SLOVENSKÁ ŠTATISTIKA a DEMOGRAFIA

SLOVAK STATISTICS and DEMOGRAPHY

2/2021
ročník/volume 31

Recenzovaný vedecký časopis so zameraním na prezentáciu moderných štatistických a demografických metód a postupov.

Scientific peer-reviewed journal focusing on the presentation of modern statistical and demographic methods and procedures.

Článok/Article: 1

Typ článku/Type of article: vedecký článok/scientific article

Strany/Pages: 3 – 17

Dátum vydania/Publication date: 15. apríl 2021/April 15, 2021

Eva KOTLEBOVÁ,
Katedra štatistiky, Fakulta hospodárskej informatiky, Ekonomická univerzita v Bratislave
Martin SLIACKY
Swis Re, Bratislava

MOŽNOSTI POROVNÁVANIA ÚROVNE HODNÔT ZNAKU VO VIACERÝCH SÚBOROCH PRI NESPLNENÍ PREDPOKLADOV ANALÝZY ROZPTYLU

POSSIBILITIES OF COMPARING MORE THAN TWO GROUP MEANS IN A RESPONSE VARIABLE IN THE CASE OF FAILURE OF THE ASSUMPTIONS OF ANALYSIS OF VARIANCE

ABSTRAKT
Pri použití analýzy rozptylu sa v praxi javí ako najproblematickejší predpoklad homoskedasticity (rovnosti rozptylov v porovnávaných súboroch). V takom prípade (ak sú splnené ostatné predpoklady) je možnosť využiť Welchovu analýzu rozptylu, pri ktorej sa vo výpočte testovacej štatistiky rozsahy jednotlivých výberov nahradia váhami, v ktorých sú zahrnuté aj výberové rozptyly. V príspevku sa porovnávajú výhody a nevýhody klasickej a Welchovej analýzy rozptylu prostredníctvom veľkostí chýb 1. a 2. druhu pri rôznych vstupných hodnotách parametrov súborov. Výsledkom sú odporúčania pre voľbu vhodnej metódy pre rôzne kombinácie veľkostí porovnávaných súborov a ich variabilít.

ABSTRACT
When using analysis of variance in practice, the assumption of homoscedasticity (equality of variance in the compared files) appears to be the most problematic. In such case (if the other assumptions are met), it is possible to use Welch analysis of variance where the ranges of individual samples are replaced by weights, in the calculation of test statistics, including sample variances. The paper compares the advantages and disadvantages of classical and Welch analysis of variance by means of error sizes of I and II types at different input values of file parameters, the outcome of which are the recommendations for choosing a suitable method for different combinations of sizes of the compared files and their variabilities.

KĽÚČOVÉ SLOVÁ
analýza rozptylu, Welchova analýza rozptylu, hladina významnosti, sila testu, simulácie Monte Carlo

KEY WORDS
analysis of variance, Welch analysis of variance, significance level, power of a test, Monte Carlo simulations

1. ÚVOD
Analýza rozptylu je jednou z najviac používaných štatistických metód. V praxi je totiž často potrebné porovnať hodnoty číselného znaku vo viacerých súboroch, aby sme (ak sa preukážu štatisticky významné rozdiely) mohli bližšie identifikovať tie dvojice skupín, medzi ktorými existujú štatisticky významné rozdiely, a prípadne sa pokúsiť hľadať mechanizmus vplyvu faktorov, ktoré ich spôsobujú.

V príspevku sa zaoberáme najjednoduchším prípadom – jednoduchou analýzou rozptylu, v ktorej sa posudzuje vplyv jedného faktora (slovnej premennej *A*) na hodnoty spojitej číselnej závislej premennej *Y*. Faktor (kategoriálna premenná) má niekoľko obmien, podľa ktorých sú vytvorené porovnávané skupiny. Využitie metódy analýzy rozptylu si vyžaduje splnenie troch predpokladov: (1) výbery musia byť nezávislé (táto podmienka obyčajne býva v praxi splnená, vyplýva zo samotnej formulácie riešeného problému), (2) v každej skupine by mala mať závislá premenná normálne rozdelenie (táto podmienka sa na základe centrálnej limitnej vety môže čiastočne obísť, ak sú rozsahy súborov dostatočne veľké, respektíve možno využiť neparametrický Kruskallov-Wallisov test) a napokon (3) je tu najviac limitujúca podmienka rovnosti rozptylov v porovnávaných súboroch (homoskedasticita). Ak nie je splnená, nemôžeme sa totiž na výsledok analýzy rozptylu dostatočne spoľahnúť – najmä p-hodnoty blízke hranici 0,05 môžu vzbudiť isté pochybnosti.

V nasledujúcich častiach spresníme, akým spôsobom môžu rozdielne hodnoty rozptylov v porovnávaných súboroch (tzv. heteroskedasticita) ovplyvniť chyby 1. a 2. druhu pri analýze rozptylu a ako je to pri Welchovej analýze rozptylu. Využijeme pritom výsledky simulácií metódou Monte Carlo, ktoré sa realizovali pre rôzne hodnoty parametrov vstupných výberov (rozsahy, výberové rozptyly, stredné hodnoty).

2. ANALÝZA ROZPTYLU

Hodnoty závislej premennej *Y*, ktoré sú predmetom analýzy, označujeme ako y_{ij}, pričom $i = 1, 2, \dots, k$ (k je počet skupín – úrovní faktora), $j = 1, 2, \dots, n_i$ (j označuje poradie hodnoty y_{ij} v i –tej skupine, pričom počet údajov v i –tej skupine je n_i). Celkový počet údajov je $n = \sum_{i=1}^{k} n_i$.

Podstatou metódy je testovanie hypotézy, ktorá tvrdí, že faktor *A*, podľa obmien (úrovní) ktorého boli údaje rozdelené do skupín, nemá na hodnoty premennej Y vplyv, takže priemerná úroveň hodnôt znaku je vo všetkých skupinách rovnaká. Nulová hypotéza má teda tvar: H_0: μ_1=$\mu_2 = \dots = \mu_k$. Alternatívna hypotéza H_1 tvrdí, že aspoň jedna z rovností neplatí, t. j. že aspoň medzi dvomi skupinami existujú štatisticky významné rozdiely v stredných hodnotách.

Princípom testovania je porovnávanie medziskupinovej variability (ktorej zdrojom je faktor) a vnútroskupinovej variability, ktorá je determinovaná inými faktormi, pričom sem zaraďujeme aj vplyv náhody. Testovacia štatistika má tvar:

$$F = \frac{MSA}{MSE} = \frac{\frac{SSA}{k-1}}{\frac{SSE}{n-k}} = \frac{\frac{\sum_{i=1}^{k}(\bar{y}_i - \bar{\bar{y}})^2 n_i}{k-1}}{\frac{\sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_i)^2}{n-k}} \sim F_{(k-1;n-k)} \tag{1}$$

($\bar{y}_i$ označuje priemer v i – tej skupine, $\bar{\bar{y}}$ je celkový priemer hodnôt). Z jej tvaru je zrejmé, že má Fisherovo rozdelenie s počtami stupňov voľnosti $k - 1$ a $n - k$. Čím je medziskupinová variabilita (MSA) v porovnaní s vnútroskupinovou (MSE) väčšia, tým väčší je vplyv uvažovaného faktora na hodnoty závislej premennej. Znamená to, že čím je hodnota testovacej štatistiky väčšia, tým pravdepodobnejšie je zamietnutie nulovej hypotézy. Presnejšie: ak hodnota testovacej štatistiky presiahne 95. percentil jej rozdelenia, na hladine významnosti 0,05 konštatujeme štatisticky významný vplyv faktora *A* na závislú premennú *Y*. Ak takáto situácia nastane, je potrebné zistiť, medzi ktorými dvojicami skupín existujú štatisticky významné rozdiely. Na to slúžia testy mnohonásobného porovnávania (Sheffého, Tukeyho, Bonferroniho, Fisherov LSD test, Duncanov, atď. [9]).

