

# INFORMAČNÍ BULLETIN



České statistické společnosti

Ročník 23, číslo 1, březen 2012

# ZPRÁVA O ČINNOSTI ČESKÉ STATISTICKÉ SPOLEČNOSTI V ROCE 2011

**Která byla přednesená a projednaná na valné hromadě společnosti dne 3. 2. 2012.**

## 1. Základní údaje o společnosti

Uplynulý rok byl prvním rokem dvouletého funkčního období výboru České statistické společnosti, který byl zvolen na valné hromadě dne 7. 2. 2011. Předsedou byl doc. RNDr. Gejza Dohnal, CSc. (FS ČVUT v Praze), funkci prvního místopředsedy vykonávala prof. Ing. Hana Řezanková, CSc. (VŠE) a hospodářem byl Ing. Tomáš Löster (VŠE Praha).

K dnešnímu dni má ČStS 229 členů. Za uplynulý rok vstoupilo do společnosti 8 členů. V roce 2011 zemřeli 2 členové společnosti, 1 ukončil členství na vlastní žádost. V třinácti případech bylo členství ukončeno pro neplacení členských příspěvků. V zahraničí žije celkem 11 členů, z toho 6 na Slovensku (4 jsou studenti, kteří studují zde, ale mají trvalé bydliště na Slovensku).

## 2. Činnost výboru společnosti

V průběhu roku se konala dvě zasedání výboru České statistické společnosti. Kromě toho byla diskutována řada důležitých záležitostí prostřednictvím elektronické pošty a hlasování. V průběhu roku navíc proběhla řada neformálních setkání a porad při jednotlivých akcích.

Z nejvýznamnějších rozhodnutí výboru společnosti v uplynulém roce bylo připojení naší společnosti do Federace evropských národních statistických společností (FENSTATS), na jejímž založení se shodlo 14 zástupců národních statistických společností na podzim 2010 v Paříži a jejíž statut byl diskutován v průběhu roku 2011. Hlavní zásluhu na jeho přípravě má Maurizio Vichi, předseda italské statistické společnosti. Na 58. světovém statistickém kongresu ISI v Dublinu v srpnu 2011 byla připravena konečná verze a v prosinci 2011 byla tato Federace ustavena.

Z dalších mezinárodních aktivit jmenujme schůzku skupiny V6 v maďarském Visegrádu. Tentokrát byla spojena s konferencí organizovanou maďarským statistickým úřadem a Eurostatem při příležitosti ukončení maďarského předsednictví Evropské Unii. Jako hlavní téma bylo zvýšení reputace statistiky ve společnosti. Na zasedání skupiny V6, kterého se zúčastnila i ředitelka ISI paní Ada Van Krimpen, byla sepsána deklarace za zvýšení nezávislosti statistiky na politických a komerčních vlivech.

Mezi nejdůležitější domácí události, na jejichž organizaci se výbor společnosti účastnil, patří konference Stakan, společná akce naší společnosti a Slovenské štatistickej a demografickej spoločnosti. Tato konference proběhla na přelomu září a října (28. 9. – 2. 10.) v Železné Rudě na Šumavě a byla koncipována jako dvoukonference spolu s konferencí T<sub>E</sub>Xperience ve spolupráci se společností  $\zeta$ TUG a vydatnou zásluhou Pavla Stríže (a jeho rodiny). Toto spojení se ukázalo jako velmi přínosné.

Neúspěch jsme zaznamenali s přihláškou našeho časopisu do databáze Scopus, kde jsme byli odmítnuti s možností opakované žádosti až za tři roky.

Na adresu společnosti na MFF UK, Sokolovská 83, docházejí pravidelně dva zahraniční časopisy: Austrian Statistical Journal a Manchester Statistical Society Transactions.

### 3. Odborná aktivita společnosti

- Dne 7. 2. 2011 se konala v budově VŠE v Praze valná hromada společnosti, na které byl zvolen předseda a výbor společnosti. Na valné hromadě přednesl odbornou přednášku místopředseda ČSÚ Ing. Stanislav Drápal na téma Sčítání lidí, domů a bytů v roce 2011.
- Ve dnech 28. 9. – 2. 10. 2011 se konala dvoukonference STAKAN 2011 a T<sub>E</sub>Xperience pořádaná společně naší společností, Slovenskou štatistickou a demografickou spoločnosťou, sdružením  $\zeta$ TUG a Univerzitou Tomáše Bati ve Zlíně. Konference se konala v Železné Rudě na Šumavě s výletem na bavorský Velký Javor (Grosser Arber).
- Konference REQUEST se konala 14. – 15. 12. 2011 na strojní fakultě ČVUT v Praze Dejvicích. Příspěvky z této konference vyjdou v Informačním Bulletinu.
- Čtvrtý Mikuklášský statistický den ČStS zorganizovala dne 6. 12. 2011 v respiriu MFF UK v pražském Karlíně. Přišel Mikukláš, přinesl dárky a celodenní odborný program ukončila skupina FAB, s. r. o.
- Internetové stránky společnosti byly pravidelně udržovány a aktualizovány díky práci kolegy doc. Jiřího Žváčka. Bohužel, zatím nedošlo k původně plánované změně grafické úpravy těchto stránek.
- V roce 2011 vyšla dvě čísla Informačního Bulletinu, další dvě jsou připravena k výrobě a vyjdou v tomto roce (pravděpodobně jako dvoučíslo).
- ČStS formálně spolupracovala na vydávání časopisu Statistika.

## 4. Plán aktivit pro rok 2012

- Na červen 2012 se plánuje a připravuje statistický den na zámku v Nových Hradech.
- 9. – 14. září 2012 se bude konat konference ROBUST 2012 na Moravě v Němčičkách nedaleko Velkých Pavlovic v okrese Břeclav.
- Na podzim tohoto roku se bude konat další setkání skupiny V6, tentokrát v Bratislavě.
- V rámci možností se budeme podílet na organizaci statistických konferencí u nás i v zahraničí.
- Mikuklášský den bude v prosinci v Praze.

V Praze, dne 1. 2. 2011  
Doc. RNDr. Gejza Dohnal, CSc.  
předseda společnosti

## Validity Estimation in Questionnaires

### Validita dotazníku

### Zdeněk Půlpán

*Adresa:* Přírodovědecká fakulta, Univerzita Hradec Králové  
Rokitanského 62, 500 03 Hradec Králové 3

*E-mail:* Zdenek.Pulpan@uhk.cz

**Abstract:** Validity is the attribute of questionnaire. This article presents the one from many eventualities of the validity estimation.

**Keywords:** Questionnaires, Validity, Validity Estimation.

**Abstrakt:** Článek představuje jednu z mnoha eventualit ověření validity dotazníku použitého v dotazníkových šetřeních.

**Klíčová slova:** Dotazníková šetření, ověření validity.

## 1. Úvod

Získáváme-li informaci ze speciálně konstruovaného dotazníku, je dotazník měřicím prostředkem a spolu s podmínkami a metodou vyhodnocení je i normou, ke které se všechny výroky o měření vztahují. Možnosti a rozsah interpretace měření pomocí dotazníku za určitých, přesněji vymezených okolností, jsou dány jeho validitou.

**Validita je mírou kvality vztahu mezi dotazníkem, zkoumanou populací, podmínkami a záměrem zkoumání.** Dotazník je tím více validní, čím lépe za daných okolností měří jen to, co měřit má a ne něco jiného ([1], str. 47, [5], str. 21).

Jsou pokusy odhadnout validitu dotazníku jediným číselným ukazatelem získaným převážně empiricky, např. statisticky. Mohou k tomu sloužit jak výsledky dotazníku ve výběru zkoumané populace nebo data z expertních posudků (zvláště u kriteriální nebo predikční validity). Ukazuje se však, že spolehlivější cesta k odhadu validity dotazníku vede k jeho dokonalé a všestranné analýze (zvláště u obsahové, resp. pojmové validity). Apriorní kvalitativní analýza neumožňuje sestavit požadovaný kvantitativní ukazatel. Vadou empirických metod, které však často vedou k požadovanému kvantitativnímu ukazateli validity, je jejich určitá jednostrannost (ukazatel hodnotí jen určitou stránku jevu) a závislost na experimentálním souboru, na kterém se validita ověřuje.

Zde navrhuje jednu kompromisní, kvalitativně-quantitativní metodu, odvozenou z Hellingerovy divergence dvou statistických rozdělení. ([2], str. 35, vztah (56))

## 2. Metody posuzování validity dotazníku

*Kvalitativní posouzení validity* předpokládá definování atributů prověřované vlastnosti a odhad jejich účinku na výsledek dotazníku (který je reprezentován buď celkovým hrubým skóre nebo jen rozložením četností na jednotlivých položkách) v reprezentativní populaci. Validitu pak odhadujeme jako míru předpokládaného vlivu prověřované vlastnosti (reprezentované atributy) na výsledek dotazníku v kategoriích např. nízká, střední, vysoká.

Při odhadu *obsahové validity* jsou atributy základní prvky obsahu. V dotazníkových položkách musí být v určitém poměru zastoupeny podstatné prvky sémantické struktury (základní pojmy a vztahy mezi nimi). K tomu se užívají různé techniky zobrazovací prvků sémantické struktury (např. orientované multigrafy, algebraická teorie relací, ...).

Při odhadu *kriteriální validity* může být dán vzorový, dostatečně validní dotazník podle kterého se validita námi konstruovaného dotazníku posuzuje nejprve hledáním a pak porovnáváním odpovídajících položek. Validita takto zjišťovaná je úsudkem o míře podobnosti (s hlediska jistého systému atributů) dotazníku tvořeného nebo ověřovaného s dotazníkem referenčním, dostatečně validním. Jindy může být kritériem soubor jistých požadavků, jejichž splnění má daný dotazník ověřit.

Požadujeme-li, aby dotazník správně identifikoval zkoumanou vlastnost v souboru těch, kteří ji mít mají (tj., aby dotazník byl dostatečně *senzibilní*) a aby také identifikoval správně ty, kteří zkoumanou vlastnost mít nemají, hledá se taková *volba skórování dotazníku*, která obě kategorie respondentů od sebe oddělí.

Můžeme také zkoumat tzv. podobnost resp. nepodobnost dvou dotazníků z určitých hledisek; pak mluvíme o *konvergentní validitě*, která odhaduje vlastně *senzitivitu* nebo o *divergentní validitě*, která odhaduje *specificitu* dotazníku.

Některé aspekty validity lze však také posuzovat i *kvantitativně* ([1], str. 47). Validitu obsahovou lze zjišťovat pouze z expertních posudků, hodnotících reprezentativnosti souboru položek dotazníku. Předmětem statistického zkoumání nebo fuzzy přístupu jsou pak expertní posudky. Pouze z expertních odhadů lze vycházet u *konstruktové* neboli *teoretické validity*. Experti posuzují jak dotazník odráží jisté atributy (které mohou být např. psychologickými charakteristikami jako je úzkost, strach, labilita, ...). Kvantitativně lze odhadnout míru shody (nebo neshody) jejich rozhodnutí a z toho usuzovat na stupeň jednoznačnosti expertního posouzení.

