

Rakendustarkvara: R. Sügis 2017, 9. praktikum*

1 Tulemuste vormistamine

Sageli kõige tüütum andmeanalüüsi juures on tulemuste vormistamine ilusaks dokumendiks. R-i väljund on tavaliselt kas konsolis või graafikaaknas. Isegi kui tulemused on konsoliaknas tabelina, on seda Wordi vms programmi ümber tõsta tülikas. Õnneks on ka selle probleemi lahendamiseks R-is mitmeid erinevaid pakette, mille tööpõhimõte on sarnane:

1. Kirjutada oma analüüsi tekst koos vahepealsete koodiridadega ühte faili.
2. See fail kompileerida sobivasse formaati (nt html, tex või pdf).

Käesolevas kursuses tutvume paketiga **knitr**¹. Selle paketi kõige olulisem käsk on **knit(.)**, millele tuleb ette anda punktis 1. koostatud faili nimi. Selles failis on mingis formaadis tekst ja R-i kood vaheldumisi, näiteks nii:

```
---
output: html_document
---
# Esimene arvutus
```{r,echo=FALSE}
mass <- read.table("http://kodu.ut.ee/~annes/Rkursus/mass.txt", sep = "\t", header = T)
```

Massachusettsi andmestikus on kokku `r nrow(mass)` inimest; sooline jaotus on järgnev:
```{r,echo=FALSE}
table(mass$SEX)
```

## Järgmine punkt
Nüüd on lisaks tulemusele endale näha, milline koodijupp leiab keskmise vanuse:
```{r,echo=TRUE}
mean(mass$AGEP)
```
```

Oletame, et ülaltoodud ridu sisaldava faili nimi on “minuraport.Rmd”. Kui R-i konsoolile anda käsk

```
library(knitr) # kui pole arvutis, siis enne installeerida
knit("minuraport.Rmd")
```

tunneb R ära, et tegemist on **markdown**-märgenduskeeles koodiga (täpsemini R-i-spetsiifilise markdown keelega) ning töökausta genereeritakse fail **minuraport.md**, kus R-i käsud on asendatud tulemustega (nt **mean(x)** on asendatud: 3.5). Selle faili saab teegis **markdown**² oleva käsu **markdownToHTML(.)** abil muuta HTML-failiks:

```
library(markdown) # kui pole arvutis, siis enne installeerida
markdownToHTML("minuraport.md", output = "valmisraport.html")
browseURL("valmisraport.html")
```

Sama protseduuri on palju lihtsam teostada programmis RStudio: valida menüüst **File -> New file -> R Markdown**, panna faili kirja oma tekst ja R-i kood, mis teeb arvutused ning klõpsata nupul *Knit to HTML*, *Knit to PDF* (eelnevalt vaja arvutisse installeerida näiteks MikTeX) või *Knit to Word*. Vaata ka <http://rmarkdown.rstudio.com/>

*Praktikumijuhendid põhinevad aine MTMS.01.092 Rakendustarkvara: R (2 EAP) materjalidel, mille autorid on Mait Raag ning Raivo Kolde

¹<http://yihui.name/knitr/>

²<http://rmarkdown.rstudio.com/>

Ülalolevas näites on teksti vahele pikitud R-i koodi nn juppide ehk *chunk*idega. Jupp võib olla nn reasisene ehk *inline*, või täiesti eraldiseisev. Reasisene R-i koodijupp tuleb piirata graavise sümboliga (*backtick*), mida eesti paigutusega klaviatuuril leiab *backspace* klahvi kõrvalt; reasisese koodijupi alguses peab olema täht **r**.

Sageli on mõttekas kasutada täiesti eraldiseisvaid koodiplokke, mis tuleks piirata kolmekordse graavisega. Ploki alguses tuleb loogelistes sulgudes kõigepealt kirjutada täht **r**, seejärel võib kirjutada komaga eraldatult lisaargumente³. Mõned olulisemad:

- **echo** – kas väljundis peaks ka R-i kood olema näha (TRUE/FALSE)
- **eval** – kas väljundis peaks näha olema R koodi tulemus (TRUE/FALSE)
- **fig.width**, **fig.height** – kui koodiplokis tehakse joonis, siis mis mõõtmetega see peab olema.
- **results** – kuidas vormindada väljundit (konsoli väljatrükki); väärtus **'asis'** on paslik sel juhul, kui kasutame mingit spetsiifilist vorminduskäsku (nt **xtable(.)**).

Oluline on, et eraldiseisva koodiploki algus ja lõpumärgised oleks eraldi real ja ilma eelneva tühikuta. RStudio on koodiploki märgendeid lihtne lisada klahvikombinatsiooni **Ctrl + Alt + I** abil või vajutades vastavat nuppu (*Insert*) menüüreal.

