13. Intervalové odhady parametrů

Jednou ze základních úloh statistiky je stanovení hodnot parametrů rozdělení, ze kterého máme k dispozici náhodný výběr. Nejčastěji hledáme odhady dvou druhů:

-**bodový odhad** je odhad parametru pomocí statistiky (funkce náhodného výběru), jejíž hodnotu pro datový soubor považujeme za hledanou hodnotu neznámeho parametru rozdělení;

-**intervalový odhad** je stanovení intervalu, ve kterém se hodnota neznámého parametru vyskytuje s požadovanou pravděpodobností blízkou jedné.

13.1. Intervalový odhad. Jestliže je θ neznámý parametr zkoumaného rozdělení, pak hledáme statistiky T_d a T_h takové, že pro **koeficient spolehlivosti** $(1-\alpha)$ platí:

$P(T_d \leq \theta \leq T_h) = 1-\alpha$, (**oboustranný odhad**) přičemž obvykle ještě požadujeme $P(\theta < T_d) = P(\theta > T_h) = \frac{\alpha}{2}$.

Intervalovým odhadem (oboustranným) parametru θ je interval (T_d, T_h).

Někdy hledáme pouze **jednostranné odhady**. Je pak:

$\theta \in (T_d, \infty)$, kde $P(\theta \geq T_d) = 1-\alpha$ a $P(\theta < T_d) = \alpha$;

$\theta \in (-\infty, T_h)$, kde $P(\theta \leq T_h) = 1-\alpha$ a $P(\theta > T_h = \alpha$.

Obvykle volíme $\alpha = 0,05$. Spolehlivost odhadu je pak $(1-\alpha) = 0,95$. To znamená, že v 95%, případů leží hodnota parametru v uvedeném intervalu spolehlivosti. Vyjímečně volíme $\alpha = 0,01$, nebo $0,1$.

Intervalové odhady parametrů některých rozdělení.

13.2. Normální rozdělení.

A) Odhadujeme parametr μ v rozdělení $N(\mu, \sigma^2)$ při známem rozptylu σ^2. Zde použijeme statistiku $\overline{X}$ (výběrový průměr). Víme, že náhodná veličina

$U = \dfrac{\overline{X}-\mu}{\sigma}\sqrt{n}$ má normované normální rozdělení $N(0,1)$. Potom je

$$P(|U| \leq u_{1-\frac{\alpha}{2}}) = 1-\alpha \Leftrightarrow -u_{1-\frac{\alpha}{2}} \leq \frac{\overline{X}-\mu}{\sigma}\sqrt{n} \leq u_{1-\frac{\alpha}{2}},$$

kde symbolem u_p, $0 < p < 1$ označujeme $p-$kvantil normovaného normálního rozdělení $N(0,1)$. Odtud dostaneme, že

$$T_d = \overline{X} - \frac{\sigma}{\sqrt{n}}u_{1-\frac{\alpha}{2}} \leq \mu \leq T_h = \overline{X} + \frac{\sigma}{\sqrt{n}}u_{1-\frac{\alpha}{2}}.$$

Jednostrannými odhady jsou

$$\mu \leq T_h = \overline{X} + \frac{\sigma}{\sqrt{n}}u_{1-\alpha}, \qquad \text{resp.} \qquad \mu \geq T_d = \overline{X} - \frac{\sigma}{\sqrt{n}}u_{1-\alpha}.$$

B) Odhadujeme funkci σ^2 při známé střední hodnotě μ. Zde použijeme skutečnosti, že má náhodná veličina $U_i = \frac{X_i-\mu}{\sigma}$ normované normální rozdělení $N(0,1)$. Potom má náhodná veličina $\sum\limits_{i=1}^{n}\left(\dfrac{X_i-\mu}{\sigma}\right)^2$ rozdělení $\chi^2(n)$. Je pak

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2 = \frac{\sigma^2}{n}\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2.$$

Má tudíž statistika $V = \frac{ns^2}{\sigma^2}$ rozdělení $\chi^2(n)$. Pro oboustranný odhad dostaneme

$$P(v_1 \leq V \leq v_2) = 1 - \alpha \Rightarrow v_1 = \chi^2_{\frac{\alpha}{2}}(n) \quad \text{a} \quad v_2 = \chi^2_{1-\frac{\alpha}{2}}(n),$$

kde symbolem $\chi^2_p(n)$ označujeme $p-$ kvantil rozdělení $\chi^2(n)$. Odtud plyne odhad

$$\frac{ns^2}{\chi^2_{1-\frac{\alpha}{2}}(n)} \leq \sigma^2 \leq \frac{ns^2}{\chi^2_{\frac{\alpha}{2}}(n)}.$$

Obdobně dostaneme jednostranné odhady

$$\sigma^2 \leq \frac{ns^2}{\chi^2_{\alpha}(n)}, \quad \text{resp.} \quad \sigma^2 \geq \frac{ns^2}{\chi^2_{1-\alpha}(n)}.$$

C) Odhadujeme střední hodnotu μ za podmínky, že rozptyl uvažovaného rozdělení není znám. Ke stanovení intervalu spolehlivosti použijeme statistiku $T = \frac{\overline{X} - \mu}{S}\sqrt{n}$, o které víme, že má Studentovo $t-$ rozdělení $t(n-1)$ o $(n-1)$ stupních volnosti. Interval spolehlivosti určíme z podmínky

$$P\left(|T| \leq t_{1-\frac{\alpha}{2}}(n-1)\right) = 1 - \alpha.$$

Odtud je

$$-t_{1-\frac{\alpha}{2}} \leq \frac{\overline{X} - \mu}{S}\sqrt{n} \leq t_{1-\frac{\alpha}{2}},$$

tudíž

$$\overline{X} - \frac{S}{\sqrt{n}} t_{1-\frac{\alpha}{2}} \leq \mu \leq \overline{X} + \frac{S}{\sqrt{n}} t_{1-\frac{\alpha}{2}}$$

je oboustraný interval spolehlivosti pro parametr μ.

