9 Testy o rozdílu středních hodnot

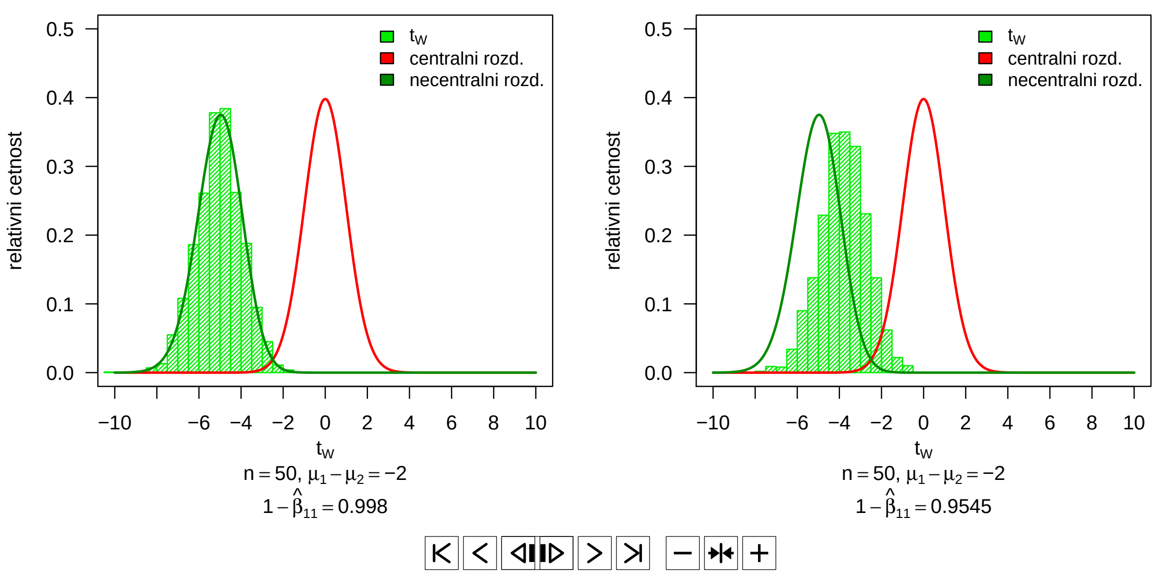
Příklad 9.1. Rozdělení testovací statistiky pro test o rozdílu středních hodnot $\mu_1 - \mu_2$
Nechť náhodný výběr X_1 pochází z normálního rozdělení, $X_1 \sim N(\mu_1, \sigma_1^2)$ a nechť náhodný výběr X_2 pochází z normálního rozdělení, $X_2 \sim N(\mu_2, \sigma_2^2)$. Pomocí simulační studie v porovnejte rozdělení testovací statistiky T_W pro klasický dvouvýběrový t-test nulové hypotézy H_0: $\mu_1 - \mu_2 = \mu_0$, kde $\mu_0 = 0$ (alternativní hypotéza $H_{11} : \mu_1 - \mu_2 \neq \mu_0$). Parametry zvolte $\mu_1 = 4$, $\mu_2 = 2.5$, $\sigma_1^2 = \sigma_2^2 = 2^2$, $n_1 = 50$, $n_2 = 50$.

Nechť dále X_1 pochází ze směsi dvou normálních rozdělení, t.j. $X_1 \sim [pN(4, 2^2) + (1 - p)N(4, 6^2)]$, kde $p = 0.8$ a nechť $X_2 \sim N(2.5, 2^2)$. Proveďte simulační studii popsanou výše také pro tento náhodný výběr.

Nasimulujte $M = 2\,000$ pseudonáhodných výběrů X_1 a X_2 a pro každé $m = 1, \ldots, 2\,000$ vypočítejte realizaci testovací statistiky $t_{W,\lambda}^{(m)}$ pro nulovou hypotézu H_0: $\mu_1 - \mu_2 = 0$ oproti $H_{11} : \mu_1 - \mu_2 \neq 0$. Vykreslete histogram testovacích statistik $t_{W,\lambda}^{(m)}$ a superponujte jej jednak křivkou hustoty Studentova rozdělení $t_{df,\lambda}$ s parametrem necentrality λ ($\lambda = \frac{\mu_1 - \mu_2 - \mu_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$, kde μ_1 a μ_2 jsou skutečné střední hodnoty (relevantní za platnosti H_{11}) a σ_1^2, σ_2^2 jsou skutečné rozptyly), a jednak křivkou hustoty centrálního Studentova rozdělení t_{df}, $df = n_1 + n_2 - 2$. Obě křivky vzájemně okometricky porovnejte.

