

Interna IT izobraževanja KIS

Uporaba statističnega programa R

Jana Obšteter

Ljubljana, oktober 2020

Kazalo vsebine

1 Osnovni podatki.....	1
2 Modul A: Osnove R-a, R-studia, pisanja kode in dela z datotekami.....	2
2.1 Spoznavanje z R-studiem.....	2
2.2 Spoznavanje s kodo.....	2
2.3 Osnove organizacije in dela z datotekami.....	2
2.4 Upravljanje s tabelami.....	2
2.5 Pomoč za R.....	2
3 Modul B: Osnovna obdelava podatkov ter pisanje kode v R.....	3
3.1 Povzemanje podatkov.....	3
3.2 Vizualizacija podatkov z lattice / ggplot2.....	3
3.3 Osnovne zanke in izrazi.....	3
3.4 Osnovna statistika – preverjanje hipotez.....	3
3.5 Shranjevanje delovnega prostora v R-u.....	3
4 Modul C: Naprednejša statistika z R, naprednejša vizualizacija in pisanje funkcij...4	4
4.1 Linearna regresija.....	4
4.2 Logistična regresija.....	4
4.3 Pisanje funkcij.....	4
4.4 Napredna vizualizacija z R.....	4
4.5 Druge možnosti R-a.....	4

1 Osnovni podatki

Modul A:

Predvideno število ur: 2

Termin: november 2020

Predpogoj: naložen R in Rstudio, prenesen material (programska koda in datoteke)

Material: predstavitev v pdf obliki, primeri kode (Rmd ali R) na platformi Basecamp in github-u, potrebne datoteke za delo

Modul B:

Predvideno število ur: 2 – 2,5

Termin: januar 2021

Predpogoj: naložen R in Rstudio, prenesen material (programska koda in datoteke)

Material: predstavitev v pdf obliki, primeri kode (Rmd ali R) na platformi Basecamp in github-u, potrebne datoteke za delo

Modul C:

Predvideno število ur: 2 – 2,5

Termin: februar 2021

Predpogoj: naložen R in Rstudio, prenesen material (programska koda in datoteke)

Material: predstavitev v pdf obliki, primeri kode (Rmd ali R) na platformi Basecamp in github-u, potrebne datoteke za delo

2 Modul A: Osnove R-a, R-studia, pisanja kode in dela z datotekami

2.1 Spoznavanje z R-om in R-studiem

- struktura R-studia
- pisanje programske kode in izvrševanje ukazov
- nameščanje paketov in nalaganje knjižnic
- samodejno dopolnjevanje

2.2 Spoznavanje s kodo

- podatkovni tipi
- podatkovne strukture
- osnovne računske operacije in operatorji
- osnovne "vgrajene" funkcije

2.3 Osnove organizacije in dela z datotekami

- kako nastaviti delovni imenik
- branje / pisanje različnih tipov datotek
- lastnosti prebranih podatkov
- primer branja podatkov
- primer ustvarjanja tabele

2.4 Preurejanje podatkov

- spajanje / uparjanje podatkov
- 'vrtenje' tabel
- naprednejše preurejanje: `tidyr`/`tidyverse` funkcije
- cevovodna obdelava podatkov

2.5 Povzemanje podatkov

- Srednje vrednosti in razpršenost podatkov
- povzemanje po skupinah
- povzemanje podatkov s funkcijami paketa `dplyr`

2.5 Pomoč za R

- programska koda
- R pomoč in zgledi kode
- spletni tečaji
- najpogostejše napake
- dobra praksa pisanja kode
- pomoč za shranjevanje kode: `github`

3 Modul B: Osnovna obdelava podatkov ter pisanje kode v R

3.1 Iteriranje

- zanke: »for« in »while« zanka
- pogojni stavek
- iteriranje skozi tabele z `apply` družino funkcij

3.2 Vizualizacija podatkov z `lattice` / `ggplot2`

- histogram
- gostota podatkov
- škatlasti grafikon
- stolpčni grafikon
- razstresen grafikon
- črtni grafikon
- shranjevanje grafikonov

3.4 Osnovna statistika – preverjanje hipotez

- stopnja zaupanja, kritična vrednost, intervali zaupanja
- t-test
- hi-kvadrat
- ANOVA
- osnovni post-hoc testi

3.5 Shranjevanje delovnega prostora v R-u

4 Modul C: Naprednejša statistika z R, naprednejša vizualizacija in pisanje funkcij

4.1 Linearna regresija

- preverjanje predpostavk modela
- priprava modela
- koeficienti in rezultati modela
- preverjanje ustreznosti modela
- napovedovanje
- post-hoc test
- večrazsežna ("multiple") linearna regresija

4.2 Logistična regresija

4.3 Pisanje funkcij

- Forma funkcije
- parametri funkcije
- uporaba funkcij

4.4 Napredna vizualizacija z R

- Združevanje podatkov
- prilagajanje grafičnih parametrov
- razčlenitev in združevanje grafov

4.5 Druge možnosti R-a

- Razvoj interaktivnih aplikacij: `shiny` R
- Priprava pdf dokumentov: `knit`
- Knjižnice za analizo genomskih podatkov