Metodami regresní analýzy lze zkoumat tzv. *predikční validitu*. Predikční validitou se rozumí schopnost dotazníku jisté předpovědi směrem do budoucnosti. ([6], [7]) Při aplikaci dvou rozdílných dotazníků (v téže populaci a za stejných podmínek), z nichž jeden je dostatečně validní, lze validitu druhého odhadovat například ze vzájemné korelace výsledků (posuzuje se tak jeden aspekt konvergentní validity).

### 3. Odhad ukazatele validity

Validita dotazníku musí být vztažena k určité úžeji vymezené vlastnosti nebo jevu (týkajícího se například sémantiky položek vzhledem k určité normě, blízkosti či odlišnosti obsahu, ...). Deklarovaný jev nebo vlastnost podmiňuje účinek, který má dotazník detekovat. Validita, vymezená vzhledem k účinku  $M$ , musí být vymezena i vzhledem k absenci účinku („ne  $M$ “), a to tak, že ho v tom případě nedetekuje. Účinek  $M$  nechť je reprezentován

souborem  $M_r$  všech jeho atributů, které jsme schopni formulovat:

$$M_r = \{m_1, m_2, \dots, m_r\}. \quad (1)$$

Ze zkušenosti víme, že v mnoha případech je možné uvažovaný účinek reprezentovat fuzzy množinou  $\underline{M}$  na  $M_r$ :

$$\underline{M} = \{m_1/\mu_M(m_1), m_2/\mu_M(m_2), \dots, m_r/\mu_M(m_r)\}. \quad (2)$$

Hodnoty měr věrohodnosti  $\mu_M(m_i)$ ,  $i = 1, 2, \dots, r$ , závisí na tom, jaký význam se příslušnému atributu přiřazuje,

$$0 \leq \mu_M(m_i) \leq 1, \quad i = 1, 2, \dots, r \quad (3)$$

(čím vyšší hodnota  $\mu_M$ , tím větší význam atributu).

Uvažujme, že dotazník se skládá ze dvou druhů položek: těch, které jsou schopny detekovat vliv aspoň jednoho atributu z  $M_r$ , a pak i z těch, které žádný vliv některého z atributů nedetekují.

Zastoupení atributů z  $M_r$  v jednotlivých dotazníkových položkách je možné expertně odhadovat. Každé položce přiřadíme součet měr věrohodnosti atributů, o nichž se domníváme, že jsou položkou při svém působení detekovány. Označíme-li pro  $i$ -tou položku uvedený součet  $s_i$ , můžeme pak určit čísla  $p_i$  podle (4):

$$p_i = \frac{s_i}{\sum_{j=1}^n s_j}, \quad i = 1, 2, \dots, k. \quad (4)$$

Z experimentu za působení nebo nepůsobení účinku  $M$  získáme pro každou z  $k$  položek dotazníku experimentální četnost  $n_i$  pozitivních reakcí. Z experimentálních četností  $n_i$  pak určíme normované experimentální hodnoty  $q_i$  podle (5):

$$q_i = \frac{n_i}{\sum_{j=1}^k n_j}, \quad i = 1, 2, \dots, k. \quad (5)$$

Získáme tak dvě fiktivní rozdělení  $P = \{p_i\}_{i=1, \dots, k}$ ,  $Q = \{q_i\}_{i=1, \dots, k}$ . Má-li účinek  $M$  vliv na výsledek zadání dotazníku v reprezentativní populaci, pak obě rozdělení musí být podobná. Podobnost rozdělení  $P, Q$  ohodnotíme Hellingerovou divergencí  $D_H(P, Q)$ :

$$0 \leq D_H(P, Q) = 2 \left( 1 - \sum_{i=1}^k \sqrt{p_i \cdot q_i} \right) \leq 2. \quad (6)$$

Protože tato divergence je shora omezená ([2], str. 35, vztah (56)), můžeme za odhad validity dotazníku považovat číslo

$$v = \sum_{i=1}^k \sqrt{p_i \cdot q_i}, \quad 0 \leq v \leq 1 \quad (7)$$

Hodnoty ukazatele  $v$ , blízké 1, svědčí o vysoké validitě dotazníku.

Validitu dotazníku musíme následně také posoudit i za nepřítomnosti účinku  $M$ . V tomto případě by žádný atribut z  $M_r$  neměl mít na reakce subjektů vliv, proto fiktivní rozložení  $P$  by mělo být rovnoměrné, tedy  $p_i = \frac{1}{k}$  pro  $i = 1, 2, \dots, k$ . Dotazník zadáme reprezentativnímu vzorku respondentů, na které nepůsobí účinek  $M$ . Podobně stanovíme z četností reakcí na položky nové fiktivní rozdělení  $\underline{Q} = \{q_i\}_{i=1,2,\dots,k}$  a vypočteme nový ukazatel  $w$  podle vztahu (8)

$$w = \sum_{i=1}^k \sqrt{\frac{1}{k} \cdot q_i}. \quad (8)$$

Validitou dotazníku pak budeme rozumět ukazatel

$$V = \sqrt{v \cdot w}, \quad V \in \langle 0; 1 \rangle \quad (9)$$

Čím je hodnota ukazatele  $V$  větší, tím lepší je validita dotazníku (vzhledem k okolnostem jeho konstrukce).

*Příklad:* Následující (zkrácený) dotazník má odhadovat velikost únavy. Byly proto formulovány následující atributy únavy ([1], str. 109):

$m_1$	..... ospalost	.....	$\mu(m_1) = 0,8$
$m_2$	..... nechuť k přemýšlení	.....	$\mu(m_2) = 1,0$
$m_3$	..... pocit těžkých rukou nebo nohou	.....	$\mu(m_3) = 0,7$

Dotazník:

1. Odmítáte komunikovat? (1,6)
2. Cítíte se ospalý? (1,2)
3. Máte pocit otupělosti? (2,0)
4. Nemůžete se soustředit? (1,8)
5. Bolí vás oči? (0,2)
6. Bolí vás v kříži? (0,2)
7. Máte pocit svědění zad? (0,0)

Ospalost může být důvodem nechuti komunikovat, není to však důvod jediný. Míra vlivu ospalosti na kladnou odpověď byla expertem ohodnocena



skórem  $0,5 \cdot 0,8$ ; nechuť k přemýšlení je významnou podmínkou nechuti ke komunikaci, expert to ohodnotil skórem 1; pocit těžkých rukou na nechuti komunikovat je většinou dost malý, expert to ohodnotil skórem  $0,2 \cdot 1$ . Celkem tedy první položka dotazníku získala skór:

$$0,5 \cdot 0,8 + 1 + 0,2 \cdot 1 = 1,6.$$

Podobně bylo postupováno u dalších dotazníkových položek. (Výsledky jsou uvedeny v závorkách u každé položky.) Z takto získaných dat byly určeny hodnoty  $p_i$ ,  $i = 1, 2, \dots, 7$ :  $p_1 = \frac{1,6}{7} = 0,23$ ;  $p_2 = 0,17$ ,  $p_3 = 0,29$ ,  $p_4 = 0,26$ ,  $p_5 = 0,03$ ,  $p_6 = 0,03$ ,  $p_7 = 0$ .

Experiment v reprezentativní populaci zřejmě unavených jedinců vedl k následujícím četnostem reakcí, viz tabulka 1.

Tabulka 1: Četnosti reakcí zřejmě unavených jedinců

Položka – $i$	1	2	3	4	5	6	7
Četnost	32	26	20	30	10	8	3
$q_i$	0,25	0,20	0,16	0,23	0,08	0,06	0,02

Výpočet  $v$  podle (7):

$$\begin{aligned} v &= \sqrt{0,23 \cdot 0,25} + \sqrt{0,17 \cdot 0,20} + \sqrt{0,29 \cdot 0,16} + \sqrt{0,26 \cdot 0,23} + \\ &+ \sqrt{0,03 \cdot 0,08} + \sqrt{0,03 \cdot 0,06} + \sqrt{0 \cdot 0,02} = \\ &= 0,98 \end{aligned}$$

Pak byl experiment opakován ve skupině neunavených jedinců s následujícím výsledkem, viz tabulka 2.

Tabulka 2: Četnosti reakcí zřejmě neunavených jedinců

Položka – $i$	1	2	3	4	5	6	7
Četnost	13	16	11	18	10	11	9
$q_i$	0,15	0,18	0,13	0,20	0,11	0,13	0,10

Výpočtem podle vztahu (8):

$$\begin{aligned} w &= \sqrt{\frac{1}{7} \cdot 0,15} + \sqrt{\frac{1}{7} \cdot 0,18} + \sqrt{\frac{1}{7} \cdot 0,13} + \sqrt{\frac{1}{7} \cdot 0,20} + \\ &+ \sqrt{\frac{1}{7} \cdot 0,11} + \sqrt{\frac{1}{7} \cdot 0,13} + \sqrt{\frac{1}{7} \cdot 0,10} = \\ &= 0,99 \end{aligned}$$

jsme získali rovněž vysokou hodnotu ukazatele  $w$ . Za odhad validity uvedeního dotazníku budeme však považovat hodnotu

$$V = \sqrt{0,98 \cdot 0,99} = 0,98. \quad \blacksquare$$

## 4. Závěr

Dotazník, který není dostatečně validní se k měření nehodí. Špatnou validitu dotazníku nezlepšíme statistickými prostředky (na rozdíl např. od reliability), ale jen jeho podstatnou úpravou. Výběr položek dotazníku musí být takový, aby při působení atributů se rozdělení výsledků podstatně odlišovalo od toho, které odhadneme při absenci jejich působení. I když originální metodika vede ke kvantitativním údajům, doporučuje se odhadovat validitu jen v několika úrovních, např. nízká, střední, vysoká. Tyto úrovně je možné modelovat pro každý dotazník fuzzy množinami na škále hodnot  $V$ .

## Literatura

- [1] Půlpán, Z.: *K problematice zpracování empirických šetření v humanitních vědách*, Academia, Praha 2004. ISBN 80-200-1221-4.
- [2] Půlpán, Z.: *Ztráty informace v důsledku restrikce měřicí škály*, UPOL, Olomouc, 2006. ISBN 80-244-1504-6.
- [3] Lord, F.M.: *Application of Item Response Theory to Practical Testing Problems*, Hillsdale, 1980. ISBN 0-89859-006-X.
- [4] Longford, N. T.: *Models for Uncertainty in Educational Testing*, Springer-Verlag, New York, Inc., 1995. ISBN 978-0387945132.
- [5] Komenda, S., Mazuchová, J.: *Pravděpodobnostní rozdělení entropie (nit), Tvorba a testování testu*, UPOL, Olomouc 1995.
- [6] Zvára, K.: *Biostatistika*, Karolinum, Praha, 2002. ISBN 978-80-246-0739-9.
- [7] Rubešová (Forstová), J.: *Statistické metody pro hodnocení predikční validity*, disertace PŘF UK Praha, 2009.