V súvislosti s testami mnohonásobného porovnávania, ale aj so samotnou analýzou rozptylu, je aktuálna otázka, prečo ich nemožno nahradiť párovými *t*-testami stredných hodnôt (testovali by sa nulové hypotézy $H_0: \mu_r = \mu_s; r = 1, 2, \ldots, k; s = 1, 2, \ldots, k; r \neq s$). Argumentom na odmietnutie tejto možnosti nie je samotný počet týchto testov (pre k skupín ich je spolu $\binom{k}{2}$), ale ich realizácia by podstatne zvýšila pravdepodobnosť chyby 1. druhu, t. j. pravdepodobnosť zamietnutia nulovej hypotézy, aj keď v skutočnosti platí. Voláme ju hladina významnosti a označujeme ako α.

Ak by sme pomocou analýzy rozptylu porovnávali len tri skupiny, tak pri každom z troch ($\binom{3}{2}$) párových t-testov je pravdepodobnosť zamietnutia platnej nulovej hypotézy α, teda pravdepodobnosť, že ju príjmeme, je $1 - \alpha$. Pravdepodobnosť, že sa v troch nezávislých t-testoch nedopustíme chyby 1. druhu, je $(1 - \alpha)^3$ (pre hladinu významnosti 0,05 je to 0,8574, takže súčasné zamietnutie troch platných nulových hypotéz má pravdepodobnosť až 14,26%, pre $k > 3$ je táto pravdepodobnosť ešte vyššia). Pri realizácii viacerých testov sa zvyčajne uvažuje tzv. skupinová miera chýb 1. druhu (familywise error – FWER), ktorá je definovaná ako pravdepodobnosť toho, že sa pri sérii testov dopustíme aspoň jednej chyby 1. druhu [14]. Toto však nie je predmetom nášho príspevku.

Ako sme už uviedli v úvode, z predpokladov využitia metódy analýzy rozptylu sa ukazuje ako najproblematickejší predpoklad homoskedasticity – rovnosti rozptylov premennej Y v porovnávaných súboroch. Testovanie hypotézy na overenie tohto predpokladu by malo predchádzať samotnej analýze rozptylu. K dispozícii sú známe testy:

Bartlettov test – testovacia štatistika má tvar [8]:

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

v ktorom $\tilde{s}_i^2$ označuje výberový rozptyl premennej *Y* v i-tom súbore,

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

a $\tilde{s}^2$ označuje združený rozptyl

$$\tilde{s}^2 = \frac{\sum_{i=1}^{k}\tilde{s}_i^2(n_i-1)}{n-k} \tag{4}$$

Pre $n \to \infty$ má rozdelenie chí-kvadrát s počtom stupňov voľnosti $k - 1$.

Leveneov test – testovacia štatistika má tvar [4]:

$$W = \frac{\frac{\sum_{i=1}^{k} n_i \cdot (\bar{Z}_i - \bar{\bar{Z}})^2}{k-1}}{\frac{\sum_{i=1}^{k} \sum_{j=1}^{n_i} (Z_{ij} - \bar{Z}_i)^2}{n-k}}, \tag{5}$$

ktorý je rovnaký ako má testovacia štatistika samotnej analýzy rozptylu (aj s rovnakým rozdelením), ale vstupnými hodnotami sú údaje premennej Z, ktorá je transformáciou analyzovanej premennej Y: $Z_{ij} = |Y_{ij} - \bar{Y}_i|$, pričom podľa okolností sa skupinové priemery $\bar{Y}_i$ môžu nahradiť aj skupinovými mediánmi, prípadne 10%-nými zostrihnutými priemermi [2].

O´Brienov test [11] je podobný ako Leveneov test; takisto je analýzou rozptylu transformovanej premennej Z, len transformácia má inú podobu. Hodnoty premennej sa určia pomocou vzťahu:

$$z_{ij} = \frac{(n_{ij} - 1{,}5) \cdot n_i \cdot (y_{ij} - \bar{y}_i)^2 - 0{,}5 s_i^2 \cdot (n_i - 1)}{(n_i - 1) \cdot (n_i - 2)} \tag{6}$$

kde s_i^2 je výberový rozptyl premennej *Y* v i-tej skupine

Cochranov test – testovacia štatistika má tvar [8]:

$$C = \frac{\max\limits_{i=1,2,\dots,k} (\tilde{s}_i^2)}{\sum_{i=1}^{k} \tilde{s}_i^2} \tag{7}$$

Hartleyho test – testovacia štatistika má tvar [8]:

$$H = \frac{\max(\tilde{s}_i^2)}{\min(\tilde{s}_i^2)} \tag{8}$$

Posledné dva testy majú svoje špeciálne rozdelenie, ktoré závisí od počtu porovnávaných skupín. Pre všetky testy homoskedasticity platí, že pre heteroskedastické údaje má testovacia štatistika veľkú hodnotu, takže kritická oblasť leží na pravom chvoste rozdelenia.

Ak podmienka homoskedasticity nie je splnená, možno hodnoty premennej Y transformovať pomocou vhodnej funkcie tak, aby transformované hodnoty vyhovovali tejto požiadavke. Voľba transformačnej funkcie závisí od tvaru rozdelenia hodnôt premennej Y [7]. Aj keď sa transformáciou môže dosiahnuť splnenie podmienky homoskedasticity, treba pripomenúť, že platí pre transformované údaje, takže interpretácia výsledkov analýzy rozptylu môže byť problematická.

3. WELCHOVA ANALÝZA ROZPTYLU

Táto metóda [5, 13] sa môže použiť vtedy, keď sa pri klasickej analýze rozptylu nepotvrdí predpoklad homoskedasticity. V testovacej štatistike tohto testu sa namiesto rozsahov výberov n_i používajú ako váhy veličiny $w_i = \frac{n_i}{\tilde{s}_i^2}$, takže sa do určitej miery eliminujú súbory s veľkou variabilitou. Čitateľ testovacej štatistiky je podobne ako pri klasickej analýze rozptylu priemerným štvorcom medziskupinovej variability (MSA), pričom váhy w_i sa použijú nielen na vyjadrenie súčtu štvorcov odchýlok, ale pomocou nich sa vypočíta aj celkový priemer:

$$\bar{y}_{WELCH} = \frac{\sum_{i=1}^{k} \bar{y}_i \cdot w_i}{\sum_{i=1}^{k} w_i} \tag{9}$$

$$SSA_{WELCH} = \sum_{i=1}^{k} (\bar{y}_i - \bar{\bar{y}})^2 \cdot w_i \tag{10}$$

$$MSA = \frac{SSA}{k-1} \tag{11}$$

Testovacia štatistika má tvar:

$$F = \frac{MSA}{1 + \frac{2\Lambda(k-2)}{3}} \tag{12}$$

v ktorom

$$\Lambda = \frac{3\sum_{i=1}^{k} \frac{\left(1 - \frac{w_i}{\sum_{i=1}^{k} w_i}\right)^2}{n_i - 1}}{k^2 - 1} \tag{13}$$

Tak ako pri klasickej analýze rozptylu má Fisherovo rozdelenie s počtami stupňov voľnosti $k-1$ a $\frac{1}{\Lambda}$. Kritickou oblasťou je pravý chvost rozdelenia, takže nulovú hypotézu na hladine významnosti α zamietneme, ak hodnota testovacej štatistiky presiahne $(1-\alpha)$ –ty percentil rozdelenia.

Nevýhodou tohto testu v porovnaní s klasickou analýzou rozptylu je jeho menšia sila (pravdepodobnosť zamietnutia nulovej hypotézy, ak v skutočnosti neplatí), preto ho používame len ako náhradnú možnosť pre heteroskedastické údaje. Najmä ak sú rozsahy súborov malé, výsledok testu treba brať s rezervou.

V nasledujúcich častiach porovnáme klasickú a Welchovu analýzu rozptylu z hľadiska pravdepodobností chýb 1. a 2. druhu. Využijeme pritom výsledky simulácií metódou Monte Carlo.

