

# Семинарски рад из Статистичког софтвера 1

## *Истраживачки пројекат*

Стефан Малбашић

2. фебруар, 2026.

### 1 Увод

Семинарски рад представља индивидуални мини-истраживачки пројекат чији је циљ примена статистичких метода и техника анализе података у реалном контексту, коришћењем програмског језика R. За разлику од задатака са унапред дефинисаним скуповима података, овај рад захтева од студената да самостално одаберу базу података, формулишу истраживачка питања и спроведу комплетан аналитички поступак, од припреме података до интерпретације резултата.

Основни циљ семинарског рада је развој способности статистичког размишљања и критичке анализе података, уз истовремену примену програмских алата обрађених на курсу *Статистички софтвер 1*. Кроз овај рад студенти треба да покажу разумевање структуре података, адекватан избор статистичких метода у складу са природом проблема, као и способност да добијене резултате аргументовано протумаче у контексту посматране појаве. Посебан акценат стављен је на јасно образложење аналитичких корака, правилну употребу симулационих и инференцијалних метода, као и на повезивање теоријских концепата са практичном анализом података.

### 2 Избор базе података

Студент самостално бира реалну базу података која мора испуњавати следеће услове:

- база садржи најмање 300 опсервација;
- база садржи најмање 6 променљивих (предиктора), од чега најмање 3 нумеричке, 2 категоријске и 1 бинарну променљиву;
- подаци потичу из реалног извора (отворени подаци, јавне базе, истраживачки скупови, на пример: [Kaggle](#), [World Health Organization](#), [Републички завод за статистику Србије](#), [Народна банка Србије](#) и други);
- база није коришћена у оквиру вежби, тестова или претходних семинарских радова;
- извор базе података мора бити јасно наведен.

### 3 Задаци

#### Задатак 1 – Формулација истраживачког проблема и опис података

На основу изабране базе података формулисати једно главно истраживачко питање и најмање две помоћне хипотезе које се могу проверити статистичком анализом. Описати контекст базе података, значење појединачних променљивих, као и њихове типове. Образложити зашто је изабрана база погодна за анализу постављеног проблема.

*Сугестија:* Приликом учења базе података потребно је проверити тип објекта и структуру појединачних променљивих. Користити функције за приказ основних статистика и дистрибуције података. Категоријске променљиве дефинисати као факторе, уз контролу редоследа нивоа где је то релевантно за даљу анализу. Користити основне функције за рад са подацима у R-у ради увида у структуру базе.

## Задатак 2 – Припрема и обрада података

Извршити припрему података за даљу анализу у складу са постављеним истраживачким питањима. Анализа треба да обухвати идентификацију и обраду недостајућих вредности (уколико постоје), креирање најмање једне нове променљиве, као и издвајање релевантног подскупа података.

*Сугестија:* Обраду података спроводити коришћењем функција за селекцију, филтрирање, сортирање и трансформацију колона. Креирање нових променљивих вршити применом аритметичких и логичких операција над постојећим колонама. Препоручује се употреба функција из пакета `dplyr`.

## Задатак 3 – Истраживачка визуализација

Графички истражити односе између кључних променљивих и проверити да ли визуелни прикази подржавају постављене хипотезе. Приказати најмање три различита типа графика, од којих бар један укључује поређење по категоријама.

*Сугестија:* Графички прикази треба да укључе хистограме и процене функције густине за нумеричке променљиве, као и `boxplot` приказе ради поређења група дефинисаних категоријалним променљивим. За анализу односа између две нумеричке променљиве користити `scatter plot`, уз могућност груписања по факторима. Сви графици треба да буду реализовани коришћењем пакета `ggplot2`.

## Задатак 4 – Статистичка инференција

Изабрати најмање две хипотезе и проверити их одговарајућим статистичким методама. Јасно формулисати нулту и алтернативну хипотезу, проверити претпоставке примене теста и интерпретирати резултате у контексту истраживачког проблема.

*Сугестија:* Приликом статистичке анализе користити функције за израчунавање интервала поверења и спровођење параметарских и непараметарских тестова. Избор теста вршити на основу типа променљивих и претпоставки о расподели података.

## Задатак 5 – Симулациони модел

Осмислити симулациони сценарио који описује реалну појаву повезану са посматраним подацима. Применом Монте Карло симулација анализирати понашање модела и приказати добијене резултате.

*Сугестија:* Потребно је дефинисати најмање два различита сценарија, на пример кроз различите вредности параметара расподеле или кроз различите линеарне комбинације независних случајних величина. За сваки сценарио спровести Монте Карло симулације са довољним бројем понављања и анализирати добијене резултате графички и нумерички, уз поређење расподела и основних статистичких мера у контексту постављеног проблема.

## Задатак 6 – Аутоматизација анализе

Дефинисати најмање једну корисничку функцију која аутоматизује део анализе и може се применити на сличне базе података. Функција треба да има јасно дефинисане аргументе и повратну вредност.

*Сугестија:* Потребно је дефинисати најмање две корисничке функције које као аргументе примају податке и релевантне параметре, а као излаз враћају статистике или резултате анализе. Унутар функције користити векторизоване операције, као и `apply`, `lapply` или `sapply` фамилију функција.

## Закључак

У закључку сумирати главне резултате анализе, прокоментарисати потврђеност постављених хипотеза, указати на ограничења базе података и предложити могуће правце даљег истраживања.

## 4 Напомене

### Напомена 1

- Семинарски рад се ради самостално. Дозвољено је коришћење материјала за курс, као и додатне литературе која је на истим истакнута.
- Студенти су у обавези да, пре почетка израде семинарског рада, доставе на мејл предметног асистента изабрану базу података (или линк ка бази), ради евидентирања и спречавања преклапања тема и скупова података између различитих радова.
- Рок за израду семинарског рада је крај испитних рокова. Одбрана семинарских радова могућа је у сваком испитном року. Семинарски рад је потребно доставити најкасније два дана пре термина одбране, уз јасну напомену да планирате излазак у конкретном испитном року.
- Радове слати на мејл [malbasic.matf@gmail.com](mailto:malbasic.matf@gmail.com) са ознаком ”СС1 - семинарски рад”.
- На усменој одбрани семинарског рада добија се и сет питања из градива које је излагано у току курса.

### Напомена 2

Рад је могуће послати у два формата:

- RMarkdown документ, у коме би били интегрисани код и детаљна анализа заједно (препоручени формат), детаљније на линку: [rmarkdown.rstudio.com](http://rmarkdown.rstudio.com);
- LaTeX документ и сл, у коме би било описано шта је урађено са детаљним закључцима, а искоментарисан код као додатак у обичном .r фајлу.