# THE EIGENVALUES SPACING DISTRIBUTION OF THE TWO-BY-TWO HERMITIAN RANDOM MATRICES OVER CAYLEY-DICKSON ALGEBRAS

## ROZDĚLENÍ VZDÁLENOSTÍ VLASTNÍCH ČÍSEL HERMITOVSKÝCH NÁHODNÝCH MATIC TYPU $2 \times 2$ NAD CAYLEY-DICKSONOVÝMI ALGEBRAMI

Martin Veselý\*, Oxana Gerasimchuk\*\*

*Adresa:* ČVUT, FJFI, KSE, Trojanova 13, 120 00 Praha 2

*E-mail:* veselm21@fjfi.cvut.cz\*, gerasoxa@fjfi.cvut.cz\*\*

**Abstract:** In this report, we investigate the construction of the random matrices with the eigenvalues spacing distribution described by the Wigner's surmise with  $\beta = 2^k$ ,  $k \in \mathbb{N}_0$ , exactly. The hermitian matrices over the arbitrary Cayley-Dickson algebra are established. Properties of the left real eigenvalues of them are investigated. The spacing distribution of these eigenvalues is derived. We have shown that the dimension of the Cayley-Dickson algebra is equal to the spectral repulsion parameter  $\beta$  from the Wigner's surmise. The theoretical results are verified by the numerical simulation and  $\chi^2$  goodness of fit test. In addition, we bring the formula for transformation of the standard normal distribution to the Wigner's one.

**Keywords:** Random Matrix, Hermitian Matrix, Wigner's Surmise, Wigner's Distribution, Cayley-Dickson Construction, Cayley-Dickson Algebras.

**Abstrakt:** Tento článek se zabývá konstrukcí náhodných matic, jejichž vzdálenost vlastních čísel je exaktně popsána Wignerovou domněnkou s  $\beta$  tvaru  $2^k$ ,  $k \in \mathbb{N}_0$ . Jsou zavedeny hermitovské matice nad libovolnou Cayley-Dicksonovou algebrou. Dále jsou studovány vlastnosti reálných levých vlastních čísel těchto matic. Následně je odvozeno rozdělení vzdáleností zmíněných vlastních čísel a je ukázáno, že dimenze Cayley-Dicksonovy algebry, z níž pocházejí prvky matice, je totožná s parametrem repulze  $\beta$  figurujícím ve Wignerově domněnce. Teoretické závěry jsou následně numericky ověřeny pomocí  $\chi^2$  testu dobré shody. Vedlejším produktem práce je předpis transformace Gaussova rozdělení na Wignerovo.

**Klíčová slova:** Náhodná matice, Hermitovská matice, Wignerova domněnka, Wignerovo rozdělení, Cayley-Dicksonova konstrukce, Cayley-Dicksonovy algebry.

## 1. Úvod

Rozdělení vzdálenosti uspořádaných vlastních čísel náhodných matic typu GOE, GUE a GSE<sup>1</sup> bylo velmi detailně studováno v [1]. Dále v pracích [3] a [4] je zmíněné rozdělení rozebíráno v souvislosti s malými maticemi (řádu 2 a 4). Navíc v [4] jsou studovány vlastnosti spekter matic tvořených oktoniony. Ve všech zmíněných pracech je hustota pravděpodobnosti tohoto rozdělení popsána Wignerovou domněnkou

$$f(r) = Ar^\beta e^{-Br^2}, \quad (1)$$

kde  $A$  a  $B$  jsou normalizační konstanty zajišťující, že vztah je hustota pravděpodobnosti a střední hodnota rozdělení je jednotková.  $\beta$  je tzv. parametr spektrální repulze popisující spektrální vlastnosti dané matice. Pro matice tříd GOE, GUE, resp. GSE nabývá  $\beta$  hodnot 1, 2, resp. 4.

V tomto článku navážeme na práce J. M. Nieminena a rozšíříme výše citované poznatky pro hermitovské matice řádu 2 tvořené prvky z libovolné Cayley-Dicksonovy algebry. Dále odvodíme předpis generátoru Wignerova rozdělení s libovolným kladným celočíselným parametrem  $\beta$ .

## 2. Cayley-Dicksonovy algebry

Je známo, že komplexní čísla jsou dvourozměrným rozšířením čísel reálných. Komplexní čísla je možné dále rozšířit pomocí tzv. Cayley-Dicksonovy konstrukce<sup>2</sup>. Jejím výsledkem jsou algebry tzv. hyperkomplexních čísel tvaru

$$a = \sum_{i=1}^n a_i \alpha_i, \quad (2)$$

kde  $n$  je mocnina čísla dvě a  $a_i \in \mathbb{R}$ . Prvky  $\alpha_i$  nazýváme základní jednotky (např. algebra komplexních čísel obsahuje základní jednotky 1 a  $i$ ). V dalším textu budeme obecnou Cayley-Dicksonovu algebru dimenze  $n$  značit symbolem  $\mathbb{CD}_n$ . Poznamenáváme, že pro  $n = 1, 2, 4, 8, 16$  získáváme po řadě reálná čísla ( $\mathbb{R}$ ), komplexní čísla ( $\mathbb{C}$ ), kvaterniony ( $\mathbb{H}$ ), oktoniony ( $\mathbb{O}$ ) a sedeniony ( $\mathbb{S}$ ). Každý prvek Cayley-Dicksonovy algebry dimenze  $n$  lze psát ve tvaru

$$a = a' + a''\mathbf{e}, \quad (3)$$

kde  $a', a'' \in \mathbb{CD}_{\frac{n}{2}}$  a  $\mathbf{e}$  je další základní jednotka, která se v algebře nižší dimenze nevyskytuje. Sčítání prvků libovolné  $\mathbb{CD}_n$  je definováno po složkách.

---

<sup>1</sup>Základní definice a vlastnosti těchto matic lze nalézt v [7].

<sup>2</sup>Více o Cayley-Dicksonově konstrukci lze nalézt v [2].

Pro všechny prvky  $\mathbb{CD}_n$  definujeme tzv. konjugovaný prvek předpisem

$$\bar{a} = \overline{a'} - a'' \mathbf{e}, \quad (4)$$

kde  $a', b'' \in \mathbb{CD}_{\frac{n}{2}}$ . Pro reálná čísla platí  $\bar{a} = a$ . Dále  $\forall a, b \in \mathbb{CD}_n$  definujeme za pomoci (3) a (4) násobení

$$ab = (a'b' - \overline{b''a''}) + (b''a' + a''\overline{b'})\mathbf{e}, \quad (5)$$

kde opět  $a', a'', b', b'' \in \mathbb{CD}_{\frac{n}{2}}$ . Násobení s rostoucí dimenzí algebry ztrácí „rozumné“ vlastnosti. Např. pro kvaterniony již není komutativní, pro oktoniony není asociativní, ale pouze alternativní, tj.  $(xy)y = x(yy) \forall x, y \in \mathbb{CD}_n$ , a obecně je pouze mocninně alternativní, tj.  $(xx)x = x(xx) \forall x \in \mathbb{CD}_n$ .

Dále  $\forall a \in \mathbb{CD}_n$  platí

$$a \bar{a} = \bar{a}a = \sum_{i=1}^n a_i^2 \in \mathbb{R}. \quad (6)$$

Toto lze dokázat matematickou indukcí. Pro  $n = 1$  tvrzení zřejmě platí, nechť tedy platí pro algebru o dimenzi  $n$ . Uvažme  $a \in \mathbb{CD}_{2n}$ , pak z (3), (5) a indukčního předpokladu plyne<sup>3</sup>

$$a \bar{a} = a' \overline{a'} - \overline{(-a'')} a'' + (-a'' a' + a'' a') \mathbf{e} = a' \overline{a'} + \overline{a''} a'' = \sum_{i=1}^n a_i^2 + \sum_{i=n+1}^{2n} a_i^2 \quad (7)$$

$$\bar{a}a = \overline{a'} a' - \overline{a''} (-a'') + (a'' \overline{a'} - a'' \overline{a'}) \mathbf{e} = a' \overline{a'} + \overline{a''} a'' = \sum_{i=1}^n a_i^2 + \sum_{i=n+1}^{2n} a_i^2, \quad (8)$$

což dává požadované tvrzení. Na základě této věty můžeme pro každou Cayley-Dicksonovu algebru definovat normu<sup>4</sup> jejich elementů vztahem

$$\|a\| = \sqrt{a \bar{a}} = \sqrt{\sum_{i=1}^n a_i^2}, \quad (9)$$

kde  $a \in \mathbb{CD}_n$ . Norma je shodná s euklidovskou normou na prostoru  $\mathbb{R}^n$ . Důvodem je totiž fakt, že každý prvek Cayley-Dicksonovy algebry lze reprezentovat pomocí vektoru z  $\mathbb{R}^n$ . Avšak obě algebraické struktury nelze vždy zaměnit, neboť na euklidovských prostorech není definováno např. násobení vektorů.

<sup>3</sup>Poznamenejme, že ještě využíváme vztahu  $\overline{-a} = \bar{a}$ ,  $\forall a \in \mathbb{CD}_n$ , který plyne z následujícího:  $\overline{-a} = -(\sum_{i=1}^n (-a_i \alpha_i)) = -(-a_1 + \sum_{i=2}^n a_i \alpha_i) = a_1 - \sum_{i=2}^n a_i \alpha_i = \bar{a}$ , kde  $\alpha_i$  představuje  $i$ -tou základní jednotku v dané Cayley-Dicksonově algebře, přičemž  $\alpha_1 = 1$ .

<sup>4</sup>Důkaz, že jde skutečně o normu lze provést za pomoci Minkowského nerovnosti.

