

**EKONOMICKÁ UNIVERZITA
V BRATISLAVE**

FAKULTA HOSPODÁRSKEJ INFORMATIKY

Evidenčné číslo: 103005/I/2020/421000214186

**TESTY HYPOTÉZ POROVNÁVAJÚCICH
PARAMETRE VIAC AKO DVOCH ZÁKLADNÝCH
SÚBOROV**

Diplomová práca

2020

Bc. Martin Sliacky

**EKONOMICKÁ UNIVERZITA
V BRATISLAVE
FAKULTA HOSPODÁRSKEJ INFORMATIKY**

**TESTY HYPOTÉZ POROVNÁVAJÚCICH
PARAMETRE VIAC AKO DVOCH
ZÁKLADNÝCH SÚBOROV**

Diplomová práca

Študijný program: Štatistické metódy v ekonómii

Študijný odbor: Ekonómia a manažment

Školiace pracovisko: Katedra štatistiky

Vedúci záverečnej práce: RNDr. Eva Kotlebová, PhD.

Bratislava 2020

Bc. Martin Sliacky

Čestné vyhlásenie

Vyhlasujem, že diplomovú prácu som vypracoval samostatne a uviedol som všetky použité zdroje a literatúru.

Dátum:

Martin Sliacky

PodĎakovanie

Na tomto mieste sa chcem poďakovať RNDr. Eve Kotlebovej, PhD., vedúcej mojej diplomovej práce, za odbornú konzultáciu, usmernenie a pomoc pri vypracovávaní tejto práce.

ABSTRAKT

SLIACKY, Martin. *Testy hypotéz porovnávajúcich parametre viac ako dvoch základných súborov*. Ekonomická univerzita v Bratislave. Fakulta hospodárskej informatiky; Katedra štatistiky. – Vedúci záverečnej práce: RNDr. Eva Kotlebová, PhD. – Bratislava: FHI EU, 2020, 83 s.

Predkladaná diplomová práca sa zaoberá problematikou testovania hypotéz porovnávajúcich parametre viac ako dvoch základných súborov. Práca pozostáva zo 4 kapitol, obsahuje 11 grafov a 15 tabuliek. V prvej kapitole je stručne popísaná metóda analýzy rozptylu, problematika chýb štatistických testov, ich robustnosť a sila. V druhej časti sú uvedené hlavné a čiastkové ciele práce. Hlavným cieľom práce je porovnať klasickú verziu analýzy rozptylu a Welchovu analýzu rozptylu z hľadiska robustnosti a sily testu, a súčasne objasniť, pri akých parametroch údajov je vhodnejšie použiť Welchovu analýzu rozptylu a kedy klasickú analýzu rozptylu. V tretej časti je vysvetlená metodika práce. V podkapitolách tretej časti je podrobne popísaná jednofaktorová analýza rozptylu a Welchova analýza rozptylu. V ďalšej podkapitole sú vysvetlené predpoklady použitia uvedených metód a spôsoby ich overovania. V závere tretej kapitoly je popísaný vplyv nesplnenia predpokladov na robustnosť štatistických metód a následne postup na určenie mier chýb štatistických testov pomocou simulácií Monte Carlo. Štvrtá časť obsahuje výsledky práce. Metódy popísané v metodologickej časti sú aplikované na reálne údaje, popisujúce ceny rôznych konfigurácií počítačových zostáv. V podkapitole štvrtej časti je pomocou Monte Carlo simulácií analyzovaný vplyv heteroskedasticity na silu a robustnosť klasickej jednofaktorovej analýzy rozptylu a Welchovej analýzy rozptylu. Na základe výsledkov analýzy robustnosti a sily testov sa kvantifikuje vplyv heteroskedasticity na skúmané štatistické metódy a uvádza sa, pri akých vstupných údajov je výhodnejšie použiť Welchovu analýzu rozptylu namiesto klasickej analýzy rozptylu.

Kľúčové slová: analýza rozptylu, Welchova analýza rozptylu, homoskedasticita, sila testu, robustnosť, Monte Carlo simulácie

ABSTRACT

SLIACKY, Martin. *Hypothesis tests comparing parameters of more than two populations*. University of Economics in Bratislava. Faculty of Economic Informatics; Department of Statistics. – Thesis supervisor: RNDr. Eva Kotlebová, PhD. – Bratislava: FHI EU, 2020, 83 p.

The following thesis is focused on the topic of the comparison of parameters of more than two populations. It consists of four chapters alongside with 11 charts and 15 tables. In the first chapter we provide the overview of analysis of variance, topic of errors in statistical tests, their robustness and power. In the second chapter we state the main and the partial objectives of the thesis. The thesis' main objective is to compare the classic version of analysis of variance and Welch's analysis of variance in terms of their robustness and power. Simultaneously, we would like to assess, when, in terms of the input data's parameters, it is advisable to use Welch's analysis of variance instead of classic analysis variance. The third chapter describes methodology used in the subsequent analysis. We describe the one-way analysis of variance and Welch's analysis of variance. The following chapter is focused on the assumptions of the tests and their verification methods. Subsequently, we demonstrate how the violation of the tests' assumptions affects their robustness. Then we explain the estimation of tests' type I and II error rates, using Monte Carlo simulations. The results of the thesis are presented in the fourth chapter, where the methods described in the thesis' methodology are applied on the pricing data of computer configurations. In this chapter, using the Monte Carlo simulations, we analyse the effect of heteroscedasticity on the robustness and power of classic one-way analysis of variance and Welch's analysis of variance. Based on the results of the analysis, we quantify the effect of heteroscedasticity on the tests, and propose recommendations how to choose between them.

Key words: analysis of variance, Welch's analysis of variance, homoscedasticity, power of a test, robustness, Monte Carlo simulations

OBSAH

| | |
|---|----|
| Úvod | 10 |
| 1 Súčasný stav riešenej problematiky doma a v zahraničí..... | 12 |
| 1.1 Analýza rozptylu | 12 |
| 1.2 Chyby štatistických testov..... | 13 |
| 2 Cieľ práce..... | 16 |
| 3 Metodika práce | 17 |
| 3.1 Rozdiel medzi analýzou rozptylu a dvojvýberovým t-testom..... | 17 |
| 3.2 Dvojvýberový F-test zhody rozptylov | 18 |
| 3.3 Analýza rozptylu | 19 |
| 3.3.1 Zápis modelu | 19 |
| 3.3.2 Zápis štatistických hypotéz..... | 21 |
| 3.3.3 Určenie stupňov voľnosti | 25 |
| 3.3.4 Tabuľka analýzy rozptylu..... | 26 |
| 3.4 Welchova analýza rozptylu | 27 |
| 3.5 Kruskal-Wallisov test..... | 29 |
| 3.6 Predpoklady použitia analýzy rozptylu | 31 |
| 3.6.1 Normalita..... | 32 |
| 3.6.2 Nezávislosť pozorovaní..... | 33 |
| 3.6.3 Homoskedasticita (zhoda rozptylov)..... | 33 |
| 3.6.4 Hartleyov test..... | 34 |
| 3.6.5 Bartlettov test..... | 35 |
| 3.6.6 Leveneov test..... | 36 |
| 3.6.7 O'Brienov test | 37 |
| 3.7 Transformácia údajov..... | 38 |
| 3.8 Robustnosť analýzy rozptylu | 39 |

| | | |
|-------|--|----|
| 3.8.1 | Porušenie normality | 40 |
| 3.8.2 | Porušenie homoskedasticity | 41 |
| 3.9 | Simulácie a metóda Monte Carlo | 42 |
| 3.9.2 | Stanovenie mier chýb I. a II. druhu pomocou simulácií..... | 44 |
| 4 | Výsledky práce | 47 |
| 4.1 | Aplikácia analýzy rozptylu | 48 |
| 4.1.1 | Údajová základňa | 48 |
| 4.1.2 | Overenie homoskedasticity | 49 |
| 4.1.3 | Posúdenie vplyvu druhu diskov na cenu zostavy | 50 |
| 4.1.4 | Porovnanie výsledkov | 51 |
| 4.2 | Porovnanie robustnosti analýzy rozptylu a Welchovej analýzy rozptylu .. | 53 |
| 4.3 | Výsledky analýzy robustnosti | 57 |
| 4.3.1 | Analýza rozptylu | 57 |
| 4.3.2 | Robustnosť Welchovej analýzy rozptylu | 58 |
| 4.3.3 | Zhodnotenie analýzy robustnosti..... | 60 |
| 4.4 | Porovnanie sily testu analýzy rozptylu a Welchovej analýzy rozptylu..... | 60 |
| 4.4.1 | Výsledky analýzy sily testu | 64 |
| 4.4.2 | Zhodnotenie analýzy sily testov | 68 |
| | Záver..... | 70 |
| | Zoznam použitej literatúry | 74 |
| | Knižné publikácie..... | 74 |
| | Články v časopise..... | 75 |
| | Internetové zdroje..... | 77 |
| | Prílohy | 78 |
| | 1. Tabuľka odhadnutých mier I. druhu..... | 78 |
| | 2. Tabuľka odhadnutej sily testov | 79 |
| | 3. Zdrojový kód | 81 |

Zoznam ilustrácií a tabuliek

| | |
|--|----|
| Graf 3.1: Histogram zobrazujúci pravostranne zošikmené údaje | 32 |
| Graf 3.2: Intervaly spoľahlivosti stredných hodnôt porovnávaných skupín..... | 34 |
| Graf 4.1: Stredné hodnoty zostáv rozčlenené podľa obmien faktora <i>druh diskov</i> | 52 |
| Graf 4.2: Miera chýb I. druhu klasickej analýzy rozptylu | 57 |
| Graf 4.3: Miera chýb I. druhu Welchovej analýzy rozptylu | 59 |
| Graf 4.4: Sila klasickej analýzy rozptylu pri rovnomernom počte pozorovaní | 64 |
| Graf 4.5: Sila Welchovej analýzy rozptylu pri rovnomernom počte pozorovaní | 65 |
| Graf 4.6 Sila analýzy rozptylu pri priamej úmernosti medzi variabilitou a počtom pozorovaní | 66 |
| Graf 4.7: Sila Welchovej analýzy rozptylu pri priamej úmernosti medzi variabilitou a počtom pozorovaní | 67 |
| Graf 4.8: Sila analýzy rozptylu pri nepriamej úmernosti medzi variabilitou a počtom pozorovaní | 67 |
| Graf 4.9: Sila Welchovej analýzy rozptylu pri nepriamej úmernosti medzi variabilitou a počtom pozorovaní | 68 |
| | |
| Tabuľka 1.1: Rozdelenie chýb štatistických testov | 14 |
| Tabuľka 3.1: Zápis modelu jednofaktorovej analýzy rozptylu..... | 21 |
| Tabuľka 3.2: Tabuľka výsledkov analýzy rozptylu | 27 |
| Tabuľka 3.3: Najčastejšie používané transformácie údajov..... | 38 |
| Tabuľka 4.1: Úrovně faktorov skúmanej premennej | 49 |
| Tabuľka 4.2: Výsledky Leveneovho testu homoskedasticity | 50 |
| Tabuľka 4.3: Výsledky jednofaktorovej analýzy rozptylu | 50 |
| Tabuľka 4.4: Výsledky Welchovej analýzy rozptylu | 51 |
| Tabuľka 4.5: Popisné charakteristiky vstupných údajov rozčlenené podľa skúmanej premennej | 52 |
| Tabuľka 4.6: Parametre údajov použitých pri analyzovaní robustnosti | 55 |
| Tabuľka 4.7: Parametre údajov použitých pri analyzovaní sily testu..... | 62 |
| Tabuľka 4.8: Parametre údajov použitých pri analyzovaní sily testu..... | 62 |
| Tabuľka 4.9: Parametre údajov použitých pri analyzovaní sily testu..... | 63 |

Úvod

Vzájomné porovnávanie viacerých súborov je jedným z najčastejšie využívaných štatistických postupov snád' vo všetkých vedných odboroch a oblastiach ich aplikácie. Na tento účel sú používané štatistické testy, ktoré nám umožňujú potvrdiť alebo vyvrátiť naše hypotézy o parametroch porovnávaných súborov. Táto diplomová práca sa zaoberá práve problémom porovnávania viacerých základných súborov a ich parametrov štatistickými testami.

Prvá časť práce je venovaná prehľadu súčasného stavu riešenej problematiky. Stručne je tu vysvetlené akými metódami sa budeme zaoberať hlbšie v teoretickej časti práce. Venujeme sa analýze rozptylu, čo je jedna z najpoužívanějších štatistických metód v oblasti porovnávania viacerých základných súborov. Je tu spomenutá aj problematika chýb v štatistických testoch, prečo by sme im mali venovať pozornosť, a spôsoby akými sa zisťujú.

V metodologickej časti práce sa ďalej venujeme metóde analýzy rozptylu. Najprv ukazujeme, prečo nie sú na problém porovnávania viacerých súborov vhodné jednoduchšie štatistické testy. Vysvetľujeme jej základnú myšlienku a pri akých skúmaných problémoch je vhodné jej použitie. Podrobne opisujeme, ako prebieha rozklad variability údajov na jednotlivé zložky a ich úpravami sa dostávame k testovacej štatistike. V nasledujúcej kapitole popisujeme variáciu analýzy rozptylu zvanú Welchova ANOVA, ktorú je vhodné použiť ak nie je splnený jeden z predpokladov využitia klasickej analýzy rozptylu – homoskedasticita, teda zhoda rozptylov porovnávaných súborov. Ďalej je vysvetlená neparametrická alternatíva k analýze rozptylu v prípade ak analyzované údaje nepochádzajú z normálneho rozdelenia.

Predpoklady na použitie analýzy rozptylu a spôsoby ich overovania si priblížime v samostatnej časti. Bližšie sa pozrieme na podmienku rovnosti rozptylov medzi porovnávanými skupinami. Uvedieme štatistické testy, ktorými je túto podmienku možné overiť, a ktoré z nich sú považované za najvhodnejšie. Uvedieme aj možnosti ako transformovať údaje za účelom priblíženia sa k požadovaným predpokladom na správne použitie analýzy rozptylu.

V ďalšej kapitole metodologickej časti práce vysvetlíme, aké chyby sa môžu udiť pri štatistickom testovaní hypotéz. Objasníme pojmy ako miera chýb I. druhu a sila testu.

Vysvetlíme, ako môžeme v inferenčnej štatistike využiť simulačné metódy Monte Carlo a ako pomocou nich odhadnúť miery chýb v štatistických testoch.

V tretej časti sú popísané ciele, ktoré chceme v práci dosiahnuť a spôsob, akým metódy vysvetlené v metodologickej časti budú slúžiť na tento účel.

Štvrtá časť je zameraná na naplnenie cieľov spomenutých v predošlej časti práce. V prvej časti aplikujeme metódy z teoretickej časti analýzy rozptylu na údajoch, ktoré popisujú ceny počítačových zostáv. Naším cieľom bude zistiť, či na cenu zostáv vplyva zvolený faktor. Overíme predpoklady na využitie analýzy rozptylu a následne našu hypotézu budeme overovať dvomi rôznymi variantmi analýzy rozptylu. Posúdime, či je rozdiel v prijatých záveroch pri použití bežnej analýzy rozptylu oproti Welchovej analýze rozptylu.

V druhej kapitole výsledkovej časti práce budeme porovnávať klasickú analýzu rozptylu a Welchovu analýzu rozptylu z hľadiska miery chýb I. druhu a sily testu. Zameriame sa na analýzu vplyvu heteroskedasticity porovnávaných údajov. Budeme sa snažiť určiť, pri akých parametroch analyzovaných údajov je vhodnejšie použiť jednu alebo druhú metódu. Na odhad miery chýb a sily testov použijeme metódu Monte Carlo.

1 Súčasný stav riešenej problematiky doma a v zahraničí

Pri vedeckom výskume je jedným z častých postupov vzájomné porovnávanie viacerých (viac ako dvoch) skupín, ktoré sa navzájom odlišujú v určitom znaku. Mnohokrát nás zaujíma, či je medzi skupinami štatisticky významný rozdiel, teda či je rozdiel medzi nimi spôsobený určitým faktorom. Z toho následne vyvstávajú ďalšie podstatné otázky. Napríklad: ktoré konkrétne skupiny sa odlišujú, aká je miera tejto odlišnosti alebo aké veľké by mali byť skúmané vzorky na to, aby sme dokázali zistiť odlišnosť. Zisťovanie rozdielov medzi skupinami sa vykonáva často vo vedeckých výskumoch, ekonomickej praxi alebo technických procesoch. Napríklad ak sú na pacientoch testované rozdielne spôsoby liečby, zaujíma nás, ktorá z nich prináša želané výsledky. Vo výrobe môžeme použiť viaceré odlišné metódy na výrobu určitého produktu a chceme zistiť, ktorá metóda je najvhodnejšia.

1.1 Analýza rozptylu

Práve na otázku, či sa porovnávané skupiny významne líšia, odpovedá *analýza rozptylu*, známa tiež pod skratkou ANOVA (z anglického analysis of variance). Umožňuje určiť vplyv jedného alebo viacerých faktorov na skúmanú závislú premennú. Posudzujeme ňou významnosť jednotlivých zdrojov variability v lineárnych štatistických modeloch. Celkovú variabilitu v skúmaných údajoch môžeme rozložiť na dve špecifické zložky. Na zložku variability medzi skupinami, ktorú získame, keď údaje rozdelíme podľa obmien skúmaného faktora a zložku variability vo vnútri skupín, ktorá je spôsobená inými faktormi alebo náhodnými vplyvmi. Analýzu rozptylu po prvýkrát spomína Ronald Fisher vo svojej práci *Štatistické metódy pre výskumných pracovníkov* z roku 1925.

ANOVA je metódou štatistického testovania hypotéz. V typickej aplikácii analýzy rozptylu nulová hypotéza tvrdí, že všetky porovnávané skupiny pochádzajú z tej istej populácie. Pri aplikácii v medicínskom výskume sa často skúma vplyv rôznych metód liečby na vzorkách pacientov. V takom prípade nulová hypotéza tvrdí, že všetky metódy liečby majú rovnaký efekt. Zamietnutie nulovej hypotézy (prijatie alternatívnej) znamená, že pozorované rozdiely v efektoch rôznych metód liečby nie sú spôsobené náhodnými vplyvmi.

ANOVA je často používaným nástrojom pri *navrhovaní a vyhodnocovaní experimentov*, kde je cieľom zistiť efekt intervencie na určitý objekt či jav (napríklad

metóda liečby). Pre posúdenie efektu intervencie je potrebné porovnať objekty vystavené intervencii s objektami, ktoré jej vystavené neboli. Objekty nevystavené intervencii nazývame *kontrolná vzorka*. Práve ANOVA sa používa na určenie vplyvu tejto intervencie.

Montgomery (2017) hovorí o analýze rozptylu ako o “pravdepodobne najužitočnejšej metóde z odboru inferenčnej štatistiky”. Hojné využívanie analýzy rozptylu vo vedeckej komunite aj mimo nej však prináša so sebou riziká. Mnoho výskumníkov používa alebo interpretuje jej výsledky nesprávne. Štúdia od Wu a kol. (2011) ukázala, že v 10 popredných čínskych medicínskych časopisoch v roku 2008 sa v 446 článkoch použila analýza rozptylu. Z nich až v 56 % bola použitá nesprávne.

Použitie analýzy rozptylu predpokladá splnenie niektorých predpokladov o analyzovaných údajoch. Konkrétne ide o nezávislosť porovnávaných skupín, normalitu údajov a heteroskedasticitu (zhodu rozptylov). Porušenie týchto predpokladov má vplyv na spoľahlivosť výsledkov a vyslovené závery analýzy. V prípade nesplnenia predpokladu normality je možné použiť na posúdenie vplyvu rozdielnej úrovne faktora neparametrický Kruskal-Wallisov test. Ak je porušený predpoklad o zhode rozptylov, existuje modifikácia analýzy rozptylu, nazývaná Welchova ANOVA, ktorá berie do úvahy rozdielnu variabilitu porovnávaných skupín.

1.2 Chyby štatistických testov

Závery testov sú platné vždy len s určitou pravdepodobnosťou, preto je v štatistike zavedený pojem hladina významnosti. Obyčajne volíme 5 % hladinu významnosti. Znamená to, že ak by sme vykonali 100 náhodných výberov z tej istej populácie, o ktorej vieme, že platí nulová hypotéza, tak aplikáciou testu na každý z výberov získame maximálne 5 prípadov, kedy nulovú hypotézu zamietneme. V týchto 5 prípadoch je teda možné, že učiníme nesprávne rozhodnutie, teda spravíme chybu. Pri štatistických testoch môžu nastať 2 typy chýb, kedy sa dostaneme k chybným záverom. Prvou je chyba I. druhu, ktorá nastane, keď zamietneme hypotézu H_0 , ktorá je v skutočnosti pravdivá. Druhú voláme chyba II. druhu, ktorou je prijatie hypotézy H_0 , ktorá v skutočnosti neplatí. Tieto informácie môžeme zhrnúť do tabuľky 1.1.

Tabuľka 1.1: Rozdelenie chýb štatistických testov

| Druh chyby | Popis | Pravdepodobnosť chyby |
|-------------------|-----------------------------|------------------------------|
| Chyba I. druhu | Zamietnutie pravdivej H_0 | α |
| Chyba II. druhu | Prijatie nepravdivej H_0 | β |

Zdroj: PACÁKOVÁ a kolektív. *Štatistická indukcia pre ekonómov a manažérov*. Bratislava: Wolters Kluwer, 2015, str. 69. ISBN 978-80-8168-081-6.

Ideálna situácia by nastala, ak by sme použili taký test, ktorý minimalizuje oba druhy spomenutých chýb. Vo všeobecnosti však platí, že minimalizovaním jedného druhu chýb sa zvyšuje riziko chyby druhého druhu.

Pri použití metódy ANOVA je nutné pamätať na to, že naše údaje nebudú vždy dokonale spĺňať jej predpoklady. Vlastnosť štatistického testu byť citlivý na porušenie jeho predpokladov sa nazýva *robustnosť*. Mali by sme preto vedieť, v akej miere je použitý test citlivý na takéto porušenia. Ak je test robustný, znamená to, že miera chýb I. druhu (pravdepodobnosť zamietnutia pravdivej nulovej hypotézy) sa s porušením týchto predpokladov výrazne nemení.

Robustnosť sa vzťahuje na chyby I. druhu, môže však nastať aj situácia, keď dôjde k chybe II. druhu (nezamietne sa nepravdivá nulová hypotéza). *Sila testu* je číslo, ktoré udáva pravdepodobnosť, že pri neplatnosti nulovej hypotézy dôjde k jej zamietnutiu. To znamená, že silu testu môžeme vyjadriť aj pomocou miery chýb II. druhu ako

$$1 - \text{miera chýb II. druhu} = 1 - \beta. \quad (1.1)$$

Vo všeobecnosti chceme, aby mal test čo najvyššiu silu. Tá je ovplyvnená viacerými okolnosťami. Okrem typu zvoleného testu je to najmä vzorka reprezentujúca skúmanú populáciu a jej veľkosť. Vo vedeckých štúdiách sa preto odporúča pri použití konkrétneho testu zvoliť rozsah výberu tak, aby bola zaručená minimálna požadovaná sila testu.

Na skúmanie sily štatistických testov a ich robustnosti sa často používajú simulácie, takzvané Monte Carlo metódy. Sú založené na myšlienke, že poznatky o analyzovanej charakteristike môžeme získať mnohonásobným vykonaním náhodného vyberania z toho istého rozdelenia alebo populácie a následným pozorovaním, ako sa hodnota tejto charakteristiky mení medzi výbermi. Opakovaný náhodný výber môžeme uskutočniť pomerne jednoducho a rýchlo. Pomocou štatistických softvérov a programovacích jazykov

ako SAS, R či Python vieme získať náhodný výber zo zvoleného rozdelenia so špecifikovanými parametrami a opakovať ho v rádoch tisícov aj viac.

Ak pomocou Monte Carlo metódy odhadujeme mieru chýb I. druhu, tak generujeme náhodné výbery z populácie, v ktorej platí nulová hypotéza. To v prípade analýzy rozptylu znamená, že generujeme údaje, kde majú skupiny s rôznou úrovňou faktora rovnakú strednú hodnotu. Následne vykonáme štatistické testovanie na každom vygenerovanom náhodnom výbere. Mieru chýb I. druhu dostaneme, ak spočítame, v koľkých prípadoch test nesprávne zamietol nulovú hypotézu o zhode stredných hodnôt, a toto číslo vydáme celkovým počtom vykonaných testov.

