiti slične obrasce. Džons je sproveo niz višestrukih regresija metodom najmanjih kvadratnih distanci sa standardnim mjerama sedam tačaka političke ideologije, zajedno sa četiri kontinuirane mjere za različite kulturne vrste (individualista, hijerarhijsta, egalitarista i fatalista) među nezavisnim varijablama. On je takođe uključio kontrole za obrazovanje, pol[[48]](#endnote-40) i činjenično znanje o klimatskim promjenama. Njegove zavisne varijable, od kojih su neke prikazane u tabeli 7.3[[49]](#endnote-41) tražile su od ljudi da izraze svoj nivo saglasnosti sa određenim izjavama koje se odnose na klimatske promjene, u rasponu od „u potpunosti se ne slažem“ (0) do „potpuno se slažem“ (10). Iste je tretirao kao kontinuirane mjere.[[50]](#endnote-42)

 Kao što se očekivalo, varijabla ideologije bila je statistički značajna i negativna. Što su konzervativnije osobe bile, manje je bilo vjerovatno da će vjerovati u klimatske promjene, da je ljudska aktivnost izazvala klimatske promjene, te da je reakcija vlade neophodna. Rezultati koji nisu prikazani ovdje takođe pokazuju da su konzervativci manje vjerovali u to da klimatske promjene predstavljaju lični ili društveni rizik. Rezultati koji su bili od većeg značaja ticali su se četiri kulturološka tipa. Nekoliko ih je bilo statistički značajno, čak i uz kontrolu političkih stavova. Štaviše, nisu svi išli u istom pravcu. Dok je individualizam bio u obrnutoj vezi sa većinom zavisnih varijabli, hijerarhija i egalitarizam su često bili u direktnoj vezi. Individualisti su manje vjerovali da je do klimatskih promjena došlo zahvaljujući ljudskoj aktivnosti (B= -.128), ili da je važno za SAD da smanje emisiju gasova sa efektom staklene bašte (B= -.116). Hijerarhisti i egalitaristi snažnije su vjerovali da su ljudi bili uzrok klimatskih promjena i da je akcija bila potrebna. Istraživanje koje je sproveo Džons pokazalo je da linija razdvajanja, kada su u pitanju stavovi prema klimatskim promjenama, nije sadržana iskjlučivo u odnosu liberala i konzervativaca. Zanimljivo je da ispitanici nisu bili kao podijeljeni oko pitanja mogućih rješenja u vidu politika, kao što su obnovljivi izvori energije, nuklearna energija i politike ograničavanja i trgovine emisijama. Na taj način, kreatori politike možda žele izbjeći rasprave o obimu i uzroku klimatskih promjena i više se fokusirati na lijekove.[[51]](#endnote-43)

|  |
| --- |
| **Tabela 7.3. Višestruka regresija: Stav prema klimatskim promjenama** |
|  | **Vjerovanje u klimatske promjene** | **Klimatske promjene uzrokovane ljudskim faktorom** | **Neophodnost djelovanja** | **Obnovljiva energija** |
| Ideologija | -.483\*\* | -671\*\* | -.542\*\* | -.305\* |
| Individualizam | -.051 | -.128\*\* | -.116\*\* | .067 |
| Hijerarhija | .103\* | .176\*\* | .111\*\* | .095 |
| Egalitarizam | .092\* | .171\*\* | .244\*\* | .021 |
| Fatalizam | -.032 | -.079 | -.150\*\* | .009 |
| Obrazovanje | -.025 | .025 | .055 | -.137 |
| Pol | -.331 | -.369 | -.355 | -.814\* |
| Svijest o klimatskim promjenama | .339’ | .239’ | .308\* | -.009 |
| Pril. R2 | .180 | 388 | 395 | .061 |
| *N* | 262 | 262 | 263 | 255 |

Izvor: Michael D. Jones, “Leading the Way to Compromise? Cultural Theory and Climate Change Opinion,” PS: Political Science and Politics 44, no. 4 (October 2011): 720– 25.

