Univerzita Karlova v Praze
Matematicko-fyzikální fakulta

BAKALÁŘSKÁ PRÁCE

Michaela Kurková

Dvouvýběrový T-test v případě nestejných rozptylů

Katedra pravděpodobnosti a matematické statistiky

Vedoucí bakalářské práce: Ing. Marek Omelka
Studijní program: Obecná matematika, Matematická statistika

2006

Děkuji Ing. Markovi Omelkovi za cenné rady při psaní bakalářské práce, trpělivost a ochotu.

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

V Praze dne 23.5.2006 Michaela Kurková

Obsah

Název práce: Dvouvýběrový T-test v případě nestejných rozptylů
Autor: Michaela Kurková
Katedra (ústav): Katedra pravděpodobnosti a matematické statistiky
Vedoucí bakalářské práce: Ing. Marek Omelka
e-mail vedoucího: omelka@karlin.mff.cuni.cz

Abstrakt: V této práci se zabýváme problémem testování shody průměrů dvou nezávislých náhodných výběrů, které mají normální rozdělení. Zaměřujeme se především na případ, kdy se rozptyly těchto výběrů neshodují. Zjišťujeme asymptotické vlastnosti dané testové statistiky a na základě toho určujeme, jak je ovlivněn výsledek testu, pokud jsou velikosti výběrů nestejné a zároveň jsou rozptyly různé. Popíšeme metodu, na které jsou založeny modifikace t-testu, z nichž konkrétně uvedeme Welchův a Satterthwaiteův test. Na závěr uvedeme jedno moderní řešení, známé pod pojmem bootstrap.

Klíčová slova: dvouvýběrový t-test, Behrens-Fisherův problém, Welchův test

Title: Two-sample T-test in the case of unequal variances
Author: Michaela Kurkova
Department: Department of Probability and Mathematical Statistics
Supervisor: Ing. Marek Omelka
Supervisor's e-mail address: omelka@karlin.mff.cuni.cz

Abstract: In this work we are concerned with the problem of testing the equality of the means from the two independent normally distributed populations. We concentrate mainly on situation when the population variances are unequal. We study the asymptotic characteristics of test criterion T and on the basis of these results we determine robustness of the two-sample t-test when the sample size are unequal. In this situation it is appropriate the Welch test or Satterthwaite test, eventually the bootstrap test, which can be combined with Welch test.

Keywords: two-sample t-test, Behrens-Fisher problem, Welch test

Kapitola 1

Úvod

Řešíme situaci, kdy máme dva nezávislé náhodné výběry o velikosti m, n, které mají normální rozdělení se středními hodnotami μ_1, μ_2 a rozptyly σ_1^2, σ_2^2. Za situace, kdy se rozptyly neshodují, testujeme nulovou hypotézu H_0, že $\mu_1 = \mu_2$ na předem stanovené hladině. Tento problém je též známý jako Behrens-Fisherův.

Pro dvouvýběrový t-test jsou nezbytné předpoklady nezávislosti, normality a shody rozptylů, aby byla zaručena spolehlivost testu. Nicméně to se stává v praxi zřídka. Budeme se tedy zabývat otázkou, jak porušení těchto předpokladů, zejména shody rozptylů, ovlivňuje spolehlivost dvouvýběrového t-testu. V literatuře se odpovědi na tyto otázky různí. Můžeme najít příklady, viz Bradley (1980), kdy je test značně nerobustní. Nicméně simulace ukazují, že především pro výběry stejných velikostí je t-test dostatečně robustní.

Pro případ, kdy neznáme rozptyly, se používají různé modifikace t-testu. V této práci se budeme blíže zabývat dvěma modifikacemi, známými jako Satterthwaiteův a Welchův test. Nedoporučuje se rozhodovat, zda použít dvouvýběrový t-test nebo nějakou z jeho modifikací na základě předběžného testu rozptylů (viz Moser, Stevens, 1992).

Nakonec zmíníme další ze způsobů, jak testovat nulovou hypotézu. Kombinaci bootstrapové metody s Welchovým testem.

Kapitola 2

Dvouvýběrový t-test

Nechť $X_1, \dots, X_m$ je náhodný výběr z $\mathrm{N}(\mu_1, \sigma_1^2)$ a nechť $Y_1, \dots, Y_n$ je náhodný výběr z $\mathrm{N}(\mu_2, \sigma_2^2)$. Nechť přitom $m \geq 2, n \geq 2, \sigma_1^2 > 0, \sigma_2^2 > 0$. Předpokládejme, že oba výběry jsou na sobě nezávislé a $\sigma_1^2 = \sigma_2^2$. Označme

$$\overline{X} = \frac{1}{m}\sum_{i=1}^{m} X_i, \qquad \overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

$$S_X^2 = \frac{1}{m-1}\sum_{i=1}^{m}(X_i - \overline{X})^2 \tag{2.1}$$

$$S_Y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \overline{Y})^2 \tag{2.2}$$

Potom platí, že náhodná veličina

$$T = \frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{\sqrt{(m-1)S_X^2 + (n-1)S_Y^2}} \sqrt{\frac{m\,n(m+m-2)}{m+n}}$$

má rozdělení t_{m+n-2} (viz Anděl, 1998, str.87).

