

Lineární statistické modely II

*Pokyny k domácímu úkolu
jarní semestr 2023*

Stanislav Katina a Vojtěch Šindlář

25. dubna 2023

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 **R**-skript a flexibilní systém Sweave. Odevzdává se jeden pdf soubor nazvaný `LSM2-2023.pdf` (obsahuje řešení příkladů, tabulky, obrázky, komentáře a náhled **R**-kódu), jeden zdrojový soubor naprogramovaných funkcí `funkce-LSM2-2023.R` a jeden Sweave soubor `LSM2-2023.Rnw`, z něhož byl vygenerován výsledný pdf soubor a který využívá zdrojový soubor naprogramovaných funkcí. V **R** Sweave se při používání **L**^AT_EX šablon postupuje identicky jako v **L**^AT_EXu.

K vygenerování **R**-kódu v požadované formě použijte v **L**^AT_EX-ovské hlavičce `Rnw` dokumentu balíček `listings`. Následujícím kódem umístěným taktéž v **L**^AT_EX-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 pozadí R-kódu
3 \definecolor{llgray}{gray}{0.98} % barva pozadí R-vystupu
4
5 \lstdefinestyle{Rstyle}{% nastavení vzhledu R-kódu
6 language=R, % nastavení jazyka R
7 basicstyle=\ttfamily\small, % typ a velikost písma R-kódu
8 backgroundcolor=\color{lgray}, % barva pozadí R-kódu
9 commentstyle=\ttfamily\small\itshape\color{dgray}, % barva komentáře k funkcím
10 showstringspaces=false, % zakaz zvyrazňování mezer
11 numbers=left, % číslování vlevo
12 numberstyle=\ttfamily\small, % typ písma a velikost číslování
13 stepnumber=1, % číslování po kroku jedna
14 firstnumber=last, % kumulativní číslování řádku v po sobě následujících Chunk prostedích
15 breaklines=T} % automatické zalamování kódu na konci řádku
16
17 \lstdefinestyle{Routstyle}{% nastavení vzhledu R-vystupu
18 language=R, % nastavení jazyka R
19 basicstyle=\ttfamily\small, % typ a velikost písma R-vystupu
20 backgroundcolor=\color{llgray}, % barva pozadí R-vystupu
21 showstringspaces=true, % zakaz zvyrazňování mezer
22 numbers=right, % číslování vpravo
23 numberstyle=\ttfamily\small, % typ písma a velikost číslování
24 firstnumber=last, % kumulativní číslování řádku v po sobě následujících Chunk prostedích
25 breaklines=T} % automatické zalamování kódu na konci řádku



```

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 -kódu automaticky zobrazí ve výsledném pdf souboru, pokud v hlavičce Chunk prostředí, obsahujícím -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](#).

Při psaní -kódu postupujte podle instrukcí v prezentaci [Standards of programming in R: R style guide](#).

2. Forma L^AT_EX

Tento způsob kombinuje k vytvoření řešení RSkript a L^AT_EX. Odevzdává se jeden pdf soubor nazvaný `LSM2-2023.pdf` (obsahuje řešení příkladů, tabulky, obrázky, -kód napsaný v L^AT_EXu), jeden zdrojový soubor naprogramovaných funkcí `funkce-LSM2-2023.R` a jeden soubor -kódu konkrétních řešení DÚ `LSM2-2023.R`, který používá tento zdrojový kód. Na psaní -kódu použijte L^AT_EX-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 -kód vkládáme do prostředí `\begin{lstlisting}` a `\end{lstlisting}`. Při psaní -kódu postupujte podle instrukcí v prezentaci [Standards of programming in R: R style guide](#).

Pokud nemáte nainstalovaný L^AT_EX, můžete pro vygenerování souboru `LSM2-2023.pdf` s textem domácího úkolu použít **Overleaf**.

- Na Overleaf si vyberte **template UWE dissertation report**. Tím dojde k vytvoření projektu, který pojmenujte např. `DU-LSM2`. Automaticky se vytvoří adresář `files` s několika soubory, z nichž některé je nutné přejmenovat a následně nahradit jejich obsah obsahem vzorových souborů a jiné vymazat takto:

- přejmenujte `main.tex` na `LSM2-2023.tex`,
- přejmenujte `Chapter1.tex` na `LSM2-text.tex`,
- přejmenujte `titlepage.tex` na `LSM2-title-page.tex`,
- obsah souboru `LSM2-2023.tex` nahraďte obsahem souboru `stka-vzor-du-template.tex`,
- obsah souboru `LSM2-text.tex` nahraďte obsahem souboru `stka-vzor-text-template.tex`,
- obsah souboru `LSM2-title-page.tex` nahraďte obsahem souboru `stka-vzor-title-page-template.tex`,
- vymažte soubory `Abstract.tex`, `biblio.bib` a `references.bib`.

- V souboru LSM2-title-page.tex modifikujte následující

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

- V souboru LSM2-2023.tex vyplňte následující

```
19 \fancyhead[L]{Nazev predmetu} %% hlavicka vlevo
20 \fancyhead[R]{Jmeno Prijmeni} %% hlavicka vpravo
```

- V souboru LSM2-2023.tex modifikujte následující

```
21 \input{stka-vzor-title-page-template} %% nacteni souboru s titulni strankou
22 \input{stka-vzor-text-template} %% nacteni souboru s hlavnim textem ukolu
```

zaměňte za své názvy souborů

```
23 \input{LSM2-title-page} %% nacteni souboru s titulni strankou
24 \input{LSM2-text} %% nacteni souboru s hlavnim textem ukolu
```

- Pro psaní ve slovenštině v hlavičce souboru LSM2-2023.tex namísto

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

použijte

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

- V souboru LSM2-text.tex je zapotřebí postupovat takto:

- text svého projektu pište 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

