ROBUST'98, 89 – 101

ANOVA – JEJÍ ROBUSTNOST A SÍLA

Eva JAROŠOVÁ a Jiří REIF[1]

VŠE KSTP a ZČU KMA

Abstract. The first part of the paper is devoted to the robustness of the classical analysis of variance with respect to various departures from the standard hypotheses. Some Monte Carlo simulations are carried out for the one-way layout. For the two-way layout we present a brief survey of results comparing the power of parametric and nonparametric methods. In the second part the power of several tests for the two sample location-scale problem is compared via a simulation and a new nonparametric test is suggested which appears to be quite powerful in many cases.
Резюме: Первая часть работы посвящена робастности дисперсионного анализа по разным отклонениям от классических предположений. Во второй части мы изучаем мощность разных критериев для двухвыборочных задач, в которых распределения могут отличаться совместно в положении (сдвиге) и в рассеянии (масштабе). Для такой задачи мы вводим один новый критерий, который довољно мощным в многих случаях. Для сравнений испољзуется метод Монте Карло.

1. Generátor pseudonáhodných čísel

Výsledky, které uvádíme, byly vypočteny za pomoci generátoru Wichmanna a Hilla podle popisu publikovaného Antochem (1990) vzhledem k referencím na jeho dobré vlastnosti a dlouhou periodu. Standardně jsme používali 10 000 simulací. Vyzkoušeli jsme i jiné generátory (mj. generátor Turbo Pascalu 7.0), výsledky se lišily jen v rámci příslušných intervalů spolehlivosti pro odhadovaný parametr alternativního rozdělení.

2. Uvažovaná rozdělení pravděpodobnosti

Rozdělení, ze kterých jsme uvažovali náhodné výběry, zahrnovala hustoty symetrické i nesymetrické s různou délkou chvostů, hustoty bimodální i hustoty neomezené. Ve výčtu, který následuje, uvádíme definiční obor hustot jen tehdy, není-li roven celé reálné ose.

1. normální normalizované; hustotu označíme $\varphi(x)$

2–3. směs normálních $f(x) = \big[\varphi(x) + \varphi(x-d)\big]/2$ pro $d = 5$ a $d = 100$

4. logistické $f(x) = e^{-x}(1+e^{-x})^{-2}$

5. Cauchyho $f(x) = \big[\pi(1+x^2)\big]^{-1}$

6. rovnoměrné na (0, 1)

7. exponenciální

[1]Druhý autor byl podporován grantem MŠMT VS 97159

8–9. Weibullovo $f(x) = cx^{c-1}\exp(-x^c),\ x > 0$, pro $c = 2$ a $c = 0.5$

10. dvojitě exponenciální $f(x) = (1/2)\exp(-|x|)$

11. Simpsonovo (trojúhelníkové) $f(x) = 1 - |x|$ pro $x \in (-1, 1)$

12. arcsinové $f(x) = (1/\pi)(1 - x^2)^{-1/2}$ pro $x \in (-1, 1)$

13–14. mocninné $f(x) = cx^{c-1},\ x \in (0, 1)$, pro $c = 2$ a $c = 0.5$

15–16. Paretovo $f(x) = cx^{-(1+c)},\ x > 1$, pro $c = 1$ a $c = 3$

17. symetrizované Paretovo pro $c = 3$, $f(x) = (3/2)(1 + |x|)^{-4}$

Všimněme si, že směs normálních rozdělení s parametrem $d = 100$ je v jistém ohledu blízká k dvouhodnotovému diskrétnímu rozdělení. Pro Cauchyho rozdělení a Paretovo rozdělení s parametrem 1 neexistuje střední hodnota a rozptyl. Pro tato dvě rozdělení jsme sice mohli vyšetřovat pravděpodobnost chyby 1. druhu, avšak při zkoumání síly testu jsme tato dvě rozdělení již z úvah vynechali, neboť jsme pro jednotlivá rozdělení uvažovali posuny středních hodnot vyjádřené v násobcích směrodatné odchylky.

3. Vliv nenormality na pravděpodobnost chyby I. druhu při jednoduchém třídění

Uvažujme základní model analýzy rozptylu jednoduchého třídění, viz např. Anděl (1978), str. 151. V tomto modelu se předpokládá, že výběry pocházejí z normálního rozdělení. Pearson (1931) provedl simulační studii vlivu nenormality rozdělení na chybu I. druhu a odhadnul velikost této chyby pro rozdělení s různými hodnotami koeficientů šikmosti a špičatosti. Dále se touto problematikou zabývali např. Gayen (1950), David a Johnson (1951), Box a Watson (1962) a Tiku (1964), kteří používali různých aproximací pro vyjádření testové statistiky v závislosti na momentech skutečného rozdělení. Donaldson (1968) provedl simulační studii pro exponenciální a lognormální rozdělení. Výčet těchto prací samozřejmě není úplný.

Splňuje-li rozdělení, ze kterých provádíme výběry, předpoklady centrální limitní věty, nemá nenormalita na chybu I. druhu klasické analýzy rozptylu asymptoticky vliv (viz např. Scheffé, 1959). Proto se naše simulační studie zaměřila zejména na výběry malého rozsahu.

