

## Seznámení s R

---

### 1 Úvod

- 1) Postup prací pro cvičení se bude postupně objevovat na webu:

<https://web.natur.cuni.cz/uamvt/turcm6am>

- 2) V počítači si vytvořte složku k tomuto cvičení, v tomto textu ji nazveme např. `biostat`.
- 3) Ve složce `biostat` si vytvořte složku `data` a do ní zkopírujte soubor `Deti23.csv`, který je ke stažení na stránkách výše (viz položka `data` u materiálů k 1. cvičení).
- 4) Do složky `biostat` si dále zkopírujte soubor `PrF01.R` (položka `skript` mezi materiály k 1. cvičení), který můžete též stáhnout z webové stránky výše.
- 5) Spust'te `RStudio` (stručný návod na jeho instalaci najdete na stránkách výše v záložce "Odkazy").
- 6) V `RStudio` nastavte coby pracovní adresář složku `biostat`, a to jedním z následujících způsobů:

- (a) Napište a pomocí `Enter` odešlete v okně `Console` příkaz

```
setwd("popis_cesty/biostat")
```

kde `popis_cesty` nahrad'te za popis cesty ke složce `biostat`. (Lomítka v popisu cesty musejí být dopředná, tj. `"/`, nikoli zpětná.)

- (b) V horní nabídce postupně zvolte

`Session` ➔ `Set Working directory` ➔ `Choose directory...`

a pomocí myši najdete složku `biostat`.

- 7) Zjistěte, zda se vše povedlo. Po zadání příkazu

```
getwd()
```

by se mělo vypsát `popis_cesty/biostat`.

- 8) Pomocí horní nabídky zaveďte do skriptového okna (alias `Script Window`, vlevo nahoře) skriptový soubor `PrF01.R`:

`File` ➔ `Open File...`

## Několik poznámek k práci se skriptovým souborem

- ✧ Při psaní příkazů v [Script Window](#) jsou vám automaticky nabízeny příkazy začínající na daná písmena. Pokud se vám některý z těchto příkazů zamlouvá, vyberte ho pomocí šipky na klávesnici a poté potvrďte klávesou **Enter**.
- ✧ Znak **#** odděluje poznámky, tj. něco, co chcete v souboru vidět vy, ale nechcete, aby to vidělo **R**.
- ✧ Různé příkazy musí být na různých řádcích (doporučuji), nebo na stejném řádku oddělené středníkem.
- ✧ Když chceme, aby se provedl příkaz na konkrétním řádku, nastavíme kurzor na příslušný řádek a zmáčkneme klávesovou zkratku **Ctrl-Enter**, nebo klikneme na tlačítko **Run** (v pravém horním rohu [Script Window](#)).
- ✧ Když chceme, aby se provedlo více příkazů najednou, označíme je myší jako blok a zmáčkneme klávesovou zkratku **Ctrl-Enter**, nebo klikneme na tlačítko **Run** (v pravém horním rohu [Script Window](#)).
- ✧ Čas od času si (doplňovaný) skriptový soubor uložte (stačí sem tam stisknout klávesovou zkratku **Ctrl-S**) nebo kliknout na ikonu diskety vlevo nahoře.
- ✧ Nápoředu k libovolnému příkazu vyvoláme pomocí **help(prikaz)** nebo **?prikaz**. (Nápoředa se pak zobrazí v pravém dolním okně.)

## Další poznámky pro práci s R / RStudiem

- ✧ Všechny požadované příkazy je nutné napsat do [Script Window](#) nebo do [Console](#).
- ✧ Po odeslání požadovaného příkazu se příkaz v tichosti provede a v [Console](#) se objeví výsledky výpočtu (jsou-li nějaké). V případě chybného příkazu se objeví červená chybová hláška.
- ✧ Předchozí příkazy je možné v [Consoli](#) vyvolávat pomocí "šipky nahoru" na klávesnici. Vyvolaný příkaz lze dále upravit a odeslat klávesou **Enter**.
- ✧ Chceme-li provést ve výpočtu drobnou změnu (např. zpracovat proměnnou **vaha** místo proměnné **vyska**), můžeme si příslušnou část ve skriptovém okně zkopírovat níže, upravit a odeslat.
- ✧ Pro pozdější použití vytvořeného skriptu je dobré si do něj sem tam napsat nějaký komentář (za znak **#**).

