Uporaba testov, IZ

Nataša Kejžar

Povzetek

Test χ^2, goodness of fit

Naj bo O_i opazovano število primerov v i-ti celici kontingenčne tabele in E_i pričakovano število (tj. pričakovano število pod H_0).

Testni statistiki (za test χ^2)

$$T = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

in za posplošeni test razmerja verjetij za primer deležev

$$2 \log \Lambda = 2 \sum_{i=1}^{k} O_i \log \frac{O_i}{E_i}$$

sta asimptotsko ekvivalentni.

Izračun postane nestabilen, ko so E_i enake 0, pri drugi testni statistiki pa tudi, ko velja $O_i = 0$.

Večkratno preizkušanje domnev

Izračun napake I. vrste (tj. verjetnost za zavrnitev testa, čeprav H_0 drži), ko naredimo n neodvisnih testov (stopnja značilnosti 0,05; omejimo se na primer, ko zavračamo za velike vrednosti testne statistike):

$$\begin{aligned} P(T_1 > c_1 \vee T_2 > c_2 \vee \ldots \vee T_n > c_n) &= 1 - P(T_1 < c_1 \wedge T_2 < c_2 \wedge \ldots \wedge T_n < c_n) \\ &= 1 - P(T_1 < c_1)P(T_2 < c_2) \ldots P(T_n < c_n) \\ &= 1 - 0,95^n \end{aligned}$$

V primeru npr. 10 izvedenih testov, je verjetnost napake I. vrste tako $1 - 0,95^{10} = 0,401$.

Naloge

1. Raziskava javnega mnenja v ZDA je pokazala, da 47% ljudi podpira Busha, 44% ljudi pa Al Gore-a. V vzorec so vzeli 2207 ljudi. Testirajte pri $\alpha = 0.05$, ali se podpora obema kandidatoma razlikuje.

   a. Zapišite ničelno hipotezo.
   b. Zapišite rezultat raziskave javnega mnenja kot spremenljivko, porazdeljeno po multinomski porazdelitvi.
   c. Za multinomsko porazdelitev velja, da je $cov(X_i, X_j) = -np_ip_j$. Izračunajte populacijsko varianco za razliko v deležu ljudi, ki podpirajo Busha in ki podpirajo Al Gore-a.
   d. Testirajte ničelno hipotezo s pomočjo posplošenega testa razmerja verjetij.
   e. Testirajte ničelno hipotezo s pomočjo testa hi-kvadrat.

2. Geissler (1889) je preučeval zdravniške popise na Saškem in iz njih izbral podatke o razmerju po spolu.V datoteki **data_Geissler.csv** so podatki o številu moških potomcev v 6115 družinah z 12 otroki.
   Če je spol pri vsakem naslednjem otroku neodvisen od prejšnjih in so verjetnosti konstantne skozi čas, bi moralo biti število moških, ki se rodi v posamezni družini z 12 otroki porazdeljeno po binomski porazdelitvi (z 12 poskusi in znano verjetnostjo p). Če je verjetnost moškega potomca za vse družine enaka, potem je v datoteki 6115 binomsko porazdeljenih opazovanj.

   a. Preverite s statističnim testom, ali se podatki ujemajo s tem modelom (verjetnost za rojstvo dečka je 105/205).
   b. Zakaj je/ni model ustrezen?
   c. S simulacijami pokažite, da je model, ki ste ga dobili, za H_0 ustrezen.
   d. Spreminjajte verjetnosti za število moških potomcev za posamezno družino in poskusite, kako se obnaša model.

3. Na strani statističnega urada lahko dobimo podatke o rojstnih dnevih prebivalcev Slovenije. Ti podatki so shranjeni v datoteki **data_RD.csv**.

   a. Preberite datoteko v R in narišite stolpični diagram za rojstne dneve. Kaj je s prvim in z zadnjim datumom? Je to lahko plod naključja?
   b. Postavite ničelno domnevo: $H_0: $ vsi rojstni dnevi so enako verjetni
      Preverite jo. Kateri test boste uporabili? (pred tem izvzemite zadnji datum, ki je 29. februar)
   c. Izračunajte test na 2 načina, preverite njune vrednosti in izračunajte meje zavrnitve. Kakšen je sklep? Zakaj?
   d. Izvzemite najekstremnejše vrednosti iz podatkov in poskusite narediti test še enkrat. Komentirajte rezultate.
   e. Kako bi simulirali podatke, da bi pokazali, da ima test v primeru, ko H_0 drži, pravo velikost? (uporabite največ 100 ponovitev)

Za neko japonsko mesto vemo, da je imelo v letih 2000–2014 precej potresov. V tabeli je predstavljena opazovana pogostost. Stopnjo značilnosti nastavimo na $\alpha = 0.05$.