Pri určovaní sily testu postupujeme obdobne. Generujeme náhodné výbery z populácie, v ktorej neplatí nulová hypotéza. Nulovú hypotézu overíme štatistickým testom v každom vygenerovanom súbore. Mieru chýb II. druhu vypočítame ako podiel testov, ktoré nesprávne prijali nulovú hypotézu, a celkového počtu vykonaných testov. Silu testu zistíme, ak mieru chýb II. druhu odčítame od 1.

2 Cieľ práce

V diplomovej práci sa zaoberáme problematikou vzájomného porovnávania viacerých výberových súborov a ich parametrov. Hlavným cieľom práce je porovnanie klasickej verzie analýzy rozptylu a Welchovej analýzy rozptylu z hľadiska robustnosti a sily testu a objasnenie, pri akých parametroch údajov je vhodnejšie použiť Welchovu analýzu rozptylu, a kedy klasickú analýzu rozptylu. Na splnenie hlavného cieľu práce boli zadané nasledujúce čiastkové ciele:

- Objasniť metódu analýzy rozptylu, jej teoretické základy, postup testovania.
- Vysvetliť aký vplyv má porušenie predpokladov použitia analýzy rozptylu na kvalitu a spoľahlivosť jej výsledkov, a aké metódy je vhodné použiť v prípade nesplnenia týchto predpokladov.
- Popísať použitie Monte Carlo simulácií na odhad mier I. a II. druhu štatistických testov.
- Aplikovať analýzu rozptylu na posúdenie vplyvu vybranej premennej v počítačových zostavách na ich predajnú cenu. Porovnávať tu výsledky testovania dvoch rôznych variantov analýzy rozptylu – jej klasickej verzie a Welchovej analýzy rozptylu.
- Pomocou simulačných metód Monte Carlo posúdiť miery chýb I. a II druhu dvoch spomenutých testov.
- Posúdiť aký je vplyv heteroskedasticity vstupných údajov na robustnosť a silu testov.

3 Metodika práce

3.1 Rozdiel medzi analýzou rozptylu a dvojvýberovým t-testom

Predtým, než začneme vysvetľovať princípy, na ktorých je ANOVA založená, pokúsime sa objasniť, prečo na zisťovanie rozdielov medzi skupinami nie je vhodné použiť dvojvýberový t-test. Týmto testom zisťujeme, či sa 2 skupiny odlišujú v priemernej hodnote určitého znaku. Hypotézy testu zapíšeme nasledovne:

$$\begin{aligned}H_0: \mu_1 &= \mu_2 \\ H_1: \mu_1 &\neq \mu_2.\end{aligned}\tag{3.1}$$

Prvý problém je v tom, že ak chceme porovnávať viac ako 2 rôzne skupiny pozorovaní, pre každú dvojicu skupín musíme overiť zhodu stredných hodnôt zvlášť. Uvažujme situáciu, keď máme navzájom porovnávať nie dve, ale tri skupiny (a, b, c) a chceme zistiť, či medzi týmito skupinami existuje štatisticky významný rozdiel v priemernej úrovni sledovaného znaku. Chceme vykonať t-test na každom páre skupín, musíme preto postupne porovnať skupiny a a b , a a c a nakoniec b a c . Ak zvýšime počet skupín v našej analýze, narastie aj počet párov skupín, za ktoré budeme musieť vykonať jednotlivé t-testy (Field, 2017). Počet týchto testov môžeme vypočítať ako počet kombinácií druhej triedy z n prvkov:

$$C_2(n) = \frac{n!}{k!(n-k)!}.\tag{3.2}$$

Ďalším problémom pri takomto viacnásobnom štatistickom testovaní je *miera chýb I. druhu* (α), teda miera zamietnutia nulovej hypotézy, ktorá je pravdivá. Naša nulová hypotéza v t-teste hovorí o rovnosti stredných hodnôt medzi skupinami, chyba I. druhu bude mať preto za následok nepravdivé zamietnutie tejto rovnosti. Niekedy sa chybe I. druhu hovorí *chyba nadmerného skepticizmu* (Shermer, 2002).

Miera chýb I. druhu nám hovorí chybách len v jednom teste, nemôžeme ju preto v tejto podobe použiť pri viacnásobnom testovaní. Použijeme *skupinovú mieru chýb I. druhu*, v angličtine „*familywise error rate*“ (FWER). Tá určuje pravdepodobnosť vyskytnutia aspoň jednej chyby prvého druhu v n testoch:

$$FWER = P(A > 0),\tag{3.3}$$

kde A je počet chýb I. druhu v n vykonaných testoch (Zar, 2010).

Takto definovaná FWER je výhodná v prípade, keď potrebujeme silný dôkaz alebo počet testov n je malý. Pri väčšine prípadov je však takáto podmienka príliš prísna. Preto je používanější takzvaná *zovšeobecnená FWER* (vzťah 3.4), ktorá pripúšťa určené množstvo (k) chýb I. druhu vo všetkých testoch (Zar, 2010).

$$FWER = P(A > k), k \in \mathbb{N}. \quad (3.4)$$

Ak by sme porovnávali 3 skupiny a pri každom teste by sme uvažovali s 5% hladinou významnosti, tak pravdepodobnosť chyby I. druhu v každom teste by bola 5%. Z toho vyplýva, že pravdepodobnosť toho, že sa nevyskytne ani jedna chyba I. druhu je v každom teste 95%. Ak uvažujeme, že testy sú navzájom nezávislé (takže pravdepodobnosti môžeme navzájom násobiť), tak celková pravdepodobnosť, že v ani jednom teste nie je chyba I. druhu, je $0,95^3$, čo je 0,857 %. Z toho vyplýva, že $FWER = 1 - 0,857$. Dokopy je tak vo všetkých testoch miera chýb I. druhu 14,3 %, čo je vyššie ako zvolená hladina významnosti 5% (Field, 2017).

Ak by sme porovnávali ešte viac skupín, počet vykonaných testov by sa zvýšil, a s ním takisto aj miera chýb I. druhu. Z tohto dôvodu nie sú t-testy vhodné na skupinové porovnávanie.

3.2 Dvojvýberový F-test zhody rozptylov

Videli sme, prečo nie sú obyčajné párové t-testy vhodné na porovnávanie stredných hodnôt viacerých súborov. Test používaný pri overovaní predpokladu homoskedasticity v analýze rozptylu je takzvaný dvojvýberový F-test zhody rozptylov. Ten porovnáva rozptyly dvoch nezávislých výberov za predpokladu normálneho rozdelenia. Jeho hypotézy zapíšeme ako

$$\begin{aligned} H_0: \sigma_X^2 &= \sigma_Y^2, \\ H_1: \sigma_X^2 &> \sigma_Y^2, \end{aligned} \quad (3.5)$$

kde σ_X^2 je rozptyl základného súboru, X_1, X_2, \dots, X_n a σ_Y^2 je rozptyl základného súboru Y_1, Y_2, \dots, Y_m .

Testovaciu štatistiku vypočítame ako

$$F = \frac{\tilde{s}_X^2}{\tilde{s}_Y^2}, \quad (3.6)$$

kde \tilde{s}_X^2 predstavuje výberový rozptyl výberu X_1, X_2, \dots, X_n a \tilde{s}_Y^2 je výberový rozptyl výberu Y_1, Y_2, \dots, Y_m .

Hypotézu H_0 zamietneme, ak sa výberové rozptyly príliš líšia, teda pokiaľ sa hodnota testovacej štatistiky bude nachádzať v kritickej oblasti

$$F \geq F_{1-\alpha}(n-1; m-1), \quad (3.7)$$

kde $F_{1-\alpha}(n-1; m-1)$ je $(1-\alpha)$ -kvantil Fisherovho rozdelenia s $(n-1)$ a $(m-1)$ stupňami voľnosti (Pacáková a kol., 2015).

V ďalšej časti v ktorej vysvetľujeme princíp analýzy rozptylu ukážeme, ako sa tento test aplikuje na overovanie zhody stredných hodnôt viac ako dvoch súborov.

3.3 Analýza rozptylu

Analýza rozptylu, tak ako si ju predstavujeme a používame dnes, bola vyvinutá Ronaldom A. Fisherom, keď sa snažil o lepšie vyhodnocovanie dát z oblasti poľnohospodárstva. Práve on do štatistiky zaviedol pojmy *rozptyl* a *analýza rozptylu*. Henry Scheffé ponúka približnú definíciu analýzy rozptylu vo svojej knihe, príznačne pomenovanej *Analýza rozptylu*: „Analýza rozptylu je štatistická technika určená na analyzovanie meraní závislých na rôznych druhoch vzájomne pôsobiacich faktorov tak, aby bolo možné rozhodnúť sa, ktoré z faktorov sú dôležité, a odhadnúť ich vplyv.“ Jej podstatou je určiť, či existujú štatisticky významné rozdiely medzi strednými hodnotami závislej premennej v skupinách, ktoré majú odlišné obmeny (úrovne, kategórie) nezávislej premennej. To sa docieli rozkladom variability závislej premennej na zložku variability medzi skupinami, ktoré sú tvorené jednotlivými úrovňami faktora, a na zložku variability vnútri spomínaných skupín, ktorá je spôsobená inými činiteľmi a nie sledovaným faktorom.

3.3.1 Zápis modelu

Uvažujme, že bolo uskutočnených k nezávislých výberov, pochádzajúcich z k základných súborov, ktoré majú normálne rozdelenie

$$\begin{aligned} Y_{11}, \dots, Y_{1n_1} &\sim N(\mu_1, \sigma_1^2) \\ Y_{21}, \dots, Y_{2n_2} &\sim N(\mu_2, \sigma_2^2), \\ &\dots \\ Y_{k1}, \dots, Y_{kn_k} &\sim N(\mu_k, \sigma_k^2). \end{aligned} \quad (3.8)$$

Môžeme taktiež uvažovať o k podmnožinách jedného základného súboru s normálnym pravdepodobnostným rozdelením. Pričom pre rozsahy výberových súborov platí, že $n_1 \geq 2, n_2 \geq 2, \dots, n_k \geq 2$ a pre ich rozptyly platí $\sigma_1^2 \geq 0, \sigma_2^2 \geq 0, \dots, \sigma_k^2 \geq 0$. Ďalej predpokladáme, že všetky náhodné výbery majú rovnaký rozptyl σ^2 . Nie sú známe stredné hodnoty zo základných súborov $\mu_1, \mu_2, \dots, \mu_k$, ani hodnoty rozptylov základných súborov $\sigma_1^2, \sigma_2^2, \dots, \sigma_k^2$.

Ak máme n_i pozorovaní v každej z k skupín (populácií), celkový počet pozorovaní je $n (= \sum_{i=1}^k n_i)$. Potom y_{ij} vyjadruje i -te pozorovanie v j -tej skupine.

Výberový priemer závislej premennej v j -tej skupine \bar{y}_j vyjadríme vzťahom

$$\bar{y}_j = \frac{1}{n} \sum_{i=1}^{n_j} y_{ij}. \quad (3.9)$$

Celkový priemer $\bar{y}_..$ závislej premennej, za hodnoty zo všetkých výberov vyjadríme ako

$$\bar{y}_.. = \frac{1}{n} \sum_{i=1}^k \sum_{j=1}^{n_i} y_{ij} \quad (3.10)$$

(Pacáková a kol., 2012).

Údaje získané z výberového zisťovania potom môžeme zapísať vo forme nasledujúcej tabuľky.

Tabuľka 3.1: Zápis modelu jednofaktorovej analýzy rozptylu

| Obmena faktora | Zistené hodnoty | Rozsah výberového súboru | Súčet hodnôt v riadku | Priemer |
|----------------|-----------------------------------|----------------------------|---|--------------------|
| 1 | $y_{11}, y_{12}, \dots, y_{1n_1}$ | n_1 | $y_{1\cdot} = \sum_{j=1}^{n_1} y_{1j}$ | $\bar{y}_{1\cdot}$ |
| 2 | $y_{21}, y_{22}, \dots, y_{2n_2}$ | n_2 | $y_{2\cdot} = \sum_{j=1}^{n_2} y_{2j}$ | $\bar{y}_{2\cdot}$ |
| \vdots | \vdots | \vdots | \vdots | \vdots |
| i | $y_{i1}, y_{i2}, \dots, y_{in_i}$ | n_i | $y_{i\cdot} = \sum_{j=1}^{n_i} y_{ij}$ | $\bar{y}_{i\cdot}$ |
| \vdots | \vdots | \vdots | \vdots | \vdots |
| k | y_{k1}, \dots, y_{kn_k} | n_k | $y_{k\cdot} = \sum_{j=1}^{n_k} y_{kj}$ | $\bar{y}_{k\cdot}$ |
| Spolu | | $n = \sum_{j=1}^{n_i} n_i$ | $\bar{y}_{..} = \sum_{i=1}^k \sum_{j=1}^{n_i} y_{ij}$ | $\bar{y}_{..}$ |

Zdroj: PACÁKOVÁ, Viera, Viera LABUDOVÁ, Ľubica SIPKOVÁ a Iveta STANKOVIČOVÁ. *Štatistická indukcia pre ekonómov a manažérov*. Bratislava: Wolters Kluwer, 2015, s 140. ISBN 978-80-8168-081-6.

3.3.2 Zápis štatistických hypotéz

Chceme zistiť, či stredná hodnota závislej premennej je rovnaká pre všetkých k obmien skúmaného faktora. Testujeme teda, či sa stredné hodnoty závislej premennej rovnajú v jednotlivých k skupinách

$$H_0: \mu_1 = \mu_2 = \dots = \mu_k, \quad (3.11)$$

oproti alternatívnej hypotéze

$$H_1: \text{non } H_0, \quad (3.12)$$

teda, že aspoň jedna z k stredných hodnôt je rozdielna.

Aj keď existuje konečný počet faktorov ovplyvňujúci jav, ktorý vo výskume sledujeme, musíme si uvedomiť, že budeme schopní merať iba určitý počet týchto faktorov. Zvyšok buď nebudeme môcť pozorovať a merať, alebo ich jednoducho nezohľadníme. To je aj podstatou analýzy rozptylu. Celkový rozptyl údajov vstupujúcich do analýzy rozdelíme na objasnenú časť (známe zdroje variability) a neobjasnenú časť,

ktorú považujeme za náhodnú (tu sú zahrnuté aj faktory, ktoré nie sú náhodné, ale v našom modeli od nich abstrahujeme).

Keď usporiadame hodnoty $y_{11}, \dots, y_{1n_1}, y_{21}, \dots, y_{2n_2}, y_{k1}, \dots, y_{kn_k}$ do vektorov, tak úlohu môžeme zapísať pomocou matíc ako

$$\begin{bmatrix} y_{11} \\ \vdots \\ y_{1n_1} \\ y_{21} \\ \vdots \\ y_{k1} \\ \vdots \\ y_{kn_k} \end{bmatrix} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & 1 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_k \end{bmatrix} + \begin{bmatrix} \varepsilon_{11} \\ \vdots \\ \varepsilon_{1n_1} \\ \varepsilon_{21} \\ \vdots \\ \varepsilon_{k1} \\ \vdots \\ \varepsilon_{kn_k} \end{bmatrix}, \quad (3.13)$$

kde ε_{ij} predstavuje náhodnú chybu, ktorá má normálne rozdelenie $N(0, \sigma^2)$. Náhodné chyby ε_{ij} sú nezávislé pre $i = 1, 2, \dots, k; j = 1, 2, \dots, n_i$ (Parellada, 2015).

Hodnotu závislej premennej y_{ij} potom zapíšeme ako

$$y_{ij} = \mu_i + \varepsilon_{ij}, \quad (3.14)$$

kde μ_i predstavuje strednú hodnôu závislej premennej s i -tou úrovňou faktora a ε_{ij} predstavuje náhodnú chybu, ktorá má normálne rozdelenie $N(0, \sigma^2)$. Náhodné chyby ε_{ij} sú nezávislé pre $i = 1, 2, \dots, k; j = 1, 2, \dots, n_i$.

Keďže náhodné chyby ε_{ij} majú strednú hodnotu $E(\varepsilon_{ij}) = 0$, tak pre všetky obmeny faktora je stredná hodnota závislej premennej y_{ij} rovnaká

$$E(y_{ij}) = \mu_i. \quad (3.15)$$

Ak má μ_i konštantnú hodnotu, tak môžeme povedať, že rozptyl hodnôt okolo strednej hodnoty je pre každú obmenu faktoru rovnaký

$$D(y_{ij}) = D(\varepsilon_{ij}) = \sigma^2. \quad (3.16)$$

Efekt i -tej obmeny faktora zapíšeme ako α_i . Potom stredná hodnota i -tej obmeny faktora bude

$$\mu_i = \mu + \alpha_i. \quad (3.17)$$

Zo vzťahu 3.17 potom môžeme zapísať hodnotu y_{ij} závislej premennej ako

$$y_{ij} = \mu + \alpha_i + \varepsilon_{ij}. \quad (3.18)$$

V prípade, že faktor nepôsobí na hodnoty závislej premennej, platí $\alpha_1 = \alpha_2 = \dots = \alpha_k = 0$. Tým pádom môžeme model zapísať ako

$$y_{ij} = \mu + \varepsilon_{ij}. \quad (3.19)$$

Odhadmi stredných hodnôt pre i -tu úroveň faktora, budú výberové priemery $\hat{\mu}_i = \bar{y}_{j.}$, podľa vzťahu 3.9 a odhadom strednej hodnoty modelu bude celkový priemer za všetky výbery $\hat{\mu} = \bar{y}_{..}$, podľa vzťahu 3.10.

Reziduálnu zložku ε_{ij} odhadneme ako $\varepsilon_{ij} = y_{ij} - \hat{y}_{ij}$, Čo zo vzťahu 3.15 môžeme zapísať ako $\varepsilon_{ij} = y_{ij} - \mu_i$. Náhradou μ_i jeho bodovým odhadom z predošlého odstavca, čo predstavuje výberový priemer pre i -tu obmenu faktora, dostávame vzťah

$$\hat{\varepsilon}_{ij} = y_{ij} - \bar{y}_{j.}. \quad (3.20)$$

Dosaďme do vzťahu 3.18 namiesto strednej hodnoty celého modelu μ jeho odhad $\bar{y}_{..}$. Súčasne zapíšme efekt i -tej obmeny faktora α_i ako $y_{ij} - \bar{y}_{j.}$ ako vo vzťahu 3.20 a reziduálnu zložku ε_{ij} ako $y_{ij} - \mu_i$:

$$y_{ij} = \bar{y}_{..} + (y_{ij} - \bar{y}_{..}) + (y_{ij} - \bar{y}_{j.}). \quad (3.21)$$

Presunom celkového priemeru $\bar{y}_{..}$ na druhú stranu dostávame na ľavej strane vzťah pre celkovú odchýlku, a na strane pravej dostávame súčet odchýlky i -tej obmeny faktora a reziduálnej zložky

$$y_{ij} - \bar{y}_{..} = (\bar{y}_{j.} - \bar{y}_{..}) + (y_{ij} - \bar{y}_{j.}). \quad (3.22)$$

Po úprave umocnením a sčítaním všetkých pozorovaní naprieč skupinami dostávame

$$\begin{aligned} & \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{..})^2 \\ &= \sum_{i=1}^k \sum_{j=1}^{n_i} (\bar{y}_{j.} - \bar{y}_{..})^2 + \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{j.})^2 + 2 \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{j.})(\bar{y}_{j.} - \bar{y}_{..}) \end{aligned} \quad (3.23)$$

Vzťah 1.13 môžeme ďalej zjednodušiť a odstrániť posledný člen, keďže platí

$$\sum_{j=1}^{n_i} (y_{ij} - \bar{y}_j)(\bar{y}_j - \bar{y}_{..}) = \sum_{j=1}^{n_i} \bar{y}_j (y_{ij} - \bar{y}_j) - \sum_{j=1}^{n_i} \bar{y}_{..} (y_{ij} - \bar{y}_j) = 0, \quad (3.24)$$

pretože

$$\sum_{j=1}^{n_i} \bar{y}_j \varepsilon_{ij} = 0 \quad (3.25)$$

a zároveň

$$\sum_{j=1}^{n_i} \varepsilon_{ij} = (y_{ij} - \bar{y}_j) = 0. \quad (3.26)$$

Potom platí:

$$\sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{..})^2 = \sum_{j=1}^{n_i} (\bar{y}_j - \bar{y}_{..})^2 n_i + \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_j)^2. \quad (3.27)$$

Členy tohto vzťahu predstavujú variabilitu modelu rozloženú na jednotlivé zložky (Pacáková a kol., 2012).

Celkový súčet štvorcov (sum of squares total) všetkých pozorovaní sa vypočíta ako celková suma štvorcov takto:

$$SST = \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{..})^2 \quad (3.28)$$

Tento súčet môžeme rozdeliť na dve zložky:

1. *Medziskupinový súčet štvorcov (sum of squares of treatment):*

$$SSA = \sum_{j=1}^{n_i} n_i (\bar{y}_j - \bar{y}_{..})^2. \quad (3.29)$$

Predstavuje časť variability, ktorá je vysvetlená rozdielmi v skupinách s rozdielnymi obmenami faktora.

2. *Vnútroskupinový súčet štvorcov (sum of squares of error):*

$$SSE = \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2. \quad (3.30)$$

Predstavuje variabilitu, ktorá je vnútri skupiny s príslušnou obmenou faktora. Nazývame ju aj nevysvetlená variabilita (Pacáková a kol., 2012).

Vzťah 3.27 teda môžeme definovanou symbolikou zapísať skrátené ako $SST = SSA + SSE$.

3.3.3 Určenie stupňov voľnosti

Pri overovaní hypotézy, že stredné hodnoty pre všetky obmeny faktora μ_i sa rovnajú, budeme porovnávať priemerné hodnoty štvorcov medziskupinových odchýlok oproti priemerným hodnotám štvorcov odchýlok náhodnej zložky. Na to, aby sme dostali tieto priemerné hodnoty, musíme predeliť spomínané odchýlky príslušným počtom *stupňov voľnosti*. Tento počet určíme ako rozdiel medzi počtom pozorovaní vo výberovom súbore a počtom štatistík, ktoré boli zo súboru vypočítané. Inak povedané, je to počet nezávislých sčítancov, z ktorých sme vypočítali sumu ich štvorcov.

Pri vnútroskupinovom súčte štvorcov SSE uvažujeme odchýlky y_{ij} od \bar{y}_i . Pri výpočte každého \bar{y}_i bolo použitých $n_i - 1$ nezávislých sčítancov. Pri k sčítancoch bude počet stupňov voľnosti SSE

$$df_{SSE} = \sum_{i=1}^k (n_i - 1) = n - k. \quad (3.31)$$

Medziskupinový súčet štvorcov SSA bol vypočítaný z k súborov, takže obsahuje $k - 1$ nezávislých sčítancov

$$df_{SSA} = k - 1. \quad (3.32)$$

Pri výpočte celkového priemeru za všetky výbery bolo pri výpočte použitých n pozorovaní y_{ij} , takže počet nezávislých sčítancov, z ktorých bol vytvorený celkový súčet štvorcov SST, je

$$df_{SST} = n - 1. \quad (3.33)$$

Všimnime si, že tak ako celkový súčet štvorcov SST je súčtom SSA a SSE, takisto aj súčet stupňov voľnosti SSA a SSE sa rovná počtu stupňov voľnosti SST

$$n - 1 = (n - k) + (k - 1). \quad (3.34)$$

Potom môžeme definovať priemer štvorcov pre zdroje rozptylu ako sumu štvorcov delenú príslušnými stupňami voľnosti, ktoré sme v predošlej časti určili:

$$MSA = \frac{SSA}{k - 1}, \quad (3.35)$$

$$MSE = \frac{SSE}{n - k}. \quad (3.36)$$

Ukázali sme ako celkovú variabilitu údajov SST môžeme rozložiť na SSA a SSE. Priemer \bar{y}_i je odhadom μ_i a $\bar{y}_.$ je odhadom celkovej strednej hodnoty μ , preto pri platnosti hypotézy H_0 bude SSA malé vzhľadom na SSE. Ak je SSA veľké vzhľadom na SSE, znamená to, že priemery jednotlivých skupín sa od seba významne odlišujú a môžeme zamietnuť hypotézu o rovnosti stredných hodnôt. Testovacia štatistika bude teda porovnávať rozptyl výberových priemerov SSA a rozptyl vnútri jednotlivých skupín SSE. Všimnime si, že testovacia štatistika má tvar dvojvýberového F-testu (3.6), ktorý porovnáva rozptyly dvoch súborov.

$$F = \frac{\frac{SSA}{k - 1}}{\frac{SSE}{n - k}}. \quad (3.37)$$

H_0 teda zamietneme ak

$$F \geq F_{1-\alpha}(k - 1; n - k), \quad (3.38)$$

kde $F_{1-\alpha}(k - 1; n - k)$ je $(1 - \alpha)$ -kvantil Fisherovho rozdelenia s $(k - 1)$ a $(n - k)$ stupňami voľnosti (Pacáková a kol., 2012).

