Univerzita Karlova v Praze
Matematicko-fyzikální fakulta

BAKALÁŘSKÁ PRÁCE

Hana Jelínková

Analýza rozptylu dvojného třídění pro heteroskedastická data

Katedra pravděpodobnosti a matematické statistiky

Vedoucí bakalářské práce: Mgr. Peter Bubelíny
Studijní obor: Obecná matematika

2008

Děkuji vedoucímu bakalářské práce Mgr. Peterovi Bubelínymu za konzultace, poskytnutí literatury, cenné rady a kritické připomínky, které mi výrazně pomohly při psaní této práce.

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

V Praze dne 8. prosince 2008 Hana Jelínková

Obsah

Název práce: Analýza rozptylu dvojného třídění pro heteroskedastická data
Autor: Hana Jelínková
Katedra (ústav): Katedra pravděpodobnosti a matematické statistiky
Vedoucí bakalářské práce: Mgr. Peter Bubelíny
e-mail vedoucího: bubeliny@karlin.mff.cuni.cz

Abstrakt: V předložené práci se zaměřujeme na porovnání několika testů na testování hypotéz o nulovosti řádkových a sloupcových efektů a interakcí v modelu analýzy rozptylu dvojného třídění. Pro testy těchto hypotéz uvažujeme celkem 6 testů. Při testování hypotéz se soustředíme na vyvážené třídění, předpokládáme normální rozdělení populací a uvažujeme homoskedastické i heteroskedastické případy. Testy jsou porovnávány na základě provedených simulací, při nichž zkoumáme, jak dodržují hladinu významnosti α a jakou mají sílu za platnosti různých alternativ.

Klíčová slova: analýza rozptylu, dvojné třídění, heteroskedasticita

Title: Two-way analysis of variance for heteroscedastic data
Author: Hana Jelínková
Department: Department of Probability and Mathematical Statistics
Supervisor: Mgr. Peter Bubelíny
Supervisor's e-mail address: bubeliny@karlin.mff.cuni.cz

Abstract: In the present work we target the comparison of tests for testing the hypotheses of no row and column effects and interactions in the Two-Way ANOVA model. For testing these hypotheses we consider 6 tests. We focus on balanced data, assume the normal distribution of populations and consider homoscedastic cases as well as heteroscedastic. Tests are compared on the base of simulations in which we test if they hold the nominal level α and how powerful these tests are under the assumption that one of different alternatives is valid.

Keywords: two-way analysis of variance, heteroskedasticity

Kapitola 1

Úvod

Analýza rozptylu je metoda statistického vyhodnocování experimentů, u nichž zkoumáme vliv jednoho či více faktorů na hodnotu zkoumaného znaku. Pokud je počet faktorů roven dvěma, potom hovoříme o modelu dvojného třídění.

Předpokládejme, že máme nezávislé náhodné veličiny $Y_{ij1}, \ldots Y_{ijn_{ij}}$ s neznámou střední hodnotou μ_{ij} a s neznámým a konečným rozptylem $\sigma^2_{ij} > 0$, $i = 1, ..., I$, $j = 1, ..., J$, kde I je počet úrovní prvního faktoru a J počet úrovní faktoru druhého. Náhodný vektor $(Y_{ij1}, ..., Y_{ijn_{ij}})'$ nazýváme (i, j)*-tou podtřídou* (nebo také *populací)* a n_{ij} potom *počet pozorování* v (i, j)-té podtřídě. Pokud je n_{ij} pro všechny dvojice (i, j) stejné, pak hovoříme o tzv. *vyváženém třídění,* v opačném případě o *třídění nevyváženém.* Dále předpokládejme, že se náhodné veličiny $Y_{ij1}, \ldots Y_{ijn_{ij}}$ řídí modelem

$$Y_{ijp} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijp} \qquad \text{(model dvojného třídění s interakcemi).}$$

V tomto modelu μ nazýváme *obecná střední hodnota,* α_i *řádkové efekty (efekt i-té úrovně prvního faktoru),* β_j *sloupcové efekty (efekt j-té úrovně druhého faktoru).* Parametrům γ_{ij} se říká *interakce.* Dále e_{ijp}, $i = 1, ..., I$, $j = 1, ..., J$, $p = 1, ..., n_{ij}$ jsou nezávislé náhodné veličiny s nulovou střední hodnotou a kladným rozptylem σ^2_{ij}. Poznamenejme, že střední hodnotu (i, j)-té populace můžeme zapsat ve tvaru $\mu_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij}$, $i = 1, ..., I$, $j = 1, ..., J$.

V modelu dvojného třídění testujeme tyto tři základní hypotézy:

$$H_A\colon \alpha_1 = \ldots = \alpha_I = 0,$$

$$H_B\colon \beta_1 = \ldots = \beta_J = 0,$$

$$H_{AB}\colon \gamma_{ij} = 0,\ i = 1, ..., I,\ j = 1, ..., J,$$

proti jejich přirozeným alternativám H^1_A, H^1_B a H^1_{AB}.

V případě hypotézy H_A tedy testujeme významnost řádkových efektů bez ohledu na významnost sloupcových efektů a interakcí, obdobně v případě hypotézy H_B. Pokud platí hypotéza H_{AB}, mluvíme o modelu dvojného třídění bez interakcí: $Y_{ijp} = \mu + \alpha_i + \beta_j + e_{ijp}$.

Abychom mohli parametry μ, α_i, β_j a γ_{ij} ($i = 1, ..., I$, $j = 1, ..., J$) jednoznačně odhadnout, zavedeme *reparametrizační rovnice:* $\sum_{i=1}^{I} \alpha_i = 0$, $\sum_{j=1}^{J} \beta_j = 0$, $\sum_{i=1}^{I} \gamma_{ij} = 0$ $(j = 1, ..., J)$ a $\sum_{j=1}^{J} \gamma_{ij} = 0$ $(i = 1, ..., I)$.

Na závěr ještě zavedeme pojmy *homoskedasticita* a *heteroskedasticita.* O homoskedasticitě mluvíme, jestliže rozptyly σ_{ij}^2 náhodných veličin Y_{ijp} jsou shodné pro každé $i = 1, ..., I$ a $j = 1, ..., J$. V opačném případě hovoříme o heteroskedasticitě.

Naším úkolem bude uvést testy pro testování hypotéz H_A, H_B a H_{AB}, které jsou odvozené bez předpokladu shody rozptylů (tj. bez předpokladu homoskedasticity) pro vyvážené třídění. Dále tyto testy porovnat, a to na základě simulací, při nichž budeme zjišťovat, zda testy dodržují hladinu významnosti a který z nich má největší sílu.

Kapitola 2

Značení

Uvažujme vyvážený model dvojného třídění s interakcemi: $Y_{ijp} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijp}$, $i = 1, ..., I$, $j = 1, ..., J$, $p = 1, ..., P$, $P \geq 2$ a testujme hypotézy H_A: $\alpha_i = 0$, H_B: $\beta_j = 0$ a H_{AB}: $\gamma_{ij} = 0$, $i = 1, ..., I$, $j = 1, ..., J$. Počet všech pozorování budeme označovat N, tj. $N = IJP$.

Nejprve definujeme

$$
\begin{array}{ll}
Y_{ij.} = \sum_{p=1}^{P} Y_{ijp}, & \bar{Y}_{ij.} = Y_{ij.}/P, \\
Y_{i..} = \sum_{j=1}^{J} \sum_{p=1}^{P} Y_{ijp}, & \bar{Y}_{i..} = Y_{i..}/JP, \\
Y_{.j.} = \sum_{i=1}^{I} \sum_{p=1}^{P} Y_{ijp}, & \bar{Y}_{.j.} = Y_{.j.}/IP, \\
Y_{...} = \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{p=1}^{P} Y_{ijp}, & \bar{Y}_{...} = Y_{...}/N
\end{array}
$$

a $S_{ij}^2 = \frac{1}{(P-1)} \sum_{p=1}^{P} (Y_{ijp} - \bar{Y}_{ij.})^2$ (výběrový rozptyl (i, j)-té populace).

Dále zavedeme následující značení: $\mathbf{1}_d$ je označením pro d-dimenzionální sloupcový vektor jedniček. Jednotkovou d-dimenzionální matici označíme $\mathbf{I}_d$. Pak $\mathbf{J}_d = \mathbf{1}_d \mathbf{1}_d'$ a $\mathbf{P}_d = \mathbf{I}_d - \frac{1}{d}\mathbf{J}_d$.

Dále je potřeba definovat binární operaci $\otimes$. Nechť $\mathbf{A}$ je matice typu (m, n) a $\mathbf{B}$ matice typu (p, q). Potom $\mathbf{A} \otimes \mathbf{B} = \begin{pmatrix} a_{11}\mathbf{B} & \cdots & a_{1n}\mathbf{B} \\ \vdots & \ddots & \vdots \\ a_{m1}\mathbf{B} & \cdots & a_{mn}\mathbf{B} \end{pmatrix}$. Výsledná matice $\mathbf{A} \otimes \mathbf{B}$ je typu (mp, nq).

Protože se celá tato práce týká testování hypotéz, zopakujeme si ještě některé pojmy z jeho teorie. Předpokládejme, že $\mathbf{X} = (X_1, X_2, ..., X_n)^T$ je náhodný výběr z rozdělení, které je prvkem rodiny $\{F_\theta; \theta \in \Theta\}$. Dále předpokládejme, že o parametru θ existují dvě navzájem si konkurující hypotézy. První z nich je hypotéza H_0: $\theta \in \Theta_0 (\subset \Theta)$, tu nazýváme nulová hypotéza. Druhou je tzv. alternativní hypotéza: H_1: $\theta \in \Theta_1 (= \Theta - \Theta_0)$. Naším úkolem je na základě náhodného výběru $\mathbf{X}$ rozhodnout, zda přijmout hypotézu H_0, nebo ji zamítnout ve prospěch alternativy H_1. Postup, při kterém zjišťujeme, zda hypotézu přijmout či zamítnout, se nazývá testování hypotézy. Naše rozhodnutí ale nemusí být vždy správné. Pokud zamítneme hypotézu H_0, ačkoli je správná, pak se dopustíme chyby prvního druhu. Když naopak tuto hypotézu nezamítneme, ačkoli není správná, dopustíme se chyby druhého druhu. Dále zavedeme pojem kritický obor testu. Jde o takovou množinu $W \subset R^n$, že je-li $\mathbf{X} \in W$, potom hypotézu H_0 zamítneme a přijmeme její alternativu H_1, v opačném případě hypotézu H_0 přijmeme. Kritický obor volíme tak, aby pravděpodobnost chyby prvního druhu byla menší nebo rovna zvolenému malému kladnému číslu α. Za α obvykle volíme 0.05. Hodnotu $sup_{\theta \in \Theta_0} P_\theta(\mathbf{X} \in W)$ nazýváme hladinou významnosti testu. Pro spojitá rozdělení F_θ můžeme vždy zvolit test, jehož hladina významnosti je právě rovna α. Silofunkcí testu budeme rozumět funkci proměnné θ, která udává pravděpodobnost, že zamítneme nulovou hypotézu za podmínky že neplatí, tj. $\beta(\theta) = P(\mathbf{X} \in W | \theta \in \Theta_1)$. Řekneme, test T_1 má větší sílu než test T_2, pro nějaké pevné $\tilde{\theta} \in \Theta_1$ na hladině významnosti α, jestliže oba testy mají hladinu významnosti α a silofunkce testu T_1 je v bodě $\tilde{\theta}$ větší než silofunkce testu T_2 v tomtéž bodě.

Kapitola 3

Zavedení testových statistik a testů

3.1 F-test

F-test je nejznámějším a nejpoužívanějším testem pro testování hypotéz o nulovosti řádkových a sloupcových efektů a o nulovosti interakcí. Předpokládá jak normalitu, tak i homoskedasticitu populací. Jeho odvození najdeme například v [1].

Uvažujme tedy model dvojného třídění s interakcemi a definujme součty čtverců S_A, S_B a S_{AB}:

$$S_A = JP \sum_{i=1}^{I} \bar{Y}_{i..}^2 - N\bar{Y}_{...}^2, \qquad S_B = IP \sum_{j=1}^{J} \bar{Y}_{.j.}^2 - N\bar{Y}_{...}^2,$$

$$S_{AB} = S_T - S_A - S_B - S_e,$$

kde $S_e = \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{p=1}^{P} Y_{ijp}^2 - P \sum_{i=1}^{I} \sum_{j=1}^{J} \bar{Y}_{ij.}^2$ je *reziduální součet čtverců a* $S_T = \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{p=1}^{P} Y_{ijp}^2 - N\bar{Y}_{...}^2$ *celkový součet čtverců.*

Potom hypotézu H_A testujeme pomocí testové statistiky

$$F_A = \frac{S_A}{(I-1)} \frac{1}{s^2},$$

kde $s^2 = \frac{S_e}{N-IJ}$.

