7. Parametrické úlohy o dvou nezávislých náhodných výběrech z normálních rozložení

7.1. Motivace: V této situaci je naším úkolem porovnat střední hodnoty či rozptyly dvou normálních rozložení na základě znalosti dvou nezávislých náhodných výběrů pořízených z těchto rozložení. Zpravidla konstruujeme intervaly spolehlivosti pro rozdíl středních hodnot respektive hodnotíme shodu středních hodnot pomocí dvouvýběrového t-testu či dvouvýběrového z-testu a shodu rozptylů pomocí F-testu.

7.2. Věta: Rozložení statistik odvozených z výběrových průměrů a výběrových rozptylů

7.3. Věta: Vzorce pro meze 100(1-α)% empirických intervalů spolehlivosti pro parametrické funkce $\mu_1 - \mu_2$ a σ_1^2/σ_2^2

a) Interval spolehlivosti pro $\mu_1 - \mu_2$, když σ_1^2, σ_2^2 známe

b) Interval spolehlivosti pro $\mu_1 - \mu_2$, když σ_1^2, σ_2^2 neznáme, ale víme, že jsou shodné

c) Interval spolehlivosti pro společný neznámý rozptyl σ^2

d) Interval spolehlivosti pro podíl rozptylů σ_1^2/σ_2^2

Upozornění: Není-li v 7.3. (b) splněn předpoklad o shodě rozptylů, lze sestrojit aspoň přibližný 100(1-α)% interval spolehlivosti pro $\mu_1 - \mu_2$. V tomto případě má statistika T přibližně rozložení t(ν), kde počet stupňů volnosti $\nu = \dfrac{\left(s_1^2/n_1 + s_2^2/n_2\right)^2}{\dfrac{\left(s_1^2/n_1\right)^2}{n_1-1} + \dfrac{\left(s_2^2/n_2\right)^2}{n_2-1}}$. Není-li ν celé číslo, použijeme v tabulkách kvantilů Studentova rozložení lineární interpolaci.

7.4. Příklad: Ve dvou nádržích se zkoumal obsah chlóru (v g/l). Z první nádrže bylo odebráno 25 vzorků, z druhé nádrže 10 vzorků. Byly vypočteny realizace výběrových průměrů a rozptylů: $m_1 = 34{,}48$, $m_2 = 35{,}59$, $s_1^2 = 1{,}7482$, $s_2^2 = 1{,}7121$. Hodnoty zjištěné z odebraných vzorků považujeme za realizace dvou nezávislých náhodných výběrů z rozložení $N(\mu_1, \sigma^2)$ a $N(\mu_2, \sigma^2)$. Sestrojte 95% empirický interval spolehlivosti pro rozdíl středních hodnot $\mu_1 - \mu_2$.

Řešení:

Úloha vede na vzorec 7.3. (b). Vypočteme vážený průměr výběrových rozptylů a najdeme odpovídající kvantily Studentova rozložení:

$$s_*^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2} = \frac{24 \cdot 1{,}7482 + 9 \cdot 1{,}7121}{33} = 1{,}7384,\ t_{0,975}(33) = 2{,}035.$$

Dosadíme do vzorců pro dolní a horní mez intervalu spolehlivosti:

$$d = m_1 - m_2 - s_* \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\, t_{1-\alpha/2}(n_1+n_2-2) = 34{,}48 - 35{,}59 - \sqrt{1{,}7384} \cdot \sqrt{\frac{1}{25} + \frac{1}{10}} \cdot 2{,}035 = -2{,}114$$

$$h = m_1 - m_2 + s_* \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\, t_{1-\alpha/2}(n_1+n_2-2) = 34{,}48 - 35{,}59 + \sqrt{1{,}7384} \cdot \sqrt{\frac{1}{25} + \frac{1}{10}} \cdot 2{,}035 = -0{,}106$$

Zjistili jsme, že $-2{,}114$ g/l $< \mu_1 - \mu_2 < -0{,}106$ g/l s pravděpodobností aspoň 0,95.

7.5. Příklad: V příkladu 7.4. nyní předpokládáme, že dané dva náhodné výběry pocházejí z rozložení $N(\mu_1, \sigma_1^2)$ a $N(\mu_2, \sigma_2^2)$. Sestrojte 95% empirický interval spolehlivosti pro podíl rozptylů.