4. SIMULÁCIE MONTE CARLO A ICH VYUŽITIE NA ODHADY CHÝB PRI TESTOVANÍ

Simulácie číselných náhodných veličín s konkrétnymi hodnotami parametrov sú vhodným nástrojom na riešenie širokého spektra teoretických (to je náš prípad), ale aj praktických problémov, keď je potrebné odhadnúť hodnoty kľúčových charakteristík vývoja, hospodárskych výsledkov, investičných stratégií, hroziacich rizík atď. Prakticky všetky softvéry využívané v štatistike umožňujú simulácie hodnôt veličín so známym rozdelením s konkrétnymi hodnotami parametrov.

V induktívnej štatistike majú simulácie kľúčový význam. Umožňujú nám napríklad odhadnúť rozdelenie testovacej štatistiky (ak ho nevieme analyticky vyjadriť), porovnávať výsledky rôznych metód pri rôznych vstupných podmienkach, porovnávať kvalitu odhadov, ale aj chýb pri testovaní, čo sme využili v našom príspevku.

Pri **odhade miery chýb 1. druhu** (pravdepodobnosť zamietnutia platnej nulovej hypotézy) vo všeobecnosti postupujeme takto [6]:

1. Vytvorí sa hypotetická populácia (teoretické rozdelenie) s určitými vlastnosťami tak, aby zodpovedalo nulovej hypotéze.
2. Metódou Monte Carlo [3] sa z rozdelenia vygeneruje náhodný výber dopredu stanoveného rozsahu.
3. Pomocou vygenerovaných údajov realizujeme test.
4. Zistíme, či nastala chyba 1. druhu, teda či bola nulová hypotéza zamietnutá (údaje boli generované z rozdelenia spĺňajúceho podmienky nulovej hypotézy).
5. Výsledok zaznamenáme pomocou binárnej veličiny

$$I_i = \begin{cases} 1 & \text{ak nastala chyba 1. druhu} \\ 0 & \text{ak nenastala chyba 1. druhu} \end{cases}$$

6. Kroky 2. – 5. $M \times$ zopakujeme.
7. Pravdepodobnosť chyby 1. druhu odhadneme ako podiel testov, v ktorých nastala chyba, zo všetkých M testov:

$$\hat{\alpha} = \frac{1}{M} \sum_{i=1}^{M} I_i \qquad (14)$$

Je optimálne, ak je odhad blízky hodnote 0,05 (v štatistických testoch sa len výnimočne využíva iná hladina významnosti ako 0,05), resp. ak pochádza z intervalu 0,025 – 0,075. Príliš malé hodnoty (pod 0,025) znamenajú, že test je konzervatívny (je vyššia pravdepodobnosť chyby 2. druhu). Ak je odhad chyby väčší ako 0,075, test je príliš liberálny, t. j. má vyššiu pravdepodobnosť zamietnutia nulovej hypotézy, ako je štandardne nastavená hladina významnosti 0,05.

Mieru chýb druhého druhu (a z nej odvodenú **silu testu**) odhadneme na základe postupu [6]:

1. Vytvorí sa hypotetická populácia (teoretické rozdelenie) s určitými vlastnosťami tak, aby nulová hypotéza neplatila.
2. Metódou Monte Carlo sa z rozdelenia vygeneruje náhodný výber dopredu stanoveného rozsahu.
3. Pomocou vygenerovaných údajov realizujeme test.
4. Zistíme, či nastala chyba 2. druhu, teda či bola nulová hypotéza prijatá (údaje boli generované z rozdelenia, ktoré nespĺňalo podmienky nulovej hypotézy).
5. Výsledok zaznamenáme pomocou binárnej veličiny

$$I_i = \begin{cases} 1 & \text{ak nastala chyba 2. druhu} \\ 0 & \text{ak nenastala chyba 2. druhu} \end{cases}$$

6. Kroky 2. – 5. $M \times$ zopakujeme.
7. Pravdepodobnosť chyby 2. druhu odhadneme ako podiel testov, v ktorých nastala chyba, zo všetkých M testov:

$$\hat{\beta} = \frac{1}{M} \sum_{i=1}^{M} I_i \qquad (15)$$

8. Odhadom sily testu je potom $1 - \hat{\beta}$.

Za optimálny výsledok považujeme odhad sily testu blízky hodnote 1 (pravdepodobnosť chyby 2. druhu blízka 0).

Pri odhade miery chýb 1. druhu, ktoré môžu nastať pri realizácii klasickej, resp. Welchovej analýzy rozptylu, preto vytvárame skupiny populácií s rovnakou strednou hodnotou (tak, aby platila nulová hypotéza) a veľkosť chyby 2. druhu odhadujeme pomocou skupín populácií s odlišnými strednými hodnotami.

Okrem stredných hodnôt (rovnakých, vs. rôznych) sme v nami definovaných rozdeleniach použili aj rôzne rozsahy výberových súborov (vyvážené s rovnakými počtami, resp. nevyvážené) a variabilitu (homoskedastické, vs. nehomoskedastické populácie). Tým sme dostali odpovede na otázku, ktoré z týchto okolností a v akej miere ovplyvňujú pravdepodobnosti chýb.

5. ODHAD CHÝB 1. A 2. DRUHU PRI KLASICKEJ A WELCHOVEJ ANALÝZE ROZPTYLU

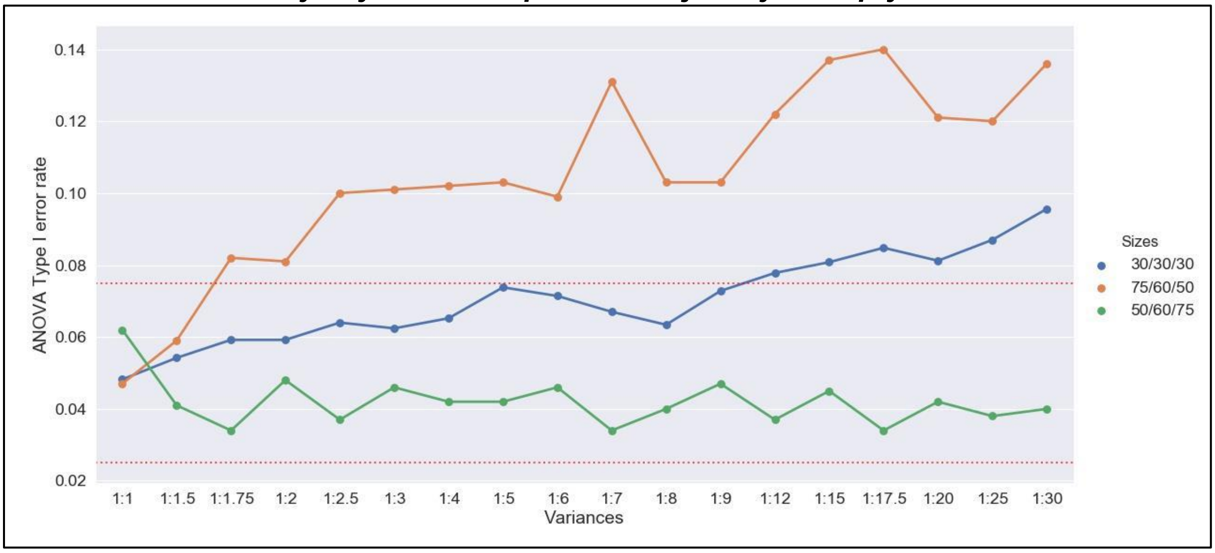
Pri odhadoch pravdepodobností chýb obidvoch druhov sme sledovali faktor s tromi úrovňami, takže sme generovali trojice súborov s rozsahmi n_1, n_2, n_3 s určitými vopred stanovenými vlastnosťami. Premenná mala vo všetkých súboroch normálne rozdelenie $N(\mu_i;\ \sigma_i^2); i = 1, 2, 3$.

