

Univerzita Karlova v Praze
Matematicko-fyzikální fakulta

DIPLOMOVÁ PRÁCE



Eliška Krausová

Vlastnosti dvoufázových testovacích procedur

Katedra pravděpodobnosti a matematické statistiky

Vedoucí diplomové práce: doc. Mgr. Michal Kulich, Ph.D.
Studijní program: Pravděpodobnost, matematická statistika
a ekonometrie

2010

Na tomto místě bych chtěla poděkovat vedoucímu své práce doc. Mgr. Michalu Kulichovi, Ph.D., za jeho trpělivost a za rady, které mi poskytl.

Prohlašuji, že jsem svou diplomovou práci napsala samostatně a výhradně s použitím citovaných pramenů. Souhlasím se zapůjčováním práce a jejím zveřejňováním.

V Praze dne 14.4.2010

Eliška Krausová

Obsah

1	Úvod	5
2	Dvoufázové testovací procedury	7
2.1	Kde se procedury doporučují	7
2.2	Popis procedury	11
3	Statistické testy	14
3.1	Jednovýběrové testy	14
3.2	Dvouvýběrové testy	16
3.3	Testy pro ověření předpokladů	19
4	Asymptotické vlastnosti dvoufázových procedur	23
4.1	Celková pravděpodobnost zamítnutí	24
4.2	Asymptotická hladina	26
4.3	Asymptotická síla	28
5	Simulace	29
5.1	Šikmé normální rozdělení	29
5.2	Předpoklad normality - 1 výběr	31
5.3	Předpoklad normality - 2 výběry	39
5.4	Předpoklad shody rozptylů - 2 výběry	46
6	Závěr	56
	Literatura	57

Název práce: Vlastnosti dvoufázových testovacích procedur

Autor: Eliška Krausová

Katedra (ústav): Katedra pravděpodobnosti a matematické statistiky

Vedoucí diplomové práce: doc. Mgr. Michal Kulich, Ph.D.

e-mail vedoucího: kulich@karlin.mff.cuni.cz

Abstrakt: V předložené práci studujeme dvoufázové testovací procedury, které nejprve formálně ověří předpoklady nějakým testem a podle jeho výsledku vyberou jednu z více variant testové statistiky. Na začátku popíšeme dvoufázové testovací procedury a zmíníme literaturu, kde jsou doporučovány. Pokusíme se odvodit jejich hladinu a sílu v jednoduchých případech. Zjištěné vlastnosti nakonec demonstrujeme na simulovaných datech.

Klíčová slova: dvoufázové testovací procedury, ověření předpokladů, hladina testu, síla testu

Title: Properties of two-phase testing procedures

Author: Eliška Krausová

Department: Department of Probability and Mathematical Statistics

Supervisor: doc. Mgr. Michal Kulich, Ph.D.

Supervisor's e-mail address: kulich@karlin.mff.cuni.cz

Abstract: In the present work we study properties of two-phase testing procedures which formally verify assumptions by performing some test (first phase) and subsequently calculate a test statistic selected according to the results of the previous test (second phase). In the beginning we describe two-phase testing procedures and mention some literature, in which they are recommended. We try to derive formula for the combined level and power of the whole two-phase testing procedure. After that we illustrate their properties through simulation studies.

Keywords: two-phase testing procedures, verification of assumptions, level of test, power of test

Kapitola 1

Úvod

Statistická literatura často doporučuje ověřit předpoklady nějakého testu (např. normalitu, rovnost rozptylů) provedením formálního testu jejich platnosti a podle výsledku tohoto testu vybrat jednu z více variant testové statistiky (např. t-test versus Wilcoxonův test). Tato testová procedura jako celek ovšem nemusí nutně dodržovat požadovanou hladinu ani mít optimální sílu.

V této práci se budeme zabývat dvoufázovými testovacími procedurami, zejména jejich teoretickými vlastnostmi, které se pokusíme demonstrovat na simulovaných datech. Také zmíníme praktické problémy související s jejich použitím.

V druhé kapitole podrobněji popíšeme dvoufázové procedury, uvedeme příklady nejčastěji doporučovaných procedur a literaturu, kde se tyto procedury vyskytují a porovnáme různá doporučení z této literatury.

Ve třetí kapitole popíšeme testy, které se v dvoufázových procedurách nejčastěji objevují. Půjde zejména o testy středních hodnot pro jeden nebo dva výběry a testy předpokladů normality a shody rozptylů.

Ve čtvrté kapitole se budeme zabývat vlastnostmi dvoufázových testovacích procedur, pokusíme se odvodit jejich asymptotickou hladinu a sílu v jednoduchých případech.

V páté kapitole provedeme simulační studii některých běžných dvoufázových procedur a zaměříme se na to, jak dodržují předepsanou hladinu a jakou mají sílu proti různým alternativám.

Kapitola 2

Dvoufázové testovací procedury

2.1 Kde se procedury doporučují

Předpokládejme, že chceme testovat nějakou hypotézu H_0 o rozdělení dat $\mathcal{L}(X)$. Pro test hypotézy máme na výběr dva testy, z nichž první test T_1 můžeme bez dalších omezení použít, zatímco druhý test T_2 vyžaduje splnění nějakého dalšího předpokladu, ale může dávat přesnější výsledky. K ověření tohoto dodatečného předpokladu máme k dispozici test T_X .

Dvoufázové procedury jsou nejčastěji doporučovány v prakticky orientované statistické literatuře, určené například pro biology, lékaře nebo psychology. Vyskutují se zde ve formě návodů nebo tabulek a diagramů pro jednotlivé situace.

Mezi nejčastěji doporučované dvoufázové procedury patří volba mezi t-testem a Wilcoxonovým testem na základě testu normality (Shapirův test, Lilieforsův test) (např. Hendl [6] nebo Dowdy and Wearden [5]), volba mezi klasickým t-testem a Welschovou variantou t-testu pro různé rozptyly na základě testu shody rozptylů (Fisherův F-test, Leveneův test) (např. Hendl [6]) nebo volba mezi t-testem a Wilcoxonovým testem na základě testu shody rozptylů (např. Zvára [9]).

Příkladem může být diagram (1), který uvádí Dowdy and Wearden [5], nebo na první pohled podobný diagram (2), který uvádí Hendl [6]. V obou

těchto diagramech vybíráme vhodný test středních hodnot na základě typu škály, počtu pozorování, tvaru rozdělení a shody rozptylů. Všimněme si však několika rozdílů. Zatímco Hendl [6] doporučuje vícefázovou proceduru pro větší výběry ($n > 10$, $n_1 \cdot n_2 > 10$) a pro malé výběry rovnou doporučuje neparametrický test (většinou Wilcoxonův), Dowdy and Wearden [5] doporučují tyto procedury pro menší počet pozorování ($n < 30$, $n_1 < 30$, $n_2 < 30$), zatímco pro větší počet pozorování využívají centrální limitní věty a t-statistiku srovnávají s normálním rozdělením.

Ve více teoreticky zaměřené literatuře určené přímo pro statistiky jsou autoři opatrnější a často dávají přednost použití grafů a zkušeností statistika před klasickým testem předpokladu (např. qq-plot jako ověření předpokladu normality) jako např. Zvára [9] nebo popisují všechny testy a jejich vlastnosti, ale konkrétní návody už neuvádějí jako např. Anděl [1].

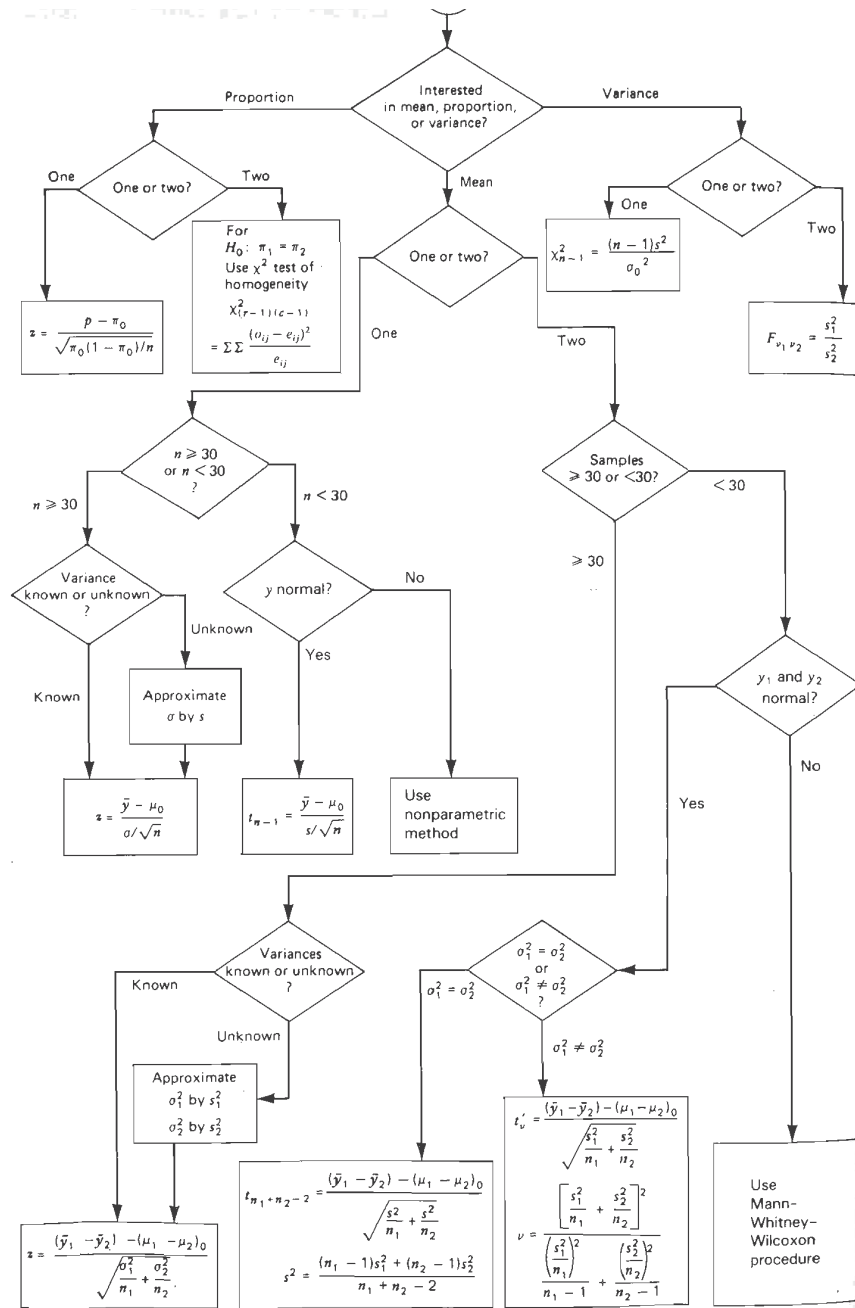
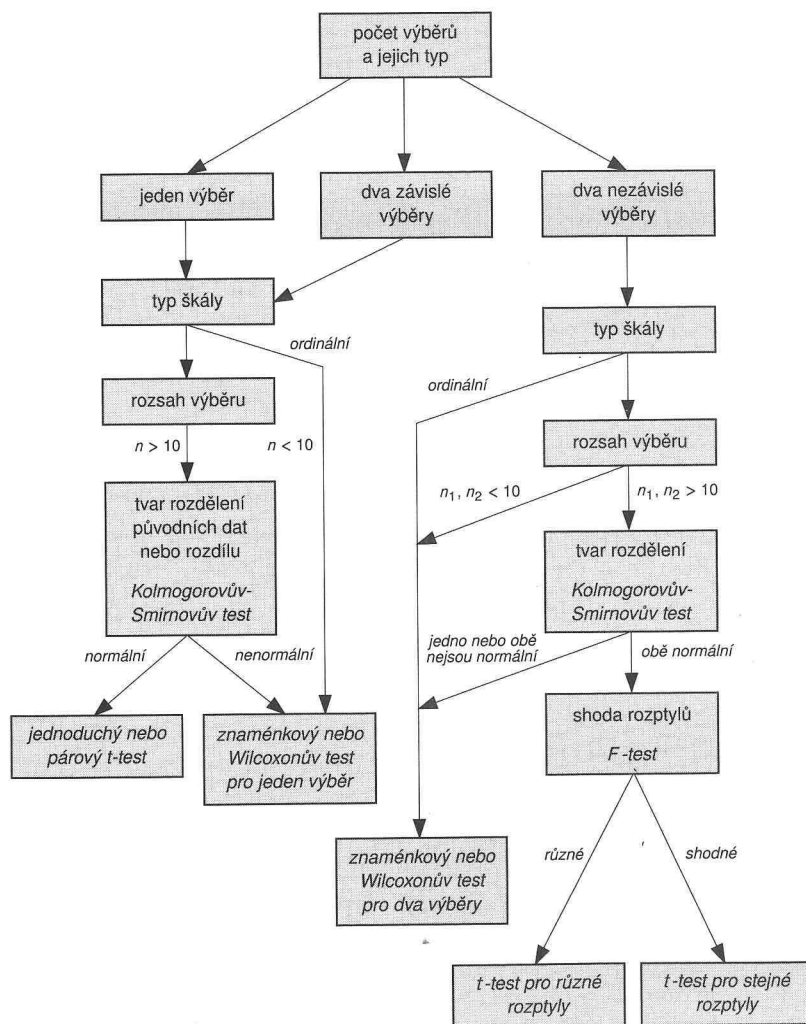


FIGURE 8.4. Flowchart of test statistics.