3.3.4 Tabuľka analýzy rozptylu

Výsledky analýzy rozptylu sa veľmi často uvádzajú vo forme nasledujúcej tabuľky:

Tabuľka 3.2: Tabuľka výsledkov analýzy rozptylu

| Zdroj variability | Súčet štvorcov odchýlok | Stupne voľnosti | Priemerný súčet štvorcov | Testovacia štatistika F |
|-------------------|---|-----------------|---------------------------|-------------------------|
| Faktor | $SSA = \sum_{j=1}^{n_i} n_i (\bar{y}_i - \bar{y}_{..})^2$ | $k - 1$ | $MSA = \frac{SSA}{k - 1}$ | $\frac{MSA}{MSE}$ |
| Náhodné vplyvy | $SSE = \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2$ | $n - k$ | $MSE = \frac{SSE}{n - k}$ | |
| Spolu | $SST = \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{..})^2$ | $n - 1$ | — | — |

Zdroj: KOTLEBOVÁ, Eva, Erik ŠOLTÉS a Daniela SIVAŠOVÁ. *Štatistická indukcia v príkladoch*. Bratislava: Vydavateľstvo EKONÓM, 2015, s. 97. ISBN 978-80-225-4135-0.

3.4 Welchova analýza rozptylu

Klasická ANOVA, využívajúca bežný F-test, ktorý je vysvetlený v kapitole 3.2, predpokladá, že porovnávané skupiny majú zhodné rozptyly, t.j. sú homoskedastické. Z tohto dôvodu nie je optimálne používať klasickú ANOVU pri údajoch, ktoré nespĺňajú podmienku homoskedasticity.

Welchov F-test (1951) je určený na testovanie rovnosti vnútroskupinových priemerov, hlavne v prípadoch, keď nie je splnená jedna z podmienok použitia klasického F-testu a to homogenita rozptylu. Zovšeobecňuje testovaciu štatistiku F (3.6) a aproximuje jej rozdelenie pravdepodobnosti, spolu s použitými stupňami voľnosti.

Welch spolu s Aspinom už skôr prišli s aproximatívnym riešením overovania zhody stredných hodnôt v prípade rozdielnych rozptylov pre dve porovnávané skupiny ($k = 2$). Welch sa snažil zistiť, či táto metóda aproximácie bude mať využitie aj pre všeobecnejší problém testovania zhody stredných hodnôt viac ako dvoch súborov ($k > 2$). Vo svojej práci (1951) ukazuje, ako sa dá v tomto prípade odvodiť testovacia štatistika F-testu a príslušné počty stupňov voľnosti.

Testovacia štatistika Welchovho F-testu obsahuje váhu w_i . Hodnota tejto váhy závisí od početnosti i -teho porovnávaného súboru n_i a súčasne od príslušného rozptylu s_i . Váha sa vypočíta ako

$$w_i = \frac{n_i}{s_i}. \quad (3.39)$$

Tieto váhy následne použijeme pri výpočte upraveného celkového priemeru \bar{y}_{WELCH}

$$\bar{y}_{Welch} = \frac{\sum_{i=1}^k w_i \bar{y}_{i.}}{\sum_{i=1}^k w_i}, \quad (3.40)$$

kde $\bar{y}_{i.}$ je priemer k-teho súboru

$$\bar{y}_{i.} = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij}. \quad (3.41)$$

Takisto ako pri klasickej analýze rozptylu, aj v tomto prípade celkový súčet štvorcov (SST) rozložíme na medziskupinový súčet štvorcov (SSA) a vnútroskupinový súčet štvorcov (SSE). V SSA použijeme na výpočet medziskupinových odchýlok upravený celkový priemer \bar{y}_{Welch} . Druhá mocnina týchto odchýlok je ešte prevážaná váhou w_k .

$$SSA_{Welch} = \sum_{i=1}^k w_i (\bar{y}_{i.} - \bar{y}_{Welch})^2 \quad (3.42)$$

Priemer štvorcov odchýlok potom vypočítame ako

$$MSA_{Welch} = \frac{SSA_{Welch}}{k-1}. \quad (3.43)$$

Testovaciu štatistiku definujeme ako

$$F_{Welch} = \frac{\frac{SSA_{Welch}}{k-1}}{1 + \frac{2\Lambda(k-2)}{3}}, \quad (3.44)$$

kde výraz Λ vypočítame ako

$$\Lambda = \frac{3 \sum_{i=1}^k \frac{(1 - \frac{w_i}{\sum_{i=1}^k w_i})}{n_i - 1}}{k^2 - 1}. \quad (3.45)$$

Štatistika F_{Welch} pochádza z Fisherovho rozdelenia so stupňami voľnosti:

$$df_1 = (k-1); df_2 = \frac{1}{\Lambda}. \quad (3.46)$$

Testovaná hypotéza je taká istá, ako v prípade klasickej analýzy rozptylu. Budeme teda overovať, či platí, že všetky stredné hodnoty porovnávaných súborov sú si rovné, t.j. $\mu_i = \mu$ ($i = 1, 2, \dots, k$)

$$H_0: \mu_1 = \mu_2 = \dots = \mu_k, \quad (3.47)$$

oproti alternatívnej hypotéze, že aspoň dve stredné hodnoty sú rozdielne

$$H_1: \text{non } H_0. \quad (3.48)$$

H_0 zamietame ak sa testovacia štatistika F nachádza v kritickej oblasti

$$F \geq F_{1-\alpha}(k-1; \frac{1}{\Lambda}), \quad (3.49)$$

kde $F_{1-\alpha}(k-1; \frac{1}{\Lambda})$ je $(1-\alpha)$ -kvantil Fisherovho rozdelenia s $(k-1)$ a $(\frac{1}{\Lambda})$ stupňami voľnosti.

Pri porušení predpokladu o homoskedasticite porovnávaných súborov je Welchova ANOVA spoľahlivejšou alternatívou analýzy rozptylu. Jeho testovacia štatistika má menej stupňov voľnosti, preto sa považuje za menej silný test. Pozor by sme si mali dávať, najmä v prípade malého počtu pozorovaní, pretože váhy w_i v testovacej štatistike závisia od početnosti súboru a jeho rozptylu. Výsledky v tom prípade môžu byť nestabilné (Hangcheng, Liu, 2015).

3.5 Kruskal-Wallisov test

Kruskal Wallisov poradový neparametrický test možno použiť v prípadoch, keď chceme zistiť, či rozdielne úrovne faktora medzi porovnávanými skupinami vplývajú na stredné hodnoty skúmanej závislej premennej Y , ale údaje nepochádzajú z normálneho rozdelenia, teda je porušená podmienka použitia analýzy rozptylu.

Podmienkou jeho využitia je, že porovnávané výberové súbory sú navzájom nezávislé, sledovaná náhodná premenná X je spojitá a stupnica merania hodnôt premennej X je najmenej ordinálna. Možno ho použiť aj v prípade, ak je stupnica hodnôt výberového súboru intervalová, ale budeme k nej pristupovať ako keby bola ordinálna. Aj keď pri Kruskal-Wallisovom teste nemusí byť splnený predpoklad normality, podmienka homoskedasticity by mala byť splnená (McDonald, 2014).

Nulová hypotéza Kruskal-Wallisovho testu tvrdí, že mediány v skupinách sú zhodné

$$H_0: \tilde{X}_1 = \tilde{X}_2 = \dots = \tilde{X}_k; k \geq 3, \quad (3.50)$$

oproti alternatíve, že mediány aspoň dvoch skupinách nie sú zhodné

$$H_1: \text{non } H_0. \quad (3.51)$$

Môžeme sa stretnúť aj s alternatívnym znením nulovej hypotézy, že výberové súbory pochádzajú zo základného súboru s rovnakým rozdelením. To znamená, že poloha rozdelení vo všetkých k podsúboroch je rovnaká, oproti alternatívnej hypotéze, že aspoň dve rozdelenia sa líšia polohou. Ak sú rozdelenia vo všetkých súboroch rovnaké, potom platí v k súboroch rovnosť všetkých charakteristík.

Zoradíme všetky n hodnôt x_1, x_2, \dots, x_n z k výberových súborov do jedného súboru s rozsahom $n = n_1 + n_2 + \dots + n_k$ a tieto hodnoty usporiadajme do neklesajúcej postupnosti

$$x_i: x_1, x_2, \dots, x_n \rightarrow x_{(i)}: x_{(1)}, x_{(2)}, \dots, x_{(n)}, \quad (3.52)$$

kde

$$x_{(i-1)} \leq x_{(i)} \leq x_{(i+1)}; i = 1, 2, \dots, n. \quad (3.53)$$

Hodnotám v spojenom súbore priradíme poradia, kde poradie 1 bude priradené najmenej zistenej hodnote a poradie n bude priradené najväčšej hodnote. Ak nastane situácia, keď budú dve hodnoty rovnaké, tak obom priradíme priemerné poradie. V prípade platnosti nulovej hypotézy by poradia mali byť rovnomerne rozdelené do k výberových súborov (Pacáková a kol., 2015).

Rovnomernosť rozdelenia poradí medzi výberovými súbormi budeme posudzovať podľa súčtu určených poradí v každom výberovom súbore. Tento súčet označíme T_j , pre j -ty výberový súbor. Testovacia štatistika má tvar

$$KW = \left[\frac{12}{n(n+1)} \sum_{j=1}^k \frac{T_j^2}{n_j} \right] - 3(n+1). \quad (3.54)$$

Ak sa v postupnosti vyskytnú priemerné poradia, vypočítame koeficient korekcie

$$c = 1 - \frac{\sum_{i=1}^p (h_i^3 - h_i)}{n^3 - n}, \quad (3.55)$$

kde p je počet skupín s priemerným poradím a h_i je počet priemerných poradí v i -tej skupine. Upravenú testovaciu štatistiku KW' dostávame po predelení koeficientom c :

$$KW' = \frac{KW}{c}. \quad (3.56)$$

Ak sú podiely súčtov poradí v jednotlivých k výberových súboroch len nevýznamne rozdielne vzhľadom na rozsah podsúborov, bude mať testovacia štatistika KW malú hodnotu. Ak sú ale naopak rozdiely v podieloch súčtov poradí významne veľké, štatistika bude nadobúdať veľkú hodnotu.

Pri splnení podmienok $k \geq 4$ a $n_i \geq 5$, má testovacia štatistika χ^2 -rozdelenie s $k - 1$ stupňami voľnosti. Pri nesplnení týchto podmienok musíme použiť špeciálne tabuľky pre Kruskal-Wallisov test (Kotlebová a kol., 2015). Nulovú hypotézu zamietame, ak sa hodnota testovacej štatistiky KW nachádza v kritickej oblasti

$$KW \geq \chi_{\alpha}^2(k - 1), \quad (3.57)$$

kde $\chi_{\alpha}^2(k - 1)$ je α -kvantil χ^2 rozdelenia s $k - 1$ stupňami voľnosti.

Pri Kruskal-Wallisovom teste pracujeme len s poradiami, a tak strácame časť informácií z pôvodných hodnôt, čo môže spôsobiť, že miera chýb II. druhu je väčšia (sila testu je menšia) ako pri parametrickej analýze rozptylu.

3.6 Predpoklady použitia analýzy rozptylu

Dvojvýberový F-test zhody rozptylov a teda aj ANOVA, ktorá z neho vychádza, vyžaduje splnenie týchto predpokladov:

- Populácie, z ktorých pochádzajú porovnávané výberové súbory, majú normálne rozdelenie.
- Porovnávané výberové súbory sú navzájom nezávislé.
- Všetky základné súbory majú zhodné rozptyly (podmienka homoskedasticity).

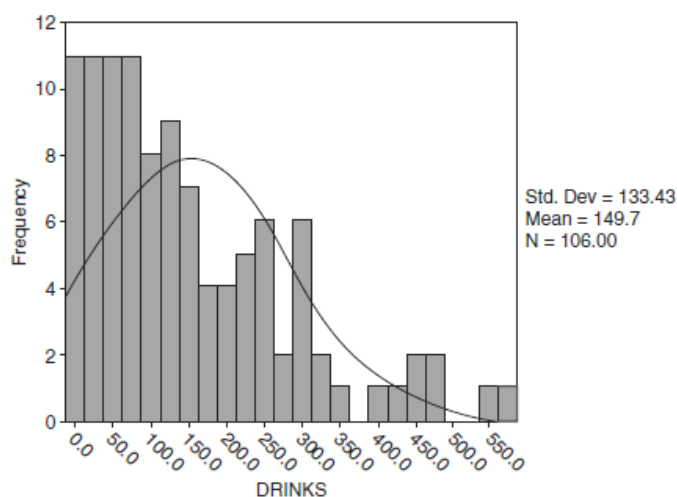
3.6.1 Normalita

Predpoklad normality znamená, že pozorovania vo výberových súboroch pochádzajú z populácie, ktorá má normálne rozdelenie pravdepodobnosti. Tieto populácie však nemáme k dispozícii, máme z nich iba náhodný výber. Vďaka centrálnej limitnej vete vieme, že ak naše výbery budú dostatočne veľké (väčšie ako 30 pozorovaní), tak rozdelenie výberového priemeru sa bude blížiť k normálnemu rozdeleniu.

Normalitu môžeme overovať viacerými spôsobmi. Jedným z nich je grafické posúdenie. Pomocou grafického zobrazenia môžeme vidieť, či sa rozdelenie analyzovaných údajov približuje ku krivke normálneho rozdelenia. Najviac používané typy grafov na zisťovanie normality sú histogram a takzvaný kvantilový Q-Q graf (quantile-quantile). Q-Q graf porovnáva vzostupne zoradené údaje z výberu, oproti hodnotám z teoretického rozdelenia.

V mnohých prípadoch nám pri overovaní normality pomôže jednoduchý histogram z našich údajov. Vidíme tak, či je rozdelenie empirických údajov podobné normálnemu rozdeleniu. Ako jasne vidno na grafe 3.1, zobrazené údaje sú pravostranne zošikmené a líšia sa od symetrickej krivky normálneho rozdelenia (Maxwell, Delaney, 2018).

Graf 3.1: Histogram zobrazujúci pravostranne zošikmené údaje



Zdroj: MAXWELL, Scott E.; DELANEY, Harold D.; KELLEY, Ken. *Designing experiments and analyzing data: A model comparison perspective*. Routledge, 2017. s. 138

Ďalšou možnosťou na overenie normality sú štatistické testy, ako napríklad Kolmogorovov-Smirnovov test a Shapiro-Wilkov test. Ak naše údaje obsahujú menej ako

50 pozorovaní odporúča sa použiť Shapiro-Wilkov test. V prípade väčšieho počtu pozorovaní je vhodnejšie použiť Kolmogorovov-Smirnovov test (Maxwell, Delaney, 2017).

3.6.2 Nezávislosť pozorovaní

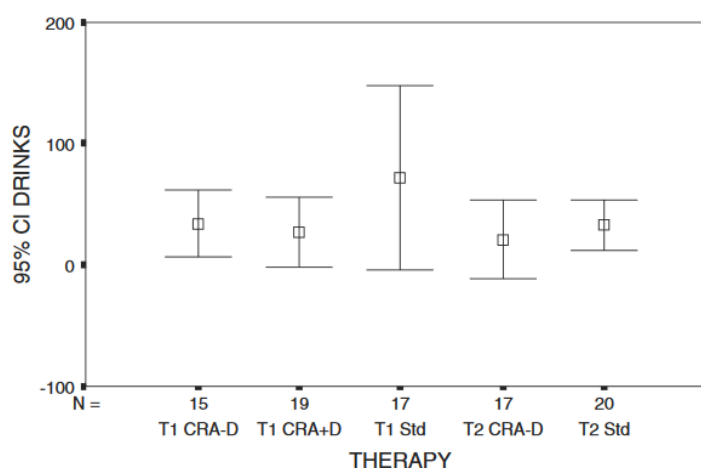
Nezávislosť výberov znamená, že hodnota jedného pozorovania z náhodného výberu neovplyvňuje hodnotu druhého pozorovania a naopak. Pri analýze rozptylu môžeme túto podmienku ešte rozčleniť. Musí byť splnená nezávislosť pozorovaní v rámci jednej skupiny, a súčasne musia byť nezávislé aj pozorovania medzi skupinami. Závislosť medzi pozorovaniami je často spôsobená nesprávnym získavaním údajov alebo zlým návrhom štúdie (McDonald, 2014).

Predpoklad nezávislosti je dôležitý. Porušenie tohto predpokladu má za následok podstatné zvýšenie miery chýb I. druhu. Scariano a Davenport (1987) ukázali, že ak vykonávame analýzu rozptylu so zvolenou hladinou významnosti 0,05, tak pri závislých pozorovaniach (s hodnotou Pearsonovho korelačného koeficientu 0,5) je miera chýb I. druhu až 0,74.

3.6.3 Homoskedasticita (zhoda rozptylov)

Len zriedkakedy si môžeme byť istí homoskedasticitou analyzovaných údajov. Preto je dôležité, aby sme vedeli použiť štatistické testy, ktorými overíme, či majú dáta zhodné rozptyly. Grafy, ktoré zobrazujú intervaly spoľahlivosti pre skupinové priemery spolu so smerodajnými odchýlkami taktiež môžu byť nápomocné na predbežné vyvrátenie alebo potvrdenie predpokladu homoskedasticity v našich dátach. Pri pohľade na graf 3.2, ktorý zobrazuje 95% intervaly spoľahlivosti stredných hodnôt porovnávaných skupín, vidíme, že najdlhší interval je takmer 3 krát dlhší ako najkratší. Napovedá nám to, že predpoklad zhody rozptylov je porušený (Maxwell, Delaney, 2017).

Graf 3.2: Intervaly spoľahlivosti stredných hodnôt porovnávaných skupín



Zdroj: MAXWELL, Scott E.; DELANEY, Harold D.; KELLEY, Ken. *Designing experiments and analyzing data: A model comparison perspective*. Routledge, 2017. s. 140

Pri testovaní homoskedasticity (zhody rozptylov) vo všeobecnosti overujeme platnosť nasledovnej nulovej hypotézy:

$$H_o: \sigma_1^2 = \sigma_2^2 = \dots \sigma_i^2 = \dots = \sigma_k^2, \quad (3.58)$$

oproti alternatívnej hypotéze

$$H_1: \text{non } H_o \text{ (aspoň jeden rozptyl } \sigma_i^2 \text{ je rôzny)}. \quad (3.59)$$

V nasledujúcom texte uvedieme najpoužívanéjšie testy na overenie homoskedasticity.

3.6.4 Hartleyov test

Test má veľmi jednoduchú testovaciu štatistiku, ktorou je pomer najväčšieho výberového rozptylu a najmenšieho výberového rozptylu. Na jeho použitie musí byť splnená podmienka normality testovaných údajov, čo dokazujú aj štúdie skúmajúce jeho robustnosť, (Neter et al., 1990) a navyše sú potrebné aj rovnaké rozsahy skupín.

$$H = \frac{\max \tilde{s}_i^2}{\min \tilde{s}_i^2}. \quad (3.60)$$

Hodnoty testovacej štatistiky blízke 1 podporujú hypotézu o homoskedasticite, naopak, vysoké hodnoty podporujú alternatívnu hypotézu o heteroskedasticite. Kritická oblasť tejto testovacej štatistiky sa vyjadrí ako

$$H_\alpha = \{H; H \geq H_{1-\alpha}(k; \bar{n} - 1)\}, \quad (3.61)$$

kde $H_{1-\alpha}(k; \bar{n} - 1)$ je za predpokladu platnosti nulovej hypotézy $(1 - \alpha)$ -kvantil rozdelenia H s k a $\bar{n} - 1$ stupňami voľnosti. Pre n platí, že $n_1 = n_2 = \dots n_k = \bar{n}$ (Pacáková a kol., 2012). Hodnota H je určená z tabuľky kritických hodnôt pre špeciálne rozdelenie. V tabuľke nájdeme hodnoty len pre 1. a 5. percentil. Zistíme z nej hodnoty pre 2 až 12 výberov pričom stupne voľnosti musia byť z množiny $\{2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 15, 20, 30, 60, \infty\}$ (Kanji, 1999).

3.6.5 Bartlettov test

Výhodou Bartlettovho testu je, že ho možno použiť aj pre súbory s rozdielnymi početnosťami skupín. Podobne ako Hartleyov test je však citlivý na porušenie normality. Jeho testovacia štatistika má tvar

$$B = \frac{1}{c} \left[(n - k) \ln \tilde{S}^2 - \sum_{i=1}^k (n_i - 1) \ln \tilde{S}_i^2 \right], \quad (3.62)$$

kde n predstavuje súčet početností všetkých k skupín ($n = \sum_{i=1}^k n_i$). \tilde{S}_i^2 predstavuje rozptyl v i -tej skupine s n_i pozorovaniami, čo môžeme vyjadriť ako

$$\tilde{S}_i^2 = \frac{\sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2}{n_i - 1}. \quad (3.63)$$

\tilde{S}^2 predstavuje priemer zo skupinových výberových rozptylov

$$\tilde{S}^2 = \frac{\sum_{i=1}^k (n_i - 1) \tilde{S}_i^2}{n - k} \quad (3.64)$$

a c sa vypočíta pomocou vzťahu

$$c = 1 + \frac{\sum_{i=1}^k \frac{1}{n_i - 1} - \frac{1}{(n - k)}}{3(k - 1)} \quad (3.65)$$

Ak platí nulová hypotéza o zhode rozptylov, testovacia štatistika B má približne rozdelenie χ^2 s $k - 1$ stupňami voľnosti. Zároveň je nutné, aby každá zo skupín mala aspoň 6 pozorovaní. Kritická oblasť testu je

$$B_\alpha = \{B; B \geq \chi_{1-\alpha}^2(k - 1)\} \quad (3.66)$$

(Pacáková a kol., 2012).

3.6.6 Leveneov test

Robustnejšiu alternatívu k Bartlettovmu a Hartleyovmu testu ponúka Leveneov test homoskedasticity. Je menej citlivý k porušeniu normality. V podstate pri ňom vykonávame analýzu rozptylu odchýlok pozorovaných hodnôt od celkového priemeru. Testovacia štatistika W má tvar

$$W = \frac{\frac{\sum_{i=1}^k (\bar{z}_i - \bar{z}_{..})^2}{(k-1)}}{\frac{\sum_{i=1}^k \sum_{j=1}^{n_i} (z_{ij} - \bar{z}_i)^2}{(n-k)}}. \quad (3.67)$$

Môžeme vidieť, že štatistika W má veľmi podobný tvar ako F test. V čitateli je priemerný súčet štvorcov rozdielu skupinových priemerov a celkového priemeru (SSA) a v čitateli priemerný súčet štvorcov rozdielu jednotlivých hodnôt od skupinového priemeru (SSE).

Vzťah 1.40 obsahuje premennú z , ktorá môže byť vypočítaná rôzne, v závislosti na rozdelení údajov.

$$z_{ij} = |y_{ij} - \bar{y}_j|, \quad (3.68)$$

ako absolútna hodnota rozdielu y_{ij} a skupinového priemeru,

$$z_{ij} = |y_{ij} - \tilde{y}_j|, \quad (3.69)$$

ako absolútna hodnota rozdielu y_{ij} a skupinového mediánu,

$$z_{ij} = |y_{ij} - \widetilde{y}_j^*|, \quad (3.70)$$

ako absolútna hodnota rozdielu y_{ij} a 10 % zastreihnutého skupinového priemeru.

