

---

## DOMÁCÍ ÚLOHY KE ZKOUŠCE Z M6120 LINEÁRNÍ STATISTICKÉ MODELY II

---

### Instrukce

Odevzdává se jeden pdf soubor nazvaný `prijmeni-jmeno-text-linmod-2017.pdf` (obsahuje řešení příkladů, obrázky,  $\mathbb{R}$ -kód napsaný v  $\text{\TeX}$ u), jeden zdrojový soubor naprogramovaných funkcí `prijmeni-jmeno-source-linmod-2017.r` a jeden soubor  $\mathbb{R}$ -kódu konkrétních řešení příkladů z DÚ `prijmeni-jmeno-priklady-linmod-2017.r`, který využívá zdrojový kód naprogramovaných funkcí. Dejte si záležet na přehlednosti programovaného kódu, na doplnění komentářů a vhodného užití zavedených pravidel, které máte k dispozici v prezentaci *Standards of programming in R: R style guide*. Také věnujte svou pozornost a čas dostatečným popisům vašich úvah, zvolených postupů a interpretacím výsledků, vždy také okomentujte rozdíly při použití různých přístupů. I to bude součástí celkového hodnocení. Na psaní  $\mathbb{R}$ -kódu doporučujeme  $\text{\TeX}$  balíček `listings` a vytvoření prostředí v hlavičce dokumentu pomocí následujícího kódu:

```
1 \lstset{language=R, % nastaveni jazyka R
2 basicstyle=\footnotesize\ttfamily, % typ pisma R-kodu
3 commentstyle=\ttfamily\color{barva1}, % barva komentare
4 numberstyle=\color{barva2}\footnotesize, % format cislovani
5 numbers=left, % cislovani vlevo
6 stepnumber=1, % cislovani s krokem jedna
7 frame=leftline, % vytvoreni leve hranicni cary
8 breaklines=true} % zalomeni radku
```

V textu potom kód vkládáme do prostředí `\begin{lstlisting}` a `\end{lstlisting}`.

Pokud jste neabsolvovali předmět M5751 Elektronická sazba a publikování v  $\text{\TeX}$ u nebo podobný, dostanete individuální instrukce na napsání DÚ v Overleaf: <https://www.overleaf.com/>. Odevzdání DÚ v MS Word není možné.

Úlohy se odevzdávají do Odevzdáárny 5 dnů před vybraným termínem zkoušky.

---

### Zadání

**Cvičení 1.** Datový soubor `more-samples-variances-clavicle.txt` obsahuje měření největší délky pravé klíční kosti v mm `cla.L` na lidských kostrách pocházejících ze čtyř populací: anglické (**eng**) indické z Amritsaru (**ind1**), indické z Varansi (**ind2**) a řecké (**gre**) z Atén. Budeme se zabývat studiem závislosti délky klíční kosti na populaci, ze které kostra pochází.

- (a) Načtete data do  $\mathbb{R}$ , prohlédněte si je a případně upravte do tvaru, ve kterém se vám s nimi bude lépe pracovat.  
(b) Jaký je rozsah souboru, který můžete použít pro analýzu závislosti délky klíční kosti na populaci, ze které kostra pochází?
- Udělejte si o závislosti délky klíční kosti na populaci první představu za pomoci vhodného grafu. Co na jeho základě soudíte?

3. Testujte hypotézu, že střední hodnota délky klíční kosti nezávisí na populaci, ze které kostra pochází.
- (a) Formulujte příslušnou nulovou a alternativní hypotézu.
  - (b) Zvolte vhodný test, formulujte jeho předpoklady a okomentujte, do jaké míry se zdají být splněny.
  - (c) Interpretujte výsledek testu. Formulujte statisticky korektní tvrzení.
4. Pomocí vhodné skupiny testů najděte dvojice populací, mezi kterými se délka klíční kosti prokazatelně liší. Testujte na celkové hladině 5 %.
- (a) Formulujte příslušnou skupinu dvojic nulových a alternativních hypotéz
  - (b) Zvolte vhodnou skupinu testů, formulujte jejich předpoklady a okomentujte, do jaké míry se zdají být splněny.
  - (c) Vysvětlete, jak jste zaručili, že celková hladina procedury nepřekročí 5 %, a zdůvodněte, proč jste si daný přístup vybrali.
  - (d) Interpretujte výsledek a formulujte statisticky korektní tvrzení.
  - (e) Souhlasí vaše závěry s tím, co jste očekávali před tím, než jste se podívali na data?
  - (f) Souhlasí vaše závěry s tím, co jste očekávali po tom, co jste vyřešili část 2.?
5. (a) Formulujte model analýzy rozptylu, za pomoci kterého můžete studovat rozdíly mezi délkou klíční kosti ve třech populacích: řecké a dvou indických.
- (b) Formulujte předpoklady modelu a okomentujte, do jaké míry se zdají být splněny.
- (c) Ilustrujte výstupy modelu za pomoci vhodného obrázku.
- V řešení této úlohy se neočekává použití dat o anglické populaci.*
6. Pro následující střední hodnoty
- (i) střední hodnota délky klíční kosti v dané populaci pro každou ze tří populací,
  - (ii) střední hodnota rozdílu mezi délkou klíční kosti ve dvou populacích pro každou dvojici populací,
  - (iii) střední hodnota rozdílu mezi řeckou populací a průměrem délek klíční kosti ve dvou indických populacích

řešte následující úlohy

- (a) vyjádřete je jako funkce parametrů modelu z části 5,
- (b) spočtěte jejich odhady,
- (c) spočtěte 95 % konfidenční interval pro každou z těchto středních hodnot,
- (d) spočtěte pro každou ze středních hodnot v části (ii) konfidenční interval tak, aby jejich společné pokrytí nebylo menší než 95 % (zvolte si vhodnou metodu a její výběr zdůvodněte).
- (e) spočtěte pro každou z těchto středních hodnot konfidenční interval tak, aby jejich společné pokrytí nebylo menší než 95 % (zvolte si vhodnou metodu a její výběr zdůvodněte).