Obrázek 2.1: diagram uvedený v Dowdy and Wearden [5]

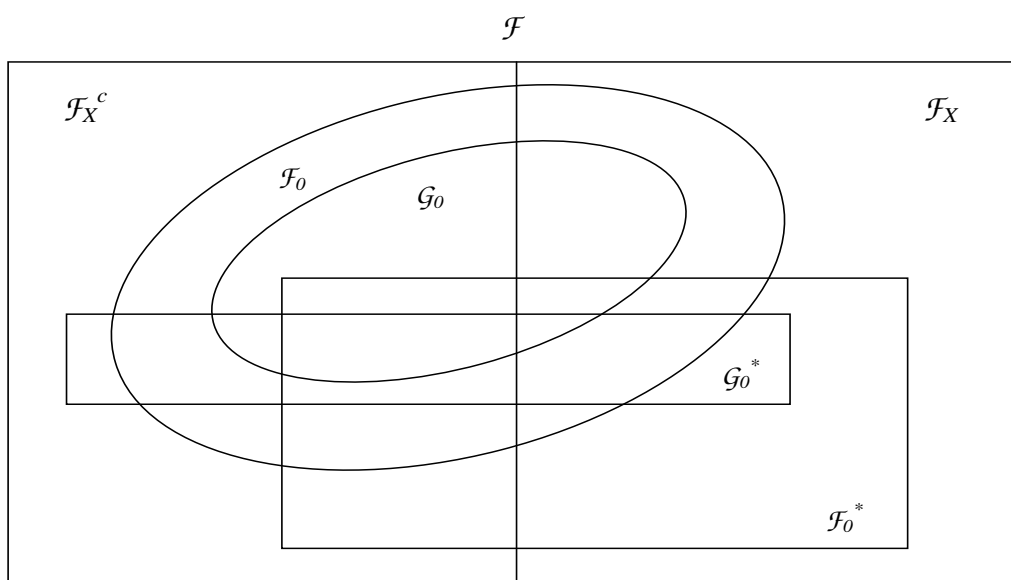


Obrázek 2.2: diagram uvedený v Hendl [6]

2.2 Popis procedury

V prvním kroku dvoufázové procedury otestujeme daný předpoklad pomocí testové statistiky T_X . Pokud hypotézu H_X o platnosti předpokladu nezamítneme, použijeme v druhém kroku testovou statistiku T_2 , v případě zamítnutí použijeme testovou statistiku T_1 . Přitom testy T_1 a T_2 nemusí nutně testovat stejnou hypotézu. Nyní popíšeme proceduru podrobněji v vztahu k modelu na obrázku 2.3.

Nechť $\mathcal{L}(X) \in \mathcal{F}$, kde \mathcal{F} je množina všech uvažovaných rozdělení.



Obrázek 2.3: Množiny rozdělení, na kterých platí nulové hypotézy nebo na kterých nedochází k zamítání hypotéz.

Testová statistika T_X testuje hypotézu

$$H_X : \mathcal{L}(X) \in \mathcal{F}_X$$

a slouží pro výběr testu.

Rozhodování:

$$T_X \in C_X(\alpha_X) \rightarrow \text{zamítáme } H_X, \text{ použijeme test } T_1$$

$T_X \notin C_X(\alpha_X) \rightarrow$ nezamítáme H_X , použijeme test T_2

Hypotéza H_X je testována na hladině $P_{H_X} [T_X \in C_X(\alpha_X)] \leq \alpha_X$ a C_X je kritický obor testu.

Test T_1 testuje hypotézu

$$H_0 : \mathcal{L}(X) \in \mathcal{G}_0.$$

Rozhodování:

$$T_1 \in C_1(\alpha) \rightarrow \text{zamítáme } H_0$$

$$T_1 \notin C_1(\alpha) \rightarrow \text{nezamítáme } H_0$$

Testujeme na hladině $P_{H_0} [T_1 \in C_1(\alpha)] \leq \alpha$. Množinu rozdělení, která splňují tuto podmínku, označíme \mathcal{F}_0 . Tedy

$$\mathcal{F}_0 = \{F \in \mathcal{F} : P_F [T_1 \in C_1(\alpha)] \leq \alpha\}.$$

V ideálním případě by platilo $\mathcal{F}_0 = \mathcal{G}_0$, může ale také nastat situace, kdy H_0 neplatí a podmínka je splněna, tedy $\mathcal{G}_0 \subset \mathcal{F}_0$.

Test T_2 testuje hypotézu

$$H_0^* : \mathcal{L}(X) \in \mathcal{G}_0^*.$$

Rozhodování:

$$T_2 \in C_2(\alpha) \rightarrow \text{zamítáme } H_0^*$$

$$T_2 \notin C_2(\alpha) \rightarrow \text{nezamítáme } H_0^*$$

Testujeme na hladině $P_{H_0^*} [T_2 \in C_2(\alpha)] \leq \alpha$. Množinu rozdělení, která splňují tuto podmínku, označíme \mathcal{F}_0^* . Tedy

$$\mathcal{F}_0^* = \{F \in \mathcal{F} : P_F [T_2 \in C_2(\alpha)] \leq \alpha\}.$$

U tohoto testu ovšem máme zaručeno dodržení hladiny jen pro rozdělení $\mathcal{L}(X) \in \mathcal{F}_X$, neboli za platnosti hypotézy H_X . Nemusí tedy platit vztah $\mathcal{G}_0^* \subseteq \mathcal{F}_0^*$, ale platí $\mathcal{F}_X \cap \mathcal{G}_0^* \subseteq \mathcal{F}_0^*$.

Příklad k popsanému modelu

Mějme náhodnou veličinu X , jejíž rozdělení $\mathcal{L}(X)$ pochází z množiny všech spojitých symetrických rozdělení. Tuto množinu označíme \mathcal{F} . Chceme testovat hypotézu $E X = \mu$. Použijeme proceduru, kde nejprve otestujeme normalitu pomocí Shapiro-Wilkova testu T_X a na základě jeho výsledku použijeme buď Wilcoxonův test T_1 (testuje hypotézu $H_0 : \tilde{x} = \mu$) nebo dvouvýběrový t-test T_2 (testuje hypotézu $H_0^* : E X = \mu$).

$$\begin{aligned}\mathcal{F}_X &= \{F \in \mathcal{F} : X \text{ má normální rozdělení} \} \\ \mathcal{F}_0 &= \left\{ F \in \mathcal{F} : P_F [X < \mu] = \frac{1}{2} \right\} \\ \mathcal{G}_0 &= \left\{ F \in \mathcal{F} : P_F [X < \mu] = \frac{1}{2} \right\} \\ \mathcal{F}_0^* &= \left\{ F \in \mathcal{F} : \left| \frac{\bar{X} - \mu}{S} \sqrt{n} \right| \leq t_{n-1}(\alpha) \right\} \\ \mathcal{G}_0^* &= \{F \in \mathcal{F} : E X = \mu\}\end{aligned}$$

V tomto jednoduchém příkladě, díky vhodně zvolené množině \mathcal{F} , vyjdou množiny $\mathcal{F}_0, \mathcal{G}_0$ a \mathcal{G}_0^* totožné. Dále bude platit, že $\mathcal{F}_X \cap \mathcal{G}_0^* = \mathcal{F}_X \cap \mathcal{F}_0^*$.

Kapitola 3

Statistické testy

V této kapitole popíšeme jednotlivé testy, které se nejčastěji vyskytují v dvoufázových testovacích procedurách a které budeme používat v simulacích v dalších kapitolách. Hlavně se budeme zabývat testy střední hodnoty nebo mediánu.

3.1 Jednovýběrové testy

Následující formulace jednovýběrových testů jsou převzaty z knihy Anděl [1].

Jednovýběrový t-test

Mezi nejpoužívanější testy střední hodnoty patří jednovýběrový T-test.

Věta 3.1. *Nechť X_1, \dots, X_n je výběr z $N(\mu, \sigma^2)$, kde $n \geq 2$ a $\sigma^2 > 0$. Pak*

$$T = \frac{\bar{X} - \mu}{S} \sqrt{n} \sim t_{n-1}.$$

Definice 3.2. *Nechť $T \sim t_k$. Kritická hodnota $t_k(\alpha)$ je definována vzorcem $P\{|T| \geq t_k(\alpha)\} = \alpha$.*

Nechť X_1, \dots, X_n je výběr z $N(\mu, \sigma^2)$, kde $n \geq 2$. Parametr $\sigma^2 > 0$ není znám. Je třeba testovat hypotézu $H_0 : \mu = \mu_0$ proti $H_1 : \mu \neq \mu_0$, kde μ_0 je dané číslo. Hypotéza H_0 se bude zamítat, bude-li \bar{X} hodně vzdáleno od μ_0 . Za platnosti H_0 máme $P\{|\bar{X} - \mu_0| \sqrt{n}/S \geq t_{n-1}(\alpha)\} = \alpha$. Hypotézu H_0 tedy zamítneme na hladině α , když bude platit $|\bar{X} - \mu_0| \sqrt{n}/S \geq t_{n-1}(\alpha)$.

Je známo, že T-test není příliš citlivý na předpoklad normality.

Jednovýběrový Wilcoxonův test

Neparametrickou obdobou jednovýběrového t-testu je jednovýběrový Wilcoxonův test.

Definice 3.3. *Nechť $T \sim N(0, 1)$. Kritická hodnota $u(\alpha)$ je definována vzorcem $P\{T \geq u(\alpha)\} = \alpha$. Tedy $u(\alpha) = \Phi^{-1}(1 - \alpha)$.*

Nechť X_1, \dots, X_n je náhodný výběr ze spojitého rozdělení s hustotou f , která je symetrická kolem bodu a a kladná v jeho okolí. Platí tedy $f(a+x) = f(a-x)$. Z toho plyne že a musí být rovno mediánu \tilde{x} . Chceme testovat hypotézu $H_0 : \tilde{x} = x_0$ proti alternativě $H_1 : \tilde{x} \neq x_0$.

— Předpokládejme, že žádná z veličin X_i není rovna x_0 . Položme $Y_i = X_i - x_0$. Veličiny Y_i seřadíme do neklesající posloupnosti podle jejich absolutní hodnoty

$$|Y|_{(1)} \leq |Y|_{(2)} \leq \dots \leq |Y|_{(n)}.$$

Budiž R_i^+ pořadí veličiny Y_i . Označme

$$S^+ = \sum_{Y_i \geq 0} R_i^+, \quad S^- = \sum_{Y_i < 0} R_i^+.$$

Přitom platí $S^+ + S^- = n(n+1)/2$. Je-li číslo $\min(S^+, S^-)$ menší nebo rovno tabelované kritické hodnotě $w_n(\alpha)$, zamítáme H_0 .

Z učiněných předpokladů vyplývá, že při platnosti H_0 jsou Y_1, \dots, Y_n nezávislé stejně rozdělené náhodné veličiny, jejichž rozdělení je symetrické kolem nuly.

Věta 3.4. *Platí-li H_0 , pak*

$$E S^+ = \frac{1}{4}n(n+1), \quad \text{var } S^+ = \frac{1}{24}n(n+1)(2n+1).$$

Dá se dokázat, že za platnosti H_0 má S^+ asymptoticky normální rozdělení. Test hypotézy lze tudíž také založit na veličině

$$U = \frac{S^+ - E S^+}{\sqrt{\text{var } S^+}}.$$

Vyjde-li $|U| \geq u(\frac{\alpha}{2})$, zamítneme H_0 na hladině, která se s rostoucím n blíží číslu α .

Wilcoxonova testová statistika je tvořena součtem pořadí nezáporných pozorování mezi absolutními hodnotami všech pozorování. U Wilcoxonova testu je důležitý předpoklad symetrie rozdělení. K zamítnutí H_0 tedy může oprávněně dojít i tehdy, je-li medián roven x_0 , ale hustota f je výrazně nesymetrická.

3.2 Dvouvýběrové testy

Mezi nejpoužívanější testy střední hodnoty patří dvouvýběrový t-test. Jeho formulace je převzata z knihy Anděl [1].

Dvouvýběrový T-test

Věta 3.5. *Nechť X_1, \dots, X_n je výběr z $N(\mu_1, \sigma^2)$ a nechť Y_1, \dots, Y_m je výběr z $N(\mu_2, \sigma^2)$. Nechť tyto dva výběry jsou na sobě nezávislé. Předpokládejme, že $n \geq 2$, $m \geq 2$ a $\sigma^2 > 0$. Označme \bar{X}, S_X^2 a \bar{Y}, S_Y^2 charakteristiky těchto výběrů. Pak*

$$T = \frac{\bar{X} - \bar{Y} (\mu_1 - \mu_2)}{\sqrt{(n-1)S_X^2 + (m-1)S_Y^2}} \sqrt{\frac{nm(n+m-2)}{n+m}} \sim t_{n+m-2}$$

Chceme-li za uvedených předpokladů při neznámém $\sigma^2 > 0$ testovat hypotézu $H_0 : \mu_1 - \mu_2 = 0$ proti $H_1 : \mu_1 - \mu_2 \neq 0$, vypočteme

$$T = \frac{\bar{X} - \bar{Y}}{\sqrt{(n-1)S_X^2 + (m-1)S_Y^2}} \sqrt{\frac{nm(n+m-2)}{n+m}}.$$

Jestliže $|T| \geq t_{n+m-2}(\alpha)$ zamítneme H_0 na hladině α .

Dvouvýběrový t-test není příliš citlivý na nesplnění předpokladu normality, ale je citlivý na nesplnění předpokladu shody rozptylů, zejména pro různé počty pozorování v jednotlivých výběrech.

Welschův test

Pokud u t-testu není dodržen předpoklad shody rozptylů, je možné použít jeho variantu, tzv. Welschův test. Ten je popsán například v Dowdy and Wearden [5].

Věta 3.6. Necht' X_1, \dots, X_n je výběr z $N(\mu_1, \sigma_1^2)$ a necht' Y_1, \dots, Y_m je výběr z $N(\mu_2, \sigma_2^2)$. Necht' tyto dva výběry jsou na sobě nezávislé. Předpokládejme, že $n \geq 2$, $m \geq 2$, $\sigma_1^2 > 0$ a $\sigma_2^2 > 0$. Označme \bar{X}, S_X^2 a \bar{Y}, S_Y^2 charakteristiky těchto výběrů. Pak přibližně platí

$$T' = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{S_X^2}{n} + \frac{S_Y^2}{m}}} \sim t_\nu$$

kde

$$\nu = \frac{\left(\frac{S_X^2}{n} + \frac{S_Y^2}{m}\right)^2}{\frac{\left(\frac{S_X^2}{n}\right)^2}{n-1} + \frac{\left(\frac{S_Y^2}{m}\right)^2}{m-1}}$$

Chceme-li za uvedených předpokladů testovat hypotézu $H_0 : \mu_1 - \mu_2 = 0$ proti $H_1 : \mu_1 - \mu_2 \neq 0$, vypočteme statistiku T' a pokud $|T'| \geq t_\nu(\alpha)$ zamítneme H_0 na hladině α .

Dvouvýběrový Wilcoxonův test

Neparametrickou obdobou dvouvýběrového t-testu je Wilcoxonův test nebo jeho varianta někdy zvaná Mannův-Whitneyův test. Oba testy jsou popsány v knize Anděl [1].

