

ТЕОРИЈА УЗОРАКА (3В) - Писмени испит
11.6.2024.

1. У датотеци *cities_by_gdp* налазе се подаци о бруто домаћем производу (*GDP.billion.US*) за 2022. годину у милијардама долара за велика градска подручја широм света, као и подаци о броју становника у тим подручјима (*Metropolitan.Population*).
 - а) Изабрати узорак са понављањем обима 100 са вероватноћама избора пропорционалним броју становника користећи Лахиријев метод. Оценити просечан бруто домаћи производ Хансен-Хурвицовом оценом. На основу изабраног узорка, оценити и просечан бруто домаћи производ градских подручја у САД Хансен-Хурвицовом оценом.
 - б) Изабрати узорак са понављањем обима 100 са вероватноћама избора пропорционалним броју становника користећи метод кумулант. Оценити просечан бруто домаћи производ Хорвиц-Томпсоновом оценом. На основу изабраног узорка, оценити и просечан бруто домаћи производ градских подручја у САД Хорвиц-Томпсоновом оценом.
 - в) Упоредити оцене просечног бруто домаћег производа за градска подручја широм света из делова а) и б) у средње квадратном смислу.
2. Датотека *heart_disease* садржи податке о испитаницима код којих су измерени параметри који могу бити индикатори ризика за коронарну болест. Променљива *CHDRisk* говори о томе да ли је испитаник под високим ризиком да у наредних 10 година оболи од коронарне болести.
 - а) Пушаче сместити у један стратум, а непушаче у други (*smokingStatus*). Изабрати стратификован случајан узорак без понављања обима 300 користећи пропорционални распоред, па оценити удео испитаника који су под високим ризиком да у наредних 10 година оболе од коронарне болести. Одредити оцену дисперзије те оцене.
 - б) Ако се као помоћно обележје користи индекс телесне масе (*BMI*), одредити посебну по стратумима регресиону оцену просечног систолног крвног притиска (*sysBP*) и приближну дисперзију ове оцене на основу узорка из дела а). Такође, одредити посебну по стратумима регресиону оцену и њену дисперзију ако је познато да су регресиони коефицијенти по стратумима $b_{01} = 4$ и $b_{02} = 2$.
 - в) Изабрати прост случајан узорак без понављања обима 300 испитаника. Одредити 90% интервал поверења за удео испитаника који су под високим ризиком да у наредних 10 година оболе од коронарне болести.