Hypotézu H_A o nulovosti řádkových efektů zamítneme, pokud hodnota této statistiky je větší nebo rovna hodnotě $(1-\alpha)$-kvantilu F-rozdělení o $I-1$ a $N-IJ$ stupních volnosti (tj. pokud $F_A \geq F_{I-1,N-IJ}^{-1}(1-\alpha)$). Analogicky, a to sice za použití testových statistik

$$F_B = \frac{S_B}{(J-1)} \frac{1}{s^2}$$

a

$$F_{AB} = \frac{S_{AB}}{(I-1)(J-1)} \frac{1}{s^2},$$

testujeme hypotézy H_B a H_{AB}. Testové statistiky F_B a F_{AB} mají F-rozdělení o $J-1$ a $N-IJ$ resp. o $(I-1)(J-1)$ a $N-IJ$ stupních volnosti. A tedy hypotézu H_B resp. H_{AB} zamítneme, jestliže $F_B \geq F^{-1}_{J-1,N-IJ}(1-\alpha)$ resp. $F_{AB} \geq F^{-1}_{(I-1)(J-1),N-IJ}(1-\alpha)$.

3.2 Bathkeův test

Další test pro testování hypotéz H_A, H_B a H_{AB} odvodil v článku z roku 2003 Arne Bathke (viz [4]). Tento test je odvozen pro případ, kdy $I \to \infty$, a je určen pro vyvážené třídění. Nepředpokládá normalitu dat, klade však jistou podmínku na shodnost rozptylů. Vyžaduje totiž, aby $Var(e_{ijp}) = \sigma_j^2$. To znamená, že rozptyly mohou být různé pouze pro různé úrovně druhého faktoru. Arne Bathke ve svém článku navrhl testovat hypotézy pomocí stejných testových statistik, jaké využívá F-test. Jejich rozdělení, stejně jako návod jak testovat hypotézy H_A a H_{AB}, nalezneme v následující větě, uvedené a dokázané v [4] jako věta 1.

Věta 1 *Nechť* $Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk}$ *je model analýzy rozptylu dvojného třídění s faktory A a B,* $i = 1, ..., I$, $j = 1, ..., J$ *a počtem pozorování P v každé (i,j)-té podtřídě. Tedy* $\mathbf{Y} = (Y_{111}, ..., Y_{IJP})'$ *(v lexikografickém pořadí), kde* $Y_{111}, Y_{112}, ...$ *jsou nezávislé náhodné veličiny. Nechť* $E(Y_{ijp}) = \mu_{ij}$ *a* $Var(Y_{ijp}) = \sigma_j^2 < \infty$*. To znamená, že* $Var(\mathbf{Y}) = \mathbf{I}_I \otimes \mathbf{D} \otimes \mathbf{I}_P$*, kde* $\mathbf{D} = diag(\sigma_1^2, ..., \sigma_J^2)$*. Nechť dále* $\boldsymbol{\mu} = (\mu_{11}, ..., \mu_{IJ})'$*.*
Dále definujme

$$F_A = \frac{1}{I-1} Q_A / \frac{1}{IJ(P-1)} Q_\epsilon \ a$$

$$F_{AB} = \frac{1}{(I-1)(J-1)} Q_{AB} / \frac{1}{IJ(P-1)} Q_\epsilon,$$

kde $Q_A = \sum_{i=1}^{I} JP(\bar{Y}_{i..} - \bar{Y}_{...})^2$, $Q_{AB} = \sum_{i=1}^{I} \sum_{j=1}^{J} P(\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...})^2$ *a* $Q_\epsilon = \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{p=1}^{P} (Y_{ijp} - \bar{Y}_{ij.})^2$.
F_A *je testová statistika pro testování hypotézy* H_A *a* F_{AB} *je testová statistika pro testování hypotézy* H_{AB}*. Potom*

a) $(\boldsymbol{P}_I \otimes \boldsymbol{J}_J)\boldsymbol{\mu} = 0 \quad \Rightarrow \quad \sqrt{I}(F_A - 1) \xrightarrow{d} N(0, \tau_1^2)$ *pro* $I \to \infty$ *a J, P pevné, kde* $\tau_1^2 = 2(1 + \frac{1}{(P-1)}(\sum_{j=1}^{J} \sigma_j^2)^{-2} \sum_{j=1}^{J} \sigma_j^4)$,

b) $(\boldsymbol{P}_I \otimes \boldsymbol{P}_J)\boldsymbol{\mu} = 0 \quad \Rightarrow \quad \sqrt{I}(F_{AB} - 1) \xrightarrow{d} N(0, \tau_2^2)$ *pro* $I \to \infty, J, P$ *jsou pevné, kde* $\tau_2^2 = 2(\frac{1}{(J-1)^2} + (\frac{P}{(P-1)} - \frac{1}{(J-1)^2})(\sum_{j=1}^{J} \sigma_j^2)^{-2} \sum_{j=1}^{J} \sigma_j^4)$.

Poznámka 1: Neznámý parametr σ_j^2 ve znění věty 1 nahradíme jeho nestranným odhadem $\frac{1}{I}\sum_{i=1}^{I} S_{ij}^2$ a neznámý parametr σ_j^4 odhadem $(\frac{1}{I}\sum_{i=1}^{I} S_{ij}^2)^2$.

Poznámka 2: Výraz $(\mathbf{P}_I \otimes \mathbf{J}_J)\boldsymbol{\mu}= 0$ resp. $(\mathbf{P}_I \otimes \mathbf{P}_J)\boldsymbol{\mu}= 0$ si pro lepší představu, co znamená, můžeme přepsat z maticového tvaru do tvaru po složkách. Definice $\mathbf{P}_I$, $\mathbf{P}_J$, $\mathbf{J}_J$ a definici binární operace $\otimes$ nalezneme v kapitole 2. Jejich použitím a provedením několika úprav přepíšeme první z výrazů do tvaru $\sum_{j=1}^{J} \mu_{kj} - \frac{1}{I}\sum_{i=1}^{I}\sum_{j=1}^{J} \mu_{ij} = 0$, $k = 1, 2, ..., I$ a druhý do tvaru $\mu_{kl} - \frac{1}{J}\sum_{j=1}^{J} \mu_{kj} - \frac{1}{I}\sum_{i=1}^{I} \mu_{il} + \frac{1}{IJ}\sum_{i=1}^{I}\sum_{j=1}^{J} \mu_{ij} = 0$, kde $k = 1, 2, ..., I$ a $l = 1, 2, ..., J$. Pokud si nyní uvědomíme, že $\mu_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij}$, $i = 1, ..., I$, $j = 1, ..., J$, a využijeme reparametrizačních podmínek, dostaneme, že první rovnost platí, jestliže $\alpha_k = 0$, $k = 1, ..., I$, a druhá, jestliže $\gamma_{kl} = 0$, $k = 1, ..., I$, $l = 1, ..., J$. Předpoklad věty 1 v části a) můžeme tedy nahradit předpokladem platnosti hypotézy H_A a předpoklad v části b) předpokladem platnosti hypotézy H_{AB}.

Jestliže hypotéza H_A neplatí, F_A nabývá velkých hodnot (to samé platí pro hypotézu H_{AB}), proto podle věty 1 zamítneme hypotézu H_A v případě, že $\sqrt{I}(F_A - 1)/\sqrt{\tau_1^2} \geq u_{(1-\alpha)}$ a hypotézu H_{AB}, pokud $\sqrt{I}(F_{AB} - 1)/\sqrt{\tau_2^2} \geq u_{(1-\alpha)}$, kde $u_{(1-\alpha)}$ označuje $(1-\alpha)$-kvantil normovaného normálního rozdělení. Definice τ_1^2 a τ_2^2 nalezneme ve znění předchozí věty.

Z článku [4] není však zřejmé, jak testovat hypotézu H_B. Zkusíme tedy následující postup. Definujme Q_B a F_B analogicky jako Q_A a F_A a zamítněme hypotézu H_B, jestliže $\sqrt{I}(F_B - 1)/\sqrt{\tau_1^2} \geq u_{(1-\alpha)}$. Tato metoda ale nebude fungovat, což vyplyne z provedených simulací.

3.3 Wang-Akritasův test

V článku [2] z roku 2002 odvodili Lan Wang a Michael G. Akritas několik testů pro testování hypotéz H_A, H_B a H_{AB}, a to jak pro vyvážený, tak pro nevyvážený případ. Tyto testy nepředpokládají ani shodu rozptylů ani normalitu a jsou odvozeny pro případ $I \to \infty$ a J pevné.

My se budeme soustředit na testování hypotéz H_A a H_{AB} pro vyvážené třídění. Zabývejme se nejprve testováním hypotézy H_A. Pro její testování budeme používat testovou statistiku

$$Q_A = \frac{MST_A}{MSE},$$

kde

$$MST_A = \frac{JP}{I-1}\sum_{i=1}^{I}(\bar{Y}_{i..} - \bar{Y}_{...})^2,$$

$$MSE = \frac{1}{N - IJ} \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{p=1}^{P} (Y_{ijp} - \bar{Y}_{ij.})^2.$$

Analogicky test hypotézy H_{AB} je založen na statistice

$$Q_{AB} = \frac{MST_{AB}}{MSE},$$

kde MST_{AB} je definováno vztahem:

$$MST_{AB} = \frac{P}{(I-1)(J-1)} \sum_{i=1}^{I} \sum_{j=1}^{J} (\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...})^2,$$

Poznamenejme, že testové statistiky Q_A a Q_{AB} se shodují se statistikami používanými při F-testu. Jaké Wang a Akritas použili rozdělení těchto statistik se dozvíme z následující věty, která je v článku [2] uvedena jako Věta 3.3 spolu se svým důkazem.

Věta 2 *Nechť Y_{ijp}, $i = 1, ..., I$, $j = 1, ..., J$, $p = 1, ..., P$, $P \geq 2$, jsou nezávislé náhodné veličiny s $0 < Var(Y_{ijp}) = \sigma_{ij}^2 < \infty$. Předpokládejme dále, že v každé (i,j)-té podtřídě jsou Y_{ijp} nezávislé a stejně rozdělené. Nechť navíc při $I \to \infty$ je*

$$\frac{1}{IJ} \sum_{i=1}^{I} \sum_{j=1}^{J} \sigma_{ij}^2 \to \sigma^2, \quad \frac{1}{IJ} \sum_{i=1}^{I} \sum_{j=1}^{J} \sigma_{ij}^4 \to \ s^4, \quad \frac{1}{IJ} \sum_{i=1}^{I} \sum_{j_1 \neq j_2}^{J} \sigma_{ij_1}^2 \sigma_{ij_2}^2 \to \gamma^4$$

a předpokládejme, že všechny limity jsou kladné.
Jestliže pro nějaké $\delta > 0$ je $\operatorname{limsup} \frac{1}{IJ} \sum_{i=1}^{I} (\sum_{j=1}^{J} E \,|\, Y_{ij1} - \mu|^{2+\delta})^2 < \infty$, potom za předpokladu platnosti hypotézy H_A

$$I^{1/2}(Q_A - 1) \xrightarrow{d} N\left(0, \frac{2Ps^4}{J(P-1)\sigma^4} + \frac{2\gamma^4}{J\sigma^4}\right), \ pro\ I \to \infty,$$

a za platnosti hypotézy H_{AB}

$$I^{1/2}(Q_{AB} - 1) \xrightarrow{d} N\left(0, \frac{2Ps^4}{J(P-1)\sigma^4} + \frac{2\gamma^4}{J(J-1)^2\sigma^4}\right) \ pro\ I \to \infty.$$

Poznámka: Protože rozptyly populací σ_{ij}^2, $i = 1, ..., I$, $j = 1, ..., J$ jsou neznámé parametry, nahradíme je ve větě 2 jejich odhady S_{ij}^2, $i = 1, ..., I$, $j = 1, ..., J$. Z téhož důvodu namísto druhé mocniny rozptylů σ_{ij}^4,$i = 1, ..., I$, $j = 1, ..., J$ použijeme odhady $(S_{ij}^2)^2$, $i = 1, ..., I$, $j = 1, ..., J$.

Věta 2 nám dává návod, jak testovat hypotézy H_A a H_{AB}. Za platnosti hypotézy H_A víme, že platí $E(MSE) = E(MST_A)$, to tedy znamená, že za platnosti hypotézy H_A je podíl MSE/MST_A roven jedné. Pokud však hypotéza H_A neplatí, je $MST_A > MSE$. Proto, pokud jsou splněny předpoklady předchozí věty, hypotézu H_A o nulovosti řádkových efektů zamítneme, jestliže $I^{1/2}(Q_A - 1)/\sqrt{\frac{2Ps^4}{J(P-1)\sigma^4} + \frac{2\gamma^4}{J\sigma^4}} \geq u_{(1-\alpha)}$, kde $u_{(1-\alpha)}$ je $(1-\alpha)$-kvantil normovaného normálního rozdělení. Podobně jako hypotézu H_A testujeme za použití předchozí věty i hypotézu H_{AB}.

Na závěr si ještě popíšeme, jak budeme postupovat při testování hypotézy H_B. Testovou statistiku Q_B definujeme analogicky jako Q_A a budeme jí porovnávat s $(1-\alpha)$-kvantilem chí kvadrát rozdělení s jedním stupněm volnosti, neboť, jak je uvedeno v článku [2], má testová statistika Q_B asymptoticky chí kvadrát rozdělení. Hypotézu H_B tedy zamítneme, jestliže $Q_B \geq \chi^2_{1-\alpha,1}$.