Necht' X_1, \dots, X_n je náhodný výběr ze spojitého rozdělení s distribuční funkcí F a necht' Y_1, \dots, Y_m je na něm nezávislý náhodný výběr ze spojitého rozdělení s distribuční funkcí G . Je třeba testovat hypotézu $H_0 : F = G$ proti alternativě $H_1 : F \neq G$.

Všech $m + n$ veličin $X_1, \dots, X_n, Y_1, \dots, Y_m$ (tzv. *sdrúžený výběr*) uspořádáme vzestupně podle velikosti. Označme T_1 součet pořadí hodnot X_1, \dots, X_n a T_2 součet pořadí hodnot Y_1, \dots, Y_m . Je jasné, že

$$T_1 + T_2 = \frac{1}{2} (m + n) (m + n + 1).$$

Věta 3.7. *Platí-li H_0 , pak*

$$\begin{aligned} \mathbb{E} T_1 &= \frac{1}{2} m (m + n + 1), \\ \text{var } T_1 &= \frac{1}{12} m n (m + n + 1). \end{aligned}$$

Je dokázáno, že při $n \rightarrow \infty$ a $m \rightarrow \infty$ má veličina T_1 asymptoticky normální rozdělení.

Mannův-Whitneyův test

Místo veličiny se zpravidla používá veličina

$$U_1 = mn + \frac{1}{2} m (m + 1) - T_1.$$

Testu založenému na U_1 se pak říká Mannův-Whitneyův test. Dále se zavede označení

$$U_2 = mn + \frac{1}{2} n (n + 1) - T_2.$$

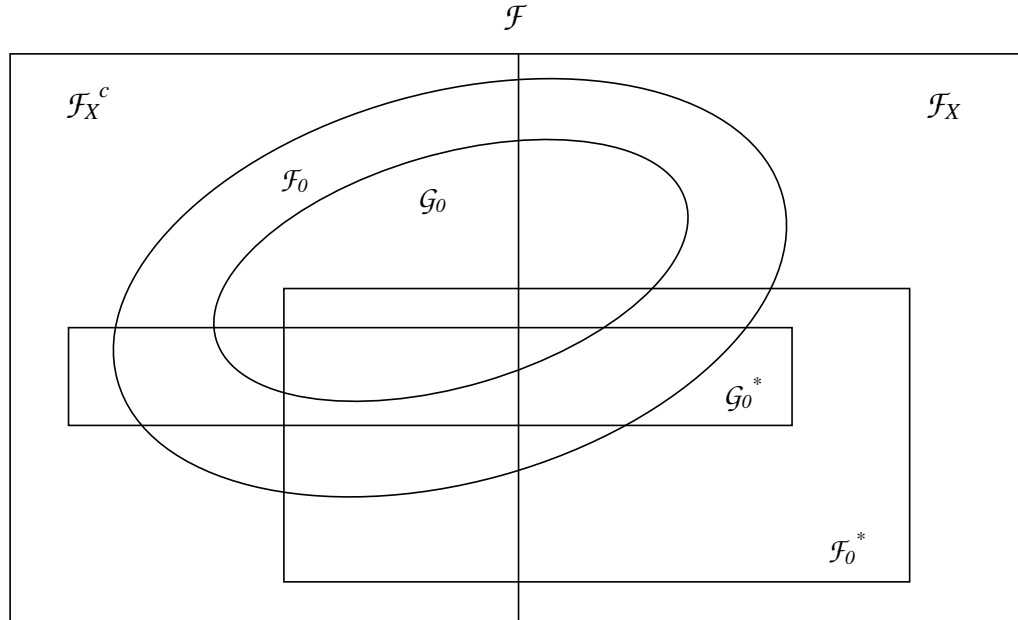
Přitom platí $U_1 + U_2 = mn$. Pokud $\min(U_1, U_2)$ je menší nebo rovno tabulované hodnotě, zamítá se nulová hypotéza. Platí

$$\mathbb{E} U_1 = \frac{1}{2} mn, \quad \text{var } U_1 = \frac{1}{12} mn (m + n + 1).$$

Testová statistika U tedy udává počet dvojic (X_i, Y_j) , kde $X_i < Y_j$.

Vztah k modelu v kapitole 2

Mějme náhodné veličiny X, Y jejichž sdružené rozdělení $\mathcal{L}(X, Y)$ pochází z nějaké množiny rozdělení \mathcal{F} (například všechna spojitá dvojrozměrná rozdělení, kde $\text{var } X = \text{var } Y$). Chceme testovat hypotézu $\mathbb{E} X = \mathbb{E} Y$. Použijeme proceduru, kde nejprve otestujeme normalitu pomocí Shapiro-Wilkova testu T_X a na jeho základě použijeme buď Wilcoxonův test T_1 (testuje hypotézu H_0) nebo dvouvýběrový t-test T_2 (testuje hypotézu H_0^*).



Obrázek 3.1: Množiny rozdělení, na kterých platí nulové hypotézy nebo na kterých nedochází k zamítní hypotéz.

Jednotlivé množiny z obrázku 3.1 budou vypadat takto:

$$\begin{aligned}
 \mathcal{F}_X &= \{F \in \mathcal{F} : X \text{ a } Y \text{ mají normální rozdělení} \} \\
 \mathcal{F}_0 &= \left\{ F \in \mathcal{F} : P_F[X < Y] = \frac{1}{2} \right\} \\
 \mathcal{G}_0 &= \{F \in \mathcal{F} : \mathcal{L}(X) = \mathcal{L}(Y)\} \\
 \mathcal{F}_0^* &= \left\{ F \in \mathcal{F} : \left| \frac{\bar{X} - \bar{Y}}{\sqrt{(n-1)S_X^2 + (m-1)S_Y^2}} \sqrt{\frac{nm(n+m-2)}{n+m}} \right| \leq t_{n+m-2}(\alpha) \right\} \\
 \mathcal{G}_0^* &= \{F \in \mathcal{F} : \mathbb{E}X = \mathbb{E}Y\}
 \end{aligned}$$

3.3 Testy pro ověření předpokladů

Testy normality

Dowdy and Wearden [5] uvádí pro testování normality Lillieforsův test (jako modifikaci Kolmogorova testu) a Shapiro-Wilkův test.

Lillieforsův test normality

Mějme náhodný výběr X_1, \dots, X_n z neznámého rozdělení s distribuční funkcí $F(x)$. Spočítáme normované hodnoty výběru Z_i , definované jako

$$Z_i = \frac{X_i - \bar{X}}{S} \quad i = 1, 2, \dots, n,$$

kde \bar{X} je výběrový průměr a S^2 výběrový rozptyl. Označme $S(x)$ empirickou distribuční funkci normovaného výběru a $\Phi(x)$ distribuční funkci normálního rozdělení $N(0, 1)$. Lillieforsova statistika T_1 je definována vzorcem

$$T_1 = \sup_x |\Phi(x) - S(x)|$$

Hypotézu H_0 o tom, že výběr pochází z normálního rozdělení s neznámou střední hodnotou a neznámým rozptylem zamítáme, pokud testová statistika T_1 překročí kritickou hodnotu uvedenou v tabulce v Dowdy and Wearden [5].

Shapiro-Wilkův test normality

Mějme náhodný výběr X_1, \dots, X_n z neznámého rozdělení s distribuční funkcí $F(x)$. Nejdříve spočítáme jmenovatele D testové statistiky

$$D = \sum_{i=1}^n (X_i - \bar{X})^2,$$

kde \bar{X} je výběrový průměr. Potom seřadíme pozorování podle velikosti

$$X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)},$$

kde $X_{(i)}$ značí i -tou pořadovou statistiku. Z tabulky uvedené v Dowdy and Wearden [5] pro velikost výběru n dostaneme koeficienty a_1, a_2, \dots, a_k , kde k je přibližně $n/2$. Testová statistika T_2 je dána vzorcem

$$T_2 = \frac{1}{D} \left[\sum_{i=1}^k a_i (X^{(n-i+1)} - X^{(i)}) \right]^2$$

Tato statistika bývá často označována písmenem W a test se často nazývá W -test.

Statistika T_2 v podstatě označuje druhou mocninu korelačního koeficientu, kde Pearsonův korelační koeficient je počítán mezi pořadovými statistikami $X_{(i)}$ a skóry a_i , které reprezentují, jak by měla statistika vypadat, pokud by rozdělení X_i bylo normální. Proto pokud je T_2 blízko 1, výběr se chová jako výběr z normálního rozdělení. Hypotézu H_0 o tom, že výběr pochází z normálního rozdělení s neznámou střední hodnotou a neznámým rozptylem zamítáme, pokud je testová statistika T_2 menší než příslušný kvantil této testové statistiky uvedený v tabulce v Dowdy and Wearden [5].

Další testy normality založené především na šikmosti a špičatosti uvádí např. Anděl [1].

Testy shody rozptylů

Mezi nejznámější testy shody rozptylů patří Fisherův F-test, který je ovšem velmi citlivý na porušení předpokladu normality. Vhodnou alternativou může být například neparametrický Siegel-Tukeyho test nebo Moodyho test používaný v programu R.

Fisherův F test

Formulace testu je převzata z knihy Anděl [1].

Chceme testovat hypotézu $H_0 : \text{var } X = \text{var } Y$ proti alternativě $H_1 : \text{var } X \neq \text{var } Y$.

Definice 3.8. *Nechť $Z \sim F_{m,n}$. Kritická hodnota $F_{m,n}(\alpha)$ je definována vzorcem $P\{Z \geq F_{m,n}(\alpha)\} = \alpha$.*

Věta 3.9. *Nechť X_1, \dots, X_n je výběr z $N(\mu_1, \sigma_1^2)$ a nechť Y_1, \dots, Y_m je výběr z $N(\mu_2, \sigma_2^2)$. Nechť tyto dva výběry jsou na sobě nezávislé. Předpokládejme, že $n \geq 2$, $m \geq 2$, $\sigma_1^2 > 0$ a $\sigma_2^2 > 0$. Označme \bar{X}, S_X^2 a \bar{Y}, S_Y^2 charakteristiky těchto výběrů.*

Platí-li $\sigma_1^2 = \sigma_2^2$, pak $Z = S_X^2/S_Y^2 \sim F_{n-1, m-1}$.

Nechť platí $S_X^2 \geq S_Y^2$. Pak hypotézu H_0 zamítáme, jestliže $S_X^2/S_Y^2 \geq F_{n-1, m-1}(\frac{\alpha}{2})$.

Moodyho test

Tento neparametrický test, používaný v programu R, je uveden v Conover [4].

Nechť X_1, \dots, X_n je náhodný výběr ze spojitého rozdělení s distribuční funkcí F a necht' Y_1, \dots, Y_m je na něm nezávislý náhodný výběr ze spojitého rozdělení s distribuční funkcí G .

Testujeme nulovou hypotézu

$$H_0 : \mathcal{L}(X) = \mathcal{L}(Y)$$

proti alternativě

$$H_1 : \text{var } X \neq \text{var } Y.$$

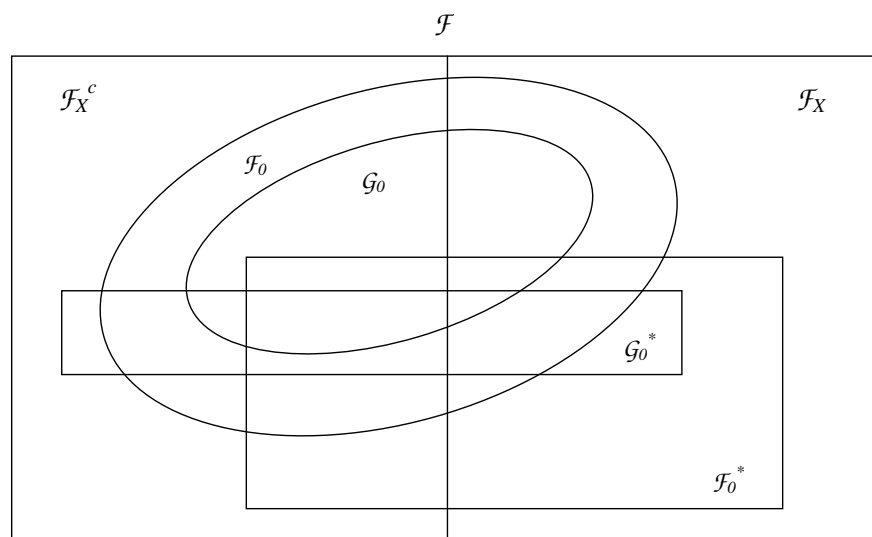
Test je založen na testové statistice

$$T = \sum_{i=1}^n \left[R(X_i) - \frac{n+m+1}{2} \right]^2,$$

kde $R(X_i)$ je pořadí pozorování X_i ve sdruženém výběru $X_1, \dots, X_n, Y_1, \dots, Y_m$. Přesné tabulky jsou uvedeny v Laubscher, Steffens, and DeLange [7].

Kapitola 4

Asymptotické vlastnosti dvoufázových procedur



Obrázek 4.1: Množiny rozdělení, na kterých platí nulové hypotézy nebo na kterých nedochází k zamítní hypotéz.

U popsaných testů lze jejich testové statistiky upravit tak, aby měly pro dostatečně velké množství pozorování asymptoticky normální rozdělení. Díky tomu se můžeme pokusit obecně vyjádřit asymptotickou hladinu a asymptotickou sílu dvoufázových testovacích procedur.

