Interna IT izobraževanja KIS

Uporaba statističnega programa R

Jana Obšteter

Ljubljana, avgust 2020

Kazalo vsebine

[1. Osnovni podatki 2](#__RefHeading___Toc6145_3356873930)

[2. Modul A: Osnove R-a, R-studia, pisanja kode in dela z datotekami 2](#__RefHeading___Toc5945_3356873930)

[2.1 Spoznavanje z R-studiem 2](#__RefHeading___Toc5947_3356873930)

[2.2 Spoznavanje s kodo 2](#__RefHeading___Toc5949_3356873930)

[2.3 Osnove organizacije dela z datotekami 2](#__RefHeading___Toc5951_3356873930)

[2.4 Osnovne zanke in izrazi 3](#__RefHeading___Toc5953_3356873930)

[2.5 Pomoč za R 3](#__RefHeading___Toc6086_3356873930)

[3. Modul B: Osnovna obdelava podatkov ter pisanje kode v R 3](#__RefHeading___Toc5957_3356873930)

[3.1 Povzemanje podatkov 3](#__RefHeading___Toc5959_3356873930)

[3.2 Upravljanje s tabelami 3](#__RefHeading___Toc5961_3356873930)

[3.3 Vizualizacija podatkov z lattice / ggplot2 3](#__RefHeading___Toc5963_3356873930)

[3.4 Osnovna statistika – preverjanje hipotez 3](#__RefHeading___Toc5965_3356873930)

[3.5 Shranjevanje delovnega prostora v R-u 4](#__RefHeading___Toc5967_3356873930)

[4. Modul C: Naprednejša statistika z R, naprednejša vizualizacija in pisanje funkcij 4](#__RefHeading___Toc5969_3356873930)

[4.1 Linearna regresija 4](#__RefHeading___Toc5971_3356873930)

[4.2 Logistična regresija 4](#__RefHeading___Toc5973_3356873930)

[4.3 Pisanje funkcij 4](#__RefHeading___Toc5975_3356873930)

[4.4 Napredna vizualizacija z R 4](#__RefHeading___Toc5977_3356873930)

[4.5 Druge možnosti R-a 4](#__RefHeading___Toc5979_3356873930)

# 1. Osnovni podatki

**Modul A:**

Predvideno število ur: 2

Termin: konec septembra / oktober 2020

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

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

**Modul B:**

Predvideno število ur: 2 – 2,5

Termin: konec oktobra / začetek novembra 2020

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

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

**Modul C:**

Predvideno število ur: 2 – 2,5

Termin: konec novembra / začetek decembra 2020

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

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

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

### 2.1 Spoznavanje z R-studiem

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

## 2.2 Spoznavanje s kodo

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

## 2.3 Osnove organizacije dela z datotekami

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

## 2.4 Osnovne zanke in izrazi

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

## 2.5 Pomoč za R

* programska koda
* R pomoč in zgledi kode
* pomoč za shranjevanje kode: github

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

## 3.1 Povzemanje podatkov

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

## 3.2 Upravljanje s tabelami

* združevanje tabel
* preurejanje tabel
* tidyr/tidyverse funkcije

## 3.3 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
  + 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

## 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