3.4 Krutchkoffův test

Testy pro hypotézy H_A, H_B a H_{AB} nalezneme taktéž v článku Krutchkoff (1989) (viz [5]), i když pouze intuitivně odvozené. Jejich předpokladem je normalita populací, ale ne shoda rozptylů. Nechť je $\hat{\sigma}^2_{ij}$ bodový odhad σ^2_{ij}. Nejčastěji za $\hat{\sigma}^2_{ij}$ volíme výběrový rozptyl (i,j)-té populace, tedy $\hat{\sigma}^2_{ij} = S^2_{ij} = \frac{1}{(P-1)} \sum_{p=1}^{P} (Y_{ijp} - \bar{Y}_{ij.})^2$. Dále definujme:

$$\bar{Y}_i^* = \sum_{j=1}^{J} (\bar{Y}_{ij.}/\hat{\sigma}^2_{ij}) / \sum_{j=1}^{J} (1/\hat{\sigma}^2_{ij}) \qquad \text{(vážený řádkový průměr s váhami } 1/\hat{\sigma}^2_{ij}\text{)},$$

$$\bar{Y}_j^* = \sum_{i=1}^{I} (\bar{Y}_{ij.}/\hat{\sigma}^2_{ij}) / \sum_{i=1}^{I} (1/\hat{\sigma}^2_{ij}) \qquad \text{(vážený sloupcový průměr s váhami } 1/\hat{\sigma}^2_{ij}\text{)},$$

$$\bar{Y}^* = \sum_{i=1}^{I} \sum_{j=1}^{J} (\bar{Y}_{ij.}/\hat{\sigma}^2_{ij}) / \sum_{i=1}^{I} \sum_{j=1}^{J} (1/\hat{\sigma}^2_{ij}) \qquad \text{(vážený celkový průměr s váhami } 1/\hat{\sigma}^2_{ij}\text{)}.$$

Testy hypotéz jsou založené na testových statistikách definovaných následujícími předpisy. Pro testování nulovosti sloupcových efektů použijeme statistiku

$$KC = \sum_{j=1}^{J} [(\bar{Y}_j^* - \bar{Y}^*)^2 \sum_{i=1}^{I} (P/\hat{\sigma}^2_{ij})]/(J-1),$$

pro testování nulovosti řádkových efektů statistiku

$$KR = \sum_{i=1}^{I} [(\bar{Y}_i^* - \bar{Y}^*)^2 \sum_{j=1}^{J} (P/\hat{\sigma}_{ij}^2)]/(I-1)$$

a konečně pro testování nulovosti interakcí testovou statistiku

$$KI = \sum_{i=1}^{I} \sum_{j=1}^{J} [(\bar{Y}_{ij.} - \bar{Y}_i^* - \bar{Y}_j^* + \bar{Y}^*)^2 P/\hat{\sigma}_{ij}^2] / [(I-1)(J-1)] \, .$$

Krutchkoff se však ve svém článku nesnažil najít rozdělení těchto statistik. Kritické hodnoty pro tyto statistiky nalézal pomocí simulací, a to následovně. Nejprve pro každé $i = 1, ..., I$, $j = 1, ..., J$ a $p = 1, ..., P$ vygenerujeme nezávislé stejně rozdělené náhodné veličiny Y_{ijp} z normovaného normálního rozdělení a potom pro tyto náhodné veličiny spočítáme statistiky KC, KR a KI. Tento postup k-krát zopakujeme. Jako kritické hodnoty vezmeme $KC\,[(1-\alpha)k]$, $KR\,[(1-\alpha)k]$ a $KI\,[(1-\alpha)k]$, což jsou $(1-\alpha)k$-té pořádkové statistiky statistik KC, KR a KI. Za k obvykle volíme nějaké velké číslo. My použijeme $k = 100000$.

Hypotézu H_A zamítneme za podmínky, že hodnota testové statistiky KR bude větší nebo rovna kritické hodnotě $KR[(1-\alpha)k]$, tedy pokud $KR \geq KR[(1-\alpha)k]$. Dále hypotézu H_B zamítneme, jestliže $KC \geq KC[(1-\alpha)k]$ a hypotézu H_{AB} v případě, že $KI \geq KI[(1-\alpha)k]$.

3.5 Brown-Forsytheův test

Dalším testem, který testuje hypotézy H_A, H_B a H_{AB} je Brown-Forsytheův test odvozený v článku Brown-Forsythe (1974) (viz [3]). Jeho předpokladem je opět normalita populací, ne však jejich homoskedasticita.

K testování řádkových efektů použijeme statistiku

$$F_A^* = \frac{PJ \sum_{i=1}^{I} (\bar{Y}_{i..} - \bar{Y}_{...})^2}{(I - 1/JI) \sum_{i=1}^{I} \sum_{j=1}^{J} S_{ij}^2},$$

která má přibližně F-rozdělení s $I-1$ a f stupni volnosti, kde f je definováno rovnicí

$$\frac{1}{f} = \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{c_{ij}^2}{(P-1)}$$

a

$$c_{ij} = \frac{S_{ij}^2}{\sum_{i=1}^{I} \sum_{j=1}^{J} S_{ij}^2} \quad .$$

To tedy znamená, že hypotézu H_A o nulovosti řádkových efektů zamítneme na hladině α, pokud $F_A^* \geq F_{I-1,f}^{-1}(1-\alpha)$, tedy v případě, kdy hodnota testové statistiky F_A^* je větší nebo rovna hodnotě kvantilu F-rozdělení s $I-1$ a f stupni volnosti v bodě $1-\alpha$.

Hypotézy H_B a H_{AB} o nulovosti sloupcových efektů a interakcí testujeme stejně jako hypotézu H_A. Hypotézu H_B za použití testové statistiky

$$F_B^* = \frac{PI\sum_{j=1}^{J}(\bar{Y}_{.j.} - \bar{Y}_{...})^2}{(J-1/JI)\sum_{i=1}^{I}\sum_{j=1}^{J}S_{ij}^2},$$

která má F-rozdělení s $J-1$ a f stupni volnosti, a hypotézu H_{AB} pomocí statistiky

$$F_{AB}^* = \frac{P\sum_{i=1}^{I}\sum_{j=1}^{J}(\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}...)^2}{((J-1)(I-1)/JI)\sum_{i=1}^{I}\sum_{j=1}^{J}S_{ij}^2},$$

jež má F-rozdělení s $(I-1)(J-1)$ a f stupni volnosti.

Na závěr uveďme, že testové statistiky používané Brown-Forsytheovým testem jsou v tomto případě opět shodné se statistikami používanými při F-testu.

3.6 Chenův test

Chenův test je dalším z testů, který předpokládá normalitu populací, ale nevyžaduje shodnost rozptylů. Jeho odvození nalezneme v článku [6].

Nechť $P > 2$. Označme $p_0 = P - 1$ a spočítejme

$$\bar{Y}_{ij}(p_0) = \sum_{p=1}^{p_0} Y_{ijp}/p_0,$$

$$S_{ij}^2(p_0) = \sum_{p=1}^{p_0}(Y_{ijp} - \bar{Y}_{ij}(p_0))^2/(p_0-1).$$

Potom váhy jednotlivých pozorování v podtřídách (i,j) jsou

$$U_{ij} = \frac{1}{P} + \frac{1}{P}\sqrt{\frac{1}{P-1}\left[S_{[p]}^2/S_{ij}^2(p_0) - 1\right]},$$

$$V_{ij} = \frac{1}{P} - \frac{1}{P}\sqrt{(P-1)\left[S_{[p]}^2/S_{ij}^2(p_0) - 1\right]},$$

kde $S_{[p]}^2$ označuje maximální hodnotu z $S_{ij}^2(p_0)$, $i = 1,...,I$, $j = 1,...,J$.

Nyní definujme pro každou (i,j)-tou podtřídu vážený výběrový průměr

$$\tilde{Y}_{ij.} = \sum_{p=1}^{P} W_{ijp} Y_{ijp},$$

kde

$$W_{ijp} = \begin{cases} U_{ij}, & 1 \leq p \leq P-1 \\ V_{ij}, & p = P \end{cases}.$$

Spočtěme ještě

$$\tilde{Y}_{i..} = \frac{1}{J}\sum_{j=1}^{J} \tilde{Y}_{ij.}, \quad \tilde{Y}_{.j.} = \frac{1}{I}\sum_{i=1}^{I} \tilde{Y}_{ij.}, \quad \tilde{Y}_{...} = \frac{1}{IJ}\sum_{i=1}^{I}\sum_{j=1}^{J} \tilde{Y}_{ij.}.$$

Hypotézu H_A budeme testovat pomocí testové statistiky

$$\tilde{F}_1^1 = J \sum_{i=1}^{I} \left(\frac{\tilde{Y}_{i..} - \tilde{Y}_{...}}{S_{[p]}/\sqrt{P}} \right)^2$$

a hypotézy H_B a H_{AB} pomocí statistik $\tilde{F}_2^1$ a $\tilde{F}_3^1$, které jsou definovány vzorci

$$\tilde{F}_2^1 = I \sum_{j=1}^{J} \left(\frac{\tilde{Y}_{.j.} - \tilde{Y}_{...}}{S_{[p]}/\sqrt{P}} \right)^2$$

a

$$\tilde{F}_3^1 = \sum_{i=1}^{I}\sum_{j=1}^{J} \left(\frac{\tilde{Y}_{ij.} - \tilde{Y}_{i..} - \tilde{Y}_{.j.} + \tilde{Y}_{...}}{S_{[p]}/\sqrt{P}} \right)^2.$$

Z [6] víme, že lze dokázat, že

$$t_{ij} = \frac{\tilde{Y}_{ij.} - (\mu + \alpha_i + \beta_j + \gamma_{ij})}{S_{[p]}/\sqrt{P}},$$

$i = 1, ..., I$, $j = 1, ..., J$ má Studentovo rozdělení s $p_0 - 1$ stupni volnosti, tedy že $t_{ij} \sim t_{p_0-1}$, a pokud si uvědomíme, že $p_0 - 1 = P - 2$, potom $t_{ij} \sim t_{P-2}$.

Nechť

$$\bar{t}_{i.} = \frac{1}{J}\sum_{j=1}^{J} t_{ij}, \quad \bar{t}_{.j} = \frac{1}{I}\sum_{i=1}^{I} t_{ij}, \quad \bar{t}_{..} = \frac{1}{IJ}\sum_{i=1}^{I}\sum_{j=1}^{J} t_{ij},$$

pak můžeme testové statistiky přepsat do tvaru

$$\tilde{F}_1^1 = J \sum_{i=1}^{I} \left(\bar{t}_{i.} - \bar{t}_{..} + \frac{\alpha_i}{S_{[p]}/\sqrt{P}} \right)^2,$$

$$\tilde{F}_2^1 = I \sum_{j=1}^{J} \left(\bar{t}_{.j} - \bar{t}_{..} + \frac{\beta_j}{S_{[p]}/\sqrt{P}} \right)^2$$

a

$$\tilde{F}_3^1 = \sum_{i=1}^{I} \sum_{j=1}^{J} \left(t_{ij} - \bar{t}_{i.} - \bar{t}_{.j} + \bar{t}_{..} + \frac{\gamma_{ij}}{S_{[p]}/\sqrt{P}} \right)^2 .$$

A za platnosti hypotéz H_A, H_B a H_{AB} víme, že platí

$$\tilde{F}_1^1 \sim J \sum_{i=1}^{I} (\bar{t}_{i.} - \bar{t}_{..})^2 = Q^1,$$

$$\tilde{F}_2^1 \sim I \sum_{j=1}^{J} (\bar{t}_{.j} - \bar{t}_{..})^2 = Q^2,$$

$$\tilde{F}_3^1 \sim \sum_{i=1}^{I} \sum_{j=1}^{J} (t_{ij} - \bar{t}_{i.} - \bar{t}_{.j} + \bar{t}_{..})^2 = Q^3.$$

Kritické hodnoty statistik $\tilde{F}_1^1$, $\tilde{F}_2^1$ a $\tilde{F}_3^1$ musíme vypočítat pomocí simulací, a to tak, že k-krát vygenerujeme náhodné veličiny t_{ij} a pokaždé spočítáme statistiky Q^1, Q^2 a Q^3. Jako kritické hodnoty testů pak vezmeme $(1-\alpha)k$-té pořádkové statistiky statistik Q^1, Q^2 a Q^3 a označíme je $Q^1_{(1-\alpha)k}$, $Q^2_{(1-\alpha)k}$ a $Q^3_{(1-\alpha)k}$ (my opět zvolíme $k = 100000$). Hypotézu H_A zamítneme, jestliže testová statistika $\tilde{F}_1^1$ je větší nebo rovna hodnotě $Q^1_{(1-\alpha)k}$, tedy pokud $\tilde{F}_1^1 \geq Q^1_{(1-\alpha)k}$. Kritické obory hypotéz H_B a H_{AB} určíme analogicky.

