Úvod do teorie odhadu
(vybráno z přednášky Martiny Litschmannové)

Jak odhadnout parametry populace?

- Bodový odhad - parametr základního souboru aproximujeme jediným číslem
- Intervalový odhad – parametr populace aproximujeme intervalem, v němž s velkou pravděpodobností příslušný populační parametr leží.

Co je co v terminologii intervalových odhadů?

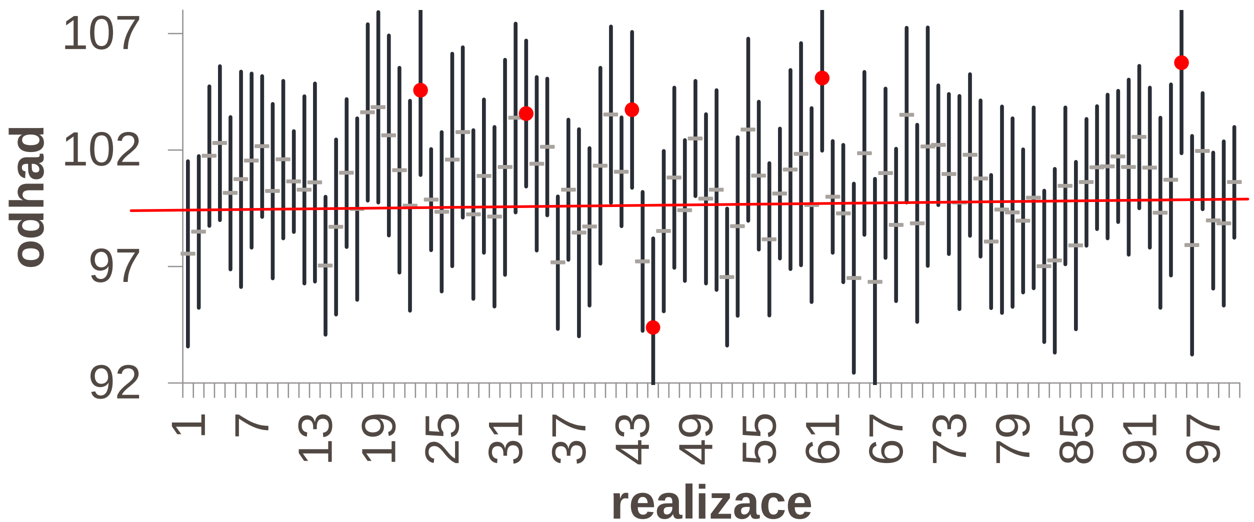
$$P(T_D \leq \theta \leq T_H) = 1 - \alpha$$

hledaný parametr
(konstanta, kterou nejsme schopni přesně určit)

spolehlivost odhadu,
tj. pravděpodobnost s níž hledaný parametr θ leží v intervalu $\langle T_D; T_H \rangle$

meze intervalu spolehlivosti
(náhodné veličiny)

Co to znamená, že spolehlivost odhadu je 1-α?



Simulace intervalových odhadů střední hodnoty (spolehlivost 0,95) získaných na základě opakovaných výběrů o rozsahu 30 z populace se střední hodnotou 100. 6 intervalů ze 100 neobsahuje skutečnou střední hodnou.

Jaké jsou typy intervalů spolehlivosti?

- oboustranné

$$P(\theta < T_D) = P(\theta > T_H) = \frac{\alpha}{2}$$

Tyto dvě podmínky zaručují, že $P(T_D \leq \theta \leq T_H) = 1 - \alpha$.

- jednostranné (odhadujeme-li například délku života nějakého zařízení, je pro nás důležitá pouze dolní mez)
    - levostranné: $P(\theta \geq T_D^*) = 1 - \alpha$
    - pravostranné : $P(\theta \leq T_H^*) = 1 - \alpha$

Vybrané intervalové odhady parametrů rozdělení náhodné veličiny

Intervalový odhad střední hodnoty náhodné veličiny s normálním rozdělením

a) známe-li rozptyl σ^2

Předpokládejme, že sledovaná náhodná veličina X má normální rozdělení s neznámou střední hodnotou μ a známým rozptylem σ^2. Vyberme vzorek z dané populace. Nechť má tento výběrový soubor rozsah n $(n < 0{,}05N)$ a průměr $\bar{x}$.

