Univerzita Karlova v Praze
Matematicko-fyzikální fakulta

Diplomová práce

Peter Bubelíny

Analýza rozptylu pro heteroskedastická data

Katedra pravděpodobnosti a matematické statistiky

Vedoucí: **Mgr. Michal Kulich, Ph.D.**

Studijní obor: **Matematická statistika**

Poděkování

Děkuji vedoucímu diplomové práce Mgr. Michalovi Kulichovi, Ph.D. za hodnotné rady a připomínky, které výrazně přispěly ke zlepšení této práce.

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

V Praze, dne 20. dubna 2007 Peter Bubelíny

Obsah

Název práce: Analýza rozptylu pro heteroskedastická data

Autor: Peter Bubelíny

Katedra (Ústav): Katedra pravděpodobnosti a matematické statistiky

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

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

Abstrakt: Hlavní zaměření této diplomové práce je porovnat několik testů na testování hypotézy o shodě středních hodnot několika nezávislých populací. Pro test této hypotézy uvažujeme 10 testů publikovaných v literatuře. Jako možné rozdělení těchto populací jsou uvažována čtyři rozdělení: normální rozdělení, rovnoměrné rozdělení, dvojitě exponenciální rozdělení a logaritmicko-normální rozdělení. Porovnání testů je provedeno jak pro homoskedastické případy, tak také pro heteroskedastické případy. Také jsou tyto testy zkoumané pro vyvážené případy, ale také pro nevyvážené případy. Pro každý uvažovaný test je zkoumáno, jak test v různých nastaveních dodržuje hladinu významnosti a jakou má sílu za platnosti tří různých alternativ. Hladiny významnosti a síly testů jsou odhadnuty jako relativní četnosti zamítnutí hypotézy pomocí simulování těchto populací za předpokladu, že známe druh rozdělení populací a víme, že buď platí hypotéza nebo určitá alternativa. Na základě porovnávání odhadovaných hladin významnosti a sil testů je určeno, který z uvažovaných testů má pro danou alternativu největší sílu mezi testama, které dodrží hladinu významnosti.

Klíčová slova: heteroskedasticita, analýza rozptylu

Title: Analysis of variance for hereroskedastic data

Author: Peter Bubelíny

Department: Department of Probability and Mathematical Statistics

Supervisor: Mgr. Michal Kulich, Ph.D.

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

Abstract: The main goal of this thesis is to compare tests for testing the hypothesis of equality of means in several independent populations. For testing this hypothesis, 10 published tests are considered. As possible distributions of these populations, the normal distribution, the uniform distribution, the double exponential distribution and the logarithmic-normal distribution are selected. These tests are compared in both homoskedastic and heteroskedastic cases, for both balanced and unbalanced data. For each test it is tested if the test holds the nominal level α for different situations and how powered this test is if one of three different alternatives occurs. Nominal levels and powers of tests are estimated as relative frequencies of rejection of hypothesis by simulating these populations provided known distribution of these populations and knowing that either the hypothesis holds or the hypothesis is rejected by a known alternative. On the base of holding nominal levels and powers of tests it is specified which of these tests has the best power for specific alternative among the tests which hold the nominal level.

Keywords: heteroskedasticity, analysis of variance

1 Úvod

Predpokladajme, že máme I nezávislých výberov (populácií) z nejakých rozdelení s neznámou strednou hodnotou μ_i a s neznámym, kladným a konečným rozptylom σ_i^2, $i = 1, ..., I$. Našou úlohou je testovať hypotézu zhody stredných hodnôt týchto populácií proti alternatíve, že hypotéza neplatí, tj. testovať hypotézu $H_0 : \mu_1 = ... = \mu_I$ proti alternatíve $A_0 : H_0$ *neplatí*. Ak označíme Y_{ij} ako i-te pozorovanie v j-tej populácii, potom toto pozorovanie môžeme rozpísať ako: $Y_{ij} = \mu_i + \epsilon_{ij} = \mu + \alpha_i + \epsilon_{ij}$, kde μ je celková stredná hodnota všetkých populácií, α_i je odchýlka i-tej populácie od celkovej strednej hodnoty všetkých populácií a ϵ_{ij} je náhodná veličina z rovnakého druhu rozdelenia ako i-ta populácia a má nulovú strednú hodnotu a rozptyl σ_i^2. Namiesto μ_i teraz máme $\mu + \alpha_i$. Preto môžeme hypotézu H_0 prepísať na tvar $H_0 : \mu + \alpha_1 = ... = \mu + \alpha_I$, čo môžeme tiež zapísať ako $H_0 : \alpha_1 = ... = \alpha_I$. Pri rozpise $\mu_i = \mu + \alpha_i$, $i = 1, ..., I$ sme ale zaviedli o jeden parameter viac ako je potreba. To má za následok, že parametre μ, α_1,..., α_I nie sú odhadnuteľné. Preto musíme zaviesť ešte reparametrizačnú rovnicu, aby sme mohli tieto parametre jednoznačne odhadnúť. Voľbou reparametrizačnej rovnice $\sum_{i=1}^{I} \alpha_i = 0$ dostávame, že testovať hypotézu H_0 proti alternatíve A_0 je ekvivalentné s testovaním hypotézy $H_0' : \alpha_1 = ... = \alpha_I = 0$ proti alternatíve $A_0' : H_0'$ *neplatí*.

Za predpokladu normality a homoskedasticity (zhody rozptylov) populácií vieme, že F-test je rovnomerne najsilnejší test hypotézy H_0 proti alternatíve A_0. Otázkou pre nás je: čo sa deje, keď sa poruší aspoň jeden z týchto dvoch predpokladov? V ktorých prípadoch ešte F-test dodržuje hladinu významnosti? Poznáme ešte nejaké iné testy hypotézy H_0 proti alternatíve A_0, ktoré dodržujú hladinu významnosti? Ktorý z týchto testov má najväčšiu silu?

2 Označenie

Budeme uvažovať problém porovnania stredných hodnôt I populácií s rôznymi rozptylmi, tj. testovať hypotézu $H_0 : \mu_1 = ... = \mu_I$ proti alternatíve A_0 : H_0 *neplatí*. Predpokladajme, že pre každú populáciu máme k dispozícii náhodný výber velikosti n_i. V prípade, že budeme mať pre všetky populácie rovnaký počet pozorovaní, budeme takéto nastavenie nazývať vyvážený prípad a počet pozorovaní v každej populácii budeme pre jednoduchosť značiť písmenom n. Keď je počet pozorovaní v populáciách rôzny, budeme toto nastavenie nazývať nevyvážený prípad. Zaveďme označenie Y_{ij}, $i = 1, ..., I$, $j = 1, ..., n_i$ pre j-te pozorovanie z i-tej populácie. Ďalej predpokladajme, že Y_{ij} sú medzi sebou nezávislé náhodné veličiny z rozdelenia so strednou hodnotou μ_i a rozptylom σ_i^2, $i = 1, ..., I$. Symbolmi $\overline{Y}_{i.}$, resp. S_i^2 budeme označovať výberový priemer, resp. výberový rozptyl pre i-tu populáciu, tj.

$$\begin{aligned}
\overline{Y}_{i.} &= \frac{1}{n_i}\sum_{j=1}^{n_i} Y_{ij} \\
S_i^2 &= \frac{1}{(n_i-1)}\sum_{j=1}^{n_i}(Y_{ij}-\overline{Y}_{i.})^2 \quad i = 1, ..., I.
\end{aligned}$$

Počet všetkých pozorovaní budeme označovať N, tj. $N = \sum_{i=1}^{I} n_i$, celkový priemer všetkých náhodných veličín Y_{ij} budeme označovať ako

$$\overline{Y}_{..} = \frac{1}{N}\sum_{i=1}^{I}\sum_{j=1}^{n_i} Y_{ij} = \frac{1}{N}\sum_{i=1}^{I} n_i\overline{Y_{i.}}.$$

Ďalej ešte zavedieme označenie MST a MSE pre priemerný súčet štvorcov medzi populáciami a priemerný súčet štvorcov v populáciách, tj.

$$\begin{aligned}
MST &= \frac{1}{I-1}\sum_{i=1}^{I} n_i(\overline{Y}_{i.}-\overline{Y}_{..})^2 \\
MSE &= \frac{1}{N-I}\sum_{i=1}^{I}\sum_{j=1}^{n_i}(Y_{ij}-\overline{Y}_{i.})^2.
\end{aligned}$$

Užitočné značenie, ktoré nám uľahčí a zjednoduší zápis hlavne v tabuľkách je značenie $x : y$, resp. $2{*}(x : y)$. Pod týmto značením budeme rozumieť aritmetickú postupnosť od x do y s krokom jedna (v prípade $x > y$ to bude krok mínus jedna), resp. aritmetickú postupnosť od $2x$ do $2y$ s krokom dva (v prípade $x > y$ to bude krok mínus dva). Ďalšie užitočné značenie pre postupnosť čísel, ktoré sa uplatní hlavne v tabuľkách je napr. 9x30,20, resp. 20,9x30. Pod týmto značením budeme rozumieť postupnosť dĺžky desať, ktorá má na prvých deviatich miestach tridsiatku a na poslednom desiatom mieste dvadsiatku, resp. postupnosť dĺžky desať, ktorá má na prvom mieste dvadsiatku a na posledných deviatich miestach tridsiatku.

Ešte si zopakujme niektoré pojmy z testovania hypotéz. Predpokladajme, že $\boldsymbol{X} = (X_1, ..., X_n)^T$ je náhodný výber z rozdelenia, ktoré patrí do rodiny $\{F_\theta, \theta \in \Theta\}$. Našou úlohou je na základe náhodného výberu $\boldsymbol{X}$ rozhodnúť, či platí hypotéza $H : \theta \in \Theta_0 (\subset \Theta)$ alebo platí alternatíva $A : \theta \in \Theta \backslash \Theta_0 (= \Theta_1)$. Naše rozhodnutie nemusí byť vždy správne. Môžeme zamietnuť hypotézu H, aj keď táto hypotéza platí (chyba prvého druhu), alebo nezamietneme hypotézu H ,aj keď táto hypotéza neplatí (chyba druhého druhu). Testom hypotézy budeme rozumieť rozhodovacie pravidlo, na základe ktorého rozhodneme, či prijmeme hypotézu alebo túto hypotézu zamietneme. Kritickým oborom testu rozumieme takú množinu $K \subset R^n$, že keď $\boldsymbol{X} \in K$, potom zamietame hypotézu H, v opačnom prípade prijmeme hypotézu H. Pre test definujme hladinu významnosti α ako suprémum cez $\theta \in \Theta_0$ z podmienenej pravdepodobnosti zamietnutia hypotézy na základe náhodného výberu $\boldsymbol{X}$ za podmienky θ, tj. $\alpha = \sup_{\theta \in \Theta_0} \mathrm{P}(\boldsymbol{X} \in K | \theta)$. Silofunkciou testu budeme označovať takú funkciu s premennou θ, ktorá je vyjadrená ako podmienená pravdepodobnosť zamietnutia hypotézy H za podmienky, že hypotéza neplatí, tj. $\beta(\theta) = \mathrm{P}(\boldsymbol{X} \in K | \theta \in \Theta_1)$. Povieme, že test T_1 je silnejší (má väčšiu silu) ako test T_2 pre nejaké ľubovoľné pevné $\theta' \in \Theta_1$ na hladine významnosti α, ak oba testy majú hladinu významnosti α a silofunkcia testu T_1 v bode θ' je väčšia ako silofunkcia testu T_2 v bode θ', tj. P(H *zamietame* testom $T_1|\theta'$) >P(H *zamietame* testom $T_2|\theta'$). Ak bude zrejmé o akú hodnotu θ' sa jedná, tak namiesto silofunkcie testu T v bode θ' budeme pre jednoduchosť hovoriť o sile testu T.

Našou úlohou bude medzi nami uvažovanými testami hľadať čo najsilnejší test na hladine významnosti α pre nami zvolené alternatívy.

3 Zavedenie testových štatistík a testov

3.1 F-test

Tento test predpokladá normalitu populácií a zhodu rozptylov pre všetky populácie. Je založený na podiele priemerného súčtu štvorcov medzi populáciami a priemerného súčtu štvorcov v populáciách, tj. na štatistike:

$$F = \frac{MST}{MSE}.$$

Táto štatistika má za predpokladu hypotézy H_0 (viď Anděl (2001)) F-rozdelenie s $I-1$ a $N-I$ stupňami voľnosti. Ak hypotéza H_0 neplatí, priemerný súčet štvorcov medzi populáciami MST nadobúda väčších hodnôt ako priemerný súčet štvorcov v populáciách MSE. Za platnosti hypotézy H_0 sú veličiny MST a MSE takmer zhodné. Preto testovanie H_0 prebieha tak, že túto hypotézu zamietame, ak štatistika F prekročí $(1-\alpha)$-kvantil F-rozdelenia s $I-1$ a $N-I$ stupňami voľnosti, tj. keď $F > F_{I-1,N-I}(1-\alpha)$. Tento test budeme v ďalšom texte označovať ako **F-test**.

3.2 Akritasov nevážený test a Akritasov vážený test

M.G. Akritas a N. Papadatos vo svojom článku Akritas and Papadatos (2004) odvodili pre rôzne situácie niekoľko testov na testovanie hypotézy H_0. Tieto testy nepredpokladajú zhodu rozptylov ani normalitu populácií. Sú odvodené pre $I \to \infty$. Heteroskedasticita sa prejavuje ako aj v testových štatistikách, tak aj pri voľbe kritických hodnôt. Ja som sa rozhodol pre testovanie H_0 vybrať dva testy, ktoré sú z týchto testov najvšeobecnejšie. Prvý test je založený na štatistike T_I definovanej predpisom

$$T_I = I^{-1/2} \sum_{i=1}^{I} [n_i(\overline{Y}_{i.} - \overline{Y}_{..})^2 - (1 - \frac{n_i}{N})S_i^2],$$

ktorá je vo vyvážených prípadoch priamo úmerná rozdielu priemerného súčtu štvorcov medzi populáciami MST a priemerného súčtu štvorcov v populáciách MSE, tj.

$$T_I = (1 - \tfrac{1}{I})(I^{1/2}(MST - MSE)).$$

Druhý test je založený na štandardizovanom súčte štvorcov medzi populáciami, tj. na štatistike

$$T_W = \sum_{i=1}^{I} \frac{n_i}{\sigma_i^2}\overline{Y}_{i.}^2 - \frac{1}{\sum_{i=1}^{I} n_i/\sigma_i^2}(\sum_{i=1}^{I} \frac{n_i}{\sigma_i^2}\overline{Y}_{i.})^2 = \sum_{i=1}^{I} \frac{n_i}{\sigma_i^2}(\overline{Y_{i.}} - \overline{Y_s})^2, \tag{3.1}$$

kde $\overline{Y_s}$ je vážený priemer výberových priemerov v populáciách s váhami $\frac{n_i}{\sigma_i^2}$, tj.

$$\overline{Y_s} = \frac{\sum_{i=1}^{I} \frac{n_i}{\sigma_i^2} \overline{Y_{i.}}}{\sum_{i=1}^{I} \frac{n_i}{\sigma_i^2}}.$$

Pretože rozptyly populácií σ_i^2, $i = 1, ..., I$ nie sú známe, budeme H_0 testovať pomocou štatistiky $\hat{T}_W$, ktorá vznikne z T_W nahradením rozptylov σ_i^2, $i = 1, ..., I$ ich nestrannými odhadmi S_i^2, $i = 1, ..., I$. Preto môžeme $\hat{T}_W$ písať v tvare:

$$\hat{T}_W = \sum_{i=1}^{I} \frac{n_i}{S_i^2} \overline{Y}_{i.}^2 - \frac{1}{\sum_{i=1}^{I} n_i / S_i^2} (\sum_{i=1}^{I} \frac{n_i}{S_i^2} \overline{Y}_{i.})^2.$$

Odpoveď na otázku ako testovať hypotézu H_0 pomocou štatistík T_I, resp. $\hat{T}_W$ nám dávajú nasledujúce vety.

Veta 3.1. *Nech Y_{ij}, $i = 1, ..., I$, $j = 1, ..., n_i$ je dvojitá postupnosť nezávislých náhodných veličín s $EY_{ij} = \mu$ a $0 < Var\ Y_{ij} = \sigma_i^2 < \infty$, $n_i \geq 2$ pre všetky i. Predpokladajme, že v každej populácii i sú náhodné veličiny Y_{ij}, $j = 1, ..., n_i$ rovnako rozdelené. Naviac ešte predpokladajme, že*

$$\frac{1}{I} \sum_{i=1}^{I} \sigma_i^4 \xrightarrow{I \to \infty} s^4 \in (0, \infty)$$

$$\frac{1}{I} \sum_{i=1}^{I} \frac{\sigma_i^4}{n_i - 1} \xrightarrow{I \to \infty} \gamma^4 \in (0, \infty)$$

a pre nejaké $\delta > 0$ je $sup_{I \geq 1} \frac{1}{I} \sum_{i=1}^{I} n_i^{2+\delta} < \infty$ a $sup_{i \geq 1} E|\ Y_{i1}\ |^{2+\delta} < \infty$. Potom:

$$T_I \xrightarrow{d} N(0, 2(s^4 + \gamma^4)) \quad ako \quad I \to \infty.$$

V ďalšej vete ešte naviac použijeme značenie $n_i(I)$, ak $n_i \to \infty$ ako $I \to \infty$.

Veta 3.2. *Nech Y_{ij} sú nezávislé náhodné veličiny jednotne obmedzené nejakou konštantou M s $EY_{ij} = \mu$ a $0 < Var\ Y_{ij} = \sigma_i^2 < \infty$, $i = 1, ..., I$, $j = 1, ..., n_i(I)$. Predpokladajme, že pre každú populáciu i sú náhodné veličiny Y_{ij}, $j = 1, ..., n_i(I)$ rovnako rozdelené. Potom nech pre nejaké $\delta > 0$ je $sup_{i \geq 1} E|\ (Y_{i1} - \mu)/\sigma_i\ |^{4+\delta} < \infty$. Nechajme $n(I) = \min(n_i(I);\ i = 1, ..., I)$, $\overline{n} = \overline{n}(I) = I^{-1/2} \sum_{i=1}^{I} n_i$. Ak*

$\sum_{i=1}^{I} exp(-(\log n_i)^2) \to 0$ *a naviac*
(a): $I^{-1/2} \sum_{i=1}^{I} \frac{(\log n_i)^4}{n_i} = o(1)$,
(b): $\frac{I^{1/2}}{\overline{n}} = o(1)$,
(c): $n(I) \to \infty$,
potom:

$$I^{-1/2}(\hat{T}_W - (I-1)) \xrightarrow{d} N(0,2) \quad ako \quad I \to \infty.$$

Poznámka 3.3. Vo vete 3.1 sa používajú aj druhé mocniny rozptylov populácií $\sigma_i^4, i = 1, ..., I$. Pretože $\sigma_i^2, i = 1, ..., I$ sú neznáme, sú neznáme tiež $\sigma_i^4, i = 1, ..., I$. Autori tejto vety odporučujú pre odhad σ_i^4, $i = 1, ..., I$ použiť nestranný odhad založený na U-štatistike. Ja som ale pre σ_i^4, $i = 1, ..., I$ použil konzistentný odhad $(S_i^2)^2$, $i = 1, ..., I$. Dôvodom použitia tohto odhadu bolo, že test odvodený z tejto vety za predpokladu normality populácií lepšie dodržujú hladinu významnosti hypotézy H_0 s použitím môjho odhadu pre σ_i^4, $i = 1, ..., I$.

Z týchto viet dostávame 2 testy:
(1) **Akritasov nevážený test**
Za hypotézy H_0 nadobúda priemerný súčet štvorcov medzi populáciami MST približne takých hodnôt ako priemerný súčet štvorcov v populáciách MSE. Preto za platnosti hypotézy H_0 je rozdiel MST a MSE (a teda aj T_I) takmer rovný nule, ak hypotéza neplatí, potom je $MST > MSE$ a teda aj "$T_I \gg 0$". Preto zamietame hypotézu H_0, ak $\frac{T_I}{\sqrt{2(s^2+\gamma^2)}} > u_{(0,1)}(1-\alpha)$, kde $u_{(0,1)}(1-\alpha)$ je $(1-\alpha)$-kvantil normovaného normálneho rozdelenia, tj. kvantil $N(0,1)$.
(2) **Akritasov vážený test**
Za platnosti hypotézy H_0 je stredná hodnota štatistiky T_W rovná počtu populácií zmenšeného o jedna, tj. $\mathrm{E}_{H_0} T_W = I - 1$. Ak táto hypotéza neplatí, potom štatistika T_W nadobúda väčších hodnôt ako za platnosti hypotézy. Preto hypotézu H_0 zamietame, ak $\frac{\hat{T}_W-(I-1)}{\sqrt{2I}} > u_{(0,1)}(1-\alpha)$, kde $u_{0,1}(1-\alpha)$ je $(1-\alpha)$-kvantil normovaného normálneho rozdelenia, tj. kvantil $N(0,1)$.

3.3 Krutchkoffov test

V článku Krutchkoff (1988) je intuitívne odvodený test pre našu hypotézu H_0. Tento test predpokladá normalitu populácií, ale nepredpokladá zhodu rozptylov týchto populácií. Heteroskedasticita je zahrnutá pri výpočte testovej štatistiky. Ako testovú štatistiku tento test používa vážený súčet štvorcov rozdielu výberového priemeru v populácii a váženého celkového priemeru zo všetkých populácií, pričom ako váhy sú použité podiely počtu pozorovaní v populáciách a výberového rozptylu v jednotlivých populáciách, tj.

$$K = \sum_{i=1}^{I} \frac{(\overline{Y}_{i.} - \overline{Y})^2 n_i}{S_i^2}, \tag{3.2}$$

kde

$$\overline{Y} = \sum_{i=1}^{I} \frac{\overline{Y}_{i.} n_i}{S_i^2} / \sum_{i=1}^{I} \frac{n_i}{S_i^2}.$$

Krutchkoff sa nesnažil nájsť rozdelenie štatistiky K. Avšak navrhol, ako môžeme nájsť kritické hodnoty pre túto štatistiku pomocou simulácií. Simulácie kritických hodnôt prebiehajú tak, že pre dané n_i, $i = 1, ..., I$ nasimulujeme nezávislé rovnako rozdelené náhodné veličiny X_{ij} z normovaného normálneho rozdelenia (tj. z $N(0,1)$) pre všetky $i = 1, ..., I$, $j = 1, ..., n_i$. Pre tieto veličiny spočítame štatistiku K definovanú vzorcom 3.2. Celý tento proces zopakujeme 100000-krát. Ako kritickú hodnotu potom zoberieme $K_{[(1-\alpha)100000]}$, čo je $(1-\alpha)100000$ poriadková štatistika štatistiky K.

Za platnosti hypotézy H_0 sú stredné hodnoty výberových priemerov zhodné, preto sa výberové priemery a teda aj celkový priemer nebudú veľmi líšiť. V prípade, že hypotéza H_0 neplatí, budú tieto rozdiely oveľa väčšie. Z toho dôvodu má štatistika K menšie hodnoty za hypotézy H_0, ako ak táto hypotéza neplatí. Preto testovanie H_0 prebieha tak, že zamietame túto hypotézu pre $K > K_{[(1-\alpha)100000]}$. Tento test budeme značiť **Krutchkoffov test**.

