Cvičení ke 12. sadě

Vzorečky Náhodný výběr n hodnot $X_1, \ldots, X_n$ náhodné veličiny X. Předpokládáme, že náhodná veličina X má střední hodnotu $E(X) = \mu$, rozptyl σ^2.

Řešení spočívá v převedení X na transformovanou normovanou náhodnou veličinu Z, která má standardizované rozložení se střední hodnotou 0. Podle toho, jestli známe rozptyl σ^2 (resp. jeho odmocninu σ = směrodatnou odchylku), nebo ne, jde o normální rozdělení, nebo o Studentovo rozdělení se stupněm volnosti $n-1$. Podrobnosti tohoto převodu (= teorie) je v materiálech o statistice (*matikaIV_statistika13_hypotezy.zip*). Nyní k praktickým výpočtům:

Počítáme $k\%$ interval spolehlivosti pro skutečnou hodnotu náhodné veličiny X. To znamená interval, ve kterém se náhodná veličina na $k\%$ vyskytuje. Předpokládáme, že je distribuční funkce Φ_X symetrická podle přímky $x = E(X)$. Takže když najdeme interval, který je umístěný symetricky podle této přímky a pravděpodobnost, že X padne vlevo od něj je $(100 - \frac{k}{2})\%$, víme, že pravdepodobnost, že X padne vpravo od něj je také $(100 - \frac{k}{2})\%$, a X padne do intervalu s pravděpodobností $k\%$.

K počítání intervalu spolehlivosti potřebujeme:

$$M = \frac{1}{n}(X_1 + \ldots + X_n)$$

Když neznáme rozptyl měření, tak místo σ^2 vezmeme výběrový rozptyl S^2 spočítaný z výběru takto:

$$S^2 = \frac{1}{n-1}\left((X_1 - M)^2 + \ldots + (X_n - M)^2\right)$$

Vzoreček pro $k\%$ interval spolehlivosti (D, H) (D - dolní mez, H - horní mez, označíme riziko $\alpha = 1 - \frac{k}{100}$):

- Pokud známe σ:

  $$D = M - \frac{\sigma}{\sqrt{n}} \cdot u_{1-\frac{\alpha}{2}}$$
  $$H = M + \frac{\sigma}{\sqrt{n}} \cdot u_{1-\frac{\alpha}{2}},$$

  kde $u_{1-\frac{\alpha}{2}}$ je takové číslo $u \in (0,1)$, pro které je hodnota distribuční funkce normálního standardizovaného rozložení $\Phi(u) = 1 - \frac{\alpha}{2}$. Toto číslo najdeme v tabulkách pro normální rozdělení.

- Pokud neznáme σ:

  Pokud neznáme σ, použijeme místo toho S a místo normálního rozdělení použijeme Studentovo rozdělení se stupněm volnosti $n - 1$.

  $$D = M - \frac{S}{\sqrt{n}} \cdot t(n-1)_{1-\frac{\alpha}{2}}$$
  $$H = M + \frac{S}{\sqrt{n}} \cdot t(n-1)_{1-\frac{\alpha}{2}},$$

  kde $t(n-1)_{1-\frac{\alpha}{2}}$ je takové číslo $t \in (0,1)$, pro které je hodnota distribuční funkce Studentova rozložení se stupněm volnosti $n-1$ rovna $\Phi(u) = 1 - \frac{\alpha}{2}$. Toto číslo najdeme v tabulkách pro Studentovo rozložení.

Jak hledat v tabulkách, které máte v ISu

- **Normální rozdělení:**

  V názvech řádků jsou desetinná místa z, v názvech sloupců setiny z, uvnitř tabulky hodnoty $P(0 \leq Z \leq z) = \Phi(z) - \frac{1}{2}$.

- **Studentovo rozdělení:** V i.tém řádku jsou absolutní hodnoty kvantilů ($|K_{0,1}| = -K_{0,1}$, $|K_{0,05}| = -K_{0,05}$, $\ldots$) Studentova rozdělení s i stupni volnosti. Aneb hodnoty $-\Phi^{-1}(0,1), \ldots$. Důvod toho mínusu je to, že Studentovo rozložení je uvažováno se střední hodnotou v nule a jde symetricky do záporných a kladných čísel. (Kdyby tedy byly v tabulce uvedeny přímo kvantily, musely by být napsané včetně minusu).

