

# Семинарски рад из предмета Статистички софтвер 1

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 категоријске и бар 2 нумеричке променљиве (колоне).

На почетку рада укратко описати базу података. За категоријске променљиве је потребно нацртати одговарајуће `barplot`-ове и `piechart`-ове и јасно их означити. У случају категоријских 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 најважније статистике и графички представити податке (нацртати хистограме, `boxplot` дијаграме,...) Ако 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 становништво у селу и граду), може се урадити тестирање у обе варијанте. Протумачити графике и резултате тестирања.

На крају рада треба да се налази закључак са сажетом и прегледном интерпретацијом добијених резултата.

При раду користити програме написане у *R*-у, а њихове кодове обавезно укључити као део рада (фонт кодова може бити мањи у односу на остатак текста). Број страна није ограничен.

Семинарски рад (у `pdf`-у) и програмске делове рада треба ставити у фолдер са именом "`ImePrezime_brojIndeksa.zip`". Тај фолдер је потребно спаковати у `.zip` формат. Студент у мејлу шаље само један фајл и то управо `.zip`-овану верзију фолдера.

---

<sup>1</sup>допунити наслов

Радови се раде самостално, и сви треба да раде са различитим базама, зато када нађете базу која вам одговара, пошаљите ми мејл да проверим да није већ заузета.

Неке од многих могућности за претрагу одговарајуће базе података:

- <https://vincentarelbundock.github.io/Rdatasets/datasets.html>
- <http://gseacademic.harvard.edu/willetjo/datasets.htm>
- <http://www.stat.ufl.edu/~winner/datasets.html>
- <http://data.princeton.edu/wws509/datasets/default.html>
- [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)

Рок за предају семинарског рада је: **14. децембар 2016.**