3.4 Weerahandiho test

S. Weerahandi v článku Weerahandi (1995) odvodil test na testovanie hypotézy H_0 založený na zobecnenej p-hodnote. Tento test predpokladá normalitu populácií, ale nepredpokladá homoskedasticitu medzi populáciami. Heteroskedasticita sa prejaví pri výpočte p-hodnoty tohto testu. Pre odhady neznámych rozptylov σ_i^2, $i = 1, ..., I$ Weerahandi používa maximálne vierohodné odhady, tj. odhady:

$$W_i^2 = \frac{1}{n_i} \sum_{j=1}^{n_i} (Y_{ij} - \overline{Y}_{i.})^2 = \frac{(n_i - 1)}{n_i} S_i^2 \quad i = 1, ..., I.$$

Ďalej definujme funkciu $\widetilde{S}_b$ predpisom:

$$\widetilde{S}_b = \widetilde{S}_b(x_1, ..., x_I) = \sum_{i=1}^{I} \frac{n_i \overline{Y}_{i.}^2}{x_i} - [(\sum_{i=1}^{I} \frac{n_i \overline{Y}_{i.}}{x_i})^2 / \sum_{i=1}^{I} \frac{n_i}{x_i}].$$

Povšimnime si, že štatistika T_W definovaná vzorcom 3.1 a použitá pri jednom z Akritasových testov je rovná $\widetilde{S}_b(\sigma_1^2, ..., \sigma_I^2)$. Weerahandi odvodil, že p-hodnota príslušná pre testovanie hypotézy H_0 je rovná rozdielu jednotky a strednej hodnoty distribučnej funkcie F-rozdelenia s $I-1$ a $N-I$ stupňami voľnosti v bode $\widetilde{S}_b(n_1 z_1, ..., n_I z_I)$, kde z_i, $i = 1, ..., I$ sú rovné podielu maximálne vierohodného odhadu rozptylu i-tej populácie a súčinu $I - i + 1$ náhodných veličín z beta rozdelenia, presnejšie povedané:

$$\begin{aligned} p &= 1 - E(F_{I-1,N-I}(\frac{N-I}{I-1}\widetilde{S}_b[n_1 z_1, ..., n_I z_I])) \qquad (3.3) \\ z_1 &= \frac{W_1^2}{B_1 B_2 ... B_{I-1}} \\ z_i &= \frac{W_i^2}{(1-B_{i-1})B_i B_{i+1}...B_{I-1}} \quad i = 2, ..., I, \end{aligned}$$

kde $F_{I-1,N-I}$ označuje distribučnú funkciu F-rozdelenia s $I-1$ a $N-I$ stupňami voľnosti a veličiny B_j, $j = 1, ..., I-1$ sú náhodné veličiny z beta rozdelenia definované predpismi:

$$B_j \sim beta(\sum_{i=1}^{j} \frac{(n_i - 1)}{2}, \frac{(n_{j+1}-1)}{2}), \quad j = 1, ..., I-1. \qquad (3.4)$$

Vo vzorci 3.3 môžeme strednú hodnotu a teda aj p-hodnotu odhadnúť pomocou Monte Carlo metódy. To urobíme tak, že 1000-krát nasimulujeme náhodné veličiny B_j definované predpisom 3.4 a v každom nastavení dosadíme B_j do vzorca 3.3. Potom spočítame príslušnú hodnotu distribučnej funkcie F-rozdelenia a $\mathrm{E}(F_{I-1,N-I}(.))$ aproximujeme výberovým priemerom týchto hodnôt, tj.

$$EF_{I-1,N-I}(\widetilde{S}_b(.)) \doteq \frac{1}{1000}\sum_{i=1}^{1000} F_{I-1,N-I}(\widetilde{S}_b(.)), \quad j = 1, ..., 1000.$$

Ak je táto p-hodnota $< \alpha$, potom zamietame hypotézu H_0. Tento test budeme značiť ako **Weerahandiho test**.

3.5 Brownov-Forsytheov test

Ďalší test, ktorý predpokladá normalitu, ale nepredpokladá zhodu rozptylov medzi populáciami navrhli Brown a Forsythe. Heteroskedasticita sa prejavuje aj v testovej štatistike, ale aj v kritických hodnotách tejto štatistiky. Brown a Forsythe navrhli testovať hypotézu H_0 pomocou podielu $(I-1)$-násobku priemerného súčtu štvorcov medzi populáciami a váženého súčtu odhadov rozptylov vážených váhami $\frac{N-n_i}{N}$, tj. testovať hypotézu H_0 pomocou štatistiky F^* definovanej predpisom

$$F^* = \sum_{i=1}^{I} n_i(\overline{Y}_{i.} - \overline{Y}_{..})^2 / \sum_{i=1}^{I}(1 - \frac{n_i}{N})S_i^2.$$

V článku Brown and Forsythe (1974) je odporučené túto štatistiku aproximovať F-rozdelením s $I-1$ a f stupňami voľnosti, kde f je definované predpisom

$$\frac{1}{f} = \sum_{i=1}^{I} \frac{c_i^2}{n_i - 1}, \quad \text{kde} \quad c_i = \frac{(1-\frac{n_i}{N})S_i^2}{\sum_{j=1}^{I}(1-\frac{n_j}{N})S_j^2}.$$

Poznamenajme, že v prípade vyvážených dát je štatistika F^* zhodná s F štatistikou F-testu.

V prípade, že hypotéza H_0 neplatí, nadobúda čitateľ štatistiky F^* väčších hodnôt ako v prípade platnosti H_0. Z toho dôvodu hypotézu H_0 zamietame, ak $F^* > F_{I-1,f}(1-\alpha)$, kde $F_{I-1,f}(1-\alpha)$ označuje $(1-\alpha)$-kvantil F-rozdelenia s $I-1$ a f stupňami voľnosti. Tento test budeme označovať ako **Brownov-Forsytheov test**.

3.6 Boxov test

Ďalší test, ktorý testuje hypotézu H_0, je odvodený v článku Box (1954). Tento test, tak isto ako F-test, používa ako testovú štatistiku podiel priemerného súčtu štvorcov medzi populáciami a priemerného súčtu štvorcov v populáciách, tj.

$$F = \frac{MST}{MSE}.$$

Rozdiel oproti F-testu nastáva v použitej kritickej hodnote, kde je zohľadnená aj prípadná heteroskedasticita. Box odporučil túto štatistiku aproximovať b násobkom F-rozdelenia s h' a h stupňami voľnosti (alebo inak napísané: F/b má približne F -rozdelenie s h a h' stupňami voľnosti), kde b, h', h sú definované predpismi

$$\begin{aligned}
b &= \frac{N-I}{N(I-1)} \sum_{i=1}^{I}(N-n_i)\sigma_i^2 / \sum_{i=1}^{I}(n_i-1)\sigma_i^2 \\
h' &= [\sum_{i=1}^{I}(N-n_i)\sigma_i^2]^2 / [(\sum_{i=1}^{I} n_i\sigma_i^2)^2 + N\sum_{i=1}^{I}(N-2n_i)\sigma_i^4] \\
h &= [\sum_{i=1}^{I}(n_i-1)\sigma_i^2]^2 / [\sum_{i=1}^{I}(n_i-1)\sigma_i^4].
\end{aligned}$$

Pretože σ_i^2, $i = 1, ..., I$ sú neznáme parametre, nahradíme ich odhadmi S_i^2, $i = 1, ..., I$. Namiesto σ_i^4, $i = 1, ..., I$ použijeme konzistentné odhady $(S_i^2)^2$, $i = 1, ..., I$.

Ak hypotéza H_0 neplatí, priemerný súčet štvorcov medzi populáciami MST nadobúda väčších hodnôt ako priemerný súčet štvorcov v populáciách MSE. Za platnosti hypotézy H_0 sú veličiny MST a MSE podobné. Preto zamietame H_0, ak $F > bF_{h',h}(1-\alpha)$, kde $F_{h',h}(1-\alpha)$ označuje $(1-\alpha)$-kvantil F-rozdelenia s h' a h stupňami voľnosti. Tento test budeme označovať **Boxov test**.

3.7 Welshov test

Tiež B.L. Welsh sa zaoberal problematikou testovania H_0 za predpokladu normality a heteroskedasticity populácií. V jeho článku Welsh (1951) odvodil test, ktorý ako testovú štatistiku používa vážený súčet štvorcov rozdielu výberových priemerov a váženého celkového priemeru, pričom váhy závisia na počte pozorovaní a výberovom rozptyle populácií, tj.

$$v^2 = \frac{\sum_{i=1}^{I} w_i(\overline{Y}_{i.} - \widehat{Y})^2/(I-1)}{[1 + \frac{2(I-2)}{(I^2-1)} \sum_{i=1}^{I} \frac{1}{f_i}(1 - \frac{w_i}{W})^2]},$$

kde

$$\begin{aligned} w_i &= \frac{n_i}{S_i^2} \\ f_i &= n_i - 1 \quad i = 1, ..., I \\ W &= \sum_{i=1}^{I} w_i \\ \widehat{Y} &= \sum_{i=1}^{I} (w_i \overline{Y}_{i.})/W. \end{aligned}$$

Welsh odporučil štatistiku v^2 aproximovať F-rozdelením s $I-1$ a $\widehat{f_2}$ stupňami voľnosti, kde $\widehat{f_2}$ je definované predpisom

$$\widehat{f_2} = [\frac{3}{(I^2-1)} \sum_{i=1}^{I} \frac{1}{f_i}(1 - \frac{w_i}{W})^2]^{-1}.$$

Za platnosti H_0 sa výberové priemery a teda aj vážený celkový priemer líšia len o málo. Ak hypotéza H_0 neplatí, sú tieto rozdiely oveľa väčšie. Preto zamietame H_0, ak štatistika v^2 prekročí $(1-\alpha)$-kvantil F-rozdelenia s $I-1$ a $\widehat{f_2}$, tj. keď $v^2 > F_{I-1,\widehat{f_2}}(1-\alpha)$, kde $F_{I-1,\widehat{f_2}}(1-\alpha)$ označuje $(1-\alpha)$-kvantil F-rozdelenia s $I-1$ a $\widehat{f_2}$ stupňami voľnosti. Tento test budeme označovať ako **Welshov test**.

3.8 Chenov test

Ďalší test pre testovanie hypotézy H_0, ktorý predpokladá normalitu a heteroskedasticitu populácií, je odvodený v článku Chen and Chen (1998). Tento test je odvodený iba pre vyvážený prípad, tj. pre $n_1 = ... = n_I = n$. Pracuje na obdobe dvojkrokového výberu, pričom ako druhý výber berie n-té pozorovanie z každej populácie. Heteroskedasticita sa prejavuje iba v testovej štatistike.

Položme $n_0 = n - 1$ a definujme pre každú populáciu orezaný priemer a orezaný výberový rozptyl z prvých n_0 pozorovaní z každej populácie predpisom

$$\begin{aligned} \overline{Y}_{i.}(n_0) &= \sum_{j=1}^{n_0} \frac{Y_{ij}}{n_0} \\ S_i^2(n_0) &= \sum_{j=1}^{n_0} \frac{(Y_{ij} - \overline{Y}_{i.}(n_0))^2}{(n_0 - 1)} \qquad i = 1, ..., I. \end{aligned}$$

Ďalej ako $S_{[I]}^2$ budeme označovať najväčší orezaný výberový rozptyl, tj.

$$S_{[I]}^2 = \max_{i=1,...,I} S_i^2(n_0).$$

Ešte dodefinujme váhy

$$\begin{aligned} U_i &= \frac{1}{n} + \frac{1}{n}\sqrt{\frac{1}{n_0}[\frac{S_{[I]}^2}{S_i^2(n_0)} - 1]} \\ V_i &= \frac{1}{n} - \frac{1}{n}\sqrt{n_0[\frac{S_{[I]}^2}{S_i^2(n_0)} - 1]} \end{aligned}$$

$$W_{ij} = \begin{cases} U_i & , \quad 1 \le j \le n_0 \\ V_i & , \quad j = n \end{cases}$$

a definujme populačný vážený priemer a celkový vážený priemer

$$\begin{aligned} \widetilde{Y}_{i.} &= \sum_{j=1}^{n} W_{ij} Y_{ij} \\ \widetilde{Y}_{..} &= \sum_{i=1}^{I} \frac{\widetilde{Y}_{i.}}{I}. \end{aligned}$$

Hypotézu H_0 budeme testovať pomocou štatistiky $\widetilde{F}^1$, ktorá je založená na súčte štvorcov rozdielu váženého priemeru v populáciách a celkového váženého priemeru násobených podielom počtu pozorovaní a najväčšieho orezaného výberového rozptylu, tj.

$$\widetilde{F}^1 = \sum_{i=1}^{I} \frac{n(\widetilde{Y}_{i.} - \widetilde{Y}_{..})^2}{S_{[I]}^2}$$

Chen a Chen vo svojom článku tiež odvodili, že štatistika $\widetilde{F}^1$ má rovnaké rozdelenie ako $I - 1$ násobok výberového rozptylu I nezávislých rovnako rozdelených náhodných veličín z t-rozdelenia s $n_0 - 1 = n - 2$ stupňami voľnosti, tj. ako

$$Q = \sum_{i=1}^{I} (t_i - \bar{t})^2, \tag{3.5}$$

kde

$$t_i \sim t_{n-2} \quad \text{a} \quad \bar{t} \ = \frac{1}{I} \sum_{i=1}^{I} t_i.$$

Kritickú hodnotu $\widetilde{F}^1$ a teda aj Q musíme vypočítať pomocou simulácií. To urobíme tak, že 100000-krát vygenerujeme nezávislé rovnako rozdelené náhodné veličiny $t_i \sim t_{n-2}$, $i = 1, ..., I$ a pre každé opakovanie spočítame štatistiku Q definovanú vzorcom 3.5. Ako kritickú hodnotu testu potom zoberieme $Q_{[(1-\alpha)100000]}$, čo je $(1-\alpha)100000$ poriadková štatistika simulácie.

Pretože za hypotézy H_0 sú vážené priemery v populáciách $\widetilde{Y}_{i.}$, $i = 1, ..., I$ podobné, preto sa od nich aj celkový vážený priemer $\widetilde{Y}_{..}$ nebude veľmi líšiť. Ak hypotéza neplatí, potom sú rozdiely v čitateli sumy štatistiky $\widetilde{F}^1$ oveľa väčšie. Preto hypotézu H_0 zamietame pre $\widetilde{F}^1 > Q_{[(1-\alpha)100000]}$. Tento test budeme označovať ako **Chen test**.

3.9 Kombi test

Tento test vznikol kombináciou F-testu a Krutchkoffovho testu (tento test je najuniverzálnejší medzi uvažovanými testami za platnosti normality populácií). V tomto teste využijeme ešte Bartlettov test na testovanie zhody rozptylov populácií, ktorý je citlivý na poruchy normality.

Ak Bartlettov test zamietne hypotézu o zhode rozptylov populácií (na hladine významnosti α), budeme hypotézu H_0 testovať pomocou Krutchkoffovho testu, v opačnom prípade použijeme na testovanie H_0 F-test. Tento test budeme označovať **Kombi test**.

Tabuľka 1: Zhrnutie predpokladov testov.

test	homoskedasticita	normalita	$I \to \infty$	simulácie	$n_1 = ... = n_I$
F-test	áno	áno	nie	nie	nie
Akr.nevážený test	nie	nie	áno	nie	nie
Akr.vážený test	nie	nie	áno	nie	nie
Krutchkoffov test	nie	áno	nie	áno	nie
Weerahandiho test	nie	áno	nie	áno	nie
Boxov test	nie	áno	nie	nie	nie
B-F test	nie	áno	nie	nie	nie
Welshov test	nie	áno	nie	nie	nie
Chenov test	nie	áno	nie	nie	áno
Kombi test	nie	áno	nie	nie	nie

4 Nastavenie simulácií

Budeme generovať nezávislé náhodné veličiny Y_{ij} z rozdelenia so strednou hodnotou μ_i a rozptylom σ_i^2, $i = 1, ..., I$, $j = 1, ..., n_i$. Pretože chceme skúmať testy nielen za predpokladu normality, ale aj pri poruche normality, budeme okrem normálneho rozdelenia populácií uvažovať ešte aj rovnomerné, dvojne exponenciálne a logaritmicko-normálne rozdelenie populácií. Strednú hodnotu i-tej populácie môžeme zapísať v tvare $\mu_i = \mu + \alpha_i$, $i = 1, ..., I$. Našou úlohou je testovanie hypotézy $H_0 : \mu_1 = ... = \mu_I$, ktorá je ekvivalentná s testovaním hypotézy $H_0' : \alpha_1 = ... = \alpha_I = 0$. Pre všetky uvažované testy budeme skúmať ich dodržiavanie hladiny významnosti $\alpha = 0.05$ pre test H_0. Tiež budeme skúmať aj to, akú majú tieto testy silu za platnosti rôznych alternatív. V celej simulácii môžeme nastaviť veľa parametrov, preto budeme uvažovať len niektoré prípady.

Pre počet populácií I budeme uvažovať štyri prípady. Sú to $I = 3, 5, 10, 25$. Tiež sa budeme musieť obmedziť na voľbu počtu pozorovaní pre jednotlivé populácie. Budeme uvažovať ako aj vyvážený, tak aj nevyvážený prípad. Pre vyvážené prípady budeme uvažovať štyri nastavenia n a to $n = n_1 = ... = n_I = 5, 10, 25, 50$. Pre nevyvážené dáta budeme uvažovať dva rôzne prípady. Prvý prípad je, že jedno n_i sa líši od ostatných (pre $I = 3, 5$ som pre $n_1, ..., n_I$ zvolil štyri kombinácie 5 a 10, pre $I = 10$ štyri kombinácie 15 a 20 a pre $I = 25$ štyri kombinácie 30 a 35). V druhom prípade budeme uvažovať, že $n_1, ..., n_I$ tvorí aritmetickú postupnosť s krokom jedna alebo dva (výnimkou je prípad pre tri populácie), pričom hodnoty jednotlivých n_i sa budú meniť podľa toho, aké I zvolíme (pre $I = 3$ som pre $n_1, ..., n_I$ zvolil postupnosť $5, 8, 10$ resp. $10, 15, 20$, pre $I = 5$ som pre $n_1, ..., n_I$ zvolil postupnosť $6 : 10$ resp. $2 * (6 : 10)$, pre $I = 10$ som pre $n_1, ..., n_I$ zvolil postupnosť $16 : 25$ resp. $2 * (16 : 25)$ a pre $I = 25$ som pre $n_1, ..., n_I$ zvolil postupnosť $11 : 35$ a $26 : 50$). Ďalší parameter, ktorý môžeme v simuláciách ľubovoľne nastaviť, je stredná hodnota jednotlivých populácií. Pretože všetky testové štatistiky sú invariantné voči parametru posunutia, preto bez újmy na všeobecnosti budeme vždy uvažovať, že $\mu = 0$ (u logaritmicko-normálneho rozdelenia to bolo dosiahnuté posunutím o strednú hodnotu, takže medzi logaritmicko-normálne rozdelenie budem zaraďovať aj posunuté logaritmicko normálne rozdelenie). Okrem logaritmicko-normálneho rozdelenia sú ostatné tri uvažované rozdelenia symetrické okolo nuly, preto sa obmedzíme na prípad, že $\alpha_i \geq 0$. Pri testovaní sily rôznych testov budeme uvažovať tri prípady nastavenia $\alpha_1, ..., \alpha_I$. Prvý prípad bude, že práve jedno $\alpha_i > 0$ a ostané sú rovné nule. Túto situáciu budeme označovať ako alternatíva A_1. V druhom prípade budeme uvažovať, že medzi α_i je lineárna závislosť a tento prípad budeme značiť ako alternatíva A_2. Pod alternatívou A_3 budeme rozumieť tretí prípad nastavenia $\alpha_1, ..., \alpha_I$. Budeme ho uvažovať iba pre $I = 10, 25$ a položíme $\alpha_1 = ... = \alpha_{\lfloor I/2 \rfloor} = 0$ a $\alpha_{\lfloor I/2 \rfloor + 1} = ... = \alpha_I > 0$, kde $\lfloor x \rfloor$ označuje celú časť čísla x. Silofunkciu testov pre všetky tri alternatívy budeme odhadovať v troch bodoch. Rôzne hodnoty nenulových α_i budeme voliť tak, aby silofunkcia F-testu pri voľbe

$\sigma_1^2 = ... = \sigma_I^2 = 1$ a normálne rozdelených populácií bola približne:
1.) medzi 0,175 a 0.275
2.) medzi 0,375 a 0.475
3.) medzi 0,775 a 0.900.
Posledný parameter, ktorý môžeme ľubovoľne meniť, je rozptyl populácií. Budeme uvažovať homoskedastický, ale aj heteroskedastický prípad. Pretože pre náhodné veličiny $Y_{ij}^* = k * Y_{ij}$, kde $k \in R$ je ľubovoľná konštanta, sú všetky testové štatistiky zhodné s testovými štatistikami vypočítanými z pôvodných veličín Y_{ij}, preto pre homoskedastický prípad budeme uvažovať iba prípad, že $\sigma_1^2 = ... = \sigma_I^2 = 1$. Pre heteroskedastické prípady budeme uvažovať dva druhy heteroskedasticity. Prvý prípad je, že buď $\sigma_1^2 = 5$ alebo $\sigma_I^2 = 5$ a ostatné sú rovné jednej. V druhom prípade $\sigma_1^2, ..., \sigma_I^2$ tvorí aritmetickú postupnosť od 1 do I.

Pre generovanie náhodných veličín Y_{ij} bol použitý software **R 2.4.0**. Hladiny významnosti α hypotézy H_0 a sily testov za platnosti jednej z alternatív A_1, A_2 alebo A_3 boli odhadnuté tak, že sme 1000-krát vygenerovali náhodné veličiny Y_{ij}. Ako odhady hladiny významnosti alebo sily testov boli použité relatívne početnosti zamietnutia hypotézy H_0 jednotlivými testami. Okrem klasickej sily testov som ešte navyše nasimuloval silu týchto testov v prípade, že by sme ich kritické hodnoty nahradili $(1-\alpha)$-výberovým kvantilom ich testovej štatistiky v prípade platnosti hypotézy H_0. Všetky takto nasimulované odhady boli uložené do tabuliek tak, že všetky odhady v jednom stĺpci tabuľky boli odhadnuté z rovnakých nasimulovaných náhodných veličín Y_{ij}. To nám umožňuje jednotlivé odhady v rovnakom stĺpci tabuľky porovnávať. Preto, aj keď nemáme presné hodnoty odhadov síl testov (resp. hladiny významnosti), môžeme sily testov (resp. hladiny významnosti) v rámci jednej simulovanej situácie porovnávať. Pretože simulovaných prípadov je strašne veľa, všetky tabuľky s výsledkami by sa do tejto diplomovej práce nevošli. Z toho dôvodu môžeme všetky moje výsledky simulácií nájsť na priloženom CD. V tejto práci sú vybrané len niektoré simulované prípady. Navyše všetky odhady uvedené v tejto práci boli presimulované na 10000 opakovaní, aby sa dosiahlo väčšej presnosti odhadov. Na priloženom CD taktiež môžeme nájsť zdrojový kód k programu **R** pre všetky simulácie.

Poznámka 4.1. Z uskutočnených simulácií je ťažké povedať, či daný test presne dodržuje hladinu významnosti. U normálne rozdelených homoskedastických populácií by sa to ešte dalo určiť pomocou F-testu, ktorý je presný a preto by mal vždy dodržať hladinu významnosti. V ostatných prípadoch už žiadny test nie je presný. Preto to, či test dodržuje hladinu významnosti môžeme chápať ako testovanie hypotézy $H' : \alpha = 0.05$ proti alternatíve $A' : \alpha \neq 0.05$. Interval spoľahlivosti pre α môžeme odhadnúť nasledujúcou úvahou. Za platnosti hypotézy H' môžeme každú simuláciu chápať ako výber z alternatívneho rozdelenia s parametrom $p = 0.05$. Preto s pomocou centrálnej limitnej vety môžeme určiť interval spoľahlivosti pre $\alpha = 0.05$ ako $(0.05 - \sqrt{\frac{0.05(1-0.05)}{n'}} u_{0,1}(\frac{\alpha'}{2}), 0.05 + \sqrt{\frac{0.05(1-0.05)}{n'}} u_{0,1}(\frac{\alpha'}{2}))$, kde n' je počet simulácií, α' je hladina významnosti pre test H' a $u_{0,1}(\frac{\alpha'}{2})$ je $(1 - \frac{\alpha'}{2})$-kvantil normovaného normálneho rozdelenia, tj. $N(0,1)$. Pre hodnoty $\alpha' = 0.01$ a $n' = 1000$ je tento interval spoľahlivosti približne rovný (0.032, 0.068), pre $\alpha' = 0.01$ a $n' = 10000$ je približne rovný (0.044, 0.056). Preto ak poviem, že test dodržuje hladinu významnosti, tak to nemusí znamenať, že jeho hladina významnosti testovania hypotézy H_0 je presne rovná teoretickej hladine významnosti $\alpha = 0.05$, ale znamená to, že sa od tejto teoretickej hodnoty významne nelíši. Hodnotu hladiny významnosti α, ktorú spočítame pomocou simulácie, budeme nazývať odhadovaná hladina významnosti. Ak budem hovoriť o teste, že podhodnocuje (resp. nadhodnocuje alebo prekračuje) hladinu významnosti, tak tým budeme rozumieť, že odhadovaná hladina významnosti tohto testu je významne menšia (resp. väčšia) ako teoretická hodnota, ktorú by sme chceli dosiahnuť.