Pro rozdělení vyjmenovaná v odst. 2 jsme zkoumali nejprve dva výběry s rozsahy $1 \le n_1 \le n_2 \le 10$ (samozřejmě s výjimkou případu $n_1 = n_2 = 1$) pro každé uvedené rozdělení, tedy 54 dvojic výběrů, viz např. níže uvedená matice výsledků pro Cauchyho rozdělení. Dále jsme vyšetřovali tři dvojice výběrů se stejnými rozsahy 20, 30 a 50, tři trojice výběrů se stejnými rozsahy 2, 3 a 10 a tři pětice výběrů se stejnými rozsahy 2, 3 a 10. Celkem tedy bylo provedeno pro každé rozdělení 63 voleb počtu výběrů a přísl. rozsahů a pro každou z uvedených voleb provedeno 10 000 simulací. Pro kontrolu, zda bylo generováno správné rozdělení, jsme zaznamenávali aritmetický průměr a rozptyl generovaných hodnot a porovnávali je s teoretickými hodnotami pro přísl. rozdělení. Samozřejmě, tato kontrola nebyla možná pro Cauchyho rozdělení a Paretovo rozdělení s parametrem 1, pro která střední hodnota a rozptyl nejsou definovány, takže při různých volbách (a také různých generátorech pseudonáh. čísel) vychází naprosto odlišné průměrné simulované výběrové charakteristiky. Věnovali jsme se pouze případu, kdy kritický obor testu byl volen tak, aby pro normální rozdělení odpovídal hladině významnosti 0.05. Odhady pravděpodobnosti chyby I. druhu budeme dále označovat $\hat{\alpha}$.

```
ANOVA (one-way) - simulovana  chyba I. druhu v procentech
pocet vyberu: 2   rozsahy jsou uvedeny v 1. radku a 1. sloupci
pocet simulaci =  10 000    $\alpha$ = 0.05
pouzita pseudonahodna cisla: generator Wichmanna a Hilla
rozdeleni:  Cauchyho
```

	1	2	3	4	5	6	7	8	9	10
1	-	8.4	11.2	11.4	11.0	10.5	10.0	9.8	9.0	9.2
2		2.9	2.3	2.6	3.5	4.1	5.3	8.4	8.9	8.8
3			2.2	2.0	2.3	2.8	3.3	3.5	3.9	4.5
4				2.0	1.9	2.2	2.1	2.5	3.3	3.4
5					2.2	2.1	2.3	2.1	2.7	2.9
6						2.0	2.0	2.3	2.3	2.5
7							1.9	2.0	2.3	2.2
8								2.2	1.9	2.1
9									2.2	2.0
10										2.2

Pro normální rozdělení vyšlo pro vyšetřované případy nejmenší $\hat{\alpha}$ rovno 0.045, největší pak 0.055, což není v rozporu s příslušnými intervaly spolehlivosti pro $\hat{\alpha}$ odvozenými pomocí binomického rozdělení.

Odhad pravděpodobnosti chyby I. druhu $\hat{\alpha}$ překročil 0.10 pouze v některých případech pro směsi normálních rozdělení (pro parametr $d = 100$ i $d = 5$), Weibullovo rozdělení s parametrem $c = 0.5$, Cauchyho rozdělení a Paretova rozdělení (pro $c = 1$ i $c = 3$). V případě, že rozsahy všech výběrů byly alespoň 2, překročil odhad $\hat{\alpha}$ hodnotu 0.10 pouze pro směs normálních rozdělení s parametrem $d = 100$, kdy $\hat{\alpha} = 0.129$ pro $n_1 = n_2 = 2$. Pro rozsahy výběrů alespoň 3 bylo vždy $\hat{\alpha} \leq 0.084$ (po vynechání směsi normálních rozdělení s parametrem $d = 100$ bylo dokonce vždy $\hat{\alpha} \leq 0.065$).

Všimněme si podrobněji výsledků pro Cauchyho rozdělení a Paretovo rozdělení s parametrem $c = 1$, která nesplňují předpoklady centrální limitní věty. Výsledky se pro tato dvě rozdělení od sebe lišily jen velmi málo. Pro Cauchyho rozdělení uvádíme tabulku hodnot $\hat{\alpha}$ pro $1 \leq n_1 \leq n_2 \leq 10$; pro ostatní uvažované rozsahy výběrů vyšlo vždy $\hat{\alpha}$ hluboko pod 0.05, např.

$\hat{\alpha} = 0.020$ pro $n_1 = n_2 = 50$ ($\hat{\alpha} = 0.019$ pro Paretovo rozdělení a tytéž rozsahy výběrů). Překvapivé je, že na rozdíl od nestabilních výběrových momentů jsou výsledky analýzy rozptylu pro tato rozdělení stabilní, jak ukázaly pokusy provedené též s jinými generátory pseudonáhodných čísel.

Celkově lze říci, že klasická ANOVA je robustní vůči odchylkám od normality. Kompletní výsledky je možno si vyžádat od autorů.