Napomena: Koeficijenti nisu standardizovani

\*Značajna na nivou .05

\*\* Značajna na nivou .01

 Ovi primjeri samo su vrh ledenog brijega multivarijatnosti Sve ove varijable ponašale su se prilično normalno. U drugim istraživanjima, neke varijable će biti asimetrične i biće neophodno pretvoriti ih u logaritamsku skalu. Nezavisne varijable mogu se izraziti pojedinačno ili u interakcijskom obliku. Nisu sve zavisne varijable sa kojima politikolozi rade kontinuirane kao što su to stopa izlasnosti birača ili skor na termometru osjećanja. Neke od njih su dihotomne, sa samo dvije moguće vrijednosti: na primjer, pojedinci jesu ili nisu glasali na posljednjim izborima; zakonodavci su glasali za ili protiv određenog zakona; zemlje su bile u miru ili u ratu. Ova vrsta pitanja često se rješava sa logističkim modelom regresije, statističkim rođakom višestrukom regresijom, tačnije metodom najmanjih kvadratnih distanci, ili modelom proporcionalne vjerovatnoće. Moja dva primjera koristila su presjeke podataka poredeći mnoga zapažanja u istom trenutku. Kada istraživači rade sa podacima vremenskih serija ili panel podacima, neophodne su im naprednije statističke tehnike. Rosova analiza prokletstva resursa zasnovana je na objedinjenim podacima iz više od stotinu zemalja između 1971. i 1997. godine. Pročitajte njegov članak pažljivo i vidjećete da je on koristio verziju regresije poznatu kao izvodljiva uopštena metoda najmanjih kvadrata.[[52]](#endnote-44) Regresiona analiza, takođe, nije jedina opcija; neki analitičari preferiraju statističku tehniku koja se zove procjena maksimalne vjerovatnoće. Sada je očigledno zašto je potrebno više od jednog semestra slušanja predmeta iz oblasti statistike da bi se istinski ovladalo materijom.

**Završna razmatranja**

U ovom poglavlju sam u velikoj mjeri izbjegava onu nezgodnu riječ na U - *uzročnost*. Koristeći neke primjere statističke analize, ukazao sam na empirijske pravilnosti: da li su moje varijable bile povezane, pridružene, u vezi, u korelaciji. Brojni politikolozi bi pregledajući moje podatke došli i do jačih zaključaka. Rasa je imala uticaja na stavove prema smrtnoj kazni. U prosjeku, porast od 10 procenata u pogledu broja svršenih srednjoškolaca vodi povećanja od 3,6 procenata u izlaznosti birača. Sama po sebi, varijabla obrazovanja objašnjavala je skoro 20 procenata varijacije u izlaznosti na izbore. Mnogo različitih načina je da se ukaže na vezu između uzroka i posljedice.

 Ustezao sam se da koristim ovu vrstu jezika zbog glasova koje čujem u glavi.[[53]](#footnote-9) Prvi glasovi tvrde da samo pažljivo osmišljeni eksperimenti mogu ukazati na uzročnost (vidi poglavlje 4). Koliko god da se trudim da ou obzir uzmem i druge uticaje, znam da +nikada ne mogu kontrolisati sve sa opservacionim dizajnom istraživanja. Neka neopažena varijabla može zbrisati, ili u najmanju ruku ozbiljno oslabiti, svaku vezu koju sam možda otkrio. Potrebno je samo na tren pomisliti na demografske varijable (npr. starosnu dob, dohodak) i varijable u pogledu stavova (npr. povjerenje u vladu, orijentaciju prema budućnosti) koje bi mogle uticati na nečiji stav o klimatskim promjenama, pa ipak – sve one su izostavljene iz Džonsove studije. U principu, ti glasovi mi govore da je najbolje što možemo učiniti s opservacionim podacima jeste da isključimo određene odnose ako ne prođu test statističke značajnosti, što je korisno saznanje, i da istaknemo odnose koje bi i drugi naučnici mogli detaljnije testirati eksperimentima.

 Drugi glasovi pripadaju stvarno pametnim ljudima koji su vješti u statističkim analizama a opet zavise od višestrukih dokaza ukoliko žele da utvrde postoanje uzročno-posljedične veze. Neki od njih su istaknuti matematički statističari[[54]](#endnote-45) Drugi se pominju poglavlju 4 gdje smo naveli njihove tvrdnje da su detaljne studije slučaja pogodnije za utvrđivanje uzročne putanje i mehanizma od statističkih poređenja sa velikim *n*.[[55]](#endnote-46) Henri Brejdi, politikolog koji zna ponešto o brojevima, ima sjajnu ilustraciju ograničenosti statističkih modela. Ubrzo nakon predsjedničkih izbora 2000. godine, jedan regresiona analiza, koja je zavrijedila pažnju javnosti, pokazala je Džordž Buš izgubio najmanje deset hiljada glasova na Floridi jer su mediji objavili rezultate izbora deset minuta prije nego što birališta su zatvorena u jednom dijelu države. Dakle, Buš je trebalo da lako osvoji Floridu, što bi te izbore učinilo mnogo manje spornim. Pažljivim procesom analize, koristeći kombinaciju dokumenata i brojeva, Brejdi je pokazao da je velika većina stanovnika Floride već glasala prije nego što su se medijske mreže oglasile, da većina ljudi koji nisu glasali nisu ni čuli objavu, i da većina ljudi koji nisu glasali sve i da su čuli istu, vjerojatno ipak ne bi odlučili da glasaju. Pravi broj glasova koje je izgubio je bliže cifri od pedeset, što je daleko od prvobitne procjene.[[56]](#endnote-47)