```
27 \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

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

- použití tabulek – příklad

```
34 %% prostredi tabulky
35 %% zarovnani vpravo (r), pocet pismen "r" predstavuje pocet sloupctu
36 %% h - here, na tomto miste, t - top, v horni casti stranky
37 %% velikost pisma \footnotesize (10pt), \scriptsize (8pt)
38 \begin{table}[ht]
39 \caption{Popisek ...}
40 \footnotesize
41 \centering
42 \begin{tabular}{r|rrr|rrr}
43 %% telo tabulky
44 \end{tabular}
45 \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`

```
46 %% prostredi pro R-kod
47 \begin{lstlistings}
48 %% R kod
49 \end{lstlistings}
```

Ukázku vloženého kódu najdete v souboru `stka-vzor-text-template.tex`.

- Po dokončení domácího úkolu exportujete celý adresář DU-LSM2 (obsahující zdrojové soubory, obrázky) kliknutím na šipku pod ikonou `DOWNLOAD AS ZIP`, kde vyberete možnost `Input and Output Files`.
- Bližší informace o \LaTeX -u najdete např. zde: The Not So Short Introduction to \LaTeX .

DÚ je nejprve po formální stránce hodnocen cvičícím. Toto hodnocení zahrnuje:

1. *přítomnost 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 a dodržování instrukcí v prezentaci Standards of programming in R: R style guide.*

Pro udělení zápočtu je třeba odevzdat vyřešený domácí úkol.

Zadání

Příklad 1 Mějme lineární regresní model tvaru


$$Y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \varepsilon_i, \quad i = 1, 2, \dots, n,$$


kde pro náhodné chyby platí $\varepsilon_i \stackrel{iid}{\sim} N(0, \sigma^2)$. Na základě Scheffého věty odvoďte tvar pásů spolehlivosti tak, aby pravděpodobnost pokrytí byla $100 \cdot (1 - \alpha)\%$, kde $\alpha \in (0; 1)$ je předem zvolená hladina významnosti.

Příklad 2 V datovém souboru `baseball_hit.Rdata` jsou zaznamenány údaje o baseballových odpalech. Máme k dispozici proměnnou vzdálenost (`distance`), která udává horizontální vzdálenost (v metrech) mezi pálkařem a dopadem míčku, a proměnnou úhel (`angle`), jež zachycuje velikost úhlu (ve stupních) mezi trajektorií míčku a zemí při odpalu.

- Najděte vhodný lineární regresní model pro popis závislosti délky odpalu na úhlu odpalu. V následujících bodech pracujte s tímto modelem.
- Sestrojte pásy spolehlivosti s pravděpodobností pokrytí 95 %.
- Predikujte (bodově i intervalově) vzdálenost dopadu míčku od pálkaře pokud odpálí míček pod úhlem 55° .
- Vypočtěte simultánní intervalový odhad střední hodnoty délky odpalu pro úhly odpalu 20° , 25° a 30° . Zvolte hladinu významnosti $\alpha = 0,05$ a využijte Bonferroniho adjustaci.
- Zakreslete výsledky z předchozích bodů do jednoho obrázku. Tj. vykreslete bodový graf pozorování, přidejte křivku vašeho modelu, zvýrazněte pás spolehlivosti a požadované intervalové odhady i predikce.

Příklad 3 Pracujte s antropometrickými údaji studentů vysokých škol uloženými v souboru `lrm-foot.txt`. Data obsahují proměnné: pohlaví (`sex`), délku chodidla v milimetrech (`foot.L`) a tělesnou výšku v milimetrech (`body.H`). Zajímá nás efekt pohlaví na tělesnou výšku adjustovaný na délku chodidla.

- Prohlédněte si data – pokud obsahují nějaké chybějící hodnoty, odstraňte příslušné řádky. Vykreslete do obrázku krabicové diagramy popisující výšku studentů v závislosti na pohlaví.
- Modelujte závislost střední hodnoty tělesné výšky na pohlaví a délce chodidla. Vyzkoušejte různé varianty složitosti modelu: model se vzájemnou interakcí pohlaví a délky chodidla (1 – všeobecný, různé sklony přímk), model bez interakce (2 – ANCOVA, stejné sklony přímk) a model bez vlivu proměnné pohlaví (3 – jedna přímka). Vyberte ten nejvhodnější a své rozhodnutí zdůvodněte a podpořte příslušným výstupem z . Vypište také odhad vektoru β vašeho modelu.
- Vykreslete všechna pozorování jako bodový graf, kde na ose x bude délka chodidla a na ose y tělesná výška. Barevně rozlište muže i ženy a do obrázku umístěte i jednoduchou legendu. Do grafu postupně přidejte také přímky pro modely (1), (2) a (3), přičemž všechny modely mezi sebou barevně odlište.

- (d) *Dále vytvořte graf, který bude obsahovat všechna pozorování spolu s regresní přímkou odpovídajícími modelu (3). Vypočtěte z dat minimální a maximální délku chodidla a v rámci tohoto rozsahu zkonstruujte a nakreslete v  95% Scheffého pás spolehlivosti pro tuto regresní přímku. Dále odhadněte střední hodnotu tělesné výšky pro jedince s délkou chodidla 230, 250 a 270 mm pomocí 95% simultánních oboustranných intervalů spolehlivosti se Šidákovou adjustací. Vámi vypočtené intervaly zakreslete do grafu.*
- (e) *Pomocí 95% intervalu spolehlivosti predikujte výšku chlapce, jenž má délku chodidla 240 mm. Vypište dolní a horní hranici tohoto IS a do obrázku z (d) příslušný interval znázorněte.*