8 Výběrové charakteristiky

Statistická indukce

= metoda, která dovoluje stanovit vlastnosti celku (základního souboru) na základě pozorování jeho části (náhodného výběru)

- např. ve výběru 100 lidí bude průměr IQ 110 → střední hodnota IQ bude v okolí 110 (V jakém okolí? S jakou pravděpodobností?)

Výběrové charakteristiky

populační parametr (konstanta)	výběrová charakteristika (náhodná veličina)
střední hodnota μ	výběrový průměr $\overline{X}$
medián $x_{0.5}$	výběrový medián$\widetilde{X}_{0.5}$
směrodatná odchylka σ	výběrová směrodatná odchylka S
pravděpodobnost π	relativní četnost p

Odhady parametrů populace

- bodový odhad
- intervalový odhad

Limitní věty

Slabý zákon velkých čísel Mějme nekonečný náhodný výběr $X_1, X_2, ...$ z rozdělení se střední hodnotou μ a konečným rozptylem, kde $X_1, X_2, ...$ jsou nekorelované náhodné veličiny. Potom platí, že výběrový průměr $\overline{X_n}$ vypočítaný z prvních n pozorování se pro $n \to \infty$ blíží ke střední hodnotě μ, což zapisujeme

$$\lim_{n\to\infty} \left[P\left(|\overline{X}_n - \mu|\right) < \varepsilon\right] = 0 \ \ \forall \varepsilon > 0.$$

Centrální limitní věta Jsou-li X_i nezávislé náhodné veličiny se stejnou střední hodnotou μ a se stejným konečným rozptylem σ^2, pak výběrový průměr má při dostatečně velkém počtu pozorování přibližně normální rozdělení, ať už X_i pocházejí z libovolného rozdělení.

$$\overline{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

Důsledky centrální limitní věty

Součet n hodnot:

$$\overline{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right) \to \overline{X}_n \cdot n = \sum_{i=1}^{n} X_i \sim N\left(n\mu, n\sigma^2\right)$$

Relativní četnost p binomického rozdělení:

$$p \sim N\left(\pi; \frac{\pi\left(1-\pi\right)}{n}\right)$$

Rozdíl výběrových průměrů:

$$\overline{X}_1 - \overline{X}_2 \sim N\left(\mu_x - \mu_y, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$

Rozdíl relativních četností $p_1 - p_2$ binomického rozdělení:

$$p_1 - p_2 \sim N\left(\pi_1 - \pi_2; \frac{\pi_1 (1 - \pi_1)}{n_1} + \frac{\pi_1 (1 - \pi_1)}{n_2}\right)$$

Spojitá rozdělení využívaná při statistické indukci

χ^2 - rozdělení (Pearsonovo)

- Definován jako suma n kvadrátů normovaných normálních rozdělení

$$X_1, ..., X_n \sim N(0;1): \sum_{i=1}^{n} X_i^2 \sim \chi_n^2$$

- Má jediný parametr - počet stupňů volnosti
- Použití při odhadu rozptylu

$$\frac{S^2}{\sigma^2}(n-1) \sim \chi_{n-1}^2$$

Studentovo (t) rozdělení

- Máme-li $Z \sim N(0;1)$ a $V \sim \chi_k^2$ pak náhodná veličina

$$T = \frac{Z}{\sqrt{V/k}}$$

  má Studentovo t rozdělení.
- Má jediný parametr - počet stupňů volnosti
- Použití při odhadu střední hodnoty bez přesné znalosti rozptylu

$$\frac{\overline{X} - \mu}{S}\sqrt{n} \sim t_{n-1}$$

  (při znalosti střední hodnoty používáme normální rozdělení)
- Pro vysoká n platí $t_n \to N(0;1)$

Fisher-Snedecorovo (F) rozdělení

- Definován jako normovaný podíl dvou χ^2 rozdělení

$$V \sim \chi_m^2, w \sim \chi_n^2: \frac{\frac{V}{m}}{\frac{W}{n}} \sim F_{m,n}$$

- Má dva parametry - počty stupňů volnosti obou χ^2 rozdělení
- Používá se k testování shody rozptylů