

Reproducējams pētījums

Reproducible research

Didzis Elferts

**Kāda ir atšķirība starp
reproducējamu (reproducible)
un atkārtojamu (repeatable)
pētījumu?**

Kas ir atkārojams pētījums?

Atkārojams pētījums ir tāds pētījums, kuru cits zinātnieks izmantojot savus datus var atkārtot, balstoties uz Jūsu pētījumā aprakstītās metodes.

Kas ir reproducējams pētījums?

Reproducējams pētījums ir tāds pētījums, kuru ir iespējams pilnībā reproducēt (iegūt rezultātu vēlreiz) pētījumu autoram vai kādai citai personai (Wikipedia).

Kāpēc jāveic pētījumu atkārtošana?

- Lai pārlicinātos par iepriekšējo rezultātu pareizību
- Lai pārlicinātos, ka līdzīgi rezultāti tiek iegūti ar jauniem datiem, citos apstākļos
- Lai būtu vairāk pierādījumu sabiedrībai būtiskos jautājumos

Problēmas ar pētījumu atkārtošanu

- Laika ierobežojums
- Finanšu ierobežojums
- Situācija ir pārāk unikāla

Kāpēc ir nepieciešami reproducējami pētījumi?

- Pieaug datu daudzums, to apjoms un sarežģītība
- Datorresursi ļauj veikt arvien sarežģītākus aprēķinus
- Cilvēkiem ir jābūt iespējai pārlicināties par iegūto rezultātu pareizību

**Reproducejamība nenovērš
klūdas pētījumā, bet ļauj
vieglāk kādam tās atrast**

Kā padarīt sava pētījumu reproducējamu?

- Pieejami izejas dati
- Pieejami analīzes kodi (ar skaidrojumiem un atšifrējumiem)
- Skaidri aprakstīti nosacījumi (sistēmas parametri), pie kādiem veikti aprēķini

Reproducejamība un "Atvērtā zinātne" (Open Science)

- Pētījumu reproducejamība lielā mērā saistās arī ar "atvērto pieeju" (open access)
- Atvērtā zinātne balstās uz četriem būtiskiem soļiem:
 - Atvērtie dati
 - Atvērtie kodi
 - Atvērtā pieeja (brīvpieeja)
 - Atvērtie standarti

No kā jāizvairās darba gaitā

- Datu pārveidošana ar "roku" (piemēram, Excel):
 - kļūdu labošana
 - trūkstošo vērtību aizvietošana
 - jaunu mainīgo ieviešana
- Darba darīšana ar "klikšķināmajām" programmām, sevišķi, ja tās neveido "log" failus - pierakstīt visus soļus ir ļoti grūti

Versiju kontroles sistēmas

Veicot datu analīzes procesu vēlams izmantot kādu no versiju kontroles sistēmām, piemēram, git. Tas nodrošina (ja pareizi izmanto):

- Visu veikto izmaiņu fiksēšanu
- Iespēju atgriezties "vēsturē", ja kaut kas izdarīts nepareizi
- Iespēju vienlaicīgi izmēģināt alternatīvus ceļus (branch)
- Var nodrošināt sadarbību starp vairākiem zinātniekiem

Reproducējams dokuments

R markdown

- Programmā R (ne tikai) var veidot īpašus dokumentus (literāte programming), kas satur gan cilvēkam saprotamu tekstu, gan arī programmai saprotamas un izpildāmas komandu rindas
- Viens no veidiem tā nodrošināšanai ir R markdown valoda
- RStudio šādu dokumentu veidošana jau ir integrēta, izmantojot failus ar paplašinājumu .Rmd
- Par teksta noformējumu atbild markdown, bet par izpildāmajiem kodiem pakete knitr, formātu maiņu veic pandoc

R markdown priekšrocības

- Viegla sintakse
- Izejas dokuments ir viegli lasāms
- Viss noformējums redzams
- Iekļauti aprēķina kodi

Tekstu noformējums

**italic* teksts slīprakstā*

****bold** trekns teksts**

Virsraksti:

1.līmenis

2. līmenis

3. līmenis

Saraksti

Vienkāršs saraksts

* Ieraksts 1

* Ieraksts 2

- Ieraksts 2a

- Ieraksts 2b

Numurēts saraksts

1. Ieraksts 1

2. Ieraksts 2

R kodi

- Kā programmas kodi tiek uztverti tikai speciālās "zīmēs" iekļautās komandu rindas, pārējais ir tikai teksts
- Iekļauts kods sastāv no koda sākuma apzīmējuma un koda nobeiguma apzīmējuma

```
```\r}
```