Vytvořte animaci zobrazující odchýlení rozdělení testovací statistiky T_W od centrálního t-rozdělení. Zvolte $\mu_1 = 4$ a $\mu_2 \in \{6, 5.5, \ldots, 3, 2.5, 2\}$.

Poznámka: U směsi křivka hustoty necentrálního rozdělení nesuperponuje histogram statistik t_W dostatečně. Zamyslete se nad tím, proč, a co z toho pro nás do praxe vyplývá. Odpověď uveďte formou komentáře.



Obrázek 1: Rozdělení testovací statistiky T_W pro klasický dvouvýběrový t-test pro (a) normální rozdělení; (b) smíšené normální rozdělení

Příklad 9.2. Pravděpodobnost pokrytí klasického a věrohodnostního dvouvýběrového testu (nejen) pro směs

Pomocí simulační studie ($M = 2\,000$) vypočítejte pravděpodobnost pokrytí 95 % DIS pro rozdíl $\mu_1 - \mu_2$, a to jako podíl $\frac{\sum_{m=1}^{M} I(|t_{W,m}|<t_{df}(1-\alpha/2))}{M}$, kde t_W, m jsou testovací statistiky (1) klasického dvouvýběrového t-testu, (2) dvouvýběrového t-testu s Welchovou aproximací; (3) věrohodnostního testu za předpokladu, že rozptyly σ_1^2 a σ_2^2 jsou neznámé ale shodné; (4) věrohodnostního testu za předpokladu, že rozptyly σ_1^2 a σ_2^2 jsou neznámé a různé. Hodnoty parametrů volte následující:

(a) $X_j \sim N(\mu_j, \sigma^2)$, kde $j = 1, 2$, $\mu_1 = 20$, $\mu_2 = 20$ a $\sigma^2 = 9^2$;

(b) $X_j \sim N(\mu_j, \sigma_j^2)$, kde $j = 1, 2$, $\mu_1 = 20$, $\mu_2 = 20$, $\sigma_1^2 = 9^2$ a $\sigma_2^2 = 12^2$;

(c) $X_j \sim pN(\mu_j, \sigma^2) + (1-p)N(\mu_j, \sigma_a^2)$, kde $j = 1, 2$, $\mu_1 = 20$, $\mu_2 = 20$, $\sigma^2 = 9^2$ a $\sigma_a^2 = 18^2$, $p = 0.8$;

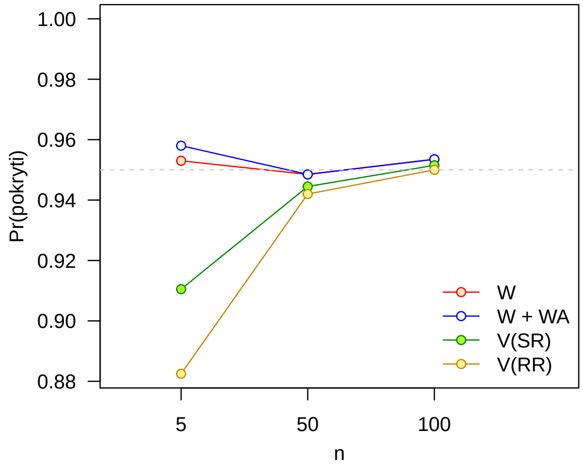
(d) $X_j \sim pN(\mu_j, \sigma_j^2) + (1-p)N(\mu_j, \sigma_{ja}^2)$, kde $j = 1, 2$, $\mu_1 = 20$, $\mu_2 = 20$, $\sigma_1^2 = 9^2$, $\sigma_2^2 = 12^2$, $\sigma_{1a}^2 = 18^2$, $\sigma_{2a}^2 = 22^2$, $p = 0.8$.

Rozsahy náhodných výběrů zvolte (i) $n_1 = n_2 = 5$; (ii) $n_1 = n_2 = 50$; (iii) $n_1 = n_2 = 100$. Pro každou situaci (a)–(d) vykreslete spojitý diagram zachycující pravděpodobnost pokrytí pro DIS (1)–(4) při volbách rozsahů n_1 a n_2 (i)–(iii). Jednotlivé typy DIS v grafu barevně odlište.

Nakonec zhodnoťte uvedené typy DIS ((1)–(4)) podle pravděpodobnosti pokrytí a uveďte, který DIS má z hlediska pravděpodobnosti pokrytí nejlepší a který naopak nejhorší vlastnosti. V úvahu vezměte jednak změnu vzhledem k rozsahu náhodného výběru a jednak chování pravděpodobnosti pokrytí v různých situacích (a)–(d).

