

1 NAVODILA ZA IZDELAVO SEMINARSKE NALOGE

Vsako seminarsko nalogo delata **dva študenta**, da lahko analizo sproti predebatirata in jo na koncu skupaj predstavita.

Izbirate lahko med naslednjimi nalogami (dokler posamezna naloga ni izbrana, "kdor prvi pride, prvi melje"):

1. Na podlagi spremenljivk v podatkovnem okviru **Credit** iz paketa **ISLR** analizirajte odvisnost spremenljivke **Balance** od ostalih spremenljivk. Izberite optimalen model za napoved te spremenljivke in ga obrazložite. (**Julija Rupnik in Ela Vidovič**)
2. V podatkovnem okviru **flowers** v paketu **GLMsdata** so podatki iz poljskega poskusa, kjer so želeli ugotoviti, kako jakost svetlobe (**Light**) in čas osvetlevanja (**Timing**) vplivata na število cvetov na rastlino (**Flowers**). Analizirajte rezultate poskusa. Za izbrani linearni model primerjajte 95 % intervale zaupanja za parametre modela s 95 % bootstrap intervale zaupanja za parametre modela. (**Eva Lavrenčič in Špela Plemelj**)
3. V podatkovnem okviru **fishfood** iz paketa **GLMsData** so podatki o prehrani različnih vrst rib. Zanima nas odvisnost **FoodCon** od ostalih spremenljivk v podatkovnem okviru. Z modelom želimo napovedovati **FoodCon**. Izberite optimalni model in ga obrazložite. (**Gašper Antolin in Andrej Sokolič**)
4. V podatkovnem okviru **wheatrain** iz paketa (**GLMsData**) so podatki o letnih padavinah (**AR**) na 24 lokacijah v New South Wales v Avstraliji. Zanima nas odvisnost **AR** od ostalih spremenljivk v podatkovnem okviru. V modeliranju upoštevajte, da je podatek za povprečno letno količino padavin izračunan na podlagi različno dolgih časovnih obdobj. Izberite ustrezen linearni model, zanj naredite diagnostiko in rezultate obrazložite. (**Andraž Košuta in Anja Žavbi Kunaver**)
5. V podatkovnem okviru **Boston** iz paketa **MASS** analizirajte odvisnost spremenljivke **nox** od spremenljivke **dis**. Uporabite polinomske regresije, regresije zlepkov in regresije naravnih zlepkov. Za izbiro ustrezne stopnje polinoma, oziroma ustreznih stopinj prostosti regresijskih zlepkov (stopnja polinoma po odsekih in število vozlišč) uporabite postopek K-kratnega navzkrižnega preverjanja. Primerjajte izbrane modele na podlagi RMSE in za model z najboljšo kakovostjo napovedi naredite diagnostiko modela in ga obrazložite. (**Bijec Janez in Vesna Zupanc**)
6. V podatkovnem okviru **Boston** iz paketa **MASS** analizirajte odvisnost **medv** od ostalih spremenljivk v podatkovnem okviru. Modelirajte z namenom napovedovanja vrednosti spremenljivke **medv**. Za izbrani model naredite diagnostiko in ga obrazložite. (**Luka Androjna in Robert Dovžan**)
7. V podatkovnem okviru **College** iz paketa **ISLR** analizirajte odvisnost **Outstate** od ostalih spremenljivk v podatkovnem okviru. Podatke na začetku razdelite v učni in testni podvzorec. S sekvenčnimi metodami izbire primernega modela izberite model za učni vzorec in

ovrednotite napako napovedi na podlagi testnega podvzorca. Postotpek modeliranja ponovite na treh različnih delitvah enot v učni in testni vzorec. Primerjajte in obrazložite rezultate modeliranja.

8. V podatkovnem okviru `Auto` iz paketa `ISLR` analizirajte odvisnost `mpg` od ostalih spremenljivk v podatkovnem okviru. Modelirajte z namenom napovedovanja vrednosti spremenljivke `mpg`. Za izbrani model naredite diagnostiko in ga obrazložite. (**Jan Jezeršek in Mojca Klemenčič**)

Splošna navodila:

- a) kratko opišite vsebinsko ozadje problema in vir podatkov (`help()`);
- b) opredelite namen statistične analize (ali gre za pojasnjevanje vpliva izbranih napovednih spremenljivk na odzivno spremenljivko, ali gre za testiranje domnev, ali gre za napovedovanje);
- c) naredite ustrezne grafične prikaze;
- d) v skladu z namenom statistične analize prikažite postopek modeliranja. Če je vsebinsko smiselno, v model vključite tudi interakcijske člene. Če je potrebno, nelinearnost modelirajte s polinomsko regresijo ali regresijo zlepkov. Za izbrani model izvedite diagnostiko;
- e) grafično prikažite napovedi izbranega modela z intervali zaupanja za povprečne napovedi;
- f) napišite jedrnat povzetek vaših ugotovitev.

Predstavitve seminarskih nalog bodo predvidoma 16. 1. 2019 (o tem se lahko še dogovorimo).