TECHNICKÁ UNIVERZITA V LIBERCI

Nové možnosti rozvoje vzdělávání na Technické univerzitě v Liberci

Specifický cíl A3: Tvorba nových profesně zaměřených studijních programů

NPO_TUL_MSMT-16598/2022

Matematická statistika (statistická indukce) – část 1

Ing. Vladimíra Hovorková Valentová, Ph.D.

Financováno Evropskou unií
NextGenerationEU

Národní plán obnovy

MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

Úvod

- Na základě výběrových dat usuzujeme na obecnější skutečnosti, týkající se základního souboru (dále ZS), tj. provádíme zevšeobecňující (induktivní) úsudek.
- Induktivní usuzování pomocí matematicko-statistických metod se nazývá *statistická indukce*.
- Induktivní uvažování s sebou vždy nese riziko nesprávného úsudku (= riziko omylu).
- Výběrová data musí být pořízena náhodným výběrem (o tom, zda určitá jednotka ZS bude vybrána, rozhoduje pouze náhoda).

Statistické indukce zahrnuje:

a) teorii odhadu,
b) testování statistických hypotéz.

Teorie odhadu

- metody, kterými lze z napozorovaných hodnot náhodného výběru získat co nejlepší odhady neznámých parametrů jejího rozdělení.

1. Bodový odhad

- Spočívá v nahrazení neznámé hodnoty parametru ZS hodnotou vhodné výběrové charakteristiky, která bude sloužit jako dobrá náhrada neznámého parametru.
- Vhodnost jednotlivých odhadů posuzujeme podle několika vlastností.

Vlastnosti bodového odhadu:

1. nevychýlenost (nestrannost, nezkreslenost),
2. konzistence,
3. vydatnost,
4. výběrová charakteristika má být postačující.

Symbolika: Parametry v ZS značíme obecně Θ (konkrétně např. $\mu, \sigma, \dots$).

Výběrové charakteristiky značíme obecně *t* (např. $\bar{x}, s_x, \dots$).

$t - \Theta$ je *výběrová chyba.*

Symbolický zápis bodového odhadu: $est\,\Theta = t$ nebo $t \sim \Theta$.

2. Intervalový odhad

- Jedná se o odhad neznámého parametru ZS pomocí intervalu, který s pravděpodobností $P = 1 - \alpha$ bude obsahovat skutečnou hodnotu odhadované charakteristiky ZS Θ.

Spolehlivost odhadu $1 - \alpha$

- Je to pravděpodobnost; $0 < \alpha < 1$.
- Volíme vždy číslo blízké 1, nejčastěji 0,95 (event. 0,99 nebo 0,9).
- Čím vyšší spolehlivost žádáme, tím je, za jinak stejných podmínek, interval spolehlivosti (dále IS) širší.

Riziko odhadu α

- Udává, v kolika případech ze 100 (v jakém % případů) nebude IS pokrývat odhadovaný parametr Θ.

Intervaly spolehlivosti mohou být konstruovány jako:

1. oboustranné $\Theta_d < \Theta < \Theta_h$; kde Θ_h je horní mez, Θ_d je dolní mez.

2. jednostranné: pravostranné $\Theta < \Theta_h$

   levostranné $\Theta_d < \Theta$.

Odhad parametru μ (střední hodnoty) normálního rozdělení

1. Bodový odhad

Bodovým odhadem střední hodnoty $\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$ je *výběrový průměr* $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$.

Je to nevychýlený odhad, neboť platí: $E(\overline{x}) = \mu$.

2. Intervalový odhad

Při konstrukci IS pro parametr μ rozlišujeme 3 případy:

I. Velký výběr z normálního rozdělení se známým rozptylem σ²:

Oboustranný IS:

$$P\left(\bar{x} - u_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + u_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

Pravostranný IS:

$$P\left(\mu < \bar{x} + u_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

Levostranný IS:

$$P\left(\bar{x} - u_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}} < \mu\right) = 1 - \alpha$$

$\Delta = u_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$ je *přípustná chyba odhadu.*

II. Velký výběr z normálního rozdělení s neznámým rozptylem σ²:

Při řešení praktických úloh obvykle neznáme rozptyl ZS σ². Odhadujeme jej pomocí *výběrového rozptylu* $s_x'^2$:

$$s_x'^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}.$$

Oboustranný IS:

$$P\left(\bar{x} - u_{1-\frac{\alpha}{2}} \cdot \frac{s_x'}{\sqrt{n}} < \mu < \bar{x} + u_{1-\frac{\alpha}{2}} \cdot \frac{s_x'}{\sqrt{n}}\right) = 1 - \alpha$$

Pravostranný IS:

$$P\left(\mu < \bar{x} + u_{1-\alpha} \cdot \frac{s'_x}{\sqrt{n}}\right) = 1-\alpha$$

Levostranný IS:

$$P\left(\bar{x} - u_{1-\alpha} \cdot \frac{s'_x}{\sqrt{n}} < \mu\right) = 1-\alpha$$

III.Malý výběr z normálního rozdělení s neznámým rozptylem σ²:

Kvantily rozdělení N (μ; σ²) nahradíme kvantily Studentova rozdělení t s $\nu = n-1$ stupni volnosti.