 Patnamova analiza italijanskih regionalnih vlada nije stala nakon utvrđene djelimične korelacije. Nakon mjerenja njegovih ključnih varijabli i proračunavanja njihovih veza, on je promijenio brzinu, posvetivši posljednja dva poglavlja istorijski obaviještenoj studiji slučaja Italije. Ako ništa drugo, htio je pokazati da građanske zajednice utiču na rad institucija a ne obrnuto. Vrijeme i poredak su među ključnim preprekama sa kojima se suočavaju uzročni argumenti. Osim toga, Patnam želio da objasni kako građanstvo može poboljšati učinak vlasti. On je tvrdio da je „društveni kapital“ bio ključni element koji povezuje ove dvije varijable.[[57]](#endnote-48) Dok neki naučnici sumnjaju u Patnamov argument, moja poenta je u tome da je shvatio da su njegovi brojevi mogli da otkriju samo toliko.

 Moja statistička analiza nije mi pokazala kako rasa utiče na stavove prema smrtnoj kazni. Mogao sam amisliti bilo koji broj razloga, kao što su mišljenje o pravičnosti sistema krivičnog pravosuđa, vjerska uvjerenja o oduzimanju života drugoj osobi, percepciju prijetnje koja dolazi od nasilnih kriminalaca u jednoj zajednici, i istorijsko nasljeđe ropstva i segregacije. Svi su prihvatljivi, ali nijedan nije testiran. Dok neko ne shvati kako rasa oblikuje stavove prema smrtnoj kazni, iznošenje snažne tvrdnje o uzročnosti se čini preuranjenim (možda čak i nepouzdanim). U međuvremenu, ja ću pokušati da se uzdržim od toga da bijem svoje podatke dok ne priznaju.

 Ovo poslednje poglavlje, stoga, nije baš pravi zaključak. Lekcije koje se nalaze u ovom vodiču ne kulminiraju u vidu sofisticiranog modela višestruke regresije. Niti kulminiraju najdetaljnijom studijom slučaja ili domišljatim eksperimentom. U stvari, ova knjiga uopšte ne bi trebalo da ima zaključak. Čitav ovaj projekat predstavlja širi uvod u proučavanje i praksu političkih nauka. Šta se dalje dešava zavisi od vas, čitalaca. Svijet politike je komplikovan, često konfuzan, a ponekad i apsurdan. Bilo bi primamljivogledati na njega sa zbunjenošću ili prezirom da ulog nije bilo tako prokleto visok. Nadam se da će vas poznavanje nekih osnovnih pojmova i opštih strategija učiniti spremnijim da razumijete smisao politike u budućnosti, a možda čak i inspirisati da proširite i unaprijedite vaše analitičke vještine. Svet nije baš očajnički željan profesionalnih politikologa – ali mu uvijek dobro dođu ljudi koji pažljivo i sistematski promišljaju o politici.

 **VJEŽBA: INSPEKCIJA**

1. Provjerite da li možete da protumačite numeričke tabele, brojke i grafikone u publikacijama koje su navedene u nastavku. Fokusirajte se na statistički značaj, snagu i pravac svakog odnosa.

