

Statistika

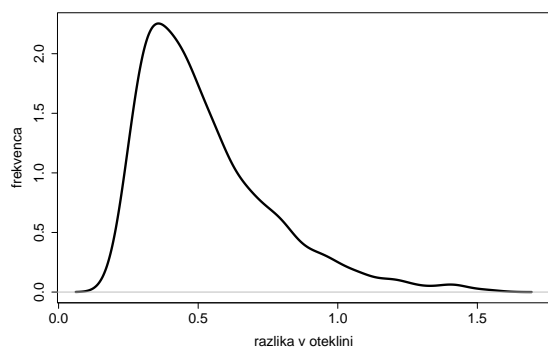
13. 6. 2017

Ime in priimek: _____ Vpisna številka: _____ Število točk: _____ FIZ521

Pri vprašanjih z več možnimi odgovori, je vedno pravilen natanko en odgovor. V primeru več zapisanih odgovorov se vprašanje točkuje z nič točkami. Odgovore nalog zapišite v spodnjo preglednico, sicer se NE UPOŠTEVAJO.

NALOGA	1	2	3	4	5	6	7
(I) oteklina				X	X	X	X
(II) Hodgkinov limfom				X	X	X	X
(III) Hiroshima							
(IV) Cooper, kajenje							
(V) Cooper						X	X

- (I) Raziskovalci so zbrali podatke o tenisačih in atletih, ki so si poškodovali komolec in so jim oteklino oskrbeli s kineziološkim trakom. Za vsakega so zabeležili njegov šport, obseg oteklina (cm) in *razliko v obsegu oteklina* po 2 dneh (cm). Slika prikazuje porazdelitev *razlike oteklina*.



- 1 Kateri graf bi izbrali za prikaz porazdelitve *razlike v oteklini* komolca glede na *šport*?
 - (a) Histogram za spremenljivko šport.
 - (b) Dva okvira z ročaji za spremenljivko *razlika v obsegu oteklina* (vsakega za svoj šport).
 - (c) Dva stolpična diagrama za spremenljivko obseg oteklina (vsakega za svoj šport).
 - (d) Okvir z ročaji za spremenljivko šport.
 - (e) Nič od zgoraj naštetega.
 - 2 Glede na sliko na prvi strani, s katero mero središčnosti in mero variabilnosti bi predstavili spremenljivko *razlika v oteklini*?
 - (a) Mediana in standardni odklon.
 - (b) Povprečje in standardni odklon.
 - (c) Povprečje in standardna napaka.
 - (d) Mediana in interkvartilni razmik.
 - (e) Delež in število enot.
 - 3 Katera trditev velja glede na sliko na prvi strani?
 - (a) Mediana spremenljivke je manjša kot njeno povprečje.
 - (b) Standardni odklon spremenljivke je manjši od standardne napake.
 - (c) Spremenljivka ima končno mnogo možnih vrednosti.
 - (d) Standardni odklon je enak interkvartilnemu razmiku.
 - (e) Minimuma spremenljivke se ne da določiti.
- (II) Onkološka literatura poroča, da je verjetnost ozdravitve bolnikov s Hodgkinovim limfomom 80%. Zdravnik izbere 20 naključnih bolnikov s Hodgkinovim limfomom.
- 1 Izberite pravilno trditev.
 - (a) Verjetnost ozdravitve je številska slučajna spremenljivka.
 - (b) Spremenljivka, ki nas zanima, je, ali je bolnik zbolel za Hodgkinovim limfomom in je opisna spremenljivka.
 - (c) Spremenljivka, ki nas zanima, je, ali se je bolnik s Hodgkinovim limfomom pozdravil in je opisna imenska spremenljivka.
 - (d) Raziskava opazuje zdrave in bolne osebe s Hodgkinovim limfomom.
 - (e) Populacija, ki jo zdravnik proučuje, je 20 oseb s Hodgkinovim limfomom.
 - 2 Koliko bolnikov bo najverjetneje ozdravelo?
 - (a) 10
 - (b) 16
 - (c) 17
 - (d) 20
 - (e) Nič od naštetega.
 - 3 Kako verjetno bo ozdravelo 17 bolnikov?
 - (a) 0.205
 - (b) 0.794
 - (c) < 0.001
 - (d) 0.159
 - (e) 0.657