```
x <- 1:10
```

```
plot(x)
```

```
```\r}
```

R kodi

- R kodu un tā rezultātu var iekļaut arī teksta rindā, lietojot zīmes `r`
- Failā mtcars ir `r nrow(cars)` rindiņas
- Failā mtcars ir 50 rindiņas

Paslēpts kods

Iekļaujot argumentu `echo=FALSE`, tiek rādīts tikai rezultāts, nerādot pašu kodu gala dokumentā.

```
```{r,echo=FALSE}
```

```
x <- 1:10
```

```
plot(x)
```

```
```
```

Kods, kas netiek izpildīts

Arguments `eval=FALSE` nodrošina, ka kods pārādās dokumentā, bet tas netiek izpildīts.

```
```{r,eval=FALSE}
```

```
x <- 1:10
```

```
plot(x)
```

```
```
```

Brīdinājumi un paziņojumi

Brīdinājumus un paziņojumus, kurus parāda atsevišķas r funkcijas vai paketes, var paslēpt/parādīt ar argumentiem:

- `message=TRUE` vai `message=FALSE`
- `warning=TRUE` vai `warning=FALSE`

Attēlu parametri

Katram attēlam atsevišķi var mainīt izmērus, attēlu parakstus

```
`` `{r,fig.width=8,fig.height=4.5,fig.cap="Attēla paraksts"}
```

```
x <- 1:10
```

```
plot(x)
```

```
`` `
```

Tabula

Tabulu veidošanai (kolonnu atdalīšanai) izmanto | zīmi

Viens | Divi | Trīs

- - -|- - -|- - -

12 | 56 | 89

Viens

Divi

Trīs

12

56

89

Tabulas noformējums R objektiem

Lai dabūtu "skaistāku" tabulu noformēju, var izmantot funkciju `kable` no paketes `knitr`, pirms tam pievienojot argumentu `results=`asis`` r koda sākumam.

```
knitr::kable(head(cars))
```

| speed | dist |
|-------|------|
| 4 | 2 |
| 4 | 10 |
| 7 | 4 |
| 7 | 22 |
| 8 | 16 |
| 9 | 10 |

Papildus paketes

Visas papildus paketes ir jāpievieno dokumentā, jo katrs dokuments tiek izpildīts savā "vidē". Tas pats attiecas uz visiem darba failiem un ārējiem kodiem.

Rezultātu saglabāšana

Pēc noklusējuma katreiz izpildot komandu Knit, visi aprēķini tiek veikti no jauna. Ja tas prasa pārāk daudz laika, var noteikt, ka rezultāti jā saglabā un nav jāpārrēķina, ja nemainās nosacījumi.

Faila augšpusē jāiekļauj sekojošas komandu rindas.

```
```{r setup, include=FALSE}  
knitr::opts_chunk$set(cache=TRUE)
```
```

Dokumentu veidi

Izmantojot R markdown var veidot trīs veidu dokumentus un trīs veidu prezentācijas, attiecīgi kā output: norādot:

- `pdf_document` - pdf dokuments (lai tos veidotu nepieciešama TEX sistēma uz datora)
- `html_document` - html dokuments
- `word_document` - docx dokuments
- `slidy_presentation` - slidy prezentācija
- `ioslides_presentation` - ioslaidu prezentācija (problēmas uz Windows ar latviešu burtiem)
- `beamer_presentation` - LATEX beamer prezentācija (lai to veidotu nepieciešama TEX sistēma uz datora)

Latviešu valoda

Lai varētu veidot latviešu valodas tekstus, izejas failam (.Rmd) ir jābūt UTF-8 kodējumā.

Veidojot pdf dokumentus, jāizmanto xelatex sistēma.

Dokumentu paraugi

Word dokumentiem un html dokumentiem var izveidot stilu paraugus, kurus izmantot veidojot jaunus dokumentus. html dokumentiem stilus maina ar css palīdzību, word dokumentiem izveido paraugu kā .docx dokumentu.

Pēc tam metadatos jānorāda sekojošā rindiņa:

```
reference_docx: stili.docx
```

R markdown

Vairāk informācijas par R markdown iespējām programmā RStudio atrodama mājaslapā: <http://rmarkdown.rstudio.com/>

Par automātisku attēlu/tabulu numerāciju: <https://gist.github.com/benmarwick/f3e0cafe668f3d6ff6e5>

Par automātisku attēlu/tabulu numerāciju un markdown pamatiem: <http://galahad.well.ox.ac.uk/repro/>