Heather K. Evans, Victoria Cordova and Savannah Sipale, 'Twitter Style: An Analysis of How House Candidates Used Twitter in Their 2012 Campaigns," *PS Political Science and Politics* 47, No. 2 (April 2014): 454-62;

M. Steven Fish, "Islam and Authoritarianism," *World Politics* 55, No. 1 (October 2002): 4-37;

Ronald Inglehart and Christian Wetzel, "Changing Mass Priorities: The Link between Modernization and Democracy," *Perspectives and Politics* 8, No. 2 (June 2010): 551-67;

Gary C. Jacobson, "It's Nothing Personal: The Decline of the Incumbency Advantage in US House Elections: Journal of Politics 77, No. 3 (July 2015): 861-73;

Michael D. Jones, "Leading the Way to Compromise? Cultural Theory and Climate Change Opinion," PS. *Political Science and Politics* 44, No. 4 (October 2011): 720-2S;

Edward D. Mansfield, Diana C. Mutz i Laura R. Silver, "Men, Women, Trade, and Free Markets," *International Studies Quarterly* 59, No. 2 (June 2015): 303-15;

Pippa Norris, "Does Television Erode Social Capital? A Reply to Putnam: PS. *Political Science and Politics* 29, br. 3 (September 1996): 474-80;

Robert D. Putnam, *Making Democracy Work: Civic Traditions in Modern Italy* (Princeton, NJ: Princeton University Press, 1993.), poglavlje 4;

Craig Volden and Alan E. Wiseman, *Legislative Effectiveness in the United States Congress: The Lawmakers* (New York: Cambridge University Press, 2014.), poglavlje 3.

**VJEŽBA: IZGRADNJA**

1. Za svaki od sljedećih parova varijabli navedite da li bi ukrptanje, korelacija ili razlika aritmetičkih sredina bila najprikladnija statistička tehnika za upotrijebiti:

Etnička grupa; povjerenje u vladu (nisko, srednje, visoko)
Dohodak po glavi stanovnika zemlje; procenat žena poslanica u parlamentu
Region UK; partijska pripadnost u Velikoj Britaniji (konzervativci, zeleni, laburisti, liberalne demokrate, škotski nacionalisti, indipendentisti, ostali)
Nivo obrazovanja (manje od srednje škole, srednja škola, pohađao koledž, svršeni visokoškolac); pogled na istopolne brakove (snažno se suprotstavlja, suprotstavlja se, podržava, snažno podržava)

Starosna dob; skor na termometru osjećanja za bilo koju političku ličnost

Etnička grupa; skor na termometru osjećanja za bilo koju političku ličnost

1. Lizbet Hohe je željela da utvrdi da li je rad za međunarodnu organizaciju uticao na podršku jedne osobe određenim nadnacionalnim normama. Ako jeste, da li bi se ta veza održala, kontrolišući druge moguće uticaje? Ona je anketirala najviše zvaničnike Evropske komisije o njihovim stavovima prema Evropskoj uniji, a matrica korelacije ispod pokazuje samo mali dio ovih rezultata. Kako biste tumačili ove bivarijatne odnose? Razmišljajte u kontekstu statističkog značaja, pravca i snage.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Nadnacionalizam** | **Ideologija** | **Veličina zemlje** | **Međunarodno obrazovanje** | **Radni vijek** |
| Nadnacionalizam | 1 | -.17\* | .01 | .22\*\* | .16\* |
| Ideologija |  | 1 | -.14 | -.05 | -.03 |
| Veličina zemlje |  |  | 1 | -.04 | .40\*\* |
| Int. obrazovanje |  |  |  | 1 | .01 |
| Radni vijek |  |  |  |  | 1 |

*Izvor:* Liesbet Hooghe, "Several Roads Lead to International Norms, but Few via International Socialization: A Case Study of the European Commission," *International Organization* 59, No. 4 (October 2005): 861-98.

Napomene: Nadnacionalizam mjeri stepen do kojeg ispitanici smatraju da Evropska unija, a ne pojedinačne zemlje, treba da preuzme vođstvo u upravljanju Evropom; ideologija je samoprijavljena i kreće se od ljevičarske do desničarske; veličina zemlje se odnosi na populaciju u milionima; međunarodno obrazovanje ispitanika se kreće od niskog do visokog; radni vijek jednak je broju godina rada za Evropsku komisiju.

\*p < .05.

\*\*p<.01.