Chceme testovat hypotézu $H_0 : \mu_1 = \mu_2$ proti její alternativě $H_1 : \mu_1 \neq \mu_2$. Jestliže je $|T| \geq t_{m+n-2}(\alpha)$ zamítáme hypotézu H_0 na zvolené hladině α.

Je známo, že test založený na statistice T je za daných předpokladů "nejlepší" (přesněji stejnoměrně nejsilnější nestranný), jak uvádí Jurečková (1982), str. 80. Optimalita tohoto testu je těsně svázána s předpokladem normality a shodnosti rozptylů. Zabývejme se nyní otázkou, jak se chová T v případě porušení těchto předpokladů.

Posten (1984) poukazuje na to, že můžeme v různé literatuře najít nejasnosti týkající se podmínek, při kterých jsou testy robustní a při kterých ne. Například některé články ukazují, že dvouvýběrový t-test je zcela robustní vzhledem k odchylkám od normality i vzhledem k odchylkám od shody rozptylů (zejména pokud jsou velikosti výběrů stejné nebo téměř stejné). Na druhé straně můžeme najít případy, kdy je t-test nerobustní. Jedná se však o situace, kdy je rozdělení již velmi vzdáleno normálnímu rozdělení. Přes tyto výjimky, pokud nedojde u náhodných výběrů k extrémním odchylkám od normality, simulace ukazují, že dvouvýběrový t-test není příliš citlivý na odchylky od normality.

2.1 Vlastnosti testové statistiky T v případě porušení předpokladu shody rozptylů

Pro základní představu o chování statistiky T v případě nestejných rozptylů nejprve zjistíme, jaké rozdělení bude mít asymptoticky.
Víme, že

$$\frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}} \sim N(0,1)$$

Podle Dupač, Hušková (2003), Věta 5.1. a Věta 5.9. platí, že $S_X^2 \xrightarrow{s.j.} \sigma_1^2$ a proto $\frac{S_X^2}{\sigma_1^2} \xrightarrow{s.j.} 1$. Obdobně $\frac{S_Y^2}{\sigma_2^2} \xrightarrow{s.j.} 1$.
Předpokládejme, že $m, n \to \infty$ a $\frac{m}{n} \to \lambda$. Potom

$$\frac{(m-1)S_X^2 + (n-1)S_Y^2}{m+n-2} = \frac{(m-1)S_X^2}{m+n-2} + \frac{(n-1)S_Y^2}{m+n-2} \xrightarrow{s.j.} \frac{\lambda}{\lambda+1}\sigma_1^2 + \frac{1}{\lambda+1}\sigma_2^2$$

a dále

$$\frac{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}{\frac{m+n}{mn}} \to \frac{\sigma_1^2 + \lambda\sigma_2^2}{\lambda+1}$$

S využitím předchozího a Věty 4.14 z Dupač, Hušková (2003) dostáváme

$$\begin{aligned}
T &= \frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{(m-1)S_X^2 + (n-1)S_Y^2}{m+n-2}}\sqrt{\frac{m+n}{mn}}} \\
&= \frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}} \frac{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}}{\sqrt{\frac{(m-1)S_X^2 + (n-1)S_Y^2}{m+n-2}}\sqrt{\frac{m+n}{mn}}} \\
&\xrightarrow{d} N\left(0, \frac{\sigma_1^2 + \lambda\sigma_2^2}{\lambda\sigma_1^2 + \sigma_2^2}\right) \qquad (2.3)
\end{aligned}$$

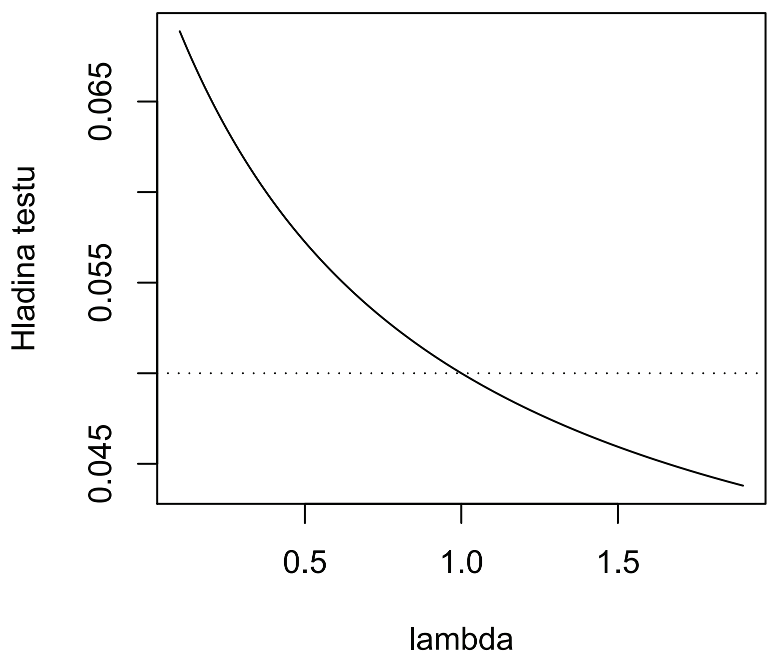
Z (2.3) je patrno, že abychom mohli sestavit test, který alespoň asymptoticky dodržuje předepsanou hladinu, stačilo by nám znát poměr rozptylů $\theta := \sigma_1^2/\sigma_2^2$. Tato situace je však v praxi velmi řídká.

