

Ústav matematiky a statistiky  
Přírodovědecká fakulta  
Masarykova univerzita

---

## Aplikovaná statistika II – cvičení

*Pokyny k domácímu úkolu  
jarní semestr 2021*

Stanislav Katina, Veronika Bendová,  
Zdeňka Geršlová, Markéta Janošová a Vojtěch Šindlář

23. května 2021

## Instrukce

**Vzorový domácí úkol.** V souboru `stka-vzor-du-template.pdf` máte k dispozici vzorovou šablonu pro domácí úkol, vygenerovanou z následujících `*.tex` souborů:

1. `stka-vzor-du-template.tex`,
2. `stka-vzor-text-template.tex` a
3. `stka-vzor-title-page-template.tex`.

V odevzdávaném pdf souboru s domácím úkolem zachovejte styl použitý v šablonách.

Domácí úkol odevzdejte v jedné ze dvou níže uvedených forem. V názvech souborů nepoužívejte diakritiku a dodržujte velká a malá písmena podle návodu.

### 1. Forma Sweave

Tento způsob kombinuje k vytvoření řešení RSkript a flexibilní systém Sweave. Odevzdává se jeden pdf soubor nazvaný `UCO-prijmeni-jmeno-predmet-2021.pdf` (obsahuje řešení příkladů, tabulky, obrázky, komentáře a náhled `R`-kódu), jeden zdrojový soubor naprogramovaných funkcí `UCO-prijmeni-jmeno-funkce-predmet-2021.R` (jen pokud používáte vlastní naprogramované funkce) a jeden Sweave soubor `UCO-prijmeni-jmeno-predmet-2021.Rnw`, z něhož byl vygenerován výsledný pdf soubor a který využívá zdrojový soubor naprogramovaných funkcí. V R Sweave se při používání  $\LaTeX$  šablon postupuje identicky jako v  $\LaTeX$ u.

K vygenerování `R`-kódu v požadované formě použijte v  $\LaTeX$ -ovské hlavičce Rnw dokumentu balíček `listings`. Následujícím kódem umístěným taktéž v  $\LaTeX$ -ovské hlavičce Rnw dokumentu upravíte původní nastavení vzhledu `R`-kódu a `R`-výstupů do požadované formy.

```

1 \definecolor{dgray}{gray}{0.35} % barva textu komentaru
2 \definecolor{lgray}{gray}{0.95} % barva pozadi R-kodu
3 \definecolor{llgray}{gray}{0.98} % barva pozadi R-vystupu
4
5 \lstdefinestyle{Rstyle}{ % nastaveni vzhledu R-kodu
6 language=R, % nastaveni jazyka R
7 basicstyle=\ttfamily\small, % typ a velikost pisma R-kodu
8 backgroundcolor=\color{lgray}, % barva pozadi R-kodu
9 commentstyle=\ttfamily\small\itshape\color{dgray}, % barva komentare k funkcim
10 showstringspaces=false, % zakaz zvyraznovani mezer
11 numbers=left, % cislovani vlevo
12 numberstyle=\ttfamily\small, % typ pisma a velikost cislovani
13 stepnumber=1, % cislovani po kroku jedna
14 firstnumber=last, % kumulativni cislovani radku v po sobe nasledujicich Chunk prostedich
15 breaklines=T} % automaticke zalamovani kodu na konci radku
16
17 \lstdefinestyle{Routstyle}{ % nastaveni vzhledu R-vystupu
18 language=R, % nastaveni jazyka R
19 basicstyle=\ttfamily\small, % typ a velikost pisma R-vystupu
20 backgroundcolor=\color{llgray}, % barva pozadi R-vystupu
21 showstringspaces=true, % zakaz zvyraznovani mezer
22 numbers=right, % cislovani vpravo
23 numberstyle=\ttfamily\small, % typ pisma a velikost cislovani
24 firstnumber=last, % kumulativni cislovani radku v po sobe nasledujicich Chunk prostedich
25 breaklines=T} % automaticke zalamovani kodu na konci radku

