

Opet ta statistika

Again with the statistics

Gordana Marić-Brumini

Uredniku:

Možete li zamisliti biomedicinsko istraživanje, primjerice RCT (engl. *Random Control Trial*) bez primjene statističkih metoda? Već odabir varijabli i mjerenje njihovih vrijednosti zahtijeva poznavanje mjernih ljestvica i poznavanje pogrešaka mjerenja, a da važnost prikaza rezultata i kritičkog promišljanja rezultata ni ne spominjemo. Izaberemo li i manje spektakularno istraživanje, kvalitativno istraživanje, za koje će neki reći kako se radi o jednostavnom prebrojavanju neke pojave ili nekih entiteta, opet se susrećemo sa statističkim metodama; od prikupljanja podataka, prikaza rezultata do statističke analize podataka. Neupitno je kako se moderna znanost svakodnevno susreće s gomilom podataka, dok istovremeno računalna tehnologija potpomognuta računalnim potporama omogućuje pohranu i laku dostupnost podacima. No s tim bogatstvom podataka često se nađemo u nedoumici u odabiru statističkih metoda koje će nam pomoći dati odgovore na postavljena pitanja. I evo nas, za tren, ponovno u statistici.

Nadamo se da ćete nakon pročitano g članka, ako ne zavoljeti statistiku, onda potvrditi postojeća znanja o statistici, možda naučiti nešto novo i naći pokoji razlog zašto je statistika važna i lijepa.

POČECI STATISTIKE

Kada su svećenici u župnim uredima (u stoljećima prije našeg) bilježili broj rođenih i umrlih ili pratili i bilježili broj oboljelih od neke bolesti kroz vrijeme nisu mogli ni zamisliti kako su njihove evidencije predstavljale nešto što mi danas nazivamo statističkim pregledima. Pišući o počecima statističkog načina mišljenja zgodno je spomenuti probleme kockara u nastojanju da pronađu ponavljanja određene karte ili broja. Možda pravi početci statistike nastaju u trenutku kada su se matematičari zainteresirali za njihove probleme i pokušali te ugrađene slučajnosti ponavljanja objektivno i pravilno protumačiti. S vremenom, posebice poslije Drugoga svjetskog rata, geometrijska progresija znanstvenih spoznaja u prirodnim i društvenim područjima uključuje statistički način mišljenja u rješavanju problema i na neki način uvjetuje razvoj zasebne znanstvene discipline primijenjene matematike – statistike. U moru različitih definicija o tome što je to statistika, najjednostavnije i najslikovitije statistiku opisuje definicija matematičara i statističara Borisa Petza: Statistika je obrada brojčanih podataka radi jasnijeg prikazi-

Dopisni autor:

Izv. prof. dr. sc. Gordana Marić-Brumini
Zavod za medicinsku informatiku
Medicinski fakultet Sveučilišta u Rijeci
Braće Branchetta 20, 51 000 Rijeka
e-mail: gordana.brumini@medri.uniri.hr

<http://hrcak.srce.hr/medicina>

vanja. (Nematematičari i nestatističari vjerojatno se neće složiti s ovim „jasnije”, ali razumljivo nam postaje onog trena kada znamo što i zašto radimo.)

PRIMJENA STATISTIKE

Danas je statistička metodologija dio znanstvene kulture svakog znanstvenika i postoji u baš svakom djeliću znanosti. U znanstvenim istraživanjima tri su osnovne razine primjene statistike: praćenje stručne i znanstvene literature, statistička obrada rezultata koja uključuje prikaz rezultata i analizu, te zaključivanje i poopćavanje rezultata. Čitajući stručnu i znanstvenu literaturu zasigurno ste naišli na čitav niz precizno određenih statističkih termina i simbola: aritmetičku sredinu, median, koeficijent korelacije, granice pouzdanosti... Kako živimo u radnom svijetu koji u biomedicini zahtijeva cjeloživotno obrazovanje, praćenje struke i znanosti nezamislivo je bez razumijevanja pročitanog koje neizbježno i barem dijelom sadrži statistiku.

Mjerenjem neke pojave, događaja ili obilježja dobivamo podatke temeljem kojih je teško nešto zaključiti i donijeti valjane odluke. Stoga, kako bi podaci postali korisni, slijedi deskripcija i analiza rezultata. Statističkom obradom prikupljenih podataka dobivamo rezultate s kojima želimo upoznati širu akademsku ili neakademska zajednicu, polazeći od možda najvažnijeg cilja znanosti – njene uporabivosti. Često se dogodi da početnik s velikom marljivošću i srcem skupi mnoštvo podataka s kojima ni sam ne zna što bi ih uredi i prikaže nepregledno i nerazumljivo, pa mukotrpan posao završi u ladici. Potrebno je znati opisati podatke tako da ih uredimo i sažmemo kako bi oni bili što pregledniji, pri čemu vodimo računa o vrsti podataka i onome što želimo prikazati. Primjerice, učestalost objavljivanja znanstvenih ili stručnih radova tijekom proteklih pet godina možemo prikazati aritmetičkom sredinom i standardnom devijacijom u tablici ili u stupčastom dijagramu. Koji ćemo odabrati ovisi prvenstveno o tome s kojom svrhom to istražujemo. Tablice i slikovni prikazi dva su osnovna načina prikaza rezultata, jasno prenose poruku ako su jasne, jednostavne, pregledne, samorazumljive i ako im je svrha jasno određena. Slikovni prikazi koje vidamo u bolnič-