Koodiplokkide vahelise teksti puhul mõned olulisemad märgendusvõtted⁴:

- **#** Pealkiri – esimese taseme pealkiri
- **##** alapealkiri – teise taseme pealkiri jne (kuni 6 taset)
- Uue lõigu alustamiseks jätta üks tühi rida vahele
- Nummerdamata loetelu elementide ette näiteks – või *****; loetelu ees peab olema tühi rida
- Nummerdatud loetelu kõigi elementide ette **1.** (kindlasti mitte **2.**)
- Käsitsi saab tabelit vormistada miinusemärgi ja püstkriipsude abil
- allmärkuseid saab nii: **mingitekst[^][allmärkuse tekst]**

Nii on võimalik R-i väljundit mugavasti ühte faili saada, et ei peaks pidevalt kopeerima-kleepima. Kuidas aga R-i produtseeritavad tabelid ilusaks saada? Lihtsamad tabelid saab kujundada **knitr** käsu **kable** abil. Näiteks kui eelnevalt on moodustatud tabel **tabel0**:

```
library(dplyr)
tabel0 <- mass %>% group_by("Kodakondsus" = CIT) %>%
  summarise("Keskmise vanus" = mean(AGEP),
            "Vanuse standardviga" = sd(AGEP)/sqrt(n()),
            "Soo ülekaal" = ifelse(sum(SEX == "Male") > n()/2, "Mehed", "Naised"))
```

siis tabeli esitamiseks väljundis on vaja lisada koodiplokk kujul (**format** määrab väljundi tüübi näiteks **latex**, **html**):

```
```{r koodijupp tabeli tegemiseks 1, results = 'asis'}
library(knitr)
kable(tabel0, digits = c(1, 2), align = c("lccr"), format = "markdown")
```
```

| Kodakondsus | Keskmise vanus | Vanuse standardviga | Soo ülekaal |
|---|----------------|---------------------|-------------|
| Born abroad of American parent(s) | 38.60 | 2.8 | Naised |
| Born in Puerto Rico, Guam, the U.S. Virgin Islands, | 45.04 | 2.1 | Naised |
| Born in the U.S. | 38.95 | 0.3 | Naised |
| Not a citizen of the U.S. | 37.91 | 0.8 | Mehed |
| U.S. citizen by naturalization | 50.27 | 0.9 | Naised |

³<http://yihui.name/knitr/options>

⁴http://rmarkdown.rstudio.com/authoring_basics.html

On olemas ka spetsiaalseid pakette R tabelite esitamiseks ja näiteks mudelite väljundi ilusaks esitamiseks. Üks sellistest pakettidest on **xtable**, mille kõige olulisem käsk `xtable(.)` produtseerib etteantud tabelist (või vähegi tabelit meenutavast objektist, nt `data.frame`'ist) sobiliku LaTeX või HTML koodi:

```
```{r koodijupp tabeli tegemiseks 2, results='asis'}
library(xtable)
tabel0 <- table(mass$CIT, mass$SEX)
tabel <- xtable(tabel0, caption = "Sagedustabel", align = c("r|cc"))
print(tabel, type = "latex", comment = FALSE) # type = "latex" tabel LaTeX koodis
```
```

| | Female | Male |
|---|--------|------|
| Born abroad of American parent(s) | 31 | 31 |
| Born in Puerto Rico, Guam, the U.S. Virgin Islands, | 46 | 33 |
| Born in the U.S. | 2785 | 2637 |
| Not a citizen of the U.S. | 211 | 237 |
| U.S. citizen by naturalization | 226 | 187 |

Table 2: Sagedustabel

Lineaarse mudeli koefitsentide esitus tabeli kujul:

```
```{r mudeli kordajad, results = 'asis'}
mudel <- lm(WAGP ~ WKHP, data = mass)
print(xtable(mudel))
```
```

| | Estimate | Std. Error | t value | Pr(> t) |
|--|-------------|------------|---------|----------|
| (Intercept) | -71544.4589 | 11165.5063 | -6.41 | 0.0000 |
| AGEP | 3143.3746 | 345.3799 | 9.10 | 0.0000 |
| I(AGEP^2) | -31.2506 | 3.9803 | -7.85 | 0.0000 |
| LANXYes, speaks another language | -8156.2759 | 3144.4080 | -2.59 | 0.0095 |
| CITBorn in Puerto Rico, Guam, the U.S. Virgin Islands, | -14376.4311 | 12470.3254 | -1.15 | 0.2490 |
| CITBorn in the U.S. | 451.5616 | 8806.3954 | 0.05 | 0.9591 |
| CITNot a citizen of the U.S. | 533.0460 | 9357.1599 | 0.06 | 0.9546 |
| CITU.S. citizen by naturalization | 1321.1767 | 9383.0085 | 0.14 | 0.8880 |
| SEXMale | -12695.2778 | 5180.4127 | -2.45 | 0.0143 |
| WKHP | 1112.2084 | 95.5204 | 11.64 | 0.0000 |
| SEXMale:WKHP | 652.1225 | 128.8311 | 5.06 | 0.0000 |