število potresov (k)	0	1	2	3	4	5
število let, v katerih je bilo k potresov	3	5	2	4	0	1

4. Za število potresov **vemo**, da se porazdeljuje približno po Poissonovi porazdelitvi. Radi bi pokazali, da je parameter za hitrost potresov λ za to japonsko mesto večji od 1.

   $$P(X = x) = \frac{\lambda^x \exp(-\lambda)}{x!}; \qquad X \sim Pois(\lambda),$$

   vsota neodvisnih Poissonovo porazdeljenih spremenljivk je spet porazdeljena po Poissonu ($Pois(\sum_{i=1}^n \lambda_i)$).

   a. Zapišite ničelno in alternativno domnevo za vaš primer. Komentirajte ju.

b. Za testiranje **ne bomo uporabljali aproksimacije**. Kaj bo torej vaša testna statistika (razložite, zakaj) in kako je ta statistika porazdeljena?
c. Določite območje zavrnitve.
d. Izračunajte testno statistiko za vaš primer in vsebinsko interpretirajte vaš rezultat.
e. Določite, kakšen tip napake bi (lahko) zagrešili, če bi prava vrednost parametra znašala $\lambda_1 = 0.8$ oziroma $\lambda_2 = 1.4$, zbran pa bi imeli enako velik vzorec. Podajte intuitivno razlago, nato pa še izračunajte obe verjetnosti.
f. Kako velik vzorec (in kaj je v našem primeru sploh enota v vzorcu?) bi morali nabrati, da bi lahko dokazali (z 90% močjo), da je λ za to japonsko mesto večji od 1, če je pravi $\lambda_{PRAVI} = 1.2$. Komentirajte rezultat glede na ugotovitve iz prejšnje točke.

5. Preverili bi radi, ali se število potresov porazdeljuje po Poissonovi porazdelitvi.

   a. Zapišite ničelno in alternativno domnevo za vaš primer. Komentirajte ju.
   b. Kaj bo vaša testna statistika (razložite, zakaj) in kako je ta statistika porazdeljena?
   c. Izračunajte testno statistiko za vaš primer in vsebinsko interpretirajte vaš rezultat.
   d. Kako se spremeni testna statistika in njena porazdelitev, če nas zanima, ali se število potresov porazdeljuje po Poissonovi porazdelitvi z $\lambda = 1$?
   e. S simulacijami preverite velikost testa za dva različna primera, ko generirate podatke iz $Pois(\lambda = 1)$:
      i. Število kategorij določite glede na generirane podatke.
      ii. Število kategorij je fiksno in je enako 5 (od 0 potresov do 4 ali več).
   f. Katera simulacija je pravilnejša in zakaj?

6. Zanima nas, ali je verjetnost, da se v enem mesecu zgodi potres, večja od 7% (pri tem zanemarimo, da se v enem mesecu lahko zgodita tudi dva potresa ali več, za naše izračune predpostavimo, da so potresi med sabo neodvisni).

   a. Zapišite ničelno in alternativno domnevo za vaš primer. Komentirajte ju.
   b. Kaj bo vaša testna statistika (razložite, zakaj) in kako je ta statistika porazdeljena?
   c. Izračunajte testno statistiko za vaš primer in vsebinsko interpretirajte vaš rezultat.

7. Lastnik trgovine označi vsak trgovalni dan kot dober ali slab. Vsak teden prešteje dobre dneve (trgovalni teden ima 6 dni). Podatki iz enega leta so v tabeli spodaj. Testiraj ($\alpha = 0,05$) ali so dobri in slabi dnevi enako verjetni?

   a. Kaj je ničelna domneva?
   b. Kaj je spremenljivka, ki nas zanima? Kako je porazdeljena?
   c. Izpeljite testno statistiko na podlagi posplošenega testa razmerja verjetij. Kako je porazdeljena?
   d. Izračunajte vrednost testne statistike in interpretirajte rezultat.
   e. Kakšna bi bila testna statistika, če bi nas zanimalo, ali je verjetnost dobrih dni v enem tednu enaka za vse "kategorije" tednov? Kaj bi bila ničelna in alternativna domneva?
   f. Kako je s predpostavkami obeh testov? Kaj lahko naredimo?