5 Výsledky simulácií za predpokladu normality populácií

5.1 Dodržiavanie hladiny významnosti α

F-test

Pre homoskedastické prípady tento test dodržuje hladinu významnosti α. Problém začína nastávať, keď sa v dátach začína objavovať heteroskedasticita. Pri vyvážených prípadoch nie je vplyv heteroskedasticity až taký veľký, ale je badateľný, pre nevyvážené prípady už tento test prakticky vôbec nedodržuje hladinu významnosti. Preto by som ho určite neodporučil použiť v nevyvážených prípadoch, kde som si nie istý homoskedasticitou.

Tabuľka 2: Dodržiavanie hladiny významnosti α=5% F-testu vo vyváženom prípade s normálnymi populáciami pre rôzne rozptyly.

I	**I=3**		**I=5**		**I=25**	
rozptyl	n=5	n=50	n=5	n=50	n=5	n=50
1, ... ,1	5.07	4.91	5.16	5.26	5.17	4.86
1,...,1,5	6.94	6.71	8.22	7.06	7.79	7.29

Tabuľka 3: Dodržiavanie hladiny významnosti α=5% F-testu v nevyváženom prípade s normálnymi populáciami pre rôzne rozptyly.

I	**I=3**				**I=25**			
n=(...)	5,5,10	10,10,5	5,8,10	10,15,20	24x30,35	24x35,30	11:35	26:50
$\sigma^2 = (1, ..., 1)$	4.95	5.25	5.05	5.19	4.73	4.86	4.56	5.35
$\sigma^2 = (1, ..., 1, 5)$	2.33	14.62	3.51	3.19	6.45	8.74	5.29	6.12
$\sigma^2 = (5, 1, ..., 1)$	10.78	4.42	13.38	12.31	7.59	7.91	11.61	10.38

Akritasov nevážený test

Pripomeňme, že Akritasov nevážený test bol odvodený za predpokladu, že počet populácií ide do nekonečna. Tento test vo väčšine simulovaných prípadov mierne nadhodnocuje hladinu významnosti. S rastúcim počtom pozorovaní alebo rastúcim počtom populácií sa toto nadhodnocovanie zmenšuje a pre veľký počet pozorovaní alebo počet populácií už tento test dodržuje hladinu významnosti. Aj keď za predpokladu homoskedasticity dodržiava α menej ako F-test, v heteroskedastických prípadoch sa správa podstatne lepšie ako F-test (často dodržiava hladinu významnosti aj tam, kde F-test zlyháva). Ak sme si nie istí homoskedasticitou, tak by som ho určite uprednostnil pred F-testom.

Tabuľka 4: Dodržiavanie hladiny významnosti α=5% Akritasovho neváženého testu vo vyváženom prípade s normálnymi populáciami pre rôzne rozptyly.

I	**I=3**			**I=10**			**I=25**	
rozptyl	n=5	n=10	n=50	n=5	n=10	n=50	n=5	n=50
1, ... ,1	6.22	5.76	4.67	5.65	6.07	5.66	4.32	5.18
1,...,1,5	7.14	5.61	4.50	7.66	6.60	5.47	5.93	5.69

Tabuľka 5: Dodržiavanie hladiny významnosti α=5% Akritasovho neváženého testu v nevyváženom prípade s normálnymi populáciami pre rôzne rozptyly.

I	**I=3**				**I=25**			
n=(...)	5,5,10	10,10,5	5,8,10	10,15,20	24x30,35	24x35,30	11:35	26:50
$\sigma^2 = (1, ..., 1)$	6.09	5.42	5.93	5.47	5.71	5.87	5.76	5.45
$\sigma^2 = (1, ..., 1, 5)$	3.97	9.23	4.56	4.10	6.26	6.53	6.07	5.92
$\sigma^2 = (5, 1, ..., 1)$	9.09	5.11	8.95	6.70	5.97	6.19	6.48	5.89

Akritasov vážený test

Aj Akritasov vážený test bol odvodený za predpokladu, že počet populácií ide do nekonečna. Avšak navyše predpokladá, že počet pozorovaní ide tiež do nekonečna a to dokonca rýchlejšie ako počet populácií. Preto tento test má obrovskú nevýhodu v tom, že potrebuje aby bol počet pozorovaní mnohonásobne vyšší ako je počet populácií k tomu, aby dodržal hladinu významnosti, napr. pre tri populácie by mal byť počet pozorovaní v každej populácii aspoň päťdesiat. Preto je použiteľný iba v experimentoch, kde máme veľký počet pozorovaní pre každú populáciu.

Tabuľka 6: Dodržiavanie hladiny významnosti α=5% Akritasovho váženého testu vo vyváženom prípade s normálnymi populáciami pre rôzne rozptyly.

I	**I=3**			**I=5**		**I=25**	
rozptyl	n=10	n=25	n=50	n=25	n=50	n=25	n=50
1, ... ,1	7.82	5.86	5.73	7.58	6.37	11.32	9.18
1,...,1,5	8.05	6.00	5.47	7.08	6.72	12.06	8.52

Krutchkoffov test

Tento test má problém s dodržiavaním hladiny významnosti iba v nevyvážených heteroskedastických prípadoch s malým počtom pozorovaní v populáciách. Chyby v nedodržaní hladiny významnosti môžu byť spôsobené simuláciami, ktoré sa používajú nie len pri simulovaní dodržiavania hladiny významnosti, ale tiež pri počítaní kritických hodnôt Krutchkoffovho testu.

Tabuľka 7: Dodržiavanie hladiny významnosti α=5% Krutchkoffovho testu vo vyváženom prípade s normálnymi populáciami pre rôzne rozptyly.

I	**I=3**		**I=25**	
rozptyl	n=5	n=50	n=5	n=50
1, ... ,1	4.53	4.88	5.02	5.58
1,...,1,5	5.45	4.78	5.09	5.13

Tabuľka 8: Dodržiavanie hladiny významnosti α=5% Krutchkoffovho testu v nevyváženom prípade s normálnymi populáciami pre rôzne rozptyly.

I	**I=3**				**I=25**			
n=(...)	5,5,10	10,10,5	5,8,10	10,15,20	24x30,35	24x35,30	11:35	26:50
$\sigma^2 = (1, ..., 1)$	5.20	4.82	5.10	4.82	4.90	5.10	4.97	4.80
$\sigma^2 = (1, ..., 1, 5)$	4.02	6.82	4.66	4.78	4.93	4.83	4.59	5.36
$\sigma^2 = (5, 1, ..., 1)$	5.82	4.50	6.27	5.62	5.11	5.29	4.86	5.20

Weerahandiho test

Pre tri populácie a malý počet pozorovaní tento test mierne podhodnocuje hladinu významnosti. Pre päť a viac populácií a malý počet pozorovaní tento test nadhodnocuje hladinu významnosti. Toto nedodržiavanie sa s rastúcim počtom populácií a pevným počtom pozorovaní výrazne zhoršuje. Ak chceme, aby tento test aspoň približne dodržiaval hladinu významnosti α, musíme zaručiť, aby pre väčší počet populácií ako je päť, bol počet pozorovaní v každej populácii minimálne 2-krát väčší ako počet populácií I. Na druhú stranu výhodou Weerahandiho testu by mohlo byť to, že hladina významnosti odhadnutá za homoskedasticity a heteroskedasticity v prípade rovnakého počtu pozorovaní sa veľmi nelíši.

Tabuľka 9: Dodržiavanie hladiny významnosti α=5% Weerahandiho testu vo vyváženom prípade s normálnymi populáciami pre rôzne rozptyly.

I	**I=3**		**I=5**		**I=10**		**I=25**	
rozptyl	n=5	n=50	n=5	n=50	n=10	n=25	n=25	n=50
1, ... ,1	3.83	5.10	6.40	5.13	8.94	6.09	8.23	6.72
1,...,1,5	4.23	4.88	6.88	5.03	7.92	6.06	8.68	6.06

Tabuľka 10: Dodržiavanie hladiny významnosti α=5% Weerahandiho testu v nevyváženom prípade s normálnymi populáciami pre rôzne rozptyly.

I	**I=3**		**I=5**				**I=25**	
n=(...)	5,5,10	10,10,5	4x5,10	4x10,5	6:10	2*(6:10)	11:35	26:50
$\sigma^2 = (1, ..., 1)$	4.56	4.58	6.81	6.36	6.12	5.82	9.09	6.83
$\sigma^2 = (1, ..., 1, 5)$	3.84	4.97	5.71	6.40	6.08	5.58	9.63	7.57
$\sigma^2 = (5, 1, ..., 1)$	5.35	4.17	6.98	5.82	6.54	5.47	9.27	7.04

Brownov-Forsytheov test

Nie vždy sa dá povedať, že tento test dodržuje hladinu významnosti. Ale na druhú stranu sa nedá povedať, že v nejakom simulovanom prípade výrazne nedodržal hladinu významnosti. Preto sa dá použiť vo všetkých prípadoch a hlavne vtedy, ak nám nevadí malé nedodržanie hladiny α, ale veľké nedodržanie hladiny α je pre nás už problém. Ďalej si je dobré povšimnúť, že vo vyvážených prípadoch sa so zväčšujúcim počtom pozorovaní tento test výrazne blíži (v dodržiavaní hladiny významnosti) k F-testu, čo okrem iného spôsobuje jeho horšie správanie sa pri väčšom počte pozorovaní v heteroskedastických prípadoch.

Tabuľka 11: Dodržiavanie hladiny významnosti α=5% Brownovho-Forsytheovho testu vo vyváženom prípade s normálnymi populáciami pre rôzne rozptyly.

I	**I=3**		**I=5**		**I=25**	
rozptyl	n=5	n=50	n=5	n=50	n=5	n=50
1, ... ,1	4.04	4.70	4.21	4.72	4.32	5.34
1,...,1,5	5.61	5.91	6.40	7.10	6.26	7.41

Boxov test

Táto vylepšená verzia F-testu má v homoskedastických prípadoch problém akurát pre menší počet pozorovaní v populáciách ako je počet populácií, keď

Tabuľka 12: Dodržiavanie hladiny významnosti α=5% Brownovho-Forsytheovho testu v nevyváženom prípade s normálnymi populáciami pre rôzne rozptyly.

I	**I=3**				**I=25**			
n=(...)	5,5,10	10,10,5	5,8,10	10,15,20	24x30,35	24x35,30	11:35	26:50
$\sigma^2 = (1, ..., 1)$	4.82	4.78	4.75	4.92	4.89	4.71	4.85	5.42
$\sigma^2 = (1, ..., 1, 5)$	5.57	6.63	5.54	6.19	7.17	7.12	7.27	7.34
$\sigma^2 = (5, 1, ..., 1)$	5.52	5.48	6.67	7.00	7.91	7.26	7.01	7.55

podhodnocuje hladinu významnosti. Vylepšenie oproti F-testu sa uplatní hlavne v nevyvážených heteroskedastických prípadoch, kde F-test absolutne zlyháva, naopak Boxov test sa zdá, že tu dodržuje hladinu významnosti spoľahlivo až na prípady, keď máme málo populácií a v populáciách s menším počtom pozorovaní je väčší rozptyl. Preto by som tento test nepoužíval iba v prípadoch, kde $n_i < I$ pre nejaké $i = 1, ..., I$.

Tabuľka 13: Dodržiavanie hladiny významnosti α=5% Boxovho testu vo vyváženom prípade s normálnymi populáciami pre rôzne rozptyly.

I	**I=3**		**I=5**		**I=25**	
rozptyl	n=5	n=50	n=5	n=50	n=5	n=50
1, ... ,1	4.30	4.90	4.03	4.64	2.48	4.65
1,...,1,5	5.49	4.99	5.27	4.99	4.04	4.78

Tabuľka 14: Dodržiavanie hladiny významnosti α=5% Boxovho testu v nevyváženom prípade s normálnymi populáciami pre rôzne rozptyly.

I	**I=3**				**I=25**			
n=(...)	5,5,10	10,10,5	5,8,10	10,15,20	24x30,35	24x35,30	11:35	26:50
$\sigma^2 = (1, ..., 1)$	4.62	4.45	4.88	4.83	4.77	4.72	4.69	4.49
$\sigma^2 = (1, ..., 1, 5)$	4.53	7.05	4.54	4.94	4.66	5.31	4.70	4.49
$\sigma^2 = (5, 1, ..., 1)$	7.05	4.46	7.16	5.82	5.46	5.27	5.69	5.20

Welshov test

Výrazné nedodržiavanie (veľké nadhodnocovanie) hladiny významnosti α nastáva iba v prípade, keď $n_i \ll I$. Ak je $n_i \geq I$, potom už tento test hladinu významnosti α dodržuje. Výhodou tohto testu by mohlo byť dobré dodržiavanie hladiny významnosti v heteroskedastických prípadoch pri dostatočnom počte pozorovaní pre každú populáciu.

Tabuľka 15: Dodržiavanie hladiny významnosti α=5% Welshovho testu vo vyváženom prípade s normálnymi populáciami pre rôzne rozptyly.

I	**I=3**		**I=5**		**I=10**			**I=25**		
rozptyl	n=5	n=50	n=5	n=50	n=5	n=10	n=50	n=5	n=25	n=50
1, ... ,1	4.51	4.98	5.55	4.77	8.21	5.80	4.94	15.13	5.26	5.00
1,...,1,5	4.70	5.13	5.31	4.97	8.55	5.95	4.86	14.51	5.27	4.86

Tabuľka 16: Dodržiavanie hladiny významnosti α=5% Welshovho testu v nevyváženom prípade s normálnymi populáciami pre rôzne rozptyly.

I	**I=3**				**I=25**			
n=(...)	5,5,10	10,10,5	5,8,10	10,15,20	24x30,35	24x35,30	11:35	26:50
$\sigma^2 = (1, ..., 1)$	4.72	5.26	4.84	5.20	5.34	5.29	5.77	5.42
$\sigma^2 = (1, ..., 1, 5)$	4.62	5.27	4.76	5.14	5.12	5.51	5.78	5.50
$\sigma^2 = (5, 1, ..., 1)$	5.33	4.51	5.41	5.12	5.11	5.47	5.44	4.98

Chenov test

Tento test, odvodený iba pre vyvážený prípad, dodržuje hladinu významnosti ako aj za homoskedasticity, tak aj za heteroskedasticity vo všetkých uvažovaných vyvážených prípadoch.

Tabuľka 17: Dodržiavanie hladiny významnosti α=5% Chenovho testu vo vyváženom prípade s normálnymi populáciami pre rôzne rozptyly.

I	**I=3**		**I=5**		**I=10**		**I=25**	
rozptyl	n=5	n=50	n=5	n=50	n=5	n=50	n=5	n=50
1, ... ,1	5.29	5.31	4.98	5.06	5.16	5.15	5.05	5.65
1,...,1,5	5.57	5.64	4.76	4.95	5.01	4.84	4.88	4.74

Kombi test

Tento test nie vždy dodrží hladinu významnosti α. Pre malé n_i a pre homoskedasticitu tento test výrazne kopíruje F-test a teda si v týchto prípadoch ponecháva aj jeho zlé vlastnosti. Pre dosť veľké n_i alebo veľkú heteroskedasticitu zase Bartlettov test často zamieta hypotézu o zhode rozptylov. To vedie k tomu, že Kombi test je pre takéto prípady skoro zhodný s Krutchkoffovým testom. Preto mi príde lepšie nepoužívať tento kombinovaný test, ale jednotlivé testy samotné (vieme aké poruchy môžeme očakávať).

Tabuľka 18: Dodržiavanie hladiny významnosti α=5% Kombi-testu vo vyváženom prípade s normálnymi populáciami pre rôzne rozptyly.

I	**I=3**		**I=5**		**I=25**	
rozptyl	n=5	n=50	n=5	n=50	n=5	n=50
1, ... ,1	4.92	5.16	5.51	4.80	6.18	4.88
1,...,1,5	7.41	4.78	8.75	5.03	9.19	5.13

Tabuľka 19: Dodržiavanie hladiny významnosti α=5% Kombi-testu v nevyváženom prípade s normálnymi populáciami pre rôzne rozptyly.

I	**I=3**				**I=25**			
n=(...)	5,5,10	10,10,5	5,8,10	10,15,20	24x30,35	24x35,30	11:35	26:50
$\sigma^2 = (1, ..., 1)$	5.52	5.33	5.63	5.07	4.85	5.11	5.15	5.07
$\sigma^2 = (1, ..., 1, 5)$	4.16	11.49	5.16	4.82	4.94	4.91	4.91	5.36
$\sigma^2 = (5, 1, ..., 1)$	9.63	5.14	10.96	6.92	5.19	5.35	7.43	5.38

5.2 Porovnanie síl testov

Už na prvý pohľad vidieť, že Chenov test má slabšiu silu ako všetky ostatné testy. Aj keď sa s rastúcim počtom pozorovaní v populáciách tento rozdiel zmenšuje, stále je zrejmý.

Akritasov vážený test dodržuje hladinu významnosti iba pre veľký počet pozorovaní a preto ho ďalej nebudem uvažovať.

Taktiež nebudem porovnávať silu Kombi testu s ostatnými testami. Budem uvažovať iba F-test a Krutchkoffov test samostatne. Sila Kombi-testu sa aj tak (až na nejaké výnimky) pohybuje medzi silou týchto dvoch testov.

5.2.1 Platí alternatíva A_1: $\mu_1 = ... = \mu_{I-1} = 0$ a $\mu_I > 0$

Vyvážené prípady

V homoskedastických prípadoch sú najsilnejšie F-test a Akritasov nevážený test. Pre malý počet populácií a malý počet pozorovaní v nich je Akritasov vážený

Tabuľka 20: Porovnanie sily Chenovho testu a F-testu v homoskedastických vyvážených prípadoch s normálnymi populáciami a porovnanie sily Chenovho testu a Krutchoffovho testu v heteroskedastických prípadoch s normálnymi populáciami za platnosti rôznych alternatív vo vyvážených prípadoch.

	Posledná populácia iná				**Lin. závislosť populácií**				**Pol na pol**	
I	**I=3**		**I=25**		**I=3**		**I=25**		**I=10**	
$\sigma^2 = (...)$	n=5	n=50	n=5	n=50	n=5	n=50	n=25	n=50	n=25	n=50
Chen. test										
(1,...,1)	0.238	0.726	0.053	0.672	0.297	0.767	0.493	0.634	0.606	0.669
(1,...,1,5)	0.127	0.223	0.061	0.183	0.130	0.238	0.169	0.182	0.191	0.191
(5,1,...,1)	0.141	0.219	0.056	0.180	0.137	0.243	0.151	0.175	0.186	0.182
F-test										
(1,...,1)	0.825	0.809	0.845	0.884	0.857	0.843	0.859	0.866	0.880	0.814
Krut. test										
(1,...,1,5)	0.286	0.290	0.102	0.194	0.465	0.482	0.765	0.806	0.803	0.787
(5,1,...,1)	0.682	0.713	0.442	0.853	0.477	0.508	0.764	0.804	0.822	0.772

test mierne silnejší ako F-test, čo je spôsobené tým, že v týchto prípadoch tento test mierne nadhodnocuje hladinu významnosti α. Pre $n \geq 10$ alebo $I > 10$ sa už zdá, že Akritasov vážený test dodržuje hladinu významnosti a má podobnú silu ako F-test. Avšak oproti F-testu lepšie dodržuje hladinu významnosti v heteroskedastických populáciách, preto by som ho pre tieto situácie uprednostnil. Ďalej uvažujme heteroskedastický prípad. F-test, Boxov test a Brownov-Forsytheov test používajú v tejto situácii rovnakú testovú štatistiku. Keby všetky tieto testy rovnako dodržiavali hladinu významnosti, potom by museli mať rovnakú kritickú hodnotu pre ich testovú štatistiku. Z tohto dôvodu sa ich sila líši len v tom, ako ktorý test nedodržiava hladinu α. Akritasov nevážený test sa správa podobne ako tieto tri testy, má podobnú silu vo všetkých prípadoch. Označme tieto štyri testy ako testy prvej skupiny. Druhú skupinu podobných testov tvoria Krutchkoffov test, Weerahandiho test a Welshov test. Tieto testy majú lepšiu silu ako testy prvej skupiny najmä v prípade, že porucha homoskedasticity nie je v poruchovej populácii (tj. tá, kde je $\alpha_i > 0$), v prípade lineárnej poruchy homoskedasticity má poruchová populácia najmenší rozptyl. Pre $n_i \ll I$ sa táto výhoda stráca, dokonca testy prvej skupiny majú lepšiu silu vo všetkých simulovaných prípadoch. Preto by som pre $I \geq 10$ a $n_i \ll I$ použil Brownov-Forsytheov test, ktorý ako tak dodržiava hladinu významnosti spolu s Akritasovým nevážným testom, ale má málinko lepšiu silu. Z testov druhej skupiny sa ako najlepší zdá byť Krutchkoffov test vďaka tomu, že najlepšie dodržuje hladinu významnosti medzi týmito testami. Za zmienku si je ešte dobré povšimnúť, že s rastúcim počtom pozorovaní sa Welshov test výrazne a rýchlo blíži ku Krutchkoffovmu testu. Pri výbere testov z prvej skupiny by som tiež ako hlavné kritérium zvolil dodržiavanie hladiny významnosti, pretože všetky testy majú veľmi podobnú silu v prípade, že sú ich odhadované hladiny významnosti blízke.

Tabuľka 21: Porovnanie sily F-testu, Akritasovho neváženého testu a Krutchkoffovho testu vo vyváženom homoskedastickom prípade s normálnymi populáciami za platnosti alternatívy $A_1 : \mu_1 = ... = \mu_{I-1} = 0$ a $\mu_I > 0$. Symbolom $\widehat{\alpha}\%$ označujeme odhad hladiny významnosti hypotézy H_0 v percentách.

I	**I=3**				**I=25**			
	$\widehat{\alpha}\%$	n=10	$\widehat{\alpha}\%$	n=50	$\widehat{\alpha}\%$	n=5	$\widehat{\alpha}\%$	n=50
F-test	5.0	0.811	5.1	0.806	5.1	0.819	5.1	0.881
Akr.nev.test	5.6	0.821	5.0	0.806	4.6	0.803	5.9	0.894
Krutch. test	4.9	0.784	4.9	0.800	5.3	0.420	5.1	0.854

Tabuľka 22: Porovnanie sily Brownovho-Forsytheovho testu a Krutchkoffovho testu vo vyváženom heteroskedastickom prípade s normálnymi populáciami za platnosti alternatívy $A_1 : \mu_1 = ... = \mu_{I-1} = 0$ a $\mu_I > 0$. Symbolom $\widehat{\alpha}\%$ označujeme odhad hladiny významnosti hypotézy H_0 v percentách.

I	**I=3**				**I=25**			
	$\widehat{\alpha}\%$	n=5	$\widehat{\alpha}\%$	n=25	$\widehat{\alpha}\%$	n=5	$\widehat{\alpha}\%$	n=25
$\sigma^2 = (1, ..., 1, 5)$								
B-F test	5.6	0.375	6.0	0.414	6.6	0.558	7.8	0.577
Krutch. test	5.6	0.295	4.9	0.297	4.8	0.115	5.4	0.167
$\sigma^2 = (5, 1, ..., 1)$								
B-F test	5.7	0.403	5.8	0.435	5.9	0.657	7.6	0.717
Krutch. test	5.8	0.679	4.7	0.729	4.6	0.418	5.2	0.749

Jedna populácia s odlišným počtom pozorovaní

V týchto prípadoch sa testy správajú podobne ako vo vyváženom prípade. Najväčší rozdiel je v povahe Weerahandiho testu, ktorý sa mierne líši od svojej skupiny. Ak je poruchová skupina (tj. tá, kde je $\alpha_i > 0$) zaradená tam, kde je viac pozorovaní, tak má tento test o niečo väčšiu silu ako ostatné testy z jeho skupiny. Naopak, ak je v poslednej populácii menší počet pozorovaní, potom má slabšiu silu ako testy z jeho skupiny. Preto by som ho použil vtedy, keď je lepšie použiť test druhej skupiny a navyše máme podozrenie, že populácia, ktorá sa od ostatných líši je tá, pre ktorú máme viac pozorovaní. S rastúcim počtom populácií sa táto výhoda Weerahandiho testu zmenšuje. Pre $I \geq 10$ už má tento test výrazný problém s dodržiavaním hladiny významnosti a navyše sa zdá, že táto výhoda Weerahandiho testu sa už úplne stráca. Preto by som pre $I \geq 10$ radšej použil iný test ako Weerahandiho test.

Tabuľka 23: Porovnanie sily Weerahandiho testu, Krutchkoffovho testu, Welshovho testu a Boxovho testu v nevyváženom heteroskedastickom prípade s normálnymi populáciami za platnosti alternatívy $A_1 : \mu_1 = ... = \mu_{I-1} = 0$ a $\mu_I > 0$. Symbolom $\widehat{\alpha}\%$ označujeme odhad hladiny významnosti hypotézy H_0 v percentách.