Na odhad **pravdepodobnosti chýb 1. druhu** sme generovali trojice súborov s rovnakou strednou hodnotou (použili sme hodnotu 0), pričom sme pracovali s vyváženými ($n_1 = n_2 = n_3 = 30$) aj nevyváženými súbormi ($n_1 < n_2 < n_3$; $n_1 : n_2 : n_3 = 50 : 60 : 75$), homoskedastickými aj heteroskedastickými súbormi, v ktorých sa pomer najmenšieho a najväčšieho rozptylu pohyboval v rozmedzí od 1 : 1 až po 1 : 30 (spolu 18 rôznych variantov pomerov). Rozlišovali sme pritom dve možnosti: súbor s najväčšou variabilitou mal najväčší rozsah ($n_1 < n_2 < n_3,\ \sigma_1^2 < \sigma_2^2 < \sigma_3^2$) alebo to bolo naopak ($n_1 < n_2 < n_3, \sigma_1^2 > \sigma_2^2 > \sigma_3^2$). Celkovo sme tak analyzovali $18 \times 3 = 54$ skupín údajov, pričom pre každú z nich sme vytvorili 5 000 simulácií. Na základe postupu uvedeného v 3. časti sme odhadli pravdepodobnosť chyby 1. druhu pre klasickú analýzu rozptylu aj pre Welchovu analýzu rozptylu. Výsledky sme znázornili graficky.

Na grafe č. 1 sú pomocou spojnicového grafu znázornené odhady miery chýb 1. druhu klasickej analýzy rozptylu, a to pre vyvážené modely (modrá farba), pre modely, v ktorých má súbor s najvyššou variabilitou najväčší rozsah (zelená farba) a oranžovou farbou sú znázornené odhady v prípade, že súbor s najvyššou variabilitou má najmenší rozsah. Na osi x sú znázornené všetky uvažované pomery rozptylov ($\sigma_{min}^2/\sigma_{max}^2$). Prvý bod zľava pri všetkých lomených čiarach (zodpovedajúci pomeru 1 : 1) tak znázorňuje odhad chyby 1. druhu pre homoskedastické údaje. Optimálna situácia z hľadiska veľkosti chyby 1. druhu je vtedy, keď sa bod lomenej čiary nachádza medzi dvomi prerušovanými čiarami, ktoré sú na úrovni 0,025 a 0,075.

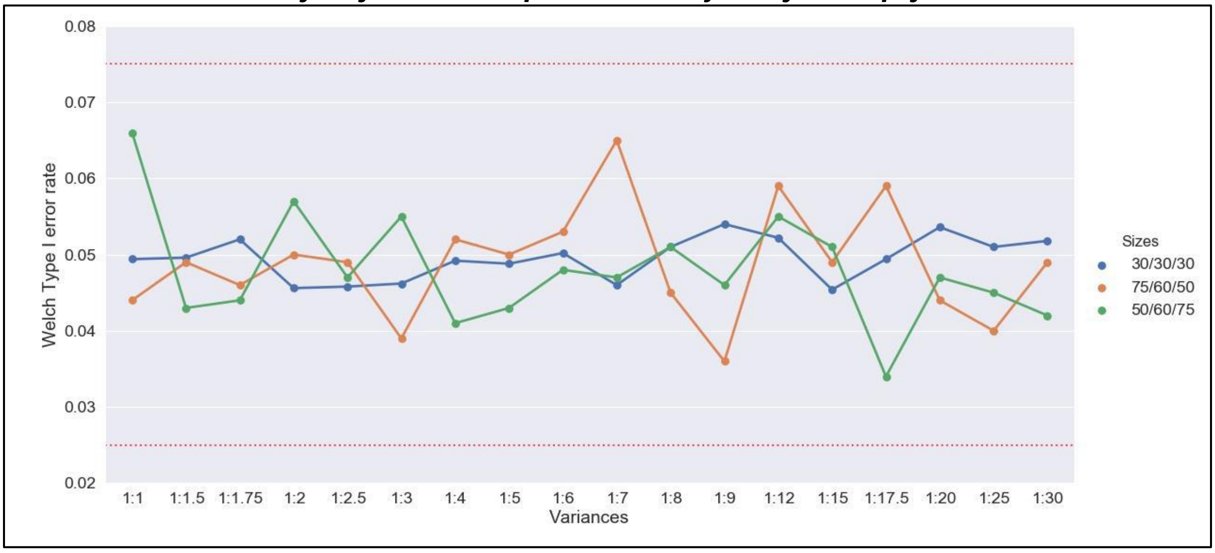
Z grafu č. 2 sú zrejmé nasledujúce skutočnosti: chyba 1. druhu sa nachádza v požadovanom intervale nielen pre údaje spĺňajúce podmienku homoskedasticity, ale aj vtedy, ak je pomer najmenšieho a najväčšieho rozptylu 1 : 1,5, a to pri vyvážených aj nevyvážených modeloch. Smerom k výraznejším rozdielom vo variabilite sa javí ako problematické, keď súbor s najväčšou variabilitou je najmenší (a naopak), ktorý už od pomeru 1 : 1,75 má hodnoty vyššie ako 0,075, čo znamená, že v takýchto prípadoch je vyššie riziko, že súbory budeme považovať z hľadiska úrovne hodnôt znaku za rôzne, aj keď sú v skutočnosti rovnaké. Pri vyvážených modeloch sa odhad chyby už pri pomere 1 : 5 blíži hranici 0,075, presiahne ju až od pomeru 1 : 12. Najpriaznivejší výsledok z hľadiska využiteľnosti klasickej analýzy rozptylu vyšiel vtedy keď najväčšiu variabilitu mal súbor s najväčším rozsahom. Všetky hodnoty vyšli v požadovanom intervale 0,025 až 0,075, aj keď sa pohybujú (s výnimkou pomeru 1 : 1) pod hranicou 0,05. Na grafe č. 2 sú znázornené odhady chýb 1. druhu pre rovnaké spektrum možností vstupných veličín pri Welchovej analýze rozptylu. Vidíme, že pri všetkých kombináciách možností početností súborov a variability sa body lomených čiar nachádzajú vo vnútri požadovaného intervalu (dokonca užšieho), takže žiadna z okolností neovplyvňuje (jedným alebo druhým smerom) pravdepodobnosť chyby 1. druhu.