# dobrih dni	0	1	2	3	4	5	6
# tednov	1	9	12	13	11	5	1

8. V igralnicah neprestano preverjajo uravnoteženost ruletnega cilindra. Na ruletnem cilindru je 37 različnih števil (od 0 do 36), želimo, da se vse pojavljajo z enako verjetnostjo.

   a. Denimo, da imamo zabeležene vrednosti, ki so se pojavile v n poskusih. Naša ničelna domneva je H_0 : cilinder je uravnotežen. Zapišite testno statistiko, ki bi jo uporabili za preverjanje te domneve. Kako je porazdeljena?
   b. Poznate eksaktno ali le asimptotsko porazdelitev? Kratko utemeljite.
   c. Pri preverjanju za določen cilinder smo dobili $p = 0,005$. Se to lahko zgodi, tudi če je cilinder uravnotežen? Utemeljite.
   d. Pri preverjanju nekega drugega cilindra smo dobili $p = 0,23$. Ali lahko z gotovostjo trdimo, da je cilinder uravnotežen? Utemeljite.

e. V igralniški hiši je 100 cilindrov. Vsakega testirajo vsak dan na osnovi podatkov za tisti dan. Pravilo je, da natančneje preverjajo cilinder, za katerega je bila vrednost $p \leq 0,01$. Če bi bili vsi cilindri idealni, torej če bi za vsakega veljala ničelna domneva, kakšen odstotek cilindrov na dan bi kontrolirali v nekem daljšem obdobju?
f. Med cilindri v igralniški hiši je tudi pokvarjen cilinder. Ga bodo na dolgi rok kontrolirali manjkrat ali večkrat ko idealne cilindre, če je pravilo enako kot v prejšnji točki? Utemeljite.

9. Čas do trenutka, ko se žarnica pokvari, lahko modeliramo z eksponentno porazdelitvijo. Navadne žarnice naj bi v splošnem delovale v povprečju 1200 ur. Kontrolni oddelek nekega proizvajalca navadnih žarnic bi rad s statističnim testom ugotovil, ali so njihove žarnice splošno primerljive.
Za eksponentno porazdelitev velja: $f(x|\lambda) = \lambda e^{-\lambda x}; x \geq 0, \lambda > 0$, ter $E(X) = 1/\lambda$.

   a. Zapišite ničelno domnevo in prostor vrednosti parametrov pod H_0 in H_A.
   b. Kateri statistični test naj kontrolni oddelek uporabi? Izpeljite/zapišite testno statistiko in zapišite njeno asimptotsko porazdelitev.

10. Problem večkratnega testiranja hipotez. S spodnjo kodo generiramo podatke za 60 normalno porazdeljenih spremenljivk ($n = 20$). Preverjamo ničelne hipoteze, da je populacijsko povprečje enako 0. Prvih 5 spremenljivk je generirano s populacijskim povprečjem, ki je različno od 0, ostale imajo populacijsko povprečje enako 0.

   a. Preverite, koliko ničelnih hipotez boste zavrnili. Izpišite, koliko H_0 je zavrnjenih pri spremenljivkah, kjer bi se to moralo zgoditi in koliko pri tistih, kjer s tem naredimo napako prve vrste. b.Poženite to kodo vsaj 1000-krat in preverite, kaj se dogaja. Preverite, kakšno je število (in delež) napačno zavrnjenih testov.
   b. Uporabite popravek za večkratno testiranje (**p.adjust**) in preverite rezultate v tem primeru.

```
n = 20; mu = 4; sd = 3
# generiraj podatke - 60 spremenljivk
data=data.frame(V1 = 1:n)
for(j in 1:5)
  data[,j] = rnorm(n,mu,sd)
for(j in 6:60)
  data[,j] = rnorm(n,0,sd)
  # preveri s t-testom
testi = numeric(60)
for(j in 1:60)
  testi[j] = t.test(data[,j])$p.value
```

11. Psihologa zanima, zakaj ženske ponavadi obrnejo zemljevid tako, da se njihovo mesto ujema s smerjo, v katero so namenjene, medtem ko moški zemljevid ohranijo v prvotni legi in si samo predstavljajo spremembo smeri. Predpostavlja, da obstaja med spoloma razlika v zmožnosti miselnega obračanja objektov, zato izvede računalniški poskus, kjer meri odzivne čase na odgovore, ki so povezani z miselno vizualno rotacijo objektov. Časi (v sekundah) 15 žensk in 15 moških so v datoteki **data_orientation.csv**. Privzemite, da je standardni odklon odzivnih časov znan in je enak 0.028 sekunde.

