

# СТАТИСТИКА ЗА АУТОМАТСКУ АНАЛИЗУ ПОДАТАКА

## Упутство за израду семинарских радова

Семинарски рад представља испитну обавезу која се састоји из писаног рада и усмене одбране. Сваки од ова два дела носи максимално 20 поена. Семинарски рад раде по 3 студента (теме 1-6), односно 2 студента (остале теме). Могући су изузеци, као нпр. да неку од тема од 1-6 раде 2 студента, или да неку од осталих тема ради 1 студент, али то не сме да буде правило.

За писани део сви студенти који су обрадили исту тему добијају исти број поена, док на одбрани остварени број поена може се разликовати у зависности од знања студената.

Рад мора да садржи:

- насловну страну са насловом рада, именима аутора и називом предмета,
- увод,
- главни део,
- закључак,
- литературу.

Теме 1-6 су теоријско-практичне. За њих рад треба да се састоји из два дела, где ће се у првом приказати теоријска оправдања за оно што се ради у другом делу. Други део ради се у R-у и у њему се илуструју методе приказане у првом делу на конкретним подацима. Наведена литература је препоручена, али није и обавезна за коришћење. Подаци могу бити било који осим оних примера који су детаљно обрађени у препорученој литератури. Неки релевантни подаци могу се, на пример, наћи у задацима и допунама на крају препоручених поглавља.

Теме 7-11 су практичне. У темама 7-10 треба укратко описати податке (шта представљају), затим одредити почетну анализу података (дескриптивну статистику). Наравно, не треба обавезно обрађивати све променљиве из базе, већ направити смислену селекцију. Након тога формирати модел вишеструке линеарне регресије, образложити која променљива се узима као зависна, а које као предиктори. Формирати најпре почетни модел (са свим променљивим), а затим приказати алтернативне моделе добијене селекцијом променљивих и/или трансформацијом. Урадити дијагностику сваког предложеног модела и навести закључке за сваки модел понаособ и одредити који је модел најприкладнији.

Сва литература која се користи мора бити правилно цитирана, тј. на крају одговарајуће реченице ставити нпр. (за више информација погледати Milton et al. (1986)).

Величина рада није унапред одређена. Радови треба да буду концизни и да приказују одговарајућу статистичку анализу.

Рад се у пдф формату предаје путем мејла на [marcone@matf.bg.ac.rs](mailto:marcone@matf.bg.ac.rs) или [bojana.todic@matf.bg.ac.rs](mailto:bojana.todic@matf.bg.ac.rs) најкасније 5 дана пре задатог термина у испитном року, док ће усмена одбрана бити у том термину. Кодови из R-а могу бити убачени у пдф документ (није обавезно), али морају бити послати у пратећем текстуалном документу путем мејла. Сви кодови који се налазе у оквиру рада морају бити исправни и не смеју давати грешке као излаз.

### Избор теме

За избор теме потребно је да један члан групе пошаље мејл на [bojana.todic@matf.bg.ac.rs](mailto:bojana.todic@matf.bg.ac.rs) у коме је јасно назначен број или назив теме, као и имена свих чланова групе која се пријављује. Саветујемо вам да се пре слања мејла договорите с колегама да се не би догодило да две групе бирају исту тему. Уколико се то ипак догоди, предност има група која је раније послала мејл.

### Усмена одбрана

Препоручује се (али није обавезно) да сви аутори рада имају одбрану у истом року. Први део одбране подразумева презентовање рада, док други део подразумева одговарање на питања професора и асистента. Питања могу бити како у вези рада, тако и целокупног градива курса. За одбрану рада потребно је припремити презентацију. Презентација треба да траје максимално 15 минута с равномерном поделом презентовања међу ауторима.

## Списак тема

1. **Дијаграми контроле** - Препоручена литература: [1, §7] и [5, §9]
2. **Уравнотежени некомплетни блок дизајн** - Препоручена литература: [2, §4] и [3, §16]
3. **Двофакторска дисперзиона анализа** - Препоручена литература: [2, §5] и [3, §15]
4. **Коваријациона анализа** - Препоручена литература: [2, §14] и [3, §12]
5. **Логистичка регресија** - Препоручена литература: [4, §2] и [6, §4]
6. **Пуасонова регресија** - Препоручена литература: [4, §3]
7. **Анализа дискриминације у платама** (подаци: *Salaries*, пакет *car*)
8. **Истраживање о аутомобилима** (подаци: *TopGear*, пакет *robustHD*)
9. **Анализа трошкова авиокомпанија** (подаци: *airline\_costs*)
10. **Слободна тема** (податке бирате сами, и шаљете их на мејл [marcone@matf.bg.ac.rs](mailto:marcone@matf.bg.ac.rs) на одобрење)
11. **Анализа резултата спортисте Име Презиме**. Изабрати неког спортисту (или евентуално спортски клуб) и за њега/њу осмислити истраживање, прикупити податке, применити одговарајућу статистичку анализу како бисте дошли до приказа резултата и шта на њих утиче. Руководити се корацима у статистичкој анализи с презентације.

**Напомена:** Теме 10 и 11 могу бирати више група студената (наравно за различите податке, односно спортисте). За све евентуалне недоумице обратити се путем мејла професору или асистенту.

## Литература

- [1] J.S. Milton, J.J. Corbet and P.M. McTeer. Introduction to Statistics, DC Heath & Company, 1986.
- [2] D.C. Montgomery. Design and analysis of experiments. John Wiley & sons, 2001.
- [3] J.J. Faraway. Linear models with R, Chapman and Hall/CRC, 2004.
- [4] J.J. Faraway. Extending the linear model with R: generalized linear, mixed effects and nonparametric regression models, CRC press, 2016.
- [5] E.L. Cano, J.M. Moguerza, and M.P. Corcoba. Quality Control with R. An ISO Standards Approach, Springer, 2015.
- [6] Б. Милошевић. Основи статистике, Математички факултет, Београд, 2021.