### 3. Hermitovské matice řádu 2 nad $\mathbb{CD}$ algebrami

Čtvercovou matici  $\mathbf{A}$  nazveme hermitovskou nad algebrou  $\mathbb{CD}_n$ , pokud její prvky splňují rovnost  $a_{ij} = \overline{a_{ji}}$ . Je ihned zřejmé, že  $\forall i a_{ii} \in \mathbb{R}$ . Dále uvažme matice řádu 2, tj. mající tvar

$$\begin{pmatrix} a & c \\ \bar{c} & b \end{pmatrix}, \quad (10)$$

kde  $a, b \in \mathbb{R}$  a  $c = \sum_{i=1}^n a_i \alpha_i \in \mathbb{CD}_n$ . Pro vlastní čísla vypočtená dle vztahu  $\det(\mathbf{A} - \lambda \mathbf{I}) = 0$  díky (6) platí

$$\Lambda = \frac{1}{2} \left\{ (a+b) \pm \sqrt{(a-b)^2 + 4c\bar{c}} \right\} = \left\{ \frac{1}{2}(a+b) \pm \sqrt{(a-b)^2 + 4 \sum_{i=1}^n a_i^2} \right\}. \quad (11)$$

Je zřejmé, že tato vlastní čísla jsou reálná. Každá hermitovská matice řádu 2 nad  $\mathbb{CD}_n$  má tedy alespoň dvě levá reálná vlastní čísla. Toto však neznamenaá, že nemůže mít ještě další vlastní čísla obecně z  $\mathbb{CD}_n$ . S tímto jevem se lze setkat např. u kvaternionových matic (viz [5]). Pro vzdálenost vlastních čísel (11), tj.  $|\Lambda_1 - \Lambda_2|$  platí

$$r = 2 \sqrt{\left(\frac{a-b}{2}\right)^2 + \sum_{i=1}^n a_i^2}. \quad (12)$$

### 4. Náhodné hermitovské matice řádu 2 nad $\mathbb{CD}_n$

Nechť platí, že  $a, b \sim \mathcal{N}(0, 2)$ ,  $\forall i a_i \sim \mathcal{N}(0, 1)$  a nechť jsou tyto náhodné veličiny statisticky nezávislé. Pak díky linearitě střední hodnoty a  $\text{var}[a(X+Y)] = a^2(\text{var}X + \text{var}Y)$  platí  $\frac{a-b}{2} \sim \mathcal{N}(0, 1)$ . Vzdálenost levých reálných vlastních čísel (10) lze psát tedy jako náhodnou veličinu<sup>5</sup>

$$r = 2 \sqrt{\sum_{i=1}^{n+1} a_i^2}, \quad (13)$$

kde  $a_i \sim \mathcal{N}(0, 1)$ . Určeme dále hustotu pravděpodobnosti rozdělení této vzdálenosti. Suma pod odmocninou má zřejmě rozdělení  $\chi^2(n+1)$ , jehož hustota pravděpodobnosti je

---

<sup>5</sup>Poznamenáváme, že vzhledem k faktu, že pro každou vygenerovanou náhodnou matici z našeho modelu lze určit pouze jednu hodnotu vzdálenosti vlastních čísel, nemá smysl provádět unfolding spektra.

$$f(x) = \theta(x) \frac{1}{2^{\frac{\nu}{2}} \Gamma\left(\frac{\nu}{2}\right)} x^{\frac{\nu}{2}-1} e^{-\frac{x}{2}}, \quad (14)$$

kde  $\nu = n + 1$  je počet stupňů volnosti. Jelikož  $2\sqrt{x}$  je pro  $x \geq 0$  monotónní transformace s inverzní funkcí  $\frac{x^2}{4}$ , užitím věty o monotónní transformaci rozdělení a dosazením za stupně volnosti  $\nu = n + 1$  dostáváme pro hustotu pravděpodobnosti veličiny (13) vztah

$$f(r) = \theta(r) \frac{1}{2^{\frac{3n+1}{2}} \Gamma\left(\frac{n+1}{2}\right)} r^n e^{-\frac{r^2}{8}}. \quad (15)$$

Dále vypočteme střední hodnotu  $ER$  této náhodné veličiny. S výhodou využijeme integrál<sup>6</sup>

$$\int_0^{+\infty} x^r e^{-ax^2} dx = \frac{\Gamma\left(\frac{r+1}{2}\right)}{2a^{\frac{r+1}{2}}}, \quad a > 0, \quad r > -1. \quad (16)$$

Máme tedy

$$ER = \frac{\Gamma\left(\frac{n+2}{2}\right)}{\Gamma\left(\frac{n+1}{2}\right)} 2^{\frac{3}{2}}. \quad (17)$$

Dále provedeme normalizaci vzdálenosti vlastní čísel střední hodnotou, tzn. transformaci náhodných veličin pomocí funkce  $h(r) = \frac{r}{ER}$ . Funkce je zřejmě monotónní. Po transformaci tedy máme

$$f_{\text{norm}}(r) = 2 \frac{[\Gamma\left(\frac{n+2}{2}\right)]^{n+1}}{[\Gamma\left(\frac{n+1}{2}\right)]^{n+2}} r^n e^{-x^2 \frac{[\Gamma\left(\frac{n+2}{2}\right)]^2}{[\Gamma\left(\frac{n+1}{2}\right)]^2}}. \quad (18)$$

Nalezli jsme tudíž hustotu pravděpodobnosti rozdělení vzdáleností reálných levých vlastních čísel hermitovských matic řádu 2 vybudovaných nad Cayley-Dicksonovými algebry o dimenzi  $n$ . Dimenzi algebry tedy můžeme ztotožnit s repulzním parametrem  $\beta$ .

Celkově tedy máme návod pro tvorbu matic, jejichž spektrum je popsitelné parametrem  $\beta = 2^k$ , kde  $k \in \{0, 1, 2, \dots\}$ . Tyto náhodné matice můžeme nazvat Cayley-Dickson-Wignerovými maticemi.

Dále jsme ukázali, že vztah (13) transformuje standardní normální rozdělení na Wignerovo nenormalizované rozdělení s parametrem  $\beta = n$ . Pokud

---

<sup>6</sup>Připomínáme definici gama funkce:  $\forall x > 0: \Gamma(x) = \int_0^{+\infty} t^{x-1} e^{-t} dt$ .

navíc získané realizace vydělíme jejich výběrovým průměrem, dostaneme realizace náhodné veličiny popsané hustotou pravděpodobnosti (18).

Navíc lze předpokládat, že Wignerova domněnka dostatečně přesně aproximuje vzdálenosti uspořádaných vlastních čísel matic velkých rozměrů nad obecnými Cayley-Dicksonovými algebrami, stejně jako je tomu v případě klasických skupin náhodných matic, tj. GOE, GUE a GSE.

## 5. Výsledky numerické simulace

Pomocí  $\chi^2$  testu dobré shody otestujeme, zda vztah (13) skutečně generuje rozdělení popsané hustotou pravděpodobnosti (1). Test provedeme pro  $\beta = 1, 2, 4, 8, 16, 24$ . Poznamenejme, že  $\beta = 24$  neodpovídá žádné Cayley-Dicksonově algebře.

Test provádíme na hladině významnosti 5 %. Odpovídající  $p$ -hodnoty testu shrnuje tabulka 1. Srovnání teoretického a empirického průběhu hustoty pravděpodobnosti lze nalézt na grafech na další straně. Poznamenáváme, že vždy bylo vygenerováno  $10^6$  realizací náhodné veličiny.

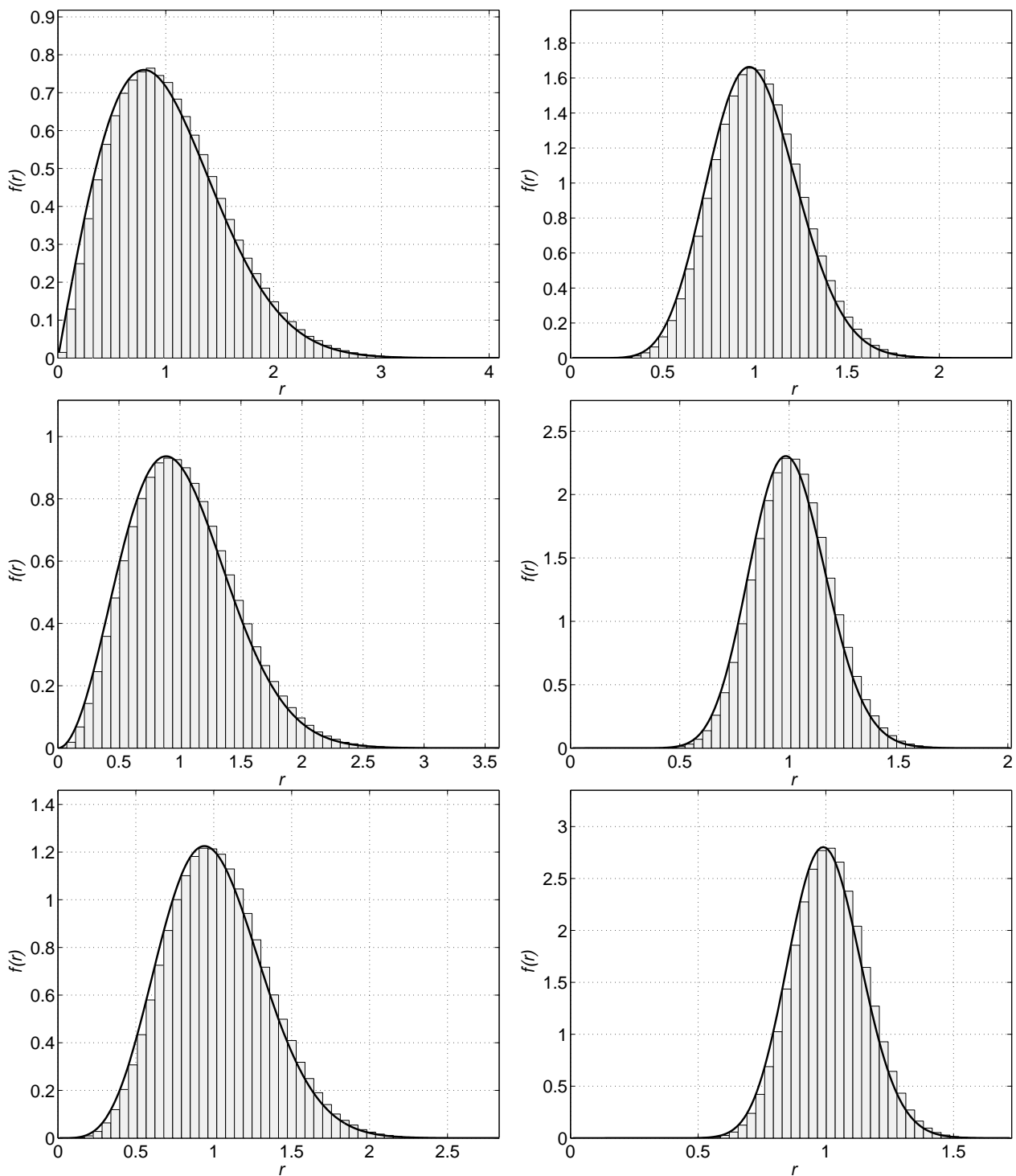
Parametr $\beta$	$p$ -hodnota
1	0,17041
2	0,98875
4	0,78498
8	0,90771
16	0,19587
24	0,46451

Tabulka 1:  $p$ -hodnoty  $\chi^2$  testu dobré shody pro Wignerovo rozdělení s různými hodnotami repulzního parametru  $\beta$ .