Nyní se zabývejme vlastnostmi statistiky T v závislosti na hodnotě θ a poměru velikostí výběrů.

Asymptotický rozptyl testové statistiky T lze dle (2.3) přepsat jako

$$\mathrm{AVar}T = \frac{\theta + \lambda}{\lambda\theta + 1}$$

Obr.1: Asymptotická hladina testu při různých hodnotách λ ($\lambda = m/n$, $\alpha = 0,05$ je předepsaná hladina testu a $\theta = 1,2$ ($\theta = 1$ tečkovaně))



Z obrázku vidíme, že pokud je poměr θ blízký 1, tak se hladina testu i při různých velikostech výběrů příliš neliší od předepsané hladiny testu.

Pokud je $\lambda = 1$ (t.j. $m = n$), potom $\mathrm{AVar}T = 1$. Tedy pokud výběry mají stejné velikosti, nerovnost rozptylů asymptoticky neovlivňuje hladinu testu. Jestliže jsou velikosti výběrů téměř rovny, t-test zachovává asymptoticky předepsanou hladinu testu i pro velké rozdíly rozptylů. Simulace potvrzují, že to platí i pro malé rozsahy výběrů, jak je vidět např. v tabulce 1.

Tabulka 1 ukazuje maximální chybu v předpokládáné hladině testu. Tedy např. pro $m = n = 15$ a $\alpha = 0,05$ maximální chyba v předpokládané dosažené hladině (0,05) je $0,0098$. Proto skutečná dosažená hladina nebude vyšší o více než 0,0098 od předpokládáné hladiny, nezávisle na poměru rozptylů θ.

Tabulka 1 (viz Posten, 1984)

	$\alpha = 0,05$	$\alpha = 0,01$		$\alpha = 0,05$	$\alpha = 0,01$
n	ε	ε	n	ε	ε
2	0,0954	0,0539	15	0,0098	0,0052
3	0,0589	0,0341	20	0,0072	0,0038
4	0,0419	0,0241	25	0,0057	0,0030
5	0,0324	0,0184	30	0,0048	0,0025
6	0,0263	0,0148	50	0,0028	0,0015
7	0,0222	0,0124	100	0,0014	0,0007
8	0,0191	0,0106	500	0,0003	0,0001
9	0,0168	0,0093	1000	0,0001	0,0001
10	0,0150	0,0082	∞	0,0000	0,000

Následující simulace ukazuje, jak se chová t-test, pokud nejsou velikosti náhodných výběrů stejné. Udává maximální oblasti hodnot θ, přes které se skutečná dosažená hladina neodchyluje od předpokládané hladiny α o více než ε. Jestliže je tento rozsah θ velký, potom je v praxi t-test robustní na hladině ε. Vidíme, že čím více se liší velikosti náhodných výběrů, tím více t-test ztrácí na robustnosti. Pokud se liší o více než 20 procent, musíme být opatrnější s použitím t-testu. Problémem také je, jestliže je větší rozptyl přidružený k menší velikosti výběru. Někdy však již při plánování experimentu víme, ve kterém souboru je větší rozptyl. Pak můžeme hladinu robustnosti významně zlepšit tím, že ze souboru s větším rozptylem vybereme více jednotek.

Z tabulky 2 můžeme vidět, že např. pro $m = 27$, $n = 33$ (tedy pokud se velikosti výběrů liší o 10 procent) je maximální oblast poměru θ 0,00–36,95. V této oblasti se tedy skutečná dosažená hladina testu neodchýlí o více než $\varepsilon = 0,03$ od předpokládané hladiny $\alpha = 0,05$.

