

STATISTIKA 2

Statistika a statistické zpracování dat

Blok 2-3

Jiří Fišer

22. listopadu 2024

Bodový a intervalový odhad – Úvodní příklad

- Jste manažerem firmy vyrábějící žárovky a chcete odhadnout průměrnou životnost (**střední hodnotu**) žárovek bez testování každé z nich.
- Vyberete náhodně 10 žárovek a změříte jejich životnost:

$$x = (850, 870, 890, 900, 920, 940, 960, 980, 1000, 1020).$$

- Na základě těchto měření chcete odhadnout průměrnou životnost všech žárovek.

Bodový odhad střední hodnoty

- Toto je bodový odhad průměrné životnosti žárovek pomocí výběrového průměru:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = 933 \text{ hodin.}$$

Intervalový odhad střední hodnoty

- Pro určení přesnosti odhadu spočítáme směrodatnou odchylku:

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} \approx 55,08 \text{ hodin.}$$

- Pro 95% interval spolehlivosti použijeme kritickou hodnotu $t \approx 2,262$ Studentova rozdělení (pro $n - 1 = 9$ stupňů volnosti).
- Intervalový odhad průměru:

$$\bar{x} \pm t \cdot \frac{s}{\sqrt{n}} = 933 \pm 39,39 \text{ hodin.}$$

- Interval: $\langle 893,61; 972,39 \rangle$ hodin.

Statistické odhady

- Odhadujeme neznámé parametry základního souboru na základě výběru.
- **Statistický odhad** je funkce výběrových dat používaná k odhadu parametrů.
- Odhady jsou náhodné veličiny a mohou se lišit při různých výběrech.

Bodový odhad

- **Bodový odhad** parametru μ je statistika $T(X)$, která aproximuje μ .
- Bodový odhad se může měnit s každým novým výběrem.
- Metody pro volbu statistiky: metoda momentů, metoda maximální věrohodnosti.

Metoda momentů

- Porovnává teoretické momenty rozdělení s výběrovými momenty (viz skripta).
- Odhad parametru získáme řešením takové rovnice.
- Nejčastěji se používají:

$$\mu \approx \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad \text{a} \quad \sigma^2 \approx s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

Příklad: Poissonovo rozdělení

Pravděpodobnostní funkce:

$$p(x; \lambda) = \frac{\lambda^x e^{-\lambda}}{x!}, \quad x = 0, 1, 2, \dots$$

Teoretický moment:

$$\mu_1 = \mathbb{E}[X] = \lambda$$

Výběrový moment:

$$m_1 = \frac{1}{n} \sum_{i=1}^n x_i$$

Odhad parametru λ :

$$\lambda \approx m_1$$

Intervalový odhad parametrů

- **Interval spolehlivosti:** interval, ve kterém se s pravděpodobností $1 - \alpha$ nachází skutečná hodnota parametru.
- **Hladina významnosti α :** pravděpodobnost, že parametr neleží v intervalu.
- **Stupeň spolehlivosti $1 - \alpha$:** pravděpodobnost, že parametr leží v intervalu.

Intervalový odhad střední hodnoty

Pro velké n ($n \geq 30$):

$$\left\langle \bar{X} - \frac{s}{\sqrt{n}} \cdot u_{1-\frac{\alpha}{2}}, \bar{X} + \frac{s}{\sqrt{n}} \cdot u_{1-\frac{\alpha}{2}} \right\rangle$$

Pro malé n ($n < 30$):

$$\left\langle \bar{X} - \frac{s}{\sqrt{n}} \cdot t_{1-\frac{\alpha}{2}}(n-1), \bar{X} + \frac{s}{\sqrt{n}} \cdot t_{1-\frac{\alpha}{2}}(n-1) \right\rangle$$

kde:

- $\bar{X}$ je výběrový průměr,
- s je výběrová směrodatná odchylka,
- $u_{1-\frac{\alpha}{2}}$ je kvantil normovaného normálního rozdělení,
- $t_{1-\frac{\alpha}{2}}(n-1)$ je kvantil Studentova t -rozdělení.

Definice 6.16. Praktický vzorec pro výpočet intervalu spolehlivosti pro střední hodnotu náhodné veličiny X při velkém počtu pozorování ($n > 30$):

$$\left\langle \bar{X} - \frac{s}{\sqrt{n}} \cdot u_{1-\frac{\alpha}{2}}, \bar{X} + \frac{s}{\sqrt{n}} \cdot u_{1-\frac{\alpha}{2}} \right\rangle,$$

kde:

- $\bar{X}$ je výběrový průměr,
- s je výběrová směrodatná odchylka,
- n je počet pozorování (velikost výběru),
- $u_{1-\frac{\alpha}{2}}$ je kvantil normovaného normálního rozdělení odpovídající zvolené hladině spolehlivosti $1 - \alpha$,
- α je hladina významnosti (obvykle $\alpha = 0,05$, což odpovídá 95% intervalu spolehlivosti).

Definice 6.17. Praktický vzorec pro výpočet intervalu spolehlivosti pro střední hodnotu náhodné veličiny X při malém počtu pozorování ($n < 30$):

$$\left\langle \bar{X} - \frac{s}{\sqrt{n}} \cdot t_{1-\frac{\alpha}{2}}(n-1), \bar{X} + \frac{s}{\sqrt{n}} \cdot t_{1-\frac{\alpha}{2}}(n-1) \right\rangle,$$

kde:

- $\bar{X}$ je výběrový průměr,
- s je výběrová směrodatná odchylka,
- n je počet pozorování (velikost výběru),
- $t_{1-\frac{\alpha}{2}}(n-1)$ je kvantil Studentova t -rozdělení s $n-1$ stupni volnosti odpovídající zvolené hladině spolehlivosti $1 - \alpha$,
- α je hladina významnosti (obvykle $\alpha = 0,05$, což odpovídá 95% intervalu spolehlivosti).

