MATEMATICKÁ STATISTIKA – 1. ČÁST

- na základě výběrových dat usuzujeme na obecnější skutečnosti, týkající se celého *základního souboru, tzv. populace*
- provádíme zevšeobecňující neboli induktivní úsudek
- usuzování pomocí matematicko-statistických metod se nazývá *statistická indukce*
- induktivní uvažování s sebou nese riziko nesprávného úsudku (riziko omylu)
- aby bylo možno správně aplikovat metody a postupy matematické statistiky, musí být výběrová data pořízena *náhodným výběrem*, což znamená, že o tom, zda určitá jednotka ZS bude vybrána nebo ne rozhoduje pouze náhoda.

Matematická statistika zahrnuje dvě oblasti:
- teorii odhadu
- testování statistických hypotéz.

Teorie odhadu

- souhrn metod a postupů, kterými lze z napozorovaných hodnot NV získat co nejlepší odhady neznámých parametrů rozdělení NV.

Bodový odhad

- spočívá v nahrazení neznámé hodnoty parametru základního souboru (dále jen ZS) hodnotou vhodné výběrové charakteristiky, která bude sloužit jako dobrá náhrada neznámého parametru
- vhodnost jednotlivých odhadů posuzujeme podle jeho vlastností.

Vlastnosti bodového odhadu:

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

Symbolika:
- parametry základního souboru značíme obecně Θ (konkrétně např. $\mu, \sigma, ...$)
- výběrové charakteristiky značíme obecně *t* (konkrétně např. $\bar{x}, s_x, ...$)
- *výběrová chyba:* $t - \Theta$
- *symbolický zápis bodového odhadu:* $est\ \Theta = t$ nebo $t \sim \Theta$.

Intervalový odhad

- spočívá v konstrukci náhodného intervalu, od něhož se zvolenou pravděpodobností $P = 1 - \alpha$ očekáváme, že bude obsahovat skutečnou hodnotu neznámého parametru
- dovoluje, abychom uvažovali pravděpodobnost, s níž lze očekávat, že odhad je správný.

Spolehlivost odhadu $\mathbf{1 - \alpha}$

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

Riziko odhadu α

- udává, v kolika případech ze 100 (tj. v jakém procentu případů) nebude IS pokrývat odhadovaný parametr Θ.

Intervaly spolehlivosti mohou být konstruovány jako:

1. oboustranné, kde $\Theta_d < \Theta < \Theta_h$
2. jednostranné: pravostranné, kde $\Theta < \Theta_h$
   levostranné, kde $\Theta > \Theta_d$.

Θ_h je horní mez, Θ_d je dolní mez.

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* $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$. Je to nevychýlený odhad střední hodnoty μ.

2. Intervalový odhad

Při konstrukci IS pro parametr μ rozlišujeme následující situace:

- **Velký výběr z normálního rozdělení se známým rozptylem σ^2:**

  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.*

- **Velký výběr z normálního rozdělení s neznámým rozptylem σ^2:**

  Při řešení praktických úloh obvykle neznáme rozptyl ZS σ^2, a proto ho odhadujeme 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:

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

- **Malý výběr z normálního rozdělení s neznámým rozptylem σ^2:**

  Kvantily normovaného normálního rozdělení ve vzorcích z předcházející situace nahradíme kvantily Studentova rozdělení s $\nu = \mathrm{n} - 1$ 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 σ^2 (rozptylu) normálního rozdělení

1. Bodový odhad

Bodovým odhadem rozptylu $\sigma^2 = \frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N}$ je *výběrový rozptyl* $s_x^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$. Je to nezkreslený a konzistentní odhad σ^2 .

2. Intervalový odhad

Při konstrukci IS pro parametru σ^2 rozlišujeme teoreticky dva případy: buď druhý parametr rozdělení (μ) známe nebo neznáme. V praxi je daleko častější případ, kdy parametr μ neznáme, proto se na něj v následujícím textu zaměříme.

Oboustranný IS:

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

Pravostranný IS:

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

Levostranný IS:

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

Odhad parametru π (relativní četnosti) alternativního rozdělení

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) $p = \frac{m}{n}$.

M ... počet jednotek se sledovanou vlastností v ZS
N ... rozsah 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

Pokud při stanovení minimálního rozsahu výběru, potřebného k dodržení zadaných podmínek, vycházíme ***ze vzorce přípustné chyby odhadu parametru μ***, jeho úpravou dostaneme:

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

Pokud neznáme rozptyl základního souboru σ^2, použijeme jeho bodový odhad s_x^2.

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

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

Pokud neznáme relativní četnost základního souboru π, použijeme její bodový odhad p.