Graf č. 1: Odhad miery chýb 1. druhu pri klasickej analýze rozptylu



Zdroj: vlastné spracovanie

Graf č. 2: Odhad miery chýb 1. druhu pri Welchovej analýze rozptylu

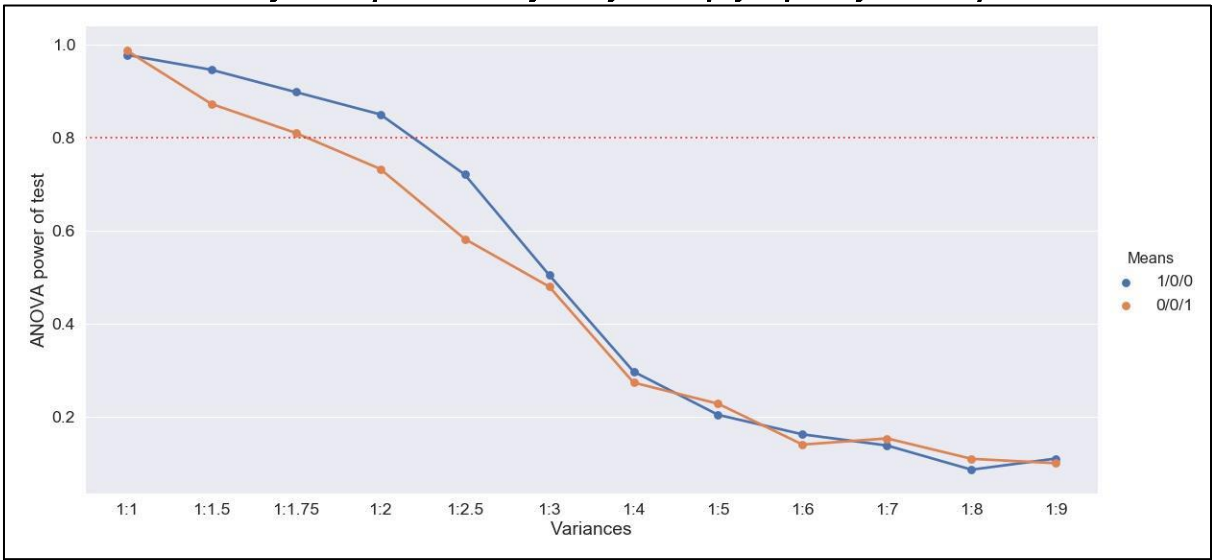


Zdroj: vlastné spracovanie

Na odhad **pravdepodobnosti chýb 2. druhu** sme generovali trojice súborov s nerovnakými strednými hodnotami – pre súbory s početnosťami n_1, n_2, n_3 sme použili trojice stredných hodnôt $(0, 0, 1)$ a $(1, 0, 0)$. Rozsahy súborov sme nastavili tak ako pri odhade pravdepodobnosti chyby 1. druhu: analyzovali sme vyvážené súbory s početnosťami $(30{:}\,30{:}\,30)$ aj nevyvážené s početnosťami $(50{:}\,60{:}\,75)$, resp. $(75{:}\,60{:}\,50)$. Použili sme 12 možností pre pomer najmenšieho aj najväčšieho rozptylu od 1 : 1 do 1 : 9. Celkovo sme analyzovali 2 x 3 x 2 x 9 = 108 skupín. Pre každú konkrétnu možnosť zvolených parametrov sme vytvorili 5000 simulácií, pričom pre každú z nich sme realizovali klasickú aj Welchovu analýzu rozptylu. Výsledky sme pre každú kombináciu možností znázornili graficky. Na každom grafe sú na osi x všetky možnosti pomeru rozptylov ($\sigma^2_{min}/\sigma^2_{max}$), y-ové súradnice bodov sú na úrovni sily testu (t. j. $1-\hat{\beta}$). Hranica na dostatočnú silu ($0{,}8$) [12] je znázornená prerušovanou čiarou, takže tie body, ktoré sa nachádzajú nad ňou, predstavujú akceptovateľnú množinu východiskových podmienok z hľadiska chyby 2. druhu.

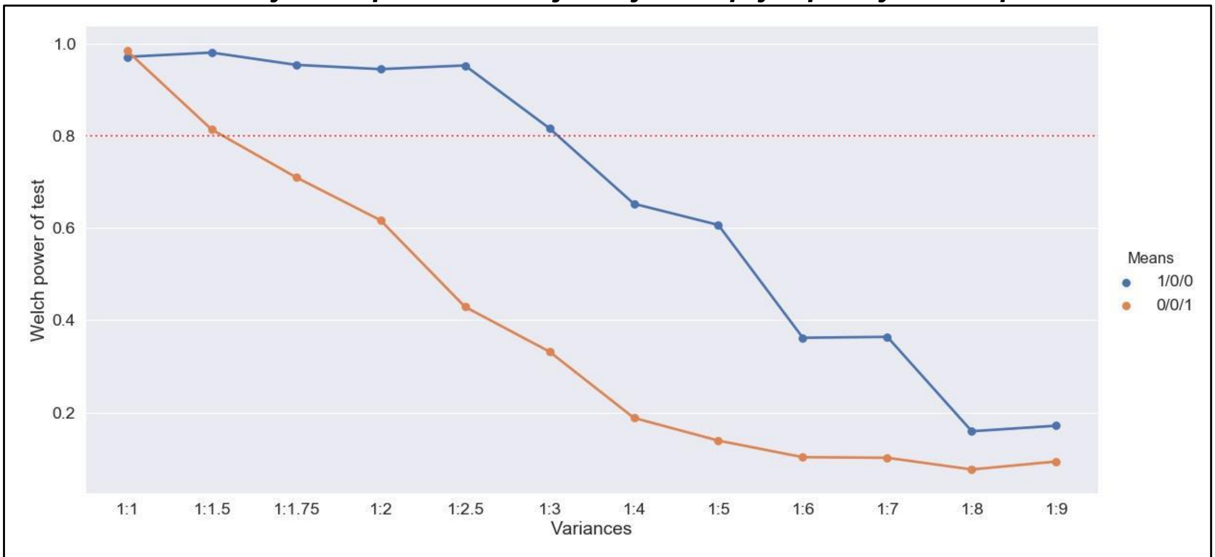
Na grafe č. 3 a č. 4 sú znázornené výsledky pre súbory s rovnakými početnosťami. Farbami sú odlíšené možnosti, ktoré vyjadrujú, či skupina s najväčším rozptylom je tá, ktorá má odlišnú hodnotu (oranžová farba), alebo má odlišnú strednú hodnotu skupina s najmenším rozptylom (modrá farba).

Graf č. 3: Odhad sily testu pri klasickej analýze rozptylu pre vyvážené pozorovania



Zdroj: vlastné spracovanie

Graf č. 4: Odhad sily testu pri Welchovej analýze rozptylu pre vyvážené pozorovania



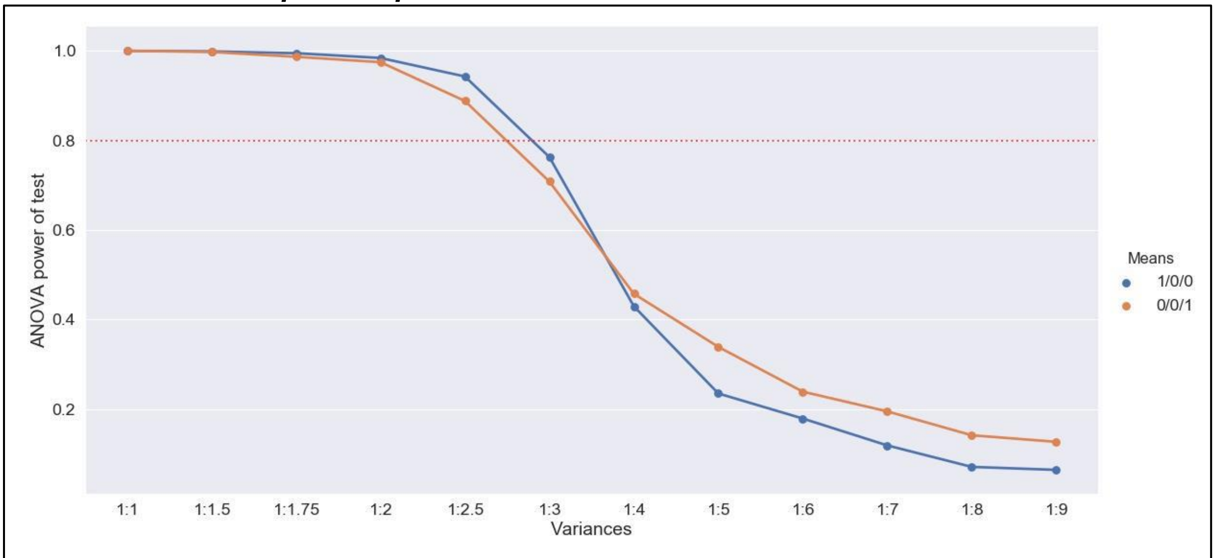
Zdroj: vlastné spracovanie

Ako vidno z grafu č. 3, pri klasickej analýze rozptylu v prípade, že skupina s odlišnou strednou hodnotou je tá, ktorá má najnižšiu variabilitu, spĺňajú podmienku prvé štyri prípady: po pomer rozptylov 1 : 2, ak má odlišnú hodnotu stredná hodnota skupiny s najväčším rozptylom, vyhovujú len prípady do pomeru 1 : 1,75. Smerom doprava (k nižším hodnotám pomerov) sa lomená čiara zvažuje nadol (toto však platí aj pre všetky nasledujúce prípady – veľké disproporcie medzi rozptylmi znižujú silu testu bez ohľadu na ostatné vstupné podmienky).