Tabulka 3.1: Shrnutí předpokladů testů; poslední sloupec obsahuje údaj o tom, zda lze příslušný test modifikovat a použít i pro nevyvážené třídění.

test	normalita	homoskedasticita	$I \to \infty$	nevyváženost
F-test	ano	ano	ne	ano
Bathkeův test	ne	částečně [1]	ano	ne
Wang-Akritasův test	ne	ne	ano	ano
Krutchkoffův test	ano	ne	ne	ano
Brown-Forsytheův test	ano	ne	ne	ne
Chenův-test	ano	ne	ne	ne

[1] Rozptyly se mohou lišit pro různé úrovně druhého faktoru ($Var(e_{ijp}) = \sigma_j^2$).

Kapitola 4

Nastavení simulací

Naším úkolem je nyní porovnat uvažované testy hypotéz H_A: $\alpha_i = 0$, H_B: $\beta_j = 0$ a H_{AB}: $\gamma_{ij} = 0$, $i = 1, ..., I$, $j = 1, ..., J$. Za tímto účelem budeme zkoumat jejich dodržování hladiny významnosti $\alpha = 0.05$ a dále nás bude také zajímat, jakou mají tyto testy sílu za platnosti různých alternativ. Hladinu významnosti α testů i jejich sílu za platnosti alternativ budeme odhadovat tak, že 1000-krát vygenerujeme náhodné veličiny Y_{ijp} z normálního rozdělení se střední hodnotou μ_{ij} a rozptylem σ^2_{ij}, $i = 1, ..., I$, $j = 1, ..., J$. Pro každé vygenerované náhodné veličiny provedeme testy hypotéz H_A, H_B a H_{AB}, jež tyto hypotézy přijmou, nebo zamítnou. Za odhady hladiny významnosti (resp. síly testů) vezmeme relativní četnosti zamítnutí hypotéz jednotlivými testy. Ke generování náhodných veličin použijeme statistický software **R 2.6.0**.

V celé simulaci můžeme nastavit mnoho parametrů. Proto budeme uvažovat jen některé případy a parametry volit tak, abychom mohli co nejlépe určit chování testů. Prvním z parametrů, který můžeme zvolit, je počet pozorování P. Uvažujme tři případy, a to P=5, P=10 a P=50. Počet úrovní prvního faktoru I volme rovno dvěma, deseti a padesáti a počet úrovní druhého faktoru J rovno třem a deseti. Další z parametrů, který můžeme volit, je rozptyl populací σ^2_{ij}, $i = 1, ..., I$, $j = 1, ..., J$. Kromě homoskedastického případu budeme uvažovat i případy heteroskedastické. Pro homoskedastický případ položíme všechny rozptyly rovny jedné (tj. σ^2_{ij}=1, $i = 1, ..., I$, $j = 1, ..., J$). Bez předpokladu homoskedasticity budeme uvažovat celkem 3 možnosti. První z nich je zvolit σ^2_{11}=5 a ostatní rozptyly položit jedné. Druhou možností je položit $\sigma^2_{ij} = 5$ pro $i = 1, ..., I$ a $j = 1$ a $\sigma^2_{ij} = 1$ pro $i = 1, ..., I$ a $j = 2, ..., J$. Jako třetí variantu vezmeme $\sigma^2_{ij} = 5$ pro $j = 1, ..., J$ a $i = 1$ a ostatní rozptyly necháme opět rovny jedné. Tyto tři heteroskedastické případy označme v pořadí, v jakém jsme je zde uvedli, jako heteroskedastický případ 1, heteroskedastický případ 2 a heteroskedastický případ 3. Posledním parametrem, který je potřeba při simulacích nastavit, je střední hodnota normálního rozdělení. Při zkoumání dodržování hladiny významnosti testů budeme střední hodnotu μ_{ij} náhodných veličin Y_{ijp} bez újmy na obecnosti vždy volit rovnu nule (tedy μ_{ij}=0, $i = 1, ..., I$, $j = 1, ..., J$). Při zjišťování síly testů za platnosti alternativ budeme bez újmy na obecnosti volit rovnu nule obecnou střední hodnotu μ a uvažovat budeme následující dvě alternativy. Jako alternativu 1 označíme alternativu H_A^1 hypotézy H_A. Budeme předpokládat, že ne všechny řádkové efekty $\alpha_1, ..., \alpha_I$ jsou nulové. Jako nenulový zvolíme parametr α_1. Alternativu H_{AB}^1 hypotézy H_{AB}, kdy parametr γ_{11} bude nenulový a ostatní parametry nulové, pojmenujeme alternativa 2. Bez újmy na obecnosti můžeme předpokládat, že parametry α_1 a γ_{11} jsou kladné. Jejich hodnoty budeme volit tak, aby neplatná hypotéza byla F-testem zamítnuta přibližně v padesáti procentech případů.

Odhady hladiny významnosti a síly testů budeme ukládat do tabulek. Tak lépe uvidíme rozdíly mezi jednotlivými testy a budeme je moci lépe porovnat. Z důvodu velkého množství simulovaných případů není možné v této práci uvést všechny tabulky s výsledky. Ty nalezneme spolu se zdrojovým kódem simulací na přiloženém CD.

Kapitola 5

Výsledky simulací

5.1 Dodržování hladiny významnosti

5.1.1 Za předpokladu homoskedasticity

V homoskedastických případech dodržují hladinu významnosti téměř všechny uvedené testy. Výjimkou jsou pouze Bathkeův test a Wang-Akritasův test, které jsou odvozeny pro $I \to \infty$, a proto při nízkém počtu úrovní prvního faktoru mají s dodržováním hladiny významnosti problém. Pro $I = 50$ při testování hypotéz H_A a H_{AB} oba dva zmíněné testy hladinu významnosti již většinou dodržují. Není tomu tak však při testování hypotézy H_B. Bathkeův test hladinu významnosti při testování hypotézy H_B nadhodnocuje (až na případ $I = 2$ a $J = 10$, kde je hladina významnosti velmi podhodnocena), naopak Wang-Akritasův test ji výrazně podhodnocuje, a to především pro větší hodnoty J.

Tabulka 5.1: Dodržování hladiny významnosti $\alpha = 0.05$ v homoskedastickém případě, $\sigma_{ij}^2 = 1$, $i = 1, ..., I$, $j = 1, ..., J$, $I = 2$, $J = 3$

Test	H_A	H_B	H_{AB}	H_A	H_B	H_{AB}	H_A	H_B	H_{AB}
	$P = 5$			$P = 10$			$P = 50$		
F-test	0.050	0.048	0.050	0.048	0.042	0.041	0.052	0.053	0.065
Bathke	0.123	0.083	0.132	0.098	0.067	0.113	0.109	0.074	0.118
Wang-Akritas	0.117	0.035	0.118	0.095	0.023	0.109	0.109	0.022	0.118
Krutchkoff	0.053	0.040	0.041	0.052	0.047	0.045	0.051	0.053	0.066
Brown-Forsythe	0.048	0.045	0.042	0.046	0.040	0.041	0.052	0.053	0.065
Chen	0.043	0.051	0.049	0.048	0.067	0.050	0.046	0.051	0.055

Tabulka 5.2: Dodržování hladiny významnosti $\alpha = 0.05$ v homoskedastickém případě, $\sigma_{ij}^2 = 1$, $i = 1, ..., I$, $j = 1, ..., J$, $I = 10$, $J = 3$

Test	H_A	H_B	H_{AB}	H_A	H_B	H_{AB}	H_A	H_B	H_{AB}
	$P = 5$			$P = 10$			$P = 50$		
F-test	0.044	0.047	0.049	0.050	0.047	0.062	0.053	0.047	0.059
Bathke	0.080	0.163	0.077	0.072	0.175	0.088	0.078	0.188	0.080
Wang-Akritas	0.067	0.022	0.060	0.069	0.019	0.078	0.077	0.019	0.078
Krutchkoff	0.056	0.042	0.052	0.051	0.039	0.059	0.057	0.046	0.065
Brown-Forsythe	0.041	0.046	0.039	0.050	0.047	0.061	0.053	0.047	0.059
Chen	0.060	0.065	0.052	0.049	0.064	0.051	0.051	0.054	0.054

Tabulka 5.3: Dodržování hladiny významnosti $\alpha = 0.05$ v homoskedastickém případě, $\sigma_{ij}^2 = 1$, $i = 1, ..., I$, $j = 1, ..., J$, $I = 50$, $J = 3$

Test	H_A	H_B	H_{AB}	H_A	H_B	H_{AB}	H_A	H_B	H_{AB}
	$P = 5$			$P = 10$			$P = 50$		
F-test	0.041	0.041	0.046	0.056	0.048	0.046	0.047	0.055	0.036
Bathke	0.053	0.268	0.070	0.071	0.270	0.055	0.059	0.287	0.041
Wang-Akritas	0.041	0.022	0.036	0.065	0.023	0.045	0.059	0.029	0.039
Krutchkoff	0.055	0.043	0.048	0.071	0.023	0.050	0.051	0.057	0.041
Brown-Forsythe	0.039	0.041	0.037	0.055	0.048	0.046	0.047	0.055	0.036
Chen	0.037	0.043	0.037	0.057	0.056	0.051	0.055	0.062	0.038

Tabulka 5.4: Dodržování hladiny významnosti $\alpha = 0.05$ v homoskedastickém případě, $\sigma_{ij}^2 = 1$, $i = 1, ..., I$, $j = 1, ..., J$, $J = 10$, $P = 10$

Test	H_A	H_B	H_{AB}	H_A	H_B	H_{AB}	H_A	H_B	H_{AB}
	$I = 2$			$I = 10$			$I = 50$		
Bathke	0.106	0.007	0.110	0.071	0.093	0.082	0.064	0.218	0.053
Wang-Akritas	0.106	0.001	0.105	0.071	0.000	0.069	0.064	0.000	0.038

5.1.2 Bez předpokladu homoskedasticity

Heteroskedastický případ 1 ($\sigma_{11}^2 = 5$, zbylé rozptyly rovny jedné)

Bez předpokladu homoskedasticity v případě, že jedna z populací má jiný rozptyl než ostatní, dodržuje vždy hladinu významnosti jen Krutchkoffův test a Chenův test. F-test zde selhává. Dodržuje hladinu významnosti pouze v těchto případech: pro velký počet úrovní prvního faktoru při testování hypotézy H_B a pro velký počet úrovní druhého faktoru při testování hypotézy H_A. Dále si povšimněme, že se zvyšujícím se počtem pozorování P se chování Brown-Forsytheova testu velmi rychle přibližuje chování F-testu. Toto přibližování (v dodržování hladiny významnosti) můžeme pozorovat i se zvětšujícím se I a J, není však tak rychlé. Wang-Akritasův test dodržuje hladinu významnosti, stejně jako v homoskedastickém případě, pouze při testování hypotéz H_A a H_{AB} pro velký počet úrovní prvního faktoru. Při testování hypotézy H_B hladinu významnosti výrazně podhodnocuje. Bathkeův test nedodržel hladinu významnosti v žádném ze simulovaných případů, ale pro J=10 je vidět, že s rostoucím I se dodržování hladiny významnosti zlepšuje v případě testování hypotéz H_A a H_{AB}. Nedodržování hladiny významnosti Bathkeova testu pro malé J i při $I \to \infty$ je zde pravděpodobně dáno porušením předpokladu kladeného na rozptyly populací.

