Kapitola 5.: Úlohy o dvou nezávislých náhodných výběrech z normálních rozložení

5.1. Motivace

V tomto případě 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 nebo podíl rozptylů respektive hodnotíme shodu středních hodnot pomocí dvouvýběrového t-testu a shodu rozptylů pomocí F-testu.

5.2. Rozložení statistik odvozených z výběrových průměru a výběrových rozptylu

Nechť $X_{11},\dots,X_{n_1 1}$ je náhodný výběr z rozložení $N(\mu_1, \sigma_1^2)$ a $X_{12},\dots,X_{n_2 2}$ je na něm nezávislý náhodný výběr z rozložení $N(\mu_2, \sigma_2^2)$, přičemž $n_1 \geq 2$ a $n_2 \geq 2$. Označme M_1, M_2 výběrové průměry a S_1^2, S_2^2 výběrové rozptyly. Pak platí:

a) Statistiky $M_1 - M_2$ a $S_*^2 = \dfrac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}$ jsou stochasticky nezávislé.

b) $M_1 - M_2 \sim N\left(\mu_1 - \mu_2, \dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}\right)$, tedy $U = \dfrac{(M_1 - M_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} \sim N(0, 1)$.

(Pivotová statistika U slouží k řešení úloh o $\mu_1 - \mu_2$, když σ_1^2 a σ_2^2 známe.)

c) Jestliže $\sigma_1^2 = \sigma_2^2 =: \sigma^2$, pak $K = \dfrac{(n_1+n_2-2)S_*^2}{\sigma^2} \sim \chi^2(n_1 + n_2 - 2)$.

(Pivotová statistika K slouží k řešení úloh o neznámém společném rozptylu σ^2.)

d) Jestliže $\sigma_1^2 = \sigma_2^2 =: \sigma^2$, pak $T = \dfrac{(M_1 - M_2) - (\mu_1 - \mu_2)}{S_* \sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} \sim t(n_1 + n_2 - 2)$.

(Pivotová statistika T slouží k řešení úloh o $\mu_1 - \mu_2$, když σ_1^2 a σ_2^2 neznáme, ale víme, že jsou shodné.)

e) $F = \dfrac{S_1^2 / S_2^2}{\sigma_1^2 / \sigma_2^2} \sim F(n_1 - 1, n_2 - 1)$.

(Pivotová statistika F slouží k řešení úloh o σ_1^2 / σ_2^2.)

5.2.1. Příklad

Nechť jsou dány dva nezávislé náhodné výběry, první pochází z rozložení N(2, 1,5) a má rozsah 10, druhý pochází z rozložení N(3, 4) a má rozsah 5. Jaká je pravděpodobnost, že výběrový průměr 1. výběru bude menší než výběrový průměr 2. výběru?

Řešení: $P(M_1 < M_2) = P(M_1 - M_2 < 0) =$

$$P\left(\frac{(M_1 - M_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} < \frac{0 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}\right) = P\left(U < \frac{-2+3}{\sqrt{\frac{1,5}{10} + \frac{4}{5}}}\right) = P(U < 1,026) = \Phi(1,026) = 0,8475.$$

S pravděpodobností přibližně 84,8% je výběrový průměr 1. výběru menší než výběrový průměr 2. výběru.

5.3. Intervaly spolehlivosti pro parametrické funkce $\mu_1 - \mu_2$, σ_1^2 / σ_2^2

Budeme zabývat speciálními případy, kdy za parametrickou funkci $h(\vartheta)$ považujeme rozdíl středních hodnot $\mu_1 - \mu_2$ nebo podíl rozptylů σ_1^2 / σ_2^2 dvou normálních rozložení. Při konstrukci intervalu spolehlivosti pro rozdíl středních hodnot buď rozptyly známe nebo neznáme a víme, že jsou shodné či nikoliv. Shodu rozptylů ověřujeme pomocí F-testu. Uvedeme jen přehled vzorců pro meze $100(1-\alpha)\%$ empirických intervalů spolehlivosti pro parametrické funkce $\mu_1 - \mu_2$, σ_1^2 / σ_2^2.