Určení přesnosti odhadu

Velikost intervalu spolehlivosti, označovaná jako Δ , nám říká, jak přesný je náš odhad střední hodnoty μ . Vyjadřuje se následovně:

$$\Delta = \frac{s}{\sqrt{n}} \cdot u_{1-\frac{\alpha}{2}}, \quad \text{nebo} \quad \Delta = \frac{s}{\sqrt{n}} \cdot t_{1-\frac{\alpha}{2}}(n-1).$$

Tento interval můžeme zapsat jako $\mu \in \langle \bar{X} - \Delta, \bar{X} + \Delta \rangle$, což znamená, že skutečná hodnota μ se nachází někde uvnitř tohoto intervalu (při zvolené hladině spolehlivosti $1 - \alpha$).

Příklad: Intervalový odhad μ

Praktický příklad

Příklad 6.18. Měřili jsme průměr vačkového hřídele na 250 součástkách. Předpokládáme, že data mají normální rozdělení. Z výsledků měření jsme určili výběrový průměr $\bar{x} = 995,6$ a výběrový rozptyl $s^2 = 134,7$. Určete interval spolehlivosti pro střední hodnotu základního souboru při hladině významnosti 5 %.

Řešení: Úlohu vyřešíme pomocí Excelu, kde kritickou hodnotu normálního rozdělení získáme pomocí funkce NORM.S.INV:

$$\Delta = \frac{s}{\sqrt{n}} \cdot u_{1-\frac{\alpha}{2}} = \frac{\sqrt{134,7}}{\sqrt{249}} \cdot \text{NORM.S.INV}(0,975) \approx 1,441558.$$

Intervalový odhad střední hodnoty je tedy:

$$\langle \bar{x} - \Delta; \bar{x} + \Delta \rangle = \langle 994,1584; 997,0416 \rangle.$$

Intervalový odhad rozptylu

$$\left\langle \frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)}, \frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2}}(n-1)} \right\rangle$$

kde:

- s^2 je výběrový rozptyl,
- χ^2 jsou kvantily chí-kvadrát rozdělení s $n - 1$ stupni volnosti.

Příklad: Intervalový odhad rozptylu

Definice 6.19. Praktický vzorec pro výpočet intervalu spolehlivosti pro rozptyl σ^2 :

$$\left\langle \frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)}, \frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2}}(n-1)} \right\rangle,$$

kde:

- s^2 je výběrový rozptyl,
- n je počet pozorování (velikost výběru),
- $\chi^2_{\frac{\alpha}{2}}(n-1)$ a $\chi^2_{1-\frac{\alpha}{2}}(n-1)$ jsou kvantily chí-kvadrát rozdělení s $n-1$ stupni volnosti odpovídající hladinám významnosti $\frac{\alpha}{2}$ a $1 - \frac{\alpha}{2}$,
- α je hladina významnosti (obvykle $\alpha = 0,05$, což odpovídá 95% intervalu spolehlivosti).

Tento interval spolehlivosti pro rozptyl σ^2 vyjadřuje, že skutečná hodnota rozptylu se s pravděpodobností $1 - \alpha$ nachází v daném intervalu.

Příklad: Intervalový odhad σ^2

Praktický příklad

Příklad 6.20. Určete oboustranný interval spolehlivosti pro rozptyl normálně rozloženého základního souboru při hladinách spolehlivosti 0,90, 0,95 a 0,99. Měření byla provedena na vzorku s velikostí $n = 12$ a bylo zjištěno, že vzorkový rozptyl je $s^2 = 0,64$. Posuďte získané výsledky.

Řešení: Známe obecný tvar intervalu spolehlivosti pro rozptyl:

$$\left\langle \frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)}, \frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2}}(n-1)} \right\rangle.$$

Dosadíme konkrétní hodnoty:

$$n = 12, \quad s^2 = 0,64, \quad n - 1 = 11.$$

Řešení pro hladinu spolehlivosti 0,90:

$$\frac{11 \cdot 0,64}{\chi^2_{0,95}(11)} \leq \sigma^2 \leq \frac{11 \cdot 0,64}{\chi^2_{0,05}(11)}$$

Příklad: Intervalový odhad σ^2

Použití Excelu: Pro výpočet kritických hodnot použijeme funkci CHISQ.INV. Tato funkce vrací kvantil rozdělení χ^2 . Syntaxe je CHISQ.INV(pravděpodobnost, stupně volnosti).

Kritické hodnoty jsou:

- $\chi_{0,95}^2(11) = \text{CHISQ.INV}(0,95; 11) = \text{CHISQ.INV.RT}(0,05; 11) = 19,675,$
- $\chi_{0,05}^2(11) = \text{CHISQ.INV}(0,05; 11) = 4,575.$

Interval spolehlivosti pro rozptyl je tedy:

$$\frac{11 \cdot 0,64}{19,675} \leq \sigma^2 \leq \frac{11 \cdot 0,64}{4,575}.$$

Výsledkem je interval: $\langle 0,358, 1,539 \rangle$.

Stejný postup použijeme pro zbývající dva případy.



- Bodové odhady poskytují konkrétní hodnotu parametru, ale nemusí být přesné.
- Intervalové odhady poskytují rozsah hodnot s určitou spolehlivostí.
- Metoda momentů a metoda maximální věrohodnosti jsou základní techniky odhadu parametrů.