1. Kako biste tumačili sljedeće rezultate iz modela višestruke regresije u kojoj je zavisna varijabla samoprijavljena politička ideologija, od ljevičarske (1) do desničarske (10), u Brazilu?

|  |  |  |  |
| --- | --- | --- | --- |
| **Koeficjenti** | **Nestandardizovano *B*** | **Standardizovano (Beta)** | **Znač.** |
| Godine | .016 | .091 | .003 |
| Dohodak | -.035 | -.026 | .393 |
| Nivo obrazovanja | -.128 | -.118 | .000 |
| Ukupan model | Kvadrat R=.032 | Prilagođeni R kvadrat =.030 | VIDI=2.770 |

*Izvor: World Values Survey wave 6 (data for Brazil, 2014).*

1. John Ishiyama, "Annual Report of the Editors of the American Political Science Review, 2013-2014," PS. Political Science and Politics 48, no. 2 (april 2015.): 396-99. Kao što je već pomenuto, dužina standardnog članka u časopisu takođe više odgovara istraživačkom pristupu orijentisanom ka varijabli i kvantitativnim tehnikama. [↑](#endnote-ref-1)
2. Ili one Tvenove: „Činjenice su tvrde, ali je statistika savitljiva.“ [↑](#footnote-ref-1)
3. Npr. Matt Ford, 'The Missing Statistics of Criminal Justice," Atlantic, 31 maj. 2015. godine; dostupno na http://www.theatlantic.com/politics/archive/2015/05/what-we-dont-know-about-mass-incarceration/394520/. [↑](#endnote-ref-2)
4. "Don't Lie to Me, Argentina: Economist, 25. Februar 2012. godine; dostupno na http://www.economist.com/21548242; Pogledati takođe Juan Forero, "A Quiet Battle over Argentina's Inflation Rate: Washington Post, 31. Oktobar 2011. godine; dostupno na http://www.washingtonpost.com/world/americas/a-quiet-battle-over-argentinas-inflation- rate/2011/10/29/qlQAEiUiYM\_story.html. [↑](#endnote-ref-3)
5. Christopher S. P. Magee i John A. Daces, "Reconsidering Regime Type and Growth: Lies, Dictatorships, and Statistics," International Studies Quarterly 59, no. 2 (Jun 2015.): 223-37. [↑](#endnote-ref-4)
6. Pogledati tabelu 6A.2 na http://www.electionstudies.org/nesquide/toptable/tab6a\_2.htm (pristupljen 29. februara 2016). Za tačnije brojke u vezi sa izlaznošću, pogledati http://www.electproject.org/national-1789-present (pristupljeno 29. februara 2016.). [↑](#endnote-ref-5)
7. U ovoj situaciji, naučnici često rade sa jednom mjerom u tekstu, ali u fusnoti ili pojašnjenju navode da druge mjere istog koncepta daju slične rezultate. [↑](#endnote-ref-6)
8. Ellen M. Immergut, "The Rules of the Game: The Logic of Health Policy-Making in France, Switzerland, and Sweden" u Structuring Politics Historical Institutionalism in Comparative Analysis, ed. Sven Steinmo, Kathleen Thelen i Frank Longstreth (New York: Cambridge University Press, 1992.). str. 57-89. [↑](#endnote-ref-7)
9. Michael L Ross, "Does Oil Hinder Democracy? ". *World Politics* 53, no. 3 (april 2001.): 325-61. [↑](#endnote-ref-8)
10. Ibid. [↑](#endnote-ref-9)
11. Zapravo, oni koji se oslanjaju na statističku analizu često koriste više izvora kako bi opravdali njihovu mjeru ključnih varijabli i tehnike nego izvore podataka [↑](#endnote-ref-10)
12. Možemo takođe naznačiti statistički značaj cijelog modela, koji je određen F testom. [↑](#endnote-ref-11)
13. John Sides i Jack Citrin, "European Opinion about Immigration: The Role of Identities, Interests and Information*," British Journal of Political Science* 37, no. 3 (Jul 2007.): 477-504. [↑](#endnote-ref-12)
14. Hajde da se na trenutak vratimo na prokletstvo resursa. Pretpostavimo da proučavamo zemlje svijeta koje najviše zavise od resursa i da želimo da znamo koliko dobro one štite politička prava. Prema skali od 7 tačaka koje je razvila organizacija *Freedom House*, mnoge od ovih zemalja će se rangirati blizu dna, bilo da je u pitanju 5 (npr. Kuvajt), 6 (npr. Brunej, Libija) ili 7 (npr. Saudijska Arabija). To su tipični slučajevi. Nekoliko zemalja se, ipak, kotira pri vrhu (npr. Norveška, 1). One su tzv. ekstremni slučajevi. [↑](#endnote-ref-13)