Tabulka 2 (viz Posten, 1984)

$\alpha = 0,05$								
$\varepsilon = 0,03$								
m	n	θ	m	n	θ	m	n	θ
5	5	0,02 - 85,63				4	6	0,00 - 2,89
10	10	0,00 - ∞	9	11	0,00 - 8,33	8	12	0,00 - 3,06
15	15	0,00 - ∞	14	16	0,00 - ∞	12	18	0,00 - 3,17
20	20	0,00 - ∞	18	22	0,00 - 17,58	16	24	0,00 - 3,25
25	25	0,00 - ∞	23	27	0,00 - ∞	20	30	0,02 - 3,30
30	30	0,00 - ∞	27	33	0,00 - 36,95	24	36	0,03 - 3,33
40	40	0,00 - ∞	36	44	0,00 - 104,22	32	48	0,06 - 3,38
50	50	0,00 - ∞	45	55	0,00 - ∞	40	60	0,05 - 3,42

$\varepsilon = 0,02$

m	n	θ	m	n	θ	m	n	θ
5	5	0,09 - 12,33				4	6	0,00 - 2,14
10	10	0,00 - ∞	9	11	0,00 - 4,02	8	12	0,21 - 2,17
15	15	0,00 - ∞	14	16	0,00 - 9,31	12	18	0,26 - 2,20
20	20	0,00 - ∞	18	22	0,00 - 4,95	16	24	0,28 - 2,21
25	25	0,00 - ∞	23	27	0,00 - 3,90	20	30	0,29 - 2,22
30	30	0,00 - ∞	27	33	0,00 - 5,55	24	36	0,30 - 2,23
40	40	0,00 - ∞	36	44	0,00 - 5,96	32	48	0,31 - 2,24
50	50	0,00 - ∞	45	55	0,00 - 6,26	40	60	0,31 - 2,25

2.2 Testy při nerovnosti rozptylů

V případě nerovnosti rozptylů se v základních učebnicích statistiky doporučuje použít Satterthwaiteův nebo Welchův test, což jsou modifikace dvouvýběrového t-testu.

Skutečným problémem ale není, jestli jsou rozptyly různé, ale jestli je známý jejich poměr. Tento poměr je však v praxi jen zřídka znám.

Při výběru vhodného testu pro testování shody středních hodnot μ_1 a μ_2 se rozhodujeme podle informací, které máme o rozptylech, a podle velikostí obou náhodných výběrů.

Pokud jsou velikosti výběrů stejné, Satterhwaiteův/Welchův test a t-test mají stejné velikosti a téměř stejné síly testu.

Kdykoli jsou velikosti výběrů nestejné a poměr θ je známý a blízký 1, t-test poskytuje velkou sílu testu. Proto je v takovémto případě t-test vhodný.

Dále jestliže nejsou velikosti výběrů stejné a poměr θ není známý nebo je různý od 1 je vhodné použít Satterhwaiteův/Welchův test, protože t-test může mít velkou velikost v případě, že θ je různé od 1.

Z toho také vyplývá, že v případě, že neznáme poměr rozptylů, není nutné provádět předběžný test jejich rovnosti, viz Moser a Stevens (1992). Nejenže je takovýto test nadbytečný, ale další nevýhodou je, že standardní testy na rovnost rozptylů založené na poměru S_X^2/S_Y^2 jsou velice citlivé na odchylky od normality a navíc nezamítnutí hypotézy H_0 u tohoto testu neznamená, že rozptyly jsou opravdu shodné (hlavně pro nevelké rozsahy výběru).

Kapitola 3

Princip modifikace t-testu pro nestejné rozptyly

Připomeňme, že předpokládáme, že $X_1, \dots, X_m$ mají normální rozdělení $\mathrm{N}(\mu_1, \sigma_1^2)$, $Y_1, \dots, Y_n$ mají normální rozdělení $\mathrm{N}(\mu_2, \sigma_2^2)$ a oba náhodné výběry jsou na sobě nezávislé.

Testujeme hypotézu $H_0 : \mu_1 = \mu_2$ proti alternativě $H_1 : \mu_1 \neq \mu_2$. Je přirozené založit test na $\overline{X} - \overline{Y}$. Tento rozdíl má za hypotézy H_0 rozdělení $N(0, \frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n})$. Nestranným a konzistentním odhadem $\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}$ je

$$S^2 = \frac{1}{m} S_X^2 + \frac{1}{n} S_Y^2, \tag{3.1}$$

kde S_X^2 a S_Y^2 je definováno v (2.1) a (2.2).

Zdá se tedy přirozené použít testovou statistiku $T^* = \frac{\overline{X} - \overline{Y}}{S}$. Neznáme však přesné rozdělení statistiky T za H_0 a proto nejsme schopni určit kritické hodnoty pro tuto statistiku. Dále nastíníme myšlenku, jak toto rozdělení aproximovat. Přepíšeme-li T^* do tvaru

$$T^* = \frac{\overline{X} - \overline{Y}}{S} = \frac{\dfrac{\overline{X} - \overline{Y}}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}}}{\sqrt{K}}, \text{ kde } K = \frac{S^2}{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}, \tag{3.2}$$

pak má čitatel $N(0, 1)$ rozdělení.