I	I=3				I=10			
$\widehat{\alpha}\%$ a n=(...)	$\widehat{\alpha}\%$	5,5,10	$\widehat{\alpha}\%$	10,10,5	$\widehat{\alpha}\%$	9x15,20	$\widehat{\alpha}\%$	9x20,15
$\sigma^2 = (1, ..., 1, 5)$								
Weerah. test	3.7	0.300	5.1	0.210	7.2	0.322	6.6	0.193
Krutch. test	4.1	0.270	6.3	0.280	4.9	0.256	5.0	0.169
Welshov test	4.7	0.289	5.9	0.261	5.4	0.267	5.2	0.174
Boxov test	4.0	0.389	7.6	0.356	4.7	0.589	5.5	0.370
$\sigma^2 = (5, 1, ..., 1)$								
Weerah. test	5.1	0.578	4.7	0.598	7.2	0.931	6.2	0.661
Krutch. test	6.0	0.590	4.9	0.690	4.9	0.910	4.4	0.652
Welshov test	5.6	0.555	5.1	0.690	5.3	0.916	4.8	0.663
Boxov test	6.8	0.347	4.8	0.390	5.3	0.720	5.0	0.392

Lineárna závislosť počtu pozorovaní

Pre n_i majúce lineárnu závislosť sa chovanie testov prakticky nemení od správania, keď sa líši len jedno n_i. Tiež úvaha odlišnosti Weerahandiho testu od Krutchkoffovho testu a Welshovho testu sa potvrdila. Navyše sa ešte ukázalo, že s rastúcim počtom pozorovaní sa tieto rozdiely medzi Weerahandiho testom a ostatnými dvomi testami zmenšujú. Preto je použitie testov rovnaké ako pre predchádzajúci simulovaný nevyvážený prípad.

Tabuľka 24: Porovnanie sily Weerahandiho testu, Krutchkoffovho testu a Boxovho testu v nevyváženom heteroskedastickom prípade s normálnymi populáciami za platnosti alternatívy $A_1 : \mu_1 = ... = \mu_{I-1} = 0$ a $\mu_I > 0$. Symbolom $\widehat{\alpha}\%$ označujeme odhad hladiny významnosti hypotézy H_0 v percentách.

$\widehat{\alpha}\%$ a n=(...)	$\widehat{\alpha}\%$	5,8,10	$\widehat{\alpha}\%$	10,15,20	$\widehat{\alpha}\%$	10,8,5	$\widehat{\alpha}\%$	20,15,10
$\sigma^2 = (1, ..., 1, 5)$								
Weerah. test	4.2	0.320	4.8	0.329	4.9	0.201	4.9	0.247
Krutch. test	4.4	0.306	5.0	0.320	6.4	0.269	5.6	0.281
Boxov test	4.5	0.408	4.9	0.417	7.2	0.346	6.4	0.356
$\sigma^2 = (5, 1, ..., 1)$								
Weerah. test	5.1	0.737	5.1	0.728	4.1	0.578	5.0	0.699
Krutch. test	6.4	0.728	5.6	0.719	4.4	0.676	5.0	0.729
Boxov test	7.6	0.405	6.1	0.353	4.9	0.395	4.8	0.443

Záver pre alternatívu A_1

Tento prístup sa dá použiť vtedy, keď máme podozrenie, že jedna populácia sa líši od ostatných. Ak sme si istí homoskedasticitou, tak najlepší test je F-test. V prípade podozrenia na heteroskedasticitu je výber testov zložitý. Ak si myslíme, že v líšiacej sa populácii je skôr väčší rozptyl a máme $n_i \geq 10\ \forall i$, je asi najvhodnejšie použiť Akritasov nevážený test, ak je pozorovaní málo, je lepšie uprednostniť Brownov-Forsytheov test, ktorý je tiež najlepší pre veľký počet populácií a malý počet pozorovaní pre všetky uvažované heteroskedastické prípady. Ak máme dostatok pozorovaní a podozrenie, že v líšiacej sa populácii je skôr menší rozptyl, tak je lepšie zvoliť iný test. V tejto situácii by som odporučil voliť jeden z trojice testov: Krutchkoffov test, Weerahandiho test a Welshov test. V simuláciách sa ukázalo, že Weerahandiho test je najlepšie použiť pre menej ako desať populácií v prípade, keď líšiaca sa populácia patrí medzi populácie s väčším počtom pozorovaní. Pre ostatné prípady by som radšej z týchto testov vybral Krutchkoffov test aj vďaka tomu, že najlepšie dodržuje hladinu významnosti medzi týmito testami. Rozdiel Weerahandiho testu sa s rastúcim počtom pozorovaní zmenšuje od ostatných dvoch testov. Navyše pre Weerahandiho test a tiež Krutchkoffov test sú k výpočtu kritickej hodnoty potrebné ďalšie simulácie. Preto pre veľký počet pozorovaní v každej populácii je lepšie pred týmito testami uprednostniť Welshov test.

5.2.2 Platí alternatíva A_2 : μ *lineárne*

Vyvážený prípad

Ak sme si istí homoskedasticitou populácií, tak najsilnejší test je F-test. Aj keď by sa mohlo zdať, že Akritasov nevážený test má o málo lepšiu silu ako F-test, je to spôsobené tým, že Akritasov nevážený test mierne nadhodnocuje hladinu významnosti okrem prípadu, keď máme veľa populácií alebo veľa pozorovaní v nich. Na druhú stranu výhodou Akritasovho neváženého testu oproti F-testu je, že lepšie dodržuje hladinu významnosti v heteroskedastických populáciách. V uvažovaných heteroskedastických prípadoch si je dobré uvedomiť, že normálne rozdelenie je symetrické okolo strednej hodnoty a tiež to, že naša alternatíva je tiež symetrická. Spolu s využitím toho, že všetky testy sú invariantné voči posunutiu, sú pre nás ekvivalentné nastavenia, keď je rozptyl $\boldsymbol{\sigma^2} = (1, ..., 1, 5)^T$ alebo $\boldsymbol{\sigma^2} = (5, 1, ..., 1)^T$ (pomocou posunutia a otočenia dostaneme rovnakú situáciu). Preto by som v prípade heteroskedasticity vyberal už iba medzi Welshovým testom, Weerahandiho testom, Krutchkoffovým testom a Brownovým-Forsytheovým testom. Brownov-Forsytheov test by som použil iba v prípade, že $I \geq 25$ a počet pozorovaní na populáciu je malý. Vtedy má najlepšiu silu medzi testami, ktoré ako tak dodržujú hladinu významnosti. V ostatných prípadoch by som testy použil podobne ako pre alternatívu A_1: $\mu_I > 0$. V ostatných prípadoch by som používal Krutchkoffov test. Tomuto testu sa s rastúcim počtom pozorovaní stáva Welshov test ekvivalentný (ako aj v dodržiavaní hladiny významnosti, tak aj v sile). Ale výhodou Welshovho testu je to, že nepotrebuje ďalšie simulácie náhodných veličín. Preto by som pre veľký počet pozorovaní radšej uprednostnil práve tento test.

Tabuľka 25: Sila F-testu, Akritasovho neváženého testu a Krutchkoffovho testu vo vyváženom homoskedastickom prípade s normálnymi populáciami za platnosti altenatívy $A_2 : \mu$ *lineárne*. Symbolom $\widehat{\alpha}\%$ označujeme odhad hladiny významnosti hypotézy H_0 v percentách.

I	I=5				I=10			
	$\widehat{\alpha}\%$	n=10	$\widehat{\alpha}\%$	n=50	$\widehat{\alpha}\%$	n=10	$\widehat{\alpha}\%$	n=50
F-test	4.7	0.808	5.1	0.873	4.9	0.836	4.9	0.841
Akr.nev.test	5.6	0.825	5.7	0.883	5.6	0.849	5.8	0.859
Krutch. test	4.7	0.765	5.2	0.871	4.9	0.761	5.0	0.835

Tabuľka 26: Porovnanie sily Krutchkoffovho testu a Brownovho-Forsytheovho testu vo vyváženom heteroskedastickom prípade s normálnymi populáciami za platnosti alternatívy $A_2 : \mu$ *lineárne*. Symbolom $\widehat{\alpha}\%$ označujeme odhad hladiny významnosti hypotézy H_0 v percentách.

I	I=3				I=25			
	$\widehat{\alpha}\%$	n=5	$\widehat{\alpha}\%$	n=10	$\widehat{\alpha}\%$	n=5	$\widehat{\alpha}\%$	n=10
$\sigma^2 = (1, ..., 1, 5)$								
Krutch. test	5.4	0.478	5.4	0.460	4.9	0.530	4.9	0.687
B-F test	5.4	0.406	6.2	0.399	6.2	0.712	6.9	0.704
$\sigma^2 = (1, 2, ..., I)$								
Krutch. test	5.4	0.539	5.1	0.510	5.3	0.091	5.1	0.106
B-F test	4.8	0.500	5.1	0.478	5.3	0.089	6.6	0.102

Jedna populácia s odlišným počtom pozorovaní

Uvažujme opäť najskôr homoskedastický prípad. Ak nám viac záleží na dodržaní hladiny významnosti α, tak by som odporučil F-test, ak nám až tak nezáleží na dodržaní hladiny významnosti, ale záleží nám viac na sile testu, tak by som odporučil Akritasov nevážený test, ktorý má lepšiu silu, ale menej dodržuje hladinu významnosti najmä pre $n_i < 10$ pre nejaké i. V heteroskedastickom prípade nedošlo k žiadnej prekvapivej zmene. Vzhľadom k symetrii alternatívy a rozdelenia populácií už nemá význam rozlišovať medzi tým, či je $\boldsymbol{\mu}$ lineárne rastúce alebo klesajúce. To má okrem iného tiež za následok , že nemá rozdiel rozlišovať medzi Krutchkoffovým testom a Weerahandiho testom podľa druhu heteroskedasticity. V tejto situácii je rozdiel medzi týmito testami hlavne v dodržiavaní hladiny významnosti. Tiež je malý rozdiel v tom, že Weerahandiho test sa trochu lepšie správa v prípade, že väčší rozptyl je v populácii s menším počtom pozorovaní. Ale napriek tomu by som medzi testami vyberal iba podľa počtu pozorovaní v populáciách rovnako ako vo vyváženom prípade a teda by som sa rozhodoval iba medzi Welshovým testom a Krutchkoffovým testom.

Tabuľka 27: Porovnanie sily F-testu a Akritasovho neváženého testu v nevyváženom homoskedastickom prípade s normálnymi populáciami za platnosti alternatívy $A_2 : \mu$ *lineárne*. Symbolom $\widehat{\alpha}\%$ označujeme odhad hladiny významnosti hypotézy H_0 v percentách.

I	**I=10**				**I=25**			
$\widehat{\alpha}\%$ a n=(...)	$\widehat{\alpha}\%$	9x15,20	$\widehat{\alpha}\%$	9x20,15	$\widehat{\alpha}\%$	24x30,35	$\widehat{\alpha}\%$	24x35,30
F-test	4.6	0.863	5.1	0.860	4.9	0.842	5.0	0.806
Akr.nev.test	5.4	0.873	5.8	0.875	5.6	0.855	5.6	0.820

Tabuľka 28: Porovnanie sily Weerahandiho testu, Krutchkoffovho testu, Welshovho testu a Boxovho testu v nevyváženom heteroskedastickom prípade s normálnymi populáciami za platnosti alternatívy $A_2 : \mu$ *lineárne*. Symbolom $\widehat{\alpha}\%$ označujeme odhad hladiny významnosti hypotézy H_0 v percentách.

I	**I=5**				**I=10**			
$\widehat{\alpha}\%$ a n=(...)	$\widehat{\alpha}\%$	4x5,10	$\widehat{\alpha}\%$	4x10,5	$\widehat{\alpha}\%$	9x15,20	$\widehat{\alpha}\%$	9x20,15
$\sigma^2 = (1, ..., 1, 5)$								
Weerah. test	5.6	0.555	6.3	0.707	7.1	0.750	6.4	0.781
Krutch. test	3.9	0.440	5.4	0.637	4.9	0.674	4.8	0.732
Welshov test	4.9	0.489	5.5	0.641	5.2	0.689	5.0	0.737
Boxov test	3.5	0.450	7.2	0.446	4.5	0.520	5.1	0.513
$\sigma^2 = (5, 1, ..., 1)$								
Weerah. test	7.4	0.640	6.5	0.612	7.3	0.778	6.5	0.768
Krutch. test	5.5	0.505	5.2	0.556	5.1	0.706	4.8	0.711
Welshov test	6.2	0.534	5.9	0.576	5.5	0.718	5.1	0.723
Boxov test	6.3	0.431	5.1	0.444	5.1	0.517	5.1	0.520

Lineárna závislosť počtu pozorovaní

Simulácie v tomto prípade potvrdili výsledky pre prípad, že jedno n_i sa líšilo od ostatných. Preto v homoskedastickom prípade je najlepší F-test. V heteroskedastickom prípade by som opäť vyberal podľa počtu pozorovaní v populáciách a to tak, že keď máme dostatok pozorovaní v každej populácii, tak by som použil Welshov test, ak je v nejakej populácii počet pozorovaní približne rovný počtu populácií, tak by som radšej použil Krutchkoffov test, ktorý lepšie dodržuje hladinu významnosti v takýchto prípadoch.

Tabuľka 29: Porovnanie sily F-testu a Akritasovho neváženého testu v nevyváženom homoskedastickom prípade s normálnymi populáciami za platnosti alternatívy $A_2 : \mu$ *lineárne*. Symbolom $\widehat{\alpha}\%$ označujeme odhad hladiny významnosti hypotézy H_0 v percentách.

$\widehat{\alpha}\%$ a n=(...)	$\widehat{\alpha}\%$	10,15,20	$\widehat{\alpha}\%$	2*(6:10)	$\widehat{\alpha}\%$	16:25	$\widehat{\alpha}\%$	11:35
F-test test	4.9	0.851	4.8	0.847	4.7	0.808	5.3	0.824
Akr.nev.test	5.0	0.852	5.5	0.858	5.4	0.823	5.9	0.833

Tabuľka 30: Porovnanie sily Weerahandiho testu, Krutchkoffovho testu a Boxovho testu v nevyváženom heteroskedastickom prípade s normálnymi populáciami za platnosti alternatívy $A_2 : \mu$ *lineárne*. Symbolom $\widehat{\alpha}\%$ označujeme odhad hladiny významnosti hypotézy H_0 v percentách.

$\widehat{\alpha}\%$ a n=(...)	$\widehat{\alpha}\%$	5,8,10	$\widehat{\alpha}\%$	10,8,5	$\widehat{\alpha}\%$	10,15,20	$\widehat{\alpha}\%$	20,15,10
$\sigma^2 = (1,1,5)$								
Weerah. test	4.2	0.421	4.6	0.474	4.5	0.473	5.0	0.498
Krutch. test	4.7	0.444	6.0	0.473	4.6	0.483	5.4	0.499
Boxov test	4.9	0.411	7.3	0.363	4.5	0.436	5.7	0.363
$\sigma^2 = (5,1,1)$								
Weerah. test	4.9	0.480	4.1	0.433	4.7	0.501	4.6	0.473
Krutch. test	6.3	0.488	4.4	0.461	5.5	0.506	4.8	0.485
Boxov test	7.4	0.380	4.5	0.418	5.7	0.369	4.9	0.438

Záver pre alternatívu A_2

Tento prípad simuluje, že očakávame lineárnu závislosť medzi strednými hodnotami populácií. V prípade, že sme si istí homoskedasticitou a chceme čo najviac dodržať hladinu významnosti, potom by som odporučil F-test. V prípade, že nám nevadia malé poruchy v nedodržaní hladiny významnosti, potom by som na homoskedastický prípad odporučil Akritasov nevážený test, ktorý má lepšiu silu a lepšie dodržuje hladinu významnosti v prípade heteroskedasticity. V heteroskedastickom prípade si je dobré uvedomiť, že sa jedná o symetrickú alternatívu a symetrické rozdelenie. Preto už nemá význam pre túto alternatívu v uvažovaných heteroskedastických prípadoch používať testy prvej skupiny. Jedinou výnimkou je prípad, keď máme k dispozícii veľký počet populácií a malý počet pozorovaní.

Vtedy majú testy druhej skupiny malú silu. Preto by som pre tento prípad odporučil Brownov-Forsytheov test. Testy druhej skupiny majú veľmi podobné sily v prípade, že dodržia hladinu významnosti. Preto sa musíme medzi týmito testami rozhodovať hlavne podľa toho, ako túto hladinu dodržujú. Najspoľahlivejšie ju dodržuje Krutchkoffov test, preto by som vo väčšine prípadov odporučil práve tento test. S rastúcim počtom pozorovaní sa rozdiely medzi jednotlivými testami druhej skupiny zmenšujú. Preto pre dosť veľký počet pozorovaní majú tieto testy takmer rovnakú silu. Avšak výhoda Welshovho testu oproti zvyšným dvom testom je v tom, že nepotrebuje dodatočné simulácie na určenie kritickej hodnoty. Nevýhodou tohto testu je to, že ak v niektorej populácii máme približne toľko pozorovaní ako je počet populácií, tak tento test môže mať problém s dodržaním hladiny významnosti. Preto by som Krutchkoffov test nahradil Welshovým testom iba v prípade dostatočného počtu pozorovaní v každej populácii.

5.2.3 Platí alternatíva $A_3 : \alpha_1 = ... = \alpha_{\lfloor I/2 \rfloor} = 0$ *a* $\alpha_{\lfloor I/2 \rfloor + 1} = ... = \alpha_I > 0$

Pripomeňme, že simulácie pre túto alternatívu boli uskutočnené len pre desať a dvadsaťpäť populácií. Preto testy, ktoré zle dodržujú hladinu významnosti v prípade veľkého počtu populácií (hlavne Weerahandiho test) budú pre túto alternatívu vynechané bez ďalšieho upozornenia.

Vyvážené prípady

Aj v tomto prípade má Akritasov nevážený test podobnú silu ako F-test. Taktiež by sa mohlo na prvý pohľad zdať, že Akritasov nevážený test je v niektorých prípadoch dokonca silnejší. Je to však spôsobené tým, že pre tieto prípady Akritasov nevážený test o málo nadhodnocuje hladinu významnosti. Pre heteroskedastické populácie by som vo väčšine prípadov použil buď Krutchkoffov test alebo Welshov test. Výhodou Krutchkoffovho testu oproti Welshovmu testu je to, že dodržuje hladinu významnosti aj pre malý počet pozorovaní na populáciu. Welshov test zase k určeniu kritickej hodnoty nepotrebuje dodatočné simulácie. Preto pre $n \geq I$ by som použil Welshov test, pre ostatné prípady Krutchkoffov test, okrem prípadu, keď je počet pozorovaní v populáciách výrazne menší ako počet populácií. Vtedy by som použil Brownov-Forsytheov test.

Jedna populácia s odlišným počtom pozorovaní

Ani v tejto uvažovanej situácii nedošlo k ničomu, čo by sa z predchádzajúcich simulácií nedalo očakávať. V homoskedastickom prípade majú opäť podobnú silu F-test a Akritasov nevážený test. V heteroskedastických populáciách sa opäť ako najlepší ukázal Krutchkoffov test, ktorý by som nahradil Welshovým testom v prípade, že máme minimálne toľko pozorovaní v každej populácii, ako je počet populácií. Dôvodom je takmer rovnaká sila, avšak Welshov test je výpočtovo jednoduchší.

Tabuľka 31: Porovnanie sily F-testu, Akritasovho neváženého testu a Krutchkoffovho testu vo vyváženom homoskedastickom prípade s normálnymi populáciami za platnosti alternatívy $A_3 : \alpha_1 = ... = \alpha_{\lfloor I/2 \rfloor} = 0$ *a* $\alpha_{\lfloor I/2 \rfloor + 1} = ... = \alpha_I > 0$. Symbolom $\widehat{\alpha}\%$ označujeme odhad hladiny významnosti hypotézy H_0 v percentách.

I	**I=10**				**I=25**			
	$\widehat{\alpha}\%$	n=10	$\widehat{\alpha}\%$	n=50	$\widehat{\alpha}\%$	n=5	$\widehat{\alpha}\%$	n=50
F-test	5.0	0.872	5.2	0.834	5.0	0.826	5.0	0.838
Akr.nev.test	5.7	0.882	5.9	0.851	4.5	0.810	6.0	0.853
Krutch. test	5.5	0.821	5.3	0.830	5.2	0.526	5.4	0.826

Tabuľka 32: Porovnanie sily Krutchkoffovho testu a Brownovho-Forsytheovho testu vo vyváženom heteroskedastickom prípade s normálnymi populáciami za platnosti alternatívy $A_3 : \alpha_1 = ... = \alpha_{\lfloor I/2 \rfloor} = 0$ *a* $\alpha_{\lfloor I/2 \rfloor + 1} = ... = \alpha_I > 0$. Symbolom $\widehat{\alpha}\%$ označujeme odhad hladiny významnosti hypotézy H_0 v percentách.

I	**I=10**				**I=25**			
	$\widehat{\alpha}\%$	n=5	$\widehat{\alpha}\%$	n=50	$\widehat{\alpha}\%$	n=5	$\widehat{\alpha}\%$	n=10
$\sigma^2 = (1, ..., 1, 5)$								
Krutch. test	5.0	0.520	4.9	0.769	5.1	0.507	4.9	0.806
B-F test	6.8	0.487	7.8	0.626	5.8	0.661	6.9	0.799
$\sigma^2 = (1, 2, ..., I)$								
Krutch. test	5.3	0.133	4.9	0.186	5.8	0.079	4.9	0.153
B-F test	4.9	0.133	6.2	0.186	5.6	0.080	6.0	0.162

Tabuľka 33: Porovnanie sily Krutchkoffovho testu, Welshovho testu a Akritasovho neváženého testu v nevyváženom heteroskedastickom prípade s normálnymi populáciami za platnosti alternatívy $A_3 : \alpha_1 = ... = \alpha_{\lfloor I/2 \rfloor} = 0$ *a* $\alpha_{\lfloor I/2 \rfloor + 1} = ... = \alpha_I > 0$. Symbolom $\widehat{\alpha}\%$ označujeme odhad hladiny významnosti hypotézy H_0 v percentách.

I	**I=10**				**I=25**			
$\widehat{\alpha}\%$ a n=(...)	$\widehat{\alpha}\%$	9x15,20	$\widehat{\alpha}\%$	9x20,15	$\widehat{\alpha}\%$	24x30,35	$\widehat{\alpha}\%$	24x35,30
$\sigma^2 = (1, ..., 1, 5)$								
Krutch. test	5.3	0.706	5.1	0.770	5.3	0.820	4.8	0.812
Welshov test	5.5	0.713	5.2	0.774	5.6	0.826	5.2	0.821
Akr.nev.test	6.2	0.526	6.7	0.559	6.3	0.712	6.0	0.700
$\sigma^2 = (5, 1, ..., 1)$								
Krutch. test	5.1	0.705	5.5	0.753	5.1	0.818	4.9	0.811
Welshov test	5.6	0.718	5.8	0.761	5.3	0.823	5.1	0.816
Akr.nev.test	6.2	0.523	5.6	0.560	6.2	0.705	5.9	0.699

Lineárna závislosť počtu pozorovaní

V homoskedastickom prípade sa zdá, že Akritasov nevážený test má málinko väčšiu silu ako F-test. F-test ale spoľahlivo dodržuje hladinu významnosti α, zatiaľ čo Akritasov nevážený test ju mierne nadhodnocuje. Ak máme v heteroskedastických prípadoch dostatok pozorovaní (aspoň $n_i \geq I \ \forall i$), potom by som uprednostnil Welshov test pred Krutchkoffovým testom. Dôvodom je jeho menšia výpočtová zložitosť. Pre menší počet pozorovaní by som ale už uprednostnil Krutchkoffov test, ktorý dodržuje hladinu významnosti aj pre malý počet pozorovaní. Ale ak je počet pozorovaní príliš malý, tak je lepšie použiť niektorý test prvej skupiny (buď Akritasov nevážený test alebo Brownov-Forsytheov test).

Tabuľka 34: Porovnanie sily F-testu a Akritasovho neváženého testu v nevyváženom homoskedastickom prípade s normálnymi populáciami za platnosti alternatívy $A_3 : \alpha_1 = ... = \alpha_{\lfloor I/2 \rfloor} = 0$ *a* $\alpha_{\lfloor I/2 \rfloor+1} = ... = \alpha_I > 0$. Symbolom $\widehat{\alpha}\%$ označujeme odhad hladiny významnosti hypotézy H_0 v percentách.

