9. prednáška

Intervaly spoľahlivosti pre parametre normálneho rozdelenia

Nech $X_1, X_2, \ldots, X_n$ je náhodný výber zo základného súboru, ktorý má normálne rozdelenie pravdepodobnosti $N(m, \sigma^2)$. Oba parametre rozdelenia sú neznáme, na ich odhad použijeme metódy bodového odhadu. Neznámy parameter m odhadneme výberovým aritmetickým priemerom a neznámy parameter σ^2 odhadneme výberovým rozptylom:

$$\widehat{m} = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad \widehat{\sigma^2} = S_{n-1}^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$$

Ukázali sme, že ak základný súbor má normálne rozdelenie $N(m, \sigma^2)$, tak $\overline{X} \sim N\left(m, \dfrac{\sigma^2}{n}\right)$ a uveríme, že $\chi^2 = \dfrac{n-1}{\sigma^2} S_{n-1}^2$ má približne χ^2–rozdelenie s *(n-1)* stupňami voľnosti: $\chi^2 \sim \chi^2(n-1)$.

Interval spoľahlivosti pre m, ak je rozptyl σ^2 známy

Ako sme už povedali, bodovým odhadom parametra m je výberový aritmetický priemer. Vytvoríme náhodnú premennú Z ako funkciu neznámeho parametra m a jeho odhadu $\widehat{m}$:

$$Z = \frac{\overline{X} - m}{\sigma}\sqrt{n} \sim N(0, 1)$$

Kritické hodnoty z_1, z_2 normálneho rozdelenia nájdeme tak, aby platilo:

$$P(z_1 < Z < z_2) = 1 - \alpha.$$

Vzhľadom na to, že normálne rozdelenie je symetrické, platí $z_1 = -z_2$. A v tejto chvíli urobíme ešte jedno preznačenie – označíme $z_2 = z_\alpha$, takže dostávame vzťahy:

$$P(-z_\alpha < Z < z_\alpha) = 1 - \alpha.$$

$$\Phi(z_\alpha) - \Phi(-z_\alpha) = 1 - \alpha$$

$$2\Phi(z_\alpha) - 1 = 1 - \alpha$$

$$\Phi(z_\alpha) = 1 - \frac{\alpha}{2} \Rightarrow z_\alpha = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$$

Interval spoľahlivosti nájdeme úpravou vzťahu

$$P\left(-z_\alpha < \frac{\overline{X} - m}{\sigma}\sqrt{n} < z_\alpha\right) = 1 - \alpha$$

na konečný tvar

$$P\left(\overline{X} - z_\alpha \frac{\sigma}{\sqrt{n}} < m < \overline{X} + z_\alpha \frac{\sigma}{\sqrt{n}}\right)$$

a odtiaľ $\quad I_{1-\alpha} = \left(\overline{X} - z_\alpha \frac{\sigma}{\sqrt{n}},\ \overline{X} + z_\alpha \frac{\sigma}{\sqrt{n}} \right),$ kde $\quad \Phi(z_\alpha) = 1 - \frac{\alpha}{2}.$

Výraz $z_\alpha \frac{\sigma}{\sqrt{n}}$ predstavuje maximálnu chybu, ktorej sa dopúšťame s pravdepodobnosťou $1-\alpha$, keď neznámy parameter m aproximujeme pomocou výberového aritmetického priemeru $\overline{X}$.

Obdobnými úvahami odvodíme aj jednostranné intervaly spoľahlivosti.

Ľavostranný interval spoľahlivosti:

$$I_{1-\alpha} = \left(\overline{X} - z_\alpha \frac{\sigma}{\sqrt{n}},\ \infty \right), \quad \text{kde} \quad \Phi(z_\alpha) = 1 - \alpha.$$

Pravostranný interval spoľahlivosti:

$$I_{1-\alpha} = \left(-\infty,\ \overline{X} + z_\alpha \frac{\sigma}{\sqrt{n}} \right), \quad \text{kde} \quad \Phi(z_\alpha) = 1 - \alpha.$$

Interval spoľahlivosti pre *m*, ak je rozptyl σ^2 neznámy

V tomto prípade budeme uvažovať tak bodový odhad neznámeho parametra m, ako aj neznámeho parametra σ^2. Ako sme už povedali, $\overline{X} \sim N\left(m, \frac{\sigma^2}{n}\right)$ a $\chi^2 = \frac{n-1}{\sigma^2} S_{n-1}^2$ má približne χ^2 rozdelenie s *(n-1)* stupňami voľnosti, čo zapíšeme: $\chi^2 \sim \chi^2(n-1)$.