4. Vliv heteroskedasticity na pravděpodobnost chyby I. druhu (jednoduché třídění)

V knize Scheffého (1959) je pro jednoduché třídění odvozeno, že je-li počet výběrů roven dvěma, nemá heteroskedasticita asymptoticky na pravděpodobnost chyby I. druhu vliv, zatímco v případě více než dvou výběrů heteroskedasticita asymptoticky tuto chybu zvětší. Pro omezený rozsah výběrů byl problém studován řadou autorů. Obecné případy nestejných rozptylů zahrnují např. již dříve zmíněné práce David a Johnson (1951) a Tiku (1964). V naší studii jsme si všímali pouze případu dvou výběrů stejných rozsahů 4, 5, 6, 8, 10, 20, 40, a to pro všechna rozdělení vyjmenovaná v odstavci 2, která měla konečný rozptyl, tedy vyloučeno bylo rozdělení Cauchyho a Paretovo s parametrem 1. Zkoumali jsme případ, kdy jeden z výběrů pocházel z rozdělení s dvojnásobnou, resp. trojnásobnou směrodatnou odchylkou. Pro $\sigma_2 = 2\sigma_1$ uvádíme výsledky pro normální rozdělení a ta rozdělení, pro která byl vliv heteroskedasticity znatelně vyšší.

Odhadovaná pravděpodobnost chyby I. druhu pro výběry z některých rozdělení, $\sigma_2 = 2\sigma_1$

rozsah výběrů	4	5	6	8	10	20	40
normální rozdělení	0.063	0.057	0.062	0.054	0.051	0.053	0.047
směs normálních, d = 5	0.094	0.073	0.066	0.061	0.057	0.053	0.053
směs normálních, d =100	0.128	0.074	0.053	0.057	0.063	0.057	0.057
rovnoměrné r.	0.077	0.068	0.062	0.064	0.059	0.055	0.046
exponenciální r.	0.082	0.080	0.078	0.067	0.066	0.063	0.054
Weibullovo r. s par. 0.5	0.151	0.144	0.138	0.122	0.114	0.099	0.086
arcsinové r.	0.087	0.075	0.064	0.058	0.051	0.058	0.053
mocninné r. s par. 0.5	0.087	0.083	0.072	0.061	0.062	0.056	0.053
mocninné r. s par. 2	0.074	0.073	0.063	0.061	0.060	0.052	0.053
Paretovo r. s par. 3	0.101	0.104	0.099	0.093	0.088	0.082	0.067

5. Vliv závislosti pozorování (příklad)

V naší studii se nezabýváme vlivem závislosti pozorování. Je však známo, že právě závislost pozorování mění v analýze rozptylu podstatně chybu I. druhu. Diskuse s doc. J. A. Víškem nás vedla k následujícímu příkladu.

Uvažujme analýzu jednoduchého třídění pro dva nezávislé výběry, kdy ANOVA je ekvivalentní oboustrannému t-testu shody teoretických středních hodnot pro výběry z normálních rozdělení se stejnými rozptyly. Mějme tedy nezávislé náhodné výběry $x_1, \dots, x_n$ a $y_1, \dots, y_n$ z rozdělení $N(\mu, \sigma^2)$, výběrové rozptyly označme $s_{x,1}^2$, $s_{y,1}^2$ a aritmetický průměr těchto rozptylů pak s_1^2. Rozšiřme nyní statistické soubory o závislá pozorování $x_{n+i} = x_i$ a $y_{n+i} = y_i$ pro $i = 1, \dots, n$. Rozsahy zvětšených souborů budou nyní $2n$, jejich průměry budou stejné jako průměry původních výběrů a výběrové rozptyly $s_{x,2}^2$, $s_{y,2}^2$ zvětšených souborů a jejich aritmetický průměr s_2^2 obdržíme tak, že odpovídající charakteristiky původních výběrů znásobíme koeficientem $c(n) = 2(n-1)/(2n-1)$. Speciálně tedy je $s_2 = \sqrt{c(n)}s_1$. Definujeme-li nyní pro $i = 1, 2$ statistiky

$$t_i = (\bar{x} - \bar{y})s_i^{-1}\sqrt{in/2},$$

pak t_1 má t-rozdělení pravděpodobnosti s počtem stupňů volnosti $2n-2$ a platí $t_2 = \sqrt{2/c(n)}t_1$. Kdybychom se zvětšenými soubory zacházeli jako s náhodnými výběry, uvažovali bychom pro t_2 kritický obor $|t_2| > t_{1-\alpha/2}(4n-2)$. Tato nerovnost je však vzhledem k závislosti měření ekvivalentní nerovnosti $|t_1| > \sqrt{c(n)/2}\, t_{1-\alpha/2}(4n-2)$, která má v limitním případě $n \to \infty$ pravděpodobnost $2\big(1 - \Phi(u_{1-\alpha/2}/\sqrt{2})\big)$. Např. pro $\alpha = 0.05$ bychom obdrželi hodnotu přibližně 0.166, tj. v uvažovaném příkladě se pravděpodobnost chyby I. druhu v důsledku závislosti pozorování více než ztrojnásobila.

6. Síla analýzy rozptylu (jednod. třídění)

Síla analýzy rozptylu při jednoduchém třídění byla vyšetřována mnoha autory. Např. Srivastava (1959) založil výpočet síly na aproximaci rozdělení statistiky F čtyřmi členy Edgeworthovy řady, Tiku (1971) využil rozvoje součtu čtverců pomocí Laguerrových polynomů. Výpočtem síly při obecném rozdělení se zabývá již výše zmíněná práce Davida a Johnsona (1951). Práce Donaldsona (1968) obsahuje odhady síly testu získané metodou Monte Carlo pro normální, lognormální a exponenciální rozdělení včetně případu nestejných rozptylů ve skupinách.