I	**I=10**				**I=25**			
$\widehat{\alpha}\%$ a n=(...)	$\widehat{\alpha}\%$	16:25	$\widehat{\alpha}\%$	2*(16:25)	$\widehat{\alpha}\%$	11:35	$\widehat{\alpha}\%$	26:50
F-test test	5.1	0.842	5.5	0.851	5.0	0.860	4.3	0.843
Akr.nev.test	5.8	0.856	6.4	0.868	5.7	0.869	4.9	0.856

Tabuľka 35: Porovnanie sily Krutchkoffovho testu a Akritasovho neváženého testu v nevyváženom heteroskedastickom prípade s normálnymi populáciami za platnosti alternatívy $A_3 : \alpha_1 = ... = \alpha_{\lfloor I/2 \rfloor} = 0$ *a* $\alpha_{\lfloor I/2 \rfloor+1} = ... = \alpha_I > 0$. Symbolom $\widehat{\alpha}\%$ označujeme odhad hladiny významnosti hypotézy H_0 v percentách.

I	**I=10**				**I=25**			
$\widehat{\alpha}\%$ a n=(...)	$\widehat{\alpha}\%$	16:25	$\widehat{\alpha}\%$	2*(16:25)	$\widehat{\alpha}\%$	11:35	$\widehat{\alpha}\%$	26:50
$\sigma^2 = (1, ..., 1, 5)$								
Krutch. test	4.5	0.785	4.9	0.795	5.7	0.805	5.3	0.815
Akr.nev.test	6.3	0.590	6.1	0.574	4.9	0.707	6.0	0.696
$\sigma^2 = (5, 1, ..., 1)$								
Krutch. test	4.9	0.773	5.2	0.805	5.0	0.796	4.9	0.802
Akr.nev.test	7.1	0.572	6.0	0.576	6.8	0.691	6.3	0.682

Záver pre alternatívu A_3

V homoskedastickom prípade sa ako najsilnejší test zdá byť Akritasov nevážený test. Je to však spôsobené tým, že mierne nadhodnocuje hladinu významnosti najmä pre malý počet pozorovaní. Preto ak si chceme byť istí dodržaním hladiny významnosti, tak by som odporučil F-test. V heteroskedastickom prípade už tento test nepatrí medzi najsilnejšie testy (okrem vyváženého prípadu s veľkým počtom populácií, keď $n \ll I$) a navyše už má problém s dodržiavaním hladiny významnosti . Preto je ho vhodné v heteroskedastickom prípade nahradiť buď Krutchkoffovým testom alebo Welshovým testom. Na rozdiel od Welshovho testu nemá Krutchkoffov test problém s dodržiavaním hladiny významnosti ani pre malý počet pozorovaní v populáciách. Pre $n_i \geq I \; \forall i$ má Welshov test takmer zhodnú silu s Krutchkoffovým testom. Jeho výhodou však je, že na určenie kritickej hodnoty nepotrebujeme ďalšie simulácie ako pre Krutchkoffov test. Preto by som v prípade, že máme dostatok pozorovaní, uprednostnil práve Welshov test. Ak je pozorovaní málo, ale nie až príliš, tak by som radšej použil Krutchkoffov test. Ak je $n_i \ll I$ tak by som asi použil Brownov-Forsytheov test, ktorý má pre tento prípad lepšiu silu ako Krutchkoffov test.

5.2.4 Záver na simulácie pre populácie z normálneho rozdelenia

Simulácie ukázali, že je užitočné rozlišovať medzi homoskedastickým a heteroskedastickým prípadom. V homoskedastickom prípade sa ako najlepší test zdá byť Akritasov nevážený test, ktorý má najlepšiu silu, ale zato má malý problém s dodržiavaním hladiny významnosti α. Preto v homoskedastických prípadoch, kde si chceme byť istí dodržaním hladiny významnosti α, by sme mali voliť F-test. V heteroskedastických prípadoch sa ukazuje ako najuniverzálnejší Krutchkoffov test. Tento test spoľahlivo dodržuje hladinu významnosti vo všetkých prípadoch a tiež patrí medzi testy s najlepšou silou medzi testami, ktoré dodržia hladinu významnosti. Prípad, keď nie je Krutchkoffov test najsilnejší, je pre $I > 10$ a $n_i \ll I$. V tomto prípade by som použil Brownov-Forsytheov test. U nesymetrickej alternatívy $A_1 : \mu_I > 0$ sa tiež ukázalo, že Krutchkoffov test nepatrí medzi testy s najlepšou silou v prípade, keď bol väčší rozptyl u odlišnej populácie. Pre malý počet pozorovaní v populáciách by som v takejto situácii použil Brownov-Forsytheov test. Ten však s rastúcim počtom pozorovaní začína nedodržovať hladinu významnosti. Dôvodom je to, že sa rýchlo približuje k F-testu, ktorý nedodržuje hladinu významnosti v prítomnosti heteroskedasticity. Preto by som ho pre viac pozorovaní nahradil Boxovým testom alebo Akritasovým neváženým testom. Nevýhodou Krutchkoffovho testu je, že pri výpočte kritických hodnôt musíme použiť simulácie. Preto pre $n_i \geq I \; \forall i$ môžeme ako vhodnú alternatívu namiesto tohto testu použiť Welshov test, ktorého kritické hodnoty sú známe a teda nemusíme použiť simulácie (čo nám môže ušetriť veľa času). Navyše pre veľký počet pozorovaní sú tieto dva testy takmer zhodné.

6 Výsledky simulácií pri poruche predpokladu normality populácií

6.1 Dodržiavanie hladiny významnosti

F-test

Pre populácie z rovnomerného rozdelenia má F-test podobné vlastnosti ako pre populácie z normálneho rozdelenia. Preto dodržuje hladinu významnosti hypotézy H_0 pre homoskedastické populácie. Pre homoskedastické dvojne exponenciálne rozdelené populácie tento test mierne podhodnocuje hladinu významnosti pre malý počet populácií a malý počet pozorovaní v nich. Pre homoskedastické prípady z logaritmicko-normálneho rozdelenia mierne podhodnocuje hladinu významnosti α pre všetky situácie s malým počtom populácií. Pre vyvážené heteroskedastické prípady tento test mierne nedodržuje hladinu významnosti a úplne zlyháva pre nevyvážené heteroskedastické prípady pre všetky tri rozdelenia.

Tabuľka 36: Dodržiavanie hladiny významnosti α=5% F-testu vo vyváženom prípade s populáciami z rovnomerného, dvojne exponenciálneho a logaritmicko-normálneho rozdelenia pre rôzne rozptyly.

	I=3		**I=25**	
	n=5	n=50	n=5	n=50
rovnomerné				
$\sigma^2 = (1, ..., 1)$	5.57	5.07	4.59	5.06
$\sigma^2 = (1, ..., 1, 5)$	8.06	6.47	8.16	7.44
dvojne exp.				
$\sigma^2 = (1, ..., 1)$	4.33	4.52	4.58	4.92
$\sigma^2 = (1, ..., 1, 5)$	5.82	6.32	6.82	7.90
log-normál.				
$\sigma^2 = (1, ..., 1)$	3.18	4.01	4.67	5.34
$\sigma^2 = (1, ..., 1, 5)$	9.60	7.96	5.88	7.18

Akritasov neválžený test

Spomeňme si, že tento test je odvodený za predpokladu, že počet populácií ide do nekonečna. Uvažujme opäť najskôr homoskedastický prípad. Pre populácie z rovnomerného rozdelenia vo väčšine prípadov mierne nadhodnocuje hladinu významnosti. Pre dvojne exponenciálne rozdelenie tento test dodržuje hladinu významnosti hypotézy H_0 okrem prípadu, keď je veľký počet populácií a malý počet pozorovaní. Vtedy Akritasov nevážený test mierne podhodnocuje hladinu významnosti α. Pre populácie z logaritmicko-normálneho rozdelenia tento test

Tabuľka 37: Dodržiavanie hladiny významnosti α=5% F-testu v nevyváženom prípade s populáciami z rovnomerného, dvojne exponenciálneho a logaritmicko-normálneho rozdelenia pre rôzne rozptyly.

	I=3		**I=25**	
	n=(5,8,10)	n=(10,15,20)	n=(11:35)	n=(26:50)
rovnomerné				
$\sigma^2 = (1, ..., 1)$	4.80	5.08	5.13	5.05
$\sigma^2 = (1, ..., 1, 5)$	3.89	3.06	4.67	5.76
$\sigma^2 = (5, 1, ..., 1)$	13.89	12.20	11.66	9.16
dvojne exp.				
$\sigma^2 = (1, ..., 1)$	4.30	4.89	5.36	5.16
$\sigma^2 = (1, ..., 1, 5)$	3.36	2.79	0.079	5.87
$\sigma^2 = (5, 1, ..., 1)$	11.33	11.82	11.28	9.64
log-normál.				
$\sigma^2 = (1, ..., 1)$	3.79	3.70	5.29	5.15
$\sigma^2 = (5, 1, ..., 1)$	13.79	13.41	9.8	8.53

nedodržuje hladinu významnosti pre žiadne nastavenie. Pre heteroskedastické populácie z rovnomerného a dvojne exponenciálneho rozdelenia tento test nedodržuje hladinu významnosti. Avšak v porovnaní s F-testom ju dodržuje lepšie.

Tabuľka 38: Dodržiavanie hladiny významnosti α=5% Akritasovho neváženého testu vo vyváženom prípade s populáciami z rovnomerného, dvojne exponenciálneho a logaritmicko-normálneho rozdelenia pre rôzne rozptyly.

	I=3			**I=25**		
	n=5	n=10	n=50	n=5	n=10	n=50
rovnomerné						
$\sigma^2 = (1, ..., 1)$	6.93	5.57	5.42	5.18	6.38	6.07
$\sigma^2 = (1, ..., 1, 5)$	8.96	6.62	4.11	7.41	6.66	5.94
dvojne exp.						
$\sigma^2 = (1, ..., 1)$	5.31	4.82	4.91	3.17	3.70	5.15
$\sigma^2 = (1, ..., 1, 5)$	6.09	4.83	3.92	4.32	4.88	5.53
log-normál.						
$\sigma^2 = (1, ..., 1)$	2.99	2.52	2.76	0.99	1.00	2.05
$\sigma^2 = (1, ..., 1, 5)$	10.74	8.64	6.28	1.43	1.65	3.78

Akritasov vážený test

Tak ako pre populácie z normálneho rozdelenia, tak aj pre populácie z rovnomerného, dvojne exponenciálneho a logaritmicko-normálneho rozdelenia potre-

Tabuľka 39: Dodržiavanie hladiny významnosti α=5% Akritasovho neváženého testu v nevyváženom prípade s populáciami z rovnomerného, dvojne exponenciálneho a logaritmicko-normálneho rozdelenia pre rôzne rozptyly.

	n=(10,10,5)	n=(10,5,5)	n=(10,8,5)	n=(20,15,10)
rovnomerné				
$\sigma^2 = (1, ..., 1)$	6.15	6.26	6.06	5.85
$\sigma^2 = (1, ..., 1, 5)$	10.25	10.20	10.22	7.52
dvojne exp.				
$\sigma^2 = (1, ..., 1)$	4.56	4.88	5.00	4.76
$\sigma^2 = (1, ..., 1, 5)$	7.02	6.81	6.63	5.54
log-normál.				
$\sigma^2 = (1, ..., 1)$	2.38	2.67	2.62	2.26
$\sigma^2 = (1, ..., 1, 5)$	12.01	10.91	11.83	11.18

buje Akritasov vážený test mnohokrát väčší počet pozorovaní v každej populácii, ako je počet populácií, k dodržaniu hladiny významnosti hypotézy H_0.

Tabuľka 40: Dodržiavanie hladiny významnosti α=5% Akritasovho váženého testu vo vyváženom prípade s populáciami z rovnomerného, dvojne exponenciálneho a logaritmicko-normálneho rozdelenia pre rôzne rozptyly.

	I=3			**I=5**		**I=10**	
	n=10	n=25	n=50	n=25	n=50	n=25	n=50
rovnomerné							
$\sigma^2 = (1, ..., 1)$	8.95	6.04	5.08	8.07	6.79	10.27	8.28
$\sigma^2 = (5, 1, ..., 1)$	8.88	6.13	5.81	8.35	6.51	10.05	7.62
dvojne exp.							
$\sigma^2 = (1, ..., 1)$	6.72	5.89	5.19	7.61	6.80	8.63	7.31
$\sigma^2 = (5, 1, ..., 1)$	6.97	5.82	5.63	7.27	6.64	8.23	7.61
log-normál.							
$\sigma^2 = (1, ..., 1)$	7.11	6.32	5.79	10.06	8.60	16.04	12.52
$\sigma^2 = (5, 1, ..., 1)$	12.18	8.82	7.56	11.50	9.69	18.17	13.95

Krutchkoffov test

Vlastnosti Krutchkoffovho testu sa výrazne zhoršili pri poruche predpokladu normality populácií. Aj keď poruchy v dodržiavaní hladiny významnosti vo vyvážených prípadoch sú už badateľné, pre nevyvážené prípady je to ešte oveľa horšie. Najhoršie je to v prípade, keď populácie pochádzajú z logaritmicko-normálneho rozdelenia. Keď máme málo populácií a málo pozorovaní v nich, tak tento test podhodnocuje hladinu významnosti za homoskedasticity a nadhodnocuje hladinu významnosti pri veľkej heteroskedasticite. S rastúcim počtom pozorovaní sa táto nestabilita zlepšuje. Avšak s rastúcim počtom populácií už tento test začína nadhodnocovať hladinu významnosti aj pre homoskedastické populácie a ak máme viac ako desať populácií, tak tento test už nedodržuje hladinu významnosti ani pre päťdesiat pozorovaní v každej populácii a preto by nemal byť v takýchto situáciách používaný. Pre rovnomerné a dvojne exponenciálne rozdelenie je situácia lepšia. Teraz už stačí k dodržaniu hladiny významnosti vo vyvážených prípadoch s rovnomerne rozdelenými populáciami, aby pre $I < 5$ bolo $n \geq 10$ a aby počty pozorovaní v každej populácii boli aspoň dvojnásobné ako počty populácií pre $I > 5$. V prípade nevyvážených dát alebo dvojne exponenciálne rozdelených populácií treba túto podmienku ešte zosilniť. V opačnom prípade pri rovnomernom rozdelení populácií Krutchkoffov test výrazne prekračuje hladinu významnosti, pre dvojne exponenciálne rozdelené populácie hladinu významnosti α podhodnocuje.

Tabuľka 41: Dodržiavanie hladiny významnosti α=5% Krutchkoffovho testu vo vyváženom prípade s populáciami z rovnomerného, dvojne exponenciálneho a logaritmicko-normálneho rozdelenia pre rôzne rozptyly.

	I=3		**I=5**		**I=25**	
	n=5	n=50	n=5	n=10	n=10	n=25
rovnomerné						
$\sigma^2 = (1, ..., 1)$	6.73	4.93	8.18	6.11	8.42	5.55
$\sigma^2 = (1, ..., 1, 5)$	7.09	4.90	8.64	5.92	8.79	5.03
dvojne exp.						
$\sigma^2 = (1, ..., 1)$	3.31	4.84	2.81	4.09	2.51	3.76
$\sigma^2 = (1, ..., 1, 5)$	4.09	4.81	2.94	3.66	2.17	3.75
log-normál.						
$\sigma^2 = (1, ..., 1)$	3.35	5.07	3.41	5.40	15.42	17.12
$\sigma^2 = (1, ..., 1, 5)$	10.70	6.92	8.26	9.09	17.17	18.08

Weerahandiho test

Podobne ako u normálneho rozdelenia populácií, tak aj pre tieto tri rôzne rozdelenia Weerahandiho test dodržuje hladinu významnosti iba pre n_i niekoľkonásobne väčšie ako je počet populácií. V rovnomernom rozdelení pre $I = 3$ tento test dodržuje hladinu významnosti už pre okolo desať pozorovaní v každej populácii. Avšak s rastúcim počtom populácií začína tento test výrazne nadhodnocovať hladinu významnosti α. Teraz uvažujme populácie z dvojne exponenciálneho rozdelenia. Ak je populácií menej ako päť a zároveň máme málo pozorovaní, potom tento test podhodnocuje hladinu významnosti. Pre päť populácií sa tento test správa dobre pre desať a viac pozorovaní v každej populácii. S rastúcim počtom

Tabuľka 42: Dodržiavanie hladiny významnosti α=5% Krutchkoffovho testu v nevyváženom prípade s populáciami z rovnomerného, dvojne exponenciálneho a logaritmicko-normálneho rozdelenia pre rôzne rozptyly.

	n=(5,8,10)	n=(6:10)	n=(16:25)	n=(11:35)
rovnomerné				
$\sigma^2 = (1, ..., 1)$	6.49	6.96	5.63	6.48
$\sigma^2 = (1, ..., 1, 5)$	5.66	6.34	5.48	6.04
$\sigma^2 = (5, 1, ..., 1)$	8.33	7.31	5.85	6.48
dvojne exp.				
$\sigma^2 = (1, ..., 1)$	3.56	3.66	3.85	3.34
$\sigma^2 = (1, ..., 1, 5)$	3.21	3.50	3.84	3.19
$\sigma^2 = (5, 1, ..., 1)$	4.49	3.56	3.90	3.89
log-normál.				
$\sigma^2 = (1, ..., 1)$	4.39	5.44	10.67	20.32
$\sigma^2 = (1, ..., 1, 5)$	6.86	7.34	10.68	20.73
$\sigma^2 = (5, 1, ..., 1)$	13.14	11.09	12.65	22.27

populácií už tento test nadhodnocuje hladinu významnosti, pokiaľ nemáme niekoľkonásobne viac pozorovaní v populáciách, ako je počet populácií. Najhoršie je to v prípade populácií z logaritmicko-normálneho rozdelenia, kde potrebujeme aby $n_i \gg I$, napríklad pre tri populácie sa už zdá, že päťdesiat pozorovaní je málo, inak tento test najmä v heteroskedastických prípadoch výrazne nadhodnocuje hladinu významnosti α.

Tabuľka 43: Dodržiavanie hladiny významnosti α=5% Weerahandiho testu vo vyváženom prípade s populáciami z rovnomerného, dvojne exponenciálneho a logaritmicko-normálneho rozdelenia pre rôzne rozptyly.

	I=3			**I=5**			**I=25**	
	n=5	n=10	n=50	n=5	n=10	n=50	n=25	n=50
rovnomerné								
$\sigma^2 = (1, ..., 1)$	5.13	4.90	5.48	10.19	6.52	5.26	9.32	6.49
$\sigma^2 = (1, ..., 1, 5)$	5.90	5.47	5.00	9.84	6.92	5.29	9.00	6.96
dvojne exp.								
$\sigma^2 = (1, ..., 1)$	2.78	3.42	4.80	4.15	4.59	4.77	7.43	6.00
$\sigma^2 = (1, ..., 1, 5)$	2.72	4.00	4.88	4.48	4.28	4.51	6.85	5.72
log-normál.								
$\sigma^2 = (1, ..., 1)$	2.79	3.59	5.00	6.43	7.34	6.91	23.41	16.76
$\sigma^2 = (1, ..., 1, 5)$	8.64	8.83	6.22	11.68	10.26	8.03	24.07	17.41

Tabuľka 44: Dodržiavanie hladiny významnosti α=5% Weerahandiho testu v nevyváženom prípade s populáciami z rovnomerného, dvojne exponenciálneho a logaritmicko-normálneho rozdelenia pre rôzne rozptyly.

	I=3		**I=25**	
	n=(10,10,5)	n=(10,15,20)	n=(24x35,30)	n=(26:50)
rovnomerné				
$\sigma^2 = (1, ..., 1)$	5.64	5.29	8.06	7.89
$\sigma^2 = (1, ..., 1, 5)$	6.38	5.08	7.07	7.54
$\sigma^2 = (5, 1, ..., 1)$	5.43	5.15	7.82	7.73
dvojne exp.				
$\sigma^2 = (1, ..., 1)$	3.23	4.03	6.47	6.16
$\sigma^2 = (1, ..., 1, 5)$	3.59	3.81	6.37	6.67
$\sigma^2 = (5, 1, ..., 1)$	3.47	4.20	6.75	6.40
log-normál.				
$\sigma^2 = (1, ..., 1)$	4.73	5.69	19.97	20.29
$\sigma^2 = (1, ..., 1, 5)$	11.44	6.91	20.58	20.97
$\sigma^2 = (5, 1, ..., 1)$	7.73	10.05	20.05	21.09

Brownov-Forsytheov test

Aj v tomto prípade môžeme Brownov-Forsytheov test brať ako vylepšenú verziu F-testu. Pre rovnomerne rozdelené populácie má Brownov-Forsytheov test mierne problémy akurát v prípadoch, kde F-test výrazne prekročí hladinu významnosti α. Horšie ako F-test sa správa iba v prípadoch, keď v populáciách s väčším počtom pozorovaní máme väčší rozptyl. Pre populácie z dvojne exponenciálneho rozdelenia tento test mierne podhodnocuje hladinu významnosti α v prípade, že máme málo pozorovaní v niektorej populácii a nadhodnocuje v prípade heteroskedasticity, keď máme veľa populácií a veľa pozorovaní v nich. Aj napriek tomu vo väčšine prípadov výrazne zlepšuje F-test (čo do dodržania hladiny významnosti). Pre populácie z logaritmicko-normálneho rozdelenia má tento test najhoršie vlastnosti. Pre malý počet populácií sa správa dosť nestabilne. Pre homoskedastické populácie výrazne podhodnocuje hladinu významnosti, pre populácie s veľkou heteroskedasticitou naopak nadhodnocuje hladinu významnosti hypotézy H_0. Túto nestabilitu zlepšuje zvyšujúci sa počet pozorovaní a tiež zvyšujúci sa počet populácií. Ďalšou vlastnosťou Brownovho-Forsytheovho testu je, že so zvyšujúcim počtom pozorovaní sa vo vyvážených prípadoch výrazne približuje F-testu (testová štatistika je rovnaká a kritické hodnoty sa rýchlo k sebe približujú). To vysvetľuje správanie sa Brownovho-Forsytheovho testu pre veľký počet pozorovaní vo vyvážených prípadoch.

Tabuľka 45: Dodržiavanie hladiny významnosti α=5% Brownovho-Forsytheovho testu vo vyváženom prípade s populáciami z rovnomerného, dvojne exponenciálneho a logaritmicko-normálneho rozdelenia pre rôzne rozptyly.

	I=3		**I=25**		
	n=5	n=50	n=5	n=25	n=50
rovnomerné					
$\sigma^2 = (1, ..., 1)$	4.75	5.45	4.34	4.85	5.32
$\sigma^2 = (1, ..., 1, 5)$	6.75	6.01	6.75	7.08	7.83
dvojne exp.					
$\sigma^2 = (1, ..., 1)$	3.08	4.83	3.23	4.74	4.86
$\sigma^2 = (1, ..., 1, 5)$	3.81	5.77	4.26	7.20	7.04
log-normál.					
$\sigma^2 = (1, ..., 1)$	1.64	3.83	1.31	4.10	4.64
$\sigma^2 = (1, ..., 1, 5)$	8.36	7.71	1.91	6.25	6.68

Tabuľka 46: Dodržiavanie hladiny významnosti α=5% Brownovho-Forsytheovho testu v nevyváženom prípade s populáciami z rovnomerného, dvojne exponenciálneho a logaritmicko-normálneho rozdelenia pre rôzne rozptyly.