```

Dále je potřeba nastavit, aby byl balíček `listings` i s výše uvedenými nastaveními použit při překládání Rnw souboru do pdf souboru. Toto nastavení již vkládáme do těla dokumentu za příkaz `\begin{document}`.

```

26 << setup >>= # Setup Chunk
27 render_listings()
28 @

```

Po přeložení Rnw souboru se náhled **R**-kódu automaticky zobrazí ve výsledném pdf souboru, pokud v hlavičce Chunk prostředí, obsahujícím **R**-kód, nastavíme argument `echo=T`.

```

29 << echo=T >>=
30 x <- 1:10
31 plot(x)
32 @

```

Další informace o systému Sweave najdete např. zde: [Chunk options and package options](#).

Snažte se **R**-kód psát co nejvíce přehledně. Případné tipy jak postupovat při úpravě kódu můžete nalézt v prezentaci [Standards of programming in R: R style guide](#).

## 2. Forma L<sup>A</sup>T<sub>E</sub>X

Tento způsob kombinuje k vytvoření řešení RSkript a L<sup>A</sup>T<sub>E</sub>X. Odevzdává se jeden pdf soubor nazvaný UCO-prijmeni-jmeno-predmet-2021.pdf (obsahuje řešení příkladů, tabulky, obrázky, **R**-kód napsaný v L<sup>A</sup>T<sub>E</sub>Xu), jeden zdrojový soubor naprogramovaných funkcí UCO-prijmeni-jmeno-funkce-predmet-2021.R (jen pokud používáte vlastní naprogramované funkce) a jeden soubor **R**-kódu konkrétních řešení DÚ UCO-prijmeni-jmeno-predmet-2021.R, který používá tento zdrojový kód. Na psaní **R**-kódu použijte L<sup>A</sup>T<sub>E</sub>X-ovský balíček `listings` k vytvoření prostředí v hlavičce dokumentu pomocí následujícího kódu:

```

1 \definecolor{dgray}{gray}{0.35} % barva textu komentaru
2 \definecolor{lgray}{gray}{0.95} % barva pozadi R-kodu
3
4 \lstset{ % nastaveni vzhledu R-kodu
5 language=R, % nastaveni jazyka R
6 basicstyle=\ttfamily\small, % typ a velikost pisma R-kodu
7 backgroundcolor=\color{lgray}, % barva pozadi R-kodu
8 commentstyle=\ttfamily\small\itshape\color{dgray}, % barva komentare k funkcim
9 showstringspaces=false, % zakaz zvyraznovani mezer
10 numbers=left, % cislovani vlevo
11 numberstyle=\ttfamily\small, % typ pisma a velikost cislovani
12 stepnumber=1, % cislovani po kroku jedna
13 firstnumber=last, % kumulativni cislovani radku v po sobe nasledujicich castech s R-kodem
14 breaklines=T} % automaticke zalamovani kodu na konci radku

```

V textu potom **R**-kód vkládáme do prostředí `\begin{lstlisting}` a `\end{lstlisting}`. Snažte se **R**-kód psát co nejvíce přehledně. Případné tipy jak postupovat při úpravě kódu můžete nalézt v prezentaci [Standards of programming in R: R style guide](#).

Pokud nemáte nainstalovaný L<sup>A</sup>T<sub>E</sub>X, můžete pro vygenerování souboru UCO-prijmeni-jmeno-predmet-2021.pdf s textem domácího úkolu použít **Overleaf**.

- Pod následujícím odkazem je vám k dispozici připravená šablona pro vytvoření domácího úkolu: <https://www.overleaf.com/latex/templates/sablona-du-bi8773/rthxjsgnfcjy>  
Klikněte na **Open as Template**, čímž dojde k otevření šablony v editovatelné podobě (pro editaci musíte být na Overleafu přihlášení, registrace je zdarma).
- V souboru UCO-prijmeni-jmeno-predmet-title-page.tex modifikujte následující

```

15 \textbf{Nazev predmetu}
16 \textbf{Jmeno Prijmeni}
17 \textbf{UCO}
18 Obor XY

```

- V souboru UCO-prijmeni-jmeno-predmet-2021.tex vyplňte následující (řádky 18 a 19)

```
19 \fancyhead[L]{Název predmetu -- domaci ukol}
20 \fancyhead[R]{Jmeno Prijmeni}
```

