

Statistička analiza podataka – međuispit

UNIZG FER, ak. god. 2016./2017.

24. travnja 2017.

Ispit traje 120 minuta i nosi 30 bodova. Svaki zadatak nosi pet bodova. Svaki zadatak rješavajte na zasebnoj stranici. Pišite uredno i čitko.

1. Podatci prikazuju broj stradalih od tornada u SAD-u u razdoblju od 1990. do 2000. godine:

53, 39, 39, 33, 69, 30, 25, 67, 130, 94, 40.

- (a) Napravite dijagram stabljika-i-list (engl. *stem and leaf*).
- (b) Nađite srednju vrijednost, medijan, kvartile, interkvartilni rang (IQR), minimalni i maksimalni podatak te skiciraj pravokutni dijagram s izdancima (engl. *box and whiskers plot*).
- (c) Ako je područje nestršećih vrijednosti definirano izdancima (engl. *whiskers*) veličine 1.5 IQR-a, utvrdite ima li stršećih vrijednosti u ovom skupu. Kako biste to komentirali? (Napomena: izdanci ne mogu prelaziti minimalni, odnosno maksimalni podatak).
- (d) Komentirajte oblik distribucije te navedite koju biste mjeru centra distribucije u ovom slučaju koristili i zašto.
2. Uređaj proizvodi metalne valjke. Izmjereni su promjeri $n = 9$ proizvedenih valjaka. Sredina tog uzorka jest 1.01 cm, a standardna devijacija uzorka jest 0.025.
- (a) Izračunajte 99%-tni interval pouzdanosti za očekivani promjer valjaka koje proizvodi navedeni uređaj uz pretpostavku da podatci dolaze iz normalne distribucije.
- (b) Bi li na gornji izračun utjecala spoznaja o varijanci populacije? Obrazložite odgovor.
- (c) Interpretirajte 99%-tni interval pouzdanosti.
3. Menadžer taksi-kompanije treba odlučiti treba li jedan tip guma zamijeniti drugim u svrhu manje potrošnje benzina. Dvanaest automobila ispitano je sa starom i novom vrstom guma bez promjene vozača i na istoj ruti. Podatci su dani u tablici (kilometri po litri):

Automobil	Nove gume	Stare gume	Automobil	Nove gume	Stare gume
1	4.2	4.1	7	5.7	5.7
2	4.7	4.9	8	6.0	5.8
3	6.6	6.2	9	7.4	6.9
4	7.0	6.9	10	4.9	4.7
5	6.7	6.8	11	6.1	6.0
6	4.5	4.4	12	5.2	4.9

- (a) Možemo li na razini značajnosti $\alpha = 0.05$ zaključiti da su nove gume bolje od starih? Obrazložite.

- (b) Koju ste pretpostavku testa ovdje iskoristili?
4. Izvagan je slučajan uzorak od 64 vrećice kokica i dobivena je prosječna težina od 5.23 dkg uz standardnu devijaciju 0.24.
- (a) Testirajte hipotezu da je prosječna težina vrećice kokica 5.5 dkg nasuprot alternative da je manja na razini značajnosti $\alpha = 0.05$.
- (b) Izračunajte snagu testa kada je alternativa 5.3 dkg (uz pretpostavku jednostranog testa) i veličina uzorka je 10, uz pretpostavku da je devijacija populacije poznata i iznosi 0.24.
- (c) Ako želimo da je snaga testa 0.9 kada je srednja vrijednost 5.2 dkg (uz pretpostavku jednostranog testa), koliko velik uzorak moramo uzeti? I u ovom slučaju koristite da je standardna devijacija populacije poznata i iznosi 0.24.
5. Istraživana je veza između pojave povišenog krvnog tlaka i pušenja tako da je ispitano 180 osoba i rezultati su prikazani tablicom:

	Nepušač	Blagi pušač	Teški pušač
Visok tlak	21	36	30
Normalan tlak	48	26	19

- Zanima nas je li pušenje povezano s povišenim krvnim tlakom, na razini značajnosti $\alpha = 0.05$.
- (a) Obrazložite koji test ćete koristiti i zašto ga smijete koristiti.
- (b) Postavite jasno hipoteze H_0 i H_1 .
- (c) Provedite test i interpretirajte rezultate testa u kontekstu zadatka.
6. Postupci ponovnog uzorkovanja mogu se koristiti u situacijama kada nije moguće primijeniti parametarske postupke izračuna intervala pouzdanosti ili testiranja hipoteza.
- (a) Na primjenjivost parametarskog postupka utječe: (1) poznavanje egzaktnje distribucije uzorkovanja procjenitelja, (2) poznavanje aproksimativne distribucije uzorkovanja procjenitelja i (3) veličina uzorka. S obzirom na ove faktore, navedite u kojim situacijama biste, umjesto parametarskog postupka, koristili postupak ponovnog uzorkovanja (npr. jackknife). Obrazložite odgovor.
- (b) Izračunajte jackknife procjenitelj i jackknife 95%-tni interval pouzdanosti varijance za uzorak $\{1, 2, 3, 4\}$. Procjene možete izračunati preko parcijalnih procjena ili preko pseudovrijednosti.
- (c) Ukratko objasnite koja je ideja procjene standardne pogreške preko pseudovrijednosti te koja je pritom uloga centralnog graničnog teorema.