

PRVI KOLOKVIJ IZ STATISTIČKOG PRAKTIKUMAZadatak [40 bodova]

Baza podataka `boston.xls` sadrži dio rezultata istraživanja o kvaliteti stanovanja u 506 Bostonskih kvartova i okolnih mjesta koja gravitiraju Bostonu:

varijabla `CRIM` za svako naselje sadrži stopu kriminala po glavi stanovnika

varijabla `CHAS` sadrži informaciju o tome je li naselje na obali rijeke Charles ili nije (1 - na obali je, 0 - nije na obali)

varijabla `RM` sadrži prosječan broj prostorija u stambenoj jedinici (kuća ili stan)

varijabla `AGE` sadrži proporciju stambenih jedinica u privatnom vlasništvu izgrađenih prije 1940.

varijabla `DIS` sadrži ponderiranu udaljenost od naselja do najbližeg od pet velikih Bostonskih poslovnih centara

varijabla `RAD` sadrži indeks pristupačnosti auto ceste iz naselja

varijabla `PTRATIO` sadrži broj učenika po nastavniku u obrazovnim ustanovama u naselju (kvocijent ukupnog broja učenika i ukupnog broja nastavnika)

varijabla `LSTAT` sadrži proporciju siromašnih stanovnika

varijabla `MEDV` sadrži medijan cijena stambenih jedinica u privatnom vlasništvu (u tisućama dolara).

Odgovorite na sljedeća pitanja i riješite sljedeće zadatke:

1. Kojim tipovima slučajnih varijabli biste modelirali podatke u bazi `boston.xls` i zašto? Ukratko opišite mjere zastupljenosti i mjere deskriptivne statistike kojima biste opisivali pojedine varijable iz ove baze (opise tih mjera napravite u skladu s tipovima varijabli koje ste identificirali u bazi). (4 boda)
2. Na prikladan način opišite podatke sadržane u varijablama `RM`, `RAD` i `MEDV`
 - a) posebno za svaku kategoriju varijable `CHAS`
 - b) zajedno za sve kategorije varijable `CHAS`.Opise popratite odgovarajućim grafičkim prikazima i interpretacijama. (6 bodova)
3. Navedite najvažnije činjenice i teorijske rezultate vezane uz normalnu distribuciju. Teorijske rezultate interpretirajte. (8 bodova)
4. Navedite procjenitelja za očekivanje slučajne varijable te navedite i interpretirajte njegova najvažnija asimptotska svojstva. Nadalje, konstruirajte interval pouzdanosti γ za procjenu očekivanja slučajne varijable. Korištenjem obiju metoda procijenite očekivanje slučajne varijable kojom modeliramo medijan cijena stambenih jedinica u privatnom vlasništvu (varijabla `MEDV`). Istu procjenu napravite samo za stambene jedinice koje pripadaju naseljima s indeksom pristupačnosti auto cesti manjim od dva. (8 bodova)
5. Navedite metode za procjenu vjerojatnosti događaja. Korištenjem navedenih metoda procijenite vjerojatnost da stambena jedinica u slučajno odabranom naselju u prosjeku ima manje od četiri prostorije (varijabla `RM`). (6 bodova)
6. Navedite procjenitelje za distribuciju i funkciju distribucije slučajne varijable. Korištenjem navedenih procjenitelja procijenite distribuciju i funkciju distribucije slučajne varijable kojom modeliramo prosječan broj prostorija u stambenoj jedinici. Što vam sugerira grafički prikaz procjene distribucije te slučajne varijable? U skladu s tim procijenite vjerojatnost iz zadatka 5. (8 bodova)