Vo vzťahoch 3.67 až 3.70 \bar{z}_i označuje priemer premennej z v i -tej skupine a $\bar{z}_{..}$ je celkový priemer zo všetkých pozorovaní (Pacáková a kol., 2012).

Zastreihnutý priemer dostaneme, keď zo vzostupne usporiadaného výberového súboru vynecháme 10 % najmenších a najväčších hodnôt (He, 2011).

Typ výpočtu z ovplyvňuje robustnosť a silu Leveneovho testu. Leveneova pôvodná práca obsahovala len výpočet s obyčajným priemerom. Brown a Forsythe (1974) rozšírili Leveneov test o možnosti s orezaným priemerom a mediánom. Vykonali Monte Carlo simulácie, ktoré indikovali, že pri údajoch pochádzajúcich z rozdelenia s ťažkými

chvostami sa najlepšie výsledky dosiahli s orezaným priemerom. Ak údaje pochádzali zo zošikmeného (napríklad χ^2) rozdelenia, tak najrobustnejšie výsledky sa dosiahli s mediánom. Použitie priemeru malo najväčšiu silu pri symetrických rozdeleniach so stredne ťažkými chvostami.

Optimálna voľba výpočtu z závisí od rozdelenia údajov, ale vo všeobecnosti sa odporúča možnosť s použitím mediánu, keďže poskytuje dobrú mieru robustnosti, pri zachovaní sily. Ak máme viac informácií o rozdelení analyzovaných údajov, môžeme zvoliť inú vhodnejšiu alternatívu.

Kritická oblasť testovacej štatistiky W (pre všetky varianty z) je množina

$$W_\alpha = \{F; F_{1-\alpha}(k-1; n-k)\}, \quad (3.71)$$

kde $F_{1-\alpha}(k-1; n-k)$ je za predpokladu platnosti nulovej hypotézy $(1-\alpha)$ -kvantil Fisherovho rozdelenia s $k-1$ a $n-k$ stupňami voľnosti (Pacáková a kol., 2012).

3.6.7 O'Brienov test

Podobne ako pri Leveneovom teste, aj pri O'Brienovom teste sa vychádza z F-testu a analýzy rozptylu, ale úprava premennej z_{ij} je mierne odlišná.

Najprv vypočítame pre každú skupinu výberový rozptyl

$$s_j^2 = \frac{\sum_{i=1}^k (y_{ij} - \bar{y}_j)^2}{n_j - 1}. \quad (3.72)$$

Hodnoty transformovanej premennej z určíme pomocou vzťahu

$$z_{ij} = \frac{(n_{ij} - 1,5)n_i(y_{ij} - \bar{y}_j)^2 - 0,5s_j^2(n_i - 1)}{(n_i - 1)(n_i - 2)}. \quad (3.73)$$

Testovacia štatistika má potom tvar podielu

$$O = \frac{\frac{\sum_{i=1}^k (\bar{z}_i - \bar{z}_{..})^2}{(k-1)}}{\frac{\sum_{i=1}^k \sum_{j=1}^{n_i} (z_{ij} - \bar{z}_i)^2}{(n-k)}}, \quad (3.74)$$

kde \bar{z}_i je skupinový priemer transformovaných premenných a $\bar{z}_{..}$ je celkový priemer všetkých transformovaných premenných. Ak platí nulová hypotéza o rovnosti rozptylov,

testovacia štatistika má rozdelenie F s počtami stupňov voľnosti $(k - 1)$ a $(n - k)$. Nulovú hypotézu zamietame, ak nadobúda hodnoty z množiny

$$O_\alpha = \{F; F_{1-\alpha}(k - 1; n - k); \infty\}. \quad (3.75)$$

Ak sú rozsahy výberových súborov veľmi nevyvážené (najväčší výber je 4 a viac násobne väčší ako najmenší výber), odporúča sa kvôli robustnosti použiť Welchov F-test namiesto klasického F-testu (Abdi, 2007).

3.7 Transformácia údajov

Ak údaje nespĺňajú podmienku normality alebo homoskedasticity, môžeme ich transformovať tak, aby sa viac priblížili k podmienkam použitia analýzy rozptylu. Transformácia znamená upravenie údajov pomocou matematickej funkcie. Treba podotknúť, že transformované sú hodnoty závislej premennej. Napríklad, keď sú údaje kladne zošikmené, tak použitie odmocninovej alebo logaritmickej transformácie má vo všeobecnosti za následok priblíženie sa k normalite, v niektorých prípadoch aj k homoskedasticite.

Existuje množstvo transformácií, z ktorých si môžeme vybrať. Jednou zo stratégií, ktorou môžeme postupovať pri voľbe transformácie, je jednoducho vyskúšať ich viacero a zvoliť tú, po ktorej sa údaje najviac priblížili k splneniu predpokladov (menšia heteroskedasticita, nižšia miera špicatosti a šikmosti). Ďalší spôsob spočíva v tom, že najčastejšie používané transformácie možno zoradiť podľa stupňa mocniny pôvodnej premennej. Takýto „*rebríček umocnenia*“ je uvedený v tabuľke 3.3 (Maxwell, Delaney, 2017).

Tabuľka 3.3: Najčastejšie používané transformácie údajov

| Exponent | Transformácia | Vysvetlenie |
|---------------|---------------|---|
| 2 | Y^2 | Umocnenie, užitočné pri zápornom zošikmení |
| 1 | Y | Pôvodné údaje, bez transformácie |
| $\frac{1}{2}$ | \sqrt{Y} | Odmocnenie, užitočné pri kladne zošikmených rozdeleniach |
| 0 | $\log_{10} Y$ | Logaritmovanie, užitočné keď je odhadnuté p blízke 0, t.j. pri extrémne kladnom zošikmení |
| -1 | $\frac{1}{Y}$ | Recipročné |

Zdroj: MAXWELL, Scott E.; DELANEY, Harold D.; KELLEY, Ken. *Designing experiments and analyzing data: A model comparison perspective*. Routledge, 2017. s. 143.

Pri logaritmovaní a odmocnení treba pamätať na to, že takéto úpravy nemôžeme spraviť, ak je pôvodná hodnota nulová. V tom prípade k nule pripočítame malú hodnotu, obyčajne 1, takže transformácia bude mať tvar $\log_{10}(Y + 1)$. Ak údaje obsahujú veľa nulových hodnôt, logaritmická transformácia neprináša spoľahlivé výsledky a analyzovanie takýchto údajov je komplikované (Childs, 2018).

Použitie transformácií však prináša aj nevýhody. Hlavnou z nich je, že interpretácia výsledkov je menej jasná a jednoznačná. Z potvrdenia nulovej hypotézy o zhode stredných hodnôt premennej Y nevyplýva, že platí hypotéza o zhode transformovaných stredných hodnôt Y . Na ilustráciu uvádzame príklad:

Majme dva výberové súbory X_1 a X_2 . Ak údaje z oboch výberov pochádzajú zo základných súborov s lognormálnym rozdelením, $LN(\mu_1, \sigma_1^2)$ a $LN(\mu_2, \sigma_2^2)$, tak stredná hodnota rozdelenia prvého výberu je $\exp(\mu_1 + \frac{\sigma_1^2}{2})$ a stredná hodnota rozdelenia druhého výberu je $\exp(\mu_2 + \frac{\sigma_2^2}{2})$. Ak použijeme dvojvýberový t-test na pôvodné údaje, tak testujeme nulovú hypotézu, že tieto 2 stredné hodnoty sú rovnaké:

$$H_0: \exp\left(\mu_1 + \frac{\sigma_1^2}{2}\right) = \exp\left(\mu_2 + \frac{\sigma_2^2}{2}\right) \quad (3.76)$$

Ak použijeme logaritmickú transformáciu na naše údaje, tak transformované údaje budú mať strednú hodnotu μ_1 a rozptyl σ_1^2 pre prvý výber a μ_2 a σ_2^2 pre 2. výber. Teda ak použijeme dvojvýberový t-test na transformované údaje, nulová hypotéza bude mať tvar

$$H_0: \mu_1 = \mu_2. \quad (3.77)$$

Na porovnaní týchto dvoch nulových hypotéz jasne vidno, že nie sú ekvivalentné (Feng, 2014).

3.8 Robustnosť analýzy rozptylu

Pod robustnosťou rozumieme nízku citlivosť štatistických metód na porušenie predpokladov ich použitia. To znamená, že miera chýb I. druhu (pravdepodobnosť zamietnutia pravdivej nulovej hypotézy) sa porušením týchto predpokladov výrazne nemení (Box, 1953).

Štúdie skúmajúce robustnosť štatistických metód vrátane analýzy rozptylu využívajú Monte Carlo simulácie. Aby sa zabezpečila porovnateľnosť týchto štúdií, je potrebné pracovať so štandardným kritériom určujúcim robustnosť. Ako najviac vhodné

kritérium sa v odbornej literatúre objavuje podmienka definovaná Bradleyom (1978). Tá hovorí, že štatistický test možno považovať za robustný, ak empirická miera chýb I. druhu je medzi 0,025 a 0,075 pre zvolenú hladinu významnosti 0,05. Ak je táto miera nad 0,075, tak test je príliš liberálny. V podmienkach analýzy rozptylu to znamená riziko záveru testovania, podľa ktorého porovnávané skupiny majú rozdielne stredné hodnoty, lenže v skutočnosti sú ich stredné hodnoty rovnaké. Naopak, keď je miera chýb I. druhu menšia ako 0,025, tak test je príliš konzervatívny. Znamená to, že je vyššia pravdepodobnosť chyby druhého druhu.

V mnohých prípadoch použitia analýzy rozptylu sú predpoklady na jej využitie splnené. Empirické skúsenosti často potvrdzujú, že údaje sú aspoň približne normálne. Takisto podmienka rovnosti rozptylov je často splnená, pretože rôzne úrovne skúmaného faktora môžu ovplyvňovať strednú hodnotu v skupinách, ale nie variabilitu. Splnenie podmienky nezávislosti, je z veľkej časti určené návrhom a vykonávaním samotného experimentu. Aj keď testované údaje nekopírujú dokonale normálne rozdelenie, môžu byť k nemu dostatočne blízke (unimodálne, symetrické, len málo extrémnych pozorovaní) a nemusíme podmienke normality prikladať veľkú váhu. Otázkou je, ako rozhodnúť, kedy je toto odchýlenie od normality a homoskedasticity ešte v tolerancii, a kedy už nie. Za účelom zistenia robustnosti analýzy rozptylu bolo vykonaných mnoho štúdií, ktorých výsledky zhrnutujeme v tejto kapitole (Maxwell, Delaney, 2017).

3.8.1 Porušenie normality

ANOVA je vo všeobecnosti robustná voči porušeniu podmienky normality a v tomto prípade je miera chýb I. druhu blízko k nominálnej (želanej) hodnote. Ak údaje v štúdií porovnávajúcej 3 skupiny po šiestich subjektoch nepochádzajú z normálneho rozdelenia, percento zistených hodnôt F-štatistiky presahujúcich kritickú hodnotu 3,68 je stále relatívne blízko k 5 %. Preto mnohí nepovažujú porušenie predpokladu normality za výraznú prekážku pri použití klasickej analýzy rozptylu. Metaanalýza simulačných štúdií, ktoré používali dáta pochádzajúce z výrazne zošikmeného ($\gamma_1 = 2$) a zošpicateného ($\gamma_2 = 6$) rozdelenia s rovnakými početnosťami v skupinách, vykazovali hladiny α veľmi blízko k nominálnej hodnote 0,05. Konkrétne $\alpha = 0,059$ so smerodajnou odchýlkou 0,026. Výraznejší rozdiel bol pri analýze rozptylu, pri skupinách s rozdielnymi početnosťami, kde $\alpha = 0,069$ so smerodajnou odchýlkou 0,048 (Lix, Keselman, Keselman, 1996). Štúdie

naznačujú aj to, že miera chýb I. druhu vo všeobecnosti rastie úmerne s tým, ako sa dáta odlišujú od normality.

Na robustnosť však nemožno pozerat' spôsobom buď – alebo. Tým ako sa údaje vzdialujú od normality, zistená miera chýb I. druhu v štúdiách typicky rastie. Napríklad Micceri (1989) analyzoval dáta z extrémne asymetrického rozdelenia, kde miery chýb I. druhu boli všeobecne veľmi blízko nominálnej miere aj pri nepriaznivých podmienkach, akými boli málo početné výbery s nerovnakými početnosťami a zároveň boli skôr konzervatívne.

Väčšina štúdií zameraných na robustnosť analýzy rozptylu sa zameriavala na chyby I. druhu, ale nie na silu testu (1- miera chýb II. druhu). Wilcox (1995) uvádza, že ANOVA je vo všeobecnosti málo robustná v sile testu pri porušení normality. Podľa neho údaje vo výskumoch často pochádzajú z rozdelenia z neočakávane ťažkými chvostami. Navyše na prvý pohľad pomocou grafických metód ani nezistíme, že dáta v našom výbere sú z takéhoto rozdelenia.

3.8.2 Porušenie homoskedasticity

Štúdie zamerané na robustnosť F-testu boli zhrnuté Glassom, Peckhamom a Sandersom (1972). Podľa nich sa výsledky týchto štúdií dajú zosumarizovať do 2 základných bodov. F-test je možné považovať za robustný, keď majú skupiny rovnaké početnosti a rozsah skupín nie je veľmi malý (väčší než 7).

Druhé zistenie je, že F-test obyčajne nie je robustný, keď majú porovnávané skupiny heterogénne rozptyly. V tom prípade, ale miera chýb I. druhu závisí na pomere rozptylov a početnosti v skupine. To znamená, že F-test má tendenciu byť konzervatívny, keď je táto kombinácia pozitívna, teda keď má skupina s najväčším rozptylom aj najväčšiu početnosť a súčasne skupina s najmenším rozptylom má najmenšiu početnosť. Naopak, ak je kombinácia negatívna – skupina s najväčším rozptylom má najmenšiu početnosť a skupina s najmenším rozptylom má najväčšiu početnosť, tak test je liberálny.

Porovnanie mnohých novších štúdií od Blanca a kol. (2017), ktoré sa zameriavajú na použitie analýzy rozptylu s rovnako početnými skupinami, však nie je vôbec jednoznačné. Uvádzajú viacero štúdií, ktoré došli k protichodným záverom, niektoré robustnosť F-testu potvrdzujú, iné zase vyvracajú. Táto nekonzistentnosť môže byť spôsobená viacerými faktormi. Štúdie nepoužili jednotné kritérium na definovanie robustnosti (napríklad spomínané Bradleyovo). Takisto použili rôzne ukazovatele na

posúdenie miery heteroskedasticity (variačný koeficient, variačný pomer, prípadne vlastné špecifické indexy). Vo všeobecnosti sa dá z výsledkov týchto štúdií usúdiť, že pri situácii keď má jedna skupina extrémne odlišnú variabilitu (1:1:30) F-test nie je robustný.

Štúdia od Blanca a kol. (2017) testovala za pomoci Monte Carlo simulácií robustnosť F-testu pri heterogénnych rozptyloch s rôznymi charakteristikami výberových súborov. Sledovali, či na robustnosť vplyva rozsah výberového súboru, rovnaké aj odlišné početnosti v skupinách, kombinácia veľkosti rozsahu súboru a veľkosti variability a takzvaného *pomeru rozptylov*. Ich výsledky ukázali, že robustnosť dokážeme spoľahlivo posúdiť pomocou pomeru rozptylu skupiny s najväčšou variabilitou a skupiny s najmenšou variabilitou:

$$v_{\sigma_{skupin.}} = \frac{\sigma_{i_{max}}}{\sigma_{i_{min}}}. \quad (3.78)$$

Celkový rozsah výberového súboru nevplyva pri porušenej podmienke rovnosti rozptylov na robustnosť. Veľký rozsah nás teda nemusí ochrániť pred efektom, ktorý má heteroskedasticita. Robustnosť F-testu klesá spolu so stúpajúcim pomerom rozptylov. Inými slovami, čím viac sa poruší podmienka homoskedasticity, tým menej je test robustný. Pri pomere rozptylov rovnému 1.5, bol F-test robustný pri všetkých možných overovaných podmienkach. Pri vyššej hodnote tohto ukazovateľa už začína byť test pri niektorých parametroch konzervatívny alebo naopak liberálny.

3.9 Simulácie a metóda Monte Carlo

Metódou Monte Carlo nazývame numerickú metódu na riešenie matematických úloh pomocou modelovania náhodných veličín a štatistického odhadu ich charakteristík. Dajú sa ňou riešiť nielen pravdepodobnostné, ale aj deterministické úlohy. Prvé pokusy, ktoré boli predchodcami metód Monte Carlo, boli známe už v 19. storočí. Takzvaná Buffonova úloha o ihle spočívala v určení čísla π , pomocou náhodného hádzania ihly na rovinu pokrytú rovnako vzdialenými rovnobežkami, pričom dĺžka ihly bola rovná vzdialenosti medzi rovnobežkami. Dá sa odvodiť, že pravdepodobnosť pretnutia rovnobežky ihlou je $\frac{2}{\pi}$ (Dřimal, Trunec, 1989).

K praktickejšiemu využitiu takýchto metód došlo až po rozšírení výpočtovej techniky a počítačov. V roku 1944 John von Neumann navrhol použitie aparátu pravdepodobnosti pomocou počítačov pri vývoji atómovej bomby, keď spolu s S. A.

Ulamom, N. Metropolisom, H. Kahnom a E. Fermim skúmali správanie neutrónov. Práve v práci Metropolisa a Ulama (1949) bol spomenutý termín *metóda Monte Carlo*.

3.9.1.1 Postup riešenia úloh metódou Monte Carlo

Pomocou metódy Monte Carlo modelujeme takú náhodnú veličinu X , ktorej stredná hodnota $E(X)$ je rovná hľadanej hodnote a , čo je riešením úlohy. Teda platí

$$E(X) = a. \quad (3.79)$$

Ak vygenerujeme n nezávislých realizácií X_1, X_1, \dots, X_n náhodnej veličiny X , tak a môžeme odhadnúť pomocou aritmetického priemeru

$$a \doteq \frac{1}{n}(X_1 + X_2 + \dots + X_n). \quad (3.80)$$

Hodnoty X_1, X_1, \dots, X_n dostaneme vo väčšine Monte Carlo metód najprv generovaním hodnôt náhodnej veličiny Y rovnomerne rozdelených na intervale $(0; 1)$. Tieto sa následne pretransformujú vhodne zvolenou funkciou f alebo vhodným algoritmom na požadované hodnoty X_1, X_1, \dots, X_n . Získané výsledky následne vhodne štatisticky spracujeme (Dřímal, Trunec, 1989).

3.9.1.2 Použitie Monte Carlo simulácie v indukčnej (inferenčnej) štatistike

Základnou myšlienkou použitia Monte Carlo simulácií v inferenčnej štatistike je to, že určité poznatky o skúmanej charakteristike môžeme získať opakovaným náhodným vyberaním z toho istého rozdelenia alebo populácie a pozorovaním, ako sa hodnota tejto charakteristiky mení medzi výbermi.

V štatistike majú Monte Carlo simulácie viacero využití. Patrí medzi ne:

- Vykonávanie inferenčných metód, keď nevieme analyticky vyjadriť rozdelenie testovacej štatistiky.
- Posudzovanie presnosti inferenčných metód, keď sú porušené predpoklady ich využitia.
- Testovanie nulových a alternatívnych hypotéz za rôznych podmienok.
- Posudzovanie sily inferenčných metód.
- Porovnávanie kvality rôznych odhadov.

Všeobecný postup, akým používame Monte Carlo simulácie v úlohách inferenčnej štatistiky je nasledovný:

1. Určíme model alebo pseudo-populáciu, ktorá bude reprezentovať skutočnú populáciu, ktorá nás zaujíma.
2. Urobíme náhodný výber na získanie vzorky z rozdelenia alebo pseudo-populácie
3. Zistíme hodnotu analyzovanej štatistiky a uložíme ju.
4. Zopakujeme kroky 2 až 3 pre M pokusov.
5. M hodnôt zistených z jednotlivých pokusov použijeme na určenie rozdelenia štatistiky.

(Martinez, Martinez, 2008)

3.9.2 Stanovenie mier chýb I. a II. druhu pomocou simulácií

Ako už bolo uvedené v prehľade súčasného stavu riešenej problematiky, pri štatistickom testovaní hypotéz môže dôjsť k dvom druhom chýb. Prvou je *chyba I. druhu*, ktorá nastane, keď štatistický test zamietne pravdivú nulovú hypotézu. Opačný problém nastane, keď prijmem nulovú hypotézu, ktorá v skutočnosti nie je pravdivá. Vtedy hovoríme, že nastala *chyba II. druhu*.

Pomerne jednoduchou procedúrou je možné zo všetkých vygenerovaných údajov určiť pre skúmaný štatistický test odhad miery chýb I. a II. druhu.

3.9.2.1 Odhad miery chýb I. druhu

Pri zisťovaní odhadu tejto miery budeme generovať údaje, v ktorých majú skupiny s rozdielnou úrovňou faktora rovnaké stredné hodnoty. Môžeme povedať, že generujeme výberové súbory z pseudo-populácie, ktorá reprezentuje hypotézu H_0 . Postup pri stanovení odhadu miery chýb I. druhu je potom nasledovný.

1. Definujeme pseudo-populáciu alebo rozdelenie, kde *platí* nulová hypotéza.
2. Vygenerujeme náhodný výber s rozsahom n z tohto rozdelenia alebo pseudo-populácie.
3. Overíme hypotézu zvoleným štatistickým testom.
4. Zistíme, či počas testu došlo k chybe I. druhu. Overíme teda, či nulová hypotéza bola na zvolenej hladine významnosti zamietnutá. Výber sme uskutočnili z rozdelenia, kde nulová hypotéza platí, preto jej zamietnutie znamená dopustenie sa chyby I. druhu.

5. Zaznamenáme výsledok pokusu ako

$$I_i = \begin{cases} 1; & \text{ak chyba I. druhu nastala} \\ 0; & \text{ak chyba I. druhu nenastala.} \end{cases} \quad (3.81)$$

6. Kroky 2 až 5 zopakujeme pre M pokusov.

7. Odhad miery I. druhu štatistického testu pri overovaní hypotéz o údajoch zo zvoleného rozdelenia je potom

$$\hat{\alpha} = \frac{1}{M} \sum_{i=1}^M I_i. \quad (3.82)$$

(Martinez, Martinez 2008)

3.9.2.2 Odhad miery chýb II. druhu a sily testu

Postup je veľmi podobný pri určení ako pri odhade miery chýb I. druhu. V tomto prípade budeme generovať údaje, v ktorých majú aspoň dve skupiny s rozdielnou úrovňou faktora rôzne stredné hodnoty.

1. Definujeme pseudo-populáciu alebo rozdelenie, kde *neplatí* nulová hypotéza.
2. Vygenerujeme náhodný výber s rozsahom n z tohto rozdelenia alebo pseudo-populácie.
3. Overíme hypotézu zvoleným štatistickým testom.
4. Zistíme, či počas testu došlo k chybe II. druhu. Overíme teda, či nulová hypotéza bola na zvolenej hladine významnosti prijatá. Výber sme uskutočnili z rozdelenia, kde nulová hypotéza neplatí, preto jej prijatie znamená dopustenie sa chyby II. druhu.
5. Zaznamenáme výsledok pokusu ako

$$I_i = \begin{cases} 1; & \text{ak chyba II. druhu nastala} \\ 0; & \text{ak chyba II. druhu nenastala.} \end{cases} \quad (3.83)$$

6. Kroky 2 až 5 zopakujeme pre M pokusov.

7. Odhad miery I. druhu štatistického testu pri overovaní hypotéz o údajoch zo zvoleného rozdelenia je potom

$$\hat{\beta} = \frac{1}{M} \sum_{i=1}^M I_i. \quad (3.84)$$

8. Silu testu následne určíme ako

$$\text{silu testu} = 1 - \hat{\beta}. \quad (3.85)$$

Neexistuje všeobecne aplikovateľná teória, pomocou ktorej by sme mohli určiť vhodný počet simulácií. Odporúča sa však postupovať podľa nasledujúceho spôsobu:

Mali by sme najprv vykonať malý počet simulácií a presvedčiť sa, že program, pomocou ktorého generujeme údaje funguje správne. Keď sme si istí, že náš simulačný program generuje želané údaje, mali by sme sa snažiť vykonať čo najväčší počet simulácií M . Ich počet závisí aj od časovej a výpočtovej náročnosti vykonávaných testov. Vo všeobecnosti by počet simulácií nemal byť nižší ako 1000 a najčastejšie sa ich v štúdiách vykonáva 10000 až 25000 (Martinez, Martinez 2008).