Obdobně dostaneme jednostrané intervaly ve tvaru:

$$\mu \leq \overline{X} + \frac{S}{\sqrt{n}} t_{1-\alpha}, \quad \mu \geq \overline{X} - \frac{S}{\sqrt{n}} t_{1-\alpha},$$

kde symbolem t_α označujeme α kvantil uvažovaného rozdělení.

D) Odhadujeme parametr σ^2 při neznámé střední hodnotě μ. Zde použijeme statistiku $Y = \frac{n-1}{\sigma^2} S^2$, která má rozdělení $\chi^2(n-1)$. Vycházíme ze skutečnosti, že pro statistiku S^2 je $E(S^2) = \sigma^2$ a může tedy sloužit jako vhodný odhad parametru σ^2. Oboustraný interval spolehlivosti dostaneme z podmínky

$$P(v_1 \leq Y \leq v_2) = 1 - \alpha \Rightarrow v_1 = \chi^2_{\frac{\alpha}{2}}(n-1), \; v_2 = \chi^2_{1-\frac{\alpha}{2}}(n-1)$$

jsou odpovídající kvantily rozdělení χ^2. Odtud plyne pro oboustraný interval spolehlivosti

$$v_1 \leq \frac{(n-1)S^2}{\sigma^2} \leq v_2 \Rightarrow \frac{(n-1)}{v_2} S^2 \leq \sigma^2 \leq \frac{(n-1)}{v_1} S^2.$$

Jednoduchou úpravou získáme jednostrané intervaly spolehlivosti ve tvaru

$$\sigma^2 \leq \frac{(n-1)}{v_1} S^2, \quad \frac{(n-1)}{v_2} S^2 \leq \sigma^2,$$

kde v_1 a v_2 jsou zde po řadě kvantily $\chi^2_\alpha(n-1)$, $v_2 = \chi^2_{1-\alpha}(n-1)$ rozdělení *chí*-kvadrát o $(n-1)$ stupních volnosti.

13.3. Exponenciální rozdělení.

Uvedeme interval spolehlivosti pro rozdělení $Ex(0;\delta)$, kde využijeme skutečnosti, že je střední hodnota $E(\overline{X}) = \delta$. Statistika $T = \frac{2n\overline{X}}{\delta}$ má totiž rozdělení $\chi^2(2n)$. Interval spolehlivosti získáme z identity

$$P(v_1 \leq T \leq v_2) = 1 - \alpha \Rightarrow v_1 \leq \frac{2n\overline{X}}{\delta} \leq v_2 \Rightarrow$$

$$\frac{2n\overline{X}}{v_2} \leq \delta \leq \frac{2n\overline{X}}{v_1},$$

kde $v_1 = \chi^2_{\frac{\alpha}{2}}(2n)$ a $v_2 = \chi^2_{1-\frac{\alpha}{2}}(2n)$ kvantil rozdělení *chí*-kvadrát.

Obdobně dostaneme jednostrané intervaly spolehlivosti ve tvaru

$$\frac{2n\overline{X}}{v_2} \leq \delta, \quad \delta \leq \frac{2n\overline{X}}{v_1},$$

kde $v_1 = \chi^2_\alpha(2n)$ a $v_2 = \chi^2_{1-\alpha}(2n)$ kvantil rozdělení *chí*-kvadrát.

13.5. Alternativní rozdělení.

Odhadujeme hodnotu parametru p, kde využíváme skutečnosti, že pro náhodný výběr z alternativního rozdělení má výběrový úhrn $\tilde{X} = \sum_{i=1}^{n} X_i$ binomické rozdělení $Bi(n,p)$. Podle centrální limitní věty lze pro dostatečně rozsáhlý výběr předpokládat, že součet má normální rozdělení. Protože je $E(\tilde{X}) = np$ a $D(\tilde{X}) = np(1-p)$, má pro $np(1-p) > 9$ výběrový úhrn $\tilde{X}$ normální rozdělení $N(np, np(1-p))$. Má potom náhodná veličina

$$Z = \frac{\tilde{X} - np}{\sqrt{np(1-p)}} = \frac{\overline{X} - p}{\sqrt{\frac{p(1-p)}{n}}} \sim N(0;1)$$

normované normální rozdělení.

Potom je

$$P(|Z| \leq u_{1-\frac{\alpha}{2}}) = 1 - \alpha \Leftrightarrow -u_{1-\frac{\alpha}{2}}\sqrt{\frac{p(1-p)}{n}} \leq \overline{X} - p \leq u_{1-\frac{\alpha}{2}}\sqrt{\frac{p(1-p)}{n}}.$$

Odtud plyne, že pro parametr p platí

$$\overline{X} - u_{1-\frac{\alpha}{2}}\sqrt{\frac{p(1-p)}{n}} \leq p \leq \overline{X} + u_{1-\frac{\alpha}{2}}\sqrt{\frac{p(1-p)}{n}}.$$

Intervalový odhad parametru p obsahuje ale hodnotu rozptylu, která závisí na p. Hodnotu rozptylu nahradíme jeho odhadem $\frac{\overline{X}(1-\overline{X})}{n}$. Pro parametr p dostaneme intervalový odhad

$$\overline{X} - u_{1-\frac{\alpha}{2}}\sqrt{\frac{\overline{X}(1-\overline{X})}{n}} \leq p \leq \overline{X} + u_{1-\frac{\alpha}{2}}\sqrt{\frac{\overline{X}(1-\overline{X})}{n}}.$$