15. Sides and Citrin, "European Opinion about Immigration." [↑](#endnote-ref-14)
16. Pogledati https://www.cia.gov/library/publications/resources/the-world-factbook/. Pristupljeno 1. Marta 2016. Godine. [↑](#endnote-ref-15)
17. Pogledati [http://www.electionstudies.orq/nesquide/toptablettab5a 1.htm. Pristupljeno 29](http://www.electionstudies.orq/nesquide/toptablettab5a%201.htm.%20Pristupljeno%20%20%20%20%20%2029). februara 2016. godine. [↑](#endnote-ref-16)
18. Za nominalne varijable, distribucija je prosto potpuna lista kategorija i njihovih vrijednosti, otprilike kao što sam ja učinio sa religijom u Singapuru. [↑](#endnote-ref-17)
19. Moguće je da srednja aritmetička vrijednost bude veća od medijane u negativnoj raspodjeli, i da medijana bude manja od aritmetičke sredine u pozitivnoj distribuciji. Ovo se može desiti ako distribucija ima nekoliko vrhova, ili ako je jedan rep kratak i debeo, a drugi dugačak i mršav. [↑](#endnote-ref-18)
20. Imajte na umu da mi prijavljujemo svaku vrstu opsega od tačke A do tačke B, a ne samo razdaljine između A i B. Ako bih vam rekao da je interkvartilni raspon prihoda domaćinstva 65.000 dolara, vi ne biste znali da li to znači između 10,000 $ i 75.000 $, između 50,200 $ i 115,200 $, ili nešto drugo. Umjesto toga, prijavljujem takava raspon od 25,000 $ do 90,000 $. [↑](#endnote-ref-19)
21. Carmen DeNavas-Walt i Bernadette D. Proctor, *Income ond Poverty in the United States: 2013*, Current Population Reports, P60-249 (Washington, DC: Government Printing Office, 2014.), tabela A-1. [↑](#endnote-ref-20)
22. Ross, "Does Oil Hinder Democracy?" [↑](#endnote-ref-21)
23. Statističari razlikuju greške tipa I ("lažni pozitivni rezultati") i greške tipa II ("lažni negativni rezultati"). Političari bi mnogo radije napravili grešlu tipa II nego grešku tipa I. [↑](#endnote-ref-22)
24. Pretpostavimo da vas je prijatelj pitao da li ste u vezi sa Tejlorom i Vi ste rekli da niste. Ako je taj prijatelj pitao da li je vaša veza slaba ili jaka, kratka ili duga, vjerovatno ćete ponoviti odgovor, sa naglaskom da nepostojanje veze znači nepostojanje veze. [↑](#footnote-ref-2)
25. Nasuprot tome, većina tabela koja se pojavljuje u medijima jednostavno prikazuje kolone sa procentima i ostavlja čitaocima da odluče da li postoje značajne razlike. [↑](#footnote-ref-3)
26. Međutim, kada je *n* mali (npr. manji od 50), neki statističari preferiraju upotrebu Fišerovog tačnog testa umjesto hi-kvadrats. [↑](#endnote-ref-23)
27. Paul M. Kellstedt i Guy D. Whitten, The Fundamentals of Political Science Research, druga edicija ed. (New York: Cambridge University Press, 2013.), str. 147-48. [↑](#endnote-ref-24)
28. Lično, više volim Kramerovo V, ali me morate odvesti na večeru prije nego što vam kažem zašto. [↑](#footnote-ref-4)
29. Dodatno koristimo Kendalov tau-b ukoliko je kontigencijska tabela u potpunosti kvadratna, a tau-c ako nije. Kvadratna tabela je ona sa jednakim brojem kolona i redova (2 x 2, 3 x 3). U mom primjeru rase i stavova prema smrtnoj kazni, tabela bi bila 2 x 4, te ćemo stoga koristiti tau-c. U praksi, vrijednosti za ove dvije verzije tau su vrlo blizu jedna drugoj. [↑](#endnote-ref-25)
30. Iako je *p*-vrijednost prijavljena kao .000, nikada nismo 100% sigurni da odbacujemo nultu hipotezu. Ako je SPSS predstavio brojeve sa više decimala, možda je prava *p*-vrijednost u ovom primjeru bila nešto oko .0003. [↑](#footnote-ref-5)
31. Kao što vidite, određene tehnike se često nazivaju po statističarima koji ih razviju. [↑](#endnote-ref-26)
32. Uprkos tome, ako naši podaci formiraju savršeno horizontalnu liniju, *r* se ne bi računalo. Kada je svaka vrijednost na *x* osi povezana sa istom vrijednošću na *y* osi, nemamo kovarijaciju; naša *y* varijabla bi bila konstanta. Ista logika se primjenjuje na savršeno vertikalnu liniju, koja bi učinila *x* varijablu konstantom. [↑](#endnote-ref-27)