	I=3		**I=25**	
	n=(5,8,10)	n=(10,15,20)	n=(11:35)	n=(26:50)
rovnomerné				
$\sigma^2 = (1, ..., 1)$	5.17	4.96	4.93	4.85
$\sigma^2 = (1, ..., 1, 5)$	6.26	5.91	7.15	7.78
$\sigma^2 = (5, 1, ..., 1)$	7.39	6.49	7.23	7.42
dvojne exp.				
$\sigma^2 = (1, ..., 1)$	3.94	4.19	4.72	4.80
$\sigma^2 = (1, ..., 1, 5)$	4.99	5.80	6.76	7.38
$\sigma^2 = (5, 1, ..., 1)$	4.29	5.25	6.36	6.93
log-normál.				
$\sigma^2 = (1, ..., 1)$	2.07	2.98	4.08	4.70
$\sigma^2 = (1, ..., 1, 5)$	7.50	7.30	5.88	7.14
$\sigma^2 = (5, 1, ..., 1)$	10.41	10.20	5.24	6.31

Boxov test

Tiež Boxov test sa správa ako vylepšená verzia F-testu. Pri rovnomernom rozdelení populácií má väčší problém s dodržiavaním hladiny významnosti iba v prípadoch, keď máme pre niektorú populáciu málo pozorovaní. Potom Boxov test podhodnocuje hladinu významnosti najmä pre homoskedastické a nadhodnocuje pre niektoré heteroskedastické populácie. Vo väčšine prípadov sa správa oveľa lepšie ako Brownov-Forsytheov test. Horší je akurát v niektorých nevyvážených prípadoch s malým počtom populácií a malým počtom pozorovaní, keď v populáciách s menším počtom pozorovaní máme väčší rozptyl. Preto by som ho za podmienky, že máme dostatok pozorovaní vždy uprednostnil pred Brownovým-Forsytheovým testom a teda aj F-testom. Pre populácie z dvojne exponenciálneho rozdelenia sa aj vlastnosti Boxovho testu zhoršili. Vo vyvážených prípadoch sa správa trochu lepšie ako v nevyvážených, ale aj tak potrebujeme väčší počet pozorovaní v každej populácii ako je počet populácií, inak tento test výrazne podhodnocuje hladinu významnosti hypotézy H_0. Najhoršie je to opäť s populáciami z logaritmicko-normálneho rozdelenia. V tomto prípade Boxov test za podmienky homoskedasticity výrazne podhodnocuje, v prípade heteroskedasticity a malého počtu populácií nadhodnocuje, v prípade heteroskedasticity a veľkého počtu populácií podhodnocuje hladinu významnosti α. Toto sa síce s rastúcim počtom pozorovaní v populáciách zlepšuje, ale tento test potrebuje obrovský počet pozorovaní, aby sme mohli povedať, že dodržuje hladinu významnosti.

Tabuľka 47: Dodržiavanie hladiny významnosti α=5% Boxovho testu vo vyváženom prípade s populáciami z rovnomerného, dvojne exponenciálneho a logaritmicko-normálneho rozdelenia pre rôzne rozptyly.

	I=3		**I=25**		
	n=5	n=50	n=5	n=25	n=50
rovnomerné					
$\sigma^2 = (1, ..., 1)$	4.17	5.06	3.23	4.76	5.01
$\sigma^2 = (1, ..., 1, 5)$	6.68	4.98	5.63	4.62	5.17
dvojne exp.					
$\sigma^2 = (1, ..., 1)$	3.04	4.61	1.55	3.93	3.94
$\sigma^2 = (1, ..., 1, 5)$	3.46	4.80	2.18	4.48	4.83
log-normál.					
$\sigma^2 = (1, ..., 1)$	1.29	3.25	0.26	1.04	1.37
$\sigma^2 = (1, ..., 1, 5)$	7.25	6.70	0.78	1.71	2.78

Welshov test

Pre populácie z rovnomerného a dvojne exponenciálneho rozdelenia sa Welshov test správa podobne. V obidvoch prípadoch potrebujeme väčší počet pozorovaní v každej populácii ako je počet populácií. Koľkonásobne má byť n_i väčšie ako počet populácií I sa nedá presne z mojich simulácií určiť. Rozdiel je avšak v povahe nedodržania hladiny významnosti. Pre rovnomerné rozdelenie a malý počet pozorovaní Welshov test výrazne prekračuje hladinu významnosti, pre dvojne exponenciálne rozdelenie a malý počet pozorovaní tento test výrazne podhodnocuje hladinu významnosti hypotézy H_0. Pre logaritmicko-normálne rozdelenie sa tento test správa podobne ako Krutchkoffov test. Pre malý počet populácií by sme ho ešte mohli použiť, aj keď pre malý počet pozorovaní mierne

Tabuľka 48: Dodržiavanie hladiny významnosti α=5% Boxovho testu v nevyváženom prípade s populáciami z rovnomerného, dvojne exponenciálneho a logaritmicko-normálneho rozdelenia pre rôzne rozptyly.

	I=3		**I=25**	
	n=(5,8,10)	n=(10,15,20)	n=(11:35)	n=(26:50)
rovnomerné				
$\sigma^2 = (1, ..., 1)$	5.50	4.78	4.86	4.63
$\sigma^2 = (1, ..., 1, 5)$	5.40	4.79	5.04	4.95
$\sigma^2 = (5, 1, ..., 1)$	9.52	6.33	5.79	5.25
dvojne exp.				
$\sigma^2 = (1, ..., 1)$	3.69	3.95	3.56	4.08
$\sigma^2 = (1, ..., 1, 5)$	3.36	4.27	4.25	4.57
$\sigma^2 = (5, 1, ..., 1)$	4.89	5.05	4.38	4.53
log-normál.				
$\sigma^2 = (1, ..., 1)$	2.06	2.29	0.88	1.26
$\sigma^2 = (1, ..., 1, 5)$	6.47	6.63	1.84	2.54
$\sigma^2 = (5, 1, ..., 1)$	10.58	9.70	1.27	2.12

podhodnocuje hladinu významnosti v homoskedastických populáciách a nadhodnocuje v heteroskedastických. To sa však s rastúcim počtom pozorovaní stabilizuje. Pre väčší počet populácií už výrazne nadhodnocuje hladinu významnosti aj v homoskedastických populáciách a veľkom počte pozorovaní. Preto by som ho pre desať a viac logaritmicko-normálnych populácií nepoužíval.

Tabuľka 49: Dodržiavanie hladiny významnosti α=5% Welshovho testu vo vyváženom prípade s populáciami z rovnomerného, dvojne exponenciálneho a logaritmicko-normálneho rozdelenia pre rôzne rozptyly.

	I=5			**I=25**		
	n=5	n=10	n=25	n=10	n=25	n=50
rovnomerné						
$\sigma^2 = (1, ..., 1)$	8.33	6.20	5.57	11.19	6.45	5.37
$\sigma^2 = (1, ..., 1, 5)$	9.17	6.19	5.39	11.02	5.92	5.12
dvojne exp.						
$\sigma^2 = (1, ..., 1)$	3.19	3.73	4.37	4.12	4.68	4.75
$\sigma^2 = (1, ..., 1, 5)$	3.54	3.75	4.73	3.91	4.55	4.68
log-normál.						
$\sigma^2 = (1, ..., 1)$	3.70	5.17	6.65	19.47	18.20	13.86
$\sigma^2 = (1, ..., 1, 5)$	9.17	6.19	5.39	21.02	18.85	14.72

Tabuľka 50: Dodržiavanie hladiny významnosti α=5% Welshovho testu v nevyváženom prípade s populáciami z rovnomerného, dvojne exponenciálneho a logaritmicko-normálneho rozdelenia pre rôzne rozptyly.

	I=5		**I=25**	
	n=(6:10)	n=2*(6:10)	n=(11:35)	n=(26:50)
rovnomerné				
$\sigma^2 = (1, ..., 1)$	6.82	6.02	7.05	5.61
$\sigma^2 = (1, ..., 1, 5)$	7.05	5.70	6.73	5.55
$\sigma^2 = (5, 1, ..., 1)$	7.10	5.68	6.85	5.68
dvojne exp.				
$\sigma^2 = (1, ..., 1)$	3.54	4.18	4.30	4.66
$\sigma^2 = (1, ..., 1, 5)$	3.04	4.11	4.32	4.64
$\sigma^2 = (5, 1, ..., 1)$	3.44	4.54	4.19	4.66
log-normál.				
$\sigma^2 = (1, ..., 1)$	5.36	7.29	21.61	16.66
$\sigma^2 = (1, ..., 1, 5)$	7.65	8.15	21.66	17.78
$\sigma^2 = (5, 1, ..., 1)$	10.16	10.61	24.14	17.76

Chenov test

Pre iné ako normálne rozdelenie populácií už má Chenov test problém s dodržiavaním hladiny významnosti hypotézy H_0. Pre populácie z rovnomerného rozdelenia sa to prejavuje tak, že s rastúcim počtom pozorovaní tento test začína podhodnocovať hladinu významnosti v heteroskedastických prípadoch. Pre populácie z dvojne exponenciálneho rozdelenia tento test s rastúcim počtom populácií začína prekračovať hladinu významnosti. Pre populácie z logaritmicko-normálneho rozdelenia nedodržuje hladinu významnosti v žiadnom simulovanom prípade aj napriek tomu, že s rastúcim počtom pozorovaní sa prekračovanie hladiny významnosti zmenšuje.

Kombi test

Myšlienkami k odvodeniu tohto testu boli dobré vlastnosti (čo do dodržania hladiny významnosti hypotézy H_0) F-testu za predpokladu homoskedasticity populácií a Krutchkoffovho testu pri poruche homoskedasticity za podmienky, že populácie pochádzajú z normálneho rozdelenia. Avšak pri poruche normality Krutchkoffov test výrazne stráca svoje dobré vlastnosti. Preto Kombi test nemá význam používať v takých heteroskedastických situáciách, kde Krutchkoffov test nedodržuje hladinu významnosti α. To nás podstatne obmedzuje v použití Kombi testu. Druhou nevýhodou je to, že ak jeden z dvojice používaných testov nedodrží hladinu významnosti, Bartlettov test na homoskedasticitu nám nezaručuje, že vyberie ten správny test. To vedie k nedodržaniu hladiny významnosti vo väčšine prípadov aj tam, kde jeden z testov dodrží hladinu významnosti a druhý nie.

Tabuľka 51: Dodržiavanie hladiny významnosti α=5% Chenovho testu vo vyváženom prípade s populáciami z rovnomerného, dvojne exponenciálneho a logaritmicko-normálneho rozdelenia pre rôzne rozptyly.

	I=3		**I=5**		**I=10**		**I=25**	
	n=5	n=50	n=5	n=50	n=5	n=50	n=5	n=50
rovnomerné								
$\sigma^2 = (1, ..., 1)$	6.80	4.88	6.18	4.87	6.51	4.58	7.07	4.17
$\sigma^2 = (1, ..., 1, 5)$	6.54	4.06	6.47	3.38	6.34	3.02	6.54	2.39
dvojne exp.								
$\sigma^2 = (1, ..., 1)$	6.15	5.58	7.57	6.06	8.73	6.71	11.00	9.30
$\sigma^2 = (1, ..., 1, 5)$	6.62	5.97	7.23	7.11	8.83	9.49	10.26	11.41
log-normál.								
$\sigma^2 = (1, ..., 1)$	14.43	8.37	18.76	12.05	27.26	17.81	42.33	27.17
$\sigma^2 = (1, ..., 1, 5)$	15.51	9.70	19.30	11.99	26.36	17.45	42.10	27.41

Preto podobne ako za predpokladu normality populácií by som radšej používal F-test a Krutchkoffov test samostatne a nie ako Kombi test.

Tabuľka 52: Dodržiavanie hladiny významnosti α=5% Kombi testu vo vyváženom aj nevyváženom prípade s populáciami z rovnomerného, dvojne exponenciálneho a logaritmicko-normálneho rozdelenia pre rôzne rozptyly.

	I=3		**I=5**	
	n=5	n=(10,10,5)	n=10	n=(4x10,5)
rovnomerné				
$\sigma^2 = (1, ..., 1)$	5.87	5.10	5.21	5.01
$\sigma^2 = (1, ..., 1, 5)$	9.31	13.56	8.78	14.92
dvojne exp.				
$\sigma^2 = (1, ..., 1)$	3.95	4.66	4.60	4.89
$\sigma^2 = (1, ..., 1, 5)$	5.53	7.39	4.68	7.33
log-normál.				
$\sigma^2 = (1, ..., 1)$	3.36	4.26	5.48	5.48
$\sigma^2 = (1, ..., 1, 5)$	11.50	13.81	9.75	13.16

Porovnanie síl testov

Pretože sa jednotlivé testy správajú rôzne pre rôzne rozdelenia populácií, budeme každé rozdelenie uvažovať samostatne. Pre každé rozdelenie budeme rozlišovať medzi vyváženým a nevyváženým prípadom. Pre každý prípad budeme porovnávať silu testov, ktoré aspoň približne dodržali v danej situácii hladinu významnosti hypotézy H_0.

6.2 Porovnanie síl testov pre rovnomerne rozdelené populácie

6.2.1 Platí alternatíva $A_1 : \mu_1 = ... = \mu_{I-1} = 0 \; a \; \mu_I > 0$

Vyvážené prípady

Za podmienky normality populácií sme sa rozhodovali medzi F-testom a Akritasovým neváženým testom. Podobne je to aj v prípade, že populácie pochádzajú z rovnomerného rozdelenia. Tiež pre homoskedastické populácie z rovnomerného rozdelenia by sme sa mali rozhodovať medzi týmito dvomi testami. Aj v tomto prípade sa zdá, že Akritasov nevážený test je o kúsok silnejší ako F-test, ale tiež mierne nadhodnocuje hladinu významnosti α. V hetoroskedastickej situácii sú iba dva testy, ktoré ako tak dodržujú hladinu významnosti α v prípade, že $n < I$. Ja by som uprednostnil Brownov-Forsytheov test pred F-testom, ktorý v tomto prípade spoľahlivejšie dodržuje hladinu významnosti hypotézy H_0. Tiež by sme mohli použiť Boxov test. Ten pri veľkej heteroskedasticite lepšie dodržuje hladinu významnosti α ako Brownov-Forsytheov test, ale za homoskedasticity ju podhodnocuje. Pre $n \geq I$ už viacej testov takmer dodržuje hladinu významnosti. Použitie testov je podobné ako pri normálne rozdelených populáciách. Opäť môžeme použiť rozdelenie do dvoch skupín rovnako ako za predpokladu normality. Prvú skupinu teda tvoria testy: F-test, Akritasov nevážený test, Boxov test a Brownov-Forsytheov test. Do druhej skupiny patria: Krutchkoffov test, Weerahandiho test a Welshov test. Ak je väčší rozptyl v populácii, ktorá sa od ostatných líši, potom majú lepšiu silu testy prvej skupiny, ak je menší rozptyl v líšiacej sa populácii, sú silnejšie testy druhej skupiny. Rozdiely v silách jednotlivých testov sú ovplyvnené hlavne dodržaním hladiny významnosti, preto by sme mali testy vyberať s ohľadom na tento fakt. Ďalšie kritérium na výber testov môže byť tiež výpočtová zložitosť, ktorá sa hlavne prejaví pri testoch druhej skupiny.

Tabuľka 53: Porovnanie sily F-testu a Akritasovho neváženého testu vo vyváženom homoskedastickom prípade s rovnomerne rozdelenými populáciami za platnosti alternatívy $A_1 : \mu_1 = ... = \mu_{I-1} = 0 \; a \; \mu_I > 0$. Symbolom $\widehat{\alpha}\%$ označujeme odhad hladiny významnosti hypotézy H_0 v percentách.

I	**I=3**				**I=25**			
	$\widehat{\alpha}\%$	n=10	$\widehat{\alpha}\%$	n=50	$\widehat{\alpha}\%$	n=10	$\widehat{\alpha}\%$	n=50
F-test	5.4	0.821	5.1	0.808	5.1	0.816	4.9	0.882
Akr.nev.test	6.0	0.838	5.2	0.809	5.8	0.834	5.8	0.895

Tabuľka 54: Porovnanie sily Akritasovho neváženého testu, Krutchkoffovho testu, Boxovho testu, Weerahandiho testu a Welshovho testu vo vyváženom heteroskedastickom prípade s rovnomerne rozdelenými populáciami za platnosti alternatívy $A_1 : \mu_1 = ... = \mu_{I-1} = 0$ *a* $\mu_I > 0$. Symbolom $\widehat{\alpha}\%$ označujeme odhad hladiny významnosti hypotézy H_0 v percentách.

I	**I=3**				**I=10**			
	$\widehat{\alpha}\%$	n=10	$\widehat{\alpha}\%$	n=50	$\widehat{\alpha}\%$	n=25	$\widehat{\alpha}\%$	n=50
$\sigma^2 = (1, ..., 1, 5)$								
Akr.nev.test	6.3	0.359	4.4	0.338	6.7	0.459	5.9	0.446
Boxov test	5.7	0.339	5.0	0.358	5.6	0.425	5.0	0.418
Krutch. test	6.1	0.258	5.4	0.282	5.1	0.186	5.1	0.186
Weerah. test	5.5	0.234	5.3	0.277	6.3	0.216	5.8	0.199
Welshov test	6.0	0.253	5.3	0.280	5.3	0.189	5.2	0.186
$\sigma^2 = (5, 1, ..., 1)$								
Akr.nev.test	6.5	0.379	4.2	0.332	6.3	0.534	6.2	0.535
Boxov test	5.9	0.349	4.7	0.363	5.2	0.474	5.0	0.487
Krutch. test	6.0	0.688	5.3	0.721	5.2	0.746	4.9	0.773
Weerah. test	5.4	0.667	5.0	0.717	6.6	0.764	5.7	0.782
Welshov test	5.8	0.683	5.1	0.719	5.4	0.753	5.1	0.777

Nevyvážené prípady

V homoskedastickom prípade sa opäť ako najlepší test zdá byť F-test. Tento test najspoľahlivejšie dodržuje hladinu významnosti a má tiež dobrú silu. Akritasov nevážený test má trošku lepšiu silu ako F-test, ale zato menej spoľahlivo dodržuje hladinu významnosti hypotézy H_0. V nevyváženom heteroskedastickom prípade môžeme rovnako ako vo vyváženom prípade rozdeliť testy do dvoch skupín podobných testov. Uvažujme iba prípad, že $n_i \geq I \; \forall i$. V opačnom prípade má každý test veľký problém dodržať hladinu významnosti a preto použitie niektorého testu nesie veľké riziko, že budeme testovať hypotézu H_0 na inej hladine významnosti ako je α. Takže predpokladajme, že máme v každej populácii aspoň toľko pozorovaní, koľko je počet populácií. Taktiež v tomto prípade platí, že testy prvej skupiny majú lepšiu silu v prípade, že väčší rozptyl je v populácii, ktorá sa od ostatných líši. Naopak testy druhej skupiny majú lepšiu silu v prípade, že je skôr menší rozptyl v líšiacej sa populácii. Ak situácia nasvedčuje, že by sme mali použiť niektorý test z prvej skupiny, tak by som vo väčšine prípadov použil Boxov test. Tento test najlepšie dodržuje hladinu významnosti z testov svojej skupiny. Výnimku tvoria akurát prípady, keď je $I \leq 5$ a $n_i \leq 20$ a väčší rozptyl je v skupinách s menším počtom pozorovaní. Pre tieto prípady Brownov-Forsytheov test lepšie dodržuje hladinu významnosti ako Boxov test. Preto by som ho pre tieto prípady uprednostnil pred Boxovým testom. V prípadoch, kde sú testy druhej skupiny silnejšie ako testy prvej skupiny, by mal byť výber testu ovplyvnený hlavne počtom populácií, ale tiež počtom pozorovaní. Pre menej ako desať populácií by som použil buď Krutchkoffov test alebo Weerahandiho test. Simulácie ukazujú, že Weerahandiho test má troška väčšiu silu ako Krutchkoffov test v prípade, že pre líšiacu sa populáciu máme viac pozorovaní. Ak máme v líšiacej sa populácii menej pozorovaní, tak má Weerahandiho test menšiu silu ako Krutchkoffov test. Pre viac populácií by som použil Welshov test v prípade, že máme viac pozorovaní v populáciách ako je počet populácií, pre počet pozorovaní v populáciách približne rovnému počtu populácií by som radšej uprednostnil Krutchkoffov test, pretože v tomto prípade lepšie dodržuje hladinu významnosti α. Welshov test by som taktiež použil v prípade malého počtu populácií a veľkého počtu pozorovaní, pretože k určeniu kritickej hodnoty nepotrebuje dodatočné simulácie ako Krutchkoffov test a Weerahandiho test. Sily týchto troch testov sú v prípade veľkého počtu pozorovaní aj tak takmer zhodné.

Tabuľka 55: Porovnanie sily Brownovho-Forsytheovho testu, Boxovho testu, Krutchkoffovho testu, Weerahandiho testu a Welshovho testu v nevyváženom heteroskedastickom prípade s rovnomerne rozdelenými populáciami za platnosti alternatívy A_1 :$\mu_1 = ... = \mu_{I-1} = 0$ *a* $\mu_I > 0$. Symbolom $\widehat{\alpha}\%$ označujeme odhad hladiny významnosti hypotézy H_0 v percentách.

I	**I=3**				**I=5**		**I=25**	
$\widehat{\alpha}\%$ a n=(...)	$\widehat{\alpha}\%$	5,8,10	$\widehat{\alpha}\%$	20,15,10	$\widehat{\alpha}\%$	2*(6:10)	$\widehat{\alpha}\%$	26:50
$\sigma^2 = (1, ..., 1, 5)$								
B-F test	5.8	0.434	6.3	0.326	7.1	0.491	7.2	0.572
Boxov test	4.8	0.385	6.2	0.316	5.3	0.423	4.8	0.513
Krutch. test	5.3	0.276	5.7	0.239	5.5	0.247	5.3	0.160
Weerah. test	5.0	0.288	5.3	0.216	6.0	0.266	7.4	0.201
Welshov test	5.7	0.285	5.5	0.232	5.6	0.249	5.4	0.163
$\sigma^2 = (5, 1, ..., 1)$								
B-F test	8.2	0.304	5.8	0.489	7.5	0.519	7.2	0.707
Boxov test	8.9	0.343	4.7	0.426	6.1	0.434	5.2	0.613
Krutch. test	7.9	0.703	5.2	0.719	5.8	0.804	4.9	0.768
Weerah. test	6.4	0.712	5.0	0.681	6.3	0.827	7.2	0.804
Welshov test	7.2	0.662	5.4	0.724	5.9	0.806	5.2	0.773

6.2.2 Platia alternatívy A_2: μ *lineárne* a $A_3 : \alpha_1 = ... = \alpha_{\lfloor I/2 \rfloor} = 0$ *a* $\alpha_{\lfloor I/2 \rfloor + 1} = ... = \alpha_I > 0$

Vyvážené prípady

Pretože všetky testy sú invariantné voči posunutiu a volili sme rovnomerné rozdelenie populácií, ktoré je symetrické okolo strednej hodnoty, preto alternatívu A_2:μ *lineárne rastúce* môžeme jednoduchou transformáciou zmeniť na alternatívu

A_2':μ *lineárne klesajúce*. Z rovnakého dôvodu nemá tiež zmysel rozlišovať pre alternatívu $A_3 : \alpha_1 = ... = \alpha_{\lfloor I/2 \rfloor} = 0$ *a* $\alpha_{\lfloor I/2 \rfloor+1} = ... = \alpha_I > 0$, ktoré populácie majú väčšiu alebo menšiu strednú hodnotu. Preto pre tieto prípady nemá zmysel uvažovať, keď je s väčšou strednou hodnotou μ_i väčší rozptyl. Z toho dôvodu nemá pre nás zmysel rozlišovať medzi použitím testov prvej a druhej skupiny podľa toho, kde je väčší alebo menší rozptyl ako v prípade nesymetrickej alternatívy. To nám do istej miery uľahčuje výber testov. V homoskedastickom prípade je opäť najlepšie zvoliť F-test. V heteroskedastických prípadoch sa musíme rozhodovať hlavne podľa počtu pozorovaní a počtu populácií. Pre $n < I$ majú testy druhej skupiny slabšiu silu ako testy prvej skupiny. Druhý dôvod prečo nepoužívať testy druhej skupiny pri malom počte pozorovaní je to, že majú problém s dodržiavaním hladiny významnosti hypotézy H_0. Pretože pre malý počet pozorovaní vyberáme testy z prvej skupiny, ktoré majú podobnú silu v prípade, že dodržia hladinu významnosti α, budeme pri výbere konkrétneho testu zohľadňovať hlavne to, ako tieto testy túto hladinu dodržiavajú. Preto pre malú heteroskedasticitu by som vybral Brownov-Forsytheov test, ktorý lepšie dodržuje hladinu významnosti ako Boxov test, ktorý túto hladinu mierne podhodnocuje. Pre veľkú hetoroskedasticitu sa naopak zdá, že Boxov test lepšie dodržuje hladinu významnosti, naopak Brownov-Forsyteov test mierne nadhodnocuje α. S rastúcim počtom pozorovaní sa spolu s dodržiavaním hladiny významnosti zlepšuje aj sila testov druhej skupiny v porovnaní s testami prvej skupiny. Ako zlom sa zdá byť počet pozorovaní v každej populácii rovný počtu populácií. Pre tieto prípady už Krutchkoffov test začína dodržiavať hladinu významnosti a začína byť silnejší ako testy prvej skupiny. Ak je počet pozorovaní väčší ako počet populácií, tak už aj Welshov test takmer dodržuje hladinu významnosti a Weerahandiho test tiež takmer dodržuje hladinu významnosti pre malý počet populácií. Sila týchto testov je v prípade dodržania hladiny významnosti podobná. Preto by som pre počet pozorovaní blízky počtu populácií použil Krutchkoffov test. Ak máme dostatok pozorovaní v každej populácii, tak by som uprednostnil Welshov test vďaka jeho výpočtovej jednoduchosti v porovnaní s Krutchkoffovým testom a Weerahandiho testom.