Alternativou ke klasické (parametrické) analýze rozptylu jsou permutační testy, neparametrické testy (např. Kruskal-Wallisův test) nebo kombinace parametrických a pořadových testů, viz např. Pitman (1937a,b), Welch (1937), Kruskal a Wallis (1959), Siegel (1956), Fraser (1957), Kemthorne (1952), Scheffé (1959), Miller (1981).

V řadě prací se porovnáná síla různých postupů. Například asymptotická relativní vydatnost (ARE) Kruskal-Wallisova testu v porovnání s parametrickou analýzou rozptylu je za předpokladu normálního rozdělení rovna $3/\pi$ (viz např. Gibbons, 1971). Jiné srovnání síly provedl O'Gorman (1997), který porovnával klasickou analýzu rozptylu, Kruskal-Wallisův test, test s normálními skóry a jistý test s tzv. adaptivními skóry. Uvažoval tři nebo více výběrů o velkém rozsahu z různých typů rozdělení, síla byla vyšetřována vzhledem k posunům středních hodntot při nezměněném rozptylu. Např. pro výběry nejnižších rozsahů, které byly v článku uvažovány ($n = 24$), neměl žádný z uvedených testů jednoznačnou převahu proti ostatním, pro výběry podstatně větších rozsahů se však začínala projevovat výhoda adaptivního přístupu.

V případě, že budeme uvažovat jen dva výběry, splývá ANOVA s oboustranným t-testem pro výběry z normálních rozdělení se stejnými rozptyly. Tohoto případu se bude týkat námi provedená simulační studie, jejíž výsledky jsou popsány v odstavci 8.

7. Dvojné třídění

Pro dvojné třídění byla rovněž studována celá řada neparametrických postupů a často jsou pak prováděna různá srovnání síly. Např. Gilbert (1972) porovnal sílu klasického postupu a tří neparametrických metod (mj. Friedmanova testu). Někteří autoři doporučují aplikovat klasickou parametrickou metodu na pořadí místo na původní data (tzv. RT test), viz např. Iman (1974) nebo Conover a Iman (1981). Iman (1974) uvádí simulační studii síly RT testu při dvojném třídění pro rozdělení normální, kontaminované normální a exponenciální. V práci Hora a Conover (1984) je odvozeno limitní rozdělení RT-statistiky při dvojném třídění. Iman, Hora a Conover (1984) srovnávají pomocí simulace sílu klasické analýzy (pro dvojné třídění), Friedmanova testu a RT testu za předpokladu normálního, lognormálního, dvojitě exponenciálního, rovnoměrného a Cauchyho rozdělení. Pro normální rozdělení uvádějí asymptotickou relativní vydatnost (ARE) Friedmanova testu $(3/\pi)k/(k+1)$, kde k je počet výběrů, a na základě provedených studií tvrdí, že pro jiná rozdělení neklesne její hodnota pod $0.864k/(k+1)$. Thompson a Ammann (1989) odvodili ARE pro RT test při dvojném třídění pro normální, logistické, rovnoměrné a dvojitě exponenciální rozdělení. Limitní vlastnosti RT statistiky při vyváženém dvojném třídění jsou studovány v práci Thompson (1991). Asymptotické srovnání síly některých neparametrických metod s klasickými provedl také Akritas (1990) a navrhuje neparametrické metody připouštějící heteroskedasticitu.

8. Síla dvouvýběrových testů shody pro nezávislé výběry stejných rozsahů

8.1. **Výběr testů a alternativ pro simulační studii.** V naší simulační studii jsme se omezili na porovnání několika testů dobré shody pouze pro dva výběry shodných rozsahů 5, 6, 8, 10, 20, 40.

Zkoumali jsme citlivost testu na současné změny střední hodnoty i rozptylu některého z rozdělení. Jeden z výběrů pocházel z rozdělení, jehož střední hodnota byla posunuta o k-násobek směrodatné odchylky postupně pro k = -2, -1, 0, 1, 2 (pro symetrická rozdělení stačí k nezáporná), pak byla směrodatná odchylka postupně násobena čísly 1, 2, 3 při nezměněné střední hodnotě. Uvažovali jsme všechny kombinace popsaných změn, tedy obecně 15 (pro symetrická rozdělení 9) různých variant.

Vyšetřovali jsme sílu následujících testů:

1. W : 2-výběrový Wilcoxonův test (je ekviv. Kruskal-Wallisově testu)
2. ST : test Siegel – Tukey
3. WST : kombinace testů W a ST
4. KS : Kolmogorov – Smirnovův test
5. CM : Cramér von Misesův test
6. MRV : nově navrhovaný pořadový test popsaný níže v odstavci 9
7. t : t-test pro výběry z normál. rozdělení se stejným rozptylem
8. F : F–test shody rozptylů pro dva výběry z normálního rozdělení.

Test Siegel – Tukey se provádí analogicky jako Wilcoxonův test s tím rozdílem, že po seřazení podle velikosti se nahrazují hodnoty přirozenými čísly podle schématu: 1 4 5 8 9 12 13 16 ... 15 14 11 10 7 6 3 2. Lze užít stejných kritických hodnot jako pro Wilcoxonův test. Test je vhodný pro testování změn rozptylu při nezměněné střední hodnotě.

Kombinace Wilcoxonova testu a testu Siegel–Tukey spočívala v provedení obou testů, každého z nich na hladině významnosti přibližně 0.025; tím dosáhneme citlivosti na změny parametrů polohy i měřítka.

