



# Ko učinkovita cepiva naletijo na statistični paradoks Simpsonov obrat *Pri izpeljavah in preračunih je potrebna dobršna mera previdnosti – Pogledati je treba več kot le številke*

**Pisalo se je leto 1973 in na univerzi Berkeley v Kaliforniji so imeli problem. Na podiplomski študij so sprejeli le 35 odstotkov prijavljenih kandidat, medtem ko je bilo sprejetih kandidatov 44 odstotkov. Takšna razlika ne more biti posledica naključja. Če vzamemo, da so moški in ženske enako sposobni, se kar sam ponuja edini možen zaključek, da je univerza ravnala diskriminatorno. Zaradi tega si je tudi nakopala tožbo.**

## MATEJ HUŠ

Sprejem kandidatov je v pristojnosti posameznih oddelkov, zato so na univerzi poglobljeno pogledali, kdo so krivci za diskriminacijo žensk. Izkazalo se je, da hujših prestopnikov ni bilo. Nekateri oddelki so sprejeli več kandidatov, drugi pa kandidatov, a večjih odstopanj ni bilo. Kaj se je zgodilo?

Univerza Berkeley je na lastni koži izkusila statistični pojav, ki je znan že več kot stoletje. Dve leti pozneje so raziskovalci natančno pojasnili, kaj se je zgodilo v tem primeru. Nekateri oddelki imajo zelo priljubljene študijske programe, za katere je zanimanje veliko, odstotek sprejetih pa nizek. Takšni programi so na primer družbene vede. Druga skrajnost sta matematika in tehnika, kjer je prijav tako malo, da sprejmejo skoraj vse. Na Berkeleyju se je zgodilo, da so se ženske večinoma prijavljale na zelo kompetitivne programe, zato je bilo sprejetih manj, moški pa v glavnem na manj oblegane. Četudi so oddelki oba spola sprejemali uravnoteženo – opazili so celo rahlo favoriziranje študentk – je bilo med vsemi prijavljenimi študentkami na vse smeri več neuspešnih. Preprosto povprečenje in računanje odstotkov ni le neuporabno, temveč izkrivi sklepe.

Analiza je pokazala, da univerza ni diskriminirala pri vpisnem postopku. To pa ne pomeni, da diskriminacije ni. Raziskovalci so v sklepu jasno zapisali, da se opažena diskriminacija zgodi že prej v izobraževalni vertikali in družbi. Očitno obstajajo dejavniki, ki ženske usmerjajo proti programom, ki so bolj kompetitivni, slabše fi-

nancirani in imajo pogosto slabše zaposlitvene možnosti.

## Paradoks, ki to ni

Berkeleyju se je zgodil Simpsonov paradoks. Ta je dobil ime po britanskem statistiku Edwardu Hughu Simpsonu, ki je leta 1951 napisal zelo suhoparen članek o tem pojavu z naslovom *Interpretacija interakcij v kontingenčnih tabelah*. Simpson ni prvi naletel na takšno obnašanje. Karl Pearson je leta 1899 raziskoval plodnost pri ljudeh in tekmovalnih konjih. Kakor Simpson je tudi Pearson opazil, da določene korelacije izginejo, če povprečimo celotno populacijo in na njej opazujemo lastnosti, namesto da bi analizirali posamezne podskupine posebej.

Svoje ime pa je Simpsonov paradoks dobil šele leta 1972, ko je kanadski statistik Colin Ross Blyth ugotovil, da včasih korelacije ne izginejo, temveč se obrnejo. Tedaj s povprečenjem po celotni populaciji pridemo do sklepa, ki ni resničen. Pojav je poimenoval Simpsonov paradoks, čeprav pri njem ni nič protislovnega – je zgolj tako neintuitiven, da se zdi človeškemu možganom nemogoč. Pravilneje ga zato imenujemo Simpsonov obrat.

Z njim so se odtlej ukvarjali številni statistiki, ki danes dobro razumejo, kdaj se pojavi, kaj pomeni in kako pravilno interpretirati rezultate. Simpsonov paradoks tudi ni napaka statistike, temveč preprosto dejstvo, da za pravilno statistično obravnavo potrebujemo razumevanje pojava, ki ga analiziramo.