Tabulka 5.5: Dodržování hladiny významnosti $\alpha = 0.05$ v heteroskedastickém případě 1, $\sigma_{11}^2 = 5$, zbylé rozptyly rovny jedné, $I = 2$, $J = 3$

Test	H_A	H_B	H_{AB}	H_A	H_B	H_{AB}	H_A	H_B	H_{AB}
	$P = 5$			$P = 10$			$P = 50$		
F-test	0.089	0.102	0.097	0.078	0.096	0.090	0.062	0.081	0.087
Bathke	0.145	0.135	0.137	0.139	0.128	0.122	0.125	0.109	0.123
Wang-Akritas	0.130	0.088	0.112	0.110	0.069	0.093	0.102	0.044	0.088
Krutchkoff	0.041	0.041	0.074	0.049	0.052	0.068	0.044	0.045	0.073
Brown-Forsythe	0.059	0.075	0.067	0.063	0.074	0.070	0.060	0.076	0.077
Chen	0.054	0.062	0.063	0.048	0.049	0.048	0.049	0.047	0.044

Tabulka 5.6: Dodržování hladiny významnosti $\alpha = 0.05$ v heteroskedastickém případě 1, $\sigma_{11}^2 = 5$, zbylé rozptyly rovny jedné, $I = 10$, $J = 3$

Test	H_A	H_B	H_{AB}	H_A	H_B	H_{AB}	H_A	H_B	H_{AB}
	$P = 5$			$P = 10$			$P = 50$		
F-test	0.123	0.067	0.150	0.108	0.043	0.145	0.113	0.064	0.151
Bathke	0.156	0.192	0.167	0.134	0.168	0.158	0.139	0.174	0.158
Wang-Akritas	0.111	0.038	0.107	0.078	0.016	0.088	0.079	0.029	0.081
Krutchkoff	0.050	0.049	0.047	0.046	0.047	0.053	0.058	0.052	0.065
Brown-Forsythe	0.094	0.053	0.105	0.090	0.038	0.112	0.111	0.059	0.145
Chen	0.057	0.060	0.045	0.055	0.048	0.036	0.049	0.055	0.039

Tabulka 5.7: Dodržování hladiny významnosti $\alpha = 0.05$ v heteroskedastickém případě 1, $\sigma_{11}^2 = 5$, zbylé rozptyly rovny jedné, $I = 50$, $J = 3$

Test	H_A	H_B	H_{AB}	H_A	H_B	H_{AB}	H_A	H_B	H_{AB}
	$P = 5$			$P = 10$			$P = 50$		
F-test	0.100	0.061	0.123	0.085	0.053	0.110	0.088	0.049	0.136
Bathke	0.116	0.256	0.136	0.097	0.258	0.122	0.099	0.243	0.144
Wang-Akritas	0.073	0.026	0.084	0.061	0.025	0.058	0.064	0.025	0.067
Krutchkoff	0.054	0.049	0.045	0.040	0.052	0.052	0.051	0.051	0.045
Brown-Forsythe	0.085	0.058	0.095	0.077	0.052	0.096	0.087	0.048	0.130
Chen	0.053	0.061	0.055	0.053	0.045	0.050	0.047	0.044	0.050

Tabulka 5.8: Dodržování hladiny významnosti $\alpha = 0.05$ v heteroskedastickém případě 1, $\sigma_{11}^2 = 5$, zbylé rozptyly rovny jedné , $I = 2$, $J = 10$

Test	H_A	H_B	H_{AB}	H_A	H_B	H_{AB}	H_A	H_B	H_{AB}
	$P = 5$			$P = 10$			$P = 50$		
F-test	0.066	0.139	0.131	0.046	0.135	0.122	0.054	0.123	0.133
Bathke	0.117	0.080	0.143	0.107	0.065	0.115	0.112	0.064	0.114
Wang-Akritas	0.104	0.039	0.121	0.091	0.024	0.093	0.096	0.022	0.085
Brown-Forsythe	0.051	0.098	0.100	0.038	0.102	0.095	0.052	0.118	0.120

Tabulka 5.9: Dodržování hladiny významnosti $\alpha = 0.05$ v heteroskedastickém případě 1, $\sigma_{11}^2 = 5$, zbylé rozptyly rovny jedné, $I = 50$, $J = 10$

Test	H_A	H_B	H_{AB}	H_A	H_B	H_{AB}	H_A	H_B	H_{AB}
	$P = 5$			$P = 10$			$P = 50$		
F-test	0.057	0.052	0.103	0.053	0.044	0.097	0.058	0.062	0.092
Bathke	0.072	0.228	0.109	0.068	0.218	0.102	0.075	0.210	0.095
Wang-Akritas	0.062	0.000	0.060	0.054	0.000	0.058	0.062	0.001	0.060
Brown-Forsythe	0.056	0.052	0.087	0.052	0.043	0.089	0.058	0.062	0.091

Heteroskedastický případ 2
($\sigma_{i1}^2 = 5$, pro $i = 1, ..., I$ a $\sigma_{ij}^2 = 1$, pro $i = 1, ..., I$, $j = 2, ..., J$)

V tomto heteroskedastickém případě se můžeme spolehnout na dodržení hladiny významnosti pro jakoukoli volbu parametrů I, J a P pouze u Chenova testu. Krutchkoffův test je spolehlivý téměř ve všech uvažovaných případech, ale jak je vidět z provedených simulací, nadhodnocuje hladinu významnosti v případě testování hypotézy H_{AB} pro $I = 50$, $J = 3$ a $P = 5$ a v případě testování hypotézy H_A pro $I = 50$, $J = 3$, $P = 10$ ji mírně podhodnocuje. F-test neselhává jen při testování nulovosti řádkových efektů. I zde si můžeme povšimnout, že se zvětšujícím se počtem pozorování se chování Brown-Forsytheova testu výrazně blíží chování F-testu. To, co bylo řečeno o dodržování hladiny významnosti Wang-Akritasova testu v předchozích případech, platí i v tomto. Zde však navíc zaznamenáváme dodržování hladiny významnosti při testování sloupcových efektů pro $J = 3$, a to již od $I = 10$. Bathkeův test v heteroskedastickém případě 2 dodržuje hladinu významnosti při testování hypotéz H_A a H_{AB} pro vysoký počet úrovní prvního faktoru.

Tabulka 5.10: Dodržování hladiny významnosti $\alpha = 0.05$ v heteroskedastickém případě 2, $\sigma_{i1}^2 = 5$, pro $i = 1, ..., I$ a $\sigma_{ij}^2 = 1$, pro $i = 1, ..., I$, $j = 2, ..., J$, $I = 2$, $J = 3$

Test	H_A	H_B	H_{AB}	H_A	H_B	H_{AB}	H_A	H_B	H_{AB}
	$P = 5$			$P = 10$			$P = 50$		
F-test	0.060	0.093	0.096	0.066	0.080	0.082	0.045	0.093	0.065
Bathke	0.116	0.131	0.138	0.123	0.112	0.112	0.100	0.108	0.107
Wang-Akritas	0.107	0.081	0.131	0.116	0.062	0.110	0.100	0.062	0.106
Krutchkoff	0.046	0.056	0.051	0.044	0.068	0.050	0.050	0.051	0.046
Brown-Forsythe	0.039	0.061	0.057	0.056	0.071	0.066	0.043	0.092	0.064
Chen	0.045	0.044	0.051	0.050	0.057	0.058	0.040	0.052	0.036

Tabulka 5.11: Dodržování hladiny významnosti $\alpha = 0.05$ v heteroskedastickém případě 2, $\sigma_{i1}^2 = 5$, pro $i = 1, ..., I$ a $\sigma_{ij}^2 = 1$, pro $i = 1, ..., I$, $j = 2, ..., J$, $I = 10$, $J = 3$

Test	H_A	H_B	H_{AB}	H_A	H_B	H_{AB}	H_A	H_B	H_{AB}
	$P = 5$			$P = 10$			$P = 50$		
F-test	0.068	0.097	0.115	0.068	0.072	0.105	0.053	0.074	0.114
Bathke	0.091	0.198	0.093	0.085	0.168	0.088	0.076	0.170	0.081
Wang-Akritas	0.066	0.064	0.069	0.079	0.053	0.075	0.076	0.048	0.080
Krutchkoff	0.036	0.047	0.075	0.039	0.066	0.069	0.040	0.052	0.045
Brown-Forsythe	0.037	0.088	0.068	0.039	0.066	0.069	0.050	0.074	0.106
Chen	0.051	0.052	0.055	0.062	0.059	0.036	0.052	0.049	0.051

Tabulka 5.12: Dodržování hladiny významnosti $\alpha = 0.05$ v heteroskedastickém případě 2, $\sigma_{i1}^2 = 5$, pro $i = 1, ..., I$ a $\sigma_{ij}^2 = 1$, pro $i = 1, ..., I$, $j = 2, ..., J$, $I = 50$, $J = 3$

Test	H_A	H_B	H_{AB}	H_A	H_B	H_{AB}	H_A	H_B	H_{AB}
	$P = 5$			$P = 10$			$P = 50$		
F-test	0.069	0.068	0.120	0.047	0.074	0.126	0.042	0.077	0.107
Bathke	0.076	0.239	0.080	0.052	0.241	0.070	0.054	0.250	0.068
Wang-Akritas	0.047	0.036	0.046	0.043	0.046	0.051	0.052	0.050	0.066
Krutchkoff	0.033	0.050	0.116	0.029	0.045	0.078	0.050	0.045	0.049
Brown-Forsythe	0.046	0.066	0.084	0.042	0.074	0.102	0.042	0.077	0.105
Chen	0.051	0.047	0.051	0.061	0.069	0.057	0.050	0.054	0.048

Heteroskedastický případ 3
($\sigma_{1j}^2 = 5$, pro $j = 1, ..., J$ a $\sigma_{ij}^2 = 1$, pro $i = 2, ..., I$, $j = 1, ..., J$)

Zde je Chenův test také jediným testem, který dodržuje hladinu významnosti ve všech uvažovaných případech. Krutchkofův test hladinu významnosti mírně podhodnocuje v případě testování hypotézy H_B pro malý počet úrovní prvního faktoru. F-test (resp. Brown-Forsytheův test) dodržuje hladinu významnosti při testování hypotéz H_A, H_B a H_{AB} pro $I = 2$, ale pro větší počet úrovní prvního faktoru již F-test (resp. Brown-Forsytheův test) dodržuje hladinu významnosti jen při testování nulovosti sloupcových efektů. Bathkeův test výrazně nadhodnocuje hladinu významnosti téměř ve všech uvažovaných případech (výjimkou je test hypotézy H_B pro $I = 2$ a $J = 10$, kde test hladinu významnosti naopak podhodnocuje), což je opět způsobené nesplněním předpokladu tohoto testu. Wang-Akritasův test mírně nadhodnocuje hladinu významnosti při testování hypotéz H_A a H_{AB} ještě pro I=50, ale dá se předpokládat, že pro ještě větší počet úrovní prvního faktoru by se dodržování hladiny významnosti zlepšilo. Při testování hypotézy H_B Wang-Akritasův test hladinu významnosti opět výrazně podhodnocuje.

Tabulka 5.13: Dodržování hladiny významnosti $\alpha = 0.05$ v heteroskedastickém případě 3, $\sigma^2_{1j} = 5$, pro $j = 1, ..., J$ a $\sigma^2_{ij} = 1$, pro $i = 2, ..., I$, $j = 1, ..., J$, $I = 2$, $J = 3$

Test	H_A	H_B	H_{AB}	H_A	H_B	H_{AB}	H_A	H_B	H_{AB}
	$P = 5$			$P = 10$			$P = 50$		
F-test	0.055	0.066	0.060	0.048	0.056	0.050	0.045	0.073	0.067
Bathke	0.108	0.109	0.141	0.102	0.086	0.115	0.108	0.102	0.134
Wang-Akritas	0.077	0.042	0.104	0.076	0.034	0.073	0.068	0.036	0.082
Krutchkoff	0.065	0.026	0.052	0.048	0.036	0.053	0.045	0.041	0.068
Brown-Forsythe	0.042	0.037	0.042	0.045	0.048	0.045	0.044	0.072	0.065
Chen	0.057	0.051	0.060	0.054	0.045	0.049	0.054	0.056	0.055

Tabulka 5.14: Dodržování hladiny významnosti $\alpha = 0.05$ v heteroskedastickém případě 3, $\sigma^2_{1j} = 5$, pro $j = 1, ..., J$ a $\sigma^2_{ij} = 1$, pro $i = 2, ..., I$, $j = 1, ..., J$, $I = 10$, $J = 3$

Test	H_A	H_B	H_{AB}	H_A	H_B	H_{AB}	H_A	H_B	H_{AB}
	$P = 5$			$P = 10$			$P = 50$		
F-test	0.136	0.066	0.175	0.156	0.076	0.183	0.144	0.062	0.183
Bathke	0.164	0.197	0.204	0.172	0.178	0.202	0.162	0.197	0.198
Wang-Akritas	0.078	0.039	0.088	0.082	0.030	0.089	0.075	0.019	0.087
Krutchkoff	0.067	0.054	0.040	0.051	0.057	0.048	0.046	0.046	0.045
Brown-Forsythe	0.106	0.049	0.101	0.135	0.066	0.149	0.140	0.061	0.178
Chen	0.044	0.049	0.048	0.048	0.043	0.049	0.043	0.045	0.049

Tabulka 5.15: Dodržování hladiny významnosti $\alpha = 0.05$ v heteroskedastickém případě 3, $\sigma_{1j}^2 = 5$, pro $j = 1, ..., J$ a $\sigma_{ij}^2 = 1$, pro $i = 2, ..., I$, $j = 1, ..., J$, $I = 50$, $J = 3$

Test	H_A	H_B	H_{AB}	H_A	H_B	H_{AB}	H_A	H_B	H_{AB}
	$P = 5$			$P = 10$			$P = 50$		
F-test	0.193	0.054	0.213	0.172	0.052	0.179	0.178	0.063	0.183
Bathke	0.204	0.248	0.220	0.176	0.263	0.185	0.187	0.268	0.187
Wang-Akritas	0.087	0.028	0.100	0.079	0.022	0.080	0.080	0.020	0.074
Krutchkoff	0.059	0.054	0.046	0.053	0.047	0.038	0.048	0.050	0.043
Brown-Forsythe	0.148	0.052	0.152	0.156	0.051	0.154	0.177	0.062	0.178
Chen	0.048	0.040	0.043	0.062	0.055	0.047	0.056	0.054	0.038

Tabulka 5.16: Dodržování hladiny významnosti $\alpha = 0.05$ v heteroskedastickém případě 3, $\sigma_{1j}^2 = 5$, pro $j = 1, ..., J$ a $\sigma_{ij}^2 = 1$, pro $i = 2, ..., I$, $j = 1, ..., J$, $I = 2$, $J = 10$

Test	H_A	H_B	H_{AB}	H_A	H_B	H_{AB}	H_A	H_B	H_{AB}
	$P = 5$			$P = 10$			$P = 50$		
Bathke	0.102	0.015	0.127	0.112	0.009	0.123	0.117	0.005	0.143
Wang-Akritas	0.083	0.000	0.077	0.080	0.000	0.076	0.081	0.000	0.083
Krutchkoff	0.073	0.025	0.087	0.060	0.031	0.070	0.061	0.037	0.060

Dodržování hladiny významnosti - závěr

Nyní se pokusíme stručně shrnout to, co je možné říci na základě provedených simulací. V homoskedastických případech je zajisté nejlepší použít F-test, ale můžeme se spolehnout i na Krutchkoffův test, Brown-Forsytheův test a Chenův test, které také dodržely hladinu významnosti ve všech simulovaných případech. Bathkeův test a Wang-Akritasův test je možné použít jen pro testování hypotéz H_A a H_{AB}, a to v případech, kdy máme velký počet úrovní prvního fakroru.

