Interna IT izobraževanja KIS

Uporaba statističnega programa R

Jana Obšteter

Kazalo vsebine

[Modul A: Osnove R-a, R-studia in pisanja kode 2](#_Toc46984003)

[1. Spoznavanje z R-studiem 2](#_Toc46984004)

[3. Osnove organizacije dela z datotekami 2](#_Toc46984005)

[Modul B: Osnovna obdelava podatkov ter pisanje kode v R 3](#_Toc46984006)

[Modul C: Naprednejša statistika z R in pisanje funkcij 4](#_Toc46984007)

# Modul A: Osnove R-a, R-studia in pisanja kode

Predvideno število ur: 2

Termin: konec septembra / oktober 2020

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

Material: predstavitev v pdf obliki, primeri kode (Rmd ali R) na mreži ali github, potrebne datoteke za delo

### 1. Spoznavanje z R-studiem

* struktura R studia
* pisanje programske kode in izvrševanje ukazov
* nameščanje paketov in nalaganje knjižnic
* spletni viri:
  + programska koda
  + R pomoč in zgledi kode

## 2. Spoznavanje s kodo

* podatkovni tipi
* podatkovne strukture
* osnovne računske operacije in operatorji
* osnovne “vgrajene” funkcije in pomoč
* dobre prakse pisane kode
* samodejno dopolnjevanje
* najpogostejše napake

## 3. Osnove organizacije dela z datotekami

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

## 4. Osnovne zanke in izrazi

* “for” zanka
* “repeat” / “while” zanka
* izstop iz zanke
* pogojni stavki

## 4. Pomoč za R

* stack overflow

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

Predvideno število ur: 2 – 2,5

Termin: konec oktobra / začetek novembra 2020

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

Material: predstavitev v pdf obliki, primeri kode (Rmd ali R) na mreži ali github, potrebne datoteke za delo

## 1. Povzemanje podatkov

* Opisna statistika
  + srednje vrednosti, razpršenost podatkov, kvantili
* povzemanje po skupinah: funkcije v base / dplyr paketih
* ponavljajoče apliciranje funkcij: apply družina funkcij

## 2. Upravljanje s tabelami

* združevanje tabel
* preurejanje tabel
* tidyr/tidyverse funkcije

## 3. Vizualizacija podatkov z lattice / ggplot2

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

## 4. Osnovna statistika – preverjanje hipotez

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

## 5. Shranjevanje delovnega prostora v R-u

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

Predvideno število ur: 2 – 2,5

Termin: konec novembra / začetek decembra 2020

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

Material: predstavitev v pdf obliki, primeri kode (Rmd ali R) na mreži ali github, potrebne datoteke za delo

## 1. Linearna regresija

* preverjanje predpostavk modela
  + LINE
* priprava modela
* koeficienti in rezultati modela
* preverjanje ustreznosti modela
  + porazdelitev ostankov, normalni kvantilni graf (QQ plot), Wald test, test razmerij verjetnosti (likelihood ratio test)
* napovedovanje
* post-hoc test
* večrazsežna (“multiple”) linearna regresija

## 2. Logistična regresija

## 3. Pisanje funkcij

* Forma funkcije
* parametri funkcije
* uporaba funkcij

## 4. Napredna vizualizacija z R

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

## 4. Druge možnosti R-a

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