Satterthwaite (1946) navrhl aproximovat rozdělení statistiky K pomocí rozdě-lení náhodné veličiny $\frac{1}{r_s} Z_{r_s}$, kde Z_{r_s} má χ^2-rozdělení o r_s stupních volnosti. Snadno nahlédneme, že $E(K) = 1$ a $E(\frac{1}{r_s} Z_{r_s}) = 1$. Počet stupňů volnosti náhodné veličiny Z_{r_s} volíme tak, aby

$$\mathsf{var} K = \mathsf{var} \left(\frac{1}{r_s} Z_{r_s} \right) \tag{3.3}$$

Pro výpočet pravé a levé strany v (3.3) připomeňme, že jestliže náhodná veličina U má rozdělení χ^2 o r stupních volnosti, pak $\mathsf{var}\, U = 2\, r$.
Levá strana (3.3) se tedy rovná

$$\mathsf{var}\, Z_{r_s} = \frac{1}{r_s^2} \cdot 2 r_s = \frac{2}{r_s}$$

Pro výpočet pravé strany (3.3) využijeme toho, že

$$\begin{aligned} \frac{(m-1)S_X^2}{\sigma_1^2} & \quad \text{má rozdělení} \quad \chi^2_{m-1} \\ \frac{(n-1)S_Y^2}{\sigma_2^2} & \quad \text{má rozdělení} \quad \chi^2_{n-1} \end{aligned} \tag{3.4}$$

Pro jednoduchost zápisu si označme

$$\xi := \left(\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n} \right) \tag{3.5}$$

Počítejme

$$\begin{aligned} \mathsf{var}\, K &= \mathsf{var} \left(\frac{S^2}{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}} \right) \\ &= \frac{1}{\xi^2} \cdot \mathsf{var}\, S^2 \\ &= \frac{1}{\xi^2} \cdot \left(\frac{1}{m^2} \mathsf{var}\, S_X^2 + \frac{1}{n^2} \mathsf{var}\, S_Y^2 \right) \\ &= \frac{1}{\xi^2} \cdot \left(\frac{1}{m^2} \mathsf{var} \left(\frac{S_X^2 (m-1)}{\sigma_1^2} \cdot \frac{\sigma_1^2}{m-1} \right) + \frac{1}{n^2} \mathsf{var} \left(\frac{S_Y^2 (n-1)}{\sigma_2^2} \cdot \frac{\sigma_2^2}{n-1} \right) \right) \\ &= \frac{1}{\xi^2} \cdot \left(\frac{1}{m^2} \cdot 2 \frac{\sigma_1^4}{m-1} + \frac{1}{n^2} \cdot 2 \frac{\sigma_2^4}{n-1} \right) \end{aligned}$$

Porovnáním $varK$ a $\mathsf{var} \left(\frac{1}{r_s} Z_{r_s} \right)$ dostaneme, že počet stupňů volnosti náhodné veličiny Z_{r_s} by měl splňovat

$$r_s = \frac{\left(\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n} \right)^2}{\frac{\sigma_1^4}{m^2(m-1)} + \frac{\sigma_2^4}{n^2(n-1)}} \tag{3.6}$$

Tedy statistika K má podle naší aproximace přibližně $(1/r_s)\chi^2_{r_s}$ rozdělení a tudíž T^* má přibližně t rozdělení o r_s stupních volnosti. Ve vzorci (3.6) se ale vyskytují neznámé rozptyly σ_1^2 a σ_2^2. Následující testy jsou vpodstatě jen různými návrhy, jak tento teoreticky optimální počet stupňů volnosti r_s odhadnout.

Kapitola 4

Konkrétní modifikace t-testu

4.1 Satterthwaiteův test

Jelikož víme, že $E\,S_X^2 = \sigma_1^2$ a $E\,S_Y^2 = \sigma_2^2$, $S_X^2 \xrightarrow{s.j.} \sigma_1^2$ a $S_Y^2 \xrightarrow{s.j.} \sigma_2^2$, navrhl Satterthwaite nahradit neznámé σ_1^2, σ_2^2 ve vzorci (3.6) jejich nestrannými a konzistentními odhady S_X^2, S_Y^2

$$\widehat{r}_s = \frac{S^4}{\dfrac{S_X^4}{m^2(m-1)} + \dfrac{S_Y^4}{n^2(n-1)}}$$

Satterhwaiteův test zamítá H_0 v případě, že $|T^*| \geq t_{\widehat{r}_s}(\alpha)$

4.2 Welchův test

Welchův test využívá následujícího odhadu r_s:

$$\widehat{r}_s = \frac{S^4}{\dfrac{S_X^4}{m^2(m+1)} + \dfrac{S_Y^4}{n^2(n+1)}} - 2 \tag{4.1}$$

Tento odhad sice není tolik intuitivní jako Satterthwaiteův, ale jeho odvození je následující. Přepišme

$$\widehat{r}_s = \frac{S^4 - 2\left(\dfrac{S_X^4}{m^2(m+1)} + \dfrac{S_Y^4}{n^2(n+1)}\right)}{\dfrac{S_X^4}{m^2(m+1)} + \dfrac{S_Y^4}{n^2(n+1)}} \tag{4.2}$$

Ukážeme, že čitatel (4.2) je nestranný odhad čitatele z (3.6) a jmenovatel (4.2) je nestranný odhad jmenovatele z (3.6).