V heteroskedastických případech se jeví jako nejlepší používat k testování hypotéz Krutchkoffův nebo Chenův test. Ty dodržují hladinu významnosti téměř ve všech uvažovaných heteroskedastických případech. V žádném případě bych nedoporučila použít F-test a ve většině případů ani Brown-Forsytheův test, jehož chování se s rostoucím I, J a P přibližuje chování F-testu. Na Brown-Forsytheův test se můžeme s jistou opatrností spolehnout pouze pro velmi malé hodnoty parametrů I, J a P. Naopak Wang-Akritasův test není dobré používat pro malé I, je odvozený za předpokladu $I \to \infty$ a pro nízký počet úrovní prvního faktoru hladinu významnosti nedodržuje. To samé můžeme říci o Bathkeově testu. Navíc určitě bych nedoporučila použít Bathkeův test, pokud si nejsme jisti jeho předpokladem, že rozptyly populací se liší jen pro různé úrovně druhého faktoru (tj. $Var(e_{ijp}) = \sigma_j^2$).

Dále Wang-Akritasův test a Bathkeův test lze použít jen pro testování hypotéz H_A a H_{AB}. Neosvědčil se ani postup navržený k testování hypotézy H_B Bathkeovým testem, ani postup navržený k testování hypotézy H_B Wang-Akritasovým testem.

5.2 Porovnání sil testů

5.2.1 Platí alternativa 1

V této části se budeme soustředit na porovnávání testů z hlediska jejich sil. Platí-li alternativa 1, znamená to, že hypotéza H_A neplatí a tedy, že alespoň jeden z parametrů α_i, $i = 1, ..., I$ je nenulový. My jsme zvolili nenulový parametr α_1. Dále předpokládáme, že všechny ostatní parametry jsou nulové, tedy že hypotézy H_B i H_{AB} platí.

Homoskedastický případ ($\sigma_{ij}^2 = 1$, $i = 1, ..., I$, $j = 1, ..., J$)

Protože síly testů má smysl porovnávat pouze u testů, které dodržují hladinu významnosti, budeme mezi sebou porovnávat ve všech simulovaných homoskedastických případech pouze síly F-testu, Krutchkoffova testu, Brown-Forsytheova testu a Chenova testu. Nejsilnějším z nich je F-test. S rostoucím počtem pozorování P se mu výrazně blíží Brown-Forsytheův a Krutchkoffův test. Můžeme říci, že pro velký počet pozorování P je síla těchto testů přibližně stejná. Dále na první pohled můžeme vidět, že nejslabším testem je Chenův test. Jeho síla se s rostoucím počtem pozorování P také zvětšuje, ale rozdíl je stále patrný. V případě, kdy počet úrovní prvního faktoru je velký, můžeme navíc síly čtyř zmíněných testů porovnat se silami Wang-Akritasova a Bathkeova testu. Z provedených simulací je patrné, že Bathkeův i Wang-Akritasův test mají vždy větší sílu než testy zbývající. To je však pravděpodobně způsobeno skutečností, že oba testy mírně nadhodnocují hladinu významnosti. Dále vidíme, že jejich síla je stejná, jestliže je počet pozorování P velký, avšak pokud máme k dispozici jen malý počet pozorování P, zdá se být Bathkeův test silnější.

Pokud se nyní ještě podíváme na dodržování hladiny významnosti testů hypotéz H_B a H_{AB}, zjistíme, že všechny testy dodržují hladinu významnosti všude, kde ji dodržovaly v případě platnosti hypotézy H_A. Neplatnost hypotézy H_A tedy nemá na dodržování hladiny významnosti testů hypotéz H_B a H_{AB} žádný vliv.

Tabulka 5.17: Porovnání sil testů v homoskedastickém případě, platí-li alternativa 1, $I = 2$, $J = 3$

Test	H_A	H_B	H_{AB}	H_A	H_B	H_{AB}	H_A	H_B	H_{AB}
	$P = 5$			$P = 10$			$P = 50$		
F-test	0.488	0.040	0.052	0.509	0.038	0.037	0.488	0.059	0.052
Krutchkoff	0.347	0.085	0.087	0.421	0.074	0.063	0.464	0.048	0.054
Brown-Forsythe	0.466	0.038	0.045	0.507	0.038	0.036	0.488	0.059	0.052
Chen	0.121	0.044	0.043	0.264	0.045	0.046	0.381	0.057	0.047

Tabulka 5.18: Porovnání sil testů v homoskedastickém případě, platí-li alternativa 1, $I = 50$, $J = 3$

Test	H_A	H_B	H_{AB}	H_A	H_B	H_{AB}	H_A	H_B	H_{AB}
	$P = 5$			$P = 10$			$P = 50$		
F-test	0.476	0.053	0.049	0.504	0.048	0.053	0.506	0.040	0.048
Bathke	0.530	0.265	0.062	0.542	0.280	0.065	0.527	0.256	0.060
Wang-Akritas	0.479	0.026	0.038	0.520	0.025	0.052	0.526	0.013	0.057
Krutchkoff	0.134	0.076	0.056	0.350	0.056	0.059	0.493	0.035	0.053
Brown-Forsythe	0.460	0.053	0.039	0.501	0.048	0.050	0.506	0.040	0.048
Chen	0.046	0.046	0.047	0.116	0.051	0.060	0.264	0.041	0.047

Heteroskedastický případ 1 ($\sigma_{11}^2 = 5$, zbylé rozptyly rovny jedné)

Vzhledem k dodržování hladiny významnosti testů má v tomto heteroskedastickém případě smysl porovnávat ve všech případech pouze sílu Krutchkoffova a Chenova testu. Krutchkoffův test se ve všech simulovaných případech ukazuje jako silnější. Jeho síla roste s rostoucím počtem pozorování P a klesá s rostoucími počty prvního i druhého faktoru. Totéž můžeme říci i o chování Chenova testu. V případě $I = 50$ a $J = 10$ můžeme ještě sílu těchto dvou testů porovnat s Wang-Akritasovým testem. Z provedených simulací je vidět, že pro malý počet pozorování je Wang-Akritasův test podstatně silnější. Co se týká chování F-testu v heteroskedastickém případě 1, z provedených simulací testující dodržování hladiny významnosti vidíme, že při testování hypotézy H_A F-test dodržoval hladinu významnosti pro $J = 10$. Podíváme se tedy, jak je na tom v těchto případech se silou. Pro malý počet pozorování P je F-test silnější než Krutchkoffův test, pro velký počet pozorování slabší. Pro jakékoli hodnoty parametrů I a P je F-test vždy silnější než Chenův test. V případě velkého počtu úrovní prvního faktoru je F-test stejně silný jako Wang-Akritasův test. Poslední test, o kterém se zde zmíníme, je Brown-Forsytheův test. Ten v tomto heteroskedastickém případě dodržoval hladinu významnosti ve stejných situacích jako F-test a z provedených simulací testujících sílu testů vidíme, že i jeho síla se rovná síle F-testu.

Co se týká dodržování hladiny významnosti testů hypotéz H_B a H_{AB} je zde situace o trochu horší než v homoskedastickém případě. Zatímco na hladinu významnosti Chenova testu nemá neplatnost hypotézy H_A žádný vliv, na dodržení hladiny významnosti Krutchkoffova testu se v případě, kdy jsou parametry I a J malé, nemůžeme spolehnout. Wang-Akritasův test, Brown-Forsytheův test a F-test dodržují hladinu významnosti při testování sloupcových efektů a interakcí ve stejných případech, jako když hypotéza H_A platí.

Tabulka 5.19: Porovnání sil testů v heteroskedastickém případě 1, platí-li alternativa 1, $I = 2$, $J = 3$

Test	H_A	H_B	H_{AB}	H_A	H_B	H_{AB}	H_A	H_B	H_{AB}
	$P = 5$			$P = 10$			$P = 50$		
Krutchkoff	0.818	0.209	0.360	0.927	0.250	0.338	0.974	0.274	0.282
Chen	0.118	0.045	0.049	0.128	0.055	0.053	0.129	0.052	0.053

Tabulka 5.20: Porovnání sil testů v heteroskedastickém případě 1, platí-li alternativa 1, $I = 10$, $J = 3$

Test	H_A	H_B	H_{AB}	H_A	H_B	H_{AB}	H_A	H_B	H_{AB}
	$P = 5$			$P = 10$			$P = 50$		
Krutchkoff	0.309	0.091	0.075	0.598	0.101	0.069	0.790	0.102	0.059
Chen	0.051	0.064	0.053	0.074	0.055	0.055	0.079	0.067	0.056

Tabulka 5.21: Porovnání sil testů v heteroskedastickém případě 1, platí-li alternativa 1, $I = 2$, $J = 10$

Test	H_A	H_B	H_{AB}	H_A	H_B	H_{AB}	H_A	H_B	H_{AB}
	$P = 5$			$P = 10$			$P = 50$		
F-test	0.515	0.158	0.155	0.497	0.116	0.130	0.514	0.131	0.136
Krutchkoff	0.493	0.086	0.137	0.660	0.075	0.094	0.813	0.052	0.063
Brown-Forsythe	0.424	0.108	0.106	0.463	0.102	0.101	0.511	0.128	0.131
Chen	0.097	0.058	0.063	0.088	0.049	0.061	0.098	0.040	0.043

Tabulka 5.22: Porovnání sil testů v heteroskedastickém případě 1, platí-li alternativa 1, $I = 50$, $J = 10$

Test	H_A	H_B	H_{AB}	H_A	H_B	H_{AB}	H_A	H_B	H_{AB}
	$P = 5$			$P = 10$			$P = 50$		
F-test	0.476	0.044	0.089	0.489	0.047	0.132	0.509	0.052	0.096
Wang-Akritas	0.468	0.000	0.056	0.497	0.000	0.079	0.516	0.000	0.061
Krutchkoff	0.075	0.051	0.057	0.359	0.048	0.050	0.526	0.044	0.052
Brown-Forsythe	0.462	0.044	0.077	0.485	0.046	0.123	0.508	0.052	0.096
Chen	0.043	0.045	0.041	0.060	0.053	0.049	0.067	0.053	0.054

Heteroskedastický případ 2
($\sigma_{i1}^2 = 5$, pro $i = 1, ..., I$ a $\sigma_{ij}^2 = 1$, pro $i = 1, ..., I$, $j = 2, ..., J$)

V heteroskedastickém případě 2 budeme porovnávat sílu Krutchkoffova testu, Chenova testu a také F-testu a Brown-Forsytheova testu, protože ty zde při testování nulovosti řádkových efektů rovněž dodržovaly hladinu významnosti ve všech simulovaných případech. V případě $I = 50$ se ještě zmíníme o síle Wang-Akritasova a Bathkeova testu.

Je patrné, že téměř ve všech simulovaných případech je Krutchkoffův test nejsilnější a Chenův test nejslabší. Výjimku tvoří případy, kdy je počet pozorování malý a počet prvního i druhého faktoru velký, potom je F-test a Brown-Forsytheův test silnější než Krutchkoffův test. V případě velkého počtu úrovní prvního faktoru, kdy všech šest uvažovaných testů dodržuje stejnou hladinu významnosti, vidíme, že F-test, Brown-Forsytheův, Bathkeův a Wang-Akritasův test mají všechny větší sílu než Chenův test a menší sílu než Krutchkoffův test (opět až na výjimku, kdy je počet pozorování malý a počet úrovní prvního i druhého faktoru velký). Z uvedené čtveřice je nejsilnější Bathkeův test. Když je počet pozorování malý, pak je jeho síla přibližně stejná jako síla F-testu. Pro velké P se mu spíše vyrovná Wang-Akritasův test, zatímco F-test je slabší.

V tomto heteroskedastickém případě můžeme říci o dodržování hladiny významnosti testů hypotéz H_B a H_{AB} víceméně totéž jako v heteroskedastickém případě 1. Navíc zde Wang-Akritasův test, stejně jako za platnosti hypotézy H_A, dodržuje hladinu významnosti i v případě testování hypotézy H_B. Na závěr doplníme informaci o dodržování hladiny významnosti Bathkeova testu, který jsme v předchozím heteroskedastickém případě neuvažovali. Bathkeův test dodržuje hladinu významnosti pouze při testování hypotézy H_{AB}, tedy stejně, jako když hypotéza H_A platí.