(III) Raziskovalci na Japonskem so raziskovali vpliv sevanja atomske bombe na pojav rakavih obolenj v časovnem intervalu 20 let ($\alpha = 0,05$). V ta namen so zbrali 666 ljudi iz Hiroshime, ki so bili v času atomskega napada stari od 20–34 let in 6840 primerljivih oseb iz ostalih mest na Japonskem. Število obolelih iz ostalih mest na Japonskem je v našem vzorcu petkrat večje od števila obolelih iz Hiroshime. Število vseh ljudi, ki niso oboleli, je bilo 7308.

1 Kateri statistični test boste uporabili za to raziskovalno vprašanje?

- (a) Test t za dva neodvisna vzorca, saj imamo dve neodvisni skupini ljudi.
- (b) Test t za odvisna vzorca, saj so ljudje med sabo odvisni.
- (c) Test t za dva neodvisna vzorca, saj je spremenljivka starost številsko spremenljivka.
- (d) Test χ^2 , saj preverjamo povezanost med dvema opisnima spremenljivkama.
- (e) Nič od naštetega.

2 Kakšna je ničelna domneva za ta primer?

- (a) V populaciji je povprečna starost ljudi iz Hirošime je enaka povprečni starosti ljudi iz ostalih mest na Japonskem.
- (b) V populaciji je število obolelih v Hirošimi petkrat manjše kot število obolelih iz ostalih mest na Japonskem.
- (c) Spremenljivki rakavo obolenje in bivanje v Hiroshimi na Japonskem sta v populaciji povezani.
- (d) Spremenljivki rakavo obolenje in bivanje v Hiroshimi na Japonskem sta v populaciji neodvisni.
- (e) Populacijski delež obolelih v Hirošimi je različen deležu obolelih iz ostalih mest na Japonskem.

3 Kolikšno število ljudi iz Hiroshime je na vzorcu obolelih?

- (a) 33
- (b) 198
- (c) 666
- (d) 133
- (e) 1386

4 Kakšen je delež obolevnosti za rakom za ljudi iz Hiroshime in za ostale na našem vzorcu (zaokroženo na 3 decimalna mesta natančno)?

- (a) 0.050 in 0.024
- (b) 0.297 in 0.028
- (c) 0.097 in 0
- (d) 0.500 in 0.240
- (e) 0.049 in 0.004

5 Vrednost testne statistike za ta primer je 15,23. Kritična vrednost za ta test je:

- (a) $\chi_{1,0.95}^2 = 3.84$
- (b) $\chi_{2,0.95}^2 = 5.99$
- (c) $t_{7504;0.975} = 1.96$
- (d) $t_{3755;0.975} = 1.96$
- (e) $\chi_{1,0.975}^2 = 5.02$

6 Statistični sklep za izveden statistični test (testna statistika za ta primer je 15,23) je naslednji:

- (a) Ničelno domnevo zavrnamo, saj je kritična vrednost manjša od testne statistike.
- (b) Ničelne domneve ne zavrnamo, saj je kritična vrednost manjša od testne statistike.
- (c) Ničelno domnevo sprejemo, saj je kritična vrednost manjša od testne statistike.
- (d) Ničelno domnevo zavrnamo, ker je vrednost p večja od α .
- (e) Nič od naštetega.

7 Kaj so raziskovalci na podlagi izvedenega statističnega testa lahko zaključili?

- (a) V populaciji se povprečni starosti ljudi iz Hiroshime in ljudi iz ostalih mest na Japonskem statistično značilno razlikujeta.
- (b) Na vzorcu je število obolelih v Hiroshimi manjše kot število obolelih iz ostalih mest na Japonskem.
- (c) V populaciji sta spremenljivki rakavo obolenje in bivanje v Hiroshimi nepovezani.
- (d) V populaciji sta spremenljivki rakavo obolenje in bivanje v Hiroshimi neodvisni.
- (e) V populaciji je delež obolelih v Hiroshimi večji od deleža obolelih iz ostalih mest na Japonskem.