## 6. Závěr

V této práci jsme představili obecný postup pro tvorbu náhodných matic řádu 2, jejichž spektrum je popsáno repulzním parametrem  $\beta$ , jež nabývá hodnot odpovídající mocninám čísla 2. Analyticky jsme dokázali podobu hustoty pravděpodobnosti vzdálenosti uspořádaných vlastních čísel těchto matic. Tento analytický výsledek byl dále ověřen numericky. Navíc se nám podařilo vytvořit transformační vztah převádějící rozdělení  $\mathcal{N}(0, 1)$  na normalizované Wignerovo rozdělení.





Obrázek 1: Srovnání teoretického průběhu hustoty pravděpodobnosti popsané Wignerovou domněnkou s výstupem generátoru Wignerova rozdělení. Shora dolů nabývá parametr  $\beta$  hodnot 1, 2, 4 (levý sloupec) a 8, 16, 24 (pravý sl.).

## Použitá literatura

- [1] Izrailev F. M. a Scharf R.: *Dyson's Coulomb gas on a circle and intermediate eigenvalue statistics*, Journal of Physics A: Mathematical and general, vol. **23**, no. **6**, (1990) 963–977.
- [2] Biss D. K., Christensen J. D., Dugger D., Isaksen D. C. *Eigentheory of Cayley-Dickson algebras*, Forum Mathematicum, vol. **21**, issue **5**, (2009) 833–851, dostupné on-line:  
<http://pages.uoregon.edu/ddugger/eigen.pdf>
- [3] Nieminen J. M.: *Eigenvalue spacing statistics of a four-matrix model of some four-by-four random matrices*, Journal of Physics A: Mathematical and Theoretical, vol. **42**, (2009).
- [4] Nieminen J. M.: *Two-by-two random matrix theory with matrix representation of octonions*, Journal of mathematical physics, vol. **51**, (2010).
- [5] Zhang F.: *Quaternions and quaternion matrices*, Linear algebra and its application, vol. **251**, (1997) 21–57.
- [6] Tian Y.: *Matrix representations of octonions and their applications*, Cornell University Library (2000), dostupné on-line:  
<http://arxiv.org/abs/math/0003166v2>
- [7] Veselý M.: *Úvod do náhodných matic*, Informační bulletin České statistické společnosti, roč. **22**, č. **1**, (2011) 5–12, dostupné online:  
<http://statapol.cz/bulletiny/ib-2011-1-web.pdf>

# A Multivariate Two-Sample Test in Nonparametric Methods

## Vícerozměrný neparametrický dvouvýběrový test

**Hidetoshi Murakami**

*Adresa:* Dept. of Mathematics, Graduate School of Science and Engineering,  
Chuo University, 1-13-27 Kasuga, Bunkyo-ku, Tokyo 112-8551, Japan

*E-mail:* murakami@gug.math.chuo-u.ac.jp

**Abstract:** In this paper, a novel nonparametric multivariate rank test based on a Baumgartner type statistic is proposed. Simulations are used to investigate the power of suggested statistics for various population distributions.

**Keywords:** Nonparametric Statistics, Multivariate Rank Test.

### 1. Introduction

The purpose of this paper is to consider a multivariate two-sample problem, which is one of the most important statistical problems. Let  $X = (X_1, \dots, X_n)$  and  $Y = (Y_1, \dots, Y_m)$  be two random samples of size  $n$  and  $m$  independent observations, each of which has a continuous distribution  $F(x)$  and  $G(y)$ , respectively. In nonparametric methods, the Wilcoxon test (Hollander and Wolfe; 1999) is a standard test for the location parameters such as  $F(x) = G(y - \theta)$ . Baumgartner *et al.* (1998) introduced a nonparametric two-sample rank test, and the power of the Baumgartner statistic is almost equivalent to the Wilcoxon test. The aforementioned authors asserted the Baumgartner statistic could be applied for a scale parameter such as  $F(x) = G(y/\sigma)$  and was more powerful than the Kolmogorov-Smirnov (Gibbons; 2003) and the Cramér-von Mises (Hájek *et al.*; 1999) tests.

Let  $R_1 < \dots < R_n$  and  $H_1 < \dots < H_m$  denote the combined-samples ranks of the  $X$ -value and  $Y$ -value in an increasing order of magnitude, respectively. The test statistic proposed by Baumgartner *et al.* is

$$B = \frac{1}{2}(B_X + B_Y),$$

where

$$B_X = \frac{1}{n} \sum_{i=1}^n \frac{\left(R_i - \frac{n+m}{n}i\right)^2}{\frac{i}{n+1} \left(1 - \frac{i}{n+1}\right) \frac{m(n+m)}{n}}$$

and

$$B_Y = \frac{1}{m} \sum_{j=1}^m \frac{\left(H_j - \frac{m+n}{m}j\right)^2}{\frac{j}{m+1} \left(1 - \frac{j}{m+1}\right) \frac{n(m+n)}{m}}.$$

Recently, Murakami (2006) defined a  $k$ -sample Baumgartner statistic. In addition, Neuhäuser (2003) suggested the Baumgartner statistic in the presence of ties. Additionally, Neuhäuser (2001) investigated the behavior of a modified Baumgartner statistic in a one-sided test. In many cases, the location and scale parameters are tested at the same time. Then Neuhäuser (2000) introduced a modified Lepage statistic, namely  $L_B$ , which was combined with the Baumgartner and Ansari-Bradley (1960) statistics. In addition, Murakami [8] suggested a modification of  $L_B$  statistic which was combined with another modified Baumgartner statistic and the Mood (1954) statistic. A modified Baumgartner statistic proposed by Murakami (2006) was defined as

$$B^* = \frac{1}{2}(B_X^* + B_Y^*),$$

where

$$B_X^* = \frac{1}{n} \sum_{i=1}^n \frac{\left(R_i - \frac{n+m+1}{n+1}i\right)^2}{\frac{i}{n+1} \left(1 - \frac{i}{n+1}\right) \frac{m(n+m+1)}{n+2}}$$

and

$$B_Y^* = \frac{1}{m} \sum_{j=1}^m \frac{\left(H_j - \frac{m+n+1}{m+1}j\right)^2}{\frac{j}{m+1} \left(1 - \frac{j}{m+1}\right) \frac{n(m+n+1)}{m+2}}.$$

The  $B^*$  statistic is used with the exact mean and variance of  $R_i$  and  $H_j$ . The  $B^*$  statistic is more powerful than the  $B$  statistic for a location parameter when sample sizes are unequal. In addition, it is also important with a statistical problem to consider a multivariate case. For a bivariate case, Murakami [9] proposed the bivariate Baumgartner statistic and derived the limiting distribution. In this paper, we propose a multivariate nonparametric rank test in Section 2. To investigate the power of the multivariate Baumgartner statistic, we carry out simulation studies of various population distributions in Section 3. All the simulations are repeated 10,000 times and there are 10,000 permutations in this paper.

## 2. A multivariate statistic

In this section, we propose a multivariate Baumgartner statistic, namely  $B_p$ . Let  $\mathbf{X} = (\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(p)})'$  and  $\mathbf{Y} = (\mathbf{y}^{(1)}, \mathbf{y}^{(2)}, \dots, \mathbf{y}^{(p)})'$ , where  $\mathbf{x}^{(d)} = (x_{d1}, \dots, x_{dn})'$ ,  $\mathbf{y}^{(d)} = (y_{d1}, \dots, y_{dm})'$ ,  $d = 1, 2, \dots, p$  are two random samples of size  $n$  and  $m$  independent observations from different populations and with  $p$ -dimensional continuous distribution  $F(\mathbf{x})$  and  $G(\mathbf{y})$ , respectively.

Suppose that  $R_1^{(d)} < \dots < R_n^{(d)}$  and  $H_1^{(d)} < \dots < H_m^{(d)}$  are the combined-sample ranks of the  $\mathbf{X}$ -value and  $\mathbf{Y}$ -value in increasing order of magnitude, respectively. This means that it is possible to obtain a separate ranking for each variable (Puri and Sen; 1971). Now we define a multivariate Baumgartner statistic as follows:

$$B_p = \sum_{d=1}^p \frac{1}{2} \left( B_X^{(d)} + B_Y^{(d)} \right),$$

where

$$B_X^{(d)} = \frac{1}{n} \sum_{i=1}^n \frac{\left( R_i^{(d)} - \frac{n+m+1}{n+1} i \right)^2}{\frac{i}{n+1} \left( 1 - \frac{i}{n+1} \right) \frac{m(n+m+1)}{n+2}}$$

and

$$B_Y^{(d)} = \frac{1}{m} \sum_{j=1}^m \frac{\left( H_j^{(d)} - \frac{m+n+1}{m+1} j \right)^2}{\frac{j}{m+1} \left( 1 - \frac{j}{m+1} \right) \frac{n(m+n+1)}{m+2}}.$$

We use the permutation test to estimate the  $p$ -value because it is difficult to calculate the exact critical values of  $B_p$  statistic.

## 3. Simulation study

Next, we investigate the behaviour of the  $B_p$  statistic. For power comparison of the statistics, we conduct a simulation study for some distributions as in different populations. In particular, we tested the hypothesis  $H_0: F(\mathbf{x}) = G(\mathbf{y})$  against  $H_1: \text{not } H_0$ . We assumed that  $F(\mathbf{x})$  and  $G(\mathbf{y})$  described the following distributions.

1.  $N(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$  and  $N(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$  : the Normal distributions
2.  $\eta(\boldsymbol{\lambda}_1)$  and  $\eta(\boldsymbol{\lambda}_2)$  : the exponential distributions

Suppose  $\Sigma = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_p)$ . Generally, the location and scale parameters of the  $\mathbf{X}$  and  $\mathbf{Y}$  samples are unequal. We examined the power at which the location and scale parameters differed. The following Tables show the results of power of the multivariate Wilcoxon test, namely  $W_p$ ,

$$W_p = \sum_{d=1}^p \left| \sum_{i=1}^n R_i^{(d)} - \frac{n(n+m+1)}{2} \right|$$

and the  $B_p$  statistics where  $n = m = 10$  and  $n = 10, m = 5$ .

For all cases, 10,000 permutations in each simulation were performed, and we simulated 10,000 times to obtain the actual significance level. We treat the case of  $p = 3, 4, \text{ and } 5$  in this paper.

