

ТЕОРИЈА УЗОРАКА (3В) - Писмени испит
24.6.2024.

1. У датотеци *youtube_data_science_channels* налазе се подаци о видео снимцима са различитих *Youtube* канала којима је главна тема *data science*.
 - а) Ако су канали (*Channel_Name*) стратуми, изабрати стратификован случајан узорак без понављања обима 1800 користећи Нејманов избор, па оценити укупан број прегледа (*Views*) свих видео снимака и одредити оцену дисперзије те оцене. Затим изабрати стратификован случајан узорак са понављањем истих обима, оценити укупан број прегледа и одредити оцену дисперзије те оцене.
 - б) Ако се као помоћно обележје користи број прегледа, на основу узорка без понављања из дела а) оценити укупан број коментара (*Comment_Count*) свих видео снимака комбинованом и посебном по стратумима количничком оценом. Израчунати приближне дисперзије ових оцена.
 - в) Изабрати систематски узорак обима 1841, па оценити број видео снимака са преко милион прегледа. Одредити дисперзију те оцене.
2. Датотека *player_stats_2019_2020* садржи статистике играча за сезону 2019/2020. за 45 кошаркашких лига широм света.
 - а) Изабрати узорак од 2 националне лиге (све лиге у бази осим Евролиге) без понављања, са вероватноћама избора пропорционалним броју играча у лиги. Оценити проценат играча рођених пре 1990. године (*birth_year*) Хорвиц-Томпсоновом оценом.
 - б) Изабрати прост случајан узорак без понављања од 10 националних лига (све лиге у бази осим Евролиге), па оценити укупан број тимова код којих је бар 5 играча рођено пре 1990. године. Одредити оцену дисперзије ове оцене.
 - в) Одредити колико би тимова НБА лиге требало изабрати у прост случајан узорак без понављања, како би се оценило колико су у просеку поена (*PTS*) постигли тимови током целе сезоне, тако да одступање не буде веће од 500 поена са вероватноћом 0.8. Изабрати прост случајан узорак без понављања добијеног обима, оценити колико су у просеку поена постигли тимови НБА лиге током целе сезоне и израчунати дисперзију те оцене.