Poznámka: V případě Waldova DIS, si můžeme vybrat, zda jmenovatel vzorce na výpočet aktuální pravděpodobnosti pokrytí vypočítáme jako počet testovacích statistik $t_{W,m}$, které náleží do kritického oboru W, nebo jako počet intervalů spolehlivosti IS_m, které pokrývají $\mu_1 - \mu_2 = \mu_0 = 0$. Oba postupy jsou ekvivalentní.

V případě věrohodnostního DIS je vhodnější vypočítat tento jmenovatel jako počet testovacích statistik $t_{W,m}$, které náleží do kritického oboru W. V případě výpočtu přes IS_m bychom k přesnému stanovení hranic IS_m potřebovali využít funkci `uniroot()`, kde bychom ale měli problém s automatizováním počátečních podmínek. Druhou možností by bylo spočítat hranice IS_m např. jako v příkladu 5.8, tj. vygenerovat posloupnost rozdílů $\mu_1 - \mu_2$, spočítat hodnoty ULR testovacích statistik pro každý rozdíl a hranice stanovit jako největší, resp. nejmenší hodnotu rozdílu, pro kterou je ULR_m menší než $\chi_1^2(1-\alpha)$. Tím bychom však získali pouze přibližné hranice IS_m. V okrajových případech by se tak mohlo stát, že rozdíl $\mu_1 - \mu_2$ by kvůli této nepřesnosti nesprávně spadl do IS, do kterého nepatří, nebo by naopak nespadl do IS, do kterého ve skutečnosti patří a aktuální pravděpodobnost pokrytí by nebyla vypočítaná správně. Z toho důvodu použijeme v obou případech (Waldova i věrohodnostního DIS) kritérium pomocí testovací statistiky a kritického oboru.



Obrázek 2: Pravděpodobnost pokrytí 95 % (1) klasického Waldova DIS (W), (2) Waldova DIS s Welchovou aproximací (W+WA); (3) věrohodnostního DIS za předpokladu, že rozptyly σ_1^2 a σ_2^2 jsou neznámé ale shodné (V(SR)); (4) věrohodnostního DIS za předpokladu, že rozptyly σ_1^2 a σ_2^2 jsou neznámé a různé (V(RR)); při volbě sady parametrů (a) vlevo nahoře; (b) vpravo nahoře; (c) vlevo dole; (d) vpravo dole

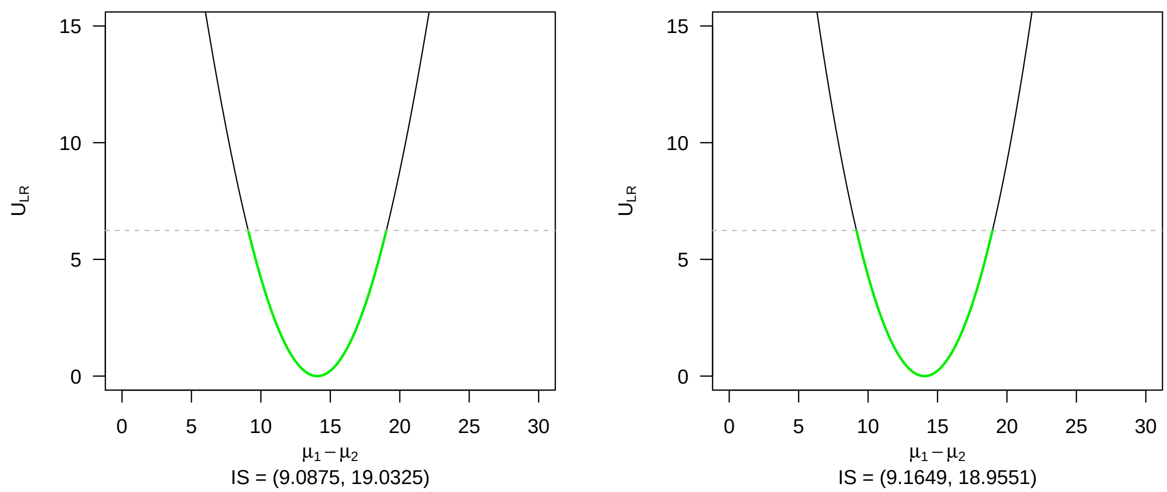
Tabulka 1: Pravděpodobnost pokrytí 95 % (1) klasického Waldova DIS (W), (2) Waldova DIS s Welchovou aproximací (W+WA); (3) věrohodnostního DIS za předpokladu, že rozptyly σ_1^2 a σ_2^2 jsou neznámé ale shodné (V(SR)); (4) věrohodnostního DIS za předpokladu, že rozptyly σ_1^2 a σ_2^2 jsou neznámé a různé (V(RR)); při volbě sady parametrů (a)–(d)