4.1 Celková pravděpodobnost zamítnutí

Celková pravděpodobnost zamítnutí dvoufázovou procedurou je pro X s distribuční funkcí F

$$P_{zam} = P_F [T_X \in C_X(\alpha_X), T_1 \in C_1(\alpha)] + P_F [T_X \notin C_X(\alpha_X), T_2 \in C_2(\alpha)].$$

Pokud je $F \in \mathcal{G}_0 \cap \mathcal{G}_0^*$ (viz. obrázek 4.1), tato pravděpodobnost by měla být menší nebo rovna α . Předpokládejme, že rozdělení dvojic statistik $\begin{pmatrix} T_X \\ T_1 \end{pmatrix}$ a $\begin{pmatrix} T_X \\ T_2 \end{pmatrix}$ konvergují k normálnímu rozdělení

$$\begin{aligned} \begin{pmatrix} T_X - \mu_X \\ T_1 - \mu_1 \end{pmatrix} &\longrightarrow N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{pmatrix}\right) \\ \begin{pmatrix} T_X - \mu_X \\ T_2 - \mu_2 \end{pmatrix} &\longrightarrow N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho_2 \\ \rho_2 & 1 \end{pmatrix}\right), \end{aligned}$$

kde μ_X , μ_1 a μ_2 obecně závisí na n . Za platnosti nulové hypotézy H_X resp. H_0, H_0^* je μ_X resp. μ_1, μ_2 nulové.

$$P_{zam} = P_F \left[|T_X| > u\left(\frac{\alpha_X}{2}\right), |T_1| > u\left(\frac{\alpha}{2}\right) \right] + P_F \left[|T_X| < u\left(\frac{\alpha_X}{2}\right), |T_2| > u\left(\frac{\alpha}{2}\right) \right]$$

Hustota dvourozměrného normálního rozdělení $N_2 \left[\begin{pmatrix} \mu_X \\ \mu_i \end{pmatrix}, \begin{pmatrix} 1 & \rho_i \\ \rho_i & 1 \end{pmatrix} \right]$ je

$$f_i(t_X, t_i) = \frac{1}{2\pi\sqrt{1-\rho_i^2}} \exp \left\{ -\frac{(t_X - \mu_X)^2 - 2\rho_i(t_X - \mu_X)(t_i - \mu_i) + (t_i - \mu_i)^2}{2(1-\rho_i^2)} \right\}.$$

Z věty 4.12 z Anděl[1] plyne, že podmíněné rozdělení T_i při daném $T_X = t_X$ je $N[\mu_i + \rho_i(t_X - \mu_X), (1 - \rho_i^2)]$.

Hustotu $f_i(t_X, t_i)$ můžeme rozepsat jako

$$f_i(t_X, t_i) = p_i(t_i|t_X) \cdot h(t_X),$$

kde $h(t_X)$ je hustota marginálního rozdělení statistiky T_X , tedy rozdělení $N(\mu_X, 1)$ a $p_i(t_i|t_X)$ je hustota podmíněného rozdělení T_i při daném $T_X = t_X$, tedy rozdělení $N[\mu_i + \rho_i(t_X - \mu_X), (1 - \rho_i^2)]$.

Označme

$$\begin{aligned} A_1 &= \left(-\infty, -u\left(\frac{\alpha_X}{2}\right)\right) \cup \left(u\left(\frac{\alpha_X}{2}\right), \infty\right) \\ A_2 &= \left(-u\left(\frac{\alpha_X}{2}\right), u\left(\frac{\alpha_X}{2}\right)\right) \\ B &= \left(-\infty, -u\left(\frac{\alpha}{2}\right)\right) \cup \left(u\left(\frac{\alpha}{2}\right), \infty\right). \end{aligned}$$

Potom

$$P_{zam} \doteq \int_{A_1} h(t_X) \left(\int_B p_1(t_1|t_X) dt_1 \right) dt_X + \int_{A_2} h(t_X) \left(\int_B p_2(t_2|t_X) dt_2 \right) dt_X.$$

Použijeme substituci:

$$y_i = \frac{t_i - \mu_i - \rho_i(t_X - \mu_X)}{\sqrt{1 - \rho_i^2}}; \quad \frac{dy_i}{dt_i} = \frac{1}{\sqrt{1 - \rho_i^2}}$$

a označíme

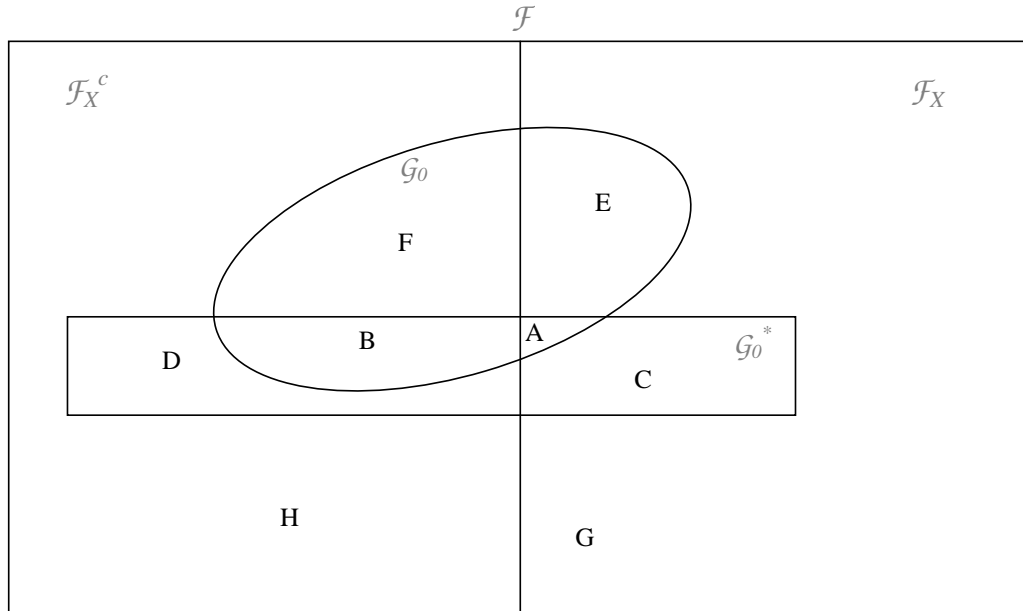
$$\begin{aligned} m_i^+(t_X) &= \frac{u\left(\frac{\alpha}{2}\right) - \mu_i - \rho_i(t_X - \mu_X)}{\sqrt{1 - \rho_i^2}} \\ m_i^-(t_X) &= \frac{-u\left(\frac{\alpha}{2}\right) - \mu_i - \rho_i(t_X - \mu_X)}{\sqrt{1 - \rho_i^2}}. \end{aligned}$$

Dostaneme

$$\int_B p_i(t_i|t_X) dt_i = \Phi(m_i^-(t_X)) + 1 - \Phi(m_i^+(t_X))$$

a celková pravděpodobnost zamítnutí se tedy rovná

$$\begin{aligned} P_{zam} &\doteq \int_{A_1} h(t_X) \cdot [\Phi(m_1^-(t_X)) + 1 - \Phi(m_1^+(t_X))] dt_X + \\ &+ \int_{A_2} h(t_X) \cdot [\Phi(m_2^-(t_X)) + 1 - \Phi(m_2^+(t_X))] dt_X. \end{aligned}$$



Obrázek 4.2: Množiny rozdělení, na kterých platí nulové hypotézy.

4.2 Asymptotická hladina

Pokud dosadíme do vzorce pro pravděpodobnost zamítnutí P_{zam} na množinách A a B za některé hodnoty μ_1, μ_2, μ_X nuly podle tabulky 4.1, dostaneme skutečnou hladinu dvoufázové procedury $\max(P_{zam}(A), P_{zam}(B))$.

Za podmínky, že testové statistiky T_X a T_1 jsou nekorelované a zároveň T_X a T_2 jsou nekorelované, tedy $\rho_1 = 0$ a $\rho_2 = 0$, dostaneme

$$P_{zam}(A) = (1 - \alpha_X) \cdot \alpha + \alpha_X \cdot \alpha = \alpha.$$

$$P_{zam}(B) = (\beta_X) \cdot \alpha^* + (1 - \beta_X) \cdot \alpha.$$

kde β_X je pravděpodobnost chyby druhého druhu testu T_X a α^* je pravděpodobnost zamítnutí hypotézy H_0^* testem T_2 při nesplnění předpokladu H_X . α^* tedy může být i větší číslo než α . Ovšem pro velká n se síla $(1 - \beta_X)$ testu T_X bude blížit k jedné a tedy hladina procedury bude asymptoticky dodržena i na množině B .

množina	H_X	H_0	H_0^*	μ_X	μ_1	μ_2
A	v	v	v	0	0	0
B	x	v	v	μ_X	0	0
C	v	x	v	0	μ_1	0
D	x	x	v	μ_X	μ_1	0
E	v	v	x	0	0	μ_2
F	x	v	x	μ_X	0	μ_2
G	v	x	x	0	μ_1	μ_2
H	x	x	x	μ_X	μ_1	μ_2

Tabulka 4.1: Nulovost parametrů μ_1, μ_2, μ_X na jednotlivých množinách obrázku 4.2.

Nulovost korelačních koeficientů ρ_1 a ρ_2 je také jediná obecná postačující podmínka pro dodržení hladiny procedury na množině $\mathcal{G}_0^* \cap \mathcal{G}_0$, tedy na množinách A a B .

Příkladem relativně vysoké korelace testových statistik může být simulace z šikmého normálního rozdělení v kapitole 5.4. Korelace t-statistiky a statistiky Moodyho testu je zde vyšší než 0.5 pro parametr šikmosti $a_1 = a_2 = 4$.

Příkladem, kdy má korelace testových statistik vliv na výslednou hladinu procedury, je simulace ze dvou shodných exponenciálních rozdělení v kapitole 5.4. Korelace t-statistiky a statistiky Moodyho testu zde přesahuje 0.5 pro všechna zadaná nastavení parametrů (To je dáno tím, že střední hodnota i rozptyl exponenciálního rozdělení jsou určeny stejným parametrem). Hladiny obou testů se pohybují kolem 5 %, ale výsledná hladina procedury je jen 4 %.

Na množinách C, D, E, F platí vždy právě jedna z hypotéz H_0 a H_0^* a není tedy jasné co je žádoucí výsledek. Procedura na těchto množinách ve výsledku neotestuje ani jednu hypotézu.

Na množinách C, E platí hypotéza H_X a procedura se bude spíše řídit testem T_2 , tedy na množině C bude přibližně dodržovat hladinu α a na množině E bude pravděpodobnost zamítnutí blízká síle testu T_2 .

Na množinách D, F hypotéza H_X neplatí a procedura se bude s rostoucím počtem pozorování spíše řídit testem T_1 , tedy na množině F se pravděpodobnost zamítnutí bude blížit hladině α a na množině D bude pravděpodobnost zamítnutí blízká síle testu T_1 .

4.3 Asymptotická síla

Pokud dosadíme do vzorce pro pravděpodobnost zamítnutí P_{zam} na množinách G a H z obrázku 4.2 za některé hodnoty μ_1, μ_2, μ_X nuly podle tabulky 4.1, dostaneme sílu dvoufázové procedury $\min(P_{zam}(G), P_{zam}(H))$.

Za podmínky, že testové statistiky T_X a T_1 jsou nekorelované a zároveň T_X a T_2 jsou nekorelované, tedy $\rho_1 = 0$ a $\rho_2 = 0$, dostaneme

$$P_{zam}(G) = (1 - \alpha_X) \cdot (1 - \beta_2) + \alpha_X \cdot (1 - \beta_1).$$

$$P_{zam}(H) = (\beta_X) \cdot (1 - \beta_2^*) + (1 - \beta_X) \cdot (1 - \beta_1).$$

kde β_X je pravděpodobnost chyby druhého druhu testu T_X , β_i je pravděpodobnost chyby druhého druhu testu T_i a β_2^* je pravděpodobnost zamítnutí testem T_2 při nesplnění předpokladu H_X .

Kapitola 5

Simulace

V této kapitole ukážeme na simulacích chování několika dvoufázových procedur pro různá rozdělení. Všechny simulace byly provedeny v programu R [8]. Pokud není uvedeno jinak, počet opakování je 3000. Nejdříve popíšeme šikmé normální rozdělení, které je v některých simulacích použito.

5.1 Šikmé normální rozdělení

Toto rozdělení zavedl Adelchi Azzalini a je popsáno v Azzalini [2]. K simulacím z tohoto rozdělení byla v programu R použita knihovna SN [3].

Definice 5.1. Řekneme, že Y má šikmé normální rozdělení SN (ξ, ω^2, α) , pokud ho lze zapsat ve tvaru $Y = \xi + \omega X$, kde X má hustotu $f(X) = 2\varphi(X) \cdot \Phi(\alpha X)$, $\varphi(X)$ je hustota a $\Phi(X)$ je distribuční funkce normálního rozdělení $N(0, 1)$.

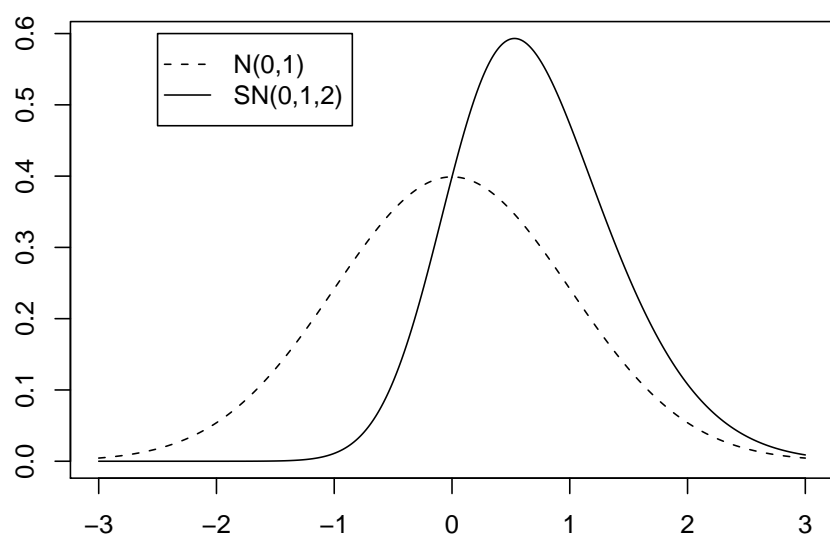
Na obrázku 5.1 můžeme vidět graf hustoty rozdělení SN $(0, 1, 2)$.

Věta 5.2. Platí

$$\begin{aligned} E X &= \sqrt{2/\pi} \cdot \delta \\ \text{var } X &= 1 - 2\delta^2/\pi \\ E Y &= \xi + \omega\sqrt{2/\pi} \cdot \delta \\ \text{var } Y &= \omega^2(1 - 2\delta^2/\pi), \end{aligned}$$

kde

$$\delta = \frac{\alpha}{\sqrt{1 + \alpha^2}}.$$



Obrázek 5.1: Hustota šikmého normálního rozdělení $SN(0, 1, 2)$.

5.2 Předpoklad normality - 1 výběr

Test hypotézy o střední hodnotě jednoho výběru provedeme pomocí dvoufázové procedury, kdy použijeme T-test nebo Wilcoxonův test podle výsledku Shapiro-Wilkova testu normality. Na základě poměru zamítnutí určíme přibližnou hladinu a sílu této testovací procedury v jednotlivých případech a porovnáme oba testy střední hodnoty.