5.3.1. Přehled vzorců

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

(využití pivotové statistiky $U = \dfrac{(M_1 - M_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0, 1)$)

Oboustranný: $(d, h) = \left(m_1 - m_2 - \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\, u_{1-\alpha/2},\ m_1 - m_2 + \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\, u_{1-\alpha/2}\right)$

Levostranný: $(d, \infty) = \left(m_1 - m_2 - \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\, u_{1-\alpha},\ \infty\right)$

Pravostranný: $(-\infty, h) = \left(-\infty, m_1 - m_2 + \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\, u_{1-\alpha}\right)$

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

(využití pivotové statistiky $T = \dfrac{(M_1 - M_2) - (\mu_1 - \mu_2)}{S_* \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t(n_1 + n_2 - 2)$)

Oboustranný:

$(d, h) = \left(m_1 - m_2 - s_* \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\, t_{1-\alpha/2}(n_1+n_2-2),\ m_1 - m_2 + s_* \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\, t_{1-\alpha/2}(n_1+n_2-2)\right)$

Levostranný: $(d, \infty) = \left(m_1 - m_2 - s_* \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\, t_{1-\alpha}(n_1+n_2-2),\ \infty\right)$

Pravostranný: $(-\infty, h) = \left(-\infty,\ m_1 - m_2 + s_* \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\, t_{1-\alpha}(n_1+n_2-2)\right)$

c) Interval spolehlivosti pro společný neznámý rozptyl σ^2

(využití pivotové statistiky $K = \frac{(n_1 + n_2 - 2)S_*^2}{\sigma^2} \sim \chi^2(n_1 + n_2 - 2)$)

Oboustranný: $(d, h) = \left(\frac{(n_1 + n_2 - 2)s_*^2}{\chi^2_{1-\alpha/2}(n_1 + n_2 - 2)}, \frac{(n_1 + n_2 - 2)s_*^2}{\chi^2_{\alpha/2}(n_1 + n_2 - 2)} \right)$

Levostranný: $(d, \infty) = \left(\frac{(n_1 + n_2 - 2)s_*^2}{\chi^2_{1-\alpha}(n_1 + n_2 - 2)}, \infty \right)$

Pravostranný: $(-\infty, h) = \left(-\infty, \frac{(n_1 + n_2 - 2)s_*^2}{\chi^2_{\alpha}(n_1 + n_2 - 2)} \right)$

d) Interval spolehlivosti pro podíl rozptylů $\frac{\sigma_1^2}{\sigma_2^2}$

(využití pivotové statistiky $F = \frac{S_1^2 / S_2^2}{\sigma_1^2 / \sigma_2^2} \sim F(n_1 - 1, n_2 - 1)$)

Oboustranný: $(d, h) = \left(\frac{s_1^2 / s_2^2}{F_{1-\alpha/2}(n_1 - 1, n_2 - 1)}, \frac{s_1^2 / s_2^2}{F_{\alpha/2}(n_1 - 1, n_2 - 1)} \right)$

Levostranný: $(d, \infty) = \left(\frac{s_1^2 / s_2^2}{F_{1-\alpha}(n_1 - 1, n_2 - 1)}, \infty \right)$

Pravostranný: $(-\infty, h) = \left(-\infty, \frac{s_1^2 / s_2^2}{F_{\alpha}(n_1 - 1, n_2 - 1)} \right)$

Upozornění: Není-li v 5.3.1. (b) splněn předpoklad o shodě rozptylů, lze sestrojit aspoň přibližný 100(1-α)% interval spolehlivosti pro μ_1-μ_2. V tomto případě má statistika T přibližně rozložení t(ν), kde počet stupňů volnosti $\nu = \frac{\left(s_1^2 / n_1 + s_2^2 / n_2\right)^2}{\frac{\left(s_1^2 / n_1\right)^2}{n_1 - 1} + \frac{\left(s_2^2 / n_2\right)^2}{n_2 - 1}}$. Není-li ν celé číslo, použijeme v tabulkách kvantilů Studentova rozložení lineární interpolaci.

5.3.2. 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 5.3.1. (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.

5.3.3. Příklad

V příklad 5.3.2. 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 5.3.1. (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.