## Od Tokia do Teksasa

Simpsonov obrat se pojavi, kadar na populacijo vpliva neka tretja spremenljivka, na katero pri analizi nismo pomislili. Čeprav mora za njegov obstoj veljati nekaj pogojev, nikakor ne gre za statistično kuriozitetu, ki bi se v resničnem življenju pojavljala redko. Nasprotno, nekateri statistiki pravijo, da bo zaradi čedalje bolj heterogenih družb postajal celo pogostejši.

Neškodljiv primer si lahko ogledamo na olimpijskih igrah, če analiziramo tekme slovenske košarkarske reprezentance v sku-

pinskem delu. Če seštejemo vse mete na koš (za dve točki, trojke in proste mete), je Žiga Dimec zadel 54 odstotkov metov (7 od 13), Zoran Dragič pa 57 odstotkov (20 od 35). A če uspešnost razčlenimo na tekme, je statistika drugačna. Proti Argentincem je imel Dimec 37,5-odstotni met (3/8), Dragič pa 33-odstotnega (4/12). Na drugi tekmi je Dimec vrgel enkrat in zadel (100 %), medtem ko je bil Dragič 75-odstoten (12/16). Na tretji tekmi je Dimec metal 75-odstotno (3/4), Dragič pa 57-odstotno (4/7). Dimec je bil na vsaki tekmi natančnejši, toda agregirano po vseh tekmah je Dragič zadel večji odstotek metov. To je seveda zgolj statistična zanimivost, ki ne pove kaj dosti o igri posameznega igralca, saj ima košarka cel kup elementov. Dobro pa pokaže, da moramo biti previdni.

Lahko pa ima Simpsonov obrat zelo škodljive vplive, če ga ne prepoznamo. V minulem desetletju se je v ZDA vnela živahna razprava o kakovosti izobraževalnega sistema. Leta 2009 so pri naravoslovju osmošolci iz Teksasa na nacionalnih testih dosegli v povprečju 150 točk, osmošolci iz Wisconsin pa 157. Podobni rezultati so bili tudi pri drugih predmetih, po čemer so nekateri sklepali, da bi morali v Teksasu uvesti čim več prijemov iz očitno kakovostnejšega Wisconsin. In se zmotili.

Učenci niso monolitna gmota. Uspešnost na ameriških standardiziranih testih je zaradi kompleksnih socioekonomskih dejavnikov korelirana z rasno oziroma etnično pripadnostjo. Razčlenimo. V Teksasu so beli učenci v povprečju dosegli 167 točk, v Wisconsinu pa 165. Temnopolti so v Teksasu dosegli 133 in v Wisconsinu 120 točk. Latinoameričani so v Teksasu dosegli 141 točk, v Wisconsinu pa 134. Osmošolci z azijskimi koreninami so v Teksasu dosegli 170 točk, v Wisconsinu pa 152. Vsaka skupina je bila v Teksasu uspešnejša. V Wisconsinu je skupno povprečje višje preprosto zato, ker ima drugačno etnično sestavo, predvsem več belcev. Iz podatkov pa ni možno sklepati, da je njihov izobraževalni sistem boljši – morda je bil celo slabši. To seveda ne pomeni, da so azijski učenci pame-

tnejši od temnopoltih. Obstaja cela vrsta socioekonomskih razlogov, zakaj se dosežki v šolah tako zelo razlikujejo med rasami, kar ameriški statistični urad seveda ve in zato objavlja rezultate tudi ločeno. Kateri graf in kakšno utemeljitev pa bi pokazal politik, je verjetno odvisno od tega, kaj zagovarja.