Ze znalosti střední hodnoty a rozptylu χ^2-rozdělení, z (3.4) a ze známé rovnosti $EX^2 = \mathsf{var}\, X + (EX)^2$ vyplývá, že

$$\begin{aligned} ES_X^4 &= E\left(\frac{S_X^2(m-1)}{\sigma_1^2}\frac{\sigma_1^2}{m-1}\right)^2 = \frac{\sigma_1^4}{(m-1)^2}\cdot\left(2\,(m-1)+(m-1)^2\right) \\ &= \frac{\sigma_1^4(m+1)}{(m-1)} \end{aligned} \tag{4.3}$$

Obdobně získáme

$$ES_Y^4 = \frac{\sigma_2^4(n+1)}{(n-1)}$$

Označme střední hodnotu čitatele (4.2) jako A a střední hodnotu jmenovatele (4.2) jako B. Pak

$$\begin{aligned} A &= E\left(\frac{1}{m}S_X^2 + \frac{1}{n}S_Y^2\right)^2 - 2E\left(\frac{S_X^4}{m^2(m+1)} + \frac{S_Y^4}{n^2(n+1)}\right) \\ &= E\left(\frac{S_X^4}{m^2} + 2\cdot\frac{S_X^2S_Y^2}{m\,n} + \frac{S_Y^4}{n^2}\right) - 2\,E\left(\frac{S_X^4}{m^2(m+1)} + \frac{S_Y^4}{n^2(n+1)}\right) \\ &= \left(\frac{\sigma_1^4(m+1)}{m^2(m-1)} + 2\frac{\sigma_1^2\sigma_2^2}{m\,n} + \frac{\sigma_2^4(n+1)}{n^2(n-1)}\right) - 2\left(\frac{\sigma_1^4}{m^2(m-1)} + \frac{\sigma_2^4}{n^2(n-1)}\right) \\ &= \frac{\sigma_1^4}{m^2} + 2\frac{\sigma_1^2\sigma_2^2}{m\,n} + \frac{\sigma_2^4}{n^2} \\ &= \left(\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}\right)^2 \\ B &= \frac{\sigma_1^4}{m^2(m-1)} + \frac{\sigma_2^4}{n^2(n-1)} \end{aligned}$$

Počet stupňů volnosti r_s z (3.6) se tedy vskutku rovná $\frac{A}{B}$.

4.3 Ověření

Nyní se pokusíme ukázat, že aproximace rozdělení statistiky T pomocí t-rozdělní s r_s stupni volnosti, které jsou dány pomocí (3.6), má opodstatnění. Toto ověření je ale pouze heuristikou (bylo by třeba ještě ukázat, že $E|S_X^2 - \sigma_1^2|^3 = O\left(1/m^{3/2}\right)$ a $E|S_Y^2 - \sigma_2^2|^3 = O\left(1/n^{3/2}\right)$).
Máme

$$T = \frac{\overline{X} - \overline{Y}}{S}, \qquad \text{kde } S = \sqrt{\frac{S_X^2}{m} + \frac{S_Y^2}{n}}$$

Pomocí (4.3) vypočteme

$$E(S_X^2 - \sigma_1^2)^2 = ES_X^4 - 2\cdot ES_X^2\sigma_1^2 + \sigma_1^4 = E(S_X^2)^2 - \sigma_1^4 = 2\cdot\frac{\sigma_1^4}{(m-1)}$$

Obdobně dostaneme $E(S_Y^2-\sigma_2^2)^2 = 2\cdot\frac{\sigma_2^4}{(n-1)}$.
Označme $H(t)$ distribuční funkci náhodné veličiny T^* (viz (3.2)).
Potom:

$$\begin{aligned}
H(t) &= P(T^*\le t) = E_S P(T^*\le t\mid S) = E_S P(\overline{X}-\overline{Y}\le tS\mid S)\\
&= E_S P\left(\frac{\overline{X}-\overline{Y}}{\sqrt{\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}}}\le \frac{t\,S}{\sqrt{\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}}}\mid S\right) = E_S\Phi\left(\frac{tS}{\sqrt{\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}}}\right)\\
&= E_S\Phi\left(\frac{t\sqrt{\frac{S_X^2}{m}+\frac{S_Y^2}{n}}}{\sqrt{\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}}}\right) \overset{\text{ozn.}}{=} g(S_X^2,S_Y^2),
\end{aligned}$$

