Cvičení z NSTP022
11. týden cvičení (29. 4. – 3. 5. 2013)

Intervalový odhad

1. Nechť $X_1, \ldots, X_n$ značí hodnoty IQ náhodně vybraných žáků osmé třídy. Předpokládejme, že $X_1, \ldots, X_n$ tvoří náhodný výběr z normálního rozdělení s neznámou střední hodnotou a rozptylem 9.
   (a) Odhadněte bodově střední hodnotu IQ žáků osmé třídy. Jaké jsou vlastnosti a rozdělení tohoto odhadu?
   (b) Sestrojte intervalový odhad o spolehlivosti $1 - \alpha$ pro střední hodnotu IQ žáků osmé třídy. Interpretujte.
   (c) Jak se bude tento interval měnit, budeme-li zvyšovat spolehlivost $1 - \alpha$?
   Jak se bude tento interval měnit, budeme-li zvyšovat počet dětí zahrnutých do experimentu (a ostatní hodnoty ve vzorcích zůstanou stejné)?
   (d) Po provedení měření jsme obdrželi následující hodnoty:

   $$111,\ 116,\ 105,\ 111,\ 110,\ 114,\ 108,\ 106,\ 112,\ 108,\ 112,\ 111,\ 105,\ 111,\ 108,\ 110.$$

   Jak byste na základě těchto konkrétních hodnot odhadli střední hodnotu IQ žáků osmé třídy? Uveďte bodový i 95%-ní intervalový odhad.
   (e) Napište 95%-ní dolní intervalový odhad pro střední hodnotu IQ žáků osmé třídy a interpretujte jej.

2. Provádíme průzkum, zda v nejmenované hospodě okrádají své hosty. Zakoupíme proto 10 piv a změříme jejich objem. Obdrželi jsme následující hodnoty (v litrech):

   $$0{,}510,\ 0{,}462,\ 0{,}491,\ 0{,}466,\ 0{,}461,\ 0{,}503,\ 0{,}495,\ 0{,}488,\ 0{,}512,\ 0{,}505.$$

   Předpokládejme, že datům odpovídají nezávislé náhodné veličiny s normálním rozdělením.
   (a) Odhadněte střední hodnotu a rozptyl objemu jednoho natočeného piva. Připomeňte, jaké jsou teoretické vlastnosti a rozdělení těchto odhadů.
   (b) Zkonstruujte intervalový odhad pro střední hodnotu objemu jednoho piva o spolehlivosti 95 %. Leží předepsaná hodnota 0,5 l v tomto intervalu?
   (c) Napište dále oba 95%-ní jednostranné (dolní a horní) intervaly spolehlivosti pro střední hodnotu objemu jednoho piva a interpretujte je. Který z nich Vám připadá „zajímavější“ pro tuto situaci?
   (d) Zkonstruujte 90%-ní intervalový odhad pro rozptyl objemu jednoho piva.

3. *Centrum pro výzkum veřejného mínění* v dubnu 2013 zveřejnilo výsledky aktuálního průzkumu mezi občany ČR staršími 15 let. Můžeme se dočíst, že z 1 059 respondentů jich 297 uvedlo, že „do korupčního jednání se zapojují téměř všichni veřejní činitelé“.
   (a) Odhadněte bodově podíl občanů ČR, kteří si myslí, že do korupčního jednání se zapojují téměř všichni veřejní činitelé. Jaké je rozdělení a asymptotické rozdělení tohoto bodového odhadu?
   (b) Určete asymptotický 95%-ní interval spolehlivosti pro podíl občanů, kteří si myslí, že do korupčního jednání se zapojují téměř všichni veřejní činitelé.
   (c) Jaký by musel být rozsah výběru n, aby intervalový odhad s asymptotickou spolehlivostí 95% pro tento podíl měl šířku nejvýše 0,03 (uvažujeme-li, že podíl občanů zahrnutých ve studii, kteří vyjadřují uvedený názor, se nezmění).

4. Chceme porovnat průměrnou výšku jedenáctiletých chlapců a dívek. Studie se zúčastnilo 25 chlapců (veličiny $X_1, \ldots, X_{25}$) a 20 dívek (veličiny $Y_1, \ldots, Y_{20}$). Obdrželi jsme následující výsledky:

   $$\overline{X} = 147{,}1, \quad \overline{Y} = 147{,}8, \quad S_X^2 = 10{,}9, \quad S_Y^2 = 14{,}7.$$

   Lze předpokládat, že výška chlapců i dívek má normální rozdělení se stejným rozptylem σ^2. Sestrojte intervalový odhad pro rozdíl průměrné výšky o spolehlivosti 95 %.

5. Zkonstruujte intervalový odhad parametru λ o asymptotické spolehlivosti 95 % pro data z příkladu 2 z minulého cvičení.

Opakování z přednášky

Model: $X_1, \dots, X_n$ je náhodný výběr (nezávislé stejně rozdělené veličiny) z rozdělení F_θ, které závisí na neznámém parametru $\theta \in \Theta$.