4 Výsledky práce

V aplikačnej časti diplomovej práce sa venujeme použitiu metód porovnávania stredných hodnôt viac ako dvoch výberových súborov popísaných v teoretickej časti. V prvej časti sa zaoberáme použitím metód analýzy rozptylu a Welchovej analýzy rozptylu na údajoch popisujúcich cenu počítačových zostáv. Zvlášť sa zameriavame na použitie týchto testov pri rozdielnych hodnotách variability porovnávaných súborov, ktorý je jedným z predpokladov použitia klasickej analýzy rozptylu. Pozrieme sa, či použitím Welchovej analýzy rozptylu, ktorá berie do úvahy aj heteroskedasticitu porovnávaných údajov, dosiahneme odlišný výsledok, ako pri bežnej analýze rozptylu.

V ďalšej kapitole aplikačnej časti sa budeme venovať vzájomnému porovnávaniu dvoch metód analýzy rozptylu (klasickej a Welchovej analýze rozptylu). Na základe predošlých vykonaných štúdií, ktoré skúmali robustnosť a silu klasickej jednofaktorovej analýzy rozptylu pri porušení predpokladu homoskedasticity, sme sa rozhodli, že bude zaujímavé vykonať obdobné analýzy aj pre Welchovu analýzu rozptylu. Aplikačná časť obsahuje porovnanie robustnosti, aj sily testu analýzy rozptylu oproti Welchovej analýze rozptylu, pri porušení predpokladu o zhode rozptylov.

Analýzy budeme vykonávať pomocou počítačových Monte Carlo simulácií. Simulačné štúdie nám umožňujú určiť presnosť a robustnosť rôznych štatistických metód, keďže v nich analyzujeme údaje, ktoré majú nami zvolené parametre a charakteristiky. V našom prípade teda vieme porovnať, koľko krát z celkového počtu simulácií štatistický test správne potvrdil alebo zamietol zhodu rozptylov, či stredných hodnôt v analyzovaných údajoch. Takýmto spôsobom bude možné analyzovať robustnosť a silu testov v prípade, keď údaje nespĺňajú predpoklady využitia príslušného testu. Vďaka možnosti generovať údaje so želanými parametrami pomerne jednoducho a rýchlo sú Monte Carlo simulácie často využívané výskumníkmi v ekonómii, financiách, sociálnych vedách či zdravotníctve (Burton et al., 2006).

Celý proces skúmania a analýzy je vykonaný pomocou procedúr v štatistickom softvéri SAS Enterprise Guide a programov vytvorených v programovacom jazyku Python a jeho vývojovom prostredí Jupyter. Kód, ktorým boli vytvorené údaje, simulácie a vykonané štatistické testy je dostupný k nahliadnutiu v prílohe 3.

Dúfame, že na základe výskumu budeme schopní určiť, pri akých charakteristikách analyzovaných údajov je vhodnejšie použiť analýzu rozptylu a kedy jej Welchovu modifikáciu.

4.1 Aplikácia analýzy rozptylu

V prvej časti sa zaoberáme použitím metód analýzy rozptylu a Welchovej analýzy rozptylu na údajoch popisujúcich cenu počítačových zostáv. Zvlášť sa zameriavame na použitie týchto testov pri porovnávaní súborov s rozdielnou variabilitou. Budeme skúmať, či použitím Welchovej analýzy rozptylu, ktorá nevyžaduje zhodu rozptylov porovnávaných údajov, dosiahneme odlišný výsledok, ako v prípade bežnej analýzy rozptylu. Prítomnosť homoskedasticity overíme Leveneovým testom.

4.1.1 Údajová základňa

V tejto časti výsledkov práce sa zaoberáme použitím metód analýzy rozptylu a Welchovej analýzy rozptylu na údajoch popisujúcich cenu počítačových zostáv. Údaje pochádzajú z internetovej stránky spoločnosti Nutanix, Inc.,¹ ktorá zabezpečuje a predáva počítačové zostavy, infraštruktúru potrebnú na úschovu a spracovanie dát a ďalšie softvérové riešenia. Údajová základňa poskytuje informácie o cenách a zložení počítačových zostáv pre inštitúcie štátnej správy v Spojených Štátoch amerických, ktoré musia byť zo zákona zverejňované. Dostupné údaje obsahujú 33 048 záznamov o predajných cenách zostáv s rôznymi parametrami počítača a jeho komponentov. Ceny sú uvedené v amerických dolároch (USD). Údaje sú rozčlenené podľa druhu pevných diskov použitých na ukladanie údajov, ktoré konkrétna zostava obsahuje.

V práci sme sa zamerali na jednofaktorovú analýzu rozptylu, takže cieľom našej analýzy bude posúdiť vplyv vybraného faktoru *druh diskov* na premennú *cena*. Vysvetľujúca premenná *druh diskov* je kategoriálna a nadobúda 3 obmeny, teda budeme porovnávať 3 výberové súbory. Závislá premenná *cena* je numerická a udávaná v amerických dolároch (USD). Označenie obmien premennej *druh diskov* je uvedené v tabuľke 4.1.

¹ Dostupné na internete: <<http://ntnxcagov.com/>>

Tabuľka 4.1: Úrovne faktorov skúmanej premennej

| Faktor | Označenie obmeny | Popis |
|-------------|------------------|-------------------------------|
| Druh diskov | All_HDD | Obsahuje len disky typu HDD |
| | All_Flash | Obsahuje len disky typu SSD |
| | Hybrid | Obsahuje kombináciu SSD a HDD |

Zdroj: Vlastné spracovanie

4.1.2 Overenie homoskedasticity

Použitie klasickej analýzy rozptylu predpokladá splnenie troch podmienok. Analyzované údaje majú pochádzať z normálneho rozdelenia, porovnávané súbory musia byť navzájom nezávislé a zároveň musí byť splnená podmienka homoskedasticity, teda zhody rozptylov.

Keďže naše porovnávané súbory majú veľké počty pozorovaní, využijeme centrálnu limitnú vetu, ktorá hovorí o tom, že ak zo základného súboru uskutočníme dostatočne veľký náhodný výber, tak výberové rozdelenie priemeru bude asymptoticky normálne.

Keďže analyzované počítačové zostavy sa skladajú z množstva komponentov a každá je zostavená z inej kombinácie týchto komponentov, predpokladáme, že pozorovania v jednotlivých súboroch a zároveň medzi porovnávanými súbormi sú vzájomne nezávislé.

Tretiu podmienku, rovnosť rozptylov porovnávaných súborov, budeme overovať Leveneovým testom. Ako už bolo spomenuté v predošlej kapitole, faktor, ktorého vplyv budeme skúmať (*druh diskov*) má 3 obmeny, čo znamená, že budeme porovnávať rozptyly 3 výberových súborov. To znamená, že na zvolenej hladine významnosti 0,05 budeme overovať zhodu rozptylov $\sigma_1^2, \sigma_2^2, \sigma_3^2$.

Výsledky Leveneovho testu homoskedasticity vidíme v tabuľke 4.2.

Tabuľka 4.2: Výsledky Leveneovho testu homoskedasticity

| Levene's Test for Homogeneity of value Variance ANOVA of Squared Deviations from Group Means | | | | | |
|---|------|----------------|-------------|---------|--------|
| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
| Druh diskov | 2 | 3.268E16 | 1.634E16 | 500.93 | <.0001 |
| Error | 3300 | 1.076E17 | 3.262E13 | | |

Zdroj: Výstup zo SAS Enterprise Guide

Na základe p-hodnoty, ktorá je menšia ako zvolená hladina významnosti 0,05, zamietame nulovú hypotézu o zhode rozptylov v porovnávaných súboroch. Môžeme teda povedať, že aspoň dva zo súborov majú rozdielne hodnoty rozptylov.

4.1.3 Posúdenie vplyvu druhu diskov na cenu zostavy

Naším cieľom je zistiť, či druh diskov použitých v zostave štatisticky významne ovplyvňuje výslednú cenu zostavy. Na posúdenie tohto vplyvu použijeme jednofaktorovú analýzu rozptylu. Keďže v predošlej časti sme zistili, že porovnávané súbory nemajú rovnakú variabilitu, použijeme bežnú analýzu rozptylu a následne našu hypotézu otestujeme pomocou Welchovej analýzy rozptylu, ktorá by mala, vzhľadom k prítomnosti heteroskedasticity, poskytovať spoľahlivejšie výsledky.

4.1.3.1 Použitie bežnej analýzy rozptylu

Ďalším krokom je overenie vplyvu premennej *druh diskov* na výslednú *cenu* zostáv. Na overenie použijeme jednofaktorovú analýzu rozptylu, kde budeme na hladine významnosti 0,05 testovať hypotézu o zhode stredných hodnôt porovnávaných súborov.

Tabuľka 4.3: Výsledky jednofaktorovej analýzy rozptylu

| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|-----------------|------|----------------|-------------|---------|--------|
| Model | 2 | 39989174 | 19994587 | 7.63 | 0.0005 |
| Error | 3300 | 8650182496 | 2621267 | | |
| Corrected Total | 3302 | 8690171670 | | | |

Zdroj: Výstup zo SAS Enterprise Guide

V ANOVA tabuľke 4.3 vidíme, že uvedená p-hodnota je výrazne nižšia ako 0,05. Na hladine významnosti 0,05 môžeme zamietnuť nulovú hypotézu o zhode stredných hodnôt. Prijímame teda alternatívnu hypotézu, že aspoň dva porovnávané súbory majú

rôzne stredné hodnoty. Z toho môžeme vyvodiť záver, že použitý druh diskov v počítačovej zostave ovplyvňuje jej výslednú cenu.

4.1.3.2 Použitie Welchovej analýzy rozptylu

Podľa výsledkov Leveneovho testu majú naše porovnávané súbory údajov rôzne rozptyly. Vplyv druhu diskov na výslednú cenu preto posúdime aj pomocou Welchovej analýzy rozptylu. V tomto prípade sa Welchova ANOVA javí ako vhodnejší test, keďže na jej použitie nemusí byť splnený predpoklad o zhodných rozptyloch.

Takisto ako pri bežnej jednofaktorovej analýze rozptylu, aj tu budeme overovať, či platí hypotéza o rovnosti stredných hodnôt všetkých porovnávaných súborov.

Tabuľka 4.4: Výsledky Welchovej analýzy rozptylu

| Welch's ANOVA for value | | | |
|-------------------------|--------|---------|--------|
| Source | DF | F Value | Pr > F |
| Druh diskov | 2.0000 | 2.94 | 0.0533 |
| Error | 1479.6 | | |

Zdroj: Výstup zo SAS Enterprise Guide

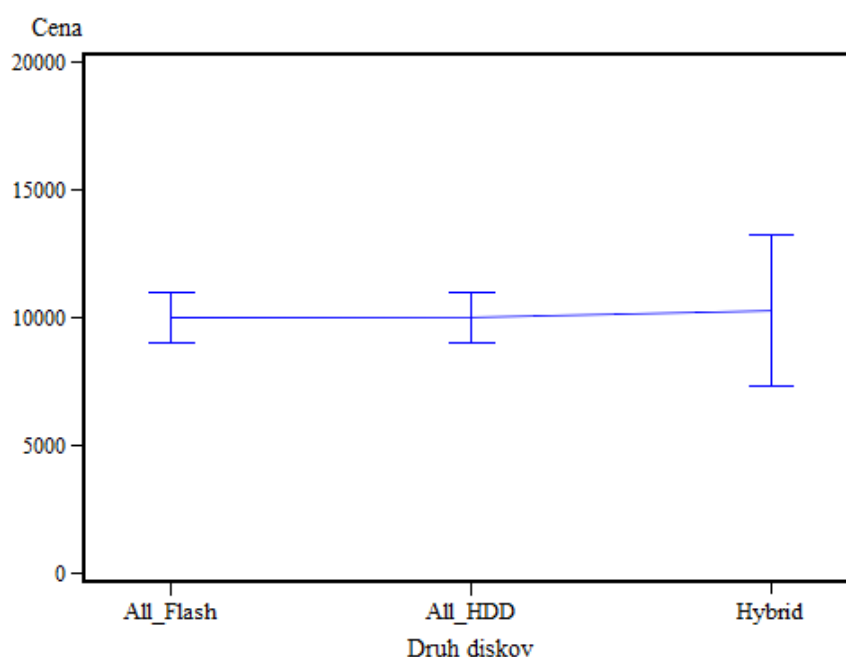
Výstup zo SAS Enterprise Guide s výsledkami Welchovej analýzy rozptylu vidíme v tabuľke 4.4. P-hodnota má v tomto prípade hodnotu 0,0533, čo je veľmi blízke zvolenej hladine významnosti 0,05. P-hodnota je vyššia ako hladina významnosti, preto na hladine významnosti 0,05 nezamietame hypotézu o zhode stredných hodnôt porovnávaných súborov. Domnievame sa, že druh použitých diskov neovplyvňuje cenu počítačovej zostavy.

4.1.4 Porovnanie výsledkov

Ako sme videli, použitím dvoch rôznych verzií analýzy rozptylu sme dostali rôzne výsledky a vyvodili odlišné závery o vplyve skúmaného faktora. V prípade bežnej jednofaktorovej analýzy rozptylu sme zamietli hypotézu o rovnosti stredných hodnôt porovnávaných súborov. Naopak pri použití Welchovej analýzy rozptylu sme túto hypotézu nezamietli a usúdili sme, že premenná *druh diskov* nemá štatisticky významný vplyv na výslednú cenu počítačovej zostavy.

Na grafe 4.1 sú zobrazené stredné hodnoty zostáv rozčlenené podľa obmien faktora *druh diskov*. Vidíme, že zostavy, ktoré majú druh diskov *All Flash* a *All HDD*, majú približne rovnaké stredné hodnoty. Stredná hodnota hybridných zostáv je oproti nim len nepatrne vyššia.

Graf 4.1: Stredné hodnoty zostáv rozčlenené podľa obmien faktora *druh diskov*



Zdroj: Výstup zo SAS Enterprise Guide

Pri pohľade na číselné vyjadrenie stredných hodnôt výberových súborov vidíme, že medzi *All Flash* a *All HDD* zostavami je rozdiel nepatrný, len 17 dolárov. Hybridné zostavy majú strednú hodnotu vyššiu o viac ako 200 dolárov.

Tabuľka 4.5: Popisné charakteristiky vstupných údajov rozčlenené podľa skúmanej premennej

| Level of Druh diskov | N | value | |
|-------------------------|------|------------|------------|
| | | Mean | Std Dev |
| All_Flash | 1052 | 9991.8018 | 1004.08312 |
| All_HDD | 1541 | 10009.4208 | 974.82192 |
| Hybrid | 710 | 10269.4740 | 2939.72223 |

Zdroj: Výstup zo SAS Enterprise Guide

Zostavy s hybridnými diskami majú aj výrazne vyššiu variabilitu, takmer trojnásobnú oproti *All Flash* zostavám. Jedným z predpokladov použitia bežnej analýzy rozptylu je zhoda rozptylov porovnávaných súborov, preto sa v tomto prípade javia

výsledky Welchovej analýzy rozptylu ako vierohodnejšie. Treba však pripomenúť, že Welchova ANOVA nezamietla hypotézu o zhode stredných hodnôt len tesne (p -hodnota = 0,0533).

4.2 Porovnanie robustnosti analýzy rozptylu a Welchovej analýzy rozptylu

V predošlej časti práce sme zisťovali vplyv vybraného faktora na stredné hodnoty porovnávaných súborov. Na zisťovanie sme použili dve metódy – bežnú jednofaktorovú analýzu rozptylu a Welchovu analýzu rozptylu. Domnievali sme sa, že spoľahlivejšie výsledky by mala poskytovať Welchova ANOVA, keďže porovnávané súbory mali rôznu variabilitu. Jej výsledky však ukázali p -hodnotu veľmi blízko zvolenej hladiny významnosti. Naopak, bežná ANOVA jednoznačne potvrdila vplyv skúmaného faktora na stredné hodnoty. Preto je na mieste otázka, pri akých podmienkach a charakteristikách údajov je vhodnejšie použiť na testovanie hypotéz o zhode stredných hodnôt konkrétnu metódu. Na tento účel využijeme metódy simulácií Monte Carlo popísané v teoretickej časti tejto práce. Budeme nimi určovať robustnosť a silu testu pre spomenuté varianty analýzy rozptylu pri rôznych hodnotách parametrov analyzovaných údajov.

Naším cieľom bude zistiť, či je možné považovať Welchovu analýzu rozptylu za robustnejšiu alternatívu klasickej analýzy rozptylu a zároveň určiť robustnosť samotných testov. Robustnosť budeme skúmať pri porušení predpokladu zhody rozptylov v porovnávaných skupinách s rôznymi úrovňami analyzovaného faktora, rozsahu porovnávaných súborov a kombinácie týchto dvoch charakteristík. Všetky vygenerované údaje použité v tejto časti analýzy pochádzali z normálneho rozdelenia $N(\mu, \sigma^2)$. Parametre rozdelenia μ a σ^2 sa líšili v každom vygenerovanom súbore a jeho skupine s rozdielnou úrovňou faktora, tak ako je to uvedené v tabuľke 4.6. Stredná hodnota vo všetkých súboroch a všetkých ich skupinách mala hodnotu 0. Výberové súbory sme generovali a rozlišovali na základe nasledujúcich charakteristík:

1. V analýze boli skúmané vždy 3 úrovne faktora, teda 1 súbor rozdelený na podsúbory s početnosťami $(n_1; n_2; n_3)$ s rozdielnou úrovňou faktora. Vytvorili sme údaje s vyváženými aj nevyváženými skupinami podľa úrovne faktora. Pri vyvážených skupinách bola početnosť každej z nich 50 pozorovaní ($n_1 = n_2 = n_3$). Pri nevyvážených skupinách boli ich početnosti v pomere 50:60:75 ($n_1 < n_2 < n_3$).

2. Ďalšia charakteristika, s ktorou sme pracovali za účelom vytvorenia rôznych kombinácií súborov údajov, bol pomer rozptylov medzi porovnávanými skupinami. Pomer rozptylov, ktorý je jednoduchou mierou heteroskedasticity, definujeme ako pomer najmenšej a najväčšej hodnoty rozptylu v porovnávaných súboroch:

$$\text{pomer rozptylov} = \frac{\sigma_{min}^2}{\sigma_{max}^2}, \quad (4.1)$$

kde σ_{min}^2 je hodnota rozptylu v súbore s najmenšou variabilitou a σ_{max}^2 je hodnota rozptylu v súbore s najväčšou variabilitou.

Keselman a kol. (1998) uvádza, že v ním analyzovaných štúdiách je priemerná hodnota pomeru rozptylov 2.0 (smerodajná odchýlka 2.6), jeho medián 1.5 a maximálna hodnota 23.8. Pomer rozptylov najčastejšie používali aj Monte Carlo analýzy robustnosti analýzy rozptylu, ktoré boli zosumarizované v Blanca et al. (2018).

Do analýzy sme zaradili spolu 8 variantov porovnávaných súborov s pomerom rozptylov od 1:1 až po 1:30.

3. Skúmaný bol aj vplyv kombinácie pomeru rozptylov a početnosti v skupinách. V analýze sú rozlíšené aj možnosti, keď má súbor s najväčšou variabilitou najväčší počet pozorovaní spomedzi všetkých súborov ($\sigma_1^2 < \sigma_2^2 < \sigma_3^2$; $n_1 < n_2 < n_3$) a možnosť, keď súbor s najväčšou variabilitou má súčasne najmenší počet pozorovaní ($\sigma_1^2 > \sigma_2^2 > \sigma_3^2$; $n_1 < n_2 < n_3$).

Skombinovaním týchto charakteristík sme vytvorili 54 súborov, na ktorých bola vykonaná klasická ANOVA, ako aj Welchova ANOVA.

Tabuľka 4.6: Parametre údajov použitých pri analyzovaní robustnosti

| Počty pozorovaní | Pomer rozptylov |
|------------------|-----------------|
| 30:30:30 | 1:1 |
| | 1:1.5 |
| | 1:1.75 |
| | 1:2 |
| | 1:2.5 |
| | 1:3 |
| | 1:4 |
| | 1:5 |
| | 1:6 |
| | 1:7 |
| | 1:8 |
| | 1:9 |
| | 1:12 |
| | 1:15 |
| | 1:17.5 |
| | 1:20 |
| | 1:25 |
| | 1:30 |

| Počty pozorovaní | Pomer rozptylov |
|------------------|-----------------|
| 50:60:75 | 1:1 |
| | 1:1.5 |
| | 1:1.75 |
| | 1:2 |
| | 1:2.5 |
| | 1:3 |
| | 1:4 |
| | 1:5 |
| | 1:6 |
| | 1:7 |
| | 1:8 |
| | 1:9 |
| | 1:12 |
| | 1:15 |
| | 1:17.5 |
| | 1:20 |
| | 1:25 |
| | 1:30 |

| Počty pozorování | Pomer rozptylov |
|------------------|-----------------|
| 75:60:50 | 1:1 |
| | 1:1.5 |
| | 1:1.75 |
| | 1:2 |
| | 1:2.5 |
| | 1:3 |
| | 1:4 |
| | 1:5 |
| | 1:6 |
| | 1:7 |
| | 1:8 |
| | 1:9 |
| | 1:12 |
| | 1:15 |
| | 1:17.5 |
| | 1:20 |
| | 1:25 |
| | 1:30 |

Zdroj: Vlastné spracovanie

Z každej kombinácie popísaných parametrov bolo vygenerovaných 5000 súborov so zhodnými strednými hodnotami jednotlivých skupín. Zhoda stredných hodnôt bola v každom z 5000 súborov štatisticky testovaná klasickou jednofaktorovou analýzou rozptylu a zároveň Welchovou analýzou rozptylu. Pri všetkých štatistických testoch vykonaných v tejto časti bola zvolená hladina významnosti 5 %.

Pri každom vygenerovanom súbore údajov sme testovali hypotézu o rovnosti stredných hodnôt medzi jeho tromi skupinami.

Vieme, že skupiny s rôznymi úrovňami faktoru majú zhodné stredné hodnoty (tak sme ich vygenerovali), empiricky teda môžeme určiť mieru chýb I. druhu. Odhad miery chýb I. druhu ($\hat{\alpha}$) sa vypočítal ako počet testov, ktoré zamietli pravdivú nulovú hypotézu ($\sum_{i=1}^M I_i$) o zhode stredných hodnôt vydelenú počtom všetkých vykonaných testov, v našom prípade 5000:

$$\hat{\alpha} = \frac{1}{5000} \sum_{i=1}^M I_i. \quad (4.2)$$

Test budeme považovať za robustný, ak odhadnutá miera chýb I. druhu bude v rozmedzí od 0,025 po 0,075, podľa kritéria definovaného Bradleyom (1978).