- Pro psaní ve slovenštině v hlavičce souboru UCO-prijmeni-jmeno-predmet-2021.tex namísto

```
21 \usepackage[czech]{babel} %% zabezpeci ceske nastaveni
```

použijte

```
22 \usepackage[slovak]{babel} %% zabezpeci slovenske nastaveni
```

- Přejmenujte všechny soubory tak, aby obsahovaly vaše UČO, příjmení a jméno (malými písmeny bez diakritiky dle vzoru) a zkratku předmětu (pro <Název předmětu> použijte zkratka).

- V souboru UCO-prijmeni-jmeno-predmet-2021.tex (řádky 48 a 50) rovněž přepište

```
23 \input{UCO-prijmeni-jmeno-predmet-title-page} %% nacteni souboru s titulni strankou
24 \input{UCO-prijmeni-jmeno-predmet-text} %% nacteni souboru s hlavnim textem ukolu
```

na své názvy souborů z předchozího kroku.

- Samotný text (včetně obrázků, tabulek atd.) píšete v souboru UCO-prijmeni-jmeno-predmet-text.tex. Je zapotřebí postupovat takto:

- text svého projektu píše buď v módu **Source** nebo **Rich Text**,

- vkládání obrázků – vedle ikony **files** je šipka a z vyrolovaného menu vyberete **Computer** a uploadujete své obrázky jako **\*.pdf**.

- použití obrázků – příklad pro  logo v textu

```
25 \includegraphics[angle=0,width=0.025\textwidth]{Rlogo.pdf}
```


Argument **width** určuje, jaká proporce šířky textu na stránce odpovídá šířce obrázku.

- použití obrázků – příklad pro samostatný obrázek

```
26 %% prostředí obrazku
27 \begin{figure}[ht]
28 \centering
29 \includegraphics[angle=0,width=0.45\textwidth]{nazev-obrazku}
30 \caption{Popisek ...}
31 \end{figure}
```

- použití tabulek – příklad

```
32 %% prostředí tabulky
33 %% zarovnani vpravo (r), pocet pismen "r" predstavuje pocet sloupcu
34 %% h - here, na tomto miste, t - top, v horni casti stranky
35 %% velikost pisma \footnotesize (10pt), \scriptsize (8pt)
36 \begin{table}[ht]
37 \caption{Popisek ...}
38 \footnotesize
39 \centering
40 \begin{tabular}{r|rrr|rrr}
41 %% telo tabulky
42 \end{tabular}
43 \end{table}
```

Export tabulek z  umožňuje knihovna **xtable** a její funkce **xtable**. Nastavení počtu desetinných míst je možné pomocí argumentu **digits**, kde první číslo vektoru je nula, neboť popis řádků je text.




– Vkládání -kódu umožňuje prostředí `listings`

```
44 %% prostředí pro R-kod
45 \begin{lstlistings}
46 %% R kód
47 \end{lstlistings}
```


Ukázku vloženého kódu, obrázku i tabulky najdete přímo v šabloně v souboru UCO-prijmeni-jmeno-predmet-text.tex, popřípadě v souboru stka-vzor-text-template.tex.

- Po dokončení domácího úkolu můžete vytvořený pdf soubor stáhnout kliknutím na šipku `Download PDF` (přímo nad náhledem pdf souboru), celý adresář včetně zdrojových souborů a obrázků exportujete kliknutím na `Menu` a zvolením možnosti `Download - Source`.
- Bližší informace o  $\text{\LaTeX}$ -u najdete např. zde: The Not So Short Introduction to  $\text{\LaTeX}$ .

### 3. Forma MS Word (nebo LibreOffice apod.)

Odevzdává se jeden doc (nebo docx) soubor nazvaný UCO-prijmeni-jmeno-predmet-2021.doc (obsahuje řešení příkladů, tabulky, obrázky a -kód), jeden zdrojový soubor naprogramovaných funkcí UCO-prijmeni-jmeno-funkce-predmet-2021.R (jen pokud používáte vlastní naprogramované funkce) a jeden soubor -kódu konkrétních řešení DÚ UCO-prijmeni-jmeno-predmet-2021.R, který používá tento zdrojový kód. -kód přehledným způsobem odlište od zbylého textu.