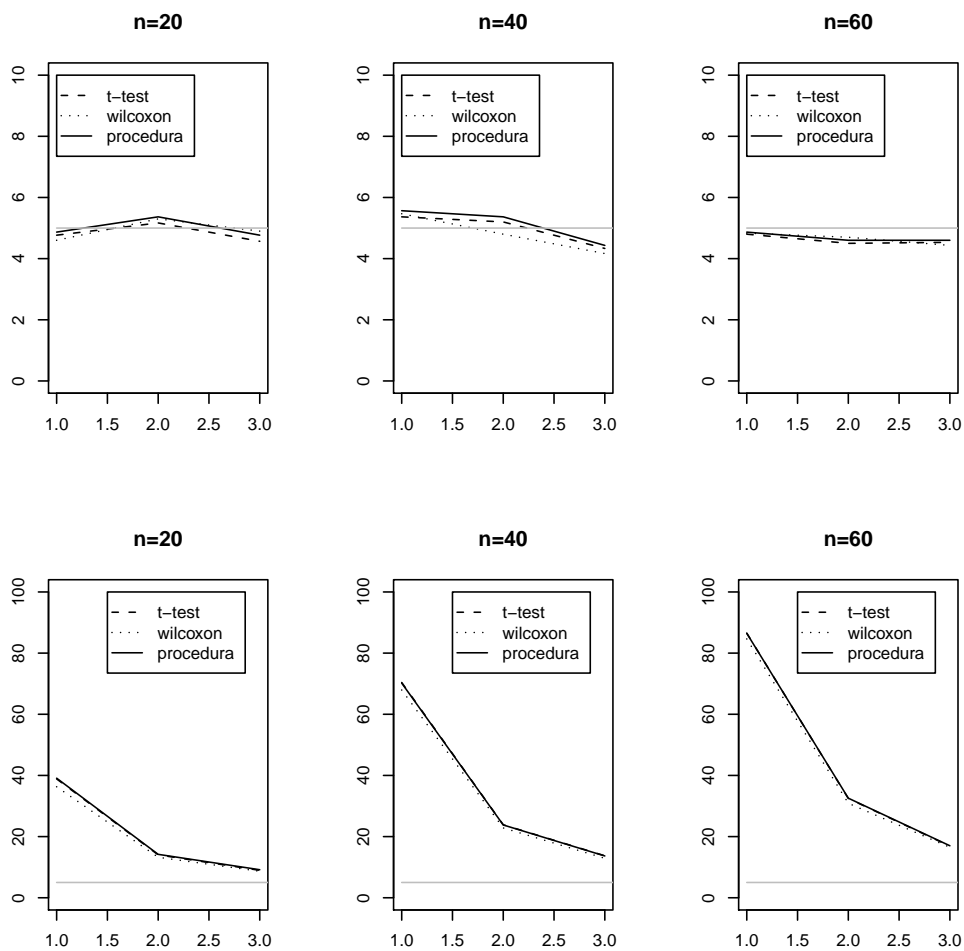
Normální rozdělení

Popis simulace

Simulujeme n pozorování ($n = 20, 40, 60$) z normálního rozdělení $N(\mu, \sigma)$. Budeme testovat hypotézu $H_0: E X = 0$, proti oboustranné alternativě nejdříve na datech se skutečnou střední hodnotou $E X = 0$ pro zjištění hladiny procedury, poté na datech se střední hodnotou $E X = 0.4$ pro zjištění síly.

Výsledky

Na obrázku 5.2 v prvním řádku vidíme, že T-test i Wilcoxonův test přibližně dodržují pětiprocentní hladinu a tedy i celá procedura se chová korektně. Na druhém řádku vidíme, že Wilcoxonův test má téměř stejnou sílu jako t-test, i když procedura rozhoduje hlavně podle t-testu.



Obrázek 5.2: Hladina a síla procedury pro jeden výběr z normálního rozdělení, kde $EX = 0$ v prvním řádku a $EX = 0.4$ ve druhém řádku pro $\sigma = 1, 2, 3$.

T-rozdělení s malým počtem stupňů volnosti

Popis simulace

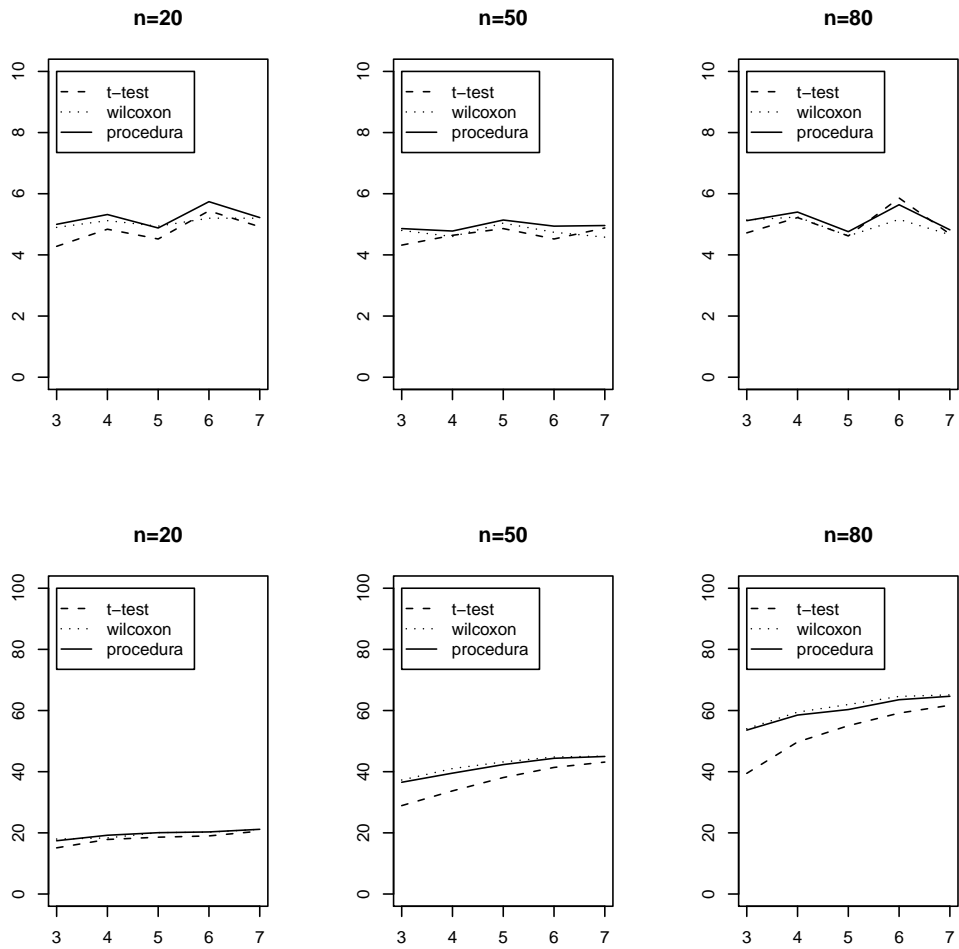
Simulujeme 1 výběr n pozorování ($n = 20, 50, 80$) z t-rozdělení o 3, 4, 5, 6 a 7 stupních volnosti. Pro větší počet stupňů volnosti se t-rozdělení blíží normovanému normálnímu rozdělení $N(0, 1)$ a tedy jsou to případy pro nás nezájímavé. V této simulaci byl použit počet opakování 5000.

Nejdříve budeme testovat hypotézu $H_0: E X = 0$, proti oboustranné alternativě pro zjištění hladiny. Dále budeme testovat hypotézu $H_1: E X = 0.3$ na výběru ze stejného rozdělení pro určení síly.

Výsledky

Jak je vidět z obrázku 5.3, T-test i Wilcoxonův test přibližně dodržují pětiprocentní hladinu a tedy i celá procedura se chová korektně nezávisle na počtu stupňů volnosti.

Pro malý počet pozorování mají t-test i Wilcoxonův test proti zvolené alternativě přibližně stejnou sílu. Pro větší počet pozorování a malý počet stupňů volnosti má t-test výrazně menší sílu než Wilcoxonův test. Dochází ale také k častějšímu zamítnutí hypotézy normality, proto procedura jako celek má sílu srovnatelnou s Wilcoxonovým testem.



Obrázek 5.3: Hladina (v prvním řádku) a síla (v druhém řádku) procedury pro výběr z t-rozdělení o 3-7 stupních volnosti.

Šikmé normální rozdělení

Popis simulace

Simulujeme 1 výběr n pozorování ($n = 20, 40, 60$) z šikmého normálního rozdělení SN (ξ, ω^2, α) .

Budeme testovat hypotézu $H_0: E X = 0$, proti oboustranné alternativě na výběrech, kde ve skutečnosti:

1. $E X = 0$
2. medián $\tilde{x} = 0$
3. $E X = 0.3$.

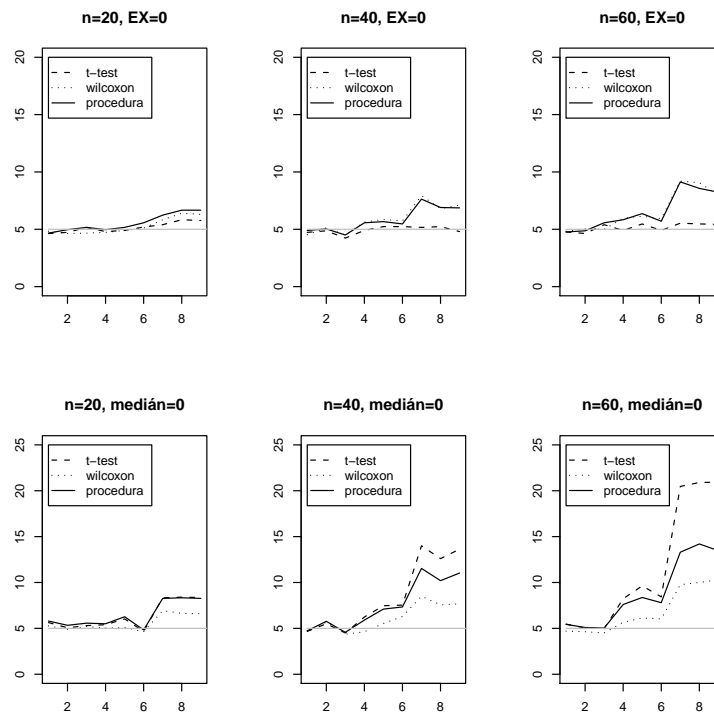
Výsledky

Na obrázku 5.4 v prvním řádku vidíme, že t-test sice dodržuje pětiprocentní hladinu, ale Wilcoxonův test ji pro více šikmé rozdělení překračuje, protože v tomto případě už netestuje stejnou hypotézu. A protože dochází k častému zamítání předpokladu normality, procedura se řídí spíše Wilcoxonovým testem a tedy pětiprocentní hladinu nedodržuje.

V druhém řádku nás nepřekvapí, že t-test zamítá ve více než pěti procentech případů, protože jeho nulová hypotéza $E X = 0$ v tomto případě neplatí. Ovšem pro více šikmá rozdělení hladinu překročí i Wilcoxonův test, přestože skutečný medián výběru je nulový. Tím se potvrzuje důležitost předpokladu symetrie rozdělení zmíněná v kapitole 3.1.

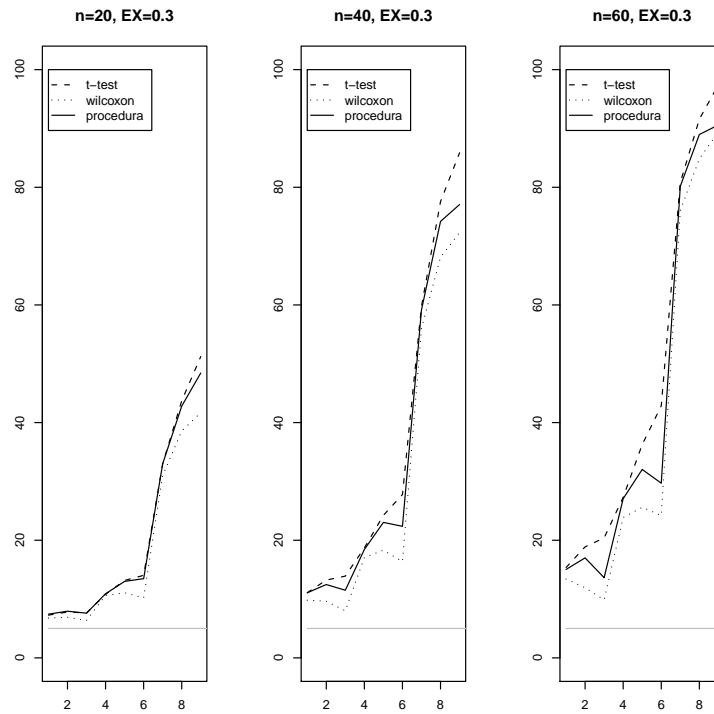
Na obrázcích 5.5 a 5.6 vidíme, že t-test a Wilcoxonův test mají velmi podobnou sílu. Wilcoxonův test má sílu nepatrně vyšší (resp. nižší) podle znaménka parametru šikmosti, tedy v případě, že vzdálenost mediánu od nuly je větší (resp. menší) než vzdálenost střední hodnoty od nuly.

	ω	α	$\text{var } X$
1	1.00	1.00	0.68
2	2.00	1.00	2.73
3	3.00	1.00	6.14
4	1.00	2.00	0.49
5	2.00	2.00	1.96
6	3.00	2.00	4.42
7	1.00	4.00	0.40
8	2.00	4.00	1.60
9	3.00	4.00	3.61



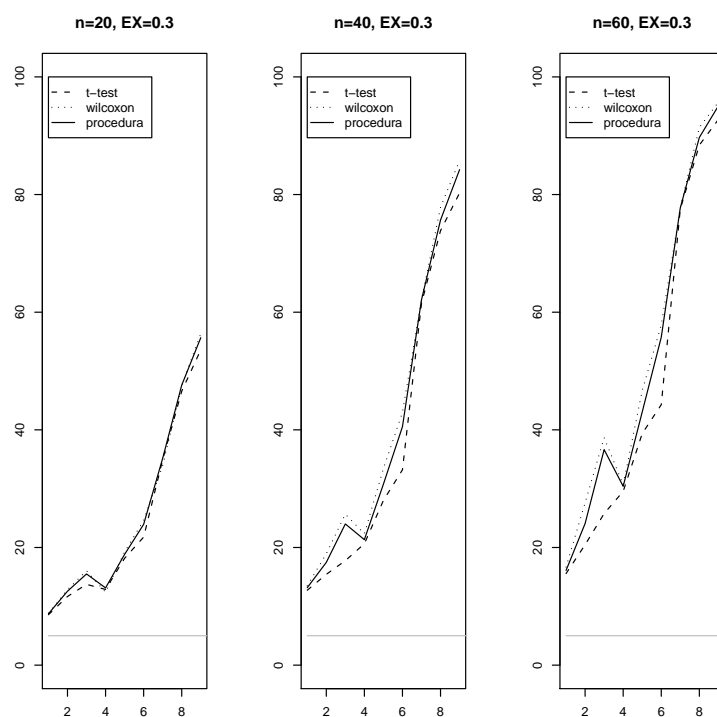
Obrázek 5.4: Hladina procedury pro jeden výběr z šikmého normálního rozdělení, kde $EX = 0$ v prvním řádku a medián $\tilde{x} = 0$ ve druhém řádku. Měnící se parametry jsou uvedeny v tabulce.