kim hodnicima i čekaonicama u zdravstvenim ustanovama daju korisne informacije ako ih razumijemo. Jedan od empirijskih kriterija „dobre” tablice ili slikovnog prikaza je njihova razumljivost i osobito osobama koje se ne bave biomedicinom i potpuno im je daleko i strano to što mi radimo.

U analitičkoj statistici problem najčešće nastaje prilikom odabira testa. Hoćemo li napraviti jednosmjernu analizu varijance (engl. *one way ANOVA*), Mann-Whitneyjev test ili ROC analizu (engl. *Receiver Operating Characteristic*) ovisi o beskonačno mnogo čimbenika i zahtijeva od nas, osim pukog statističkog znanja, pravila temeljem kojih odlučujemo o izboru testa, interpretaciju rezultata s obzirom na statistička obilježja testa i osobine naših podataka. Primjerice, multiplu regresijsku analizu možemo koristiti u ekonomiji za određivanje čimbenika koji utječu na poboljšanje proizvodnog procesa jednako vrijedno kao i kada istražujemo pridonose čimbenika na globalnu DNA metilaciju. Početni kriteriji pri odabiru statističke metode su vrsta podataka (brojčani ili kategorički podaci) i vrsta uzorka (nezavisni ili zavisni). Kada se radi o brojčanim podacima sljedeći kriterij odabira je određivanje normalnosti raspodjele (što ne znači da su podaci koji nisu normalno raspoređeni ljudi). Za normalno raspodijeljene podatke koristit ćemo parametrijske testove, a za podatke koji ne slijede normalnu raspodjelu neparametrijske. „Bolje” je kada su podaci normalno raspodijeljeni jer odabir ma kojeg parametrijskog testa pridonosi većoj snazi istraživanja. Kao i sve znanosti i statistika prati modu, pa ćete često naići na upit recenzenta vašeg rada o snazi testa. Poželjno je da se u biomedicinskim istraživanjima snaga testa kreće od 70 – 80 %. Odabir statističke metode uvijek ima i neki ali. Kada imamo mali uzorak ($n \leq 30$), neovisno postoji li normalna raspodjela podataka ili ne, rabićemo neparametrijski test.

Zaključivanje i poopćavanje temeljem rezultata dobivenih na uzorku na populaciju vjerojatno je najzahtjevniji ali i najkvalitetniji ishod primjene statističkog znanja. Nažalost, pri statističkom zaključivanju uvijek možemo napraviti dvije vrste pogrešaka: zaključiti da je naša statistička hipoteza potvrđena kada uistinu nije i suprotno, zaključiti da postavljena statistička hipoteza nije potvrđena kada uistinu jest. Zašto bi statistika bila jednostav-

na? U statističkim zaključivanjima 100 %-tna je sigurnost nemoguća, stoga kod prikazivanja naših rezultata rabimo 95 % granice pouzdanosti i kriterijska vrijednost P parametra za sve statističke testove u biomedicinskim istraživanjima iznosi 0,05. Kada nešto zaključimo uz razinu $P < 0,05$, znači da smo u 95 % slučajeva ispravno zaključili, ali u 5 % slučajeva naš zaključak je pogrešan. Općenito vrijedi pravilo da će statističko zaključivanje biti ispravnije što imamo veći broj podataka.

Poznavajući pravilnu primjenu deskriptivne i/ili analitičke statističke metode, promišljajući statistiku u svoj njejoj širini možemo se vratiti na početak i krenuti planirati istraživanje vodeći pritom posebice računa o veličini i vrsti uzorka i željenoj snazi testa.