Řešení:
Úloha vede na vzorec 7.3. (d).

$$d = \frac{s_1^2/s_2^2}{F_{1-\alpha/2}(n_1-1, n_2-1)} = \frac{1{,}7482/1{,}7121}{F_{0,975}(24{,}9)} = \frac{1{,}7482/1{,}7121}{3{,}6142} = 0{,}28$$

$$h = \frac{s_1^2/s_2^2}{F_{\alpha/2}(n_1-1, n_2-1)} = \frac{1{,}7482/1{,}7121}{F_{0,025}(24{,}9)} = \frac{1{,}7482/1{,}7121}{1/F_{0,975}(9{,}24)} = \frac{1{,}7482/1{,}7121}{1/2{,}7027} = 2{,}76$$

Dostáváme, že $0{,}28 < \frac{\sigma_1^2}{\sigma_2^2} < 2{,}76$ s pravděpodobností aspoň 0,95.

7.6. Definice: Jednotlivé typy testů pro parametrické funkce $\mu_1 - \mu_2$ a σ_1^2/σ_2^2
a) Dvouvýběrový z-test
b) Dvouvýběrový t-test
c) F-test

7.7. Příklad: V restauraci "U bílého koníčka" měřili ve 20 případech čas obsluhy zákazníka. Výsledky v minutách: 6, 8, 11, 4, 7, 6, 10, 6, 9, 8, 5, 12, 13, 10, 9, 8, 7, 11, 10, 5. V restauraci "Zlatý lev" bylo dané pozorování uskutečněno v 15 případech s těmito výsledky: 9, 11, 10, 7, 6, 4, 8, 13, 5, 15, 8, 5, 6, 8 ,7. Za předpokladu, že uvedené hodnoty pocházejí ze dvou normálních rozložení, na hladině významnosti 0,05 testujte hypotézu, že střední hodnoty doby obsluhy jsou v obou restauracích stejné.
Řešení:
Na hladině významnosti 0,05 testujeme nulovou hypotézu H_0: $\mu_1 - \mu_2 = 0$ proti oboustranné alternativě H_1: $\mu_1 - \mu_2 \neq 0$. Je to úloha na dvouvýběrový t-test. Před provedením tohoto testu je však nutné pomocí F-testu ověřit shodu rozptylů. Na hladině významnosti 0,05 tedy testujeme H_0: $\frac{\sigma_1^2}{\sigma_2^2} = 1$ proti H_1: $\frac{\sigma_1^2}{\sigma_2^2} \neq 1$. Nejprve vypočteme $m_1 = 8{,}25$, $m_2 = 8{,}13$, $s_1^2 = 6{,}307$, $s_2^2 = 9{,}41$, $s_*^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2} = \frac{19\cdot 6{,}307 + 14\cdot 9{,}41}{33} = 7{,}623$. Podle 7.6. (c) vypočteme realizaci testové statistiky: $t_0 = \frac{s_1^2}{s_2^2} = \frac{6{,}307}{9{,}41} = 0{,}6702$. Stanovíme kritický obor:

$$W = \langle 0, F_{\alpha/2}(n_1-1, n_2-1)\rangle \cup \langle F_{1-\alpha/2}(n_1-1, n_2-1), \infty) = \langle 0, F_{0,025}(19{,}14)\rangle \cup \langle F_{0,925}(19{,}14), \infty) =$$
$$= \langle 0{,}1/F_{0,925}(14{,}19)\rangle \cup \langle F_{0,925}(19{,}14), \infty) = \langle 0{,}1/2{,}649\rangle \cup \langle 2{,}8607, \infty) = \langle 0;0{,}3778\rangle \cup \langle\langle 2{,}8607, \infty), \infty)$$

Protože se testová statistika nerealizuje v kritickém oboru, nulovou hypotézu nezamítáme na hladině významnosti 0,05. Rozptyly tedy můžeme považovat za shodné.
Nyní se vrátíme k dvouvýběrovému t-testu. Podle 7.6. (b) vypočteme realizaci testové statistiky: $t_0 = \frac{m_1 - m_2 - c}{s_*\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{8{,}25 - 8{,}13}{\sqrt{7{,}623}\sqrt{\frac{1}{20}+\frac{1}{15}}} = 0{,}124$. Stanovíme kritický obor:

$$W = (-\infty, -t_{1-\alpha/2}(n_1+n_2-2)\rangle \cup \langle t_{1-\alpha/2}(n_1+n_2-2), \infty) = (-\infty, -t_{0,975}(33)\rangle \cup \langle t_{0,975}(33), \infty) =$$
$$= (-\infty, -2{,}035\rangle \cup \langle 2{,}035, \infty)$$

Protože testová statistika se nerealizuje v kritickém oboru, nulovou hypotézu nezamítáme na hladině významnosti 0,05.