	ω	α	EX	medián X	$\text{var } X$
1	3.00	1.00	0.30	0.24	6.14
2	3.00	2.00	0.30	0.12	4.42
3	3.00	4.00	0.30	0.00	3.61
4	2.00	1.00	0.30	0.26	2.73
5	2.00	2.00	0.30	0.18	1.96
6	2.00	4.00	0.30	0.10	1.60
7	1.00	1.00	0.30	0.28	0.68
8	1.00	2.00	0.30	0.24	0.49
9	1.00	4.00	0.30	0.20	0.40



Obrázek 5.5: Síla procedury pro jeden výběr z šikmého normálního rozdělení, kde $EX = 0.3$ a $\alpha > 0$. Měnící se parametry jsou uvedeny v tabulce.

	ω	α	EX	medián X	var X
1	3.00	-1.00	0.30	0.36	6.14
2	3.00	-2.00	0.30	0.47	4.42
3	3.00	-4.00	0.30	0.60	3.61
4	2.00	-1.00	0.30	0.34	2.73
5	2.00	-2.00	0.30	0.42	1.96
6	2.00	-4.00	0.30	0.50	1.60
7	1.00	-1.00	0.30	0.32	0.68
8	1.00	-2.00	0.30	0.36	0.49
9	1.00	-4.00	0.30	0.40	0.40



Obrázek 5.6: Síla procedury pro jeden výběr z šikmého normálního rozdělení, kde $EX = 0.3$ a $\alpha < 0$. Měnící se parametry jsou uvedeny v tabulce.

5.3 Předpoklad normality - 2 výběry

Test hypotézy o rovnosti středních hodnot dvou výběrů provedeme pomocí dvoufázové procedury, kdy použijeme Welschův test nebo Wilcoxonův test podle výsledku Shapiro-Wilkova testu normality. Normalitu testujeme na každém výběru zvlášť na hladině $\alpha = 0.025$, zvolené z Bonferroniho nerovnosti jako požadovaná celková hladina vydělená počtem prováděných testů, tedy $\alpha = 0.05/2$. Na základě poměru zamítnutí určíme přibližnou hladinu a sílu této testovací procedury v jednotlivých případech a porovnáme oba testy střední hodnoty.

Šikmé normální rozdělení

Popis simulace

Simulujeme 2 výběry o n_1 a n_2 pozorováních z šikmého normálního rozdělení $SN(\xi_1, \omega_1^2, \alpha_1)$ a $SN(\xi_2, \omega_2^2, \alpha_2)$. Nejdříve budeme testovat hypotézu $H_0: EX = EY$, proti alternativě $H_1: EX \neq EY$ na datech, pro která platí nulová hypotéza. Dále budeme testovat stejnou hypotézu na datech, kde ve skutečnosti platí $EX = EY - 0.3$.

Výsledky

Rozlišíme tyto tři případy:

1. $n_1 < n_2$ a $\text{var } X > \text{var } Y$
2. $n_1 > n_2$ a $\text{var } X > \text{var } Y$
3. $n_1 = n_2$ a $\text{var } X = \text{var } Y$.

V následujících grafech je popiskem "t-test" označená Welschova varianta t-testu. Podívejme se na hladinu procedury na obrázku 5.7. V prvním případě, kdy má výběr s menším počtem pozorování větší rozptyl, Wilcoxonův test zamítá zhruba v 10 % případů, to je ovšem dáno tím, že testuje shodu rozdělení a ne jen shodu jejich středních hodnot.

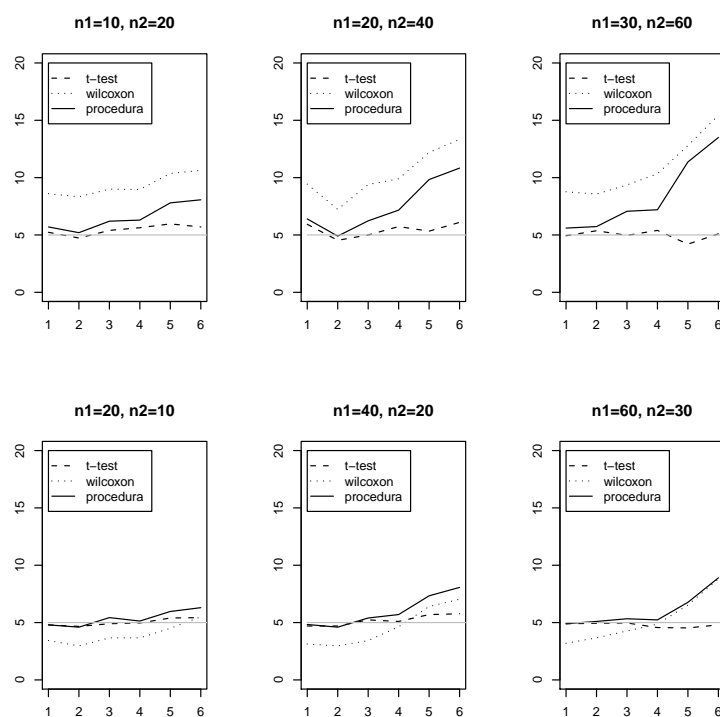
Ve druhém případě, kdy má výběr s menším počtem pozorování menší rozptyl, a ve třetím případě (obrázek 5.8), kdy jsou rozptyly stejné, Wilcoxonův test přibližně dodržuje hladinu. Jen pro větší počet pozorování a větší rozdíl

mediánů u něj procento zamítnutí roste. Welschův test dodržuje požadovanou hladinu 5 % ve všech třech případech.

Chování procedury se ve všech případech blíží chování Wilcoxonova testu a procedura tedy překračuje pětiprocentní hladinu při větším počtu pozorování ve všech třech případech.

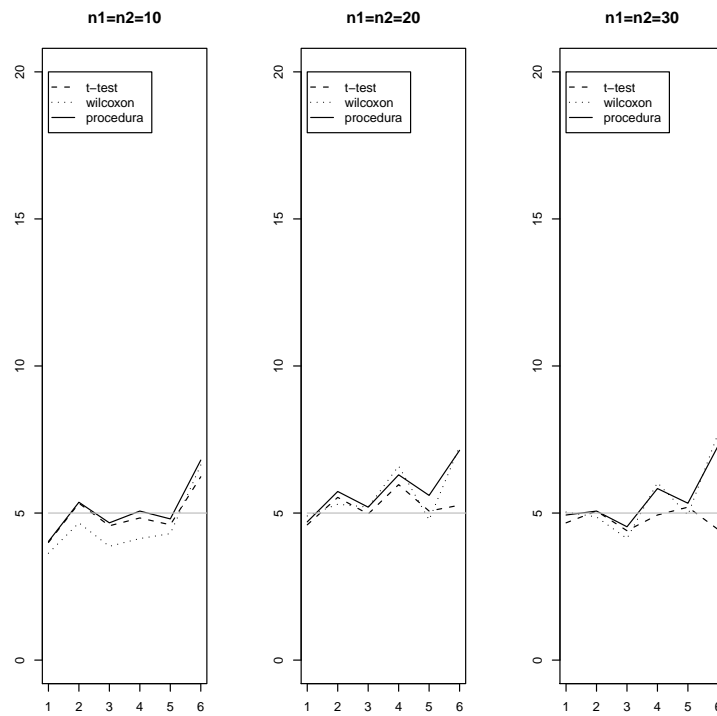
Podívejme se na sílu procedury na obrázcích 5.9, 5.10 a 5.11. Wilcoxonův test má pro větší šikmost a větší počet pozorování také větší sílu než Welschův test. Procedura se v těchto případech kvůli častějšímu zamítání normality řídí hlavně Wilcoxonovým testem, takže její chování je optimální.

	α_1	α_2	$\text{var } X$	$\text{var } Y$	medián X	medián Y
1	1.00	1.00	0.68	0.17	-0.02	-0.01
2	1.00	-1.00	0.68	0.17	-0.02	0.01
3	2.00	2.00	0.49	0.12	-0.06	-0.03
4	2.00	-2.00	0.49	0.12	-0.06	0.03
5	4.00	4.00	0.40	0.10	-0.10	-0.05
6	4.00	-4.00	0.40	0.10	-0.10	0.05



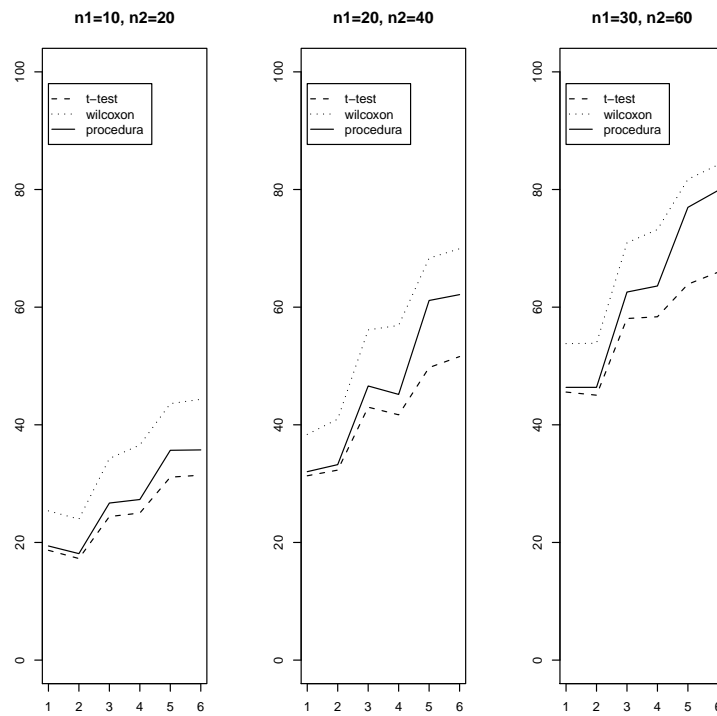
Obrázek 5.7: Hladina procedury pro dva výběry z šikmého normálního rozdělení s parametry $E X = 0, E Y = 0, \omega_1 = 1, \omega_2 = 0.5$ pro velikosti výběrů $n_1 < n_2$ v prvním řádku a $n_1 > n_2$ v druhém řádku. Měnící se parametry jsou uvedeny v tabulce.

	α_1	α_2	var X	var Y	medián X	medián Y
1	1.00	1.00	0.68	0.68	-0.02	-0.02
2	1.00	-1.00	0.68	0.68	-0.02	0.02
3	2.00	2.00	0.49	0.49	-0.06	-0.06
4	2.00	-2.00	0.49	0.49	-0.06	0.06
5	4.00	4.00	0.40	0.40	-0.10	-0.10
6	4.00	-4.00	0.40	0.40	-0.10	0.10



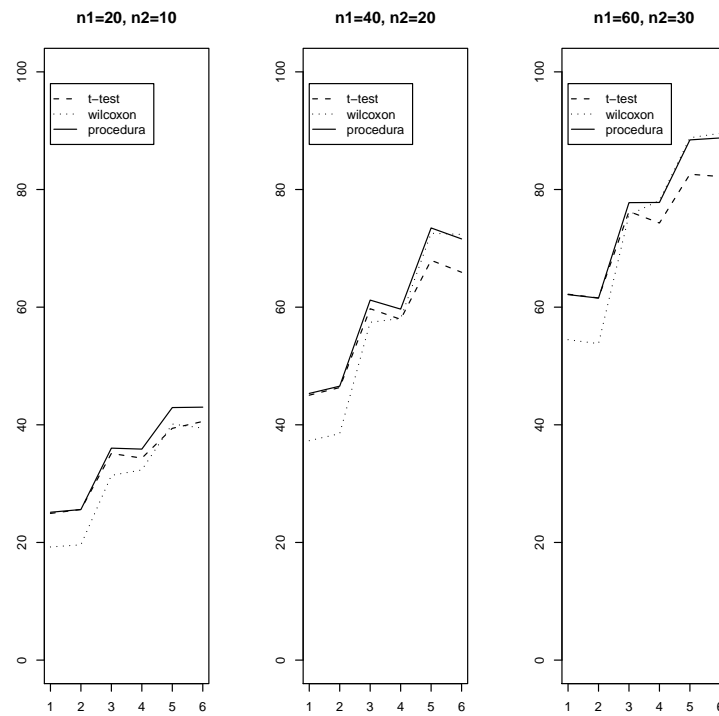
Obrázek 5.8: Hladina procedury pro dva výběry z šikmého normálního rozdělení s parametry $n_1 = n_2$, $EX = 0$, $EY = 0$, $\omega_1 = \omega_2 = 1$. Měnicí se parametry jsou uvedeny v tabulce.