Pri Welchovej analýze rozptylu badať výraznejší rozdiel medzi prípadmi znázornenými odlišnými farbami: kým modrá lomená čiara je nad hranicou 0,8 až po pomer rozptylov 1 : 3 (dokonca pri predchádzajúcich pomeroch je nad úrovňou 0,9), pre oranžovú platí, že vyhovujúca sila testu je len pre prvé dva prípady: homoskedastickom (pomer 1 : 1) a pomere 1 : 1,5. Znamená to, že z hľadiska sily testu je oveľa priaznivejšia možnosť, ak súbor s odlišnou strednou hodnotou má najmenší rozptyl.

Graf č. 5 znázorňuje odhady sily testu klasickej analýzy rozptylu pre nevyvážené súbory, pričom počty pozorovaní sú v priamom vzťahu s variabilitou ($n_1 < n_2 < n_3$ a súčasne $\sigma_1^2 < \sigma_2^2 < \sigma_3^2$). Vidíme, že lomené čiary majú veľmi podobný tvar, v oboch prípadoch (ak je odlišná stredná hodnota súboru s najmenším, resp. s najväčším rozptylom) platí, že až po pomer rozptylov 1 : 2,5 je sila testu nad hranicou 0,8, pri väčších rozdieloch medzi rozptylmi pomerne strmo klesá.

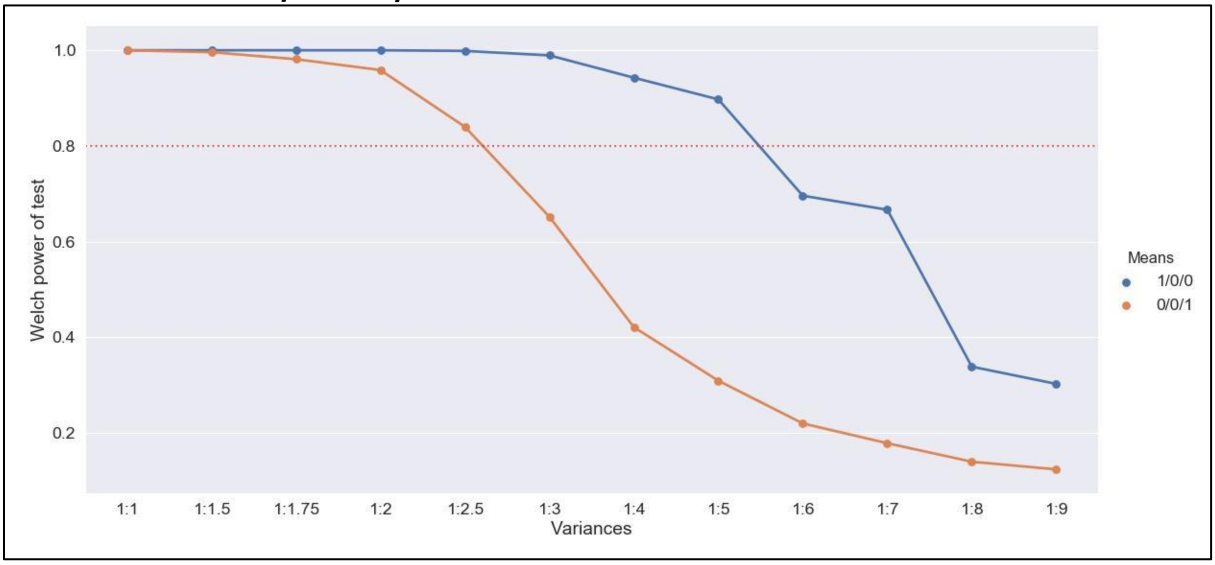
Graf č. 5: Odhad sily testu pri klasickej analýze rozptylu v prípade priameho vzťahu medzi variabilitou a počtom pozorovaní



Zdroj: vlastné výpočty

Na grafe č. 6 sú znázornené odhady sily testu Welchovej analýzy rozptylu pre analogické vzťahy medzi počtami pozorovaní a variabilitou ($n_1 < n_2 < n_3$ a súčasne $\sigma_1^2 < \sigma_2^2 < \sigma_3^2$). Oproti klasickej analýze rozptylu tu (podobne ako pri vyvážených súboroch) badať markantný rozdiel medzi odhadovanými silami testov. Kým pri súboroch, v ktorých je najnižšia variabilita v súbore, ktorý má odlišnú strednú hodnotu (modrá čiara), je sila testu nad hranicou 0,8 ešte pre pomery rozptylov 1 : 5, v prípade, že súbor s odlišnou strednou hodnotou má najvyššiu variabilitu, je posledný akceptovateľný pomer len 1 : 2,5.

Graf č. 6: Odhad sily testu pri Welchovej analýze rozptylu v prípade priameho vzťahu medzi variabilitou a počtom pozorovaní

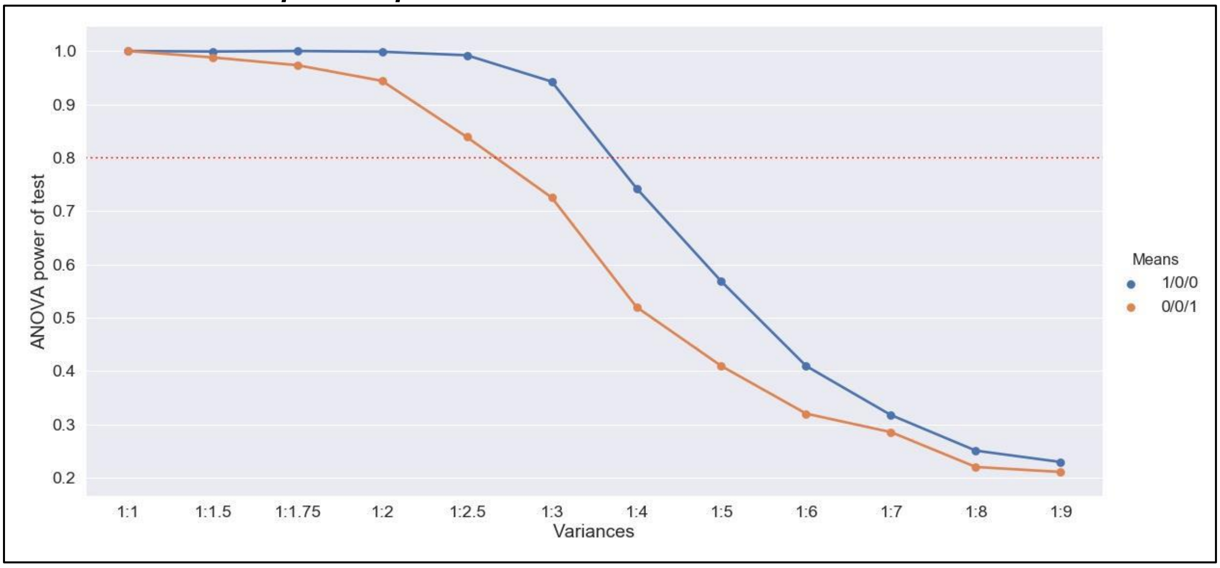


Zdroj: vlastné výpočty

Nakoniec sme sa venovali prípadu, keď medzi rozsahmi súborov a variabilitou existuje nepriamy vzťah ($n_1 < n_2 < n_3$ a súčasne $\sigma_1^2 > \sigma_2^2 > \sigma_3^2$). Výsledky sú znázornené na grafe č. 7 a č. 8.

Pri klasickej analýze rozptylu sú znovu z hľadiska sily testu lepšie vlastnosti vtedy, keď najnižšiu variabilitu má súbor s odlišnou strednou hodnotou (modrá lomená čiara) – vyhovujúci pomer rozptylov je ešte 1 : 3, vtedy, keď má súbor s odlišnou strednou hodnotou najvyššiu variabilitu, je tento pomer len 1 : 2,5.

