1. *(Cesta[1])* Cestu z kampusu na letiště je možné uskutečnit trasou A nebo trasou B. Dojezdové doby tras považujte za náhodné veličiny s normálním rozdělením pravděpodobnosti. Každá z tras byla dosud použita čtyřikrát, přičemž byly naměřeny následující dojezdové doby:

trasa A: 34,5 min, 35,0 min, 34,0 min a 34,5 min; trasa B: 33,0 min, 32,0 min, 19,0 min a 34,0 min.

Současný host se potřebuje dostat z kampusu na letiště do 35 minut. Doporučte mu vhodnou trasu pomocí 2 přístupů:

(a) Testujte hypotézu, že dojezdové doby obou tras mají stejné střední hodnoty, proti alternativě, že jedna z tras (která?) je v průměru rychlejší. Formulujte nulovou a alternativní hypotézu, zvolte správnou testovací statistiku a spočítejte její hodnotu. Pomocí kritického oboru (uveďte potřebné kvantily) na hladině významnosti 0,05 potom rozhodněte o výsledku testu a doporučte vhodnou trasu. K řešení úlohy můžete využít i vhodné funkce v *R*.

(b) Pro každou trasu na základě 4 pozorování odhadněte parametry normálního rozdělení pravděpodobnosti a spočítejte pravděpodobnost, že následující cesta na letiště bude trvat nejvýše 35 minut. Na základě těchto spočítaných pravděpodobností pak doporučte vhodnou trasu.

2. *(Diamanty[2])* V únoru roku 1992 byl v *Singapore Straits Times* uveřejněn inzerát na prodej 48 prstenů s diamanty. Datový soubor `diamanty.txt` (viz též `/Vyuka/R/M5120/data/diamanty.txt`) obsahuje hmotnosti diamantových kamenů těchto prstenů v miligramech a jejich cenu přepočtenou na CZK.

(a) Načtěte datový soubor a vykreslete histogram a boxplot cen prstenů.

(b) Zkoumejte závislost ceny prstenu, Y, na hmotnosti diamantu, x. Vyřešte 3 lineární regresní modely s následujícími regresními funkcemi (o jaké funkce se jedná?):

$$(1)\ y = \beta_0 + \beta_1 x; \qquad (2)\ y = \beta_0 + \beta_1 x + \beta_2 x^2; \qquad (3)\ y = \beta_1 x + \beta_2 x^2.$$

(c) Pro každý regresní model zapište (např. ve formě tabulky) hodnoty R^2, $\overline{R}^2$, S_e a určete významné koeficienty.

(d) Nakreslete grafy 3 odhadnutých regresních funkcí spolu s původními daty do jednoho obrázku. Rozsah osy pro hmotnost zvolte od 0 mg do 100 mg, nezapomeňte na popisky os a grafické rozlišení jednotlivých funkcí.

(e) Porovnejte výsledky. Který model byste vybrali jako nejlepší? Své rozhodnutí doprovoďte komentářem a zhodnoťte i reálnost použití jednotlivých regresních funkcí v praxi.

3. *(Transformace)* Uvažujme náhodný výběr $\boldsymbol{X} = (X_1, \dots, X_n)'$ rozsahu n z rozdělení pravděpodobnosti se střední hodnotou μ a rozptylem σ^2. Zavedeme transformované veličiny

$$U = \overline{X}; \qquad V = \sum_{k=1}^{n} \left(X_k - \overline{X}\right)^2.$$

(a) Vyjádřete veličiny U, V maticovým zápisem (lineární nebo kvadratické forma) pomocí náhodného výběru $\boldsymbol{X}$.

(b) Potom pomocí vhodných vzorců spočítejte rozptyl $\mathrm{D}U$ a střední hodnotu $\mathrm{E}V$.

4. *(Odhady)* Uvažujeme rozdělení pravděpodobnosti s hustotou pravděpodobnosti

$$f(x; h) = \frac{h}{2}\, \mathrm{e}^{-h|x|}, \qquad x \in \mathbb{R}$$

závisející na neznámém parametru $h > 0$ a náhodný výběr $\boldsymbol{X} = (X_1, \dots, X_n)'$ rozsahu n z tohoto rozdělení.

(a) Metodu momentů odvoďte odhad $\widehat{h}_\mathrm{M}$ parametru h.

(b) Metodou maximální věrohodnosti odvoďte odhad $\widehat{h}_\mathrm{ML}$ parametru h.

(c) Pro náhodný výběr $\boldsymbol{X} = (-0{,}34;\ -1{,}09;\ -0{,}10;\ 0{,}34;\ -0{,}13;\ -0{,}77;\ 1{,}09;\ 0{,}06;\ 0{,}58;\ -0{,}03)'$ číselně spočítejte oba odhady. Dále vykreslete grafy věrohodnostní a logaritmické věrohodnostní funkce a nalezněte bod lokálního maxima.

5. ***(Nepovinná úloha)*** Nalezněte maximálně věrohodný odhad neznámého parametru $\theta \in \mathbb{R}$ pro hustotu

$$f(x; \theta) = \frac{1}{2}\, \mathrm{e}^{-|x-\theta|}, \qquad x \in \mathbb{R}.$$

Klasický přístup přes nalezení stacionárního bodu (logaritmické) věrohodnostní funkce tu nebude úspěšný, pomoci může např. geometrická úvaha se zobrazením složek náhodného vývěru na reálné ose.

[1] podle Anděl, Jiří: *Matematika náhody.* MATFYZPRESS, Praha, 2007.

[2] podle Chu, Singfat: Diamond ring pricing using linear regression. *Journal of Statistics Education* **4** (3), 1996.