Nevyvážené prípady

V homoskedastických prípadoch je odhadnutá hladina významnosti hypotézy H_0 väčšia u Akritasovho neváženého testu ako u F-testu. Preto je Akritasov nevážený test silnejší ako F-test hlavne na úkor tejto dodržanej hladiny významnosti. Ďalej v homoskedastických prípadoch majú testy prvej skupiny väčšiu silu ako testy druhej skupiny. To sa naopak mení, keď sa objavuje heteroskedasticita. Testy druhej skupiny už majú väčšiu silu. Ak máme menej ako päť populácií, tak odhadovaná hladina významnosti Weerahandiho testu je nižšia ako pri Krutchkoffovom teste a Welshovom teste, čo má za následok, že tieto dva testy sú vo väčšine prípadov silnejšie ako Weerahandiho test. Výnimku tvoria akurát

Tabuľka 56: Porovnanie sily Akritasovho neváženého testu, Brownovho-Forsytheovho testu, Boxovho testu, Krutchkoffovho testu, Weerahandiho testu a Welshovho testu v nevyváženom prípade s rovnomerne rozdelenými populáciami za platnosti alternatív A_2: μ_i *lineárne* a A_3 : $\alpha_1 = ... = \alpha_{\lfloor I/2 \rfloor} = 0$ *a* $\alpha_{\lfloor I/2 \rfloor+1} = ... = \alpha_I = 0$ a s rozptylom $\sigma^2 = (1, ..., 1, 5)$. Symbolom $\widehat{\alpha}\%$ označujeme odhad hladiny významnosti hypotézy H_0 v percentách.

Alternatíva	A_2 :μ_i **lineárne**								A_3 : **pol na pol**	
I	**I=3**				**I=10**				**I=10**	
	$\widehat{\alpha}\%$	n=5	$\widehat{\alpha}\%$	n=50	$\widehat{\alpha}\%$	n=5	$\widehat{\alpha}\%$	n=25	n=5	n=25
Akr.nev.test	9.0	0.420	4.1	0.357	8.6	0.542	6.4	0.579	0.506	0.608
B-F test	6.9	0.344	6.0	0.427	7.4	0.501	7.9	0.639	0.469	0.676
Boxov test	6.5	0.321	4.8	0.383	6.2	0.417	5.3	0.525	0.386	0.553
Krutch. test	7.3	0.406	4.9	0.479	9.8	0.423	5.7	0.702	0.440	0.785
Weerah. test	6.0	0.368	4.7	0.474	19.3	0.704	6.9	0.747	0.728	0.821
Welshov test	6.7	0.388	4.8	0.476	14.1	0.536	5.8	0.708	0.557	0.789

prípady, keď väčší rozptyl je u populácie, kde máme menej pozorovaní. Potom je Weerahandiho test silnejší ako Krutchkoffov test aj napriek tomu, že jeho odhadnutá hladina významnosti je menšia ako pri Krutchkoffovom teste. Pre aspoň päť populácií už odhadovaná hladina významnosti Weerahandiho testu prevyšuje odhadovanú hladinu významnosti ostatných dvoch testov druhej skupiny. Preto je pre tieto prípady Weerahandiho test najsilnejší. Aby ale tento test dodržal hladinu významnosti pre aspoň desať populácií, potrebuje podstatne väčší počet pozorovaní ako Krutchkoffov test alebo Welshov test. Ak ešte zoberieme do úvahy výpočtovú zložitosť, tak ako najlepší test z druhej skupiny by som pre aspoň desať populácií a dostatok pozorovaní odporučil Welshov test, pre počet pozorovaní približne rovný počtu populácií zase Krutchkoffov test, ktorý má menší problém s dodržaním hladiny významnosti ako Welshov test.

6.2.3 Záver na simulácie pre populácie z rovnomerného rozdelenia

Aj pre rovnomerne rozdelené homoskedastické populácie by sme sa mali rozhodovať medzi Akritasovým neváženým testom a F-testom. Akritasov nevážený test má síce málinko väčšiu silu ako F-test, ale má tiež aj väčšiu odhadovanú hladinu významnosti ako F-test. V heteroskedastických prípadoch je opäť užitočné rozdeliť testy do dvoch skupín podobných testov. Pre malý počet pozorovaní majú všetky testy problém s dodržiavaním hladiny významnosti. Ale ak by som musel na takúto situáciu použiť nejaký test, vybral by som ho prvej skupiny, pretože tieto testy majú pre menší počet pozorovaní, ako je počet populácií väčšiu silu ako testy druhej skupiny. Ak je heteroskedasticita malá, tak by som použil Brownov-Forsytheov test, ak je heteroskedasticita veľká, použil by som Boxov test. Ak máme počet pozorovaní v populáciách aspoň taký, ako je počet populácií, tak vo

Tabuľka 57: Porovnanie sily Boxovho testu, Krutchkoffovho testu a Weerahandiho testu v nevyváženom prípade s rovnomerne rozdelenými populáciami za platnosti alternatívy A_2: μ_i *lineárne*. Symbolom $\widehat{\alpha}\%$ označujeme odhad hladiny významnosti hypotézy H_0 v percentách.

I	I=3				I=5		I=25	
$\widehat{\alpha}\%$ a n=(...)	$\widehat{\alpha}\%$	5,8,10	$\widehat{\alpha}\%$	10,15,20	$\widehat{\alpha}\%$	2*(6:10)	$\widehat{\alpha}\%$	26:50
$\sigma^2 = (1, ..., 1)$								
Boxov test	5.3	0.830	5.2	0.851	4.9	0.836	4.7	0.838
Krutch. test	6.3	0.769	5.4	0.823	5.4	0.787	5.4	0.807
Weerah. test	5.9	0.738	5.1	0.816	6.1	0.811	7.4	0.852
$\sigma^2 = (1, ..., 1, 5)$								
Boxov test	5.2	0.390	5.2	0.427	4.8	0.416	5.2	0.632
Krutch. test	5.8	0.402	5.2	0.453	5.2	0.537	5.9	0.745
Weerah. test	5.6	0.372	5.1	0.444	5.8	0.565	7.9	0.796
$\sigma^2 = (5, 1, ..., 1)$								
Boxov test	8.8	0.317	6.8	0.329	6.1	0.377	5.4	0.643
Krutch. test	7.9	0.431	6.1	0.475	6.1	0.564	5.2	0.773
Weerah. test	6.4	0.434	5.5	0.473	6.5	0.593	7.3	0.818

väčšine prípadov by som používal iba testy druhej skupiny. Situácie v simuláciách, v ktorých testy prvej skupiny mali väčšiu silu ako testy druhej skupiny (v prípade, že bol dostatok pozorovaní) boli akurát v prípade alternatívy A_1 : *posledná populácia iná* a keď väčší rozptyl bol v tejto poslednej (odlišnej) populácii. Tiež tu by som vyberal medzi Boxovým testom a Brovnovým-Forsyteovým testom podľa toho, ako ktorý test dodržuje hladinu významnosti. V ostatných heteroskedastických prípadoch už boli vždy silnejšie testy druhej skupiny. Z testov druhej skupiny by som používal Krutchkoffov test vo všetkých prípadoch, keď je počet pozorovaní v populáciách približne rovný počtu populácií. Pre väčší počet pozorovaní by som uprednostnil Welshov test, ktorý nepotrebuje dodatočné simulácie k určeniu kritickej hodnoty a má takmer zhodnú silu s Krutchkoffovým testom.

6.3 Porovnanie síl testov pre dvojne exponenciálne rozdelené populácie

6.3.1 Platí alternatíva $A_1 : \mu_1 = ... = \mu_{I-1} = 0$ *a* $\mu_I > 0$

Vyvážené prípady

Pre populácie z dvojne exponenciálneho rozdelenia by som v homoskedastickom prípade vybral F-test, ktorý najlepšie dodržuje hladinu významnosti a má tiež medzi uvažovanými testami najväčšiu silu. V heteroskedastickom prípade je opäť situácia podobná ako v predchádzajúcich prípadoch. V prípade, že väčší rozptyl je v líšiacej sa populácii, sú silnejšie testy prvej skupiny (F-test, Akritasov nevážený test, Boxov test a Brownov-Forsytheov test). V prípade, že je skôr menší rozptyl v líšiacej sa populácii, je lepšie použiť testy druhej skupiny (Krutchkoffov test, Weerahandiho test a Welshov test). Testy tej istej skupiny majú približne rovnakú silu za podmienky, že dodržia hladinu významnosti. Preto ako hlavné kritérium pre výber testu z nejakej skupiny bude dodržiavanie hladiny významnosti hypotézy H_0. Pre malý počet pozorovaní v každej populácii sa zdá, že medzi testami prvej skupiny najlepšie dodržuje hladinu významnosti Akritasov nevážený test, pre počet pozorovaní niekde medzi desať až dvadsaťpäť znovu najlepšie dodržuje hladinu významnosti Brownov-Forsytheov test, pre viac pozorovaní Boxov test. Medzi testami z druhej skupiny by som vyberal podľa počtu populácií. Pre tri populácie sa zdá byť najsilnejší Krutchkoffov test. Weerahandiho test by som uprednostnil pre počet populácií približne päť až desať, keď počet pozorovaní v každej populácii je aspoň rovný počtu populácií. Vtedy Krutchkoffov test a Welshov test mierne podhodnocujú hladinu významnosti a z toho dôvodu majú slabšiu silu ako Weerahandiho test. Ak je počet pozorovaní v populáciách menší ako počet populácií, potom Weerahandiho test mierne nadhodnocuje hladinu významnosti a preto by som radšej uprednostnil Welshov test. Taktiež pre väčší počet populácií by som použil Welshov test, ktorý je výpočtovo menej zložitý a dodržuje hladinu významnosti spoľahlivo. Ak máme pozorovaní v populáciách výrazne menší ako je počet populácií, tak sú silnejšie testy prvej skupiny ako testy druhej skupiny, preto by som použil nejaký test z prvej skupiny.

Tabuľka 58: Porovnanie sily F-testu, Krutchkoffovho testu a Weerahandiho testu vo vyváženom homoskedastickom prípade s populáciami z dvojne exponenciálneho rozdelenia za platnosti alternatívy $A_1 : \mu_1 = ... = \mu_{I-1} = 0$ *a* $\mu_I > 0$. Symbolom $\widehat{\alpha}\%$ označujeme odhad hladiny významnosti hypotézy H_0 v percentách.

I	**I=3**				**I=5**		**I=10**			
	$\widehat{\alpha}\%$	n=5	$\widehat{\alpha}\%$	n=50	$\widehat{\alpha}\%$	n=5	$\widehat{\alpha}\%$	n=10	$\widehat{\alpha}\%$	n=25
F-test	4.4	0.827	4.9	0.816	4.3	0.825	5.0	0.810	5.2	0.800
Krutch. test	3.4	0.801	5.0	0.815	2.9	0.723	3.2	0.717	3.9	0.767
Weerah. test	2.4	0.750	4.9	0.812	4.3	0.732	6.7	0.762	5.4	0.779

Nevyvážené prípady

V homoskedastických situáciách by som znovu odporučil používať F-test. V heteroskedastických populáciách sa nám oplatí rozdeliť testy do rovnakých skupín ako v predchádzajúcich prípadoch. Ak má poruchová populácia väčší rozptyl, opäť sú silnejšie testy prvej skupiny. Tu by sme mohli výber testov stanoviť podobne ako vo vyvážených prípadoch. Budeme si hlavne všímať, ako ktorý test dodržuje hladinu významnosti. Ak je počet pozorovaní v populáciách menší

Tabuľka 59: Porovnanie sily Brownovho-Forsytheovho testu, Boxovho testu, Krutchkoffovho testu, Weerahandiho testu a Welshovho testu v nevyváženom homoskedastickom prípade s dvojne exponenciálne rozdelenými populáciami za platnosti alternatívy A_1 :$\mu_1 = ... = \mu_{I-1} = 0$ a $\mu_I > 0$. Symbolom $\widehat{\alpha}\%$ označujeme odhad hladiny významnosti hypotézy H_0 v percentách.

I	**I=3**				**I=5**				**I=25**	
	$\widehat{\alpha}\%$	n=5	$\widehat{\alpha}\%$	n=25	$\widehat{\alpha}\%$	n=5	$\widehat{\alpha}\%$	n=25	$\widehat{\alpha}\%$	n=25
$\sigma^2 = (1, ..., 1, 5)$										
Akr.nev.test	6.0	0.510	4.3	0.384	6.3	0.523	5.1	0.467	5.8	0.551
B-F test	3.9	0.438	5.6	0.428	4.4	0.465	6.5	0.511	7.4	0.589
Boxov test	3.4	0.418	4.6	0.395	3.4	0.421	4.4	0.446	4.8	0.516
Krutch. test	3.8	0.371	4.6	0.323	2.6	0.265	4.2	0.281	4.1	0.175
Weerah. test	2.7	0.312	4.3	0.313	3.8	0.292	4.7	0.283	7.4	0.231
Welshov test	3.2	0.347	4.5	0.320	2.7	0.272	4.2	0.281	4.4	0.182
$\sigma^2 = (5, 1, ..., 1)$										
Akr.nev.test	5.6	0.542	4.8	0.386	6.1	0.569	5.1	0.524	5.6	0.666
B-F test	3.7	0.458	6.2	0.454	4.3	0.494	6.6	0.597	7.0	0.731
Boxov test	3.4	0.431	5.1	0.401	3.2	0.432	4.4	0.491	4.5	0.615
Krutch. test	3.7	0.740	5.0	0.740	3.0	0.705	4.5	0.829	4.0	0.768
Weerah. test	2.4	0.674	4.6	0.732	4.4	0.718	4.8	0.832	8.1	0.806
Welshov test	3.1	0.707	4.8	0.735	3.3	0.718	4.5	0.830	4.2	0.775

ako desať, tak by som odporučil Akritasov nevážený test, pre počet pozorovaní do dvadsaťpäť sa zdá byť najlepší Brownov-Forsytheov test, pre väčší počet pozorovaní zase Boxov test. Ak je naopak menší rozptyl v odlišnej populácii, testy druhej skupiny sú silnejšie ako testy prvej skupiny. Toto platí aj napriek tomu, že testy druhej skupiny podhodnocujú hladinu významnosti. Vo všetkých prípadoch s počtom populácií aspoň päť sa zdá byť Welshov test lepší ako Krutchkoffov test, aj keď tento rozdiel je minimálny. Pre tri populácie je silnejší Krutchkoffov test. Ak máme viac pozorovaní pre líšiacu sa populáciu, tak je Weerahandiho test silnejší ako ostatné dva testy jeho skupiny. Ak je naopak v líšiacej sa populácii menej pozorovaní, je lepšie uprednostniť buď Krutchkoffov test alebo Welshov test. Pre počet populácií viac ako desať už je Weerahandiho test silnejší od Welshovho testu na úkor toho, že Welshov test mierne podhodnocuje hladinu významnosti. Navyše s rastúcim počtom populácií už Weerahandiho test začína prekračovať hladinu významnosti. Preto by som pre počet populácií viac ako desať medzi testami druhej skupiny vždy uprednostnil Welshov test.

Tabuľka 60: Porovnanie sily Akritasovho neváženého testu, Brownovho-Forsytheovho testu, Boxovho testu, Krutchkoffovho testu, Weerahandiho testu a Welshovho testu v nevyváženom prípade s populáciami z dvojne exponenciálneho rozdelenia za platnosti alternatívy $A_1 : \mu_1 = ... = \mu_{I-1} = 0 \ a \ \mu_I > 0$. Symbolom $\widehat{\alpha}\%$ označujeme odhad hladiny významnosti hypotézy H_0 v percentách.

I	I=3				I=10			
$\widehat{\alpha}\%$ a n=(...)	$\widehat{\alpha}\%$	5,8,10	$\widehat{\alpha}\%$	20,15,10	$\widehat{\alpha}\%$	16:25	$\widehat{\alpha}\%$	2*(25:16)
$\sigma^2 = (1, ..., 1, 5)$								
Akr.nev.test	4.1	0.463	5.9	0.435	5.4	0.498	5.7	0.523
B-F test	4.9	0.493	5.2	0.416	7.1	0.546	7.3	0.570
Boxov test	3.7	0.453	4.9	0.408	4.4	0.465	4.6	0.491
Krutch. test	3.2	0.356	4.5	0.337	4.1	0.213	4.4	0.242
Weerah. test	3.1	0.370	4.0	0.300	5.8	0.251	5.3	0.256
Welshov test	3.5	0.363	4.0	0.324	4.2	0.218	4.5	0.246
$\sigma^2 = (5, 1, ..., 1)$								
Akr.nev.test	7.2	0.526	3.7	0.423	6.3	0.585	5.5	0.631
B-F test	4.8	0.427	5.8	0.526	7.4	0.646	7.3	0.723
Boxov test	5.4	0.458	4.4	0.469	5.1	0.528	4.6	0.590
Krutch. test	4.6	0.755	4.3	0.756	4.2	0.798	4.5	0.847
Weerah. test	3.6	0.761	4.2	0.731	6.2	0.826	5.3	0.846
Welshov test	4.1	0.723	4.4	0.758	4.4	0.804	4.6	0.848

6.3.2 Platia alternatívy A_2: μ *lineárne* a $A_3 : \alpha_1 = ... = \alpha_{\lfloor I/2 \rfloor} = 0 \ a \ \alpha_{\lfloor I/2 \rfloor + 1} = ... = \alpha_I > 0$

Vyvážené prípady

Výraznú zmenu oproti predchádzajúcim prípadom nám prináša už prípad pre homoskedastické populácie. Keď je malý počet populácií (menší ako päť), tak najsilnejší test je stále F-test. Avšak ak je počet populácií aspoň päť, potom už testy druhej skupiny majú väčšiu silu ako F-test a ostatné testy prvej skupiny. Výber konkrétneho testu z druhej skupiny urobíme ako v heteroskedastických populáciách. Pretože pre heteroskedastické populácie už k žiadnej výraznej zmene nedošlo, opäť sú najsilnejšie testy druhej skupiny. Pretože tieto testy majú veľmi podobnú silu v prípade, že dodržia hladinu významnosti, preto výber jednotlivých testov pre jednotlivé heteroskedastické situácie bude ovplyvnený hlavne dodržiavaním hladiny významnosti. Ak máme populácií menej ako päť, tak najmenej podhodnocuje hladinu významnosti Krutchkoffov test. Pre aspoň päť populácií už Weerahandiho test najmenej podhodnocuje hladinu významnosti medzi testami druhej skupiny a preto má aj najväčšiu silu. Problém tohto testu je však v tom, že pre malý počet pozorovaní a počet populácií väčší ako päť už nadhodnocuje hladinu významnosti. Preto by sme ho nemali používať, keď je počet populácií aspoň desať a počet pozorovaní v každej populácii je menší ako počet týchto populácií. Pre tieto situácie je lepšie používať Welshov test, ktorý menej podhodnocuje hladinu významnosti hypotézy H_0 ako Krutchkoffov test. Welshov test je silnejší ako testy prvej skupiny dokonca aj pre prípad, že $n \ll I$.

Tabuľka 61: Porovnanie sily F-testu, Krutchkoffovho testu a Weerahandiho testu vo vyváženom homoskedastickom prípade s populáciami z dvojne exponenciálneho rozdelenia za platnosti alternatívy A_2 : μ_i *lineárne*. Symbolom $\widehat{\alpha}\%$ označujeme odhad hladiny významnosti hypotézy H_0 v percentách.

I	**I=3**				**I=5**		**I=10**			
	$\widehat{\alpha}\%$	n=5	$\widehat{\alpha}\%$	n=50	$\widehat{\alpha}\%$	n=5	$\widehat{\alpha}\%$	n=10	$\widehat{\alpha}\%$	n=25
F-test	4.1	0.845	5.1	0.838	4.9	0.865	4.9	0.835	5.2	0.859
Krutch. test	3.1	0.837	4.8	0.838	2.7	0.836	3.2	0.838	4.4	0.868
Weerah. test	2.1	0.804	4.8	0.838	4.2	0.886	6.5	0.902	5.9	0.892

Tabuľka 62: Porovnanie sily Brownovho-Forsytheovho testu, Krutchkoffovho testu, Weerahandiho testu a Welshovho testu vo vyváženom prípade s populáciami z dvojne exponenciálneho rozdelenia a rozptylom $\sigma^2 = (1, ..., 1, 5)$ za platnosti alternatívy A_2 : μ_i *lineárne*. Symbolom $\widehat{\alpha}\%$ označujeme odhad hladiny významnosti hypotézy H_0 v percentách.

I	**I=3**				**I=10**				**I=25**	
	$\widehat{\alpha}\%$	n=5	$\widehat{\alpha}\%$	n=50	$\widehat{\alpha}\%$	n=10	$\widehat{\alpha}\%$	n=25	$\widehat{\alpha}\%$	n=5
B-F test	3.8	0.462	6.6	0.435	6.5	0.594	7.4	0.643	4.3	0.710
Krutch. test	3.9	0.565	5.3	0.501	3.1	0.725	4.4	0.770	1.4	0.711
Weerah. test	2.7	0.524	5.1	0.498	6.7	0.823	5.8	0.803	22.6	0.989
Welshov test	3.3	0.532	5.2	0.499	3.9	0.751	4.4	0.772	6.0	0.918

Nevyvážené prípady

Správanie testov v nevyvážených prípadoch sa prakticky nemení od ich chovania vo vyvážených prípadoch. Testy druhej skupiny dominujú nad testami prvej skupiny takmer vo všetkých prípadoch. Opäť je dobré vyberať testy hlavne podľa počtu populácií a podľa počtu pozorovaní v populáciách. Pre tri populácie je v homoskedastických populáciách najsilnejší F-test. V tomto nastavení testy druhej skupiny podhodnocujú hladinu významnosti (s rastúcim počtom pozorovaní sa podhodnocovanie zmenšuje). Napriek tomu sú ale aj tak silnejšie ako testy prvej skupiny v heteroskedastických populáciách. Pre tri populácie najmenej podhodnocuje hladinu významnosti z testov druhej skupiny Krutchkoffov test. Preto by som ho použil v prípade troch heteroskedastických populácií. Pre päť populácií sa už situácia zmenila. Z testov druhej skupiny najmenej podhodnocuje hladinu významnosti Weerahandiho test, ktorý je silnejší ako F-test aj v homoskedastických prípadoch. Preto by som pre päť populácií použil práve Weerahandiho test. Pre viac populácií má opäť Weerahandiho test najvyššiu odhadovanú hladinu významnosti medzi testami druhej skupiny. Problém avšak nastáva, keď máme málo pozorovaní (zdá sa, že by mohlo stačiť, aby počet pozorovaní v každej populácii bol aspoň dva krát väčší ako počet populácií pre dodržanie hladiny významnosti Weerahandiho testu). Potom už Weerahandiho test nadhodnocuje hladinu významnosti. Preto by som pre málo pozorovaní v populáciách použil Welshov test, ktorý je výpočtovo jednoduchší a aj menej podhodnocuje (rozdiel je však minimálny) hladinu významnosti ako Krutchkoffov test.

Tabuľka 63: Porovnanie sily F-tesu, Krutchkoffovho testu a Weerahandiho testu v nevyváženom prípade s dvojne exponenciálne rozdelenými populáciami za platnosti alternatívy A_2 : μ_i *lineárne*. Symbolom $\widehat{\alpha}\%$ označujeme odhad hladiny významnosti hypotézy H_0 v percentách.

