M5120 — podzim 2014 — **Zadání zápočtových úkolů** — řešení odevzdat na cvičeních, anebo do ISu do 31.12.2014

1. *(Cesta, inspirováno [1])* Cestu na letiště je možné uskutečnit dvěma trasami, A a B. Obě trasy byly dosud použity čtyřikrát, přičemž byly naměřeny následující dojezdové doby v minutách:

trasa A	34,5	35,0	34,0	34,5
trasa B	33,0	32,0	19,0	34,0

Časy považujte za náhodné veličiny s normálním rozdělením.

Současný host se potřebuje dostat na letiště do 35 minut. Pokuste se mu doporučit jednu z tras, a to na základě dvou různých přístupů:

(a) Proveďte test hypotézy, že trasy mají stejné střední dojezdové doby, proti alternativě, že jedna z trasa je v průměru rychlejší (která?). Uveďte nulovou a alternativní hypotézu. Zvolte správnou (!) testovací statistiku, určete její hodnotu a počet stupňů volnosti. Pomocí kritického oboru na hladině významnosti $\alpha = 5$ % potom učiňte závěr.

(b) Pro každou trasu odhadněte pravděpodobnost, že následující cesta na letiště trasou A, resp. B, bude trvat nejvýše 35 minut. Kterou trasu byste hostu doporučili tímto přístupem? Není výsledek v rozporu s pozorováními v tabulce?

2. *(Odhady)* Uvažujte rozdělení pravděpodobnosti s hustotou závisející na parametru $h > 0$ a náhodný výběr $(X_1, \dots, X_n)'$ rozsahu n z tohoto rozdeělení:

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

(a) odvoďte maximálně věrohodný odhad $\hat{h}_{\mathrm{ML}}$ parametru h

(b) odvoďte momentový odhad $\hat{h}_{\mathrm{M}}$ parametru h

(c) nepovinná část: Uvažujte jiný tvar hustoty závisející na parametru $\theta \in \mathbb{R}$:

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

Nalezněte maximálně věrohodný odhad parametru θ. *(Klasický přístup přes nalezení stacionárního bodu (logaritmické) věrohodnostní funkce tu nebude úspěšný; proč? Pro nalezení maxima, a tedy odvození hledaného ML-odhadu, může pomoci např. geometrická úvaha.)*

3. *(Transformace)* Náhodné veličiny v následujících příkladech zapište pomocí transformace náhodného výběru $\boldsymbol{X} = (X_1, \dots, X_n)'$ rozsahu n z rozdělení pravděpodobnosti se střední hodnotou μ a rozptylem σ^2. Pomocí vhodného vzorce poté spočítejte požadovanou číselnou charakteristiku. V řešení uveďte použité transformační matice, vzorce a výsledky.

(a) spočítejte rozptyl výběrového průměru, tzn. $\mathrm{D}(\overline{X})$

(b) spočítejte střední hodnotu $\mathrm{E}(Y)$, kde $Y = \sum_{k=1}^{n} \left(X_k - \overline{X}\right)^2$

(c) spočítejte střední hodnotu $\mathrm{E}(Z)$, kde $Z = \sum_{k=1}^{n-1} (X_k - X_{k+1})^2$

4. *(Diamanty, dle [2])* 29.02.1992 byl v *Singapore Straits Times* uveřejněn inzerát na prodej 48 prstenů s diamanty.

- Datový soubor si zkopírujte z `/Vyuka/R/M5120/data/diamanty.txt` nebo z ISu. V prvním sloupci je uvedena hmotnost diamantového kamene v miligramech, druhý sloupec souboru uvádí cenu přepočtenou na CZK.
- Nakreslete histogram hodnot cen prstenů z datového souboru.
- Zkoumejte závislost ceny prstenu, Y, na hmotnosti diamantu, x, pomocí lineárního regresního modelu s níže uvedenými čtyřmi funkcemi pro hmotnosti diamantů od 0 mg do 100 mg.

(a) Proložte data regresní přímkou. Obchodníkovi s klenoty by se výsledek jistě nelíbil, proč?

(b) Proložte data funkcí $y = \beta_1 x + \beta_2 x^2$. Co je to za funkci?

(c) Proložte exponenciální funkci $y = a\,\mathrm{e}^{b\,x + c\,x^2}$.

(d) Proložte mocninnou funkci $y = a\,x^d$.

- Nakreslete grafy všech čtyř odhadnutých regresních funkcí spolu s daty do jednoho obrázku, pro hmotnosti diamantů od 0 mg do 100 mg (příp. pro širší interval), nezapomeňte na popisky.
- Pro každou regresní funkci (a)–(d) uveďte (např. ve formě tabulky) tyto výsledky: funkční zápis odhadnuté regresní funkce, hodnoty $S_\mathrm{e}, s^2, R^2, \bar{R}^2$, a typ rozdělení pravděpodobnosti veličiny Y.
- Porovnejte výsledky. Který model byste vybrali jako nejlepší? Své rozhodnutí doprovoďte komentářem.

Použité zdroje

[1] Anděl, Jiří (2007). *Matematika náhody.* MATFYZPRESS, Praha.

[2] Chu, Singfat (1996). Diamond Ring Pricing Using Linear Regression. *Journal of Statistics Education* **4** (3).