

Време решавања теста је 90 минута.

Име и презиме: \_\_\_\_\_, бр. индекса \_\_\_\_\_, бр. поена \_\_\_\_\_.

У чувеној бази *iris.txt* (чувена, јер ју је Фишер објавио у свом раду: Fisher, R. A. (1936) The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7, Part II, 179–188. ) дати су подаци о ширини и дужини чашичних листића и латица за 50 цветова сваке од 3 врсте ириса, *Iris setosa*, *versicolor* и *virginica*. Дакле, имамо обележја *Sepal.Length*-дужина чашичног листића; *Sepal.Width*-ширина чашичног листића; *Petal.Length*-дужина латица; *Petal.Width*-ширина латица; *Species*-врста;

База се може учитати директном командом `data("iris")` (већ је сачувана у *R*-у) те вашу анализу тако и започните.

1. (1 поен) Приказати прелиминарну анализу података из базе. Која обележја бисмо моделовали непрекидним случајним величинама а која дискретним (категоричким)?
2. (2 поен) Приказати хистограме непрекидних обележја као и њихове *box – plot*-ове по категоријама врсте. Да ли на први поглед има разлике у облицима по категоријама?
3. (5 поена) Линеарно моделовати дужину латице (*Petal.Length*) преко осталих обележја (Потпун модел). Колики је коефицијент детерминације? Написати шта је нулта а шта алтернативна хипотеза *F* теста, *p*-вредност за дати модел и напишите закључак теста за праг значајности 0.01.

У задацима 4, 5, 6, 7, 8 посматра се потпун модел.

4. (2 поена) За колико би се разликовале (апсолутна разлика) тачкасте оцене дужина латица два цвета, први је врсте *setosa* а други врсте *virginica* а вредности свих осталих предиктора су им исте?
5. (2 поен) Колики је број оцењених параметара *p*? Направити интервал поверења за предикцију средње дужине цвета који има вредности предиктора *Sepal.Length*-4.4; *Sepal.Width*-3.2; *Petal.Width*-0.2; *Species*-*setosa*;
6. (2 поен) Направити 95% – *t*ни интервал поверења за коефицијент уз *Sepal.Width*. Шта се може закључити на основу тога о значаности тог коефицијента?
7. (3 поена) Приказати график резидуала преко фитованих вредности. Приказати *Q – Q* плот резидуала. Постоје ли статистички значајни аутлајери са нивоом значајности 0.05 (користите Бонферонијеву корекцију)?
8. (2 поена) Проверити да ли се у моделу јавља проблем колинеарности предиктора на основу *condition number*-а.
9. (3 поена) Моделовати *Petal.Length* преко (прве) главне компоненте потпуног модела и приказати резиме тог модела. Колики је коефицијент детерминације, и колика је *p* вредност *F* теста? Да ли је ово још један доказ да постоји јака корелација међу предикторима у потпуном моделу присутна?
10. (5 поена) Моделовати *Petal.Length* само преко *Species*. Колики је коефицијент детерминације, колика је *p*-вредност *F*-теста? Колике су средње дужине латица сваке од врста. Да ли постоји статистички значајна разлика средње дужине латица у зависности од врсте? Да ли бисте пре изабрали модел који се састоји само из главне компоненте или бисте се ипак одлучили за модел преко факторске променљиве *Species*?
11. (3 поена) Направите *Tukey's HSD* интервале поверења како бисте утврдили међу којим категоријама постоји значајна разлика у дужини латица. Извојите парове који се значајно разликују.