Popisu testu MRV (Minimum of Rank Variances) věnujeme odstavec 9.

8.2. **Výsledky pro výběry z normálního rozdělení.** Pro výběry z normálních rozdělení síla parametrických testů dominovala nad sílou neparametrických testů.

Při změnách střední hodnoty zvítězil t-test těsně před Wilcoxonovým testem, s větším odstupem následoval test Cramér von Misesův, s dalším odstupem test Kolmogorov-Smirnovův. Síla testu MRV při těchto alternativách byla ve srovnání s uvedenými testy podstatně nižší.

Při změnách rozptylu bylo prvenství F-testu bezkonkurenční. S velkým odstupem následují neparametrické statistiky, z nichž nejúspěšněji si vedl test MRV.

Při kombinovaných změnách střední hodnoty a rozptylu prvenství parametrickým statistikám zůstalo. Z neparametrických statistik jasně vítězil test MRV s výjimkou alternativy velkého posunu střední hodnoty a „malé“ změny rozptylu (μ=2, σ=2), při které byl s malým náskokem před ostatními test Cramér von Misese, avšak (zejména pro výběry většího rozsahu) i při této alternativě síla testu MRV příliš nezaostávala za sílou testu CM.

Uvedeme výsledky pro výběry rozsahu 10, kdy jsme pro všechny neparametrické testy stanovili (probráním všech možností) exaktní pravděpodnosti α chyb I. druhu pro zvolené kritické hodnoty. První výběr pocházel z rozdělení $N(0,\, 1)$, druhý z rozdělení $N(\mu,\, \sigma^2)$. Výsledky jsou pro 10 000 simulací.

Síla dvouvýběrových testů (n = 10)

	α	μ=0 σ=1	μ=1 σ=1	μ=2 σ=1	μ=0 σ=2	μ=1 σ=2	μ=2 σ=2	μ=0 σ=3	μ=1 σ=3	μ=2 σ=3
1. W	0.0524	0.053	0.554	0.984	0.065	0.276	0.733	0.067	0.171	0.469
2. ST	0.0524	0.055	0.024	0.002	0.301	0.244	0.117	0.585	0.527	0.388
3. WST	0.0464	0.046	0.402	0.959	0.224	0.312	0.661	0.472	0.495	0.593
4. KS	0.0524	0.053	0.474	0.962	0.101	0.317	0.758	0.173	0.323	0.612
5. CM	0.0498	0.046	0.520	0.979	0.081	0.306	0.760	0.134	0.262	0.576
6. MRV	0.0502	0.048	0.279	0.854	0.308	0.410	0.686	0.630	0.675	0.766
7. t	0.0500	0.047	**0.561**	**0.988**	0.054	0.271	**0.766**	0.060	0.174	0.490
8. F	0.0500	0.049	0.049	0.051	**0.497**	**0.494**	0.493	**0.874**	**0.879**	**0.873**

Nejvyšší hodnoty pro alternativy jsou vytištěny tučně.

Pro výběry většího rozsahu (20, resp. 40) zůstalo pořadí úspěšnosti testů zpravidla stejné jako pro rozsah 10. Uvedeme jen výsledky potvrzující převahu nově navrhovaného testu MRV mezi neparametrickými testy při odhalování změny rozptylu, není-li tato změna současně doprovázena velkou změnou střední hodnoty. Výsledky jsou opět pro 10 000 simulací.

Pravděpodobnost chyby I. druhu α uváděné v prvním sloupci jsou exaktní pro testy W, ST a KS, pro které bylo možno najít kritické hodnoty v tabulkách Likeš, Laga (1978). Pro test MRV je exaktní rovněž hodnota pro $n = 20$ (byla spočtena přesná d.f. probráním všech možností), pro $n = 40$ lze použít aproximace (odst. 9). Pro test WST byly kritické hodnoty a příslušná pravděpodobnost α odhadnuty aproximací Wilcoxonovy statistiky normálním rozdělením. Pro oba rozsahy v případě testu CM byly stanoveny kritické hodnoty a odpovídající α z histogramů četností metodou Monte Carlo pomocí 100 000 simulací.

Síla dvouvýběrových testů (n = 20 a n = 40)

	α	μ=0 σ=1	μ=0 σ=2	μ=1 σ=2	μ=2 σ=2	μ=0 σ=3	μ=1 σ=3	μ=2 σ=3
n = 20								
W	0.0491	0.046	0.057	0.480	0.959	0.068	0.288	0.751
ST	0.0491	0.048	0.592	0.489	0.241	0.904	0.867	0.719
WST	0.05	0.047	0.486	0.661	0.960	0.904	0.877	0.951
KS	0.0353	0.035	0.116	0.523	0.956	0.290	0.540	0.893
CM	0.050	0.052	0.134	0.598	**0.975**	0.360	0.601	0.914
MRV	0.0499	0.049	**0.664**	**0.786**	0.961	**0.950**	**0.968**	**0.989**
n = 40								
W	0.0498	0.048	0.054	0.767	0.999	0.067	0.493	0.960
ST	0.0498	0.053	0.890	0.827	0.531	0.998	0.996	0.970
WST	0.05	0.049	0.836	0.947	**1.000**	0.996	0.997	**1.000**
KS	0.0541	0.055	0.345	0.921	**1.000**	0.817	0.954	0.999
CM	0.050	0.050	0.358	0.915	**1.000**	0.864	0.962	0.999
MRV	0.050	0.048	**0.943**	**0.983**	**1.000**	**1.000**	**1.000**	**1.000**

