

Тема семинарског рада је:

ИЛУСТРАЦИЈА ИСТРАЖИВАЊА О ... <sup>1</sup>

Рад треба да представља сопствено мини (научно) истраживање на конкретном примеру. Пример треба засновати на бази података, коју је потребно пронаћи на интернету (може и међу R-овим базама). Потребно је да база података има довољан број опсервација (пожељно преко 50). База треба да садржи бар 2 категоричке и бар 2 нумеричке променљиве (колоне).

На почетку рада укратко описати базу података.

За категоричке променљиве је потребно нацртати одговарајуће `barplot`-ове и `piechart`-ове и јасно их означити. У случају категоричких вредности које већ нису представљене фактором, дефинисати одговарајуће факторе. За парове колона, где има смисла или сумњате да постоји зависност између њих, направити табеле пресека тих колона, као и графике зависности једне од друге. За сваку од табела урадити  $\chi^2$  тест независности и протумачити резултате.

За нумеричке вредности израчунати основне статистике и графички представити податке (нацртати хистограме, `boxplot` дијаграме,...). Уколико расподела узорка подсећа на неку од познатих расподела, тестирати (графички и KS тестом или неким другим) припадност тој расподели. Ако се подаци могу смислено поделити на два дела (према неком фактору), тестирати хипотезу о једнакости средње вредности нумеричких обележја на та два дела. Ако постоји више могућих подела (нпр. на мушки и женски део популације, а онда и на становиштво у селу и граду), може се урадити тестирање у обе варијанте. Протумачити графике и резултате тестирања.

Одабрати неки од статистичких тестова који има смисла применити на податке, а није рађен на часовима вежби. Описати одабрани тест (навести хипотезе, тест статистику, расподелу тест статистике при  $H_0$ ). Детаљно описати како се тај тест имплементира у R -у (функција којом се позива, аргументи те функције, излаз функције). Затим, имплементирати “пешке” тај тест и упоредити резултате са резултатима који се добијају применом уграђене функције у R -у.

Рад треба да садржи закључак са сажетом и прегледном интерпретацијом добијених резултата.

Кодове кориштене при раду обавезно укључити као део рада (фонт кодова може бити мањи у односу на остатак текста).

Број страна није ограничен.

Радови се раде самостално или у паровима.

Семинарски рад (у `.pdf` формату) и програмске делове рада (`.r` фајл) треба ставити у фолдер. Тај фолдер је потребно спаковати у `.zip` формат. У мејлу се шаље само та `.zip`—ована верзија фолдера. Рад можете писати и у `RMarkdown` -у (или `RNote` -у). У том случају се шаље такође само једна `zip`-ована верзија фолдера који садржи `.rmd` и `.html` фајл.

Неке од многих могућности за претрагу одговарајуће базе података:

- <https://vincentarelbundock.github.io/Rdatasets/articles/data.html>
- <http://www.stat.ufl.edu/~winner/datasets.html>
- <https://ec.europa.eu/eurostat/data/database/>
- <https://data.stat.gov.rs/?caller=SDDB>
- [http://college.cengage.com/mathematics/brase/understandable\\_statistics/7e/students/datasets/tvds/frames/frame.html](http://college.cengage.com/mathematics/brase/understandable_statistics/7e/students/datasets/tvds/frames/frame.html)
- <https://nces.ed.gov/ipeds/datacenter/DataFiles.aspx?gotoReportId=7&fromIped=true>

---

<sup>1</sup>допунити наслов