5.4. Testování hypotéz o parametrických funkcích $\mu_1 - \mu_2$, σ_1^2/σ_2^2

5.4.1. Přehled testů

a) Nechť $X_{11},\ldots,X_{n_1 1}$ je náhodný výběr z rozložení $N(\mu_1, \sigma_1^2)$ a $X_{12},\ldots,X_{n_2 2}$ je na něm nezávislý náhodný výběr z rozložení $N(\mu_2, \sigma_2^2)$, přičemž $n_1 \geq 2$, $n_2 \geq 2$ a σ_1^2, σ_2^2 známe. Nechť
c je konstanta. Test H_0: $\mu_1 - \mu_2 = c$ proti H_1: $\mu_1 - \mu_2 \neq c$ se nazývá dvouvýběrový z-test.

b) Nechť $X_{11},\ldots,X_{n_1 1}$ je náhodný výběr z rozložení $N(\mu_1, \sigma^2)$ a $X_{12},\ldots,X_{n_2 2}$ je na něm nezávislý náhodný výběr rozložení $N(\mu_2, \sigma^2)$, přičemž $n_1 \geq 2$ a $n_2 \geq 2$ a σ^2 neznáme. Nechť c je konstanta. Test H_0: $\mu_1 - \mu_2 = c$ proti H_1: $\mu_1 - \mu_2 \neq c$ se nazývá dvouvýběrový t-test.

c) Nechť $X_{11},\ldots,X_{n_1 1}$ je náhodný výběr z rozložení $N(\mu_1, \sigma_1^2)$ a $X_{12},\ldots,X_{n_2 2}$ je na něm nezávislý náhodný výběr rozložení $N(\mu_2, \sigma_2^2)$, přičemž $n_1 \geq 2$ a $n_2 \geq 2$. Test H_0: $\frac{\sigma_1^2}{\sigma_2^2} = 1$ proti H_1: $\frac{\sigma_1^2}{\sigma_2^2} \neq 1$ se nazývá F-test.

5.4.2. Provedení testů o parametrickcýh funkcích $\mu_1 - \mu_2$, σ_1^2/σ_2^2 pomocí kritického oboru

a) Provedení dvouvýběrového z-testu

Hypotézu H_0: $\mu_1 - \mu_2 = c$ proti H_1: $\mu_1 - \mu_2 \neq c$ (resp. H_1: $\mu_1 - \mu_2 < c$ resp. H_1: $\mu_1 - \mu_2 > c$)

zamítáme na hladině významnosti α, jestliže $\left|\dfrac{m_1 - m_2 - c}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}\right| \geq u_{1-\alpha/2}$

(resp. $\dfrac{m_1 - m_2 - c}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} \leq -u_{1-\alpha}$ resp. $\dfrac{m_1 - m_2 - c}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} \geq u_{1-\alpha}$).

b) Provedení dvouvýběrového t-testu

Hypotézu H_0: $\mu_1 - \mu_2 = c$ proti H_1: $\mu_1 - \mu_2 \neq c$ (resp. H_1: $\mu_1 - \mu_2 < c$ resp. H_1: $\mu_1 - \mu_2 > c$)

zamítáme na hladině významnosti α, jestliže $\left|\dfrac{m_1 - m_2 - c}{s_*\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}\right| \geq t_{1-\alpha/2}(n_1 + n_2 - 2)$

(resp. $\dfrac{m_1 - m_2 - c}{s_*\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} \leq -t_{1-\alpha}(n_1 + n_2 - 2)$ resp. $\dfrac{m_1 - m_2 - c}{s_*\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} \geq t_{1-\alpha}(n_1 + n_2 - 2)$)

.

c) Provedení F-testu

Hypotézu H_0: $\dfrac{\sigma_1^2}{\sigma_2^2} = 1$ proti H_1: $\dfrac{\sigma_1^2}{\sigma_2^2} \neq 1$ (resp. H_1: $\dfrac{\sigma_1^2}{\sigma_2^2} < 1$ resp. H_1: $\dfrac{\sigma_1^2}{\sigma_2^2} > 1$) zamítáme

na hladině významnosti α, jestliže $\dfrac{s_1^2}{s_2^2} \leq F_{\alpha/2}(n_1+n_2-2)$ nebo $\dfrac{s_1^2}{s_2^2} \geq F_{1-\alpha/2}(n_1+n_2-2)$

(resp. $\dfrac{s_1^2}{s_2^2} \leq F_{\alpha}(n_1+n_2-2)$ resp. $\dfrac{s_1^2}{s_2^2} \geq F_{1-\alpha}(n_1+n_2-2)$).

Podobně jako v kapitole 4 musíme ověřit normalitu dat. Pokud výběry menších rozsahů (pod 30) vykazují výraznější odchylky od normality, doporučuje se místo dvouvýběrového t-testu použít neparametrický dvouvýběrový Wilcoxonů test (viz kapitola 7).