- (IV) Zbrali smo podatke o rezultatih Cooperjevega testa (tek na 2400 metrov) za 50 moških. Zanima nas povezanost med časom teka (9 do 15 min) in kajenjem (0=NE, 1=DA).
- 1 Izberite pravilno trditev.
 - (a) Čas teka je številsko zvezna spremenljivka, kajenje pa številsko diskretna.
 - (b) Čas teka je številsko zvezna spremenljivka, kajenje pa opisna imenska.
 - (c) Kajenje je urejenostna spremenljivka, čas teka pa je imenska spremenljivka.
 - (d) Čas teka je opisna urejenostna spremenljivka, kajenje pa opisna dihotomna spremenljivka.
 - (e) Čas teka je spremenljivka, ki jo proučujemo, kajenje pa spremenljivka, ki je ne potrebujemo.
 - 2 Katero statistično metodo bi uporabili za raziskovalno vprašanje, ki nas zanima?
 - (a) test t za dva neodvisna vzorca, ker je preučeval številsko spremenljivko za dve neodvisni skupini
 - (b) test t za dva odvisna vzorca, ker je ena izmed spremenljivk odvisna
 - (c) test χ^2 , ker je preučeval dve opisni spremenljivki
 - (d) test t za dva odvisna vzorca, ker je preučeval številsko spremenljivko za dve različni skupini moških
 - (e) test t za dva neodvisna vzorca, ker je preučeval številsko spremenljivko za dve odvisni skupini
 - 3 Interval zaupanja ($\alpha = 0,05$) za kadilce je $[9,5; 11,5]$. Katera interpretacija je pravilna?
 - (a) S 95% verjetnostjo leži vzorčni povprečni čas kadilcev od 9,5 in 11,5 minut.
 - (b) Imamo 95% zaupanje, da se populacijski povprečni čas kadilcev nahaja med 9,5 in 11,5 minut.
 - (c) S 95% verjetnostjo bo kadilec tekel med 9,5 in 11,5 minut.
 - (d) S 95% verjetnostjo bo povprečni kadilec imel čas med 9,5 in 11,5 minut.
 - (e) S 95% zaupanjem trdimo, da bodo kadilci tekli od 9,5 in 11,5 minut.
 - 4 Kaj velja za 99% interval zaupanja za kadilce, če je 95% interval zaupanja $[9,5; 11,5]$.
 - (a) 99% interval zaupanja je ožji.
 - (b) Spodnja meja 99% intervala zaupanja je nižja od 9,5.
 - (c) Pri 95% intervalu zaupanja imamo večjo verjetnost, da je v njem zajeta populacijska vrednost, kot pri 99% intervalu zaupanja.
 - (d) 99% interval zaupanja ima zgornjo mejo nižjo od 11,5.
 - (e) Oba intervala zaupanja vsebujeta populacijsko povprečno vrednost časa teka za kadilce.
 - 5 Na vzorcu je bilo manj nekadilcev. Interval zaupanja za nekadilce je $[9,0; 10,0]$, za kadilce pa $[9,5; 11,5]$. Kaj iz tega lahko sklepate?
 - (a) Standardna napaka pri nekadilcih je večja kot pri kadilcih.
 - (b) Standardni odklon pri nekadilcih je manjši kot pri kadilcih.
 - (c) Standardna odklona za kadilce in nekadilce sta enaka.
 - (d) Z večjim zaupanjem lahko trdimo, da kadilci tečejo med 9,5 in 11,5 minut.
 - (e) Povprečna vrednost teka nekadilcev v populaciji je nižja kot pri kadilcih.
 - 6 Vrednost testne statistike je bila po absolutni vrednosti večja od kritične vrednosti pri $\alpha = 0,05$.
 - (a) Ničelno domnevo lahko zavrnilo, saj je v tem primeru vrednost p manjša od stopnje značilnosti.
 - (b) Ničelno domnevo lahko zavrnilo, saj je v tem primeru vrednost p večja od stopnje značilnosti.
 - (c) Ničelno domnevo sprejmemo, saj je vrednost $p > 0,05$.
 - (d) Ničelne domneve ne zavrnilo, saj je vrednost $p > 0,05$.
 - (e) Ničelno domnevo obdržimo, saj je vrednost p manjša od stopnje značilnosti.
 - 7 Kakšen je vsebinski sklep, če ničelne domneve ne zavrnilo?
 - (a) Vzorčno povprečje časa teka nekadilcev je krajše kot vzorčno povprečje teka kadilcev.
 - (b) Povprečna časa teka kadilcev in nekadilcev v populaciji sta enaka.
 - (c) Povprečna časa teka kadilcev in nekadilcev v populaciji nista različna.
 - (d) Ne moremo trditi, da sta povprečna časa kadilcev in nekadilcev v populaciji različna.
 - (e) Ne moremo trditi, da sta povprečna časa kadilcev in nekadilcev v populaciji enaka.