33. Robert D. Putnam, Making Democracy Work: Civic Traditions in Modern Italy (Princeton, NJ: Princeton University Press. 1993.), str. 85; Pippa Norris, "Does Television Erode Social Capital? A Reply to Putnam," PS. Political Science and Politics 29, no. 3 (September 1996.): 476. [↑](#endnote-ref-28)
34. U ovom poglavlju poglavlja napraviću razliku između kategoričkih (nominalnih, ordinalnih) i kontinuiranih (intervalnih, oracio) varijabli. Ovo je prilično standardna praksa među politikolozima. Neki naučnici, međutim, smatraju da intervalne i racio varijable mogu biti kontinuirane ili diskretne, što je slično razlici između realnih i cijelih brojeva; prvi se mogu izraziti kao razlomci, dok drugi ne mogu. Ovi naučnici mogu stoga suprotstaviti jedne naspram drugihkategoričke i kvantitativne varijable. [↑](#endnote-ref-29)
35. Matrica korelacije je simetrična sa obje strane dijagonale koja kreće od gornje lijeve strane, a završava na donjoj desnoj strani matrice. Dakle, samo treba da obratimo pažnju na polovinu ovih brojeva, kako su na drugoj polovini samo duplikati. [↑](#endnote-ref-30)
36. U tabeli 7.3 i u mnogim objavljenim studijama vidjećete zvjezdice koje se koriste za označavanje statističke značaja. Uobičajena praksa je da \* znači značajan na nivou .05, a \*\* znači značajan na nivou .01. Ipak, isplati se pročitati sitno štampani tekst i biti siguran. [↑](#footnote-ref-6)
37. Ovaj isti uvid odnosi se i na ukrštanje. Veza između dve ordinalne varijable može imati Kendalov tau od 0.08 i biti statistički značajna, dok drugi par varijabli iz manjeg skupa podataka ima Kendallov tau od 0.13 i nije statistički značajan. [↑](#endnote-ref-31)
38. Ova vrsta testa je češća u psihologiji i obrazovanju, gdje bi istraživači mogli da uporede rezultate muškaraca i žena, dječaka i djevojaka, crnaca i bijelaca itd. na bilo koji broj kontinuiranih mjera (npr. test čitanja za treći osnovne, skala introvertnosti-ekstrovertnosti). [↑](#endnote-ref-32)
39. Prema ANES-u iz 2000, postojala je statistički značajna razlika između toga šta su muškarci i žene osjećali u pogledu Hilari Klinton, ali to je bila prilično mala razlika. U prosjeku, žene su je gledale pozitivnije, ali su postojale značajne razlike u mišljenju među ženama i među muškarcima. [↑](#endnote-ref-33)
40. Dakle, osnovna logika je slična testu razlike aritmetičkih sredinama. [↑](#endnote-ref-34)
41. Putnam, Making Democracy Work, poglavlje 4. [↑](#endnote-ref-35)
42. Političari se ne slažu u vezi toga koliko je nezavisnih varijabli prikladno za model regresije. Neki preporučuju gornju granicu od tri, na osnovu teoretskih i statističkih ograničenja, dok mnogi drugi smatraju da broj može i treba da bude veći. Da bistestekli jasan uvid u ove argumente, pogledajte zimsko izdanje časopisa Conflict Management and Peace Science iz 2005, naročito članke Džejmsa Li Reja, Džona O'Nila i Brusa Raseta i Kristofera Ejhena. [↑](#endnote-ref-36)
43. Ubacivanje bilo kakve nezavisne varijable, koliko god teoretski irelevantne, može povećati vrijednost R-kvadrata modela višestruke regresije. Samo slučajno će se neke od varijacija u toj varijabli podudarati sa nekim varijacijama zavisne varijable. [↑](#endnote-ref-37)
44. Da li 19 procenata predstavlja značajan rezultat? To je pitanje lične percepcije. U nekim dijelovima ove discipline, naučnici bi bili zadovoljni ako njihov model predstavljao i 5 procenta varijacije u zavisnoj varijabli, sve dok su neke od njihovih nezavisnih varijabli statistički značajne. Na kraju krajeva, politika je često komplikovana, a mi ćemo možda samo povremeno izaći kao pobjednici. Drugi politikolozi bi pogledali ove rezultate i zaključili da su izostavljene varijable interesantnije od ponuđenh varijabli. [↑](#footnote-ref-7)
