MATEMATICKÁ STATISTIKA

Dana Černá

http://www.fp.tul.cz/kmd/

Katedra matematiky a didaktiky matematiky

Technická univerzita v Liberci

Intervalový odhad

$(P_1, P_2) \subset (-\infty, \infty)$ se nazývá intervalový odhad parametru θ se spolehlivostí $1 - \alpha$, jestliže $P\left(\theta \in (P_1, P_2)\right) = 1 - \alpha$.

Intervalový odhad tvaru $(\hat{\theta} + P, \hat{\theta} - P)$, $\hat{\theta},\ P \in \mathbb{R}$, se nazývá oboustranný intervalový odhad.

Intervalový odhad tvaru (P_1, ∞), $P_1 \in \mathbb{R}$, se nazývá dolní intervalový odhad.

Intervalový odhad tvaru $(-\infty, P_2)$, $P_2 \in \mathbb{R}$, se nazývá horní intervalový odhad.

Intervalový odhad střední hodnoty náhodné veličiny s normálním rozdělením pravděpodobnosti

Předpokládejme, že $X_1, \ldots, X_n$ je náhodný výběr z rozdělení $N\left(\mu, \sigma^2\right)$, kde rozptyl $\sigma^2 > 0$ známe. Potom $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ má rozdělení $N\left(\mu, \sigma^2/n\right)$.

Protože $\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$, platí $P\left(\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} < u\right) = \Phi(u)$, kde Φ je distribuční funkce rozdělení $N(0,1)$.

Pro požadovanou spolehlivost $1-\alpha$ dostaneme $P\left(\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} < u_{1-\alpha}\right) = 1-\alpha$ kde $u_{1-\alpha} := \Phi^{-1}(1-\alpha)$ se nazývá $1-\alpha$ kvantil rozdělení $N(0,1)$.

Po úpravě dostaneme $P\left(\overline{X} - \frac{\sigma}{\sqrt{n}} u_{1-\alpha} < \mu\right) = 1-\alpha$.

Z toho plyne, že $\left(\overline{X} - \frac{\sigma}{\sqrt{n}} u_{1-\alpha}, \infty\right)$ je dolní intervalový odhad parametru μ o spolehlivosti $1-\alpha$.

Podobně lze odvodit horní a oboustranný odhad střední hodnoty náhodné veličiny s normálním rozdělením pravděpodobnosti a známým rozptylem.

$\left(-\infty, \overline{X} + \frac{\sigma}{\sqrt{n}} u_{1-\alpha}\right)$ je horní intervalový odhad parametru μ.

$\left(\overline{X} - \frac{\sigma}{\sqrt{n}} u_{1-\alpha/2}, \overline{X} + \frac{\sigma}{\sqrt{n}} u_{1-\alpha/2}\right)$ je oboustranný intervalový odhad parametru μ.

Regrese

Model: $Y_i = f_\beta(x_i) + \epsilon_i$, $i = 1, \ldots, n$,

kde $x_1, \ldots, x_n \in \mathbb{R}$, $Y_1, \ldots, Y_n$ jsou náhodné veličiny, $\epsilon_1, \ldots, \epsilon_n$ jsou nezávislé náhodné veličiny reprezentující chybu modelu, $E\epsilon_i = 0$, $var\ \epsilon_i = \sigma^2 < \infty$, $\beta = (\beta_0, \ldots, \beta_k)$ je vektor parametrů.

Cíl: Pro daná data (x_i, y_i), $i = 1, \ldots, n$, určit odhad $\hat{\beta}$ parametru β. Víme, že $\beta \in M$.

Definujme reziduální součet čtverců $S_\beta = \sum_{i=1}^n (y_i - f_\beta(x_i))^2$

Takové $\hat{\beta}$, pro které S_β nabývá minima na množině M, se nazývá odhad parametru β metodou nejmenších čtverců.

Metody výpočtu $\hat{\beta}$:

1. Minimalizujeme funkci S_β, tj. pomocí prvních derivací $\frac{\partial S}{\partial \beta_i}$ určíme stacionární body S_β a pomocí druhých derivací rozhodneme, zda v nich S_β nabývá minima.
2. Je-li f_β lineární, řešíme soustavu lineárních algebraických rovnic $y_i = f_\beta(x_i)$, $i = 1, \ldots, n$, ve smyslu nejmenších čtverců.

Příklady regresních modelů:

1. Lineární regrese

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i,\ i = 1, \ldots, n$$

2. Polynomiální regrese

$$Y_i = \beta_0 + \beta_1 x_i + \ldots + \beta_k x_i^k + \epsilon_i,\ i = 1, \ldots, n$$

3. Vícerozměrná lineární regrese

$$Y_i = \beta_1 x_{i1} + \beta_2 x_{i2} \ldots + \beta_k x_{ik} + \epsilon_i,\ i = 1, \ldots, n$$

4. Nelineární model

je například model

$Y_i = \beta_0 e^{\beta_1 x_i} + \epsilon_i$, $i = 1, \ldots, n$.

Po zlogaritmování dostaneme

$ln\ Y_i \approx ln\ \beta_0 + \beta_1 x_i$, $i = 1, \ldots, n$,

což lze zapsat jako lineární model

$\tilde{Y}_i \approx \tilde{\beta}_0 + \beta_1 x_i$, $i = 1, \ldots, n$, kde $\tilde{Y}_i = ln\ Y_i$, $\tilde{\beta}_0 = ln\ \beta_0$.

Nevýhodou je, že nyní náhodná veličina reprezentující chybu nesplňuje předpoklady.