## 2 Základní operace a funkce v R

- 1) Nejprve použijte R jako kalkulačku a spusťte postupně ze skriptového okna následující příkazy:

```
2
2 + 3          # scitani
2 - 3          # odcitani
2 * 3          # nasobeni
2 / 3          # deleni
2^3           # mocnina
3^(1/2)        # (od)mocnina
sqrt(3)        # jeste jednou odmocnina
exp(3)         # exponenciela (e na neco)
log(3)         # prirodzeny logaritmus (pozor! jinde se casto znaci ln)
exp(log(3))
sin(3)         # sinus
cos(3)         # cosinus
pi             # konstanta pi
sin(pi/2)
cos(pi/2)
factorial(3)   # 3! (faktorial)
choose(3, 2)   # "3 nad 2 (kombinacni cislo)"
```

- 2) Nyní budeme pracovat s proměnnými

```
a <- 2        # ulozeni hodnoty do promenne
a = 2         # jiny zpusob ulozeni (nelze kombinovat s jinymi prikazy, viz nize)
a             # podivame se, co tam mame
b <- 3        # ulozeni do jine promenne
b
a + b         # secteni ulozenych hodnot
a - b         # odedcteni ulozenych hodnot
a * b         # vynasobeni ulozenych hodnot
a / b         # vydeleni ulozenych hodnot
```

- 3) Ukládání výsledku do dalších proměnných

```
c <- log(b)   # ulozi, ale nevytiskne vysledek
c             # vytiskne vysledek
print(c)     # tez vytiskne vysledek
```

- 4) „Úklid“ (před začátkem jiné práce)

```
ls()          # zjistime, jake promenne mame nadefinovane
rm(list=ls()) # vycistime Rko (vsechny promenne se smazou)
ls()          # zjistime, jake promenne mame nactene (snad nic)
```

- 5) Do budoucna bude vhodné, když každý vámi vytvořený skriptový soubor pro novou práci bude začínat příkazy

```
setwd("popis_cesty/biostat")
rm(list=ls())
```

kde `popis_cesty` nahradíte za popis cesty ke složce `biostat`.

### 3 Načtení dat

- ✧ Data (nějaká vlastní měření apod.) si asi většinou ukládáte v nějakém tabulkovém procesoru (LibreOffice Calc, MS Excel apod.)
- ✧ Z tabulkového procesoru lze (rozumně uložená) data uložit ve formě textového souboru (např. ve formátu `csv`), obvykle nabídkou **Save As** / **Uložit jako**.
- ✧ Z formátu `csv` (a jemu podobných) lze již data snadno načíst do **R**.
- ✧ Podívejme se na data uložená v souboru `Deti23.csv`, která obsahují následující proměnné:

<code>hoch</code>	jde o hocha? (0/1);
<code>poradi</code>	o kolikáté dítě dané matky jde;
<code>vekMatky</code>	věk matky;
<code>vekOtce</code>	věk otce;
<code>vaha</code>	váha dítěte v 1 roce ( <i>kg</i> );
<code>delka</code>	délka dítěte v 1 roce ( <i>cm</i> );
<code>hcd</code>	počet onemocnění horních cest dýchacích v prvních 12 měsících života;
<code>Pohlavi</code>	pohlaví dítěte ( <i>M/F</i> ).

- 1) Načtete data do **R**.

Bud' v pravém dolním okně v záložce **Files** klikněte na příslušná data a zvolte **Import Dataset** a nastavte:

<b>Name</b>	Deti23
<b>Delimiter</b>	Semicolon (středník)
<b>First Row as Names</b>	✓

Nebo použijte příkaz:

```
Deti23 <- read.csv2("data/Deti23.csv", header=TRUE)
```

- 2) Data si můžete prohlédnout pomocí

```
View(Deti23)
```

nebo

```
print(Deti23)
```

### 4 Výpočet nových proměnných

- 3) V datech je délka dětí uvedena v centimetrech. Co kdybychom chtěli zpracovávat délku v metrech? Vytvořme novou proměnnou, kterou nazveme `delka.m`, jež bude udávat délku dětí v metrech.

```
Deti23 <- transform(Deti23, delka.m = delka / 100) # nebo
Deti23$delka.m <- Deti23$delka / 100
```



- 4) Prohlédněte si znovu data a zjistěte, zda se vše povedlo.