(ZLO)UPORABA STATISTIKE

A je li statistika pouzdan alat? Pomoću statistike možemo svjesno ili nesvjesno ne pisati istinu, lagati. Pravilna primjena statističkih metoda bilo korištenjem jednostavnih potpora kao Excel, bilo uporabom specijaliziranih statističkih paketa, omogućuje brz i lak način statističke obrade podataka, ali istovremeno zahtijeva od korisnika poznavanje statistike. Stoga smatramo kako statističke neznalice nenamjerno lažu. Često u želji veće znanstvene kvalitete posežemo za statističkim metodama koje ne znamo rabiti, pa su dobiveni rezultati besmisleni a njihovo tumačenje nekorisno. Daleko su opasniji statistički znalci kojima su vlastiti interesi ispred struke i morala. Vješto se koriste statističkim vještinama i znanjem. Primjerice, kada se podaci prikupljaju pomoću anketa ili upitnika, kod slabog odziva ispitanika ili nepotpuno ispunjenih anketa statistički znalci znaju kako izmisliti „dobre“ podatke.

Svojim znanjem dobivene rezultate svjesno interpretiraju na očekivan i/ili željen način. Pri tome, kao svaki lažac, nalaze brdo opravdanja za takvo ponašanje. Prepravljajući ustroja istraživanja ili lažno i upitno tumačenje rezultata, zbog pritiska financijera, kao i neobjavlivanje rezultata svojih istraživanja koji su u oprečnosti s njihovim prijašnjim rezultatima, smatraju potrebnim i opravdanim. U rukama nemoralnih i neetičkih znanstvenika (ako ih uopće možemo nazivati znanstvenicima) statistika postaje ubojito oružje znanosti.

ISKUSTVA IZVRŠNE UREDNICE ZA STATISTIKU

Pri pisanju znanstvenog ili stručnog članka, nažalost, još uvijek se može naići na rad koji ne primjenjuje stečena znanja i vještine o pravilima znanstvenog pisanja i/ili ne slijedi osnovne upute časopisa pa tako ni upute na koji način prikazati podatke i statistički ih obraditi. Statističar ne može znati jesu li podaci izmišljeni ili lažni. Može prosuditi o kvaliteti prikaza rezultata, o uporabi učinkovite statističke metode i o interpretacije rezultata. Autori često, zbog nedostatnog statističkog znanja, pri objašnjavanju rezultata koriste pojmove i termine koji nisu točni ili loše pojašnjavaju dobiveno. Primjerice, kada utvrđujemo postoji li razlika u tjelesnoj masi između muškaraca i žena koji pripadaju u dobnu skupinu od 60 do 70 godina, koristimo t -test za aritmetičke sredine. Pri tome rezultat testa i pripadna P vrijednost određuju postoji li značajna razlika, ali ne znamo koji je razlog te razlike. Ali ako napravimo multiplu regresijsku analizu s prediktorima (čije su korelacije bile preduvjet da ih uvrstimo u regresijski model) i ako za visinu utvrdimo da je značajni prediktor, tada znamo da je visina varijabla koja značajno pridonosi tjelesnoj masi. Statistički test biramo ovisno o postavljenoj statističkoj hipotezi i rezultat primijenjenog statističkog testa daje jasan, jednostavan i jednoznačan odgovor: da ili ne.

Baveći se znanošću učimo znanstveno misliti, raditi i pisati. Uz znanstveno obrazovanje i stručnost jasno nam je da znanstveno iskustvo pridonosi kvaliteti znanstvenog pisanja. Ne smatramo pritom kako znanstveni početnici ne znaju, lijeni su i nepismeni, već želimo naglasiti važnost mentorske uloge koja se često nepravedno marginalizira u dobrom i/ili u lošem. Dobar mentor odgovorno sudjeluje u svim fazama rada, savjetuje, prati, provjerava, sluša prijedloge mladog znanstvenika, a katkada i svjesno propušta pogreške kako bi se na njima učilo te pritom nesebično pri-laže svoja znanja i znanstvena iskustva.

ZAKLJUČAK

Složit ćemo se da bez statistike ne ide. Poznavanje statističkih metoda razumijeva njihovu pravilnu i umješnu uporabu. Statistički se mogu obraditi svi podaci: dobri i loše prikupljeni i izmišljeni.

Stoga statistička obrada podataka ne osigurava znanstvenu vrijednost podataka. Statistika ne jamči da je istraživanje pravilno provedeno. Statistika je pouzdan alat ali statističko zaključivanje treba uključiti i zdrav razum. Znanstvenik treba biti pošten i skroman u odnosu na rezultate znanstvenog istraživanja.

Znanost od znanstvenika zahtijeva moralnost i etičnost. Namjerna zlouporaba statistike ili uporaba statistike bez statističkog znanja obmanjuje

znanstvenu zajednicu. I na kraju, dva su načina na koji se može živjeti: jedan je – kao da ništa nije čudo, drugi je – kao da je sve čudo (Einstein). No bez statistike u znanosti – ne možemo.

Izjava o sukobu interesa: autorica izjavljuje da ne postoji sukob interesa.

LITERATURA

1. Petz B. Osnovne statističke metode za nematematičare. III. dopunjeno izdanje. Jastrebarsko: Naklada Slap, 1997.