Table 1-a. Case of  $n = m = 10$  for  $N_3(\mathbf{0}, \mathbf{I}_3)$  and  $N_3(\boldsymbol{\mu}_2, \sigma \mathbf{I}_3)$

$\boldsymbol{\mu}_2$		$\sigma$				
		1.0	5.0	10.0	15.0	20.0
<b>0.0</b>	$W_p$	<b>0.05</b>	0.07	0.08	0.09	0.09
	$B_p$	<b>0.05</b>	0.20	0.43	0.58	0.69
<b>0.5</b>	$W_p$	0.28	0.13	0.11	0.11	0.11
	$B_p$	0.27	0.31	0.49	0.62	0.72
<b>1.0</b>	$W_p$	0.88	0.39	0.24	0.20	0.17
	$B_p$	0.85	0.59	0.63	0.71	0.77
<b>1.5</b>	$W_p$	1.00	0.74	0.48	0.35	0.29
	$B_p$	1.00	0.87	0.83	0.84	0.86

Table 1-b. Case of  $n = 10, m = 5$  for  $N_3(\mathbf{0}, \mathbf{I}_3)$  and  $N_3(\boldsymbol{\mu}_2, \sigma \mathbf{I}_3)$

$\boldsymbol{\mu}_2$		$\sigma$				
		1.0	5.0	10.0	15.0	20.0
<b>0.0</b>	$W_p$	<b>0.05</b>	0.11	0.14	0.16	0.17
	$B_p$	<b>0.05</b>	0.22	0.37	0.47	0.53
<b>0.5</b>	$W_p$	0.18	0.15	0.16	0.17	0.18
	$B_p$	0.19	0.29	0.41	0.49	0.55
<b>1.0</b>	$W_p$	0.67	0.31	0.25	0.23	0.23
	$B_p$	0.64	0.47	0.51	0.55	0.59
<b>1.5</b>	$W_p$	0.97	0.55	0.38	0.32	0.29
	$B_p$	0.95	0.69	0.64	0.65	0.66

Table 2-a. Case of  $n = m = 10$  for  $N_4(\mathbf{0}, \mathbf{I}_4)$  and  $N_4(\boldsymbol{\mu}_2, \sigma \mathbf{I}_4)$ 

$\boldsymbol{\mu}_2$		$\sigma$				
		1.0	5.0	10.0	15.0	20.0
<b>0.0</b>	$W_p$	<b>0.05</b>	0.06	0.08	0.08	0.09
	$B_p$	<b>0.05</b>	0.24	0.53	0.72	0.82
<b>0.5</b>	$W_p$	0.33	0.15	0.13	0.12	0.12
	$B_p$	0.32	0.38	0.61	0.75	0.84
<b>1.0</b>	$W_p$	0.95	0.46	0.28	0.22	0.19
	$B_p$	0.92	0.69	0.76	0.83	0.89
<b>1.5</b>	$W_p$	1.00	0.84	0.56	0.42	0.34
	$B_p$	1.00	0.94	0.91	0.92	0.94

Table 2-b. Case of  $n = 10, m = 5$  for  $N_4(\mathbf{0}, \mathbf{I}_4)$  and  $N_4(\boldsymbol{\mu}_2, \sigma \mathbf{I}_4)$ 

$\boldsymbol{\mu}_2$		$\sigma$				
		1.0	5.0	10.0	15.0	20.0
<b>0.0</b>	$W_p$	<b>0.05</b>	0.13	0.16	0.18	0.19
	$B_p$	<b>0.05</b>	0.27	0.47	0.58	0.66
<b>0.5</b>	$W_p$	0.21	0.18	0.19	0.20	0.21
	$B_p$	0.21	0.35	0.50	0.60	0.66
<b>1.0</b>	$W_p$	0.77	0.38	0.30	0.27	0.26
	$B_p$	0.73	0.57	0.63	0.68	0.72
<b>1.5</b>	$W_p$	0.99	0.65	0.46	0.39	0.35
	$B_p$	0.99	0.80	0.75	0.76	0.79

Table 3-a. Case of  $n = m = 10$  for  $N_5(\mathbf{0}, \mathbf{I}_5)$  and  $N_5(\boldsymbol{\mu}_2, \sigma \mathbf{I}_5)$ 

$\boldsymbol{\mu}_2$		$\sigma$				
		1.0	5.0	10.0	15.0	20.0
<b>0.0</b>	$W_p$	<b>0.05</b>	0.06	0.08	0.09	0.09
	$B_p$	<b>0.05</b>	0.28	0.64	0.82	0.91
<b>0.5</b>	$W_p$	0.39	0.16	0.13	0.13	0.13
	$B_p$	0.36	0.44	0.70	0.84	0.91
<b>1.0</b>	$W_p$	0.97	0.53	0.33	0.25	0.22
	$B_p$	0.96	0.78	0.85	0.91	0.94
<b>1.5</b>	$W_p$	1.00	0.90	0.63	0.48	0.38
	$B_p$	1.00	0.97	0.95	0.96	0.97

Table 3-b. Case of  $n = 10$ ,  $m = 5$  for  $N_5(\mathbf{0}, \mathbf{I}_5)$  and  $N_5(\boldsymbol{\mu}_2, \sigma \mathbf{I}_5)$

$\boldsymbol{\mu}_2$		$\sigma$				
		1.0	5.0	10.0	15.0	20.0
<b>0.0</b>	$W_p$	<b>0.05</b>	0.13	0.17	0.20	0.21
	$B_p$	<b>0.05</b>	0.31	0.53	0.66	0.74
<b>0.5</b>	$W_p$	0.24	0.20	0.21	0.22	0.23
	$B_p$	0.24	0.42	0.59	0.70	0.76
<b>1.0</b>	$W_p$	0.84	0.42	0.33	0.30	0.29
	$B_p$	0.81	0.64	0.70	0.76	0.81
<b>1.5</b>	$W_p$	1.00	0.72	0.52	0.43	0.39
	$B_p$	0.99	0.86	0.82	0.84	0.86

In this case, when the location (but not the scale) was shifted, the power of the  $W_p$  statistic is greater than the  $B_p$  statistic but the difference between these two statistics was small. Furthermore, the  $B_p$  statistic was more powerful for scale and location-scale parameter shifts. Therefore, the  $B_p$  statistic is more suitable than the  $W_p$  statistic for treating the parameters associated with Normal distribution.

Table 4-a. Case of  $n = m = 10$  for  $\eta_3(\mathbf{1})$  and  $\eta_3(\sigma \mathbf{1})$

	$\sigma$				
	1.0	2.0	3.0	4.0	5.0
$W_p$	<b>0.05</b>	0.42	0.82	0.96	0.99
$B_p$	<b>0.05</b>	0.42	0.82	0.96	0.99

Table 4-b. Case of  $n = 10$ ,  $m = 5$  for  $\eta_3(\mathbf{1})$  and  $\eta_3(\sigma \mathbf{1})$

	$\sigma$				
	1.0	2.0	3.0	4.0	5.0
$W_p$	<b>0.05</b>	0.28	0.63	0.82	0.92
$B_p$	<b>0.05</b>	0.30	0.65	0.84	0.93

Table 5-a. Case of  $n = m = 10$  for  $\eta_4(\mathbf{1})$  and  $\eta_4(\sigma \mathbf{1})$

	$\sigma$				
	1.0	2.0	3.0	4.0	5.0
$W_p$	<b>0.5</b>	0.50	0.91	0.99	1.00
$B_p$	<b>0.5</b>	0.49	0.91	0.99	1.00



Table 5-b. Case of  $n = 10$ ,  $m = 5$  for  $\eta_4(\mathbf{1})$  and  $\eta_4(\sigma\mathbf{1})$

	$\sigma$				
	1.0	2.0	3.0	4.0	5.0
$W_p$	<b>0.5</b>	0.33	0.73	0.91	0.96
$B_p$	<b>0.5</b>	0.35	0.74	0.92	0.97

Table 6-a. Case of  $n = m = 10$  for  $\eta_5(\mathbf{1})$  and  $\eta_5(\sigma\mathbf{1})$

	$\sigma$				
	1.0	2.0	3.0	4.0	5.0
$W_p$	<b>0.05</b>	0.56	0.96	1.00	1.00
$B_p$	<b>0.05</b>	0.56	0.95	1.00	1.00

Table 6-b. Case of  $n = 10$ ,  $m = 5$  for  $\eta_5(\mathbf{1})$  and  $\eta_5(\sigma\mathbf{1})$

	$\sigma$				
	1.0	2.0	3.0	4.0	5.0
$W_p$	<b>0.05</b>	0.37	0.79	0.94	0.98
$B_p$	<b>0.05</b>	0.39	0.80	0.95	0.99

From the results of the simulation study, the power of  $B_p$  statistic is equivalent to  $W_p$  statistic when sample sizes are equal. However, when  $n \neq m$ , the  $B_p$  statistic is more efficient than the  $W_p$  statistic. Therefore, the  $B_p$  statistic is more suitable than the  $W_p$  statistic for evaluating the parameters associated with the exponential distribution.

## Conclusion and discussion

In this paper, we proposed a multivariate nonparametric test based on the Baumgartner type statistic. The results of our simulations of using the permutation for Normal and exponential distributions indicated that the multivariate Baumgartner statistic was more suitable than the multivariate Wilcoxon test. In the future, it will be important to derive the limiting distribution of the multivariate Baumgartner statistic.

## Reference

- [1] Ansari, A. R. and Bradley, R. A. (1960), Rank sum tests for dispersion, *Annals of Mathematical Statistics*, **31**, No. 4, pp. 1174–1189. doi: 10.1214/aoms/1177705688
- [2] Baumgartner, W., Weiß, P., Schindler, H. (1998), A nonparametric test for the general two-sample problem. *Biometrics*, **54**, No. 3, 1129–1135. doi: 10.2307/2533862
- [3] Gibbons, J. D. and Chakraborti, S. (2003), *Nonparametric Statistical Inference*, 4th edition. Dekker, New York. ISBN 0-8247-4052-1.
- [4] Hájek, J., Šidák, Z. and Sen, P. K. (1999), *Theory of rank tests*, 2nd edition. Academic Press, San Diego. ISBN 0-12-642350-4.
- [5] Hollander, M., Wolfe, D. A. (1999), *Nonparametric Statistical Methods*, 2nd edition. John Wiley & Sons, New York. ISBN 978-0471190455.
- [6] Mood, A. M. (1954), On the asymptotic efficiency of certain nonparametric two-sample tests, *Annals of Mathematical Statistics*, **25**, No. 3, 514–522. doi: 10.1214/aoms/1177728719
- [7] Murakami, H. (2006), A  $k$ -sample rank test based on modified Baumgartner statistic and its power comparison. *Journal of the Japanese Society of Computational Statistics*, **19**, No. 1, 1–13. ISSN 0915-2350.
- [8] Murakami, H. (2007), Lepage type statistic based on the modified Baumgartner statistic. *Computational Statistics and Data Analysis*, Vol. 51, No. 10, pp. 5061–5067. ISSN 0167-9473. doi: 10.1016/j.csda.2006.04.026
- [9] Murakami, H. (submitted), A bivariate two-sample Baumgartner statistic. *Journal of Nonparametric Statistics*. ISSN 1048-5252.
- [10] Neuhäuser, M. (2000), An exact two-sample test based on the Baumgartner-Weiss-Schindler statistic and a modification of Lepage’s test. *Communications in Statistics – Theory and Methods*, **29**, No. 1, pp. 67–78. ISSN 0361-0926. doi: 10.1080/03610920008832469
- [11] Neuhäuser, M. (2001), One-sided two-sample and trend tests based on a modified Baumgartner-Weiss-Schindler statistic. *Journal of Nonparametric Statistics*, **13**, No. 5, 729–739. doi: 10.1080/10485250108832874
- [12] Neuhäuser, M. (2003), A note on the exact test based on the Baumgartner-Weiß-Schindler statistic in a presence of ties. *Computational Statistics and Data Analysis*, **42**, 561–568. doi: 10.1016/S0167-9473(02)00121-4
- [13] Puri, M. L. and Sen, P. K. (1971), *Nonparametric Methods in Multivariate Analysis*, John Wiley & Sons, New York. ISBN 0471702404.