Příklady

1. Máme zadané hodnoty pěti měření (neboli výběr) vlhkosti vzduchu v místnosti $70\%, 77\%, 71\%, 75\%, 72\%$. Rozptyl měření je $3\%^2$ (směrodatná odchylka je zadaná v %). Úkolem je určit 95% interval spolehlivosti skutečné vlhkosti vzduchu v místnosti.

   Řešení:

   (Procenta jsou jen jednotky: $70j, \ldots, 72j$. Hledáme 95% interval spolehlivosti. $\alpha = 1 - \frac{95}{100} = 0,05$.) Spočítáme si M

   $$M = \frac{1}{5}(0,70 + 0,77 + 0,71 + 0,75 + 0,72) = 0,73$$

   Je zadáno $\sigma = \sqrt{3j^2} = 1,732j$.

   Potřebujeme ještě najít v tabulce takové $u_{1-\frac{0,05}{2}}$, že pro distribuční funkci standardizovaného normálního rozdělení platí $\Phi(u_{1-\frac{0,05}{2}}) = 1 - \frac{0,05}{2} = 0,975$.

   Hledáme uvnitř tabulky číslo $0,975 - 0,5 = 0,475$. To se nachází ve sloupci označeném $0,06$ a řádku označeném $1,9$. Součet těchto dvou čísel je $1,96$ ($\Phi(1,96) = 0,975$).

   Všechno to dosadíme do vzorečku pro D a H:

   $$D = M - \frac{\sigma}{\sqrt{5}} \cdot u_{1-\frac{0,05}{2}} =$$

   $$= 73 - \frac{1,732}{\sqrt{5}} \cdot 1,96 = 73 - 1,5 = 71,5$$

   $$H = 73 + 1,5 = 74,5$$

   95% interval spolehlivosti je $(71,5; 74,5)$.

2. Teď ten stejný úkol, ale pro neznámý rozptyl měření.

   Řešení:

   Postupujeme stejně, jen místo σ budeme dosazovat S a budeme hledat v tabulce pro Studentovo rozdělení se stupněm volnosti $(n-1) = 5 - 1 = 4$.

To S si musíme spočítat:

$$S^2 = \frac{1}{5-1}\left((70-73)^2 + (77-73)^2 + (71-73)^2 + (75-73)^2 + (72-73)^2\right) =$$

$$= \frac{34}{4} = 8,5$$

$$S = \sqrt{8,5} = 2,915\ldots$$

V tabulce pro Studentovo rozdělení chceme najít hodnotu $t(n-1)_{1-\frac{\alpha}{2}} = t(4)_{1-\frac{0,05}{2}}$. Pro stupeň volnosti 4 hledáme na řádku označeném čtyřkou.

V tabulce uvnitř jsou čísla z, pro která je pravděpodobnost jednoho nebo dvou ocásků taková, jaký je popis sloupečku v prvním nebo druhém řádku tabulky.

Jinými slovy např. pro stupeň volnosti 1 je pravděpodobnost, že se náhodná veličina se Studentovým rozdělením stupně volnosti 1 nachází mimo interval $(-3,078; 3.078)$ - neboli v některém ze dvou ocásků - rovna $0,20$.

Pravděpodobnost, že se náhodná veličina se Studentovým rozdělením stupně volnosti 1 nachází jen vpravo (dalo by se také říci jen vlevo, je to symetrické) od intervalu $(-3,078; 3.078)$ - neboli v jednom ze dvou ocásků - rovna $0,10$.

Hledáme v řádku označeném stupněm volnosti 4 hodnotu, pro kterou je pravděpodobnost dvou ocásků (two tails) - neboli pravděpodobnost, že je náhodná veličina mimo interval $(-hodnota, +hodnota)$ rovna $\alpha = 0,05$ (nebo pravděpodobnost jednoho ocásku rovna $\frac{\alpha}{2} = 0,025$). To je právě hledaná hodnota $t(4)_{1-0,025}$.

Našli jsme třetí sloupec $t(4)_{1-0,025} = 2,776$.

Všechno to dosadíme do vzorečku pro D a H:

$$D = M - \frac{S}{\sqrt{5}} \cdot t(4)_{1-\frac{\alpha}{2}} =$$

$$= 73 - \frac{2,915}{\sqrt{5}} \cdot 2,776 = 73 - 3,62 = 69,38$$

$$H = 73 + 3,62 = 76,62$$

95% interval spolehlivosti pro neznámý rozptyl je $(69,38; 76,62)$.