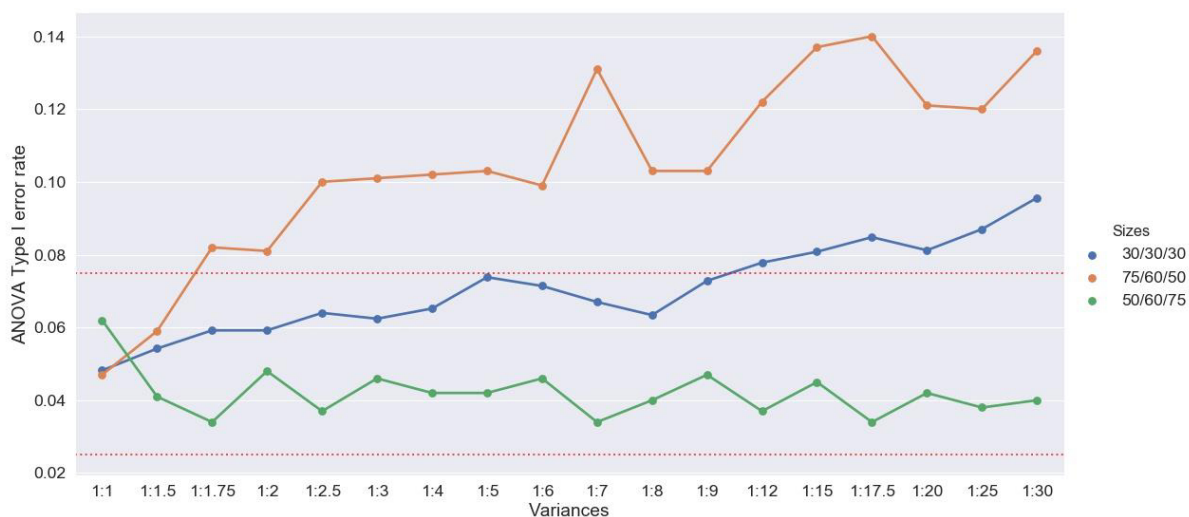
4.3 Výsledky analýzy robustnosti

Kompletná tabuľka s výsledkami odhadnutých mier chýb I. druhu na simulovaných údajoch pre klasickú analýzu rozptylu aj Welchovu analýzu rozptylu je uvedená v prílohe 1. Pre väčšiu prehľadnosť a čitateľnosť zobrazíme odhadnuté miery vo forme grafov a na základe nich vyslovíme závery o robustnosti konkrétneho testu.

4.3.1 Analýza rozptylu

Najprv sa bližšie pozrieme na zistenia v prípade použitia klasickej jednofaktorovej analýzy rozptylu. Výsledky v prílohe 1 môžeme prehľadnejšie zobrazit' pomocou spojnicového grafu. Na osi x sú súbory rozdelené podľa pomeru rozptylov medzi skupinami a na osi y je pre príslušný súbor uvedená miera chýb I. druhu. Prerušované čiary znázorňujú interval miery chýb I. druhu, v rámci ktorého môžeme hovoriť o robustnom teste (0,025 až 0,075).

Graf 4.2: Miera chýb I. druhu klasickej analýzy rozptylu



Zdroj: Vlastné spracovanie

Oranžovou farbou sú zobrazené kombinácie, kde skupina s najmenšou variabilitou mala najviac pozorovaní a súčasne skupina s najväčšou variabilitou mala najmenej pozorovaní. Zelenú farbu na grafe majú kombinácie, kde má skupina s najmenšou

variabilitou najmenej pozorovaní a súčasne skupina s najväčšou variabilitou najviac pozorovaní. Modrú farbu predstavujú kombinácie, kde majú všetky skupiny rovnako veľký počet pozorovaní.

Z grafu je zjavné, že pri splnení predpokladu o zhode rozptylov ($\sigma_1^2 = \sigma_2^2 = \sigma_3^2$) mala ANOVA mieru chýb I. druhu veľmi blízku zvolenej hladine významnosti 0,05 v prípade vyvážených aj nevyvážených počtoch pozorovaní. Takisto aj pri pomere rozptylov 1:1,5 boli vo všetkých prípadoch odhadnuté miery chýb I. druhu vo vyznačenom intervale.

V prípade nepriamej úmernosti variability a počtu pozorovaní vidíme, že už pri pomere rozptylov 1:1,75 je miera chýb I. druhu vyššia než 0,075. Hranicu 0,1 prekračuje miera chýb I. druhu pri pomere rozptylov 1:2,5. S rastúcim pomerom rozptylov má miera rastúcu tendenciu a jej hodnota dosahuje viac ako 0,12. Pri najväčšom pomere 1:30 dosahuje miera chýb I. druhu hodnotu 0,136. V prípade nepriamej úmernosti variability a počtu pozorovaní môžeme povedať, že ANOVA je robustná do okamihu, keď pomer rozptylov porovnávaných súborov nie je väčší než 1:1,5. Pri väčších pomeroch rozptylov je test príliš liberálny.

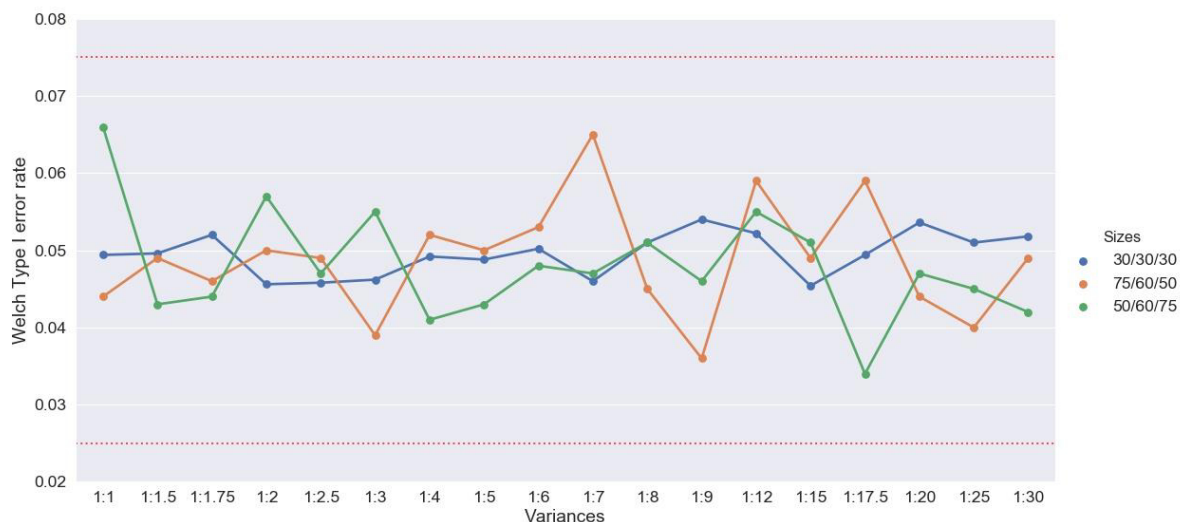
V prípade rovnakých početností v skupinách, je podľa našich zistení ANOVA robustná ak pomer rozptylov v súboroch nepresahuje 1:9. Zaujímavé je, že aj pri ďalších hodnotách pomerov je miera chýb I. druhu ešte stále pomerne blízko k hornej hranici robustnosti. V prípade pomeru 1:12 a viac už je test mimo intervalu robustnosti a stáva sa príliš liberálnym.

V situácií, keď má najmenej početný súbor najmenší rozptyl a naopak, najviac početný súbor najväčší rozptyl, sme zistili, že ANOVA je robustná pri všetkých uvažovaných pomeroch rozptylov. Vo všeobecnosti môžeme pozorovať, že miera chýb I. druhu je o niečo bližšie dolnému intervalu robustnosti.

4.3.2 Robustnosť Welchovej analýzy rozptylu

Výsledky robustnosti Welchovej analýzy rozptylu sú zobrazené podobne ako výsledky na grafe 4.2. Na osi x je zobrazený pomer rozptylov jednotlivých skupín v testovanom súbore a na osi y je zobrazená miera chýb I. druhu.

Graf 4.3: Miera chýb I. druhu Welchovej analýzy rozptylu



Zdroj: Vlastné spracovanie

Modrou farbou sú znázornené údaje v prípade, ak sú porovnávané skupiny rovnako početné. Zelenou farbou sú vyjadrené údaje, kde má skupina s najväčším počtom pozorovaní najväčšiu variabilitu a oranžová farba predstavuje údaje, kde má skupina s najmenším počtom pozorovaní najväčšiu variabilitu.

Na prvý pohľad je jasne viditeľné, že vo všetkých možných kombináciách sledovaných parametrov údajov bolo overovanie zhody stredných hodnôt Welchovou analýzou rozptylu robustné. V žiadnom prípade miera chýb I. druhu nepresiahla interval hodnôt 0,025 až 0,075. Na rozdiel od predchádzajúcej vizualizácie výsledkov klasickej analýzy rozptylu nie je badať významnejší rozdiel v prípade rôznych rozsahov porovnávaných skupín. V situácii, keď je rozsah pozorovaní v jednotlivých skupinách vyvážený, nepozorujeme tendenciu testu byť liberálnym tak, ako pri klasickej analýze rozptylu. Pri zvyšných dvoch kombináciách, kedy sú rozptyly v skupinách priamo úmerné počtu pozorovaní v skupine alebo naopak, takisto nevidno žiadny trend byť konzervatívnejším alebo liberálnejším. Vo všetkých troch prípadoch bola miera chýb, pri pomere rozptylov 1:30 I. druhu v intervale 0,04 až 0,06, čo sa len veľmi tesne líši od zvolenej hladiny významnosti 0,05.

Na základe výsledkov teda môžeme povedať, že Welchov F-test je skutočne robustný v situácii, keď majú porovnávané súbory rozdielnu variabilitu.

4.3.3 Zhodnotenie analýzy robustnosti

Z výsledkov analýzy robustnosti dvoch porovnávaných štatistických testov na overovanie zhody stredných hodnôt môžeme vyvodiť niekoľko užitočných záverov.

Pri pohľade na grafy zobrazujúce miery chýb I. druhu jasne vidno, že klasická ANOVA je vo všeobecnosti robustná, ak nie je pomer medzi skupinami s najmenším a najväčším rozptylom väčší ako 1:1,5.

Zaujímavé je aj zistenie, že klasická ANOVA je citlivá na vzťah medzi variabilitou súboru a jeho početnosťou. V situácií, keď je medzi rozptylom súborov a jeho početnosťou nepriama úmernosť, t.j. ($\sigma_1^2 < \sigma_2^2 < \sigma_3^2$; $n_1 > n_2 > n_3$), tak už pri pomere rozptylov 1:1,75 test nie je robustný a je príliš liberálny. Efekt heteroskedasticity na robustnosť testu sa zmierňuje v prípade, ak sú v porovnávaných súboroch počty pozorovaní priamo úmerné ich variabilite ($\sigma_1^2 < \sigma_2^2 < \sigma_3^2$; $n_1 < n_2 < n_3$). Miera chýb I. druhu sa vtedy nachádza v intervale robustnosti aj pri pomere rozptylov 1:30.

Welchova ANOVA dosiahla vo všetkých skúmaných kombináciách parametrov mieru chýb I. druhu blízku zvolenej hladine významnosti. Môžeme ju považovať za robustnú aj pri najvyššom testovanom pomere rozptylov (1:30).

Na základe porovnaní vo všeobecnosti môžeme považovať analýzu rozptylu pri použití Welchovho F-testu za výrazne robustnejšiu alternatívu ku klasickej analýze rozptylu, v prípade, keď je porušený predpoklad zhody rozptylov porovnávaných súborov. Jej zistené miery chýb I. druhu sú len málo odlišné od zvolenej hladiny významnosti.

4.4 Porovnanie sily testu analýzy rozptylu a Welchovej analýzy rozptylu

V nasledujúcej časti budeme porovnávať silu testu v prípade klasickej analýzy rozptylu a Welchovej analýzy rozptylu. Na určenie sily testu musíme vypočítať mieru chýb II. druhu. Budeme teda skúmať do akej miery vie test odhaliť, že údaje, ktoré sme vygenerovali majú rozdielne stredné hodnoty.

Takisto ako pri určení robustnosti štatistických testov, aj tu využijeme simulovanie údajov so želanými parametrami. Z týchto vygenerovaných údajov budeme môcť odhadnúť, akú mieru chýb II. druhu a následne akú silu, mali klasická a Welchova analýza rozptylu. Odhad miery chýb II. druhu ($\hat{\beta}$) vypočítame ako počet testov, kde sa prijala na zvolenej hladine významnosti 0,05 nulová hypotéza o rovnosti rozptylov medzi porovnávanými súbormi ($\sum_{i=1}^M I_i$), vydelení celkovým počtom vykonaných testov, ktorý v

$$\hat{\beta} = \frac{1}{5000} \sum_{i=1}^M I_i \quad (4.3)$$

Silu testu vypočítame, keď mieru chýb II. druhu odčítame od 1.

Postupne vytvoríme kombinácie údajov, kde má jedna z porovnávaných hodnôt rozdielnú strednú hodnotu ako ostatné. Rozhodli sme sa, že skupiny s rovnakými strednými hodnotami μ_i , ju budú mať na úrovni 0 a zvyšná, tretia skupina, bude mať strednú hodnotu 1. Výberové súbory sme generovali a rozlišovali na základe nasledujúcich charakteristík:

1. Parameter, ktorého vplyv budeme chcieť zistiť je variabilita medzi skupinami. Podobne ako pri analýze robustnosti, aj teraz vytvoríme viacero kombinácií údajov s pomermi rozptylov σ_i^2 medzi skupinami od hodnoty 1:1,5 až po 1:9.
2. Budeme uvažovať s dvomi možnosťami počtu pozorovaní v porovnávaných súboroch. Najprv s vyrovnanými početnosťami vo všetkých súboroch: ($n_1 = n_2 = n_3$) a následne s nevyváženými početnosťami podobne ako v prípade určovania miery chýb I. druhu ($n_1 = 50$; $n_2 = 60$; $n_3 = 75$).
3. Zvlášť vytvoríme kombinácie, kde je:
 - počet pozorovaní v skupine priamo úmerný jej variabilite ($\sigma_1^2 < \sigma_2^2 < \sigma_3^2$; $n_1 < n_2 < n_3$),
 - počet pozorovaní v skupine nepriamo úmerný jej variabilite ($\sigma_1^2 < \sigma_2^2 < \sigma_3^2$; $n_1 > n_2 > n_3$),
 - počet pozorovaní v skupinách s rôznou variabilitou vyrovnaný ($\sigma_1^2 < \sigma_2^2 < \sigma_3^2$; $n_1 = n_2 = n_3$).
4. Rozlíšime aj to, či je rozdielna stredná hodnota μ_i , v skupine s najmenším počtom pozorovaní ($n_1 < n_2 < n_3$; $\mu_1 \neq \mu_2 = \mu_3$) alebo v tej s najväčším ($n_1 > n_2 > n_3$; $\mu_1 \neq \mu_2 = \mu_3$).

Všetky uvažované kombinácie uvedených charakteristík dokopy tvoria 72 možností, na základe ktorých budeme generovať údaje.

Tabuľka 4.7: Parametre údajov použitých pri analyzovaní sily testu

| Počty pozorovaní | Stredné hodnoty | Pomer rozptylov |
|------------------|-----------------|-----------------|
| 30:30:30 | 0:0:1 | 1:1 |
| | | 1:1.5 |
| | | 1:1.75 |
| | | 1:2 |
| | | 1:2.5 |
| | | 1:3 |
| | | 1:4 |
| | | 1:5 |
| | | 1:6 |
| | | 1:7 |
| | | 1:8 |
| | | 1:9 |
| | 1:0:0 | 1:1 |
| | | 1:1.5 |
| | | 1:1.75 |
| | | 1:2 |
| | | 1:2.5 |
| | | 1:3 |
| | | 1:4 |
| | | 1:5 |
| | | 1:6 |
| | | 1:7 |
| | | 1:8 |
| | | 1:9 |

Zdroj: Vlastné spracovanie

Tabuľka 4.8: Parametre údajov použitých pri analyzovaní sily testu

| Počty pozorovaní | Stredné hodnoty | Pomer rozptylov |
|------------------|-----------------|-----------------|
| 50:60:75 | 0:0:1 | 1:1 |
| | | 1:1.5 |
| | | 1:1.75 |
| | | 1:2 |
| | | 1:2.5 |
| | | 1:3 |
| | | 1:4 |
| | | 1:5 |
| | | 1:6 |
| | | 1:7 |
| | | 1:8 |
| | | 1:9 |
| | 1:0:0 | 1:1 |
| | | 1:1.5 |
| | | 1:1.75 |
| | | 1:2 |
| | | 1:2.5 |
| | | 1:3 |
| | | 1:4 |
| | | 1:5 |
| | | 1:6 |
| | | 1:7 |
| | | 1:8 |
| | | 1:9 |

Zdroj: Vlastné spracovanie

Tabuľka 4.9: Parametre údajov použitých pri analyzovaní sily testu

| Počty pozorovaní | Stredné hodnoty | Pomer rozptylov |
|------------------|-----------------|-----------------|
| 75:60:50 | 0:0:1 | 1:1 |
| | | 1:1.5 |
| | | 1:1.75 |
| | | 1:2 |
| | | 1:2.5 |
| | | 1:3 |
| | | 1:4 |
| | | 1:5 |
| | | 1:6 |
| | | 1:7 |
| | 1:0:0 | 1:8 |
| | | 1:9 |
| | | 1:1 |
| | | 1:1.5 |
| | | 1:1.75 |
| | | 1:2 |
| | | 1:2.5 |
| | | 1:3 |
| | | 1:4 |
| | | 1:5 |
| | | 1:6 |
| | | 1:7 |
| | | 1:8 |
| | | 1:9 |

Zdroj: Vlastné spracovanie

Z každej kombinácie popísaných parametrov bolo vygenerovaných 5000 súborov s rozdielnymi strednými hodnotami porovnávaných súborov. Rovnosť ich stredných hodnôt bola v každom z vygenerovaných súborov overovaná jednofaktorovou analýzou rozptylu a zároveň Welchovou analýzou rozptylu. Uvažovaná hladina významnosti pri oboch testoch bola na úrovni 5 %. Z dôvodu prehľadnosti a vhodnosti vzájomného porovnania budeme analyzovať zvlášť výsledky pre 3 skupiny kombinácií menených parametrov:

- súbory s rovnakými početnosťami ($n_1 = n_2 = n_3$),
- súbory kde má skupina s najväčšou variabilitou najväčší počet pozorovaní a skupina s najmenšou variabilitou najmenej pozorovaní (priama úmernosť - $\sigma_1^2 < \sigma_2^2 < \sigma_3^2$; $n_1 < n_2 < n_3$),
- súbory, kde má skupina s najväčšou variabilitou najmenej pozorovaní a skupina s najmenšou variabilitou najviac pozorovaní (nepriama úmernosť - $\sigma_1^2 < \sigma_2^2 < \sigma_3^2$; $n_1 > n_2 > n_3$).

4.4.1 Výsledky analýzy sily testu

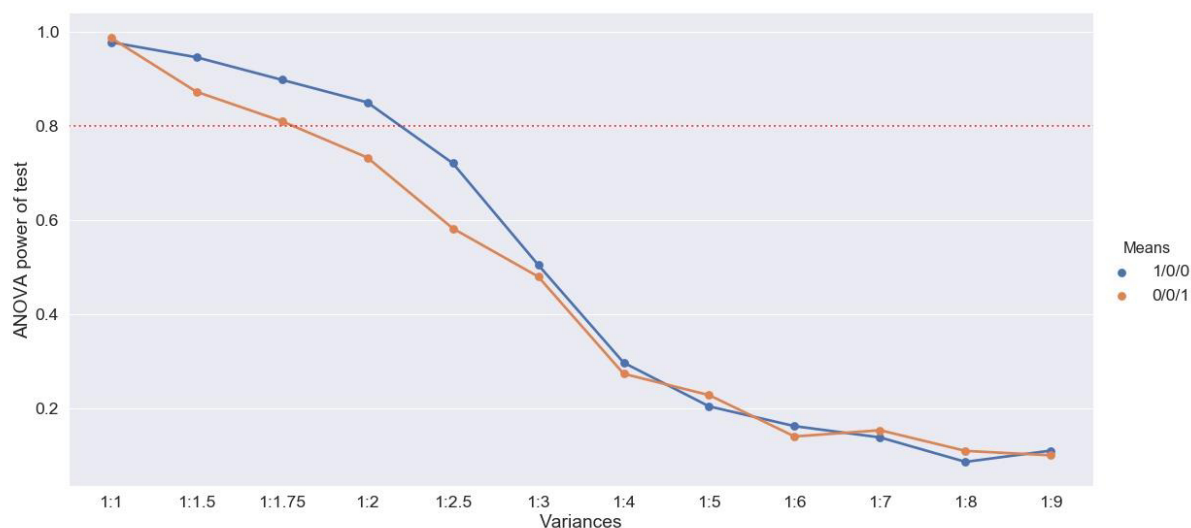
Výsledky sú zobrazené podobne ako v prípade analyzovania mier chýb I. druhu vo forme spojnicového grafu, kde na osi x sú zobrazené varianty pozorovaní podľa pomeru rozptylu v skupinách a na osi y je zobrazená hodnota sily testu. Modrá čiara predstavuje varianty, kde má skupina s odlišnou strednou hodnotou najmenšiu variabilitu. Oranžová čiara označuje varianty, kde má skupina s odlišnou strednou hodnotou najväčšiu variabilitu. Vedľa seba sú zobrazené výsledky pre klasickú analýzu rozptylu a Welchovu analýzu rozptylu. Červená priamka predstavuje hranicu sily testu 0,8. Vo väčšine oblastí výskumu sa bežne považuje sila testu pod 0,8 za príliš nízku (Walmsley, Brown, 2017).

Kompletná tabuľka s výsledkami odhadnutej sily testu na simulovaných údajoch pre klasickú analýzu rozptylu aj Welchovu analýzu rozptylu sú uvedené v časti príloh.

4.4.1.1 Rovnomerné počty pozorovaní

V nasledujúcej kapitole uvedieme výsledky analýzy sily testu v prípade, keď porovnávané súbory obsahovali rovnaký počet pozorovaní. Každý z nich obsahoval 30 pozorovaní ($n_1 = n_2 = n_3 = 30$).

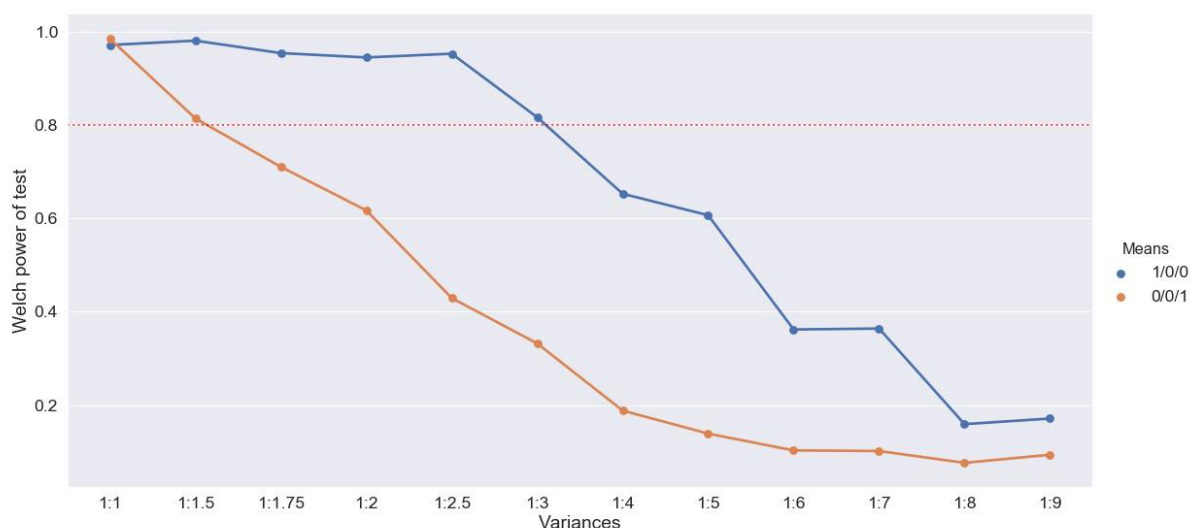
Graf 4.4: Sila klasickej analýzy rozptylu pri rovnomernom počte pozorovaní



Zdroj: Vlastné spracovanie

Vidíme, že v prípade klasickej analýzy rozptylu sú obe krivky veľmi podobné. V prípade rovnakých rozptylov (pomer rozptylov 1:1) bola zistená sila testu na veľmi vysokej úrovni, približne 98%. To znamená, že z 5000 vygenerovaných údajov, test v 100 prípadoch nesprávne potvrdil nulovú hypotézu o zhode stredných hodnôt porovnávaných súborov. V prípade, keď mal súbor s odlišnou strednou hodnotou najmenšiu variabilitu, bola sila testu až do pomeru rozptylov 1:2 nad úrovňou 80 %. Pri variante, keď súbor s odlišnou strednou hodnotou mal najväčšiu variabilitu, klesla sila testu pod 80 % o niečo skôr, už pri pomere rozptylov 1:1.75. Pri pomere rozptylov 1:3 už sila testu klesla k hodnote 50 %. Pri pomere 1:5 to bolo už len okolo 20 %.