I	**I=3**				**I=5**		**I=25**	
$\widehat{\alpha}\%$ a n=(...)	$\widehat{\alpha}\%$	5,8,10	$\widehat{\alpha}\%$	10,15,20	$\widehat{\alpha}\%$	6:10	$\widehat{\alpha}\%$	50:26
$\sigma^2 = (1, ..., 1)$								
F-test	4.7	0.852	4.7	0.847	4.4	0.856	5.0	0.848
Krutch. test	3.6	0.831	4.2	0.842	3.3	0.842	4.5	0.864
Weerah. test	3.1	0.816	4.0	0.836	4.5	0.875	6.5	0.896
$\sigma^2 = (1, ..., 1, 5)$								
F-test	3.0	0.425	3.2	0.390	5.0	0.514	9.6	0.791
Krutch. test	3.5	0.539	4.2	0.526	3.1	0.641	4.2	0.819
Weerah. test	3.4	0.513	4.1	0.517	4.3	0.685	6.4	0.869
$\sigma^2 = (5, 1, ..., 1)$								
F-test	11.6	0.616	11.6	0.592	9.7	0.654	5.8	0.691
Krutch. test	4.5	0.565	4.7	0.557	3.3	0.665	4.2	0.811
Weerah. test	3.7	0.557	4.2	0.552	4.5	0.721	6.2	0.852

6.3.3 Záver na simulácie pre populácie z dvojne exponenciálneho rozdelenia

Pre populácie z dvojne exponenciálneho rozdelenia už nastáva zmena pri chovaní testov. Testy druhej skupiny získali na sile v porovnaní s testami prvej skupiny. To sa hlavne prejavuje u symetrických alternatív A_2 a A_3. Pre aspoň päť populácií sú už testy druhej skupiny silnejšie ako testy prvej skupiny nielen v heteroskedastických, ale aj v homoskedastických situáciách. Pre tri populácie a symetrické alternatívy je v homoskedastických prípadoch stále najsilnejší F-test. Pre nesymetrickú alternatívu A_1 je situácia vo všetkých prípadoch podobná ako pre normálne rozdelené populácie. V homoskedastických prípadoch je najsilnejší

F-test, v heteroskedastických prípadoch je rozumné rozdeliť testy do dvoch skupín podobných testov. Testy prvej skupiny sú silnejšie v prípade, že v líšiacej sa populácii je väčší rozptyl, alebo keď máme výrazne menej pozorovaní v populáciách, ako je počet populácií. Naopak v prípade menšieho rozptylu v odlišnej populácii a dostatku pozorovaní majú testy prvej skupiny výrazne menšiu silu. Ak heteroskedastická situácia nasvedčuje tomu, že je lepšie použiť nejaký test z prvej skupiny, tak by som pre počet pozorovaní v každej populácii menší ako desať použil Akritasov nevážený test, pre približne desať až dvadsaťpäť pozorovaní v každej populácii by som použil Brownov-Forsytheov test, v prípade väčšieho počtu pozorovaní Boxov test. Pri výbere testov z druhej skupiny by som vyberal hlavne podľa počtu populácií. Pre tri populácie je najsilnejší Krutchkoffov test, pre približne päť až desať populácií sa zdá byť najsilnejší Weerahandiho test, ktorý pre väčší počet populácií a neveľký počet pozorovaní začína nadhodnocovať hladinu významnosti. Z toho dôvodu by som pre väčší počet populácií používal Welshov test.

6.4 Porovnanie síl testov pre logaritmicko-normálne rozdelené populácie

Pre logaritmicko-normálne rozdelené populácie májú všetky testy problém s dodržaním hladiny významnosti. Preto treba testy pre takto rozdelené populácie používať s veľkou opatrnosťou.

6.4.1 Platí alternatíva $A_1 : \mu_1 = ... = \mu_{I-1} = 0$ a $\mu_I > 0$

Vyvážené prípady

Pre homoskedastické prípady by sme mali vyberať test podľa počtu populácií. Ak je populácií menej ako desať, tak sú testy druhej skupiny silnejšie ako testy prvej skupiny. Najspoľahlivejšie medzi testami druhej skupiny dodržuje hladinu významnosti Krutchkoffov test a preto by som používal práve tento test. Pre desať a viac populácií už testy druhej skupiny začínajú strácať na sile a navyše začínajú výrazne nadhodnocovať hladinu významnosti (okrem Krutchkoffovho testu, keď máme okolo päť pozorovaní v každej populácii). Preto je určite lepšie použiť F-test, ktorý je pre takýto počet populácií najsilnejší medzi testami, ktoré takmer dodržujú hladinu významnosti. Pre heteroskedastické prípady a počet populácií menších ako desať sú tiež lepšie testy druhej skupiny ako prvej skupiny. Výnimku tvorí akurát prípad, keď je v líšiacej sa populácii väčší rozptyl. Medzi testami druhej skupiny najlepšie dodržuje hladinu významnosti Krutchkoffov test. Preto by som použil práve tento test. Pre viac populácií je lepšie vyberať už iba medzi testami prvej skupiny. Testy druhej skupiny už výrazne nadhodnocujú hladinu významnosti. V prípadoch, kde je lepšie použiť testy prvej skupiny, by som použil F-test, ktorý je najstabilnejší medzi testami prvej skupiny. Ostatné testy výrazne podhodnocujú hladinu významnosti pre malú heteroskedasticitu a malý počet pozorovaní. Tiež si je dobré ešte všimnúť, že Krutchkoffov test v prípade, že medzi rozptylmi populácií je lineárna závislosť, dodržuje hladinu významnosti horšie ako F-test.

Tabuľka 64: Porovnanie sily F-testu a Krutchkoffovho testu vo vyváženom prípade s populáciami z logaritmicko-normálneho rozdelenia za platnosti alternatívy $A_1 : \mu_1 = ... = \mu_{I-1} = 0$ a $\mu_I > 0$. Symbolom $\widehat{\alpha}\%$ označujeme odhad hladiny významnosti hypotézy H_0 v percentách.

I	**I=3**				**I=10**		**I=25**			
	$\widehat{\alpha}\%$	n=5	$\widehat{\alpha}\%$	n=50	$\widehat{\alpha}\%$	n=5	$\widehat{\alpha}\%$	n=5	$\widehat{\alpha}\%$	n=10
$\sigma^2 = (1, ..., 1)$										
F-test	3.4	0.872	4.4	0.840	4.1	0.851	4.4	0.844	4.3	0.829
Krutch. test	3.2	0.931	5.3	0.918	5.0	0.882	5.8	0.808	15.2	0.929
$\sigma^2 = (1, ..., 1, 5)$										
F-test	9.7	0.698	8.3	0.438	6.8	0.713	5.7	0.725	6.2	0.657
Krutch. test	9.8	0.631	7.0	0.317	7.2	0.419	8.2	0.369	16.5	0.472
$\sigma^2 = (5, 1, ..., 1)$										
F-test	10.0	0.699	7.9	0.540	7.5	0.765	5.7	0.788	6.2	0.767
Krutch. test	9.5	0.901	7.1	0.832	7.4	0.889	8.0	0.830	16.4	0.926
$\sigma^2 = (I, ..., 1)$										
F-test	5.1	0.719	4.9	0.604	5.2	0.317	5.9	0.111	5.9	0.103
Krutch. test	6.3	0.869	6.4	0.791	16.9	0.912	33.6	0.914	39.8	0.963

Nevyvážené prípady

V homoskedastických prípadoch sa opäť musíme rozhodovať podľa počtu populácií. Pre menej ako desať populácií je najsilnejší Krutchkoffov test, pre aspoň desať populácií F-test. V heteroskedastických populáciách je situácia zložitejšia najmä pre malý počet populácií. Ak máme desať a viac populácií, tak testy druhej skupiny výrazne nadhodnocujú hladinu významnosti hypotézy H_0. Preto ich nie je dobré používať. Pre veľký počet populácií z testov prvej skupiny najlepšie dodržuje hladinu významnosti Brownov-Forsytheov test. Z toho dôvodu sa zdá byť najrozumnejšie použiť práve tento test. Keď máme menej ako desať populácií, testy druhej skupiny sú silnejšie ako testy prvej skupiny okrem prípadu, že je veľký rozptyl v líšiacej sa populácii. Z testov druhej skupiny sa najlepšie správa Krutchkoffov test. Keď je lepšie použiť nejaký test prvej skupiny, použil by som F-test v prípade, keď máme menej pozorovaní v líšiacej sa populácii. Naopak Brownov-Forsytheov test je silnejší ako F-test, keď je viac pozorovaní v líšiacej sa populácii. Ešte si je dobré povšimnúť, že keď máme v populácii s najmenším počtom pozorovaní najväčší rozptyl, tak všetky testy výrazne nadhodnocujú hladinu významnosti hladiny významnosti α.

Tabuľka 65: Porovnanie sily F-testu, Brownovho-Forsytheovho testu a Krutchkoffovho testu v nevyváženom prípade s populáciami z logaritmicko-normálneho rozdelenia za platnosti alternatívy $A_1 : \mu_1 = ... = \mu_{I-1} = 0$ a $\mu_I > 0$. Symbolom $\widehat{\alpha}\%$ označujeme odhad hladiny významnosti hypotézy H_0 v percentách.

I	**I=3**				**I=10**			
$\widehat{\alpha}\%$ a n=(...)	$\widehat{\alpha}\%$	5,8,10	$\widehat{\alpha}\%$	20,15,10	$\widehat{\alpha}\%$	16:25	$\widehat{\alpha}\%$	2*(25:16)
$\sigma^2 = (1, ..., 1)$								
F-test	3.2	0.883	4.0	0.862	4.4	0.855	4.7	0.878
B-F test	2.6	0.833	3.0	0.872	3.7	0.827	4.5	0.881
Krutch. test	4.0	0.940	5.3	0.949	11.1	0.966	10.3	0.971
$\sigma^2 = (1, ..., 1, 5)$								
F-test	7.6	0.597	13.1	0.650	6.4	0.575	9.6	0.656
B-F test	6.9	0.655	10.7	0.464	6.5	0.589	7.7	0.610
Krutch. test	6.3	0.606	10.5	0.412	12.0	0.383	11.6	0.337
$\sigma^2 = (5, 1, ..., 1)$								
F-test	13.5	0.779	6.5	0.552	9.2	0.783	6.5	0.703
B-F test	12.9	0.620	7.2	0.637	6.9	0.695	7.3	0.971
Krutch. test	13.8	0.905	6.8	0.905	11.7	0.963	10.1	0.745

6.4.2 Platia alternatívy A_2: μ *lineárne* a $A_3 : \alpha_1 = ... = \alpha_{\lfloor I/2 \rfloor} = 0$ a $\alpha_{\lfloor I/2 \rfloor + 1} = ... = \alpha_I > 0$

Vyvážené prípady

V homoskedastických prípadoch je vo väčšine prípadov najsilnejší Krutchkoffov test. Avšak pre desať a viac populácií už výrazne nadhodnocuje hladinu významnosti hypotézy H_0 okrem prípadu, že je strašne málo pozorovaní v populáciách (okolo päť). Tiež napríklad pre päť populácií a päťdesiat pozorovaní už Krutchkoffov test nedodržuje hladinu významnosti. Preto je rozumnejšie používať F-test, ktorý oveľa lepšie dodržuje hladinu významnosti. V heteroskedastických prípadoch sú opäť silnejšie testy druhej skupiny. Medzi nimi je len malý rozdiel v odhadovanej hladine α. To spôsobuje, že je malý rozdiel aj v ich sile. Preto výberom jedného z týchto testov veľa nezískame, ale ani nestratíme. Ale sú aj situácie, kde nie je rozumné používať tieto testy. Dôvodom je výrazné prekročenie hladiny významnosti. Prvú situáciu tvorí, keď je medzi rozptylmi nejaká lineárna závislosť. Pre tieto situácie napríklad F-test dodržuje hladinu významnosti, zatiaľ čo testy druhej skupiny výrazne nadhodnotia hladinu významnosti. Druhý prípad, kde nie je rozumné používať testy druhej skupiny je, keď máme desať a viac populácií. Vtedy opäť F-test dodržuje hladinu významnosti α podstatne lepšie ako testy druhej skupiny. Preto je použitie F-testu rozumnejšie, aj keď má slabšiu silu.

Tabuľka 66: Porovnanie sily F-testu a Krutchkoffovho testu vo vyváženom prípade s populáciami z logaritmicko-normálneho rozdelenia za platnosti alternatívy $A_2 : \mu_i$ *lineárne*. Symbolom $\widehat{\alpha}\%$ označujeme odhad hladiny významnosti hypotézy H_0 v percentách.

I	**I=3**		**I=5**				**I=10**			
	$\widehat{\alpha}\%$	n=5	$\widehat{\alpha}\%$	n=5	$\widehat{\alpha}\%$	n=50	$\widehat{\alpha}\%$	n=5	$\widehat{\alpha}\%$	n=10
$\sigma^2 = (1, ..., 1)$										
F-test	3.4	0.682	3.3	0.680	4.4	0.791	3.9	0.647	3.7	0.709
Krutch. test	3.0	0.746	3.5	0.829	7.0	0.861	4.8	0.876	8.4	0.917
$\sigma^2 = (1, ..., 1, 5)$										
F-test	10.3	0.364	8.5	0.462	8.6	0.509	7.1	0.504	7.7	0.540
Krutch. test	10.2	0.436	8.0	0.617	7.6	0.634	7.6	0.769	11.5	0.828
$\sigma^2 = (1, 2, ..., I)$										
F-test	4.7	0.440	5.0	0.247	5.4	0.284	5.8	0.150	5.4	0.230
Krutch. test	6.6	0.513	10.5	0.354	8.4	0.383	17.3	0.194	20.4	0.417

Nevyvážené prípady

Pokiaľ je populácií menej ako desať, v homoskedastických populáciách je opäť najlepšie používať Krutchkoffov test. Pre desať a viac populácií už tento test výrazne prekračuje hladinu významnosti. Preto je lepšie používať F-test. V heteroskedastických populáciách je situácia podobná. Pre malý počet populácií je najlepšie používať Krutchkoffov test. Ale existujú aj výnimky, keď je lepšie použiť F-test. Prvá situácia je, keď je rozptyl lineárny taký, že väčší rozptyl je v tých populáciách, kde máme menej pozorovaní. Vtedy F-test lepšie dodržuje hladinu významnosti. Druhý prípad je, keď máme μ_i lineárne rastúce, posledná populácia má najmenej pozorovaní a zároveň najväčší rozptyl. Vtedy má F-test lepšiu silu ako Krutchkoffov test. To platí najmä pre tri populácie. Pre päť populácií to platí, keď je rozptyl lineárne rastúci. Keď je populácií viac, je lepšie používať F-test. Ten je výhodné pri veľkej heteroskedasticite nahradiť Brownovým-Forsytheovým testom, ktorý za takýchto podmienok menej nadhodnocuje hladinu významnosti. Ale na druhú stranu tento test podhodnocuje hladinu významnosti za homoskedasticity. Ešte si je dobré uvedomiť, že keď máme väčší rozptyl tam, kde máme menej pozorovaní, potom majú všetky testy výrazný problém s dodržaním hladiny významnosti.

6.4.3 Záver na simulácie pre populácie z logaritmicko-normálneho rozdelenia

Pre logaritmicko-normálne rozdelené populácie je hlavný problém v tom, že testy majú výrazný problém s dodržiavaním hladiny významnosti. Preto výber testov je zameraný hlavne na to, ako ktorý test dodržuje hladinu významnosti.

Tabuľka 67: Porovnanie sily F-testu a Krutchkoffovho testu v nevyváženom prípade s populáciami z logaritmicko-normálneho rozdelenia za platnosti alternatívy A_2 : μ_i *lineárne rastúce*. Symbolom $\widehat{\alpha}\%$ označujeme odhad hladiny významnosti hypotézy H_0 v percentách.

I	**I=3**				**I=5**			
$\widehat{\alpha}\%$ a n=(...)	$\widehat{\alpha}\%$	10,10,5	$\widehat{\alpha}\%$	10,5,5	$\widehat{\alpha}\%$	4x10,5	$\widehat{\alpha}\%$	10,4x5
$\sigma^2 = (1, ..., 1)$								
F-test	3.6	0.769	3.3	0.760	4.2	0.740	3.9	0.774
Krutch. test	4.2	0.825	3.9	0.804	5.5	0.858	3.8	0.861
$\sigma^2 = (1, ..., 1, 5)$								
F-test	14.6	0.625	12.8	0.560	12.5	0.660	10.0	0.632
Krutch. test	13.8	0.546	10.8	0.424	13.2	0.711	9.8	0.663
$\sigma^2 = (1, 2, ..., I)$								
F-test	6.3	0.598	7.0	0.616	6.4	0.338	7.8	0.424
Krutch. test	10.8	0.604	11.0	0.534	12.8	0.393	13.0	0.339

Pokiaľ je počet populácií menší ako desať, tak by som pre homoskedastické populácie používal Krutchkoffov test. Pre aspoň desať populácií už má tento test výrazný problém s dodržiavaním hladiny významnosti, preto by som ho nahradil F-testom. Tiež v heteroskedastických populáciách by som testy vyberal hlavne podľa počtu populácií. Pre počet populácií menší ako desať majú testy prvej skupiny slabšiu silu ako testy druhej skupiny okrem alternatívy A_1 : *posledná populácia iná*, keď bol väčší rozptyl v odlišnej populácii. Medzi testami druhej skupiny najlepšie dodržuje hladinu významnosti Krutchkoffov test, preto by som používal práve tento test ako reprezentanta druhej skupiny. Ak rozptyl tvoril lineárnu postupnosť, tak F-test dodržiaval hladinu významnosti výrazne lepšie ako Krutchkoffov test. Preto v prípade tejto znalosti by som ho tiež uprednostnil pred Krutchkoffovým testom. Ak by bolo populácií aspoň desať, tak by som sa rozhodoval medzi dvomi testami. F-test by som použil vo väčšine prípadov. Brownov-Forsytheov test sa zdal byť lepší iba v prípadoch veľkej heteroskedasticity, kde lepšie dodržiaval hladinu významnosti ako F-test a v nevyvážených prípadoch za platnosti alternatívy A_1, keď bolo viac pozorovaní v líšiacej sa populácii.

7 Záver

Cieľom tejto diplomovej práce bolo porovnať niekoľko testov na testovanie zhody stredných hodnôt niekoľkých nezávislých populácií. Vieme, že za predpokladu normality a homoskedasticity populácií je F-test rovnomerne najsilnejší test. Otázkou však bolo, čo sa stane pri poruche jedného z týchto dvoch predpokladov. Preto som pre niektoré takéto prípady skúmal, či F-test stále dodržuje hladinu významnosti a či poznáme nejaký iný test, ktorý sa správa lepšie ako F-test v niektorých nami zvolených prípadoch.

Môj výskum ukázal, že nie vždy je najlepšie použiť F-test. Tento test má veľký problém s dodržiavaním hladiny významnosti pri poruche homoskedasticity. Ukázalo sa, že má význam rozdeliť testy do dvoch skupín podobných testov. Medzi testami rovnakej skupiny sú rozdiely v dodržiavaní hladiny významnosti, ale zato ich sily sú takmer zhodné v prípade, že testy dodržia hladinu významnosti. Prvú skupinu tvoria F-test, Akritasov nevážený test, Boxov test a Brownov-Forsytheov test. Do druhej skupiny sú zaradené Krutchkoffov test, Welshov test a Weerahandiho test.

V normálnych homoskedastických prípadoch sa ukázalo, že aj keď je F-test rovnomerne najsilnejší test, môžeme ho vo väčšine prípadov (okrem malého počtu populácií a malého počtu pozorovaní) nahradiť Akritasovým neváženým testom. Tento test síce o niečo málo nadhodnocuje hladinu významnosti, ale zato má vo väčšine prípadov lepšiu silu ako F-test a navyše je robustnejší voči poruche homoskedasticity.

Zmeny použitia testov v homoskedastických prípadoch nastali pri zmene rozdelenia populácií. V prípade dvojne exponenciálneho rozdelenia populácií a symetrickej alternatívy sa ukázalo, že pre väčší počet populácií (aspoň päť) sú testy druhej skupiny silnejšie ako testy prvej skupiny. V prípade logaritmicko-normálneho rozdelenia sú testy druhej skupiny silnejšie pre počet populácií menej ako desať. Naopak pre rovnomerné rozdelenie populácií testy druhej skupiny strácajú na sile v porovnaní s testami prvej skupiny pre všetky tri uvažované alternatívy.

Ďalšou dôležitou vecou bolo prešetrenie testov v prípade poruchy homoskedasticity. Ukázalo sa, že testy druhej skupiny majú väčšiu silu ako testy prvej skupiny pri symetrických alternatívach ($A_2 : \mu_i$ *lineárne* a A_3 : *pol na pol*) okrem prípadu, keď máme veľa populácií a málo pozorovaní. Pre nesymetrickú alternatívu A_1 : *jedno* μ_i *iné* sa ukázalo, že je ešte viac prípadov, pre ktoré majú testy prvej skupiny väčšiu silu. Takúto situáciu tvorí aj prípad, keď je väčší rozptyl v poruchovej populácii (tj. tá, ktorá sa od ostatných líši).

Pri poruche ako normality, tak aj homoskedasticity nedošlo k žiadnej neočakávanej zmene. Najväčší problém pri výbere testov je dodržiavanie hladiny významnosti testovania hypotézy H_0. Pre symetrické rozdelenia (rovnomerné a dvojne exponenciálne rozdelenie) nedošlo k žiadnej výraznej zmene (s dodržiavaním hladiny významnosti alebo sily testov) v porovnaní s normálnym rozdelením populácií.

V nesymetrických prípadoch (logaritmicko-normálne rozdelenie) už k výrazným zmenám došlo. Všetky testy už majú veľké problémy s dodržiavaním hladiny významnosti a preto je ťažké vybrať test, ktorý dodrží hladinu významnosti a zároveň je aj najsilnejší medzi uvažovanými testami za platnosti nami vybranej alternatívy.

Príloha

Súčasťou tejto diplomovej práce je aj CD, ktoré obsahuje 13 súborov s výsledkami simulácií a 1 súbor so zdrojovým kódom k softwaru **R**.

NormA1a.pdf - *Obsahuje silu testov pre normálne rozdelené populácie za platnosti alternatívy* A_1.

NormA1b.pdf - *Obsahuje upravenú silu testov pre normálne rozdelené populácie za platnosti alternatívy* A_1.

NormA2.pdf - *Obsahuje silu testov a upravenú silu testov pre normálne rozdelené populácie za platnosti alternatívy* A_2.

NormA3.pdf - *Obsahuje silu testov a upravenú silu testov pre normálne rozdelené populácie za platnosti alternatívy* A_3.

RovnA1.pdf - *Obsahuje silu testov a upravenú silu testov pre rovnomerne rozdelené populácie za platnosti alternatívy* A_1.

RovnA2.pdf - *Obsahuje silu testov a upravenú silu testov pre rovnomerne rozdelené populácie za platnosti alternatívy* A_2.

RovnA3.pdf - *Obsahuje silu testov a upravenú silu testov pre rovnomerne rozdelené populácie za platnosti alternatívy* A_3.

DexpA1.pdf - *Obsahuje silu testov a upravenú silu testov pre dvojne exponenciálne rozdelené populácie za platnosti alternatívy* A_1.

DexpA2.pdf - *Obsahuje silu testov a upravenú silu testov pre dvojne exponenciálne rozdelené populácie za platnosti alternatívy* A_2.

DexpA3.pdf - *Obsahuje silu testov a upravenú silu testov pre dvojne exponenciálne rozdelené populácie za platnosti alternatívy* A_3.

LognA1.pdf - *Obsahuje silu testov a upravenú silu testov pre logaritmicko-normálne rozdelené populácie za platnosti alternatívy* A_1.

LognA2.pdf - *Obsahuje silu testov a upravenú silu testov pre logaritmicko-normálne rozdelené populácie za platnosti alternatívy* A_2.

LognA3.pdf - *Obsahuje silu testov a upravenú silu testov pre logaritmicko-normálne rozdelené populácie za platnosti alternatívy* A_3.

Simulacie.R - *Obsahuje zdrojový kód k simuláciám a ku všetkým uvažovaným testom pre software* **R**.

Použitá literatúra

Akritas, M. and Papadatos, N. (2004), "Heteroscedastic one-way ANOVA and lack-of-fit tests," *Journal of the American Statistical Association*, 99, 368–382.

Anděl, J. (2001), *Základy matematické statistiky*, vol. Preprint, Praha.

Box, G. E. P. (1954), "Some theorems on quadratic forms applied in the study of analysis of variance problems,I.Effect of inequality of variance in the one-way classification," *Annals of Mathematical Statistics*, 25, 290–302.

Brown, M. B. and Forsythe, A. B. (1974), "The anova and multiple comparisons for data with heterogeneous variances," *Biometrics*, 30, 719–724.

Chen, S. and Chen, H. (1998), "Single-stage analysis of variance under heteroscedasticity," *Communications in Statistics-Simulation and Computation*, 327, 641–666.

Krutchkoff, R. G. (1988), "One-way fixed effects analysis of variance when the error variances may be unequal," *Journal of Statistical Computation and Simulation*, 30, 259–271.

Weerahandi, S. (1995), "ANOVA under unequal error variances," *Biometrics*, 51, 589–599.

Welsh, B. L. (1951), "On the Comparison of Several Mean Values: An Alternative Approach," *Biometrika*, 38, 330–336.