# Protectorate Statistical Service and the 65th Anniversary of the End of the World War II

## Šedesátépaté výročí konce druhé světové války a protektorátní statistická služba

Jaroslav Češka

V roce, kdy si připomínáme šedesátépaté výročí konce druhé světové války a porážky nacistické Velkoněmecké říše v roce 1945, je účelné stručně popsat důsledky ztráty samostatnosti ČSR na úseku oficiální statistické služby vykonávané na zbylém území – v Protektorátu Čechy a Morava.

Rozbití Československa, zřízení Protektorátu Čechy a Morava výnosem A. Hitlera dne 16. března 1939 i odstoupení pohraničních území ČSR Německu v říjnu 1938 mělo dalekosáhlé důsledky pro české obyvatelstvo i činnost všech státních orgánů, včetně Státního úřadu statistického.

Území zřízeného Protektorátu představovalo jen 29 % území bývalé ČSR, ztráty podle posledního předválečného soupisu v roce 1930 u obytných budov činily 56 %, u obyvatelstva 54 %.

Státní úřad statistický musel uzpůsobit svoji působnost a činnost nové zásadně změněné situaci. Již v době tzv. druhé republiky /říjen 1938 – březen 1939/ po odstoupení pohraničních území Čech a Moravy na základě Mnichovské dohody čtyř velmocí /29. 9. 1938/ a po později odstoupeném území Slovenska a Podkarpatské Rusi Maďarsku na základě Vídeňské arbitráže /2. 11. 1938/ Státní úřad statistický zajišťoval mimořádné práce vyvo-

Tabulka 1: Území, obytné budovy a obyvatelstvo

Území	Rozloha km <sup>2</sup>	Obytné budovy	Přítomné obyv.
		Počet	
Protektorát Čechy a Morava	48 901	1 021 739	6 806 788
Území podstoupená:			
A. Německé říši	29 139	546 362	3 651 746
B. Polsku	866	26 978	231 418
C. Maďarsku	24 088	294 809	1 626 620
Z toho Podkarpatská Rus	11 085	112 457	552 124
Slovensko	37 505	409 986	2 412 964
Celkem /býv. ČSR/	140 499	2 299 874	14 729 536

lané uvedeným odstoupením území ČSR. Tyto práce zahrnovaly nová zpracování z podrobných podkladů dřívějších statistických soupisů, zejména ze sčítání lidu a bytů v roce 1930 a dalších velkých statistických šetření.

K novým naléhavým úkolům patřilo i sestavování přehledů obcí /i osad/ dotčených uvedenými územními změnami a další s tím spojené otázky, včetně tzv. spisové rozluky, tj. předání příslušných statistických podkladů a výsledků zpracování o odstoupených územích příslušným orgánům Německé říše a Maďarska. Řešení příslušných metodických a zpracovatelských problémů bylo ztíženo i tím, že nižší administrativní celky – okresy byly v některých případech územně rozděleny nově určenými hranicemi.

I když statistická služba nepatřila k úsekům určeným ve výnosu A. Hitlera k přímému řízení orgány Velkoněmecké říše, Státní statistický úřad musel po zřízení Protektorátu v březnu 1939 podřídit svoji činnost požadavkům německé správy na území Protektorátu, zejména Úřadu říšského protektora, Říšským protektorem zřízených regionálních úřadů /„Oberlandratů“/, a rovněž požadavkům Říšského statistického úřadu.

Jako ústřední úřad podřízený protektorátní vládě statistický úřad se musel řídit jednak nařízeními a závaznými pokyny obecné povahy, které se týkaly jeho činnosti a osazenstva jako ústředního úřadu, tak i novými požadavky na obsah i rozsah statistických zjišťování v jednotlivých úsecích, jejich přizpůsobení statistice říšské, sběr a zpracování údajů a předávání výsledných sestav a publikací určeným novým příjemcům ve stanovených lhůtách, včetně dalších souvisejících pracovních činností a nových postupů vyvolaných utajováním statistických dat, dvojjazyčným tiskem dotazníků a publikací apod.

Z nařízení a závazných pokynů obecné povahy lze např. uvést slib věrnosti A. Hitlerovi a dr. E. Háchovi, státnímu prezidentu Protektorátu, požadovaný od všech veřejných zaměstnanců Protektorátu, nařízení o zavedení nacistického pozdravu v protektorátních úřadech, požadavky na používání němčiny, její studium a složení příslušných zkoušek z německého jazyka zaměstnanci protektorátních úřadů, předložení důkazů od jednotlivých zaměstnanců o jejich arijském původu, postih českých občanů židovské národnosti, zaměstnávání bývalých legionářů v protektorátních úřadech a jiné.

Po zahájení druhé světové války napadením Polska jednotkami německé armády dne 1. září 1939 byla vydána řada dalších obecně závazných nařízení vyvolaná válečnými poměry k zajištění válečného hospodářství Německé říše na území Protektorátu, kterými se musel statistický úřad ve své činnosti i jeho zaměstnanci řídit.

Krátce po zřízení Protektorátu byl Státní úřad statistický přejmenován na Ústřední statistický úřad „Statistisches Zentralamt“, a to na základě závazného pokynu Úřadu říšského protektora o novém označení protektorátních

ústředních úřadů a jejich podřízených organizací v přednostním německém jazykem. Podle uvedeného pokynu nesmělo v názvech být použito slov „státní“, „český“ a slov spojených se jménem prvního prezidenta ČSR T. G. Masaryka. Obdobná změna byla uplatněna i v případě Statistické rady státní, která byla do roku 1939 usnášejícím a řídicím orgánem československé statistické služby. I když již od začátku roku 1939 Statistická rada státní nebyla svolávána a nevyvíjela žádnou činnost, její název byl pozdějším vládním nařízením upraven na „Statistická rada“.

V důsledku uvedených změn nová statistická zjišťování a související činnosti ve sběru příslušných údajů organizovaná Ústředním statistickým úřadem se neopírala o usnesení Statistické rady a výborů rady, nýbrž měla právní oporu v příslušných nařízeních Protektorátní vlády nebo, podle jejich charakteru, v nařízeních předsedy Protektorátní vlády ing. A. Eliáše. V některých výjimečných případech se činnosti statistického úřadu /např. na úseku statistiky zahraničního obchodu/ opírala o nařízení říšských ministrů vydaných v dohodě s říšským protektorem. Obdobně i pracovní plány statistického úřadu, který byl podle statistického zákona č. 49/1919 Sb. výkonným orgánem statistické služby, byly v počátečním období existence Protektorátu schvalovány předsedou protektorátní vlády.

Zřízení Protektorátu vedlo také ke zvýšenému pohybu v osazenstvu Ústředního statistického úřadu (ÚSÚ). Řada pracovníků odešla ze služeb úřadu z různých příčin. Do statistického úřadu byl naopak přijat na základě rozhodnutí protektorátní vlády značný počet pracovníků ze zrušených ministerstev po 16. březnu 1939, zejména z Ministerstva národní obrany (MNO) a Ministerstva zahraničních věcí. Značný počet převzatých pracovníků z uvedených ministerstev, zejména z MNO ve výši 220, vedl k tomu, že celkový počet pracovníků úřadu se zvýšil ze 717 v roce 1938 na 861 ke konci roku 1939.

Někteří vedoucí pracovníci ÚSÚ odcházeli do důchodu. Krátce po ustavení Protektorátu musel do předčasného důchodu odejít i doc. Dr. J. Auerhan, prezident Státního úřadu statistického a uznávaný odborník na problematiku národnostních menšin. Později ho následoval do předčasného důchodu i český viceprezident statistického úřadu doc. Dr. A. Boháč, významný statistický odborník, považovaný za zakladatele naší demografické statistiky.

Odchod českých statistických odborníků vedl současně k nástupu německých státních příslušníků do vedení statistického úřadu. Ještě v roce 1939 byl povýšen do funkce vicepresidenta ÚSÚ dr. A. Oberschall, pracovník SÚS německé národnosti, který na základě výnosu A. Hitlera se stal jako „volksdeutsche“ občanem Německé říše a který po svém jmenování ÚSÚ po určitou dobu řídil. V prosinci 1941 byl vedoucím Ústředního statistického úřadu jmenován zastupujícím Říšským protektorem R. Heydrichem dr. H. Wirth jako

„Kommissarische Leiter“, který tuto novou funkci vykonával souběžně se svou funkcí vedoucího statistického útvaru v Úřadě říšského protektora.

Příchodem dr. Wirtha celé užší vedení Ústředního statistického úřadu bylo obsazeno říšskými státními příslušníky. Protektorátní statistický úřad byl tak plně zapojen do Říšské statistické správy a stal se jejím regionálním orgánem. Dr. Wirth i ve své funkci v ÚSÚ nadále podléhal Říšskému protektoru a protektorátní vláda ztratila kontrolu nad svým statistickým úřadem. Do statistického úřadu bylo také přijato několik dalších německých občanů z říšské statistické správy.

Protektorát Čechy a Morava se stal součástí Velkoněmecké říše a jeho hospodářství bylo, i na základě příslušných právních norem, přizpůsobeno hospodářství říše. Tyto skutečnosti vedly rovněž k požadavkům na přizpůsobení protektorátní /bývalé československé/ úřední statistiky, uplatňovaným Říšským statistickým úřadem (ŘSÚ). Požadavky tohoto úřadu, jakož i požadavky jiných orgánů Říše shromažďované Říšským statistickým úřadem, byly v počátečním období Protektorátu předkládány Říšskému protektoru, jehož postavení jako nejvyššího představitele A. Hitlera a Říšské vlády v Protektorátu muselo být respektováno. Úřad říšského protektora pak uplatňoval tyto požadavky u protektorátní vlády nebo prostřednictvím svého statistického oddělení přímo u Ústředního statistického úřadu.