Z predchádzajúcich prednášok vieme, že ak $X \sim N(0,1)$ a $Y \sim \chi^2(k)$, tak $t = \frac{X}{\sqrt{\frac{Y}{k}}}$ má Studentovo rozdelenie o k stupňoch voľnosti. Pre naše potreby skonštruujeme takú premennú:

Dostali sme tak
$$t = \frac{\overline{X} - m}{S_{n-1}}\sqrt{n} \sim t(n-1)$$
Studentovo rozdelenie je symetrické rozdelenie, tak $t_1 = -t_2$ a preznačíme $t_2 = t_\alpha$, takže dostávame vzťahy:
$$P(-t_\alpha < t < t_\alpha) = 1 - \alpha.$$
$$F_{t(n-1)}(t_\alpha) - F_{t(n-1)}(-t_\alpha) = 1 - \alpha$$
$$2F_{t(n-1)}(t_\alpha) - 1 = 1 - \alpha$$
$$F_{t(n-1)}(t_\alpha) = 1 - \frac{\alpha}{2} \Rightarrow t_\alpha = F^{-1}_{t(n-1)}\left(1 - \frac{\alpha}{2}\right)$$
Interval spoľahlivosti nájdeme úpravou vzťahu
$$P\left(-t_\alpha < \frac{\overline{X} - m}{S_{n-1}}\sqrt{n} < t_\alpha\right) = 1 - \alpha$$
na konečný tvar
$$I_{1-\alpha} = \left(\overline{X} - t_\alpha \frac{S_{n-1}}{\sqrt{n}},\ \overline{X} + t_\alpha \frac{S_{n-1}}{\sqrt{n}}\right),$$
kde $\quad F_{t(n-1)}(t_\alpha) = 1 - \frac{\alpha}{2}.$

Obdobnými úvahami odvodíme aj jednostranné intervaly spoľahlivosti.

Ľavostranný interval spoľahlivosti:

$$I_{1-\alpha} = \left(\overline{X} - t_\alpha \frac{S_{n-1}}{\sqrt{n}},\ \infty\right), \text{ kde } \quad F_{t(n-1)}(t_\alpha) = 1 - \alpha.$$

Pravostranný interval spoľahlivosti:

$$I_{1-\alpha} = \left(-\infty,\ \overline{X} + t_\alpha \frac{S_{n-1}}{\sqrt{n}}\right), \text{ kde } \quad F_{t(n-1)}(t_\alpha) = 1 - \alpha.$$

Poznamenajme, že pre dostatočne veľký rozsah náhodného výberu ($n > 30$) počítame intervaly spoľahlivosti pre neznámu strednú hodnotu aj keď rozptyl nepoznáme, ako keby sme ho poznali – pracujeme s kritickými hodnotami normálneho normovaného rozdelenia.

Interval spoľahlivosti pre neznámy rozptyl σ^2

Za odhad parametra σ^2 použijeme S_{n-1}^2 . Ak predpokladáme, že náhodný výber pochádza zo základného súboru s normálnym rozdelením, tak vieme ☺, že $\chi^2 = \dfrac{(n-1)S_{n-1}^2}{\sigma^2} \sim \chi^2(n-1)$. Rozdelenie χ^2 nie je symetrické, preto kritické hodnoty χ_1, χ_2 hľadáme inak ako v predchádzajúcich prípadoch.

Obojstranný interval spoľahlivosti nájdeme úpravou vzťahu

$$P\left(\chi_1 < \frac{(n-1)S_{n-1}^2}{\sigma^2} < \chi_2\right) = 1 - \alpha$$

Konečný tvar obojstanného intervalu je:

$$I_{1-\alpha} = \left(\frac{(n-1)S_{n-1}^2}{\chi_2}, \frac{(n-1)S_{n-1}^2}{\chi_1} \right),$$

kde $\quad F_{\chi^2(n-1)}(\chi_1) = \frac{\alpha}{2}, \quad F_{\chi^2(n-1)}(\chi_2) = 1 - \frac{\alpha}{2}.$

Ľavostranný interval spoľahlivosti:

Úpravou vzťahu

$$P\left(\frac{(n-1)S_{n-1}^2}{\sigma^2} < \chi_2 \right) = 1 - \alpha$$

dostávame $I_{1-\alpha} = \left(\frac{(n-1)S_{n-1}^2}{\chi_2}; \infty \right)$, kde $F_{\chi^2(n-1)}(\chi_2) = 1 - \alpha.$

Pravostranný interval spoľahlivosti:

$$I_{1-\alpha} = \left(0;\ \frac{(n-1)S_{n-1}^2}{\chi_1} \right), \text{ kde } F_{\chi^2(n-1)}(\chi_1) = \alpha.$$

Intervaly spoľahlivosti pre odhad smerodajnej odchýlky dostaneme z intervalov spoľahlivosti pre rozptyl odmocnením hraníc intervalu.

Poznámka k intervalom spoľahlivosti pre neznámu strednú hodnotu:

Výskum ukázal, že interval spoľahlivosti

$$I_{1-\alpha} = \left(\overline{X} - t_\alpha \frac{S_{n-1}}{\sqrt{n}},\ \overline{X} + t_\alpha \frac{S_{n-1}}{\sqrt{n}} \right)$$

môžeme použiť, aj keď základný súbor nemá normálne rozdelenie. No interval spoľahlivosti je len približný. Kvalita aproximácie závisí od rozdelenia základného súboru a od rozsahu náhodného výberu. Aproximácia je väčšinou dobrá pre $n \geq 30$. Keď je rozdelenie základného súboru výrazne zošikmené alebo obsahuje odľahlé hodnoty, odporúča sa rozsah výberu aspoň 50. Keď rozdelenie základného súboru nie je normálne, ale je symetrické, môže sa dobrá aproximácia získať už pri výbere o rozsahu 15. ☺