Intervalové odhady
10.5. a 14.5. 2018

1. Nechť $X_1, \ldots, X_n$ značí hodnoty IQ náhodně vybraných žáků 8. třídy a 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 8.třídy. Jaké jsou vlastnosti a rozdělení tohoto odhadu?
   (b) Sestrojte intervalový odhad o spolehlivosti $1-\alpha$ pro střední hodnotu IQ žáků 8.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 IQ žáků 8.třídy? Uveďte bodový i 95%-ní intervalový odhad.
   (e) Napište 95%-ní dolní intervalový odhad pro střední IQ žáků 8.třídy a interpretujte.
   (f) Jak by se naše postupy a výsledky změnily, pokud bychom neznali rozptyl IQ?

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 si, 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 litru v tomto intervalu?
   (c) Napište dále oba 95%-ní jednostranné (dolní a horní) intervalové odhady 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. Počty vstřelených gólů v jednotlivých fotbalových zápasech jsou nezávislé náhodné veličiny s Poissonovým rozdělením s neznámým parametrem $\lambda > 0$. Pro data z roku 2008 máme k dispozici 306 zápasů, kde bylo vstřeleno v průměru 2.92 gólů a směrodatná odchylka vyšla 1.64.
   (a) Zkonstruujte intervalový odhad se spolehlivostí 95% pro střední počet vstřelených gólů v jednom zápase.

(b) Zkonstruujte 95%-ní intervalový odhad pro pravděpodobnost, že v zápase nepadne žádný gól.

4. Uvažujme fiktivní anketu o volebních preferencích před volbami prezidenta, ve které ze 100 náhodně dotázaných respondentů (kteří plánují jít k volbám) 35 odpovědělo, že bude v blížících se volbách volit kandidáta X.

   (a) Zkonstruujte intervalový odhad se spolehlivostí 95 % pro pravděpodobnost p, s jakou občan ČR volí kandidáta X. Je pravděpodobné, že bude kandidát X zvolen v prvním kole?
   (b) Jak se změní intervalový odhad, pokud se zeptáme 1000 osob a 350 z nich budou voliči kandidáta X?
   (c) Jestliže podíl voličů zůstane stejný jako v (a) a (b), kolik osob se musíme zeptat, aby šířka intervalu byla maximálně 2%?

5. Předpokládejme, že délky obsluhy jednotlivých zákazníků na poště jsou nezávislé náhodné veličiny z exponenciálním rozdělením s neznámým parametrem $\lambda > 0$, tj. s hustotou $f(x) = \lambda \mathrm{e}^{-\lambda x} I[x \geq 0]$. Délka obsluhy 16 po sobě jdoucích zákazníků trvala 25 minut.

   (a) Sestrojte 95%-ní intervalový odhad pro střední dobu obsluhy jednoho zákazníka na poště.
   (b) Sestrojte 95%-ní intervalový odhad pro λ.

6. Chceme porovnat průměrnou výšku 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:
$$\overline{X} = 180.6, \quad \overline{Y} = 164.9, \quad S_X^2 = 6.5, \quad S_Y^2 = 9.3.$$
   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%.

Tabulka kvantilů $\mathsf{N}(0,1)$

α	0.9	0.95	0.975	0.99	0.995
u_α	1.282	1.645	1.960	2.326	2.576

Tabulka vybraných kvantilů t_k rozdělení

$k \setminus \alpha$	0.9	0.95	0.975	0.99	0.995
2	1.886	2.920	4.303	6.965	9.925
5	1.476	2.015	2.571	3.365	4.032
9	1.383	1.833	2.262	2.821	3.250
15	1.341	1.753	2.131	2.602	2.947
20	1.325	1.725	2.086	2.528	2.845
50	1.299	1.676	2.009	2.403	2.678
100	1.290	1.660	1.984	2.364	2.626

Model: $X_1, \dots, X_n$ je náhodný výběr z rozdělení F_θ, které závisí na neznámém parametru $\theta \in \Theta$.

INTERVALOVÝ ODHAD parametrické funkce $g(\theta)$ o spolehlivosti $1-\alpha$ je interval s **náhodnými mezemi**, který **pokrývá** skutečnou (neznámou) hodnotu $g(\theta)$ s předepsanou pravděpodobností $1-\alpha$, kde $\alpha \in (0,1)$. V praxi většinou volíme $\alpha = 0.05$.

– Oboustranný intervalový odhad je dvojice náhodných veličin (L_n, U_n), kde $L_n = L_n(X_1, \dots, X_n)$ a $U_n = U_n(X_1, \dots, X_n)$ tak, že

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

– Dolní intervalový odhad je tvaru (D_n, ∞), kde náhodná veličina $D_n = D_n(X_1, \dots, X_n)$ splňuje

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

– Horní intervalový odhad je tvaru $(-\infty, H_n)$, kde veličina $H_n = H_n(X_1, \dots, X_n)$ splňuje

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

Poznámka: Všechny uvažované funkce L_n, U_n, D_n a H_n jsou borelovské funkce náhodného výběru $X_1, \dots, X_n$, jejichž funkční předpisy nezávisí na θ.