   a. Kaj je ničelna domneva?
   b. Kaj je alternativna domneva? Je enostavna/sestavljena? Je enostranska/dvostranska?
   c. Razmislite, kaj bi bila vaša testna statistika in izpeljite njeno porazdelitev.
   d. Določite meje zavrnitve pri $\alpha = 0,05$.
   e. Določite vrednost p za vaš primer.
   f. Določite $(1 - \alpha)\%$ interval zaupanja.
   g. Statistično in vsebinsko interpretirajte rezultat. Ali je rezultat strokovno pomemben?
   h. Izračunajte $(1 - \alpha)\%$ intervala zaupanja za moške in za ženske posebej. Ali se prekrivata? Kakšen pa je interval zaupanja za razliko? Komentirajte.

12. Razmislite, ali bi bilo možno dobiti take podatke, kjer bi se intervala zaupanja za populacijski povprečji posameznih neodvisnih skupin prekrivala, interval zaupanja za razliko med njima pa bi kazal na

statistično značilen rezultat. **Komentirajte.** Preberite v R podatke iz datoteke **data_CI.csv**

a. Za podatke izračunajte 95% intervale zaupanja z znanim populacijskim standardnim odklonom $\sigma = 1/1.96$. Komentirajte.
b. Kaj pa, če izračunate interval zaupanja z vzorčnim standardnim odklonom? V kakšnem primeru bi bilo vzeti σ ali s skoraj vseeno?
c. Pokažite, zakaj in kdaj lahko pride do rezultata, da se posamezna intervala zaupanja prekrivata, interval zaupanja za razliko pa kaže na statistično značilen rezultat, če privzamete, da je $n_1 = n_2$ in $\sigma_1 = \sigma_2$.
d. Pokažite to še s simulacijami (izračunajte minimalno in maksimalno razliko in ju primerjajte s teoretičnimi vrednostmi).

13. Zanima nas, ali obstaja statistično značilna razlika pri teži žensk pred zanositvijo in njihovo težo eno leto po porodu. V ta namen so bili zbrani podatki **data_teza.csv**.

    a. Zapišite H_0 in H_A.
    b. Oglejte si podatke in izberite primeren statistični test. Rezultate interpretirajte.
    c. Poglejte, kaj bi se zgodilo, če bi namesto testa, ki ste ga uporabili, uporabili napačnega (torej, če ne bi upoštevali informacije o odvisnosti). Interpretirajte rezultate in teoretično utemeljite razliko z izračunom.
    d. Kaj pa, če uporabite neparametrične teste?
    e. Generirajte podatke pod H_A za test t (normalna porazdelitev spremenljivk) in preverite moč z neparametričnimi testi. Kako pogosto zavrne?

14. Radi bi raziskali, kako je z dualnostjo binomskega eksaktnega testa ($\alpha = 0.05$) in naslednjih 95% intervalov zaupanja (IZ)

    - Clopper–Pearsonov (eksaktni) IZ
    - aproksimativni IZ za delež (z normalno porazdelitvijo)

    a. Zapišite ničelno domnevo za dvostranski binomski eksaktni test.

    b. Zapišite, kratko izpeljavo za izračun aproksimativnega IZ za delež (t.j. intervala zaupanja, ki ga izračunate, če binomsko porazdelitev aproksimirate z normalno).

    c. V R definirajte funkcijo za izračun aproksimativnega IZ

    d. S simulacijami (10.000 ponovitev) preverite, v kakšnem deležu binomski eksaktni test in obe različici intervala zaupanja zavračajo H_0. Vzorce generirajte pod H_0 za naslednji različici parametrov. Deleže zavrnitev prikažite na enem grafu in rezultate vsake različice posebej utemeljite.

        - $n = 100, \pi = 0.5$ in $n = 20, \pi = 0.5$ ter
        - $n = 100, \pi = 0.08$.

    e. Simulirajte (št. ponovitev naj bo 1000) in grafično predstavite deleže zavrnitev vseh treh metod glede na π na enem grafu. Vrednost drugega parametra naj bo $n = 100$. Utemeljite/komentirajte rezultate.

    f. Glede na rezultate, kako mislite, da bi se razlikovali (a) rezultati in (b) ugotovitve, če bi delali simulacije za 99% interval zaupanja? Utemeljite.