Oboustranný IS:

$$P\left[\bar{x} - t_{1-\frac{\alpha}{2}}(n-1)\cdot\frac{s'_x}{\sqrt{n}} < \mu < \bar{x} + t_{1-\frac{\alpha}{2}}(n-1)\cdot\frac{s'_x}{\sqrt{n}}\right] = 1-\alpha$$

Pravostranný IS:

$$P\left[\mu < \bar{x} + t_{1-\alpha}(n-1)\cdot\frac{s'_x}{\sqrt{n}}\right] = 1-\alpha$$

Levostranný IS:

$$P\left[\bar{x} - t_{1-\alpha}(n-1)\cdot\frac{s'_x}{\sqrt{n}} < \mu\right] = 1-\alpha$$

Odhad parametru π (relativní četnosti) alternativního rozdělení

Předpoklad: Je třeba mít k dispozici výběr dostatečně velkého rozsahu; to je zajištěno splněním podmínky $n\pi(1-\pi) > 9$.

1. Bodový odhad

Bodovým odhadem relativní četnosti $\pi = \frac{M}{N}$ je *výběrová relativní četnost* (výběrový podíl) $p = \frac{m}{n}$,

kde:

M .... počet jednotek se sledovanou vlastností v ZS,

N .... celkový počet jednotek ZS,

m .... počet jednotek se sledovanou vlastností ve výběrovém souboru,

n .... rozsah výběru.

2. Intervalový odhad

Oboustranný IS:

$$P\left[p - u_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{p(1-p)}{n}} < \pi < p + u_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{p(1-p)}{n}} \right] = 1-\alpha$$

$\Delta = u_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{p(1-p)}{n}}$ je *přípustná chyba odhadu.*

Pravostranný IS:

$$P\left[\pi < p + u_{1-\alpha} \cdot \sqrt{\frac{p(1-p)}{n}} \right] = 1-\alpha$$

Levostranný IS:

$$P\left[p - u_{1-\alpha} \cdot \sqrt{\frac{p(1-p)}{n}} < \pi \right] = 1-\alpha$$

Stanovení minimálního rozsahu výběru

Pro stanovení minimálního rozsahu výběru vycházíme ***ze vzorce přípustné chyby odhadu parametru*** μ , jehož jednoduchou úpravou dostaneme:

$$n \geq \frac{u^2_{1-\frac{\alpha}{2}} \cdot \sigma^2}{\Delta^2}.$$

Pokud neznáme σ^2, použijeme jeho bodový odhad s'^2_x.

Budeme-li vycházet ***ze vzorce přípustné chyby odhadu parametru*** π, dostaneme:

$$n \geq \frac{u^2_{1-\frac{\alpha}{2}} \cdot p(1-p)}{\Delta^2}.$$

Odhad parametru σ² (rozptylu) normálního rozdělení

1. Bodový odhad

Bodovým odhadem rozptylu $\sigma^2 = \frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N}$ je *výběrový rozptyl* $s'^2_x = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$.

Je to nezkreslený a konzistentní odhad.

2. Intervalový odhad

Při konstrukci IS pro parametr σ^2 rozlišujeme 2 případy - buď známe parametr μ nebo ho neznáme.
V praxi je častější případ, kdy parametr μ neznáme, proto se na něj zaměříme.

Oboustranný IS:

$$P\left[\frac{(n-1)s'^2_x}{\chi^2_{1-\frac{\alpha}{2}}(\nu)} < \sigma^2 < \frac{(n-1)s'^2_x}{\chi^2_{\frac{\alpha}{2}}(\nu)}\right] = 1-\alpha$$

Pravostranný IS:

$$P\left[\sigma^2 < \frac{(n-1)s'^2_x}{\chi^2_{\alpha}(\nu)}\right] = 1-\alpha$$

Levostranný IS:

$$P\left[\frac{(n-1)s'^2_x}{\chi^2_{1-\alpha}(\nu)} < \sigma^2\right] = 1-\alpha$$

Pozn.: $\nu = n - 1$.