kde $\Phi(t)$ je distribuční funkce normovaného normálního rozdělení.
Označme ξ stejně jako v (3.5).
Funkci $g(S_X^2,S_Y^2)$ rozvineme v Taylorovu řadu kolem bodu (σ_1^2,σ_2^2):

$$\begin{aligned}
H(t) = E_S\Big\{&\Phi(t)+\varphi(t)\frac{t}{2\,\xi}\frac{1}{m}(S_X^2-\sigma_1^2)+\varphi(t)\frac{t}{2\,\xi}\frac{1}{n}(S_Y^2-\sigma_2^2)\\
&+\frac{\varphi'(t)}{2}\frac{t^2}{(2\,\xi)^2}\frac{1}{m^2}(S_X^2-\sigma_1^2)^2-\frac{\varphi(t)}{2}\frac{t}{(2\,\xi)^2}\frac{1}{m^2}(S_X^2-\sigma_1^2)^2\\
&+\frac{\varphi'(t)t^2-\varphi(t)t}{(2\,\xi)^2}\frac{1}{m\,n}(S_X^2-\sigma_1^2)(S_Y^2-\sigma_2^2)\\
&+\frac{\varphi'(t)t^2-\varphi(t)t}{2(2\,\xi)^2}\frac{1}{n^2}(S_Y^2-\sigma_2^2)^2+O(|S_X^2-\sigma_1^2|^3+|S_Y^2-\sigma_2^2|^3)\Big\}\\
&=\Phi(t)+0+0+\frac{\varphi'(t)t^2-\varphi(t)t}{8\,\xi^2}\frac{1}{m^2}\frac{2\,\sigma_1^4}{m-1}+0\\
&+\frac{\varphi'(t)t^2-\varphi(t)t}{8\,\xi^2}\frac{1}{n^2}\frac{2\,\sigma_2^4}{n-1}+O\left(\frac{1}{m^{\frac{3}{2}}}\right)+O\left(\frac{1}{n^{\frac{3}{2}}}\right)\\
&=\Phi(t)-\frac{(t^3+t)\varphi(t)}{4}\frac{\left(\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}\right)^2}{\frac{\sigma_1^4}{m^2(m-1)}+\frac{\sigma_2^4}{n^2(n-1)}}+O\left(\frac{1}{m^{3/2}}\right)+O\left(\frac{1}{n^{3/2}}\right)
\end{aligned}$$

Tedy

$$H(t)=\Phi(t)-\frac{(t^3+t)\varphi(t)}{4}\cdot\frac{1}{r_s}+O\left(\frac{1}{m^{3/2}}\right)+O\left(\frac{1}{n^{3/2}}\right), \tag{4.4}$$

kde $\varphi(t)$ je hustota normovaného normálního rozdělení. Ve výpočtu jsme využili toho, že $\varphi'(t)=-t\cdot\varphi(t)$.

Nechť náhodná veličina Z_v má t rozdělení o v stupních volnosti. Pak obdobným postupem dostaneme

$$P(Z_v \leq t) = \Phi(t) - \frac{(t^3 + t)\varphi(t)}{4} \cdot \frac{1}{v} + O\left(\frac{1}{v^{3/2}}\right). \tag{4.5}$$

Porovnáme-li (4.4) a (4.5) v případě, že $v = r_s$, pak

$$P(T^* \leq t) - P(Z_{r_s} \leq t) = O\left(\frac{1}{m^{3/2}}\right) + O\left(\frac{1}{n^{3/2}}\right).$$

Kapitola 5

Bootstrap

Jako poslední uvedeme jednu z moderních, ale výpočetně náročných, metod.

Mějme stejnou situaci jako v úvodu kapitoly 3 s testovou statistikou T^* a výběrovými rozptyly S_X^2, S_Y^2 a testujme nulovou hypotézu $H_0 : \mu_1 = \mu_2$ proti její alternativě $H_1 : \mu_1 \neq \mu_2$ na hladině α. Jelikož za nulové hypotézy známe rozdělení $\overline{X} - \overline{Y}$ $(\sim N(0, \frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}))$, můžeme bez újmy na obecnosti předpokládat, že $\mu_1 = \mu_2 = 0$. Pokud bychom znali rozptyly σ_1^2 a σ_2^2, postupovali bychom podle následujícího bootstrapového algoritmu:

1. pro $i = 1, \ldots, B$

a. nageneruj výběr $X_1^*, \ldots, X_m^*$ z $N(0, \sigma_1^2)$ a $Y_1^*, \ldots, Y_n^*$ z $N(0, \sigma_2^2)$

b. spočti statistiku T_i^* podle

$$T_i^* = \frac{\overline{X}^* - \overline{Y}^*}{\sqrt{\frac{S_{X^*}^2}{m} + \frac{S_{Y^*}^2}{n}}},$$

$$\text{kde } S_{X^*}^2 = \frac{1}{m-1}\sum_{i=1}^{m}(X_i^* - \overline{X}^*)^2, \quad S_{Y^*}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i^* - \overline{Y}^*)^2$$

2. odhadni P-hodnotu pomocí

$$p = \frac{1 + \sum_{i=1}^{B} I\{|T_i^*| > |T|\}}{B+1}$$

3. zamítni nulovou hypotézu, pokud $p \leq \alpha$.