Intervalový odhad střední hodnoty μ se spolehlivostí $1-\alpha$ při známém rozptylu σ^2	
Oboustranný	$\left\langle \bar{x} - \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}}; \bar{x} + \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}} \right\rangle$
Levostranný	$\bar{x} - \frac{\sigma}{\sqrt{n}} z_{1-\alpha}$
Pravostranný	$\bar{x} + \frac{\sigma}{\sqrt{n}} z_{1-\alpha}$

kde z_p jsou p-kvantily $N(0;1)$

Intervalový odhad střední hodnoty náhodné veličiny s normálním rozdělením

b) neznáme-li rozptyl σ^2

Předpokládejme, že sledovaná náhodná veličina X má normální rozdělení s neznámou střední hodnotou μ a neznámým rozptylem σ^2. Vyberme vzorek z dané populace. Nechť má tento výběrový soubor rozsah n $(n < 0{,}05N)$, průměr $\bar{x}$ a výběrovou směrodatnou odchylku s.

Intervalový odhad střední hodnoty μ se spolehlivostí $1-\alpha$ při neznámém rozptylu σ^2	
Oboustranný	$\left\langle \bar{x} - \frac{s}{\sqrt{n}} t_{1-\frac{\alpha}{2}}; \bar{x} + \frac{s}{\sqrt{n}} t_{1-\frac{\alpha}{2}} \right\rangle$
Levostranný	$\bar{x} - \frac{s}{\sqrt{n}} t_{1-\alpha}$
Pravostranný	$\bar{x} + \frac{s}{\sqrt{n}} t_{1-\alpha}$

kde t_p jsou p-kvantily Studentova rozdělení s $n-1$ stupni volnosti

Intervalový odhad střední hodnoty - obecně

V obecném případě, kdy neznáme typ rozdělení, používáme tzv. **robustní (neparametrické) postupy**. Robustní postupy hodnocení náhodné veličiny typicky používáme v případech, kdy

- výběrový soubor obsahuje odlehlá pozorování, která nemohou být opravena a není vhodné je vyloučit,
- výběrový soubor nepochází z normálního rozdělení,
- výběrový soubor má velké rozptýlení dat.

Výklad robustních přístupů není součástí základního kurzu statistiky. Zájemci najdou základní informace v kapitole 4.4 (Úvod do statistiky).

Intervalový odhad rozptylu (sm. odchylky) norm. rozdělení

Předpokládejme, že sledovaná náhodná veličina X má normální rozdělení s neznámou střední hodnotou μ a neznámým rozptylem σ^2. Vyberme vzorek z dané populace. Nechť má tento výběrový soubor rozsah n $(n < 0{,}05N)$ a výběrovou směr. odchylku s.

Intervalový odhad rozptylu σ^2 se spolehlivostí $1 - \alpha$ při neznámé střední hodnotě μ	
Oboustranný	$\left\langle \frac{(n-1)s^2}{\chi_{1-\frac{\alpha}{2}}} ; \frac{(n-1)s^2}{\chi_{\frac{\alpha}{2}}} \right\rangle$
Levostranný	$\frac{(n-1)s^2}{\chi_{1-\alpha}}$
Pravostranný	$\frac{(n-1)s^2}{\chi_{\alpha}}$

kde χ_p jsou p-kvantily $\mathrm{Chí-kvadrát\ rozdělení}$ s $n-1$ $\mathrm{stupni\ volnosti}$

Intervalový odhad parametru binom. rozdělení (máme-li k dispozici dostatečně velký výběr)

standardní (Waldův) odhad

Intervalový odhad relativní četnosti π se spolehlivostí $1-\alpha$ $\left(n > 30, \frac{n}{N} < 0{,}05, n > \frac{9}{p(1-p)}\right)$	
Oboustranný	$\left\langle p - z_{1-\frac{\alpha}{2}} \sqrt{\frac{p(1-p)}{n}}; p + z_{1-\frac{\alpha}{2}} \sqrt{\frac{p(1-p)}{n}} \right\rangle$
Levostranný	$p - z_{1-\alpha} \sqrt{\frac{p(1-p)}{n}}$
Pravostranný	$p + z_{1-\alpha} \sqrt{\frac{p(1-p)}{n}}$

kde z_p jsou p-kvantily normovaného normálního rozdělení

Odhad rozsahu výběru
v případě, že odhadujeme střední hodnotu nebo parametr binom. rozdělení