Graf 4.5: Sila Welchovej analýzy rozptylu pri rovnomernom počte pozorovaní



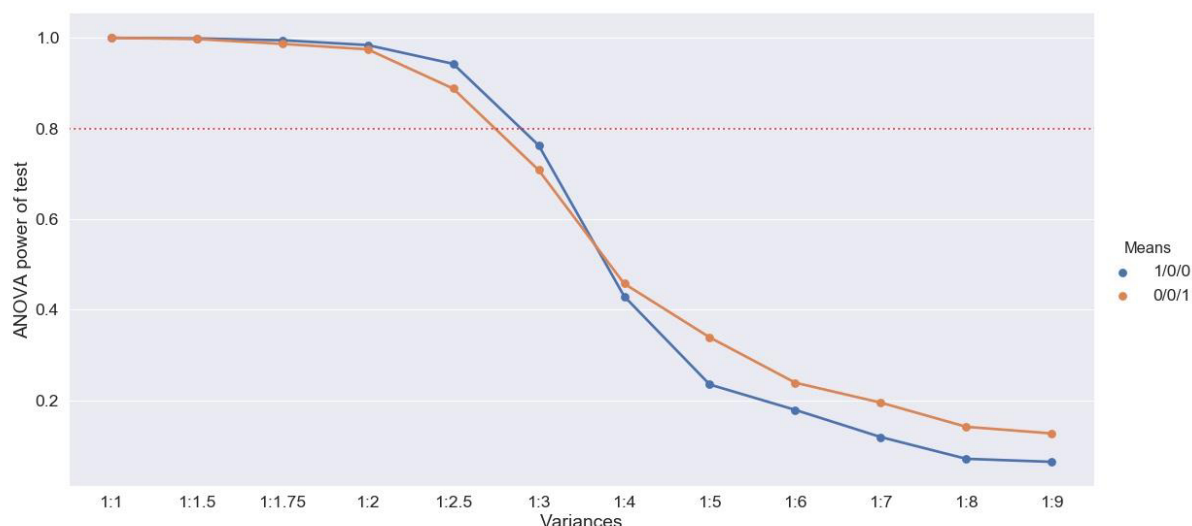
Zdroj: Vlastné spracovanie

Zaujímavé výsledky boli zistené pri analyzovaní sily testu Welchovej analýzy rozptylu. Pri situácii, keď mal súbor s rozdielnou strednou hodnotou najväčší rozptyl, sme zaznamenali podobný pokles, ako v prípade klasickej analýzy rozptylu. Pokles sily testu pod 80 % však nastal už pri pomere rozptylov 1:1,5. V prípade, keď bola rozdielna stredná hodnota v súbore s menšou variabilitou bola sila testu vyššia ako 90 % aj pri pomere rozptylov 1:2,5. Ešte pri pomere 1:3 bola sila testu vyššia ako 80 %.

4.4.1.2 Priama úmernosť medzi variabilitou a počtom pozorovaní

V tejto kapitole uvedieme výsledky analýzy sily testu v prípadoch, keď skupina údajov s najmenšou variabilitou obsahovala najmenší počet pozorovaní a súčasne skupina s najväčšou variabilitou obsahovala najviac pozorovaní ($\sigma_1^2 < \sigma_2^2 < \sigma_3^2$; $n_1 < n_2 < n_3$). Konkrétne počty pozorovaní boli vždy v pomere 50:60:75.

Graf 4.6 Sila analýzy rozptylu pri priamej úmernosti medzi variabilitou a počtom pozorovaní

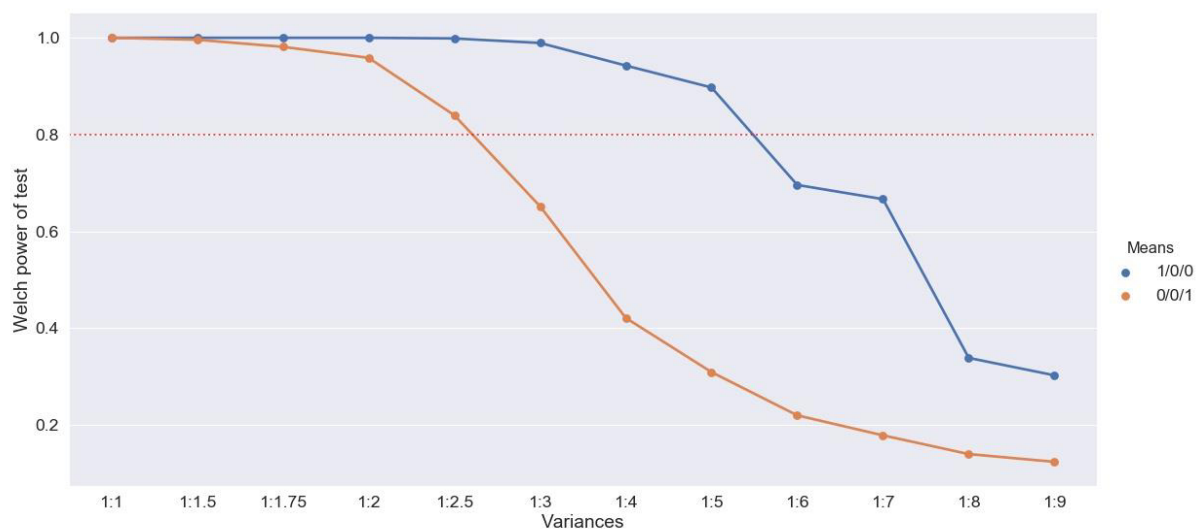


Zdroj: Vlastné spracovanie

Pri klasickej analýze rozptylu boli výsledky veľmi podobné v oboch uvažovaných umiestneniach rozdielnej strednej hodnoty. Pri pomere rozptylov 1:2 bola sila testu na úrovni približne 98 %. V prípade, keď bol pomer medzi rozptylmi 1:2,5, sila testu ešte presahovala 80 % a pri pomere 1:3 už len okolo 70 %.

V prípade Welchovej analýzy rozptylu môžeme, podobne ako pri variante s rovnomerným počtom pozorovaní, vidieť rozdiel medzi dvomi krivkami. V situácií, keď mal súbor s rozdielnou strednou hodnotou najväčší rozptyl, bola sila testu veľmi podobná bežnej analýze rozptylu a 80 % presiahla pri pomere rozptylov 1:2,5. Naopak, keď bola rozdielna stredná hodnota v súbore s najmenšou variabilitou, aj pri pomere rozptylov 1:5 bola sila testu takmer 90 %.

Graf 4.7: Sila Welchovej analýzy rozptylu pri priamej úmernosti medzi variabilitou a počtom pozorovaní

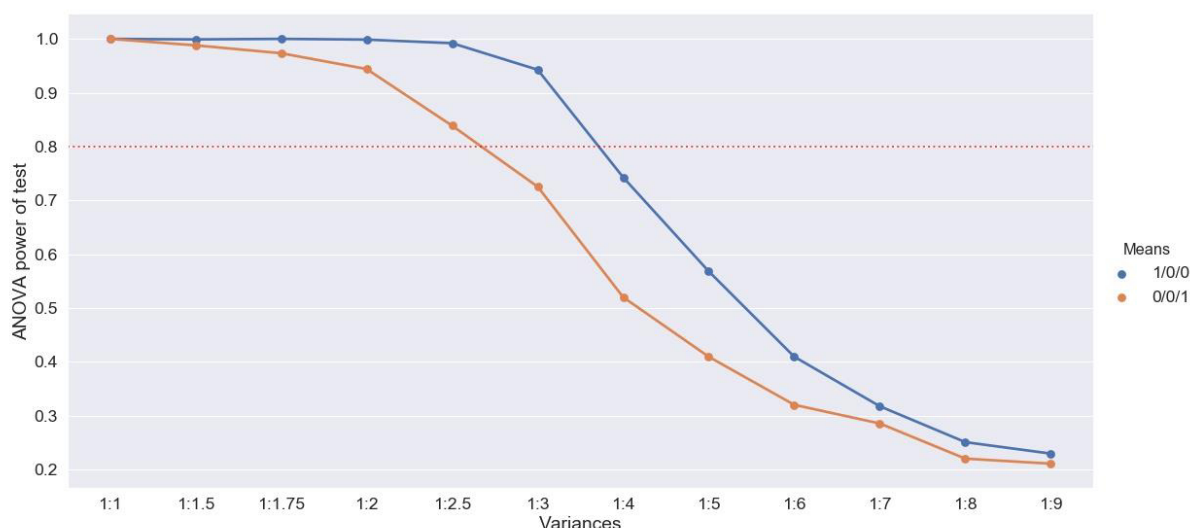


Zdroj: Vlastné spracovanie

4.4.1.3 Nepriama úmernosť medzi variabilitou a počtom pozorovaní

Nasledujúca kapitola obsahuje výsledky analýzy sily testu v prípade, keď súbor s najmenšou variabilitou obsahoval najväčší počet pozorovaní a súčasne súbor s najväčšou variabilitou obsahoval najmenej pozorovaní ($\sigma_1^2 < \sigma_2^2 < \sigma_3^2$; $n_1 > n_2 > n_3$).

Graf 4.8: Sila analýzy rozptylu pri nepriamej úmernosti medzi variabilitou a počtom pozorovaní

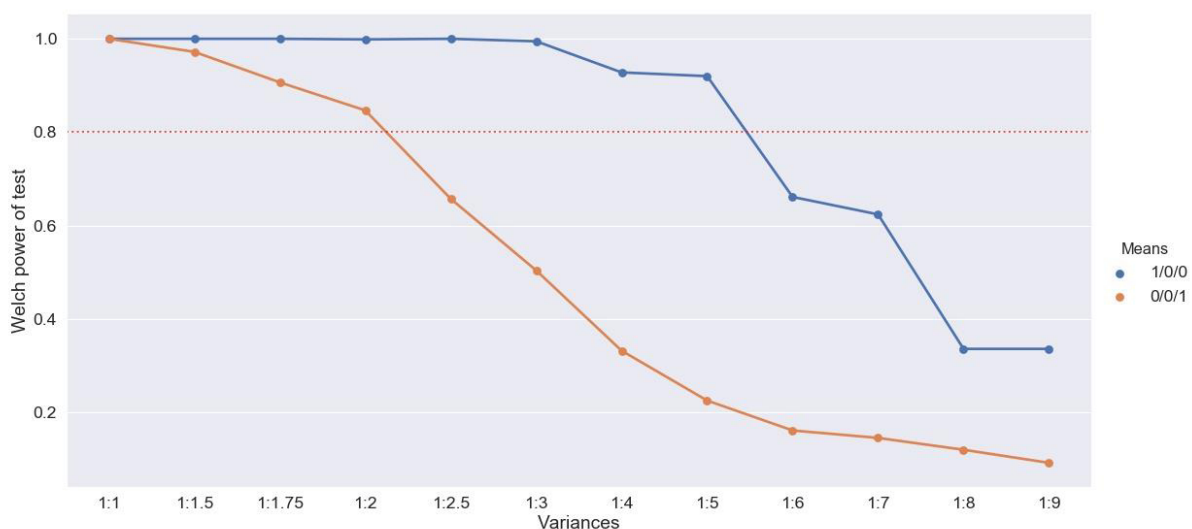


Zdroj: Vlastné spracovanie

V prípade použitia klasickej analýzy rozptylu je vidieť mierny rozdiel medzi dvoma umiestneniami rozdielnej strednej hodnoty. Pri možnosti, keď mal súbor s rozdielnou strednou hodnotou najväčší rozptyl, presiahol test hranicu 80 %, ak bol pomer rozptylov

najviac 1:2,5. V prípade, ak sa odlišná stredná hodnota nachádzala v súbore s najmenším rozptylom, tak sa sila 80 % presiahla ešte pri pomere 1:3.

Graf 4.9: Sila Welchovej analýzy rozptylu pri nepriamej úmernosti medzi variabilitou a počtom pozorovaní



Zdroj: Vlastné spracovanie

Na výsledkoch Welchovej analýzy rozptylu vidíme jasný rozdiel v dvoch možnostiach umiestnenia rozdielnej strednej hodnoty. Pri možnosti, keď bola rozdielna stredná hodnota v súbore s najväčšou variabilitou, sila testu presiahla 80 % pri pomere rozptylov 1:2. Ak mal rozdielnú strednú hodnotu súbor s najmenšou variabilitou, tak sila testu klesla pod 80 % až pri pomere rozptylov 1:6.

4.4.2 Zhodnotenie analýzy sily testov

Z uvedených výsledkov analýz je zrejmé, že vo všetkých testovaných prípadoch bolo podstatné, či mal odlišnú strednú hodnotu súbor s najväčšou alebo najmenšou variabilitou. Vo všeobecnosti sa dá povedať, že ako klasická ANOVA, tak aj Welchova ANOVA dosahujú podstatne vyššiu silu testu, ak má odlišnú strednú hodnotu súbor s menšou variabilitou.

V takomto prípade je sila Welchovej analýzy rozptylu viac ako 80 % pri pomere rozptylov 1:3, ak sú vo všetkých porovnávaných súboroch rovnaké počty pozorovaní. V situácii, keď majú súbory s vyššou variabilitou väčšiu relatívnu početnosť sa hranica sily testu posúva vyššie, až na pomer 1:5. Ak má odlišnú strednú hodnotu súbor s vyššou

variabilitou, tak sa hranica 80 % sily testu mierne znižuje. V prípade rovnomerného počtu pozorovaní v porovnávaných súboroch sa táto hranica presiahne len v prípade, ak pomer rozptylov nie je vyšší ako 1:1,5.

Klasická analýza rozptylu vo všeobecnosti dosahovala porovnateľné, až lepšie výsledky ako Welchova ANOVA, ak bola rozdielna stredná hodnota v súbore s menšou variabilitou. V prípade rovnomerného počtu pozorovaní v súboroch dosiahla aspoň 80 % silu testu pri pomere rozptylov najviac 1:1,75, čo je vyššia hodnota, ako v prípade Welchovej analýzy rozptylu (1:1,5). V ostatných dvoch prípadoch boli hranice 80 % sily testu o niečo vyššie (1:2,5). Naopak, ak bola rozdielna stredná hodnota v súbore s vyššou variabilitou, klasická ANOVA dosahovala 80 % silu testu pri nižších pomeroch rozptylov oproti Welchovej analýze rozptylu.

Ak budeme uvažovať aj s najnižším dosiahnutým výsledkom zo simulácií, tak vo všeobecnosti môžeme povedať, že Welchova ANOVA dosahuje uspokojivú silu testu (80 %), ak nie je pomer rozptylov porovnávaných skupín vyšší ako 1:1,5. O čosi lepšie je v tomto porovnaní klasická ANOVA, ktorá spoľahlivo dosahuje silu testu 80 % aj pri pomere rozptylov 1:1,75.

Záver

V diplomovej práci sme sa zaoberali problematikou vzájomného porovnávania viac ako dvoch výberových súborov a overovaním rovnosti ich parametrov. Hlavným cieľom práce bolo porovnať klasickú verziu analýzy rozptylu a Welchovu analýzu rozptylu, z hľadiska robustnosti a sily testu. Na základe poznatkov popísaných v teoretickej časti práce sme aplikovali metódy analýzy rozptylu a Welchovej analýzy rozptylu na reálne údaje popisujúce cenu počítačových zostáv. Skúmali sme vplyv druhu použitých diskov na výslednú cenu zostavy. Následne sme analyzovali, ako vplýva heteroskedasticita a ďalšie faktory na robustnosť a silu testu oboch štatistických metód. Na posúdenie robustnosti a sily testov sme miery chýb I. a II. druhu odhadli pomocou Monte Carlo simulácií.

Prvá časť práce stručne popisuje súčasný stav skúmanej problematiky. Popísali sme tu základnú myšlienku analýzy rozptylu, jej históriu a možnosti využitia. Ďalej sme sa venovali téme chýb štatistických testov. Vysvetlili sme, ako delíme chyby testov, ich miery a pojmy robustnosť a sila štatistického testu. Popísali sme, ako sa využívajú takzvané Monte Carlo simulácie na odhad mier chýb I. a II. druhu.

V druhej kapitole sme určili hlavný cieľ práce, spolu s čiastkovými cieľmi. Tretia kapitola je venovaná metodike, ktorú sme využili v aplikačnej časti práce. V podkapitole 3.1 sme ozrejmili dôvody, prečo dvojvýberový t-test nie je vhodný na overovanie rovnosti stredných hodnôt viac ako dvoch výberových súborov. Zdôraznili sme jeho vysokú mieru chýb I. druhu pri takomto použití. V podkapitole 3.2 sme sa venovali dvojvýberovému F-testu zhody rozptylov, ktorý je používaný v analýze rozptylu a jeho modifikácie sa používajú aj v ďalších skúmaných štatistických testoch. V kapitole 3.3 a jej podkapitolách je podrobnejšie popísaná metóda analýzy rozptylu. Zapísali sme model jednofaktorovej analýzy rozptylu a definovali kľúčové pojmy. Následne sme vysvetlili a odvodili rozklad celkovej variability, na variabilitu spôsobenú rozdielnymi úrovňami skúmaného faktoru a variabilitu náhodných vplyvov. Ďalej sme uviedli, ako sa na základe týchto zdrojov variability zostrojí testovacia štatistika analýzy rozptylu. Nakoniec sme znázornili, ako sa výsledky analýzy rozptylu zapíšu do prehľadnej tabuľkovej formy. Kapitola 3.4 je venovaná Welchovej analýze rozptylu, ktorá je modifikáciou bežnej analýzy rozptylu. Ozrejmili sme, ako sa v prípade Welchovej analýzy rozptylu zohľadňuje heteroskedasticita a zostrojuje sa testovacia štatistika. V kapitole 3.5 sme popísali Kruskal-Wallisov test,

ktorý je vhodnou alternatívou k analýze rozptylu, ak nie je splnený predpoklad normality údajov.

Predpokladom použitia analýzy rozptylu – normalite, nezávislosti pozorovaní a homoskedasticite, je venovaná kapitola 3.6. Stručne sme popísali, akými metódami je možné overiť normalitu údajov a prečo je dôležité, aby boli testované údaje navzájom nezávislé. Podrobnejšie sme sa venovali predpokladu homoskedasticity (zhody rozptylov). Uviedli sme viaceré štatistické testy na overovanie homoskedasticity. V kapitole 3.8 sme uviedli zistenia vedeckých štúdií, ktoré sa zaoberali vplyvom porušenia predpokladov normality a homoskedasticity na robustnosť analýzy rozptylu. V praktickej časti sme vychádzali z metodiky, ktorá bola v tejto časti popísaná. Kapitola 3.9 je venovaná simulačným metódam Monte Carlo. Vysvetlili sme na akom princípe funguje a uviedli sme všeobecný postup riešenia pomocou tejto metódy. Následne sme sa zaoberali využitím metód Monte Carlo v induktívnej štatistike. Bližšie sme sa venovali použitiu metód na stanovenie mier chýb I. a II. druhu štatistických testov a uviedli sme, ako v tomto prípade postupovať.

Vo štvrtej kapitole sme aplikovali poznatky vysvetlené v metodickej časti práce na reálne údaje popisujúce ceny rôznych konfigurácií počítačových zostáv. Analýzou rozptylu a následne Welchovou analýzou rozptylu sme skúmali vplyv druhu použitých diskov na výslednú cenu zostavy. Na výpočty a vizualizácie nám slúžil analytický softvér SAS Enterprise Guide. Použitím bežnej analýzy rozptylu sme dospeli k záveru, že použitý druh diskov v počítačovej zostave ovplyvňuje jej výslednú cenu. Rozdielny bol výsledok Welchovej analýzy rozptylu, kedy sme usúdili, že druh diskov výslednú cenu neovplyvňuje.

V kapitole 4.2 a 4.3 sme porovnali robustnosť klasickej analýzy rozptylu a Welchovej analýzy rozptylu pomocou Monte Carlo simulácií. Na generovanie potrebných údajov a stanovenie mier chýb I. a II. druhu sme využili programovací jazyk Python a jeho viaceré moduly zamerané na štatistické zisťovanie.

V analýze sme skúmali, aký vplyv má na silu testov a ich robustnosť heteroskedasticita a vzťah medzi rozptylmi porovnávaných súborov a ich početnosťami. Výsledky analýzy môžu slúžiť na určenie toho, pri akých parametroch údajov je vhodnejšie použiť Welchovu analýzu rozptylu a kedy klasickú analýzu rozptylu. Analýzou sme dospeli k viacerým zaujímavým zisteniam.

Z analýzy robustnosti a miery chýb I. druhu vyplýva, že heteroskedasticita výrazne ovplyvňuje spoľahlivosť výsledkov analýzy rozptylu. V prípade analýzy rozptylu môžeme povedať, že s rastúcou mierou heteroskedasticity (pomerom rozptylov) sa zväčšuje miera chýb I. druhu. Heteroskedasticita nemá vplyv na mieru chýb I. druhu pri Welchovej analýze rozptylu. Aj pri pomere rozptylov 1:30 sa miera chýb I. druhu približuje zvolenej hladine významnosti.

Podstatný vplyv na mieru chýb I. druhu má vzťah, medzi rozptylmi porovnávaných súborov a ich početnosťami. Pri klasickej analýze rozptylu zostáva miera chýb I. druhu blízka zvolenej hladine významnosti, ak sú v porovnávaných súboroch počty pozorovaní priamo úmerné ich variabilite.

Heteroskedasticita má významný vplyv na silu oboch skúmaných testov. So zvyšujúcim sa pomerom rozptylov sila testu klesá. Klasická analýza rozptylu dosahuje silu aspoň 80 %, ak pomer rozptylov nepresahuje hodnotu 1:1,75. Welchova analýza dosahuje 80 % silu testu ak nie je pomer rozptylov väčší ako 1:1,5.

Silu testu významne ovplyvňuje to, či má odlišnú strednú hodnotu súbor s najväčšou alebo najmenšou variabilitou. Z tohto dôvodu sa odporúča, aby sme odhadli stredné hodnoty z charakteristík výberových súborov. Vo všeobecnosti sa dá povedať, že obe skúmané metódy dosahujú podstatne vyššiu silu testu, keď má odlišnú strednú hodnotu súbor s menšou variabilitou. Tento vplyv je výraznejší pri Welchovej analýze rozptylu. Hranica 80 % sily testu je v tomto prípade posunutá na hodnotu pomeru rozptylov 1:3 až 1:5, v závislosti od vzťahu medzi rozptylmi porovnávaných súborov a ich početnosťami.

V situácií, keď sa rozhodujeme, ktorú z dvoch štatistických metód použiť, sa môžeme riadiť niekoľkými odporúčaniami.

Heteroskedasticita má významný vplyv na silu aj robustnosť oboch testov. Z tohto dôvodu je overovanie predpokladu rovnosti rozptylov dôležité pri rozhodovaní, ktorý test použijeme. Prvým krokom pri výbere testu by tak malo byť overenie zhody rozptylov porovnávaných súborov, napríklad Leveneovým testom.

Ak sa testovaním ukáže, že porovnávané súbory majú rovnaké rozptyly alebo je pomer medzi rozptylmi súborov najviac 1:1,5, môžeme použiť klasickú analýzu rozptylu. V tomto prípade analýza rozptylu dosahuje vysokú silu testu a je robustná.

V prípade, ak majú porovnávané súbory pomer rozptylov väčší ako 1:1,5, je lepšie použiť Welchovu analýzu rozptylu. Oproti klasickej analýze rozptylu dosiahneme výrazne menšiu mieru chýb I. druhu. V určitých vyššie spomínaných prípadoch môže byť nevýhodou menšia sila testu.

V práci pracujeme s mnohými faktormi, ktoré vplývajú na robustnosť a silu testu metód analýzy rozptylu. Našu analýzu by bolo možné rozšíriť aj o ďalšie faktory. Zaujímavé by bolo analyzovať robustnosť a silu testu pri viac ako troch porovnávaných skupinách. Mohli by sme skúmať aj vplyv normality na robustnosť a silu testov. V tom prípade by bolo vhodné zaradiť medzi porovnávané metódy aj Kruskal-Wallisov test.