Před provedením dvouvýběrového t-testu bychom se měli F-testem přesvědčit o shodě rozptylů. Zamítne-li F-test na dané hladině významnosti hypotézu o shodě rozptylů, musíme pro testování hypotézy o shodě středních hodnot použít speciální variantu dvouvýběrového t-testu, tzv. dvouvýběrový t-test se separovanými odhdy rozptylů.

Musíme si být vědomi rozdílu mezi dvouvýběrovým t-testem a párovým t-testem. Dvouvýběrový t-test je založen na předpokladu nezávislosti daných dvou výběrů. Pokud v situaci, která vede na párový test, použijeme dvouvýběrový t-test, můžeme dostat nepravdivé výsledky. Naopak, mají-li dva nezávislé výběry stejný rozsah a my použijeme párový t-test místo dvouvýběrového t-testu, nedopustíme se hrubé chyby, pouze méně efektivně využijeme informaci obsaženou v datech.

5.4.3. Příklad

Výrobce limonád chtěl zjistit, zda změna technologie výroby se projeví v prodeji limonád. Proto sledoval po 14 náhodně vybraných dnů před zavedením nových limonád tržby v určitém regionu a zjistil, že za den utržil v průměru 39 600 Kč se směrodatnou odchylkou 5 060 Kč. Po zavedení nových limonád prověřil stejným způsobem tržby v 11 náhodně vybraných dnech v témž regionu a zjistil průměrný příjem 41 200 Kč se směrodatnou odchylkou 4 310 Kč. Předpokládejte, že tržby za starý typ limonád se řídí rozložením $N(\mu_1, \sigma_1^2)$ a tržby za nový typ limonád se řídí rozložením $N(\mu_2, \sigma_2^2)$.

a) Na hladině významnosti 0,05 testujte hypotézu H_0: $\dfrac{\sigma_1^2}{\sigma_2^2} = 1$ proti H_1: $\dfrac{\sigma_1^2}{\sigma_2^2} \neq 1$.

b) Na hladině významnosti 0,05 testujte hypotézu H_0: $\mu_1 - \mu_2 = 0$ proti H_1: $\mu_1 - \mu_2 \neq 0$.

Řešení:

ad a) Úloha vede na F-test. Vypočteme realizaci testového kritéria:

$\dfrac{s_1^2}{s_2^2} = \dfrac{5060^2}{4310^2} = 1{,}3783$, dále najdeme příslušné kvantily:

$F_{\alpha/2}(n_1 - 1, n_2 - 1) = F_{0{,}025}(13{,}10) = 0{,}3077, F_{1-\alpha/2}(n_1 - 1, n_2 - 1) = F_{0{,}975}(13{,}10) = 3{,}5832$.

Protože testové kritérium $\dfrac{s_1^2}{s_2^2} = 1{,}3783$ se nerealizuje v kritickém oboru

$W = (0; 0{,}3077\rangle \cup \langle 3{,}5832; \infty)$, nelze na hladině významnosti 0,05 zamítnout hypotézu o shodě rozptylů.

ad b) Úloha vede na dvouvýběrový t-test. Protože jsme na hladině významnosti 0,05 nezamítli hypotézu o shodě rozptylů, můžeme rozptyly σ_1^2, σ_2^2 považovat za shodné a za jejich odhad vezmeme vážený průměr výběrových rozptylů

$$s_*^2 = \frac{13 \cdot 5060^2 + 10 \cdot 4310^2}{23} = 22548165{,}217.$$

Vypočteme realizaci testového kritéria:

$$\frac{m_1 - m_2 - c}{s_* \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{39600 - 41200}{\sqrt{22548165{,}217}\sqrt{\frac{1}{14} + \frac{1}{11}}} = -0{,}8363,\ t_{1-\alpha/2}(n_1 + n_2 - 2) = t_{0{,}975}(23) = 2{,}0687$$

Protože testové kritérium -0,8363 se nerealizuje v kritickém oboru $W = (-\infty; -2{,}0687\rangle \cup \langle 2{,}0687; \infty)$, na hladině významnosti 0,05 nelze zamítnout hypotézu o shodě středních hodnot.