Požadavky na přizpůsobení říšské statistice se týkaly řady statistických úseků a všech etap statistické práce, metod zjišťování, použití říšských vzorů výkazů a hlášení, předmětu zjišťování, okruhu zpravodajských jednotek, zjišťovaných charakteristik statistických jednotek, metod výpočtu ukazatelů, klasifikací a třídících znaků používaných ŘSÚ, zpracování, předkládání výsledných informací i statistických publikací. Požadavky na převzetí říšských postupů se týkaly i statistik, kde nahrazovaná bývalá československá oficiální statistika byla na vyšší odborné a mezinárodně uznávané úrovni, jako např. v oblasti zahraničního obchodu.

Se zavedením řízeného /válečného/ hospodářství na celém území Německé říše důraz ve statistických zjišťováních na území Protektorátu byl kladen na vyšší operativnost, opakovaná zjišťování v kratších intervalech a úplnost zahrnováním všech výrobních/ekonomických jednotek příslušného druhu. Nová zjišťování byla zaměřována na otázky důležité pro válečné hospodářství, statistiku výživy, vydávání potravin, jejich spotřebu a přidělové hospodářství, zdroje pracovních sil a jejich využívání, lepší zhodnocování a využití surovin a surovinových zdrojů, zvýšení výrobních kapacit i jejich využívání v průmyslových a řemeslných závodech a jiných.

Některá statistická hlášení, která byla považována za nedůležitá pro válečné hospodářství, byla omezována nebo i rušena. Některá zjišťování byla

ukončena, neboť příslušný předmět zjišťování byl z rozhodnutí německých orgánů zrušen. Např. v důsledku uzavření českých vysokých škol přestala mít opodstatnění příslušná statistika o této formě vzdělávání.

K redukcím statistických zjišťování docházelo i z rozhodnutí nejvyšších Říšskoněmeckých míst požadujících celkové omezování statistických hlášení a požadavků na tomto úseku v době války. Příslušné statistické orgány v Říši i Protektorátě byly za tím účelem pověřeny schvalováním statistických hlášení a výkazů organizovaných jinými orgány. V Říši vykonával tuto činnosti zřízený Ústřední statistický výbor, na území Protektorátu byl schvalováním pověřen Ústřední statistický úřad, a to ve vztahu k protektorátním úřadům a institucím. Statistická hlášení organizovaná Říšskými úřady na území Protektorátu vyžadovaly předběžný souhlas Úřadu říšského protektora. Oprávnění statistického úřadu na tomto úseku bylo převzato do poválečných statistických právních norem.

Některá jednorázová statistická šetření, např. o koncernech a koncerbových podnicích byla prováděna v Říšskoněmeckém zájmu, usnadnila průnik německého kapitálu do protektorátních koncerbových centrál a podniků, jehož objem do konce války mnohonásobně vzrostl.

Německé orgány neměly jednoznačný zájem na provádění velkých statistických soupisů na území Protektorátu během válečného období. Takový byl také osud Sčítání lidu a bytů, které podle příslušného československého zákona se mělo uskutečnit v roce 1940. I přes značně pokročilou přípravu tohoto sčítání při vzniku Protektorátu jeho provedení bylo nejdříve odkládáno a později zrušeno. Přitom však na celém území Velkoněmecké říše, včetně nově získaných území, bylo sčítání v uvedené době provedeno.

Některé demografické a jiné přehledy byly sestavovány jen podle vydaných, příp. vyřazených potravinových lístků, což obsahovalo řadu omezení, jak ukazuje přehled o českém obyvatelstvu Protektorátu podle věku v XV. zásobovacím období.

Významná omezení protektorátní statistiky vyplývala také přímo z výnosu A. Hitlera o zřízení Protektorátu. Podle článku 2 tohoto výnosu občané německé národnosti v Protektorátu se stali občany Říše a podléhali jen německé jurisdikci. Údaje o této skupině obyvatel nebyly předmětem statistických zjišťování organizovaných Ústředním statistickým úřadem. Čeští zaměstnanci statistického úřadu neměli přístup k údajům soupisů německých občanů v Protektorátě, které byly organizovány Úřadem říšského protektora, ani po zřízení Úřadu německého státního ministra K. H. Franka v listopadu 1943, kdy tato agenda byla převedena na Ústřední statistický úřad. Ze zpracování potravinových lístků podle národnosti, které mohlo být realizo-

Věková skupina	Zásobovací období /28. října – 24. listopadu 1940/
Obyvatelstvo celkem	7 399 290
v tom	
Děti	
do 1,5 roku	166 292
od 1,5 do 5 let	433 084
od 6 do 9 let	437 557
od 10 do 13 let	472 318
Mládež od 13 let a dospělé obyvatelstvo	5 890 039

váno až po válce, vyplynulo, že na území Protektorátu žilo k 8. listopadu 1943 jen 248 984 občanů německé národnosti.

Nařízení a závazné pokyny německých orgánů v průběhu války vedly rovněž k výraznému omezení dostupnosti statistických údajů, snižování počtu oprávněných příjemců statistických materiálů a publikací ÚSÚ, které později byly vydávány jen v němčině. Vydávání některých statistických publikací ke konci války bylo rovněž zastaveno. Docházelo také k postupnému zvyšování stupně utajení statistických publikací. Statistické ročenky ÚSÚ byly vydávány jako tajné publikace a příjemci měli povinnost zacházet s nimi jako s tajnými dokumenty.

Zhoršující se válečná situace Velkoněmecké říše vedla rovněž k redukci zaměstnanců ÚSÚ, k odchodu pracovníků a jejich nasazení do jiných válečně důležitých odvětví. V důsledku snížené pracovní kapacity řada statistik v posledním období války zůstala ve statistickém úřadě nezpracována. Počet stálých pracovníků úřadu dosahoval k 30. 4. 1945 jen 415 osob, tj. zhruba poloviny celkového stavu v roce 1939.

Konec činnosti ÚSÚ přinesla Pražská květnová revoluce, která propukla 5. května 1945 a do které se aktivně zapojilo i české osazenstvo úřadu. Německé vedení ÚSÚ /dr. H. Wirth a dr. A. Oberschall/ bylo zajištěno v budově úřadu a předáno české policii. Řízením statistického úřadu byl již v době Pražské revoluce pověřen Českou národní radou dr. F. Fajfr, významný statistický odborník, který obnovený Státní statistický úřad řídil až do 26. června 1961.

Až po ukončení druhé světové války v Evropě v květnu 1945 mohla statistická služba v obnovené Československé republice zjišťovat údaje o škodách



a dalších důsledcích války i publikovat dříve utajované informace. Některé z těchto statistik působí otřesně, např. statistiky vedené statistickou službou zřízenou v Terezínském ghettu o českých občanech židovské národnosti<sup>7</sup> a dalších soustředěných v Terezíně z Německa i z jiných okupovaných zemí, o vývoji jejich počtu v průběhu války, úmrtích a transportech do vyhlazovacího tábora Osvětim „Auschwitz“. Z celkového počtu stočtyřiceti tisíc deportovaných do Terezínského ghetta, devadesát tisíc bylo dopraveno do vyhlazovacího tábora, třicet tři tisíc zemřelo v ghettu. Podle posledních statistických údajů oko sedmnácti tisíc deportovaných zůstalo na živu k 30. dubnu 1945.

Také někteří pracovníci statistické služby zaplatili cenu nejvyšší. Počet 15 osob, jejichž život byl ukončen násilnou smrtí, zveřejněný po válce, zahrnoval všechny pracovníky, včetně smluvníků a důstojníků bývalého MNO, kteří byli převedeni na práce ve statistickém úřadě. Zvláštní zmínku zaslouží i případ doc. Dr. J. Auerhana, bývalého prezidenta SÚS, který byl zatčen dva dny po smrti zastupujícího Říšského protektora R. Heydricha, odsouzen k trestu smrti a zastřelen 9. června 1942 spolu s dalšími českými vlastenci s odůvodněním, že „schvalovali atentát na R. Heydricha a vyzývali k podpoře pachatelů“.

S osvobozením Československa a obnovou Československé státnosti na celém převálečném území ČSR vyvstaly před státní statistickou službou nové náročné úkoly a práce. Tyto mimořádné a náročné akce, mezi které patřila i příprava a provedení soupisů obyvatelstva v letech 1946 a 1947, se podařilo pracovníkům Československé statistické služby i přes složité poválečné poměry úspěšně zvládnout.

---

<sup>7</sup>Předseda protektorátní vlády ing. A. Eliáš odmítl zásadně protižidovské norimberské zákony na území Protektorátu uplatnit.

## Osnova

*Gejza Dohnal*

Zpráva o činnosti České statistické společnosti v roce 2011 ..... 1

*Zdeněk Půlpán*

Validita dotazníku ..... 3

*Martin Veselý*

Rozdělení vzdáleností vlastních čísel hermitovských náhodných matic typu  $2 \times 2$  nad Cayley-Dicksonovými algebrami ..... 10

*Hidetoshi Murakami*

Vícerozměrný neparametrický dvouvýběrový test ..... 18

*Jaroslav Češka*

65. výročí konce druhé světové války a protektorátní statistická služba ... 26

**Informační Bulletin** České statistické společnosti vychází čtyřikrát do roka v českém vydání. Příležitostně i mimořádné české a anglické číslo.

Časopis je zařazen do seznamu Rady pro výzkum, vývoj a inovace, více viz server <http://www.vyzkum.cz/>

The Bulletin of the Czech Statistical Society is published quarterly. Most of the contributions are published in Czech and Slovak languages.

**Předseda společnosti:** doc. RNDr. Gejza DOHNAL, CSc.  
ÚTM FS ČVUT v Praze, Karlovo náměstí 13, 121 35 Praha 2  
E-mail: [gejza.dohnal@fs.cvut.cz](mailto:gejza.dohnal@fs.cvut.cz)

**Redakční rada:** prof. Ing. Václav ČERMÁK, DrSc. (předseda), prof. RNDr. Jaromír ANTOCH, CSc., doc. Ing. Josef TVRDÍK, CSc., RNDr. Marek MALÝ, CSc., doc. RNDr. Jiří MICHÁLEK, CSc., doc. RNDr. Zdeněk KARPÍŠEK, CSc., prof. Ing. Jiří MILITKÝ, CSc., doc. RNDr. Gejza Dohnal, CSc.

**Technický redaktor:** Ing. Pavel STRÍŽ, Ph.D., [pavel@striz.cz](mailto:pavel@striz.cz)  
Informace pro autory jsou na stránkách <http://www.statspol.cz/>

**DOI:** 10.5300/IB, <http://dx.doi.org/10.5300/IB>  
**ISSN 1210–8022 (Print), ISSN 1804–8617 (Online)**

Toto číslo bylo vytištěno s laskavou podporou Českého statistického úřadu.