

2. kolokvij - Osnove teoretične statistike

22. april, 2015

1. Raziskovalci so na sistematskem pregledu pri petih letih zbrali podatke o telesnih težah dvojajčnih dvojčkov. Na teh podatkih so izvedli dva statistična testa: test t in linearno regresijo.

(a) Kateri test t bodo uporabili? **Komentirajte zakaj.**

(b) Za vsak test posebej:

- zapišite ničelno domnevo (s formulo in z besedami),
- naštejte predpostavke, razložite, kaj pomenijo v konkretnem primeru in kdaj bodo oz. ne bodo izpolnjene

(c) Izpolnite spodnji tabeli za oba testa s smiselnimi vrednostmi (pri linearni regresiji se omejimo samo na regresijski koeficient). Vrednosti znotraj istega testa naj bodo med seboj konsistentne (pri vrednosti P je dovolj samo ocena glede na $\alpha = 0,05$).

test t	t	SE	P	95% IZ

linearna regresija	b (ocena β)	SE _b	P	95% IZ

(d) Vsebinsko interpretirajte rezultate obeh testov.

(e) Ali se lahko (na takih podatkih) zgodi, da eden izmed testov da značilen rezultat, drugi pa ne? Razložite (s konkretnim primerom).

2. Avtor knjižnice v R-u občasno prejema vprašanja uporabnikov. Da bi bolje sledil količini vprašanj, je tri mesece (12 tednov) beležil tedensko število vprašanj. Podatki so naslednji: v enem tednu je dobil kar 4 vprašanja, dvakrat je dobil po dve vprašanji, pet tednov pa je bilo takih, da je zabeležil le eno vprašanje, 4 tedne pa vprašanj ni bilo. Zanima ga, ali lahko pojavljanje vprašanj uporabnikov opiše s Poissonovim modelom:

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

- (a) Zapišite ničelno domnevo, predlagajte testno statistiko in povejte, kako je porazdeljena
 - (b) Nastavite izračun testne statistike, tako da bo točno jasno, kako jo boste izračunali (ni pa je potrebno dejansko izračunati).
 - (c) Je vaša testna statistika eksaktna ali ne? Opišite, kako bi preverili ali je pri dani velikosti vzorca velikost testa ustrezna.
3. Raziskovalka želi pokazati, da je med bolniki, ki so zaradi kožnih bolezni zdravljeni s fototerapijo, večja verjetnost pojava kožnega raka kot v splošni populaciji. Iz registra raka je dobila podatek, da je delež obolelih v populaciji enak $p_0 = 0,02$, predvideva, da je v populaciji njenih bolnikov delež obolelih $p_1 = 0,04$. Z ustreznim statističnim testom bo želela pokazati, da pojavnost bolezni pri njenih bolnikih ni enaka 0,02, zanima jo, kako velik vzorec mora zbrati, da bo moč testa pri $\alpha = 0,05$ enaka $\beta = 0,8$.
- (a) Zapišite ničelno domnevo (alternativna domneva naj bo dvostranska) in predpostavke testa. Predlagajte testno statistiko in zapišite njeno porazdelitev pod ničelno domnevo.
 - (b) Jasno je, da bo potreben precej velik vzorec, zato lahko uporabimo asimptotsko aproksimacijo. Naj bo vaša testna statistika enaka \hat{p} - s katero porazdelitvijo jo lahko aproksimirate, kakšni so parametri te porazdelitve?
 - (c) Izrazite meji zavrnitve pod ničelno domnevo, pri tem uporabite aproksimativno porazdelitev.
 - (d) Zapišite izraz za moč testa in komentirajte, kateri člen v izrazu lahko zaradi majhnosti zanemarite. Nato izrazite velikost vzorca.
 - (e) Denimo, da formule v prejšnji točki ne bi izpeljali. Zapišite kodo za simulacijo s katero bi izračunali vrednost α in β pri neki velikosti vzorca z mejami, ki ste jih določili s pomočjo aproksimacije.