ASYMPTOTICKÝ INTERVALOVÝ ODHAD. Někdy je obtížné nalézt přesný intervalový odhad a spokojíme se s **asymptotickým intervalovým odhadem**, který pokrývá $g(\theta)$ s pravděpodobností konvergující k $1-\alpha$ pro $n \to \infty$. Např. pro oboustranný odhad pak požadujeme

$$\mathsf{P}_\theta(L_n < g(\theta) < U_n) \to 1 - \alpha, \quad n \to \infty \quad \text{pro všechna } \theta \in \Theta.$$

Mluvíme pak o intervalových odhadech s asymptotickou spolehlivostí $1 - \alpha$.

OBECNÁ KONSTRUKCE INTERVALOVÉHO ODHADU PRO PARAMETRICKOU FUNKCI $g(\theta)$. Najdeme funkci náhodného výběru a funkce $g(\theta)$, tj. náhodnou veličinu $h(X_1, \dots, X_n; g(\theta))$, jejíž rozdělení nezávisí na θ. Nechť $h_{\alpha/2}$ a $h_{1-\alpha/2}$ jsou kvantily tohoto rozdělení. Pak jistě platí

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

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**.
– **Asymptotické intervalové odhady** konstruované pomocí CLV.

Intervalové odhady v normálním rozdělení. Nechť $X_1, \dots, X_n$ je náhodný výběr z $\mathsf{N}(\mu, \sigma^2)$, kde $\mu \in \mathbb{R}$ je neznámé.

1. Nechť $\sigma^2 > 0$ je známé. Pak ke konstrukci intervalového odhadu pro μ využijeme, že $\overline{X}_n$ má $\mathsf{N}(\mu, \sigma^2/n)$ rozdělení. Tudíž $\sqrt{n}\frac{\overline{X}_n - \mu}{\sigma}$ má $\mathsf{N}(0,1)$ a platí
$$\mathsf{P}\left(-u_{1-\alpha/2} < \sqrt{n}\frac{\overline{X}_n - \mu}{\sigma} < u_{1-\alpha/2}\right) = 1 - \alpha,$$
kde $u_{1-\alpha/2} = \Phi^{-1}(1-\alpha/2)$ je kvantil $\mathsf{N}(0,1)$ na hladině $1-\alpha/2$. Odtud pak již snadno ekvivalentními úpravami získáme intervalový odhad μ.

2. Nechť $\sigma^2 > 0$ je neznámé. Připomeňme, že $S_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X}_n)^2$.
   - Pro odhad μ využijeme, že $T = \sqrt{n}\frac{\overline{X}_n - \mu}{S_n}$ má t-rozdělení s $n-1$ stupni volnosti. Kvantily tohoto rozdělení jsou tabelovány; pro n dostatečně velké je lze aproximovat kvantily $\mathsf{N}(0,1)$.
   - Pro odhad σ^2 využijeme, že $(n-1)S_n^2/\sigma^2$ má χ^2 rozdělení s $n-1$ stupni volnosti a tedy platí
   $$\mathsf{P}(\chi^2_{\alpha/2,n-1} < (n-1)\frac{S_n^2}{\sigma^2} < \chi^2_{1-\alpha/2,n-1}) = 1 - \alpha,$$
   kde $\chi^2_{\beta,k}$ je β-kvantil χ^2 rozdělení s k stupni volnosti.

Asymptotický intervalový odhad pro střední hodnotu. Nechť $X_1, \dots, X_n$ je náhodný výběr z rozdělení se střední hodnotou $\mu = g(\theta)$ a rozptylem $\mathsf{Var}\, X_1 \in (0, \infty)$. Pak
$$\sqrt{n} \cdot \frac{\overline{X}_n - \mu}{\sqrt{\mathsf{Var}\, X_1}} \xrightarrow[n\to\infty]{D} \mathsf{N}(0,1).$$

Pokud $\mathsf{Var}\, X_1$ známe, pak pro konstrukci intervalového odhadu μ postupujeme stejně jako v 1. u normálního rozdělení. V praxi ale většinou $\mathsf{Var}\, X_1$ neznáme. Ze Sluckého věty (věta 4.14 ve skriptech) plyne, že
$$\sqrt{n} \cdot \frac{\overline{X}_n - \mu}{\sqrt{V_n}} \xrightarrow[n\to\infty]{D} \mathsf{N}(0,1),$$
kdykoliv je V_n konzistentní odhad $\mathsf{Var}\, X_1$. Za tento konzistetní odhad V_n můžeme brát například:

- Výběrový rozptyl S_n^2 (univerzální postup).
- Pokud $\mathsf{Var}\, X_1 = v(\mu)$ je funkcí μ (např. Alt, Po, Exp), pak můžeme využít toho, že $\overline{X}_n$ je konzistentním odhadem μ, a vzít $V_n = v(\overline{X}_n)$ jako konzistentní odhad $\mathsf{Var}\, X_1$.
  Např. v $\mathsf{Alt}(p)$ je $\mathsf{Var}\, X_1 = p(1-p) = v(p)$ a tedy $V_n = v(\overline{X}_n) = \overline{X}_n(1-\overline{X}_n)$ je konzistentní odhad $\mathsf{Var}\, X_1$.