

Kolokvij 2 - Osnove teoretične statistike

20. maj, 2020

1. Naj bosta X_{1i} in X_{2i} dve meritvi na posamezniku i , $i = 1 \dots n$. Spremenljivki X_1 in X_2 naj bosta normalno porazdeljeni. Naj bo varianca ob obeh meritvah enaka. Izpeljite primerjavo med varianco za razliko povprečij teh dveh spremenljivk pri parnem t -testu in varianco pri t -testu za neodvisne vzorce na istih podatkih (kot smo to naredili na predavanjih); pri tem naj ρ označuje korelacijo med meritvama. S pomočjo te primerjave primerjajte testni statistiki teh dveh testov.

Na istih podatkih izvedete t -test za parne in t -test za neodvisne vzorce. Za vsakega od spodnjih primerov zapišite, kaj pričakujete glede velikosti in moči. Pri velikosti za vsak test posebej zapišite, kakšna je, glede na pričakovano ($\alpha = 0,05$). Pri moči zapišite, kateri izmed obeh testov ima večjo, nato pa oboje utemeljite (spodnjo tabelo si prepišite na list z rešitvami).

	velikost testa	moč testa	utemeljitev
$\rho > 0$ parni t -test t -test za neodvisna vzorca			
$\rho = 0$ parni t -test t -test za neodvisna vzorca			
$\rho < 0$ parni t -test t -test za neodvisna vzorca			

2. Raziskovalci na podganah raziskujejo učinkovitost zdravila. Vsakega od n parov podgan iz istega legla raziskovalec razdeli na (1) podgano, ki bo dobila placebo in (2) tisto, ki bo dobila preizkušano zdravilo. Po določenem času raziskovalec izmeri izboljšanje zdravstvenega stanja obeh podgan v paru. Odgovor označi kot 'pozitiven izid', če se je zdravstveno stanje bolj izboljšalo podgani z zdravilom, kot 'neodločen izid', če je stanje pri obeh približno enako in kot 'negativen' sicer. Zanima ga, ali je v populaciji verjetnost za pozitiven izid (p) različna od 0,5.

- (a) Zapišite, kako je porazdeljena spremenljivka, ki vas zanima.
- (b) Zapišite testno statistiko za posplošeno razmerje verjetij (Wilksov Λ) za testiranje $H_0 : p = 0,5$ proti $H_a : p \neq 0,5$. Komentirajte vse korake in z izračuni utemeljite vse izraze, ki ste jih pri tem uporabili. Povejte natančno, kako je ta testna statistika porazdeljena. (Kot znane lahko uporabite cenilke po metodi največjega verjetja že izpeljane na predavanjih/vajah.)
- (c) Zapišite, za kakšne vrednosti Λ bo test iz prejšnje točke zavrnil H_0 (majhne ali velike) in natanko komentirajte.
- (d) Izračunajte vrednost testne statistike za primer, ko je $n = 10$ in smo dobili 2 pozitivna izida ter 3 neodločene izide. Natančno zapišite statistični (ali zavrnilo H_0 , pomagajte si lahko s tabelami) in vsebinski sklep (kaj lahko rečemo o populaciji).

3. Na podatkih ste uporabili dvosmerni test z za en vzorec. Testirali ste ničelno domnevo, da je povprečje enako 0 in dobili 95% interval zaupanja $[0, 12]$, torej ravno na meji zavrnitve. V vzorcu je bilo 100 ljudi.

V nadaljevanju dobite idejo, da bi naknadno izračunali še moč testa (post hoc power). Pri tem za vrednost povprečja pod alternativno domnevo vzamete kar dejansko vrednost, ki ste jo ocenili na vaših podatkih. Izračunajte, kakšno moč dobite na ta način.

4. Modelske ocene kažejo, da se je v Sloveniji do sedaj z virusom SARS-COV-2 okužilo 0,5% ljudi. Ta rezultat želimo preveriti s pomočjo nacionalne raziskave, ki je delež do sedaj okuženih preverjala na reprezentativnem vzorcu 1300 posameznikov. Pri tem ocenjujemo, da je specifičnost uporabljenega testa 98% (verjetnost, da je test pozitiven, četudi posameznik ni bil okužen, je 2%), občutljivost pa 100% (verjetnost, da je test pozitiven, če je posameznik okužen).

V vzorcu je bilo pozitivnih 41 posameznikov.

- (a) Ocenite delež in število dejansko okuženih na podlagi vzorca nacionalne raziskave. Pri tem ne pozabite upoštevati specifičnosti (in občutljivosti) testa.
 - (b) Zapišite ničelno domnevo, ki jo želimo preveriti.
 - (c) Predlagajte testno statistiko, za katero poznate eksaktno porazdelitev pod ničelno domnevo. Zapišite to porazdelitev. Razložite, kako bi izračunali meje zavrnitve.
 - (d) Predlagajte še testno statistiko, za katero poznate asimptotsko porazdelitev (tj. normalno) pod ničelno domnevo. Za ta primer izračunajte dejanski meji zavrnitve za $\alpha = 0,05$.
 - (e) S pomočjo podatkov izračunajte vrednost testne statistike iz prejšnje točke in interpretirajte rezultat (naredite sklep).
 - (f) Izračunajte 95% interval zaupanja za delež okuženih, ki bo dualen testni statistiki, ki je bila dobljena s pomočjo asimptotske porazdelitve (točka (d)).
5. V prejšnji nalogi smo za preverjanje ničelne domneve obravnavali dva testa, enega z znano eksaktno in drugega z asimptotsko porazdelitvijo pod ničelno domnevo. Sedaj bi radi preverili lastnosti obeh testov še s simulacijami.
- Predlagajte načrt, kako bi se lotili simulacij, za večjo jasnost skicirajte tabelo, v kateri bi ob koncu predstavili rezultate (predvidite, kaj pričakujete in kratko utemeljite). Zapišite psevdo kodo, vendar le za jedno načrtovanih simulacij.