*DÚ je hodnocen cvičícím. Toto hodnocení zahrnuje:*

1. *přítomnost dvou (případně tří) výše zmíněných souborů a jejich názvy (při uploadu se nezaškrtně "přidat UČO, příjmení a jméno" a uploadujte jednotlivé soubory, nikoli \*.zip, \*.rar či jiné archivy),*
2. *kompletnost zpracování (každý příklad musí být vypracovaný, žádný nesmí chybět),*
3. *dostatečný opis Vašich úvah, zvoleného postupu a interpretace výsledků, ať už tabulkových nebo grafických,*
4. *přehlednost -kódu.*

*Pro udělení zápočtu je třeba odevzdat vyřešený domácí úkol.*

## Zadání

**Příklad 1** *Datový soubor diet v knihovně WRS2 v R obsahuje údaje o úbytku váhy v kilogramech pro tři různé typy diet.<sup>1</sup> Vaším úkolem je zjistit vliv typu diety (proměnná `diet.type`) a pohlaví (proměnná `gender`) na úbytek hmotnosti (proměnná `weight.loss`).*

1. Zjistěte počty pozorování, výběrové průměry a výběrové směrodatné odchylky úbytku váhy pro muže a ženy v závislosti na typu diety (tj. celkem šest kombinací).
2. Vykreslete krabicové diagramy popisující úbytek hmotnosti v závislosti na typu diety a pohlaví.
3. Modelujte závislost střední hodnoty úbytku hmotnosti na typu diety a pohlaví. Vyzkoušejte různé varianty složitosti modelu: (1) model se vzájemnou interakcí obou faktorů a (2) model bez interakce. Vyberte vhodnější z nich a své rozhodnutí zdůvodněte a podpořte příslušným výstupem z R. V následujících bodech pracujte s tímto modelem.
4. Proveďte Shapirův-Wilkův test pro ověření normality a graficky ověřte zbývající předpoklady metody. Okomentujte výsledky.
5. Proveďte mnohonásobné porovnávání pomocí Tukeyho metody. Jaké hypotézy tato metoda zamítá?
6. Vykreslete do grafu zvlášť pro ženy (červená lomená čára) a zvlášť pro muže (modrá lomená čára) odhadnuté střední hodnoty úbytku hmotnosti pro všechny druhy diet. Na ose  $x$  budou úrovně faktoru `diet.type` a na ose  $y$  úbytek hmotnosti.

Veškeré testování hypotéz provádějte na hladině významnosti 0,05.

**Příklad 2** *V souboru Howell.csv máme k dispozici kraniometrické rozměry z různých populací. Zaujímají nás ženy (kategorie  $F$  proměnné `Sex`) z populací ZULU, EGYPT a AUSTRALI a následující rozměry:*

- *NPH: výška horní části obličeje (v mm),*
- *NLH: výška nosu (v mm),*
- *ZYB: bizygomatická šířka (v mm),*
- *OBB: šířka očnice (v mm).*

*Chceme zjistit, zda vektory středních hodnot těchto proměnných jsou pro všechny tři populace stejné. Nezapomeňte, že chybějící hodnoty jsou v této databázi označeny 0.*

1. Pro každou populaci najděte počet pozorování, vektor výběrových průměrů a varianční matici.
2. Vykreslete krabicové diagramy pro jednotlivé proměnné podle populací.
3. Ověřte předpoklady metody.

---

<sup>1</sup>Další informace o datech najdete na <https://www.rdocumentation.org/packages/WRS2/versions/1.0-0/topics/diet>

4. *Otestujte hypotézu na hladině významnosti 0,05.*
5. *Pokud nulovou hypotézu zamítnete, zjistěte, které proměnné způsobují rozdíly mezi všemi populacemi.*
6. *Pokud nulovou hypotézu zamítnete, zjistěte, které populace se navzájem liší.*
7. *U populací, které se navzájem liší, zjistěte, které proměnné to způsobují.*