μ_1	μ_2	σ_{11}	σ_{21}	σ_{12}	σ_{22}	n_1	n_2	Pr_{W}(pokrytí)	$\mathrm{Pr}_{\mathrm{W+WA}}$(pokrytí)	$\mathrm{Pr}_{\mathrm{V(SR)}}$(pokrytí)	$\mathrm{Pr}_{\mathrm{V(RR)}}$(pokrytí)
20	20	9	9			5	5	0.9465	0.9500	0.9025	0.8805
20	20	9	9			50	50	0.9525	0.9525	0.9480	0.9435
20	20	9	9			100	100	0.9525	0.9525	0.9525	0.9515
20	20	9	12			5	5	0.9500	0.9585	0.9075	0.8785
20	20	9	12			50	50	0.9550	0.9550	0.9525	0.9510
20	20	9	12			100	100	0.9565	0.9570	0.9555	0.9550
20	20	9	9	18	18	5	5	0.9525	0.9575	0.9150	0.8850
20	20	9	9	18	18	50	50	0.9555	0.9565	0.9510	0.9495
20	20	9	9	18	18	100	100	0.9510	0.9510	0.9500	0.9500
20	20	9	12	18	22	5	5	0.9530	0.9580	0.9105	0.8825
20	20	9	12	18	22	50	50	0.9485	0.9485	0.9445	0.9420
20	20	9	12	18	22	100	100	0.9535	0.9535	0.9515	0.9500

Příklad 9.3. Dvouvýběrový test o rozdílu středních hodnot $\mu_1 - \mu_2$; praktický příklad
Načtěte datový soubor 03-paired-mean-clavicle.txt. Zmíněný soubor obsahuje osteometrická data o délce klíční kosti (clavicula) anglického souboru 50 mužských a 50 ženských dokumentovaných skeletů. Konkrétně jde o délku klíční kosti z pravé strany těla (length.R) a levé strany těla (length.L).

Na hladině významnosti $\alpha = 0.05$ otestujte nulovou hypotézu o shodě délky klíční kosti na pravé straně u mužů a u žen. K testování použijte (1) klasický dvouvýběrový t-test, (2) dvouvýběrový t-test s Welchovou aproximací; (3) věrohodnostní test za předpokladu, že rozptyly σ_1^2 a σ_2^2 jsou neznámé ale shodné; (4) věrohodnostní test za předpokladu, že rozptyly σ_1^2 a σ_2^2 jsou neznámé a různé. Testování proveďte pomocí (i) kritického oboru, (ii) intervalu spolehlivosti, (iii) p-hodnoty. Před testováním ověřte předpoklad normality a předpoklad shody rozptylů obou náhodných výběrů.

Vygreslete grafy 95 % věrohodnostních empirických intervalů spolehlivosti za předpokladu, že (a) rozptyly σ_1^2 a σ_2^2 jsou neznámé ale shodné, (b)) rozptyly σ_1^2 a σ_2^2 jsou neznámé a různé.



Obrázek 3: 95 % věrohodnostní empirické intervaly spolehlivosti za předpokladu, že (a) rozptyly σ_1^2 a σ_2^2 jsou neznámé ale shodné (vlevo), (b) rozptyly σ_1^2 a σ_2^2 jsou neznámé a různé (vpravo)

Tabulka 2: Výsledky klasického t-testu, t-testu s Welchovou aproximací a věrohodnostních testů pro rozdíl středních hodnot $\mu_1 - \mu_2$

	$\widehat{\mu}_1$	$\widehat{\mu}_2$	statistika	W_{hh}	W_{dh}	IS_{dh}	IS_{hh}	$p-$hodnota
klasický t-test	151.7400	137.6800	7.1019	-2.5446	2.5446	9.0223	19.0977	0.0000
t-test s Welchovou aprox.	151.7400	137.6800	7.1019	-2.5473	2.5473	9.0171	19.1029	0.0000
věrohodnostní test: $\sigma_1^2 = \sigma_2^2$	151.7400	137.6800	41.5196		6.2385	9.0875	19.0325	0.0000
věrohodnostní test: $\sigma_1^2 \neq \sigma_2^2$	151.7400	137.6800	51.4667		6.2385	9.1649	18.9551	0.0000