Odhad rozsahu výběru potřebného pro nalezení interval. odhadu se spolehlivostí $1-\alpha$ a maximální přípustnou chybou Δ_{max}	
Odhadovaný populační parametr	Požadovaný rozsah výběru
Střední hodnota μ (známe σ)	$n \geq \left(\frac{\sigma}{\Delta_{max}} z_{1-\frac{\alpha}{2}}\right)^2$
Střední hodnota μ (neznáme σ)	$n \geq \left(\frac{s_1}{\Delta_{max}} t_{1-\frac{\alpha}{2}}\right)^2$
Parametr binom. rozdělení π	$n \geq \left(z_{1-\frac{\alpha}{2}}\right)^2 \frac{p_1(1-p_1)}{\Delta_{max}^2}$
	$n \geq \left(z_{1-\frac{\alpha}{2}}\right)^2 \frac{1}{4\Delta_{max}^2}$

Intervalový odhad poměru rozptylů dvou náhodných veličin s normálním rozdělením

Intervalový odhad poměru rozptylů $\frac{\sigma_1^2}{\sigma_2^2}$ se spolehlivostí $1-\alpha$	
Oboustranný	$\left\langle \frac{1}{f_{1-\frac{\alpha}{2}}}\frac{S_1^2}{S_2^2}; \frac{1}{f_{\frac{\alpha}{2}}}\frac{S_1^2}{S_2^2} \right\rangle$
Levostranný	$\frac{1}{f_{1-\alpha}}\frac{S_1^2}{S_2^2}$
Pravostranný	$\frac{1}{f_{\alpha}}\frac{S_1^2}{S_2^2}$

kde f_p jsou p-kvantily $\text{Fisherova} - \text{Snedecorova rozdělení}$
$s\ n_1 - 1$ stupni volnosti v čitateli a $n_2 - 1$ stupni volnosti ve jmenovateli

(Vzorce a tabulky)

Intervalový odhad rozdílů středních hodnot dvou náhodných veličin s normálním rozdělením

a) známe rozptyly $\sigma_1{}^2$ a $\sigma_2{}^2$ obou populací

Mějme dvě populace s normálním rozdělením, jejichž rozptyly $\sigma_1{}^2$ a $\sigma_2{}^2$ známe. Z těchto populací jsme provedli dva nezávislé náhodné výběry o rozsahu n_1 a n_2 $(n_1 < 0{,}05N, n_2 < 0{,}05N)$, a určili jejich průměry $\bar{x}_1$ a $\bar{x}_2$.

Intervalový odhad rozdílu středních hodnot $\mu_1 - \mu_2$ se spolehlivostí $1-\alpha$ (známe σ_1, σ_2)	
Oboustranný	$\left\langle (\bar{x}_1 - \bar{x}_2) - z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}; (\bar{x}_1 - \bar{x}_2) + z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right\rangle$
Levostranný	$(\bar{x}_1 - \bar{x}_2) - z_{1-\alpha}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$
Pravostranný	$(\bar{x}_1 - \bar{x}_2) + z_{1-\alpha}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

kde z_p jsou p-kvantily normovaného normálního rozdělení

(Vzorce a tabulky)

Intervalový odhad rozdílů středních hodnot dvou náhodných veličin s normálním rozdělením

b) neznáme jejich rozptyly σ_1^2 a σ_2^2, ale víme, že $\sigma_1^2 = \sigma_2^2$

Mějme dvě populace s normálním rozdělením, jejichž rozptyly neznáme, ale víme, že jsou shodné. Z těchto populací jsme provedli dva nezávislé náhodné výběry o rozsahu n_1 a n_2 $(n_1 < 0{,}05N, n_2 < 0{,}05N)$, a určili jejich průměry $\bar{x}_1$ a $\bar{x}_2$ a výběrové směrodatné odchylky s_1 a s_2.