Minulé cvičení: bodový odhad parametrické funkce $g(\theta)$ $\leftrightsquigarrow$ náhodná veličina
+ jeho „pěkné“ vlastnosti $\leftrightsquigarrow$ nestrannost, konzistence

Intervalový odhad parametrické funkce $g(\theta)$ $\leftrightsquigarrow$ interval s náhodnými mezemi, který pokrývá skutečnou (neznámou) hodnotu $g(\theta)$ s předepsanou pravděpodobností $1-\alpha$

- Intervalový odhad parametrické funkce $g(\theta)$ o spolehlivosti $1-\alpha$ je dvojice náhodných veličin (L_n, U_n) takových, že

$$\mathsf{P}_\theta\big(L_n < g(\theta) < U_n\big) = 1-\alpha \quad \text{pro všechna } \theta \in \Theta. \tag{11.1}$$

- Veličina $D_n = D_n(X_1, \dots, X_n)$ se nazývá dolní intervalový odhad $g(\theta)$ o spolehlivosti $1-\alpha$, jestliže

$$\mathsf{P}_\theta\big(D_n < g(\theta)\big) = 1-\alpha \quad \text{pro všechna } \theta \in \Theta. \tag{11.2}$$

- Veličina $H_n = H_n(X_1, \dots, X_n)$ se nazývá horní intervalový odhad $g(\theta)$ o spolehlivosti $1-\alpha$, jestliže

$$\mathsf{P}_\theta\big(g(\theta) < H_n\big) = 1-\alpha \quad \text{pro všechna } \theta \in \Theta. \tag{11.3}$$

Podobně jako jsme měli u bodového odhadu, $L_n = L_n(X_1, \dots, X_n)$, $U_n = U_n(X_1, \dots, X_n)$, $D_n = D_n(X_1, \dots, X_n)$ a $H_n = H_n(X_1, \dots, X_n)$ jsou borelovské funkce náhodného výběru $X_1, \dots, X_n$, jejichž funkční předpisy nezávisí na θ.

Číslo $\alpha \in (0,1)$ volíme, v praxi se většinou pracuje s $\alpha = 0{,}05$.

Obecná konstrukce intervalu spolehlivosti pro parametrickou funkci $g(\theta)$**.** Najdeme funkci náhodného výběru a parametrické funkce $g(\theta)$, tj. náhodnou veličinu $h\big(X_1, \dots, X_n; g(\theta)\big)$, jejíž rozdělení nezávisí na θ. Nechť $h_{\alpha/2}$ a $h_{1-\alpha/2}$ jsou kvantily tohoto rozdělení. Pak jistě platí

$$\mathsf{P}_\theta\Big(h_{\alpha/2} < h\big(X_1, \dots, X_n; g(\theta)\big) < h_{1-\alpha/2}\Big) = 1-\alpha \quad \text{pro všechna } \theta \in \Theta. \tag{11.4}$$

Je-li možné nerovnosti v závorce převést ekvivalentními úpravami na takový tvar, že uprostřed stojí $g(\theta)$ a vlevo i vpravo je něco, co na θ nezávisí, sestrojili jsme intervalový odhad.

Speciální případy.

- Intervalové odhady pro parametry **normálního rozdělení** jsou shrnuty v tabulkách 5.1 a 5.2 ve skriptech.
- **Asymptotické intervalové odhady** (intervalové odhady s asymptotickou spolehlivostí $1-\alpha$) je možné zkonstruovat pomocí CLV - tabulka 5.3 ve skriptech.
  Intervalové odhady pro speciální případ **alternativního rozdělení** jsou rozepsány v tabulce 5.4.

Výsledky

1.(a) Odhadem střední hodnoty IQ je $\hat{\mu}_n = \overline{X}_n$. Tento odhad je nestranný a konzistentní. Jeho rozdělení je $\overline{X}_n \sim \mathsf{N}(\mu, 9/n)$.

(b) intervalový odhad pro μ se spolehlivostí $1-\alpha$ je $\left(\overline{X}_n - 3\,u_{1-\alpha/2}/\sqrt{n}, \overline{X}_n + 3\,u_{1-\alpha/2}/\sqrt{n}\right)$,

(c) pro $1-\alpha \nearrow$ se bude interval rozšiřovat a při $n \nearrow$ se bude interval zužovat,

(d) bodový odhad je 109,875, intervalový odhad se spolehlivostí 95 % je (108,405; 111,345),

(e) dolní 95%-ní intervalový odhad $(108{,}645; +\infty)$.

2.(a) $\hat{\mu}_n = \overline{X}_n = 0{,}489$, $\widehat{\sigma^2_n} = S^2_n = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X}_n)^2 = 0{,}000389$,
obecně: $\overline{X}_n$ je nestranný a konzistentní odhad μ a S^2_n je nestranný a konzistentní odhad σ^2, $\overline{X}_n \sim \mathsf{N}(\mu, \sigma^2/n)$ a $(n-1)S^2_n/\sigma^2 \sim \chi^2_{n-1}$,

(b) (0,475; 0,503),

(c) $(0{,}479; \infty)$ a $(-\infty; 0{,}501)$,

(d) 90 %-ní interval spolehlivosti pro σ^2 je $(2{,}1 \cdot 10^{-4}; 1{,}05 \cdot 10^{-3})$.

3.(a) Předpokládaný model: náhodná veličina X_i je indikátor toho, zda si i-tá osoba myslí, že do korupčního jednání se zapojují téměř všichni veřejní činitelé;
předpokládáme, že $X_1, \ldots, X_{1059}$ jsou nezávislé a stejně rozdělené s alternativním (tj. 0-1) rozdělením s neznámým parametrem $p \in (0,1)$.
Bodový odhad je $\hat{p} = 0{,}28$.

(b) (0,253; 0,308),

(c) $n \geq 3446$.

4. 95 %-ní interval spolehlivosti pro $\mu_X - \mu_Y$ je $(-2.9, 1.5)$.

5. (1,55; 2,58).