Tabulka 5.23: Porovnání sil testů v heteroskedastickém případě 2, platí-li alternativa 1, $I = 2$, $J = 10$

Test	H_A	H_B	H_{AB}	H_A	H_B	H_{AB}	H_A	H_B	H_{AB}
	$P = 5$			$P = 10$			$P = 50$		
F-test	0.481	0.157	0.146	0.487	0.144	0.135	0.496	0.141	0.124
Krutchkoff	0.578	0.101	0.130	0.835	0.083	0.070	0.915	0.051	0.045
Brown-Forsythe	0.414	0.096	0.100	0.461	0.127	0.114	0.489	0.136	0.117
Chen	0.086	0.050	0.054	0.089	0.056	0.050	0.087	0.043	0.043

Tabulka 5.24: Porovnání sil testů v heteroskedastickém případě 2, platí-li alternativa 1, $I = 50$, $J = 10$

Test	H_A	H_B	H_{AB}	H_A	H_B	H_{AB}	H_A	H_B	H_{AB}
	$P = 5$			$P = 10$			$P = 50$		
F-test	0.533	0.139	0.208	0.525	0.139	0.206	0.478	0.151	0.23
Bathke	0.546	0.226	0.062	0.549	0.233	0.068	0.513	0.236	0.068
Wang-Akritas	0.476	0.035	0.038	0.522	0.027	0.055	0.508	0.028	0.064
Krutchkoff	0.203	0.061	0.066	0.916	0.065	0.050	0.992	0.061	0.052
Brown-Forsythe	0.470	0.137	0.119	0.502	0.136	0.157	0.476	0.151	0.222
Chen	0.049	0.049	0.056	0.049	0.045	0.051	0.071	0.051	0.057

Tabulka 5.25: Porovnání sil testů v heteroskedastickém případě 2, platí-li alternativa 1, $I = 50$, $J = 3$

Test	H_A	H_B	H_{AB}	H_A	H_B	H_{AB}	H_A	H_B	H_{AB}
	$P = 5$			$P = 10$			$P = 50$		
F-test	0.523	0.077	0.110	0.490	0.094	0.115	0.472	0.075	0.110
Bathke	0.538	0.232	0.067	0.509	0.267	0.063	0.504	0.253	0.057
Wang-Akritas	0.454	0.035	0.039	0.463	0.054	0.046	0.497	0.046	0.053
Krutchkoff	0.679	0.077	0.121	0.996	0.076	0.086	1.000	0.050	0.066
Brown-Forsythe	0.449	0.077	0.069	0.456	0.094	0.084	0.469	0.075	0.107
Chen	0.039	0.052	0.044	0.077	0.053	0.051	0.084	0.054	0.032

Heteroskedastický případ 3
($\sigma_{1j}^2 = 5$, pro $j = 1, ..., J$ a $\sigma_{ij}^2 = 1$, pro $i = 2, ..., I$, $j = 1, ..., J$)

Také tady vidíme, že Chenův test je slabší než Krutchkoffův. Stejně tak i v tomto případě je zřejmé, že síla obou testů roste s rostoucím počtem pozorování P a klesá s rostoucími počty úrovní prvního a druhého faktoru. Pokud bychom porovnali tyto dva testy s Wang-Akritasovým testem v případě velkého počtu úrovní prvního faktoru, zjistili bychom, že je silnější než oba dva. Musíme ale poznamenat, že v tomto heteroskedastickém případě Wang-Akritasův test mírně nadhodnocoval hladinu významnosti ještě pro $I = 50$. F-test a Brown-Forsytheův test dodržovaly hladinu významnosti při testování nulovosti řádkových efektů pro $I = 2$. V těchto případech je můžeme porovnat s Krutchkoffovým a Chenovým testem. Vidíme, že jejich síla je pro malý počet pozorování větší než síla obou zmíněných testů. Ale protože síla Krutchkoffova testu, jak jsme uvedli, roste s rostoucím počtem pozorování P, pro $P = 50$ jsou F-test, Brown-Forsytheův test a Krutchkoffův test již přibližně stejně silné. Chenův test zůstává i při velkém počtu pozorování stále nejslabší.

Nakonec se zmíníme o dodržování hladiny významnosti testů při testování hypotéz H_B a H_{AB} za předpokladu, že hypotéza H_A neplatí. Zde všechny uvažované testy dodržují hladinu významnosti stejně jako v případě platnosti hypotézy H_A. A tedy, stejně jako v homoskedastickém případě, můžeme předpokládat, že neplatnost hypotézy H_A nemá na dodržování hladiny významnosti testů hypotéz H_B a H_{AB} vliv.

Tabulka 5.26: Porovnání sil testů v heteroskedastickém případě 3, platí-li alternativa 1, $I = 2$, $J = 3$

Test	H_A	H_B	H_{AB}	H_A	H_B	H_{AB}	H_A	H_B	H_{AB}
	$P = 5$			$P = 10$			$P = 50$		
F-test	0.522	0.056	0.063	0.497	0.071	0.071	0.499	0.050	0.057
Krutchkoff	0.355	0.052	0.082	0.421	0.045	0.061	0.500	0.047	0.063
Brown-Forsythe	0.457	0.040	0.049	0.469	0.064	0.058	0.495	0.048	0.056
Chen	0.112	0.045	0.054	0.191	0.045	0.044	0.270	0.057	0.045

Tabulka 5.27: Porovnání sil testů v heteroskedastickém případě 3, platí-li alternativa 1, $I = 10$, $J = 3$

Test	H_A	H_B	H_{AB}	H_A	H_B	H_{AB}	H_A	H_B	H_{AB}
	$P = 5$			$P = 10$			$P = 50$		
Krutchkoff	0.070	0.049	0.050	0.102	0.052	0.046	0.158	0.037	0.043
Chen	0.048	0.059	0.050	0.080	0.049	0.051	0.103	0.047	0.042

Tabulka 5.28: Porovnání sil testů v heteroskedastickém případě 3, platí-li alternativa 1, $I = 10$, $J = 10$

Test	H_A	H_B	H_{AB}	H_A	H_B	H_{AB}	H_A	H_B	H_{AB}
	$P = 5$			$P = 10$			$P = 50$		
Krutchkoff	0.062	0.053	0.045	0.105	0.046	0.048	0.125	0.048	0.054
Chen	0.056	0.049	0.047	0.078	0.063	0.057	0.099	0.039	0.037

Tabulka 5.29: Porovnání sil testů v heteroskedastickém případě 3, platí-li alternativa 1, $I = 50$, $J = 10$

Test	H_A	H_B	H_{AB}	H_A	H_B	H_{AB}	H_A	H_B	H_{AB}
	$P = 5$			$P = 10$			$P = 50$		
Wang-Akritas	0.322	0.000	0.055	0.322	0.000	0.069	0.315	0.000	0.058
Krutchkoff	0.056	0.059	0.054	0.059	0.046	0.049	0.082	0.054	0.058
Chen	0.044	0.048	0.040	0.069	0.052	0.037	0.052	0.055	0.052

Závěr pro alternativu 1

Testovali jsme sílu testů hypotézy H_A v případě platnosti alternativy 1, kdy $\alpha_1 > 0$, $\alpha_2 = ... = \alpha_I = 0$. Pokud jsme si jisti homoskedasticitou, doporučila bych k testování hypotéz použít F-test, neboť má ve všech testovaných případech největší sílu a také spolehlivě dodržuje hladinu významnosti. Za předpokladu heteroskedasticity je výběr nejlepšího testu o trochu složitější. Zkusíme si zde shrnout, co jsme z provedených simulací zjistili. Na první pohled je zřejmé, že nejslabší ve všech případech byl Chenův test. Dále, pokud byl počet pozorování P velký, měl vždy největší sílu Krutchkoffův test. V případech, kdy byl počet pozorování malý a bylo možné použít F-test, Brown-Forsytheův test, Wang-Akritasův test nebo Bathkeův test, potom tyto testy byly buď přibližně stejně silné jako Krutchkoffův test, nebo silnější. Nevýhodou Krutchkoffova testu je, že nedodržuje hladinu významnosti při testování hypotéz H_B a H_{AB}, pokud hypotéza H_A neplatí. V tomto ohledu se můžeme spolehnout pouze na Chenův test. Ostatní testy nedodržovaly hladinu významnosti v mnoha heteroskedastických případech, ani pokud hypotéza H_A platila.

5.2.2 Platí alternativa 2

Nyní budeme zkoumat síly testů za předpokladu, že neplatí hypotéza H_{AB}, tj. za podmínky, že alespoň jeden z parametrů γ_{ij}, $i = 1, ..., I$, $j = 1, ..., J$ je nenulový. My jsme jako nenulový zvolili parametr γ_{11}. Dále předpokládáme, že hypotézy H_A a H_B platí, tj. že řádkové i sloupcové efekty jsou nulové.

Homoskedastický případ ($\sigma_{ij}^2 = 1$, $i = 1, ..., I$, $j = 1, ..., J$)

V homoskedastickém případě při testování hypotézy H_{AB}, jestliže je počet úrovní prvního faktoru malý, je opět většinou nejsilnější F-test. Pouze v několika málo případech ho svou silou mírně převyšuje Krutchkoffův test. Pokud je počet úrovní prvního faktoru velký, můžeme k testování hypotéz použít i Bathkeův a Wang-Akritasův test. V tomto případě se síla F-testu zdá být menší než síla Bathkeova testu pro jakýkoli počet pozorování P a pro velký počet pozorování P i menší než síla Wang-Akritasova testu. Skutečnost, že v několika případech se některé testy zdají být silnější než F-test, je opět způsobena tím, že tyto testy nedodržují hladinu významnosti tak spolehlivě jako F-test. Chenův test se znovu ve všech situacích ukazuje jako nejslabší.

Dále se podíváme, zda testy při testování hypotéz H_A a H_B dodržují hladinu významnosti. Na první pohled vidíme, že tomu tak není. V některých případech zamítáme hypotézy H_A a H_B dokonce častěji než hypotézu H_{AB}. Situace se sice s rostoucím počtem úrovní prvního a druhého faktoru zlepšuje, avšak ani tak téměř nikde nelze říci, že by testy hladinu významnosti dodržely.

Tabulka 5.30: Porovnání sil testů v homoskedastickém případě, platí-li alternativa 2, $I = 2$, $J = 10$

Test	H_A	H_B	H_{AB}	H_A	H_B	H_{AB}	H_A	H_B	H_{AB}
	$P = 5$			$P = 10$			$P = 50$		
F-test	0.171	0.501	0.529	0.187	0.516	0.512	0.163	0.483	0.497
Krutchkoff	0.145	0.289	0.389	0.176	0.404	0.453	0.159	0.490	0.510
Brown-Forsythe	0.169	0.477	0.501	0.187	0.508	0.506	0.163	0.483	0.496
Chen	0.061	0.069	0.063	0.094	0.162	0.162	0.127	0.354	0.314

Tabulka 5.31: Porovnání sil testů v homoskedastickém případě, platí-li alternativa 2, $I = 50$, $J = 10$

Test	H_A	H_B	H_{AB}	H_A	H_B	H_{AB}	H_A	H_B	H_{AB}
	$P = 5$			$P = 10$			$P = 50$		
F-test	0.162	0.085	0.507	0.165	0.067	0.513	0.170	0.085	0.517
Bathke	0.186	0.279	0.527	0.194	0.271	0.527	0.192	0.294	0.539
Wang-Akritas	0.177	0.002	0.413	0.188	0.000	0.483	0.192	0.000	0.531
Krutchkoff	0.064	0.056	0.246	0.160	0.085	0.463	0.168	0.078	0.509
Brown-Forsythe	0.159	0.085	0.471	0.163	0.067	0.507	0.170	0.085	0.517
Chen	0.040	0.046	0.034	0.056	0.060	0.092	0.120	0.065	0.266

Heteroskedastický případ 1 ($\sigma_{11}^2 = 5$, zbylé rozptyly rovny jedné)

V heteroskedastickém případě 1 budeme ve všech případech porovnávat pouze Krutchkoffův a Chenův test. Opět vidíme, že Krutchkoffův test je silnější než Chenův. Pro $I = 2$ a $J = 3$ navíc při testování dodržování hladiny významnosti v případě hypotézy H_{AB} dodržel hladinu významnosti Brown-Forsytheův test. Srovnáme-li jeho sílu se zbylými dvěma testy, zjistíme, že je silnější. Totéž zjistíme, i pokud použijeme Wang-Akritasův test v případě velkého počtu úrovní prvního faktoru. Ostatní testy při testování hypotézy H_{AB} nedodržely hladinu významnosti, a proto je zde nebudeme uvažovat.

Co se týká dodržování hladiny významnosti při testování hypotéz H_A a H_B, je zde situace malinko lepší než v homoskedastickém případě. Toto zlepšení zaznamenáváme u Krutchkoffova testu, který dodržuje hladinu významnosti při testování hypotéz H_A a H_B ve všech simulovaných případech.