(V) Za vsakega od 50 moških smo rezultat Cooperjevega testa (tek na 2400 metrov) izmerili dvakrat – na začetku svoje kariere v službi in po 5 letih dela.

1 Kaj je v tem primeru pravilni opis podatkov?

- (a) Populacija so vse osebe, ki so zaposlene in tečejo Cooperjev test.
- (b) Populacija so vsi moški, ki so zaposleni vsaj 5 let, vzorec pa 50 moških.
- (c) Vzorec je 50 moških, ki so na začetku svoje kariere, populacija pa moški po 5 letih dela.
- (d) Populacija so vsi moški, ki so dvakrat tekli Cooperjev test.
- (e) Vzorec in populacija je 50 moških, ki so zaposleni vsaj 5 let.

2 Izberite pravilno trditev o spremenljivki razlika v času teka (razlika = po 5 letih – začetek).

- (a) Razlika v času teka je normalna spremenljivka.
- (b) Razliki v času teka ne moremo določiti standardne napake.
- (c) Spremenljivko razlika v času teka bi predstavili s stolpičnim diagramom.
- (d) Spremenljivka razlika v času teka je opisna urejenostna spremenljivka.
- (e) S povprečno vrednostjo spremenljivke razlika v času teka ugotavljamo srednjo vrednost.

3 Kakšno ničelno domnevo bi na podlagi teh podatkov lahko preverili?

- (a) V populaciji je povprečni čas teka na začetku kariere enak kot po 5 letih v službi.
- (b) Na vzorcu je povprečna razlika v času teka (razlika = po 5 letih – začetek) enaka 0.
- (c) V populaciji je povprečna razlika v času teka (razlika = po 5 letih – začetek) različna od 0.
- (d) Na vzorcu je povprečni čas teka na začetku kariere enak kot po 5 letih v službi.
- (e) V populaciji je delež moških, ki izboljšajo svoj čas teka manjši kot delež tistih, ki svoj rezultat poslabšajo.

4 Interval zaupanja za razliko v času teka (razlika = po 5 letih – začetek) je $[0, 5; 2, 1]$. Katera trditev je pravilna?

- (a) V populaciji moški po 5 letih dela tečejo v povprečju bolje.
- (b) V populaciji moški po 5 letih dela tečejo v povprečju slabše.
- (c) Na vzorcu je povprečna razlika v času teka 0,6.
- (d) V populaciji moških ne moremo trditi, da po 5 letih dela tečejo v povprečju slabše ali bolje.
- (e) V populaciji vsi moški tečejo po 5 letih slabše.

5 Interval zaupanja za razliko v času teka (razlika = po 5 letih – začetek) je $[0, 5; 2, 1]$. Katera trditev je pravilna?

- (a) Če bi izvedli test t za dva neodvisna vzorca, bi ničelno domnevo zavrnil.
- (b) Če bi izvedli test t za dva odvisna vzorca, bi ničelno domnevo zavrnil.
- (c) Če bi izvedli test t za dva neodvisna vzorca, bi ničelno domnevo obdržali.
- (d) Če bi izvedli test t za dva odvisna vzorca, bi ničelno domnevo obdržali.
- (e) Ne da bi izvedli statistični test, ne bi mogli povedati ničesar o rezultatu tega statističnega testa.