45. Isto tako, standardna greška procjene (SEE) jedva da se mienja. Neki politički analitičari skeptični su u pogledu R-kvadrata i prilagođenog R-kvadrata i preferiraju korišćenje SEE -a kada evaluiraju cijeli regresioni model. [↑](#footnote-ref-8)
46. U svrhu ove vježbe, napravio sam korak u nepoznato u razmatranju varijable koja se tiče registracije birača kao intervalne mjere, u rasponu od 1 (vrlo niski zahtjevi) do 5 (vrlo visoki). [↑](#endnote-ref-38)
47. Neki političari se udaljavaju od standardizovanog koeficijenta, dijelom zato što ne vole da izgube dragocjene informacije, kao što su jedinice svake varijable. [↑](#endnote-ref-39)
48. Pol je nominalna mjera u ovom modelu. Sa samo dvije vrijednosti, on bi se mogao smatrati dihotomnom ili „lažnom“ varijabla.Čini se da su žene obilježavane sa 0, a muškarci 1. Pošto pol ima samo dvije vrijednosti, teže je ovom metodom otkriti statistički značajniju vezu. To objašnjava zašto neki od većih *B* koeficijenata za pol nisu statistički značajni u tabeli 7.3. [↑](#endnote-ref-40)
49. Politički analitičari sve više prikazuju svoje statističke rezultate grafikonima, brojkama i mapama umjesto tabelama sa brojevima. Vizuelni prikaz informacija je važna tema u društvenim naukama uopšte, ali ne i tema za koju imam vremena u ovom vodiču. Za početak, pogledajte Jonathan P. Kastellec i Eduardo L Leoni, "Using Graphs Instead of Tables in Political Science" *Perspectives on Politics* 5, br. 4 (decembar 2007.): 755-71. [↑](#endnote-ref-41)
50. Michael D. Jones, "Leading the Way to Compromise? Cultural Theory and Climate Change Opinion," PS. Political Science and Politic 44, no. 4 (October 2011): 720-25. [↑](#endnote-ref-42)
51. Ibid. [↑](#endnote-ref-43)
52. Ross, "Does Oil Hinder Democracy?" [↑](#endnote-ref-44)
53. Verovatno ste mogli posumjati, kad god sam počinjao da pričam o građevinskim inspektorima i Glazgovskoj skali za komu. [↑](#footnote-ref-9)
54. „Ne vidim nijedan slučaj u kome su jednačine regresije, a kamoli složene metode uspjele da budu pokretači za otkrivanje uzročnih odnosa." David Freedman, ""From Association to Causation via Regression", *Advances in Applied Afathematia* 18, br. 1 (januar 1997.): 60. Slično tome, primjetite kako su dva politikologa-veterana, Jan Lilii Džonatan Negler, opisali svoju statističku analizu izlaznosti glasača: „Prilikom procjene multivariabilnog modela mi opisujemo 'zavisnu' varijablu onu koja je 'prouzrokovana'... eksplanatornom opisnom varijablom. Međutim, pretpostavka uzročnosti nije zasnovana na analizi podataka ili modelu koji je naveden, već na modelu realnog svijeta, koji pretpostavlja analitičar. Analiza podataka samo pokazuje uslovne odnose među posmatranim podacima, a ne uspostavlja uzročnost. U skladu sa konvencionalnom upotrebom iste, pomenućemo *marginalni efekat* jedne od naših deskriptivnih ... varijabli izlaznosti. Međutim, mi ne tvrdimo da takvi opservirani dokazi ukazuju na uzročnost. (Ilan Leighlei i Jonathan Nagler, *Who Votes Now?* Demographics, Issues; Inequality, and Turnout in the United States (Princeton, NJ: Princeton University Press, 2014.), str. 57). [↑](#endnote-ref-45)
55. Dara Kay Cohen, "Explaining Rape during Civil War: Cross-National Evidence (1980-2009)," *American Political Science Review* 107, br. 3 (avgust 2013.): 461-77. [↑](#endnote-ref-46)
56. Henry E. Brady, "Data-Set Observations versus Causal-Process Observations: The 2000 U.S. Presidential Election," u *Rethinking Social Inquiry: Diverse Tools; Shared Standards*; ed. Henry E. Brady i David Collier (Lanham, MD: Rowman i Littlefield, 2004.), str. 267-71. [↑](#endnote-ref-47)
57. Putnam, *Making Democracy Work*. [↑](#endnote-ref-48)