	α_1	α_2	var X	var Y	medián X	medián Y
1	1.00	1.00	0.68	0.17	-0.02	0.29
2	1.00	-1.00	0.68	0.17	-0.02	0.31
3	2.00	2.00	0.49	0.12	-0.06	0.27
4	2.00	-2.00	0.49	0.12	-0.06	0.33
5	4.00	4.00	0.40	0.10	-0.10	0.25
6	4.00	-4.00	0.40	0.10	-0.10	0.35



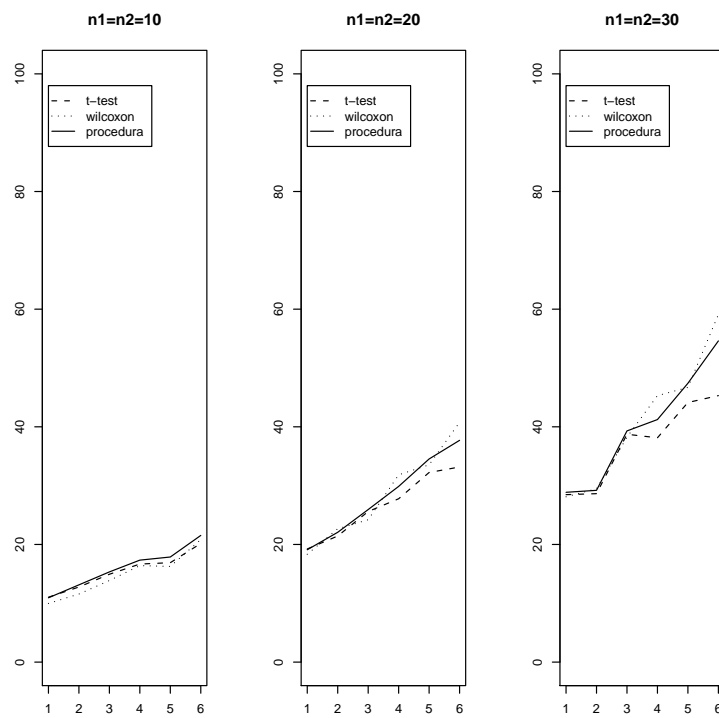
Obrázek 5.9: Síla procedury pro dva výběry z šikmého normálního rozdělení s parametry $n_1 < n_2$, $E X = 0, E Y = 0.3$, $\omega_1 = 1, \omega_2 = 0.5$. Měnící se parametry jsou uvedeny v tabulce.

	α_1	α_2	var X	var Y	medián X	medián Y
1	1.00	1.00	0.68	0.17	-0.02	0.29
2	1.00	-1.00	0.68	0.17	-0.02	0.31
3	2.00	2.00	0.49	0.12	-0.06	0.27
4	2.00	-2.00	0.49	0.12	-0.06	0.33
5	4.00	4.00	0.40	0.10	-0.10	0.25
6	4.00	-4.00	0.40	0.10	-0.10	0.35



Obrázek 5.10: Síla procedury pro dva výběry z šikmého normálního rozdělení s parametry $n_1 > n_2$, $E X = 0, E Y = 0.3$, $\omega_1 = 1, \omega_2 = 0.5$. Měnící se parametry jsou uvedeny v tabulce.

	α_1	α_2	var X	var Y	medián X	medián Y
1	1.00	1.00	0.68	0.17	-0.02	0.29
2	1.00	-1.00	0.68	0.17	-0.02	0.31
3	2.00	2.00	0.49	0.12	-0.06	0.27
4	2.00	-2.00	0.49	0.12	-0.06	0.33
5	4.00	4.00	0.40	0.10	-0.10	0.25
6	4.00	-4.00	0.40	0.10	-0.10	0.35



Obrázek 5.11: Síla procedury pro dva výběry z šikmého normálního rozdělení s parametry $n_1 = n_2$, $EX = 0$, $EY = 0.3$, $\omega_1 = 1$, $\omega_2 = 1$. Měnící se parametry jsou uvedeny v tabulce.

5.4 Předpoklad shody rozptylů - 2 výběry

Test hypotézy o rovnosti středních hodnot dvou výběrů provedeme pomocí dvoufázové procedury, kdy použijeme t-test nebo Wilcoxonův test podle výsledku Moodyho testu shody rozptylů. Na základě poměru zamítnutí určíme přibližnou hladinu a sílu této testovací procedury v jednotlivých případech a porovnáme oba testy střední hodnoty.

Šikmé normální rozdělení

Popis simulace

Simulujeme 2 výběry o n_1 a n_2 pozorováních z šikmého normálního rozdělení $SN(\xi_1, \omega_1^2, \alpha_1)$ a $SN(\xi_2, \omega_2^2, \alpha_2)$.

Nejdříve budeme testovat hypotézu $H_0: EX = EY$, proti alternativě $H_1: EX \neq EY$ na datech, pro která platí nulová hypotéza. Dále budeme testovat stejnou hypotézu na datech, kde ve skutečnosti platí $EX = EY - 0.3$.

Výsledky

Rozlišíme tyto tři případy:

1. $n_1 < n_2$ a $\text{var } X > \text{var } Y$
2. $n_1 > n_2$ a $\text{var } X > \text{var } Y$
3. $n_1 = n_2$ a $\text{var } X = \text{var } Y$.

Podívejme se na hladinu procedury na obrázku 5.12. V prvním případě, kdy má výběr s menším počtem pozorování větší rozptyl, t-test nedodrží požadovanou hladinu 5 %, ale výrazně ji překračuje. Wilcoxonův test zamítá zhruba v 10 % případů, to je ovšem dáno tím, že testuje celkovou shodu rozdělení a ne jen shodu jejich středních hodnot.

V druhém případě, kdy má výběr s menším počtem pozorování menší rozptyl, má t-test procento zamítnutí přibližně 2 %. Wilcoxonův test zamítá zhruba ve 4 % případů, ale pro větší počet pozorování a větší rozdíl mediánů u něj procento zamítnutí roste. Chování procedury se v obou případech blíží chování Wilcoxonova testu a to tím víc, čím je větší počet pozorování, tedy čím častěji je zamítnuta shoda rozptylu Moodyho testem.

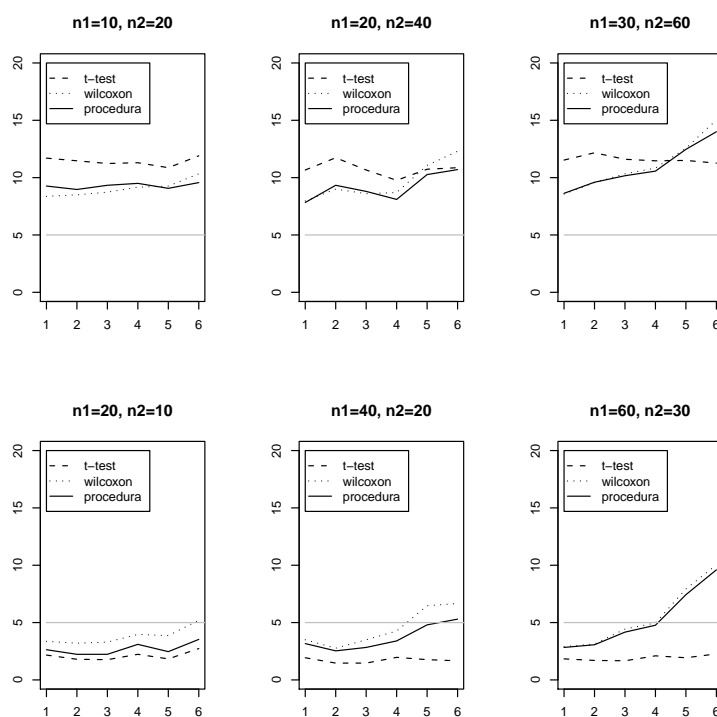
Na obrázku 5.13 vidíme, že v případě shodných rozptylů se procedura řídí hlavně podle t-testu. Pro opačně šikmá rozdělení Wilcoxonův test překračuje hladinu 5 %, protože v tomto případě netestuje rovnost středních hodnot, ale celkovou shodu rozdělení.

Podívejme se na sílu procedury na obrázek 5.15. Pokud má rozdělení s větším počtem pozorování menší rozptyl, t-test má menší větší sílu než Wilcoxonův test jen pro malý počet pozorování, případně pro menší šikmost rozdělení.

Pokud má rozdělení s větším počtem pozorování i větší rozptyl, t-test má menší sílu než Wilcoxonův test, jak je vidět na obrázku 5.15. Procedura se kvůli nerovnosti rozptylů řídila hlavně Wilcoxonovým testem, takže její chování je optimální.

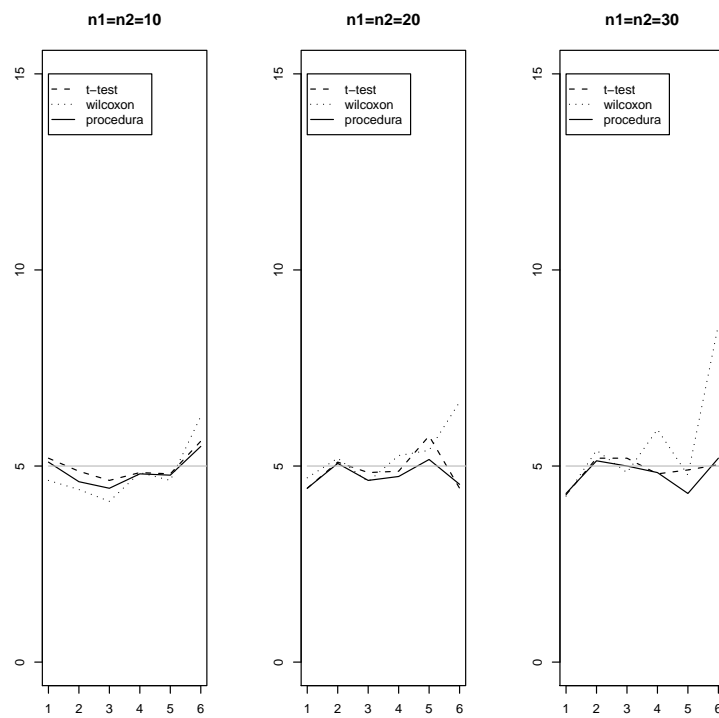
Na obrázku 5.16 vidíme, že v případě shodných rozptylů se procedura řídí hlavně podle t-testu. Pro více šikmá rozdělení má ale o něco větší sílu Wilcoxonův test.

	α_1	α_2	$\text{var } X$	$\text{var } Y$	medián X	medián Y
1	1.00	1.00	0.68	0.17	-0.02	-0.01
2	1.00	-1.00	0.68	0.17	-0.02	0.01
3	2.00	2.00	0.49	0.12	-0.06	-0.03
4	2.00	-2.00	0.49	0.12	-0.06	0.03
5	4.00	4.00	0.40	0.10	-0.10	-0.05
6	4.00	-4.00	0.40	0.10	-0.10	0.05



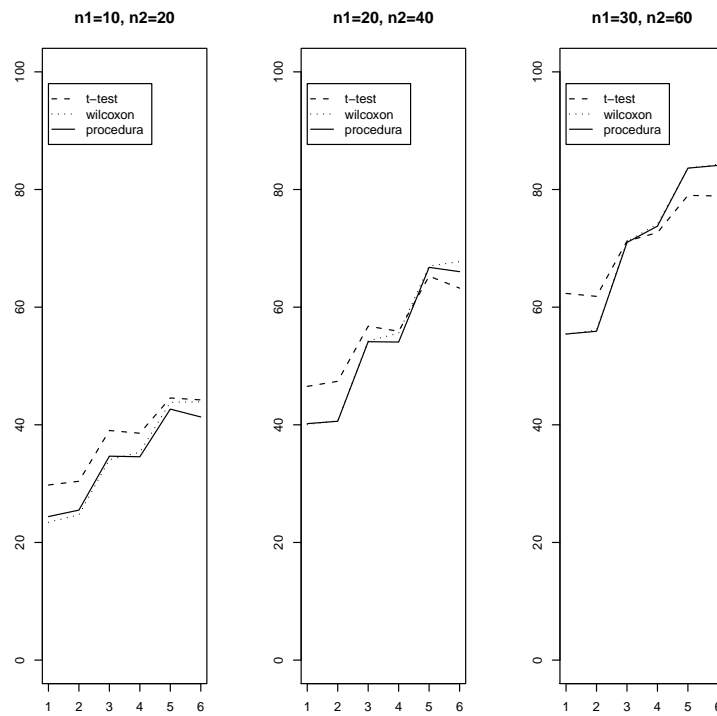
Obrázek 5.12: Hladina procedury pro dva výběry z šikmého normálního rozdělení s parametry $E X = 0, E Y = 0, \omega_1 = 1, \omega_2 = 0.5$ pro velikosti výběrů $n_1 < n_2$ v prvním řádku a $n_1 > n_2$ v druhém řádku. Měňící se parametry jsou uvedeny v tabulce.

	α_1	α_2	var X	var Y	medián X	medián Y
1	1.00	1.00	0.68	0.68	-0.02	-0.02
2	1.00	-1.00	0.68	0.68	-0.02	0.02
3	2.00	2.00	0.49	0.49	-0.06	-0.06
4	2.00	-2.00	0.49	0.49	-0.06	0.06
5	4.00	4.00	0.40	0.40	-0.10	-0.10
6	4.00	-4.00	0.40	0.40	-0.10	0.10



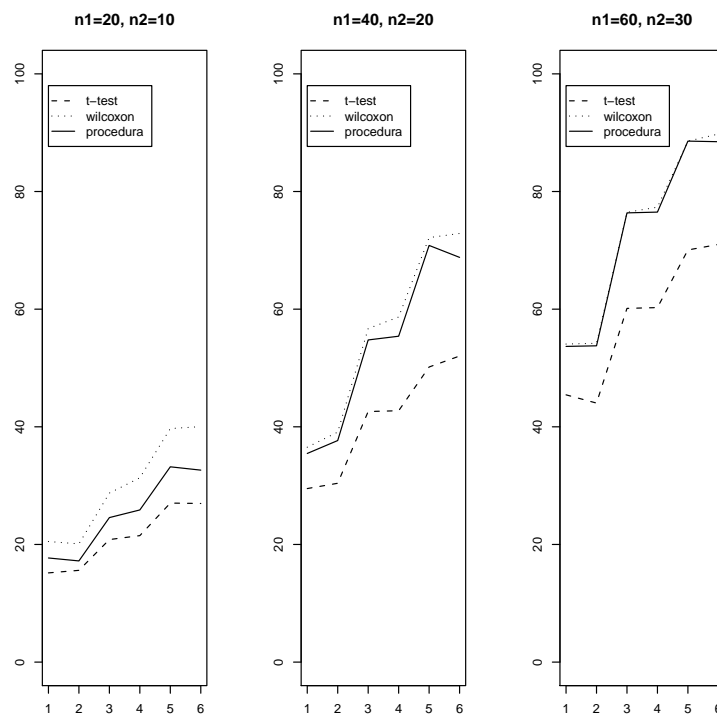
Obrázek 5.13: Hladina procedury pro dva výběry z šikmého normálního rozdělení s parametry $n_1 = n_2$, $EX = 0$, $EY = 0$, $\omega_1 = \omega_2 = 1$. Měnící se parametry jsou uvedeny v tabulce.