8.3. **Výsledky pro výběry z ostatních rozdělení.** F-test pro shodu dvou rozptylů se ukázal být značně nerobustní vzhledem k odchylkám od normality. Jestliže kritické hodnoty testu odpovídaly hladině významnosti 0.05 pro normální rozdělení, pak již pro logistické rozdělení byla jeho odhadovaná pravděpodobnost chyby I. druhu pro výběry rozsahu 10 rovna 0.089 a pro výběry rozsahu 40 pak 0.112, pro výběry z dvojitého exponenciálního rozdělení byly tyto pravděpodobnosti rovny 0.161 a 0.203 a pro symetrizované Paretovo rozdělení s parametrem 3 dokonce 0.359 a 0.504. Proto jsme F-test do srovnání pro výběry z nenormálních rozdělení nezahrnuli.

Prováděli jsme zpravidla 1000 simulací, více jen pro některá rozdělení. Kritické hodnoty statistik byly voleny stejně jako v odstavci 8.2. Protože získaná data tvoří velmi rozsáhlý materiál, uvedeme pouze „jména vítězů“, a to pro výběry rozsahu 10. V tabulce značí ps posun střední hodnoty alternativy vyjádřený v násobcích směrodatné odchylky, kf je koeficient, kterým byla násobena směrodatná ochylka. Pro úplnost jsme do tabulky zahrnuli i výsledky pro normální rozdělení vyhodnocovaného podle stejných kritérií, tj. tentokráte bez testu F.

Nejúspěšnější dvouvýběrové testy (n = 10)

	ps=1 kf=1	ps=2 kf=1	ps=0 kf=2	ps=1 kf=2	ps=2 kf=2	ps=0 kf=3	ps=1 kf=3	ps=2 kf=3
normální rozdělení	t	t	MRV	MRV	t	MRV	MRV	MRV
směs normálních, d = 5	W	t	MRV	MRV	t	MRV	MRV	MRV
směs normálních, d = 100	KS	t	1)	MRV	t	1)	1)	MRV
logistické r.	W	W	MRV	MRV	CM	MRV	MRV	MRV
rovnoměrné r.	t	t	MRV	MRV	t	MRV	MRV	MRV
exponenciální r.	KS	KS	MRV	MRV	KS	MRV	MRV	MRV
Weibullovo s par. 2	t	t	MRV	MRV	KS	MRV	MRV	MRV
Weibullovo s par. 0.5	KS	2)	MRV	MRV	3)	MRV	MRV	MRV
dvojitě exponenciální r.	CM	CM	MRV	KS	KS	MRV	MRV	MRV
Simpsonovo (trojúhelník.)	t	t	MRV	MRV	t	MRV	MRV	MRV
arcsinové r.	t	t	MRV	MRV	t	MRV	MRV	MRV
mocninné s par. 2	t	t	MRV	MRV	t	MRV	MRV	MRV
mocninné s par. 0.5	t	t	MRV	MRV	KS	MRV	MRV	MRV
Paretovo s par. 3	KS	KS	MRV	MRV	MRV	MRV	MRV	MRV
symetriz. Paret. s par. 3	KS	4)	ST	KS	KS	ST	KS	KS

	ps= -1 kf= 2	ps= -2 kf= 2	ps= -1 kf= 3	ps= -2 kf= 3
exponenciální r.	W	W	MRV	W
Weibullovo s par. 2	MRV	t	MRV	MRV
Weibullovo s par. 0.5	MRV	KS	W	4)
mocninné s par. 2	MRV	MRV	MRV	MRV
mocninné s par. 0.5	MRV	t	MRV	MRV
Paretovo s par. 3	KS	KS	MRV	KS

```
Vysvětlivky:
1) Testy WST a MRV se shodnou úspěšností  100 % .
2) Testy CM  a MRV se shodnou úspěšností  100 % .
3) Testy KS  a MRV se shodnou úspěšností  100 % .
4) Testy KS  a CM  se shodnou úspěšností  100 % .
```

Nejúspěšnější testy pro výběry větších rozsahů bylo obtížnější jednoznačně stanovit, neboť často několik testů dosáhlo úspěšnosti 100 %.

9. Test MRV dobré shody dvou nezávislých náhodných výběrů

V tomto odstavci budeme definovat statistiku MRV (Minimum of Rank Variances) navrženou druhým z autorů a test dobré shody založený na této statistice. Uvažujme dva nezávislé náhodné výběry o rozsahu n_1, n_2 ze spojitých rozdělení s distribučními funkcemi $F_1(x)$, $F_2(x)$. Označme $N = n_1 + n_2$ a buď R_{ij} ($j = 1, \dots, n_i$; $i = 1, 2$) pořadí j-tého pozorování i-tého výběru ve spojeném výběru uspořádaném podle velikosti. Definujme

$$\mathrm{RV}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} R_{ij}^2 - \Big(\frac{1}{n_i} \sum_{j=1}^{n_i} R_{ij}\Big)^2 \quad (i = 1, 2).$$

Elementární, avšak velmi zdlouhavý výpočet dává

$$\text{(1)} \qquad \mathrm{E}\,(\mathrm{RV}_1) = \frac{(n_1 - 1)N(N+1)}{12 n_1},$$

$$\text{(2)} \qquad \mathrm{var}\,(\mathrm{RV}_1) = \frac{(n_1 - 1)n_2 N(N+1)(2Nn_1 + 3N + 3n_1 + 3)}{360 n_1^3},$$

analogicky pro statistiku RV_2, a dále

$$\text{(3)} \qquad cov(\mathrm{RV}_1, \mathrm{RV}_2) = -\frac{(n_1 - 1)(n_2 - 1)N(N+1)(2N+3)}{360 n_1 n_2}.$$

Správnost odvozených vztahů byla potvrzena též numerickými výpočty, výsledky se shodovaly s (1) až (3) na 8 platných míst.