Zoznam použitej literatúry

Knižné publikácie

1. ABDI, Hervé. *O'Brien test for homogeneity of variance*. In *Encyclopedia of measurement and statistics*, London : SAGE Publication Ltd. 2007, s. 701-704. ISBN 9781412916110.
2. ĎŘÍMAL, Jiří, TRUNEC, David. *Úvod do metody Monte-Carlo*. 1. vyd. Brno: UJEP, 1989, 122 s. ISBN 80-210-0022-8.
3. FIELD, Andy. *Discovering statistics using IBM SPSS statistics: North American edition*. Sage, 2017. 816 s. ISBN 978-1526436566
4. KANJI, Gopal K. *100 Statistical Tests*. London : SAGE Publication Ltd. 1999, 215 s. ISBN 0-7619-6151-8
5. KOTLEBOVÁ, Eva, Erik ŠOLTĚS a Daniela SIVAŠOVÁ. *Štatistická indukcia v príkladoch*. Bratislava: Vydavateľstvo EKONÓM, 2015, 222 s. ISBN 978-80-225-4135-0.
6. LIU, Hangcheng. *Comparing Welch ANOVA, a Kruskal-Wallis test, and traditional ANOVA in case of heterogeneity of variance*. Richmond, Virginia: Virginia Commonwealth University. 2015. 46 s.
7. MARTINEZ, Wendy L., MARTINEZ Angel R. *Computational statistics handbook with MATLAB*. 2. vyd. Boca Raton, FL: Chapman, 2008, 767 s. ISBN 15-848-8566-1.
8. MAXWELL, Scott E., DELANEY, Harold D., KELLEY, Ken. *Designing experiments and analyzing data: A model comparison perspective*. Routledge, 2017. 1080 s. ISBN 978-1138892286.
9. MCDONALD, J.H. *Handbook of Biological Statistics (3rd ed.)*. Baltimore, Maryland: Sparky House Publishing, 2014, 305 s. ISBN 978-1257140183.
10. MONTGOMERY, Douglas C. *Design and analysis of experiments*. 9. vyd. New York, United States: John Wiley & Sons, 2017. 640 s. ISBN 978-1-119-32093-7.
11. NETER, J., WASSERMAN, W., a KUTNER, M. H. *Applied linear statistical models: Regression, analysis of variance, and experimental designs*. Homewood, IL: Irwin. 1990, 1181 s. ISBN 978-0256083385.

12. PACÁKOVÁ, Viera, Eva KOTLEBOVÁ, Viera LABUDOVÁ a Ľubica SIPKOVÁ. *Štatistická indukcia pre ekonómov*. Bratislava: Vydavateľstvo EKONÓM, 2012, 317 s. ISBN 978-80-225-3382-9.
13. PACÁKOVÁ, Viera, LABUDOVÁ, Viera, SIPKOVÁ, Ľubica, STANKOVIČOVÁ, Iveta. *Štatistická indukcia pre ekonómov a manažérov*. Bratislava: Wolters Kluwer, 2015, 365 s. ISBN 978-80-8168-081-6.
14. SHERMER, Michael. *The Skeptic Encyclopedia of Pseudoscience 2 volume set*. ABC-CLIO, 2002. 903 s. ISBN 1576076539.
15. SCHEFFÉ, Henry. *The analysis of variance*. New York, United States: John Wiley & Sons, 1999. 477 s. ISBN 978-0-471-34505-3
16. ZAR, J. H. *Biostatistical Analysis*. 5. vyd. Englewood Cliffs, NJ: Prentice Hall, Inc. 2010. 960 s. ISBN 978-0321656865.

Články v časopise

1. BLANCA, María J., et al. *Effect of variance ratio on ANOVA robustness: Might 1.5 be the limit?*. In *Behavior research methods*, 50.3, 2018, s. 937-962. ISSN 1554-351X.
2. BOX, G. E. P. *Non-normality and tests on variances*. In *Biometrika*, 40, 1953, s. 318–335. ISSN 0006-3444.
3. BRADLEY, J. V. *Robustness*. In *British Journal of Mathematical and Statistical Psychology*, 31, 1978, s. 144–152. ISSN 2044-8317.
4. BROWN, M. B., & FORSYTHE, A. B. *The ANOVA and multiple comparisons for data with heterogeneous variances*, In *Biometrics*. 1974, s. 719–724. ISSN 0006-3444.
5. BURTON, Andrea, et al. *The design of simulation studies in medical statistics*. In *Statistics in medicine*, 2006, s. 4279-4292. ISSN 1097-0258.
6. FENG, C., WANG, H., LU, N., CHEN, T., HE, H., LU, Y., & TU, X. M. *Log-transformation and its implications for data analysis*. In *Shanghai archives of psychiatry*, 2014, s. 105–109. ISSN 10020829.
7. GLASS, G. V., PECKHAM, P. D., & SANDERS, J. R. *Consequences of failure to meet assumptions underlying the fixed effects analyses of variance and covariance*. In *Review of Educational Research*, 42, 1972, s. 237– 288. ISSN 00346543.

8. KESELMAN, H. J. et al. *Statistical practices of educational researchers: An analysis of their ANOVA, MANOVA, and ANCOVA analyses*. In *Review of Educational Research*, 68, 1998, s. 350–386. ISSN 00346543.
9. LIX, L. M., KESELMAN, J. C., & KESELMAN, H. J. *Consequences of assumption violations revisited: A quantitative review of alternatives to the one-way analysis of variance F test*. In *Review of Educational Research*, 66, 1996, s. 579–619. ISSN 00346543.
10. METROPOLIS, N., & ULAM, S. *The Monte Carlo Method*. In *Journal of the American Statistical Association*, 1949, s. 335-341. ISSN 0162-1459.
11. MICCERI, T. *The unicorn, the normal curve, and other improbable creatures*. In *Psychological Bulletin*, 105, 1996, s. 156–166. ISSN 0033-2909.
12. NGUYEN DINH, He. *Analýza vybraných sociálno-ekonomických ukazovateľov s ohľadom na odľahlé údaje. Ekonomika a informatika: vedecký časopis FHI EU v Bratislave a SSHI*. Bratislava: Fakulta hospodárskej informatiky: Slovenská spoločnosť pre hospodársku informatiku, 2011, s. 149-151. ISSN 1336-3514.
13. SCARIANO, Stephen M. & DAVENPORT, James M. *The Effects of Violations of Independence Assumptions in the One-Way ANOVA*, In *The American Statistician*, 1987, s. 123-129. ISSN 0003-1305.
14. VALLAT, R. *Pingouin: statistics in Python*. In *Journal of Open Source Software*, 1026 s. 2018. ISSN 2475-9066.
15. WELCH, Bernard Lewis. *On the comparison of several mean values: an alternative approach*. In *Biometrika*, 1951, s. 330-336. ISSN 0006-3444.
16. WILCOX, R. R. *ANOVA: The practical importance of heteroscedastic methods, using trimmed means versus means, and designing simulation studies*. In *British Journal of Mathematical and Statistical Psychology*, 48, 1995, s. 99–114. ISSN 2044-8317
17. WU, Shunquan et al. *Misuse of statistical methods in 10 leading Chinese medical journals in 1998 and 2008*. In *TheScientificWorldJournal* vol. 11. Hindawi Publishing Corp., 2011. ISSN 1537-744X.

Internetové zdroje

1. CHILDS, Dylan Z. *Data Analysis and Statistics with R*. [cit. 2020-18-02]. Dostupné na internete: <<https://dzchilds.github.io/stats-for-bio/data-transformations.html#the-logarithmic-transformation>>
2. PARELLADA, Antoni. *Why is ANOVA equivalent to linear regression*. [cit. 2020-09-03]. Dostupné na internete: <<https://stats.stackexchange.com/q/175290>>
3. WALMSLEY, Angela L. E., BROWN, Michael C. *What Is Power?*. [cit. 2020-03-03]. Dostupné na internete: <<https://www.statisticsteacher.org/2017/09/15/what-is-power/>>

Prílohy

1. Tabuľka odhadnutých mier I. druhu

| Počty pozorovaní | Pomer rozptylov | ANOVA miera chýb I. druhu | Welch miera chýb I. druhu |
|------------------|-----------------|---------------------------|---------------------------|
| 30:30:30 | 1:1 | 0.0482 | 0.0494 |
| 30:30:30 | 1:1.5 | 0.0542 | 0.0496 |
| 30:30:30 | 1:1.75 | 0.0592 | 0.052 |
| 30:30:30 | 1:2 | 0.0592 | 0.0456 |
| 30:30:30 | 1:2.5 | 0.064 | 0.0458 |
| 30:30:30 | 1:3 | 0.0624 | 0.0462 |
| 30:30:30 | 1:4 | 0.0652 | 0.0492 |
| 30:30:30 | 1:5 | 0.0738 | 0.0488 |
| 30:30:30 | 1:6 | 0.0714 | 0.0502 |
| 30:30:30 | 1:7 | 0.067 | 0.046 |
| 30:30:30 | 1:8 | 0.0634 | 0.051 |
| 30:30:30 | 1:9 | 0.0728 | 0.054 |
| 30:30:30 | 1:12 | 0.0778 | 0.0522 |
| 30:30:30 | 1:15 | 0.0808 | 0.0454 |
| 30:30:30 | 1:17.5 | 0.0848 | 0.0494 |
| 30:30:30 | 1:20 | 0.0812 | 0.0536 |
| 30:30:30 | 1:25 | 0.087 | 0.051 |
| 30:30:30 | 1:30 | 0.0956 | 0.0518 |
| 75:60:50 | 1:1 | 0.047 | 0.044 |
| 75:60:50 | 1:1.5 | 0.059 | 0.049 |
| 75:60:50 | 1:1.75 | 0.082 | 0.046 |
| 75:60:50 | 1:2 | 0.081 | 0.05 |
| 75:60:50 | 1:2.5 | 0.1 | 0.049 |
| 75:60:50 | 1:3 | 0.101 | 0.039 |
| 75:60:50 | 1:4 | 0.102 | 0.052 |
| 75:60:50 | 1:5 | 0.103 | 0.05 |
| 75:60:50 | 1:6 | 0.099 | 0.053 |
| 75:60:50 | 1:7 | 0.131 | 0.065 |
| 75:60:50 | 1:8 | 0.103 | 0.045 |
| 75:60:50 | 1:9 | 0.103 | 0.036 |
| 75:60:50 | 1:12 | 0.122 | 0.059 |
| 75:60:50 | 1:15 | 0.137 | 0.049 |
| 75:60:50 | 1:17.5 | 0.14 | 0.059 |
| 75:60:50 | 1:20 | 0.121 | 0.044 |
| 75:60:50 | 1:25 | 0.12 | 0.04 |
| 75:60:50 | 1:30 | 0.136 | 0.049 |
| 50:60:75 | 1:1 | 0.062 | 0.066 |
| 50:60:75 | 1:1.5 | 0.041 | 0.043 |
| 50:60:75 | 1:1.75 | 0.034 | 0.044 |

| | | | |
|----------|--------|-------|-------|
| 50:60:75 | 1:2 | 0.048 | 0.057 |
| 50:60:75 | 1:2.5 | 0.037 | 0.047 |
| 50:60:75 | 1:3 | 0.046 | 0.055 |
| 50:60:75 | 1:4 | 0.042 | 0.041 |
| 50:60:75 | 1:5 | 0.042 | 0.043 |
| 50:60:75 | 1:6 | 0.046 | 0.048 |
| 50:60:75 | 1:7 | 0.034 | 0.047 |
| 50:60:75 | 1:8 | 0.04 | 0.051 |
| 50:60:75 | 1:9 | 0.047 | 0.046 |
| 50:60:75 | 1:12 | 0.037 | 0.055 |
| 50:60:75 | 1:15 | 0.045 | 0.051 |
| 50:60:75 | 1:17.5 | 0.034 | 0.034 |
| 50:60:75 | 1:20 | 0.042 | 0.047 |
| 50:60:75 | 1:25 | 0.038 | 0.045 |
| 50:60:75 | 1:30 | 0.04 | 0.042 |

Zdroj: Vlastné spracovanie

2. Tabuľka odhadnutej sily testov

| Počty pozorovaní | Pomer rozptylov | ANOVA sila testu | Welch sila testu | Stredné hodnoty |
|------------------|-----------------|------------------|------------------|-----------------|
| 30:30:30 | 1:1 | 0.978666667 | 0.972 | 1:0:0 |
| 30:30:30 | 1:1.5 | 0.946666667 | 0.981333333 | 1:0:0 |
| 30:30:30 | 1:1.75 | 0.898666667 | 0.954666667 | 1:0:0 |
| 30:30:30 | 1:2 | 0.850666667 | 0.945333333 | 1:0:0 |
| 30:30:30 | 1:2.5 | 0.721333333 | 0.953333333 | 1:0:0 |
| 30:30:30 | 1:3 | 0.505 | 0.817 | 1:0:0 |
| 30:30:30 | 1:4 | 0.297 | 0.653 | 1:0:0 |
| 30:30:30 | 1:5 | 0.204 | 0.607 | 1:0:0 |
| 30:30:30 | 1:6 | 0.162 | 0.362 | 1:0:0 |
| 30:30:30 | 1:7 | 0.138 | 0.364 | 1:0:0 |
| 30:30:30 | 1:8 | 0.086 | 0.159 | 1:0:0 |
| 30:30:30 | 1:9 | 0.11 | 0.171 | 1:0:0 |
| 75:60:50 | 1:1 | 1 | 1 | 1:0:0 |
| 75:60:50 | 1:1.5 | 0.999 | 1 | 1:0:0 |
| 75:60:50 | 1:1.75 | 1 | 1 | 1:0:0 |
| 75:60:50 | 1:2 | 0.998666667 | 0.998666667 | 1:0:0 |
| 75:60:50 | 1:2.5 | 0.992 | 1 | 1:0:0 |
| 75:60:50 | 1:3 | 0.942666667 | 0.994666667 | 1:0:0 |
| 75:60:50 | 1:4 | 0.742666667 | 0.928 | 1:0:0 |
| 75:60:50 | 1:5 | 0.568 | 0.92 | 1:0:0 |
| 75:60:50 | 1:6 | 0.409333333 | 0.661333333 | 1:0:0 |
| 75:60:50 | 1:7 | 0.317333333 | 0.624 | 1:0:0 |
| 75:60:50 | 1:8 | 0.250666667 | 0.336 | 1:0:0 |
| 75:60:50 | 1:9 | 0.229333333 | 0.336 | 1:0:0 |
| 50:60:75 | 1:1 | 1 | 1 | 1:0:0 |
| 50:60:75 | 1:1.5 | 0.998666667 | 1 | 1:0:0 |
| 50:60:75 | 1:1.75 | 0.994666667 | 1 | 1:0:0 |

| | | | | |
|----------|--------|-------------|-------------|-------|
| 50:60:75 | 1:2 | 0.984 | 1 | 1:0:0 |
| 50:60:75 | 1:2.5 | 0.942666667 | 0.998666667 | 1:0:0 |
| 50:60:75 | 1:3 | 0.762666667 | 0.989333333 | 1:0:0 |
| 50:60:75 | 1:4 | 0.429333333 | 0.942666667 | 1:0:0 |
| 50:60:75 | 1:5 | 0.234666667 | 0.897333333 | 1:0:0 |
| 50:60:75 | 1:6 | 0.178666667 | 0.696 | 1:0:0 |
| 50:60:75 | 1:7 | 0.118666667 | 0.666666667 | 1:0:0 |
| 50:60:75 | 1:8 | 0.070666667 | 0.338666667 | 1:0:0 |
| 50:60:75 | 1:9 | 0.064 | 0.302666667 | 1:0:0 |
| 30:30:30 | 1:1 | 0.988 | 0.985333333 | 0:0:1 |
| 30:30:30 | 1:1.5 | 0.873333333 | 0.814666667 | 0:0:1 |
| 30:30:30 | 1:1.75 | 0.810666667 | 0.710666667 | 0:0:1 |
| 30:30:30 | 1:2 | 0.733333333 | 0.617333333 | 0:0:1 |
| 30:30:30 | 1:2.5 | 0.582666667 | 0.429333333 | 0:0:1 |
| 30:30:30 | 1:3 | 0.48 | 0.332 | 0:0:1 |
| 30:30:30 | 1:4 | 0.273333333 | 0.188 | 0:0:1 |
| 30:30:30 | 1:5 | 0.228 | 0.138666667 | 0:0:1 |
| 30:30:30 | 1:6 | 0.14 | 0.102666667 | 0:0:1 |
| 30:30:30 | 1:7 | 0.153333333 | 0.101333333 | 0:0:1 |
| 30:30:30 | 1:8 | 0.109333333 | 0.076 | 0:0:1 |
| 30:30:30 | 1:9 | 0.1 | 0.093333333 | 0:0:1 |
| 75:60:50 | 1:1 | 1 | 1 | 0:0:1 |
| 75:60:50 | 1:1.5 | 0.988 | 0.972 | 0:0:1 |
| 75:60:50 | 1:1.75 | 0.973333333 | 0.906666667 | 0:0:1 |
| 75:60:50 | 1:2 | 0.944 | 0.846666667 | 0:0:1 |
| 75:60:50 | 1:2.5 | 0.838666667 | 0.657333333 | 0:0:1 |
| 75:60:50 | 1:3 | 0.725333333 | 0.504 | 0:0:1 |
| 75:60:50 | 1:4 | 0.52 | 0.332 | 0:0:1 |
| 75:60:50 | 1:5 | 0.409333333 | 0.225333333 | 0:0:1 |
| 75:60:50 | 1:6 | 0.32 | 0.161333333 | 0:0:1 |
| 75:60:50 | 1:7 | 0.285333333 | 0.145333333 | 0:0:1 |
| 75:60:50 | 1:8 | 0.22 | 0.12 | 0:0:1 |
| 75:60:50 | 1:9 | 0.210666667 | 0.092 | 0:0:1 |
| 50:60:75 | 1:1 | 1 | 1 | 0:0:1 |
| 50:60:75 | 1:1.5 | 0.997333333 | 0.996 | 0:0:1 |
| 50:60:75 | 1:1.75 | 0.986666667 | 0.981333333 | 0:0:1 |
| 50:60:75 | 1:2 | 0.974666667 | 0.958666667 | 0:0:1 |
| 50:60:75 | 1:2.5 | 0.888 | 0.84 | 0:0:1 |
| 50:60:75 | 1:3 | 0.708 | 0.652 | 0:0:1 |
| 50:60:75 | 1:4 | 0.457333333 | 0.421333333 | 0:0:1 |
| 50:60:75 | 1:5 | 0.338666667 | 0.309333333 | 0:0:1 |
| 50:60:75 | 1:6 | 0.238666667 | 0.22 | 0:0:1 |
| 50:60:75 | 1:7 | 0.194666667 | 0.178666667 | 0:0:1 |
| 50:60:75 | 1:8 | 0.141333333 | 0.14 | 0:0:1 |
| 50:60:75 | 1:9 | 0.126666667 | 0.124 | 0:0:1 |

Zdroj: Vlastné spracovanie

3. Zdrojový kód

```
# Importovanie modulov
import pandas as pd
import numpy as np
import pingouin as pg
import itertools

# Funkcie na výpočet testov
def anovaSimulation(mi_list, sigma2_list, size_list, group_list):

    group_list = [g1, g2, g3]

    concat_list = []
    for iteration, group in enumerate(group_list):
        group_level_list = []
        for elem in group:
            group_lst = [iteration, elem]
            group_level_list.append(group_lst)
            concat_list.append(group_level_list)

    list_with_factors = list(itertools.chain(concat_list[0],
        concat_list[1], concat_list[2]))

    factor_list = []
    val_list = []
    for obs in list_with_factors:

        factor_list.append(obs[0])
        val_list.append(obs[1])

    df = pd.DataFrame([factor_list, val_list]).T
    df.columns = ['factor', 'value']
    df['factor'] = df['factor'].astype(int)

    aov = pg.anova(data=df, dv='value', between='factor')

    return aov['p-unc'][0]

def welchAnovaSimulation(mi_list, sigma2_list, size_list, group_list):

    concat_list = []
    for iteration, group in enumerate(group_list):
        group_level_list = []
        for elem in group:
            group_lst = [iteration, elem]
            group_level_list.append(group_lst)
            concat_list.append(group_level_list)

    list_with_factors = list(itertools.chain(concat_list[0],
        concat_list[1], concat_list[2]))

    factor_list = []
    val_list = []
    for obs in list_with_factors:
```

```

        factor_list.append(obs[0])
        val_list.append(obs[1])

    df = pd.DataFrame([factor_list, val_list]).T
    df.columns = ['factor', 'value']
    df['factor'] = df['factor'].astype(int)

    aov = pg.welch_anova(data=df, dv='value', between='factor')

    return aov['p-unc'][0]

# Vytvorenie prázdných výsledných df
type1_results_df = pd.DataFrame(columns=[ 'Means',
                                           'Sizes',
                                           'Variances',
                                           'ANOVA Type I error rate',
                                           'Welch Type I error rate'])

power_results_df = pd.DataFrame(columns=[ 'Sizes',
                                           'Variances',
                                           'ANOVA power of test',
                                           'Welch power of test'
                                           ])

# Simulacie
n_of_trials = 5000

mi_list = [0, 0, 1]
sigma2_list = [1, 5, 9]
size_list = [50, 60, 75]

anova_p_v_results_list = []
welch_p_v_results_list = []
for iteration in range(n_of_trials):

    g1 = list(np.random.normal(loc=mi_list[0], scale=sigma2_list[0],
                               size=size_list[0]))
    g2 = list(np.random.normal(loc=mi_list[1], scale=sigma2_list[1],
                               size=size_list[1]))
    g3 = list(np.random.normal(loc=mi_list[2], scale=sigma2_list[2],
                               size=size_list[2]))

    group_list = [g1, g2, g3]

    anova_p_v_results_list.append(anovaSimulation(mi_list, sigma2_list,
                                                  size_list, group_list))
    welch_p_v_results_list.append(welchAnovaSimulation(mi_list, sigma2_list,
                                                       size_list, group_list))

# Uloženie výsledkov
pv_df = pd.DataFrame()
pv_df['ANOVA_p_values'] = anova_p_v_results_list
pv_df['Welch_p_values'] = welch_p_v_results_list

# Vypocet mier chyb I. druhu
anova_type1_error_rate =
len(pv_df[pv_df['ANOVA_p_values'] < 0.05][['ANOVA_p_values']]) / n_of_trials
welch_type1_error_rate =

```

```

len(pv_df[pv_df['Welch_p_values']<0.05][['Welch_p_values']])/ n_of_trials

print(anova_type1_error_rate)
print(welch_type1_error_rate)

part_type1_error_df = pd.DataFrame(columns=['Sizes',
                                           'Variances',
                                           'ANOVA Type I error rate',
                                           'Welch Type I error rate'
                                           ])

sizes = "{0}/{1}/{2}".format(*size_list)
variances = "{0}/{1}/{2}".format(*sigma2_list)

part_type1_error_df = part_type1_error_df.append({'Sizes': sizes,
                                                  'Variances': variances,
                                                  'ANOVA Type I error rate':
                                                  anova_type1_error_rate,
                                                  'Welch Type I error rate':
                                                  welch_type1_error_rate},
                                                  ignore_index=True)

type1_results_df = type1_results_df.append(part_type1_error_df)
type1_results_df.reset_index(drop=True, inplace=True)
type1_results_df.to_csv('new_type1_results_anova_welch.csv', index=False)

# Vypocet mier chyby II. druhu, sily testu
anova_type2_error_rate =
len(pv_df[pv_df['ANOVA_p_values']>=0.05][['ANOVA_p_values']])/ n_of_trials
welch_type2_error_rate =
len(pv_df[pv_df['Welch_p_values']>=0.05][['Welch_p_values']])/ n_of_trials

anova_power = 1 - anova_type2_error_rate
welch_power = 1 - welch_type2_error_rate

part_power_df = pd.DataFrame(columns=['Sizes',
                                      'Variances',
                                      'ANOVA power of test',
                                      'Welch power of test'
                                      ])

sizes = "{0}/{1}/{2}".format(*size_list)
variances = "{0}/{1}/{2}".format(*sigma2_list)
means = "{0}/{1}/{2}".format(*mi_list)

part_power_df = part_power_df.append({'Means': means,
                                       'Sizes': sizes,
                                       'Variances': variances,
                                       'ANOVA power of test': anova_power,
                                       'Welch power of test': welch_power},
                                       ignore_index=True)

power_results_df = power_results_df.append(part_power_df)
power_results_df.reset_index(drop=True, inplace=True)
power_results_df.to_csv('new_power_results_anova_welch.csv', index=False)

```