Tabulka 5.32: Porovnání sil testů v heteroskedastickém případě 1, platí-li alternativa 2, $I = 2$, $J = 3$

Test	H_A	H_B	H_{AB}	H_A	H_B	H_{AB}	H_A	H_B	H_{AB}
	$P = 5$			$P = 10$			$P = 50$		
Krutchkoff	0.066	0.073	0.467	0.062	0.068	0.534	0.053	0.063	0.544
Brown-Forsythe	0.467	0.539	0.546	0.545	0.634	0.639	0.596	0.696	0.682
Chen	0.126	0.118	0.138	0.176	0.199	0.184	0.164	0.198	0.230

Tabulka 5.33: Porovnání sil testů v heteroskedastickém případě 1, platí-li alternativa 2, $I = 50$, $J = 3$

Test	H_A	H_B	H_{AB}	H_A	H_B	H_{AB}	H_A	H_B	H_{AB}
	$P = 5$			$P = 10$			$P = 50$		
Wang-Akritas	0.540	0.105	0.581	0.546	0.098	0.575	0.532	0.093	0.571
Krutchkoff	0.062	0.054	0.123	0.054	0.049	0.092	0.047	0.063	0.100
Chen	0.059	0.045	0.054	0.066	0.059	0.081	0.065	0.047	0.073

Tabulka 5.34: Porovnání sil testů v heteroskedastickém případě 1, platí-li alternativa 2, $I = 2$, $J = 10$

Test	H_A	H_B	H_{AB}	H_A	H_B	H_{AB}	H_A	H_B	H_{AB}
	$P = 5$			$P = 10$			$P = 50$		
Krutchkoff	0.052	0.040	0.189	0.067	0.060	0.256	0.047	0.069	0.261
Chen	0.058	0.064	0.065	0.085	0.117	0.109	0.067	0.115	0.122

Heteroskedastický případ 2
($\sigma_{i1}^2 = 5$, pro $i = 1, ..., I$ a $\sigma_{ij}^2 = 1$, pro $i = 1, ..., I$, $j = 2, ..., J$)

Chování testů v heteroskedastickém případě 2 se velmi podobá jejich chování v heteroskedastickém případě 1. Proto si nyní popíšeme jen případné rozdíly. V tomto heteroskedastickém případě můžeme kromě Wang-Akritasova testu použít, pokud je I velké, i Bathkeův test. Síla Bathkeova testu je větší než síla Wang-Akritasova testu, a tudíž Bathkeův test je nejsilnější ze všech testů, které je možné v případě velkého počtu úrovní prvního faktoru použít.

Jedinou další odlišnost nalezneme v dodržování hladiny významnosti Krutchkoffova testu při testování hypotéz H_A a H_B za předpokladu, že H_{AB} neplatí. A sice tu, že Krutchkoffův test dodržuje hladinu významnosti jen v případě testování hypotézy H_A.

Tabulka 5.35: Porovnání sil testů v heteroskedastickém případě 2, platí-li alternativa 2, $I = 2$, $J = 3$

Test	H_A	H_B	H_{AB}	H_A	H_B	H_{AB}	H_A	H_B	H_{AB}
	$P = 5$			$P = 10$			$P = 50$		
Krutchkoff	0.073	0.270	0.296	0.052	0.312	0.329	0.051	0.350	0.348
Brown-Forsythe	0.368	0.414	0.417	0.404	0.487	0.478	0.404	0.519	0.491
Chen	0.085	0.097	0.084	0.138	0.136	0.169	0.178	0.213	0.211

Tabulka 5.36: Porovnání sil testů v heteroskedastickém případě 2, platí-li alternativa 2, $I = 50$, $J = 3$

Test	H_A	H_B	H_{AB}	H_A	H_B	H_{AB}	H_A	H_B	H_{AB}
	$P = 5$			$P = 10$			$P = 50$		
Bathke	0.382	0.289	0.403	0.356	0.326	0.391	0.389	0.342	0.424
Wang-Akritas	0.295	0.065	0.307	0.318	0.088	0.336	0.381	0.086	0.414
Krutchkoff	0.032	0.071	0.293	0.038	0.078	0.281	0.041	0.083	0.259
Chen	0.055	0.057	0.048	0.066	0.053	0.065	0.104	0.051	0.129

Heteroskedastický případ 3
($\sigma_{1j}^2 = 5$, pro $j = 1, ..., J$ a $\sigma_{ij}^2 = 1$, pro $i = 2, ..., I$, $j = 1, ..., J$)

Také tady si popíšeme jen odlišnosti od ostatních heteroskedastických případů. V heteroskedastickém případě 3 lze pro $I = 2$ použít kromě Krutchkoffova a Chenova testu i F-test a Brown-Forsytheův test. Z této čtveřice má největší sílu F-test a Brown-Forsytheův test se mu velmi rychle s rostoucím počtem pozorování přibližuje. Bathkeův test zde není možné použít, je to opět z důvodu nedodržování hladiny významnosti při testování hypotézy H_{AB}.

Už se zbývá jen zmínit o dodržování hladiny významnosti testů při testování hypotéz H_A a H_B. Zde, na rozdíl od předchozího případu, již nezaznamenáváme dodržování hladiny významnosti Krutchkoffova testu v případě testování hypotézy H_A, ale naopak pokud testujeme hypotézu H_B.

Tabulka 5.37: Porovnání sil testů v heteroskedastickém případě 3, platí-li alternativa 2, $I = 2$, $J = 10$

Test	H_A	H_B	H_{AB}	H_A	H_B	H_{AB}	H_A	H_B	H_{AB}
	$P = 5$			$P = 10$			$P = 50$		
F-test	0.157	0.488	0.481	0.164	0.514	0.523	0.159	0.485	0.487
Krutchkoff	0.148	0.051	0.413	0.166	0.049	0.520	0.154	0.048	0.486
Brown-Forsythe	0.147	0.406	0.401	0.163	0.489	0.497	0.159	0.484	0.483
Chen	0.062	0.051	0.057	0.075	0.122	0.112	0.105	0.190	0.201

Závěr pro alternativu 2

Shrňme si, co jsme z provedených simulací testujících sílu testů za předpokladu platnosti alternativy 2 zjistili. V homoskedastickém případě měl ze čtveřice F-test, Brown-Forsytheův test, Krutchkoffův test a Chenův test největší sílu ve většině případech F-test. V případě velkého počtu úrovní prvního faktoru se zdál být Bathkeův test, a pokud byl počet pozorování P velký i Wang-Akritasův test, silnější než F-test. To je však, jak jsme již uvedli, způsobeno horším dodržováním hladiny významnosti těchto testů.

V heteroskedastických případech, pokud je možné použít jen Krutchkoffův nebo Chenův test (z důvodu dodržování hladiny významnosti), bych radila použít test Krutchkoffův, neboť se ukázal nejen jako silnější, ale také lépe dodržoval hladinu významnosti v případě testování hypotéz H_A a H_B. Jestliže bylo možné v některých případech použít jiný z testů, pak tento test byl silnější při testování hypotézy H_{AB} než Krutchkoffův. Na závěr bych ještě chtěla připomenout, že většina testů při výskytu interakcí nedodržovala hladinu významnosti při testování hypotéz H_A a H_B, a proto bychom si na to při testování měli dát pozor.

Kapitola 6

Závěr

Cílem této bakalářské práce bylo uvést a porovnat několik testů na testování hypotéz o nulovosti řádkových a sloupcových efektů a interakcí v modelu analýzy rozptylu dvojného třídění pro homoskedastická a heteroskedastická data. Testy jsme porovnávali na základě simulací, při nichž jsme zkoumali, jak se testy chovají v námi předem zvolených případech. Zjišťovali jsme, které z nich dodržují hladinu významnosti a který z testů, jež hladinu významnosti dodržely, je nejsilnější za platnosti námi zvolených alternativ.

Za předpokladu homoskedasticity a normality víme, že F-test je stejnoměrně nejsilnější test. Náš výzkum ukázal, že při poruše homoskedasticity tomu tak již není, neboť v tomto případě má F-test velký problém s dodržováním hladiny významnosti.

Nyní si popíšeme, co jsme z provedených simulací zjistili o ostatních uvažovaných testech. V zásadě můžeme říci, že v homoskedastickém případě hladinu významnosti dodržovaly všechny testy za podmínky, že byly splněny předpoklady, za kterých byly tyto testy odvozeny. Na problém s dodržováním hladiny významnosti v homoskedastickém případě jsme narazili pouze při výskytu interakcí při testování hypotéz H_A a H_B. Pokud jsme si jisti homoskedasticitou, je, jak už jsme uvedli, nejlepší použít F-test. Wang-Akritasův a Bathkeův test se sice ukazují v jistých situacích jako silnější, avšak to je pravděpodobně způsobené tím, že oba mírně nadhodnocují hladinu významnosti. Navíc tyto dva testy můžeme použít jen k testování hypotézy o nulovosti těch efektů, jejichž počet úrovní příslušného faktoru je velký. Pokud je počet úrovní alespoň jednoho faktoru velký, pak je zároveň možné těmito testy testovat i hypotézu o nulovosti interakcí. Při testování hypotézy o nulovosti efektů, jejichž počet úrovní příslušného faktoru je malý, ani Bathkeův ani Wang-Akritasův test nedodržují hladinu významnosti.

V heteroskedastických případech vždy dodržovaly hladinu významnosti, pokud byly splněny jejich předpoklady, jen Bathkeův, Wang-Akritasův, Krutchkoffův a Chenův test. Výhodou Krutchkoffova a Chenova testu je, že je můžeme použít ve všech případech, tj. pro jakékoli počty pozorování P a jakékoli počty úrovní prvního a druhého faktoru I a J. Pokud bychom měli rozhodnout, který z těchto dvou testů je lepší na základě jejich sil, dala bych přednost Krutchkoffovu testu, neboť se ve všech simulovaných případech ukázal jako silnější. To, na co bychom si při použití Krutchkoffova testu měli dávat pozor, je mírné nedodržování hladiny významnosti při neplatnosti hypotézy H_A při testování zbylých hypotéz. Wang-Akritasův a Bathkeův test jsou odvozeny za předpokladu, že počet úrovní jednoho z faktorů jde do nekonečna. Proto je můžeme použít, jak už jsme se zmínili v homokedastickém případě, jen pokud je počet úrovní jednoho z faktorů velký. Navíc Bathkeův test k tomu, aby dodržoval hladinu významnosti, potřebuje, aby rozptyly náhodných veličin byly různé pouze pro různé úrovně toho faktoru, u kterého nepředpokládáme, že se jeho počet úrovní blíží nekonečnu. V těch případech, kdy tyto testy použít můžeme, jsou oba dva často silnější než Krutchkoffův test. Co se týká Brown-Forsytheova testu, ten se s rostoucími hodnotami parametrů I, J a P přibližuje velmi svým chováním F-testu, a to jak v dodržování hladiny významnosti, tak i svou silou. F-test dodržoval při poruše homoskedasticity hladinu významnosti jen v některých případech. Proto bych tyto testy používala bez předpokladu homoskedasticity jen s velkou opatrností. Síly F-testu a Brown-Forsytheova testu byly vždy větší než síla Chenova testu a pokud jsme měli k dispozici jen malý počet pozorování P, pak často i větší než síla Krutchkoffova testu. Nakonec bych ještě znovu chtěla upozornit na problém s dodržováním hladiny významnosti při výskytu interakcí při testování hypotéz H_A a H_B. Týká se všech testů v homoskedastickém i heteroskedastickém případě a je potřeba si na něj dávat při testování hypotéz pozor.

Příloha

Součástí této bakalářské práce je i CD, které kromě textu bakalářské práce obsahuje i další 4 soubory. Tři z nich obsahují výsledky simulací, ve čtvrtém souboru nalezneme zdrojový kód simulací pro software **R**.

DodrzeniHladiny.pdf - *Obsahuje tabulky s výsledky simulací testujících dodržování hladiny významnosti.*
SilyTestuA1.pdf - *Obsahuje tabulky s výsledky simulací zjišťujících sílu testů za platnosti alternativy 1.*
SilyTestuA2.pdf - *Obsahuje tabulky s výsledky simulací zjišťujících sílu testů za platnosti alternativy 2.*
Simulace.R - *Obsahuje zdrojový kód simulací pro software **R**.*

Literatura

[1] Anděl, J.: *Základy matematické statistiky*, Matfyzpress, Praha, 2005.

[2] Wang, L., Akritas, M.: *Two-way Heteroscedastic ANOVA when the Number of Levels is Large*, The Pennsylvania State University, University Park, USA, 2002.

[3] Brown M. B., Forsythe, A. B.: *The Anova and Multiple Comparisons for Data with Heterogeneous Variances*, Biometrics, 1974, 719-724.

[4] Bathke, A.: *The ANOVA F Test Can Still Be Used in Some Balanced Designs With Unequal Variances and Nonnormal Data*, University of Kentucky, 2003.

[5] Krutchkoff, R. G.: *Two-way fixed analysis of variance when the error variances may be unequal*, Journal of Statistical Computation and Simulation, 32, 1989, 177-183.

[6] Chen, S., Chen, H. J.: *Single-stage analysis of variance under heteroscedasticity*, Communication in Statistics-Simulation and Computation, 1998, 641-666.