Pro B dostatečně velké získáme velmi přesné řešení.

Protože ale většinou rozptyly neznáme, nahradíme je nám známými odhady S_X^2, S_Y^2. A tudíž v bodě **1.**a. generujeme výběr $X_1^*, \ldots, X_m^*$ z $N(0, S_X^2)$ a obdobně výběr $Y_1^*, \ldots, Y_n^*$ z $N(0, S_Y^2)$. Dá se dokázat (viz Beran, 1988), že tento test je ve smyslu dodržení předepsané hladiny stejně přesný jako test Welchův nebo Satterthwaitův.

Podle Beran (1988) zlepšíme předchozí metodu kombinací s Welchovým testem. Algoritmus takového testu by pak byl:

1. pro $i = 1, \ldots, B$

a. nageneruj výběr $X_1^*, \ldots, X_m^*$ z $N(0, \sigma_1^2)$ a $Y_1^*, \ldots, Y_m^*$ z $N(0, \sigma_2^2)$

b. spočti statistiku T_i^* a odhad stupňů volnosti h_i^*

c. spočti P-hodnotu

$$p_i^* = 2[1 - G_{h_i^*}(|T_i^*|)],$$

kde G_h je distribuční funkce t-rozdělení o h stupních volnosti (h je z (4.1))
2. odhadni skutečnou P-hodnotu pomocí

$$p = \frac{1 + \sum_{i=1}^{B} I\{p_i^* < p_T\}}{B + 1},$$

kde $p_T = 2[1 - G_{\widehat{r_s}}(|T_i^*|)]$ **3.** zamítni nulovou hypotézu, pokud $p \leq \alpha$.

Výsledky malé numerické studie provedené v Beran (1988) však nenaznačují, že by tento druhý postup, i když je teoreticky asymptoticky přesnější, přinesl nějaké podstatné zlepšení. Důvodem však je, jak naznačují četné numerické studie, že Welchův, Satterthwaiteův či bootstrapový test již fungují velmi dobře.

Kapitola 6

Závěr

Rozebrali jsme některé z možností, jak řešit Behrens-Fisherův problém. Zjistili jsme, jak nerovnost rozptylů náhodných výběrů z normálního rozdělení ovlivňuje spolehlivost t-testu a jak se na spolehlivosti odráží různé velikosti výběrů. Citované simulace potvrzují, že pro stejné velikosti výběrů je t-test dostatečně spolehlivý a zachovává asymptoticky předepsanou hladinu testu i při velkých rozdílech rozptylů a oproti svým modifikacím je v tomto případě t-test silnější. Jestliže se ale velikosti výběrů významněji liší, je již vhodné použít některou z modifikací t-testu, popř. v poslední kapitole uvedený bootstrapový test nebo kombinaci bootstrapového testu s testem Welchovým. Tyto testy sice nemají takovou sílu testu jako t-test, ale velmi dobře dodržují předepsanou hladinu i při různých rozptylech.

Dobrý test by však kromě dodržování předepsané hladiny měl mít velkou sílu. Dle našich současných znalostí však výsledky tohoto typu v literatuře buď chybějí nebo alespoň nejsou snadno objevitelné. Bylo by jistě zajímavé srovnat navržené testy dle jejich síly. To je však již nad rámec této práce.

Literatura

[1] Anděl J.: *Statistické metody*, Matfyzpress, Praha, 1998.

[2] Beran R..: *Prepivoting test statistics: A bootstrap view of asymptotic refinements*, J. Amer. Statist. Assoc. **83** (1988), 687–697.

[3] Bradley J. V.: *Nonrobustness in classical tests on means and variances: a large scale study*, Bulletin of the Psychonomic Society **15** (1980), 275–278.

[4] Dupač V., Hušková M.: *Pravděpodobnost a matematická statistika*, Nakladatelství Karolinum, Praha, 2003.

[5] Jurečková J.: *Testy parametrických hypotéz*, SPN, Praha, 1982.

[6] Moser B. K., Stevens G. R.: *Homogeneity of Variance in the Two-Sample Means Test*, The American Statistician **46** (1992), 19–21.

[7] Posten H. O.: *Robustness of the Two-Sample T-Test*, Robustness of statistical methods and nonparametric statistics, Reidel VEB Deutscher Verlag der Wissenschaften, Dordrecht, 1984, 92–99.

[8] Satterthwaite F. E.: *An Approximate Distribution of Estimates of Variance Components*, Biometrics Bulletin **2** (1946), 110–114.

[9] Welch B. L.: *The Significance of the Difference Between Two Means when the Population Variances are Unequal*, Biometrika **29** (1938), 350–362.