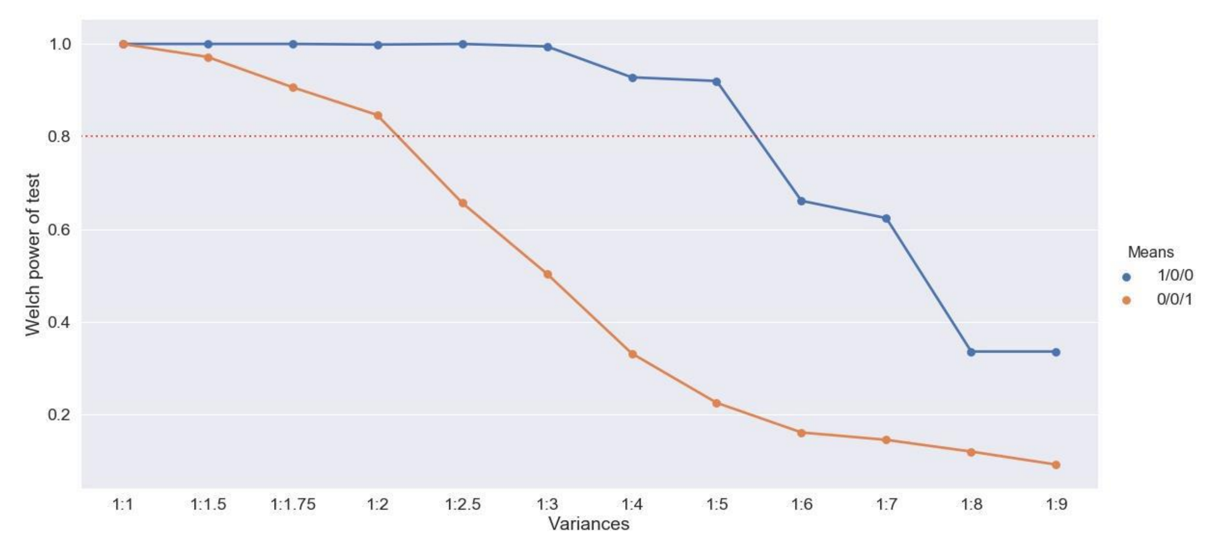
Graf č. 7: Odhad sily testu pri klasickej analýze rozptylu v prípade nepriameho vzťahu medzi variabilitou a počtom pozorovaní



Zdroj: vlastné výpočty

Výraznejší rozdiel medzi lomenými čiarami vidíme pri Welchovej analýze rozptylu (graf č. 8): modrá lomená čiara (súbor s odlišnou strednou hodnotou má najnižšiu variabilitu) sa až po pomer rozptylov 1 : 3 drží veľmi tesne pri hodnote 1, pri ďalších dvoch pomeroch (1 : 4 a 1 : 5) je ešte nad úrovňou 0,9, hodnoty nižšie ako 0,8 má odhadovaná sila testu až od pomeru 1 : 6. Pre skupiny údajov, kde súbor s odlišnou strednou hodnotou má najvyššiu variabilitu, už od pomeru rozptylov 1 : 2,5 je odhadovaná sila testu pod hranicou 0,8.

Graf č. 8: Odhad sily testu pri Welchovej analýze rozptylu v prípade nepriameho vzťahu medzi variabilitou a počtom pozorovaní



Zdroj: vlastné výpočty

6. ZÁVER

Realizovaním simulácií, v ktorých sme nastavili rôzne vstupné parametre porovnávaných troch súborov, sme pomocou odhadu chýb 1. druhu a 2. druhu (resp. sily testu) porovnali klasickú a Welchovu analýzu rozptylu.

Z hľadiska veľkosti chýb 1. druhu sa Welchova analýza rozptylu ukázala ako vhodnejšia metóda pre každý z nami zvolených pomerov rozptylov (až do pomeru najmenšieho a najväčšieho rozptylu 1 : 30) a to bez ohľadu na vzájomné vzťahy medzi rozsahmi súborov, pretože odhady chýb vo všetkých prípadoch vyšli v intervale od 0,025 po 0,075. Pri klasickej analýze rozptylu sa ukázal rozhodujúci vzťah medzi rozsahmi súborov a ich variabilitou. Ak bol priamy ($n_1 < n_2 < n_3$ a súčasne $\sigma_1^2 < \sigma_2^2 < \sigma_3^2$), pre všetky pomery rozptylov bola odhadovaná sila v požadovanom intervale. Pre súbory s rovnakým rozsahom je akceptovateľný ešte pomer rozptylov 1 : 9, ale ak sú rozsahy a rozptyly v nepriamom vzťahu, nevyhovujúce pomery sa začínajú už na úrovni 1 : 1,75.

Pri analýze veľkosti chýb 2. druhu (resp. sily testu) sme skonštatovali, že vo všetkých porovnávaných prípadoch má na odhad sily vplyv najmä to, či má súbor s odlišnou strednou hodnotou najnižšiu alebo najvyššiu variabilitu. Pre silu testov sa ukázala ako priaznivejšia možnosť, ak súbor s odlišnou strednou hodnotou mal najnižšiu variabilitu, pričom rozdiely boli oveľa výraznejšie pri Welchovej analýze rozptylu ako pri klasickej. Vzhľadom na viaceré možnosti vstupných okolností sme prehľadné porovnanie klasickej a Welchovej analýzy rozptylu zapísali do tabuľky č. 1. V nej sú uvedené tie hodnoty pomerov rozptylov, pri ktorých má príslušná metóda za uvedených okolností silu ešte nad hranicou 0,8. Priaznivejšiu možnosť (ak si máme vybrať z porovnávaných metód) sme zvýraznili. Znamená to, že za okolností zodpovedajúcich záhlaviu tabuľky má metóda z daného riadku výhodu v tom, že aj väčšia disproporcia medzi rozptylmi významne neznižuje silu testu.

Tabuľka č. 1: Porovnanie sily testu klasickej a Welchovej analýzy rozptylu (pomery rozptylov $(\sigma^2_{min}/\sigma^2_{max})$***, pre ktoré je odhad sily testu nad úrovňou 0,8)***

parametre / metóda	$n_1 = n_2 = n_3$		$n_1 < n_2 < n_3$ $\sigma_1^2 < \sigma_2^2 < \sigma_3^2$		$n_1 < n_2 < n_3$ $\sigma_1^2 > \sigma_2^2 > \sigma_3^2$	
metóda	M[1]	O[2]	M	O	M	O
klasická	1:2	**1:1,75**	1:2,5	1:2,5	1:3	**1:2,5**
Welchova	**1:3**	1:1,5	**1:5**	1:2,5	**1:5**	1:2

Zdroj: vlastné spracovanie

Pri overovaní zhody stredných hodnôt v troch súboroch v praxi je v každom prípade potrebné najskôr overiť homoskedasticitu. Ak sa potvrdí, respektíve ak je pomer najmenšieho a najväčšieho rozptylu do 1 : 1,5, možno použiť klasickú metódu analýzy rozptylu. Ak sú medzi rozptylmi výraznejšie rozdiely je vhodné zamerať sa na vzájomné vzťahy medzi rozsahmi a variabilitou, prípadne (na základe charakteristík jednotlivých súborov, ktorými sa odhadujú charakteristiky základných súborov) zistiť, aký je vzťah medzi strednými hodnotami a variabilitou tak, aby sme sa mohli (do určitej miery) riadiť kritériami na základe tabuľky č. 1.