Definujme normalizované hodnoty

$$\mathrm{RV}_i^* = [\mathrm{RV}_i - \mathrm{E}\,(\mathrm{RV}_i)][\mathrm{var}\,(\mathrm{RV}_i)]^{-1/2} \quad (i = 1, 2)$$

a nakonec

$$\mathrm{MRV} = \min\{\, \mathrm{RV}_1^*, \mathrm{RV}_2^* \,\}.$$

Kritickým oborem pro test hypotézy $F_1(x) = F_2(x)$ pro všechna reálná x proti nespecifikované alternativě bude obor $W = \{\, x \,:\, \mathrm{MRV} \le k \,\}$ pro vhodně volené číslo k.

V tomto příspěvku se omezíme na výběry shodných rozsahů $n_1 = n_2 = n$; pak MRV je lineární transformací celočíselné statistiky

$$M_n = n^2 \min\{\, \mathrm{RV}_1, \mathrm{RV}_2 \,\}.$$

Hodnoty distribuční funkce F_n veličiny M_n pro $5 \le n \le 20$ ve vybraných bodech uvedené v tomto odstavci byly spočteny probráním všech možností.

Ukazuje se, že pro $\alpha = 0.05$ nebo $\alpha = 0.01$ lze pro velká n aproximovat $100\,\alpha\,\%$-ní kvantil x_α statistiky MRV kvantilem $u_{\alpha/2}$ normálního rozdělení $N(0,\,1)$, viz níže uvedená tabulka. To je možno heuristicky bez nároků na exaktnost zdůvodnit asymptotickou separátní normalitou statistik RV_1, RV_2 a dále tím, že pro koeficient korelace ϱ mezi veličinami RV_1, RV_2 platí $\varrho \to -1$ pro $n \to \infty$, takže nízké hodnoty RV_1 jsou obvykle doprovázeny vysokými hodnotami RV_2 a naopak. Jevy $(\mathrm{RV}_1^* \le u_{\alpha/2})$, $(\mathrm{RV}_2^* \le u_{\alpha/2})$ mají pro velká n nízkou pravděpodobnost současného výskytu, tedy $\mathrm{P}(\mathrm{MRV} \le u_{\alpha/2}) \approx \mathrm{P}(\mathrm{RV}_1^* \le u_{\alpha/2}) + \mathrm{P}(\mathrm{RV}_2^* \le u_{\alpha/2}) \approx \alpha/2 + \alpha/2 = \alpha$.

n	$\mathrm{P}(\mathrm{MRV} \le u_{0.025})$	$\mathrm{P}(\mathrm{MRV} \le u_{0.005})$
10	0.0460	0.0053
20	0.0488	0.0085
40	0.0493	0.0093

Jak ukazuje tabulka, při použití aproximativních kritických hodnot $\hat{x}_\alpha = u_{\alpha/2}$ je test MRV konzervativní. Pravděpodobnosti uvedené v tabulce byly pro $n = 10$ a $n = 20$ spočteny exaktně, pro $n = 40$ metodou Monte Carlo pomocí 10^6 simulací.

Odvození asymptotického chování včetně případu různých rozsahů výběrů bude věnována připravovaná publikace [23].

Jak ukazují srovnání provedená v odstavci 8, je test MRV vhodný tam, kde si přejeme testovat shodu mezi dvěma výběry z obecně nenormálního spojitého rozdělení, přičemž lze očekávat, že teoretická rozdělení by se mohla lišit od sebe parametrem polohy nebo měřítka nebo oběma parametry.

Vybrané hodnoty d.f. $F_n(x)=P(M_n \le x)$

n	x	$F_n(x)$	x	$F_n(x)$	x	$F_n(x)$	x	$F_n(x)$
5			49	0	73	0.039683	73	0.039683
			50	0.039683	74	0.111111	74	0.111111
6			104	0	160	0.036797	184	0.086580
			105	0.012987	161	0.062771	185	0.103896
7	195	0	243	0.004079	333	0.044289	387	0.092075
	196	0.004079	244	0.011655	334	0.050699	388	0.104895
8	335	0	475	0.009635	630	0.048329	719	0.097591
	336	0.001243	476	0.011500	631	0.050039	720	0.101787
9	619	0.000370	847	0.009502	1087	0.049568	1225	0.099260
	620	0.001070	848	0.010119	1088	0.050638	1226	0.100658
10	1055	0.000801	1400	0.009504	1760	0.048691	1968	0.099483
	1056	0.001018	1401	0.010024	1761	0.050185	1969	0.100998
11	1693	0.000987	2195	0.009954	2707	0.049998	2999	0.099961
12	2570	0.000965	3274	0.009915	3991	0.049679	4379	0.099763
13	3757	0.000996	4717	0.009894	5681	0.049814	6209	0.099746
14	5311	0.000999	6600	0.009998	7863	0.049906	8551	0.099757
15	7315	0.000992	8999	0.009990	10619	0.049972	11495	0.099857
16	9855	0.000997	12014	0.009995	14047	0.049961	15150	0.099986
17	13019	0.000998	15711	0.009976	18255	0.049998	19609	0.099955
18	16888	0.000998	20224	0.009973	23335	0.049981	24988	0.099897
19	21567	0.000999	25643	0.009989	29413	0.049965	31409	0.099977
20	27175	0.000999	32090	0.009988	36610	0.049902	39003	0.099922