Intervalový odhad rozdílu středních hodnot $\mu_1 - \mu_2$ se spolehlivostí $1 - \alpha$ (neznáme σ_1^2, σ_2^2, ale víme, že $\sigma_1^2 = \sigma_2^2$)	
Oboustranný	$\left\langle (\bar{x}_1 - \bar{x}_2) - t_{1-\frac{\alpha}{2}} \sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}; (\bar{x}_1 - \bar{x}_2) + t_{1-\frac{\alpha}{2}} \sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right\rangle$
Levostranný	$(\bar{x}_1 - \bar{x}_2) - t_{1-\alpha} \sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$
Pravostranný	$(\bar{x}_1 - \bar{x}_2) + t_{1-\alpha} \sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$

kde t_p jsou p-kvantily Studentova rozdělení s $n_1 + n_2 - 2$ stupni volnosti

(Vzorce a tabulky)

Intervalový odhad rozdílů středních hodnot dvou náhodných veličin s normálním rozdělením

c) neznáme jejich rozptyly σ_1^2 a σ_2^2, a <u>nelze předpokládat</u>, že $\sigma_1^2 = \sigma_2^2$

Mějme dvě populace s normálním rozdělením, jejichž rozptyly neznáme a nelze předpokládat, že jsou shodné. Z těchto populací jsme provedli dva nezávislé náhodné výběry o rozsahu n_1 a n_2 $(n_1 < 0{,}05N, n_2 < 0{,}05N)$, a určili jejich průměry $\bar{x}_1$ a $\bar{x}_2$ a výběrové směrodatné odchylky s_1 a s_2.

Intervalový odhad rozdílu středních hodnot $\mu_1 - \mu_2$ se spolehlivostí $1 - \alpha$ (neznáme σ_1^2, σ_2^2, a **nelze předpokládat**, že $\sigma_1^2 = \sigma_2^2$)	
Oboustranný	$\left\langle (\bar{x}_1 - \bar{x}_2) - t_{1-\frac{\alpha}{2}} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} ; (\bar{x}_1 - \bar{x}_2) + t_{1-\frac{\alpha}{2}} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \right\rangle$
Levostranný	$(\bar{x}_1 - \bar{x}_2) - t_{1-\alpha} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$
Pravostranný	$(\bar{x}_1 - \bar{x}_2) + t_{1-\alpha} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$

kde t_p jsou p-kvantily Studentova rozdělení s $\left[\frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2} \right)^2}{\left(\frac{s_1^2}{n_1} \right)^2 \frac{1}{n_1+1} + \left(\frac{s_2^2}{n_2} \right)^2 \frac{1}{n_2+1}} - 2 \right]$ stupni volnosti

(Vzorce a tabulky)

Intervalový odhad pro rozdíl parametrů binom. rozdělení dvou náhodných veličin

Mějme dvě populace. Z těchto populací jsme provedli dva nezávislé náhodné výběry o rozsahu n_1 a n_2 $(n_1 < 0{,}05N, n_2 < 0{,}05N)$. Výběr z první populace obsahoval x_1 prvků se sledovanou vlastností, výběr z druhé populace obsahoval x_2 prvků se sledovanou vlastností. Výběrové relativní četnosti p_1, p_2 jsme pak určili dle vztahů $p_1 = \frac{x_1}{n_1}$, $p_2 = \frac{x_2}{n_2}$.

Intervalový odhad rozdílu relativních četností $\pi_1 - \pi_2$ se spolehlivostí $1-\alpha$ $\left(\forall i \in \{1,2\}\colon\ n_i > 30, \frac{n_i}{N_i} < 0{,}05, n_i > \frac{9}{p_i(1-p_i)}\right)$	
Oboustranný	$\left\langle (p_1 - p_2) - z_{1-\frac{\alpha}{2}}\sqrt{p(1-p)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}; (p_1 - p_2) + z_{1-\frac{\alpha}{2}}\sqrt{p(1-p)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}\right\rangle$
Levostranný	$(p_1 - p_2) - z_{1-\alpha}\sqrt{p(1-p)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$
Pravostranný	$(p_1 - p_2) + z_{1-\alpha}\sqrt{p(1-p)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$

kde z_p jsou p-kvantily normovaného normálního rozdělení

(Vzorce a tabulky)