#### »Paradoksov« je več

Simpsonov paradoks je zgolj najočitnejša in najbolj didaktična past, ki preži na rokohitrsko premetavanje števil in odstotkov, še zdaleč pa ne edina. Vzemimo primer, ki ni Simpsonov paradoks, a je prav tako neintuitiven. Predstavljajmo si – pri čemer vsaka podobnost z resničnimi dogodki ni naključna – da človeštvo napade nova bolezen. Zavoljo ilustrativnosti predpostavimo, da vsi okuženi potrebujejo bolnišnično zdravljenje. Leto pozneje človeštvo razvije cepivo, ki z 90-odstotno učinkovitostjo preprečuje okužbo. To pomeni, da če imamo dve enaki skupini in eno cepimo, druge pa ne, bo v prvi 90 odstotkov manj primerov bolezni.

Pol leta pozneje je na prvi strani časnika naslov *Polovica hospitaliziranih kljub cepljenju*. Mikavno bi bilo planiti k zaključku, da cepivo ne deluje, a za kakršenkoli sklep manjka ključen podatek, koliko ljudi se je cepilo. Če se je cepilo 91 odstotkov ljudi, je pri 90-odstotni učinkovitosti cepiva matematično dejstvo (v idealiziranem modelu), da bo polovica okuženih iz skupine cepljenih. Morda to ni intuitiven rezultat, a se da razmeroma enostavno in neizpodbitno izračunati. To pa ne pomeni, da cepivo ne deluje. V takem primeru

bi bilo v bolnišnicah 91 odstotkov cepljenih in 9 odstotkov necepljenih (enako kot delež v populaciji). Tako pa je zgolj 9 odstotkov necepljenega prebivalstva prispevalo kar polovico hospitalizacij.

Statistika in sklepi se še dodatno zapletejo, če vključimo različen potek bolezni v različnih starostnih skupinah, različno učinkovitost cepiv in različno precepljenost različnih starostnih skupin. Vse to nas zlahka zapelje v popolnoma napačne sklepe, čeprav imamo na voljo vse podatke.

Zadnji aktualni primer je statistika hospitalizacij zaradi covida v Izraelu, ki je po internetu krožila v različnih sporočilih kot »dokaz«, da cepljenje ne deluje. Po uradnih podatkih je bilo 15. avgusta v Izraelu skoraj 60 odstotkov hospitaliziranih cepljenih. Kasneje je širši javnosti profesor biostatistike s pensilvanske univerze Jeffrey S. Moris v dolgem poljudnem članku pojasnil, da se sklepi povsem spremenijo, če upoštevamo še naslednje dejavnike: visoka precepljenost, neenakomerna precepljenost po starosti, tako rekoč stoddostna precepljenost starejših, zanemarljiva precepljenost mladih, značilno različen potek bolezni pri starejših in mlajših. Pravilna analiza pokaže, da so cepiva še vedno zelo učinkovita.

Epidemija covida je kompleksen pojav, pri katerem smo vsakodnevno bombardirani s podatki, številkami in statistiko iz najrazličnejših virov. Medtem ko surovim številkam kaj dosti ne moremo oporekati, je pri izpeljavah in preračunih potrebna dobršna mera previdnosti. Tudi na internetu