*V řešení této úlohy se neočekává použití dat o anglické populaci.*

**Cvičení 2.** Datový soubor `17-anova-newborns.txt` obsahuje porodní hmotnost novorozenců v gramech `weight.C`, jejich pohlaví `sex.C` a vzdělání jejich matky `edu.M` (kódování: základní (1), střední bez maturity (2), střední s maturitou (3), vysokoškolské (4)). Budeme se zabývat studiem závislosti porodní hmotnosti na pohlaví dítěte a vzdělání matky.


- Načtěte data do  $\mathbb{R}$ , prohlédněte si je a případně upravte do tvaru, ve kterém se vám s nimi bude lépe pracovat.
  - Jaký je rozsah souboru, který můžete použít pro analýzu závislosti porodní hmotnosti na pohlaví dítěte a vzdělání matky?
- Udělejte si o závislosti porodní hmotnosti na pohlaví dítěte a vzdělání matky první představu za pomoci vhodných grafů. Co na jejich základě soudíte?
- Navrhněte vhodný model pro závislost porodní hmotnosti na pohlaví dítěte a vzdělání matky
  - Formulujte předpoklady modelu a okomentujte, do jaké míry se zdají být splněny.
  - Interpretujte koeficienty v modelu (nejste-li spokojeni s parametrizací, kterou volí  $\mathbb{R}$ , zvolte si parametrizaci, která vám vyhovuje lépe).
  - Zvolte vhodnou grafickou reprezentaci výsledků modelu.
  - Testujte významnost vlivu pohlaví dítěte a vzdělání matky. Formulujte příslušné dvojice nulové a alternativní hypotézy a vyslovte statisticky korektní závěry.
- Popište výsledky vaší analýzy intuitivně.
  - Shodují se s očekáváním, která jste měli před tím, než jste viděli data?
  - Shodují se s vašimi očekáváním po tom, co jste vyřešili část 2?
  - Okomentujte přednosti a nedostatky vaší analýzy.

**Cvičení 3.** Datový soubor `lrm-foot.txt` obsahuje tělesnou výšku ve stoje v milimetrech `body.H` a délku chodidla levé nohy v milimetrech `foot.L`. Budeme se zabývat odhadováním tělesné výšky při známé délce chodidla.

- Načtěte data do  $\mathbb{R}$ , prohlédněte si je a případně upravte do tvaru, ve kterém se vám s nimi bude lépe pracovat.
  - Jaký je rozsah souboru, který můžete použít pro analýzu vztahu mezi tělesnou výškou a délkou chodidla?
- Udělejte si o závislosti tělesné výšky na délce chodidla první představu za pomoci vhodného grafu. Co na jeho základě soudíte?
- Navrhněte vhodný model pro závislost tělesné výšky na délce chodidla.
  - Formulujte předpoklady modelu a okomentujte, do jaké míry se zdají být splněny.
  - Vybrali-li jste si model regresní přímky, interpretujte relevantní koeficient.
  - Zvolte vhodnou grafickou reprezentaci výsledků modelu.
- Zkonstruujte a zakreslete bodové konfidenční intervaly pro regresní přímku.

5. Zkonstruuje a zakreslete sdružené konfidenční pásy pro regresní přímku.
6. Okomentujte rozdíl mezi dvěma typy intervalů. Vyšly v tomto případě podle očekávání?

**Cvičení 4.** Datový soubor `ancova-cneck.txt` obsahuje tělesnou hmotnost v kilogramech `body.W`, pohlaví `sex` (kódování: muž (m), žena (f)) a obvod boků v milimetrech `hip.C`. Budeme se zabývat studiem závislosti tělesné hmotnosti na pohlaví a obvodu boků.

1. (a) Načtete data do , prohlédněte si je a případně upravte do tvaru, ve kterém se vám s nimi bude lépe pracovat.  
(b) Jaký je rozsah souboru, který můžete použít pro analýzu vztahu mezi tělesnou hmotností a obvodem boků podle pohlaví?
2. Udělejte si o závislosti tělesné hmotnosti na obvodu boků a pohlaví první představu za pomoci vhodných grafů. Co na jejich základě soudíte?
3. Navrhněte vhodný model pro závislost tělesné hmotnosti na obvodu boků a pohlaví.
  - (a) Formulujte předpoklady modelu a okomentujte, do jaké míry se zdají být splněny.
  - (b) Testujte hypotézu o tom, je-li funkční závislost střední hodnoty tělesné hmotnosti na obvodu boků stejná pro muže a pro ženy.
    - (i) Formulujte příslušnou nulovou a alternativní hypotézu.
    - (ii) Zvolte vhodný test, formulujte jeho předpoklady a okomentujte, do jaké míry se zdají být splněny.
    - (iii) Interpretujte výsledek testu. Formulujte statisticky korektní tvrzení.
  - (c) Zvolte vhodnou grafickou reprezentaci výsledků modelu.
  - (d) Interpretujte koeficienty v modelu.
  - (e) Okomentujte přednosti a slabiny modelu.
4. Zkonstruuje a zakreslete bodové konfidenční intervaly pro regresní přímky popisující závislost tělesné hmotnosti na obvodu boků podle pohlaví.
5. Zkonstruuje a zakreslete sdružené konfidenční pásy pro regresní přímky popisující závislost tělesné hmotnosti na obvodu boků podle pohlaví.