Literatura

[1] Akritas M. G. (1990), The rank transform method in some two-factor designs, *Journal of the American Statistical Association* **85**, 73 – 78

[2] Anděl J. (1978), *Matematická statistika*. SNTL/ALFA, Praha

[3] Antoch J. (1990), *Několik poznámek o generování pseudonáhodných čísel z* $R(0, 1)$. ROBUST'90, JČSMF, Praha, 14 – 36.

[4] Box G. E. P. a Watson G. S. (1962), *Robustness to Non-Normality of Regression Tests*. Biometrika **4**, 93 – 106.

[5] Conover W. J. a Iman R. L. (1981), *Rank transformations as a bridge between parametric and nonparametric statistics*. The American Statistician **35**, 124 – 129.

[6] David F. N. a Johnson N. L. (1951), *The effect of non-normality on the power function of the F-test in the analysis of variance*. Biometrika, **38**, 43 – 57.

[7] Donaldson T. S. (1968), *Robustness of the F-test to errors of both kinds and the correlation between the numerator and denominator of the F-ratio*. Journal of the American Statistical Association **63**, 660 – 676.

[8] Fraser D. A. S. (1957), *Nonparametric Methods in Statistics*, John Wiley, New York.

[9] Gayen A. K. (1950), *The distribution of the variance ratio in random samples of any size from non-normal universes*. Biometrika **37**, 236 – 255.

[10] Gibbons J. D. (1971), *Nonparametric Statistical Inference*. McGraw-Hill, New York.

[11] Hora S. C. a Conover W. J. (1984), *The F-statistics in the two-way layout with rank-score transformed data*. Journal of the American Statistical Association **79**, 668 – 673.

[12] Iman R. L. (1974), *A power study of a rank transform for the two-way classification model when interaction may be present*. The Canadian Journal of Statistics **C 2**, 227 – 239.

[13] Iman R. L., Hora S. C. a Conover W. J. (1984), *Comparison of asymptotically distribution-free procedures for the analysis of complete blocks*. Journal of the American Statistical Association, **79**, 674 – 685.

[14] Kempthorne O. (1952), *The Design and Analysis of Experiments*. John Wiley Sons, New York.

[15] Kruskal W. H. a Wallis W. A. (1952), *Use of ranks in one-criterion variance analysis*. Journal of the American Statistical Association **47**, 583 – 621.

[16] Likeš J. a Laga J. (1978), *Základní statistické tabulky*. SNTL, Praha.

[17] Mehra K. L. a Sarangi J. (1967), *Asymptotic efficiency of certain rank tests for comparative experiments*. Annals of Math. Stat. **38**, 90 – 107.

[18] Miller R. G. Jr. (1981), *Simultaneous Statistical Inference 2nd ed.*. Springer-Verlag, Berlin, Heidelberg.

[19] O'Gorman T. W. (1997), *An adaptive test for the one-way layout*. The Canadian J. of Statistics **25**, 269 – 279.

[20] Pearson E. S. (1931), *The analysis of variance in cases of nonnormal variations*. Biometrika **20**, 114 – 133.

[21] Pitman E. J. G. (1937a), *Significance tests which may be applied to samples from any population*. JRSS Suppl., **4**, 119 – 130.

[22] Pitman E. J. G. (1937b), *Significance tests which may be applied to samples from any population. III. The analysis of variance test*. Biometrika **29**, 322 – 335.

[23] Reif J. (v přípravě), On a rank statistics for two-sample location-scale problem.

[24] Scheffé H. (1959), *The Analysis of Variance*. John Wiley, New York.

[25] Siegel S. (1956), *Nonparametric Statistics for the Behavioral Sciences*. McGraw-Hill Book Company, New York.

[26] Srivastava A. B. L. (1959), *Effect of non-normality on the power of the analysis of variance test*. Biometrika **46**, 114 – 122.

[27] Thompson G. L (1991), *A note on the rank transform for interactions*. Biometrika **78**, 697 – 701.

[28] Thompson G. L. a Ammann L. P. (1989), *Efficacies of rank-transform statistics in two-way models with no interaction*. Journal of the American Statistical Association **84**, 325 – 330.

[29] Tiku M.L. (1964), *Approximating the general nonnormal variance ratio sampling distributions*. Biometrika **51**, 83 – 95.

[30] Tiku M.L. (1971), *Power function of the f-test under non-normal situations*. Journal of the American Statistical Association **65**, 913 – 916.

[31] Welch B. L. (1937), *On the Z-test in randomized blocks and latin squares*. Biometrika **29**, 21 – 52.