- 5) Samostatně vytvořte proměnnou nazvanou `bmi`, jež bude udávat BMI (body mass index = (hmotnost v kg)/(výška v m)<sup>2</sup>) jednotlivých dětí, tj. `bmi = vaha / delka.m2`.



- 6) Samostatně vytvořte proměnnou `dobaPlnolet`, která bude udávat, kolik let již byla matka při porodu plnoletá, tj. `dobaPlnolet = vekMatky - 18`.

## 5 Přístup k jednotlivým položkám dat

Abyste nemuseli při přístupu k jednotlivým proměnným stále opisovat `Deti23$`, je možné zadat

```
attach(Deti23)
```

a pak lze již názvy jednotlivých proměnných volat přímo.

- 7) Ze skriptového okna vytiskněte hodnoty proměnných `bmi` a `doBaPlnolet`:

```
bmi  
doBaPlnolet
```

Notace

```
Deti23$bmi  
Deti23$doBaPlnolet
```

přítom funguje stále!

- 8) K jednotlivým složkám vektoru `bmi` se lze dostat operátorem `[ ]`.

✧ Například BMI 10. dítěte dostaneme pomocí

```
bmi[10]
```

✧ Hodnoty BMI prvních 5 dětí lze dostat takto:

```
bmi[1:5]
```

✧ Hodnoty BMI pro 2., 5., a 7. dítě získáme zase tímto způsobem:

```
bmi[c(2, 5, 7)]
```

- 9) Obdobně se lze dostat ke konkrétním naměřeným údajům konkrétního dítěte, či konkrétních dětí.

✧ Hodnotu 5. proměnné (`vaha`) 10. dítěte získáme takto:

```
Deti23[10, 5]
```

nebo také takto

```
Deti23[10, "vaha"]
```

✧ Hodnotu 5. a 6. proměnné (`vaha` a `delka`) 10. dítěte získáme takto:

```
Deti23[10, c(5, 6)]
```

nebo také takto

```
Deti23[10, c("vaha", "delka")]
```

✧ Všechny údaje 10. dítěte (jde o 73 cm dlouhého a 9,6 kg těžkého chlapce) dostaneme takto:

```
Deti23[10, ]
```

✧ Všechny údaje 7.–11. dítěte dostaneme takto:

```
Deti23[7:11, ]
```

✧ Všechny údaje 2., 5., a 12. dítěte dostaneme takto:

```
Deti23[c(2, 5, 12), ]
```

✧ Hodnotu 5. a 6. proměnné (**vaha** a **delka**) 2., 5., a 12. dítěte dostaneme takto:

```
Deti23[c(2, 5, 12), c(5, 6)]
```

nebo také takto

```
Deti23[c(2, 5, 12), c("vaha", "delka")]
```

## 6 Uložení dat

Do datové tabulky `Deti23` jsme přidali několik nových proměnných (`delka.m`, `dobaPlnolet`, `bmi`). V budoucnu s nimi možná budeme znovu pracovat. Abychom si tyto proměnné nemuseli vytvářet znovu, uložíme si současnou formu datové tabulky `Deti23`:

1) buď ve formátu csv

```
write.table(Deti23, file = "data/Deti23upr.csv", row.names=FALSE,  
            col.names=TRUE, sep=";", dec=",")
```

2) nebo ve formě R datového formátu (přípona `RData`)

```
save(Deti23, file = "data/Deti23.RData")
```

## 7 Konec práce

Než zavřete všechna okna, nezapomeňte si uložit skriptový soubor:

**File** ➡ **Save as**

nebo klávesovou zkratkou `Ctrl+s`. (Při standardním zavírání otevřených oken budete tak jako tak dotázáni, zda chcete tak učinit).