	α_1	α_2	var X	var Y	medián X	medián Y
1	1.00	1.00	0.68	0.17	-0.02	0.29
2	1.00	-1.00	0.68	0.17	-0.02	0.31
3	2.00	2.00	0.49	0.12	-0.06	0.27
4	2.00	-2.00	0.49	0.12	-0.06	0.33
5	4.00	4.00	0.40	0.10	-0.10	0.25
6	4.00	-4.00	0.40	0.10	-0.10	0.35



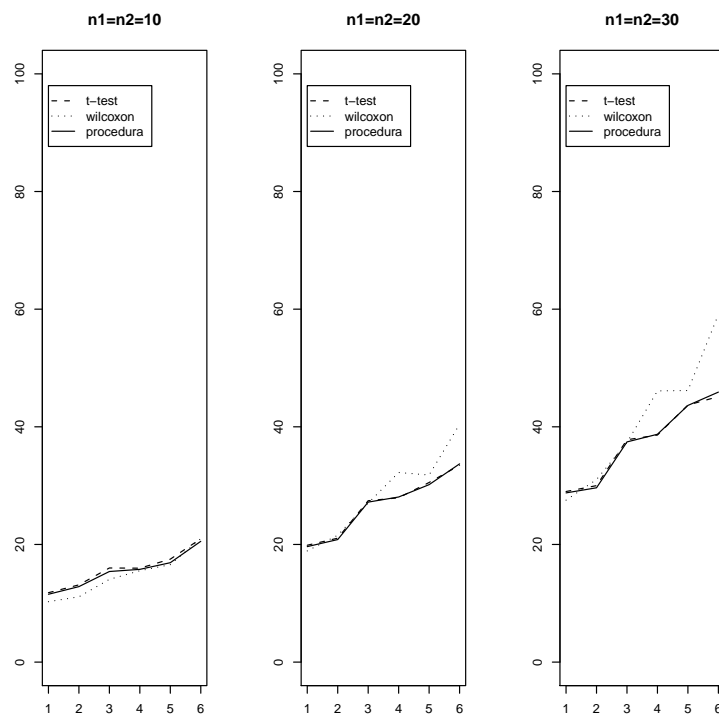
Obrázek 5.14: Síla procedury pro dva výběry z šikmého normálního rozdělení s parametry $n_1 < n_2$, $E X = 0, E Y = 0.3$, $\omega_1 = 1, \omega_2 = 0.5$. Měnící se parametry jsou uvedeny v tabulce.

	α_1	α_2	var X	var Y	medián X	medián Y
1	1.00	1.00	0.68	0.17	-0.02	0.29
2	1.00	-1.00	0.68	0.17	-0.02	0.31
3	2.00	2.00	0.49	0.12	-0.06	0.27
4	2.00	-2.00	0.49	0.12	-0.06	0.33
5	4.00	4.00	0.40	0.10	-0.10	0.25
6	4.00	-4.00	0.40	0.10	-0.10	0.35



Obrázek 5.15: Síla procedury pro dva výběry z šikmého normálního rozdělení s parametry $n_1 > n_2$, $E X = 0, E Y = 0.3$, $\omega_1 = 1, \omega_2 = 0.5$. Měnící se parametry jsou uvedeny v tabulce.

	α_1	α_2	var X	var Y	medián X	medián Y
1	1.00	1.00	0.68	0.68	-0.02	0.28
2	1.00	-1.00	0.68	0.68	-0.02	0.32
3	2.00	2.00	0.49	0.49	-0.06	0.24
4	2.00	-2.00	0.49	0.49	-0.06	0.36
5	4.00	4.00	0.40	0.40	-0.10	0.20
6	4.00	-4.00	0.40	0.40	-0.10	0.40



Obrázek 5.16: Síla procedury pro dva výběry z šikmého normálního rozdělení s parametry $n_1 = n_2$, $EX = 0$, $EY = 0.3$, $\omega_1 = 1$, $\omega_2 = 1$. Měnící se parametry jsou uvedeny v tabulce.

Exponenciální rozdělení

Popis simulace

Simulujeme 2 výběry o n_1 a n_2 pozorováních z exponenciálního rozdělení $\text{Ex}(\lambda_1)$ a $\text{Ex}(\lambda_2)$. Platí tedy $\text{E}X = \lambda_1$ a $\text{E}Y = \lambda_2$.

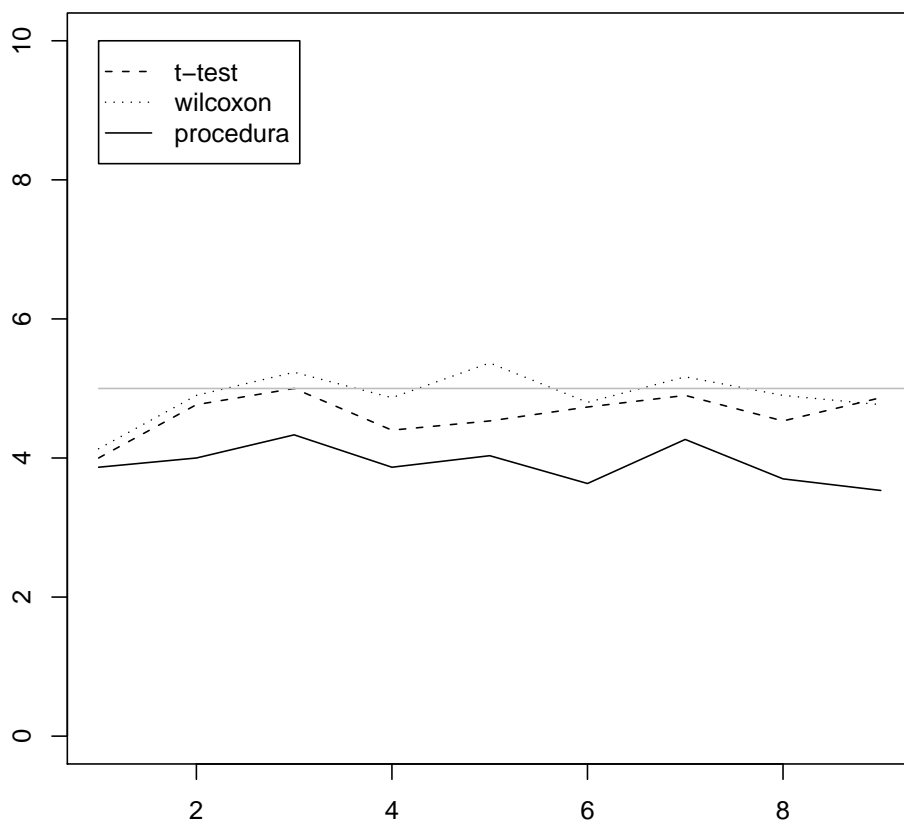
Nejdříve budeme testovat hypotézu $H_0: \text{E}X = \text{E}Y$, proti alternativě $H_1: \text{E}X \neq \text{E}Y$ na datech, pro která platí nulová hypotéza. Dále budeme testovat stejnou hypotézu na datech, kde ve skutečnosti platí $\text{E}X = \text{E}Y - 0.5$ a $\text{E}X = \text{E}Y - 1$. Testy hypotéz provedeme pomocí dvoufázové procedury, v níž použijeme t-test nebo Wilcoxonův test podle výsledku Moodyho testu shody rozptylů.

Výsledky

Testové statistiky Moodyho testu a t-testu i Moodyho testu a Wilcoxonova testu byly v tomto případě výrazně korelované, korelace byla přes 0.5. To je samozřejmě dáno tím, že střední hodnota i rozptyl exponenciálního rozdělení jsou určeny jedním parametrem. Na obrázku 5.17 vidíme, že oba testy sice přibližně dodržují hladinu 5 %, ale hladina výsledné procedury je v důsledku velké korelace testových statistik jen 4 %, tedy nižší než u každého testu zvlášť.

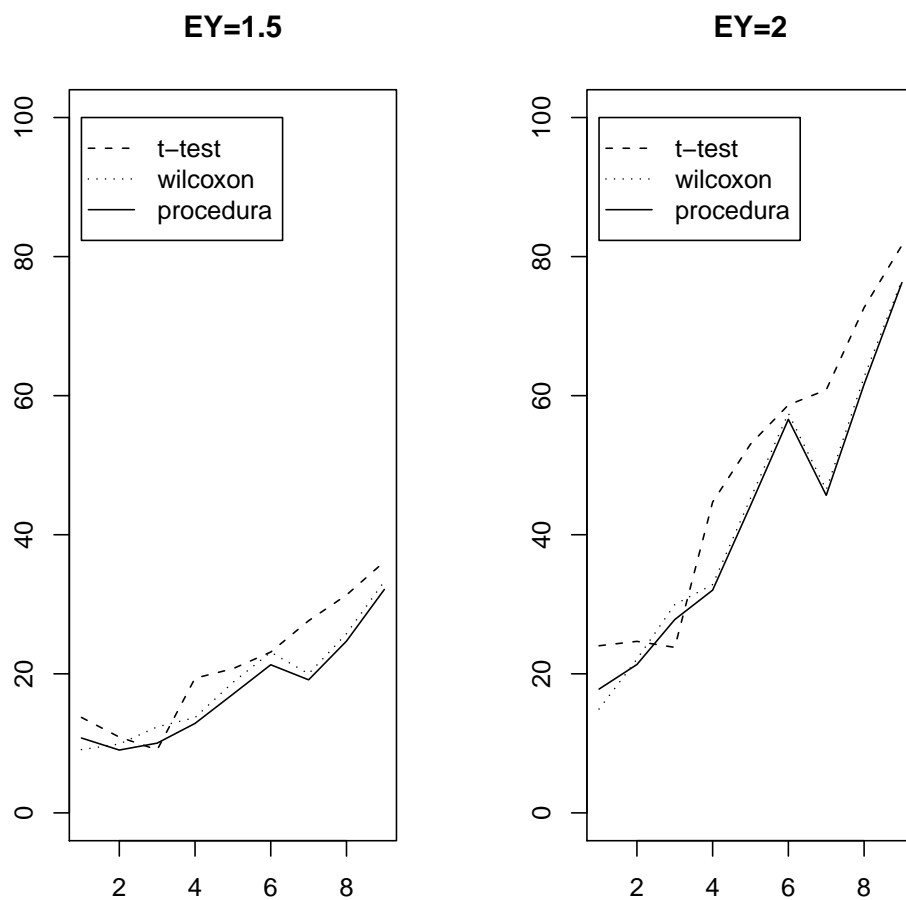
Na obrázku 5.18 vidíme, že síla t-testu je v těchto případech exponenciálního rozdělení větší než síla Wilcoxonova testu a že síla procedury je téměř totožná se silou Wilcoxonova testu, protože byl častěji vybírán procedurou kvůli nerovnosti rozptylů. Procedura tedy častěji volila test s menší silou.

	1	2	3	4	5	6	7	8	9
n_1	10	10	10	20	20	20	30	30	30
n_2	5	10	20	10	20	40	15	30	60



Obrázek 5.17: Hladina procedury pro dva výběry z exponenciálního rozdělení s parametry $E X = 0$, $E Y = 0$. Velikosti výběrů jsou uvedeny v tabulce.

	1	2	3	4	5	6	7	8	9
n_1	10	10	10	20	20	20	30	30	30
n_2	5	10	20	10	20	40	15	30	60



Obrázek 5.18: Síla procedury pro dva výběry z exponenciálního rozdělení s parametry $EX = 1, EY = 1.5$ (nalevo) a $EX = 1, EY = 2$ (napravo). Velikosti výběrů jsou uvedeny v tabulce.

Kapitola 6

Závěr

V této práci jsme popsali dvoufázové testovací procedury, uvedli jejich nejčastější příklady a odvodili jsme jejich hladinu a sílu. Zjistili jsme, že dvoufázové procedury hladinu dodržují jen pokud se shodují nulové hypotézy obou použitých testů a jen pro nekorelované statistiky testu hypotézy a testu předpokladu. Ukázali jsme také, že existuje mnoho případů, kdy každý test testuje jinou hypotézu a ve výsledku pak není jasné jakou hypotézu jsme vlastně zamítli nebo přijali.

Na simulacích jsme ukázali, že předpoklad normality má smysl testovat jen pro malé výběry, kde ale mají testy normality malou sílu. Pro velké výběry se předpoklad normality testuje lépe, ale t-test i Welschův test jsou vůči porušení tohoto předpokladu robustní. V případě jednoho výběru lze dvoufázovou proceduru doporučit jen pro symetrická rozdělení, jinak můžeme dostat velmi zkreslené výsledky. Pokud se rozhodujeme podle shody rozptylů dvou výběrů, musíme si dát pozor na výběr testu pro různé rozptyly. Pokud použijeme Wilcoxonův test, celková procedura hladinu výrazně překročí, hlavně v případě, kdy má výběr s větším počtem pozorování menší rozptyl.

Význam práce spočívá v poznatku, že při testování hypotéz nelze jen slepě následovat zjednodušené návody a diagramy ze statistických příruček, ale vždy je nutné vzít úvahu povahu konkrétních dat, využít grafů a vlastních zkušeností. Pokud se rozhodneme pro použití dvoufázové testovací procedury, je nutné pečlivě zvážit, jaké testy v ní použijeme a jaké předpoklady jsou nezbytné pro jejich správné fungování.

Literatura

- [1] Anděl J. (2002): Základy matematické statistiky. Matfyzpress, Praha.
- [2] Azzalini A.(1985): A class of distributions which includes the normal ones. Scand.J.Statist. 12, 171-178.
- [3] Azzalini A. (2009): R package sn: The skew-normal and skew-t distributions (version 0.4-12). Universita di Padova, Italia.
- [4] Conover W.J. (1999): Practical nonparametric statistics, third edition. John Wiley & Sons , New York.
- [5] Dowdy S. and Wearden S. (1991): Statistics for research. John Wiley & Sons , New York.
- [6] Hendl J. (2004): Přehled statistických metod zpracování dat, Portál, Praha.
- [7] Laubscher, N.F., Steffens, F.E., and DeLange E.M. (1968): Exact critical values for mood's distribution-free test statistic for dispersion and its normal approximation. Technometrics, 10, 497-508.
- [8] R Development Core Team (2005): R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
- [9] Zvára K. (2004): Biostatistika. Karolinum, Praha.