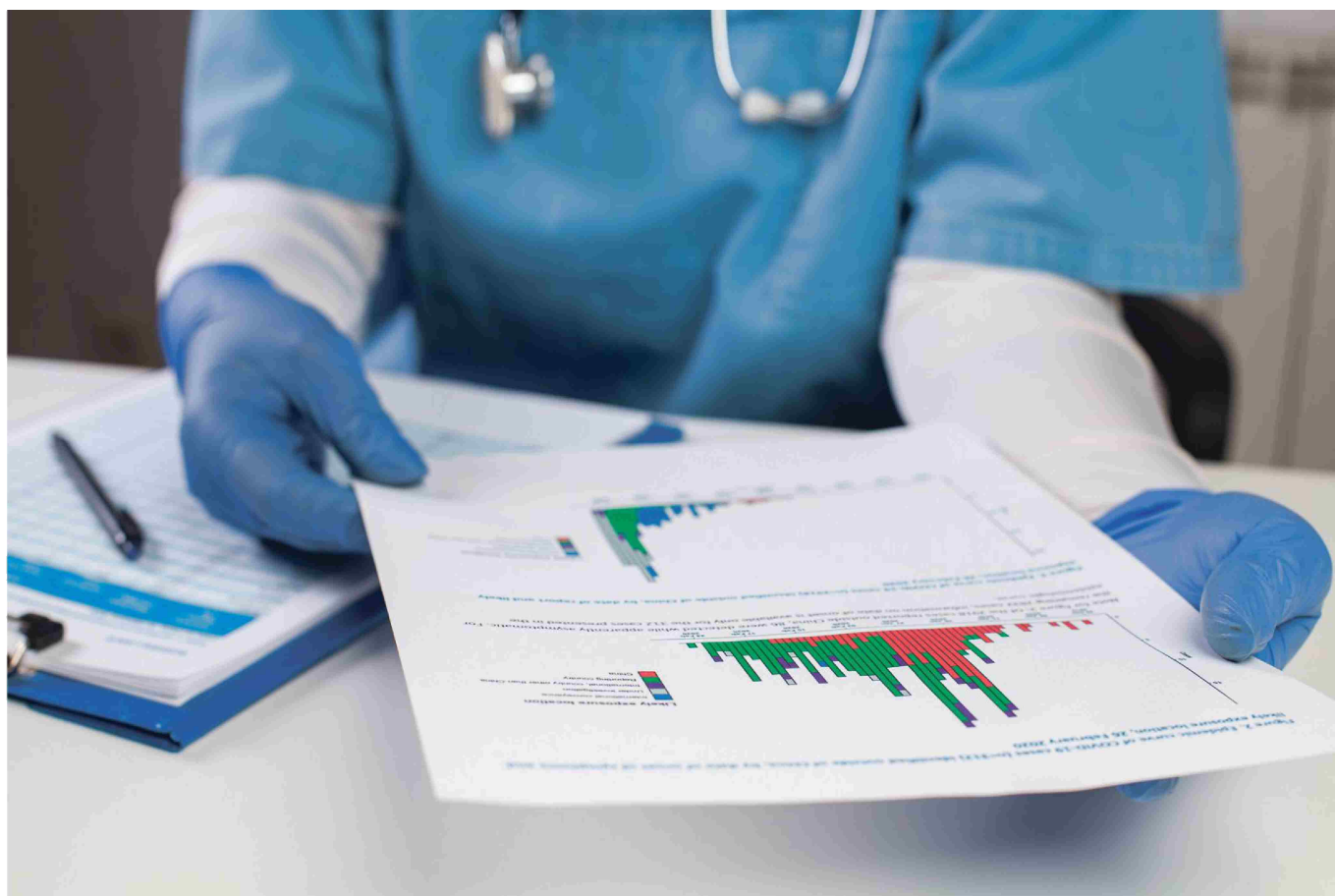
najdemo diamante, če jih znamo poiskati in obrusiti, torej razumeti, a to zahteva obilico časa, znanja in izkušenj. Epidemiologi in statistiki v uradnih institucijah, ki se s tem početjem ukvarjajo vso kariero, so zagotovo prijaznejši verodostojen vir.

Dr. Matej Huš je kot znanstveni sodelavec zaposlen na **Kemijskem inštitutu**.

*Simpsonov paradoks ni napaka statistike, temveč preprosto dejstvo, da za pravilno statistično obravnavo potrebujemo razumevanje pojava, ki ga analiziramo.*

*Pri paradoksu oziroma obratu določene korelacije izginejo, če povprečimo celotno populacijo, namesto da bi analizirali posamezne podskupine posebej.*

*Preprosto povprečenje in računanje odstotkov ni le neuporabno, temveč izkrivi sklepe.*



Različni parametri nas zlahka zapeljejo v popolnoma napačne sklepe, čeprav imamo na voljo vse podatke. FOTO SHUTTERSTOCK



Simpsonov paradoks je zgolj najočitnejša past, ki preži na rokohtersko premetavanje števil, še zdaleč pa ne edina. FOTO SHUTTERSTOCK