Kontrolní otázky

1. Které pivotové statistiky používáme při řešení úloh o rozdílu středních hodnot a podílu rozptylů dvou normálních rozložení?
2. Jaké meze má 100(1-α)% empirický interval spolehlivosti pro podíl směrodatných odchylek dvou normálních rozložení?
3. V čem spočívá rozdíl mezi dvouvýběrovým z-testem a dvouvýběrovým t-testem?
4. V jakých situacích používáme dvouvýběrový t-test a v jakých a párový t-test?
5. K čemu slouží F-test?

Příklady

1. Bylo vylosováno 11 stejně starých selat téhož plemene. Šesti z nich byla předepsána výkrmná dieta č. 1 a zbylým pěti výkrmná dieta č. 2. Průměrné denní přírůstky v Dg za dobu půl roku jsou následující:
dieta č. 1: 62, 54, 55, 60, 53, 58
dieta č. 2: 52, 56, 49, 50, 51.
Zjištěné hodnoty považujeme za realizace dvou nezávislých náhodných výběrů pocházejících z rozložení $N(\mu_1, \sigma_1^2)$ a $N(\mu_2, \sigma_2^2)$. Sestrojte 95% empirický interval spolehlivosti pro podíl rozptylů a 95% empirický interval spolehlivosti pro rozdíl středních hodnot $\mu_1 - \mu_2$.
Výsledek:

$$0{,}1872\ \mathrm{Dg}^2 < \frac{\sigma_1^2}{\sigma_2^2} < 12{,}9541\ \mathrm{Dg}^2$$ s pravděpodobností aspoň 0,95.

$0{,}99\ \mathrm{Dg} < \mu_1 - \mu_2 < 9{,}81\ \mathrm{Dg}$ s pravděpodobností aspoň 0,95.

2. Pro údaje z příkladu 1. testujte na hladině významnosti 0,05 hypotézu, že obě výkrmné diety mají stejný vliv na hmotnostní přírůstky selat.
Výsledek:
Testujeme hypotézu H_0: $\mu_1 - \mu_2 = 0$ proti H_1: $\mu_1 - \mu_2 \neq 0$
1. způsob – pomocí intervalu spolehlivosti. 95% empirický interval spolehlivosti pro $\mu_1 - \mu_2$ je interval (0,99; 9,81). Neobsahuje nulu, proto H_0 zamítáme na hladině významnosti 0,05.
2. způsob – pomocí kritického oboru. Protože testové kritérium se realizuje hodnotou 2,771, která patří do kritického oboru $(-\infty; -2{,}2622\rangle \cup \langle 2{,}2622; \infty)$, H_0 zamítáme na hladině významnosti 0,05.

3. Máme k dispozici relaizace dvou nezávislých náhodných výběrů z rozložení $N(\mu_1, \sigma^2)$ a $N(\mu_2, \sigma^2)$ o rozsazích $n_1 = 10$, $n_2 = 15$. Výběrové průměry se realizovaly hodnotami $m_1 = 120{,}56$, $m_2 = 124{,}13$, výběrové rozptyly hodnotami $s_1^2 = 9{,}14$, $s_2^2 = 8{,}95$. Lze na základě těchto výsledků zamítnout na hladině významnosti 0,1 nulovou hypotézu H_0: $\mu_1 - \mu_2 = 0$ ve prospěch oboustranné alternativy H_1: $\mu_1 - \mu_2 \neq 0$?
Výsledek: Nulovou hypotézu zamítáme na hladině významnosti 0,1.

4. (S) 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.
Na hladině významnosti 0,05 testujte hypotézu, že střední hodnoty doby obsluhy jsou v obou restauracích stejné. Uveďte hodnotu testové statistiky, počet stupňů volnosti, p-hodnotu a rozhodnutí o nulové hypotéze.
Výsledek:
Vzhled diagnostických grafů svědčí o normalitě dat v obou výběrech.
Nejprve je zapotřebí pomocí F-testu ověřit shodu rozptylů. Realizace testového kritéria = 1,493, počet stupňů volnosti = 19 a 14, odpovídající p-hodnota = 0,41044, hypotézu o shodě rozptylů tedy nezamítáme na hladině významnosti 0,05.
Shodu středních hodnot otestujeme pomocí dvouvýběrového t-testu. Realizace testového kritéria = 0,1237, počet stupňů volnosti = 33, odpovídající p-hodnota = 0,9023, hypotézu o shodě středních hodnot tedy nezamítáme na hladině významnosti 0,05.