Náš príspevok rieši problematiku porovnania dvoch metód v rámci vstupných podmienok, ktoré sme si vopred stanovili. Treba si uvedomiť, že sme pracovali len s tromi súbormi, premenná mala normálne rozdelenie, pričom sme si zvolili len určité pomery rozsahov súborov. Pri väčšom počte súborov alebo väčších disproporciách medzi rozptylmi môžu byť výsledky iné. (Napríklad [Abdi, 2007]) uvádza, že v prípade ak je najpočetnejší súbor aspoň 4-násobne väčší ako najmenší zo súborov, odporúča sa použiť Welchovu analýzu rozptylu.) Navyše, medzi strednými hodnotami neboli výrazné rozdiely, čo tiež môže zohrať úlohu pri rozhodovaní sa, ktorá z metód je vhodnejšia. Preto možno naše závery aplikovať len za predpokladov, ktoré sú podobné ako tie naše. Na komplexnejšiu analýzu by bolo vhodné pracovať s viacerými súbormi, použiť aj iné rozdelenia, voliť rôzne pomery medzi rozsahmi súborov, prípadne rozšíriť spektrum vzťahov medzi parametrami výberových súborov.

LITERATÚRA

[1] ABDI, H.: O'Brien test for homogeneity of variance. In: Encyclopedia of measurement and statistics. London: SAGE Publication Ltd. 2007, s. 701 – 704. ISBN 9781412916110.

[2] BROWN, M. B. – FORSYTHE, A. B.: The ANOVA and multiple comparisons for data with heterogeneous variances. In: Biometrics. Wiley Online Library, 1974, s. 719 – 724. ISSN 0006-3444.

[3] ĎŘÍMAL, J. – TRUNEC, D.: Úvod do metody Monte-Carlo. 1. vyd. Brno: UJEP. 1989. 122 s. ISBN 80-210-0022-8.

[4] KOTLEBOVÁ, E. – ŠOLTÉS, E. – SIVAŠOVÁ, D.: Štatistická indukcia v príkladoch. 1. vydanie. Bratislava: EKONÓM, 2015, 222 s. ISBN 978-80-225-4135-0.

[5] LIU, H.: Comparing Welch ANOVA, a Kruskal-Wallis test, and traditional ANOVA in case of heterogeneity of variance. A Thesis. Richmond, Virginia: Virginia

[1] M – modrá farba lomenej čiary (súbor s odlišnou strednou hodnotou má najnižšiu variabilitu).

[2] O – oranžová farba lomenej čiary (súbor s odlišnou strednou hodnotou má najvyššiu variabilitu).

Commonwealth University, 2015. 46 s. [online]. [cit. 5. 1. 2021]. Dostupné na: https://scholarscompass.vcu.edu/etd/3985
[6] MARTINEZ, W. L. – MARTINEZ A. R.: Computational statistics handbook with MATLAB. 2. vydanie. Boca Raton: FL: Chapman, 2008, 767 s. ISBN 15-848-8566-1.
[7] MAXWELL, S. E. – DELANEY, H. D. – KELLEY, K.: Designing experiments and analyzing data: A model comparison perspective. Boca Raton: Routledge, 2017. 1080 s. ISBN 978-1138892286.
[8] PACÁKOVÁ, V. – KOTLEBOVÁ, E. – LABUDOVÁ, V. – SIPKOVÁ, Ľ.: Štatistická indukcia pre ekonómov. 1. vydanie. Bratislava: EKONÓM, 2012. 317 s. ISBN 978-80-225-3382-9.
[9] PACÁKOVÁ, V. – LABUDOVÁ, V. – SIPKOVÁ, Ľ. – STANKOVIČOVÁ, I.: Štatistická indukcia pre ekonómov a manažérov. 1. vydanie. Bratislava: Wolters Kluwer, 2015. 365 s. ISBN 978-80-8168-081-6.
[10] SLIACKY, M.: Testy hypotéz porovnávajúcich parametre viac ako dvoch súborov. Diplomová práca. Fakulta hospodárskej informatiky, Ekonomická univerzita, Bratislava, 2020. Evidenčné číslo: 103005/I/2020/421000214186
[11] SOIKLIEW, K. – ARAVEEPORN, A.: Modifications of Levene's and O'Brien's Tests for Testing the Homogeneity of Variance Based on Median and Trimmed Mean, Bangkok: Thai Statistical Association, 2018, s. 106 – 128. ISSN 2351-0676. [online]. [cit. 5.1.2021]. Dostupné na: https://ph02.tci-thaijo.org/index.php/thaistat/article/view/135555/101285
[12] WALMSLEY, A. – L. E., BROWN, M. C.: What Is Power?. [online]. [cit. 5.1.2021]. Dostupné na: https://www.statisticsteacher.org/2017/09/15/what-is-power/
[13] WELCH, B. L.: On the comparison of several mean values: an alternative approach. In Biometrika, 1951, s. 330 – 336. ISSN 0006-3444.
[14] ZAR, J. H.: Biostatistical Analysis. 5. vyd. Englewood Cliffs: NJ: Prentice Hall, Inc. 2010. 960 s. ISBN 978-0321656865.

RESUMÉ

Príspevok sa zaoberá možnosťami porovnania stredných hodnôt spojitej kvantitatívnej premennej vo viac ako dvoch nezávislých súboroch (pracovali sme s trojicami súborov), ak je porušený predpoklad homoskedasticity pri analýze rozptylu. Cieľom bolo porovnanie klasickej a Welchovej analýzy rozptylu z hľadiska možností nastania chýb pri testovaní. Pomocou metódy Monte Carlo sme generovali trojice hodnôt súborov s vopred stanovenými vlastnosťami, na ktoré sme aplikovali klasickú aj Welchovu analýzu rozptylu. Výsledky slúžili na odhad miery chýb 1. a 2. druhu, pomocou ktorých sme porovnávali uvedené metódy. Ukázalo sa, že okrem overenia homoskedasticity pred samotným testovaním je vhodné na základe opisných charakteristík súborov zistiť vzťah medzi strednými hodnotami a variabilitou (či v súbore, ktorého stredná hodnota sa najviac odlišuje od zvyšných dvoch, je v porovnaní s ostatnými vysoká alebo nízka variabilita), pretože práve táto okolnosť môže výrazne redukovať silu testu.
Dosiahnuté výsledky môžu byť vhodným podkladom na rozhodnutie, ktorú z dvoch metód treba v danej situácii použiť.

RESUME

The paper deals with the possibilities of comparing the mean values of continuous quantitative variable in multiple independent data sets (we have worked with three data sets), if the presumption of homoscedasticity in the analysis of variance is violated. The aim was to compare the classical and Welch analysis of variance in terms of the

possibility of testing errors. Using the Monte Carlo method, three values of data sets with predetermined properties were generated, on which both classical and Welch analysis of variance were applied. The results were used to estimate the error rate of type I and l type II by means of which the above methods were compared. It has been shown that in addition to verifying homoscedasticity before the actual testing, it is appropriate to determine the relationship between mean values and variability, based on the descriptive characteristics of the data sets (whether there is a high or low variability in the data set whose mean value differs the most from the other two), since it is this circumstance that can significantly reduce the strength of the test.
The results obtained can be a suitable basis for deciding which of the two methods should be used in a given situation.

PROFESIJNÝ ŽIVOTOPIS

RNDr. Eva Kotlebová, PhD., *je absolventkou Matematicko-fyzikálnej fakulty Univerzity Komenského v Bratislave (vedecký smer matematika – teória systémov). Po ukončení vysokoškolského štúdia bola tri roky na študijnom pobyte na Katedre štatistiky Fakulty riadenia Vysokej školy ekonomickej v Bratislave. Potom pôsobila niekoľko rokov ako stredoškolská učiteľka matematiky na gymnáziu v Bratislave. Od roku 2003 pracuje na Katedre štatistiky Fakulty hospodárskej informatiky v Bratislave. V roku 2008 ukončila doktorandské štúdium. Venuje sa štatistickej indukcii, bayesovskej štatistike a aplikácii štatistických metód v poisťovníctve.*

Ing. Martin Sliacky *je absolventom Fakulty hospodárskej informatiky Ekonomickej univerzity v Bratislave (študijný program Štatistické metódy v ekonómii). Pracuje ako dátový analytik v zaisťovni Swiss Re.*

KONTAKT

eva.kotlebova@euba.sk
martin@sliacky.com