

**PRACE NAUKOWE**

Uniwersytetu Ekonomicznego we Wrocławiu

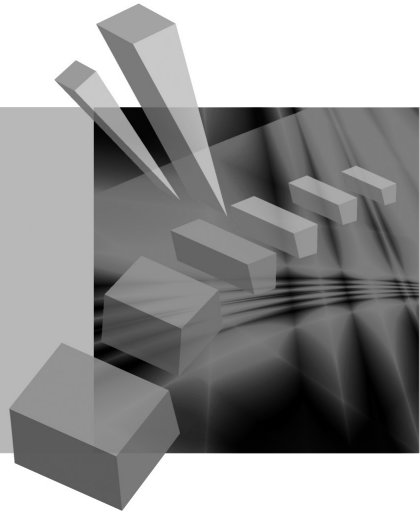
**RESEARCH PAPERS**

of Wrocław University of Economics

**278**

# **Taksonomia 20**

## **Klasyfikacja i analiza danych – teoria i zastosowania**



Redaktorzy naukowi

**Krzysztof Jajuga**

**Marek Walesiak**



Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu  
Wrocław 2013

Redaktor Wydawnictwa: Aleksandra Śliwka

Redaktor techniczny: Barbara Łopusiewicz

Korektor: Barbara Cibis

Łamanie: Małgorzata Czupryńska

Projekt okładki: Beata Dębska

Publikacja jest dostępna w Internecie na stronach:

[www.ibuk.pl](http://www.ibuk.pl), [www.ebscohost.com](http://www.ebscohost.com),

The Central and Eastern European Online Library [www.ceeol.com](http://www.ceeol.com),

a także w adnotowanej bibliografii zagadnień ekonomicznych BazEkon

[http://kangur.uek.krakow.pl/bazy\\_ae/bazekon/nowy/index.php](http://kangur.uek.krakow.pl/bazy_ae/bazekon/nowy/index.php)

Informacje o naborze artykułów i zasadach recenzowania znajdują się

na stronie internetowej Wydawnictwa

[www.wydawnictwo.ue.wroc.pl](http://www.wydawnictwo.ue.wroc.pl)

Tytuł dofinansowany ze środków Narodowego Banku Polskiego

oraz ze środków Sekcji Klasyfikacji i Analizy danych PTS

Kopiowanie i powielanie w jakiegokolwiek formie

wymaga pisemnej zgody Wydawcy

© Copyright by Uniwersytet Ekonomiczny we Wrocławiu

Wrocław 2013

**ISSN 1899-3192** (Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu)

**ISSN 1505-9332** (Taksonomia)

Wersja pierwotna: publikacja drukowana

Druk: Drukarnia TOTEM

## Spis treści

<b>Wstęp</b> .....	9
<b>Józef Pocięcha:</b> Wskaźniki finansowe a klasyfikacyjne modele predykcji upadłości firm .....	15
<b>Eugeniusz Gatnar:</b> Analiza miar adekwatności rezerw walutowych .....	23
<b>Marek Walesiak:</b> Zagadnienie doboru liczby klas w klasyfikacji spektralnej .....	33
<b>Joanicjusz Nazarko, Joanna Ejdyś, Anna Kononiuk, Anna M. Olszewska:</b> Analiza strukturalna jako metoda klasyfikacji danych w badaniach foresight .....	44
<b>Andrzej Bąk:</b> Metody porządkowania liniowego w polskiej taksonomii – pakiet <code>pllord</code> .....	54
<b>Aleksandra Łuczak, Feliks Wysocki:</b> Zastosowanie mediany przestrzennej Webera i metody TOPSIS w ujęciu pozycyjnym do konstrukcji syntetycznego miernika poziomu życia .....	63
<b>Ewa Roszkowska:</b> Zastosowanie rozmytej metody TOPSIS do oceny ofert negocjacyjnych .....	74
<b>Jacek Batóg:</b> Analiza wrażliwości metody ELECTRE III na obserwacje nietypowe i zmianę wartości progowych .....	85
<b>Jerzy Korzeniewski:</b> Modyfikacja metody HINoV selekcji zmiennych w analizie skupień .....	93
<b>Małgorzata Markowska, Danuta Strahl:</b> Wykorzystanie referencyjnego systemu granicznego do klasyfikacji europejskiej przestrzeni regionalnej ze względu na filar inteligentnego rozwoju – kreatywne regiony .....	101
<b>Elżbieta Sobczak:</b> Inteligentne struktury pracujących a efekty strukturalne zmian zatrudnienia w państwach Unii Europejskiej.....	111
<b>Elżbieta Gołata, Grażyna Dehnel:</b> Rozbieżności szacunków NSP 2011 i BAEL.....	120
<b>Iwona Foryś:</b> Wykorzystanie analizy historii zdarzeń do badania powtórnego sprzedaży na lokalnym rynku mieszkaniowym .....	131
<b>Hanna Dudek, Joanna Landmesser:</b> Wpływ relatywnej deprivacji na subiektywne postrzeganie dochodów.....	142
<b>Grażyna Łaska:</b> Syntaksonomia numeryczna w klasyfikacji, identyfikacji i analizie przemian zbiorowisk roślinnych .....	151
<b>Magdalena Osińska, Marcin Faldziński, Tomasz Zdanowicz:</b> Analiza zależności między procesami fundamentalnymi a rynkiem kapitałowym w Chinach .....	161

<b>Andrzej Bąk, Tomasz Bartłomowicz:</b> Mikroekonometryczne modele wielomianowe i ich zastosowanie w analizie preferencji z wykorzystaniem programu R .....	169
<b>Andrzej Dudek, Bartosz Kwaśniewski:</b> Przetwarzanie równoległe algorytmów analizy skupień w technologii CUDA .....	180
<b>Michał Trzęsiok:</b> Wycena rynkowej wartości nieruchomości z wykorzystaniem wybranych metod wielowymiarowej analizy statystycznej .....	188
<b>Joanna Trzęsiok:</b> Wybrane symulacyjne techniki porównywania nieparametrycznych metod regresji.....	197
<b>Artur Mikulec:</b> Kryterium Mojeny i Wisharta w analizie skupień – przypadek skupień o różnych macierzach kowariancji .....	206
<b>Artur Zaborski:</b> Analiza <i>unfolding</i> z wykorzystaniem modelu grawitacji ....	216
<b>Justyna Wilk:</b> Identyfikacja obszarów problemowych i wzrostowych w województwie dolnośląskim w zakresie kapitału ludzkiego .....	225
<b>Karolina Bartos:</b> Analiza ryzyka odejścia studenta z uczelni po uzyskaniu dyplomu licencjata – zastosowanie sieci MLP .....	236
<b>Ewa Genge:</b> Segmentacja uczestników Industriady z wykorzystaniem analizy klas ukrytych .....	246
<b>Izabela Kurzawa:</b> Wielomianowy model logitowy jako narzędzie identyfikacji czynników wpływających na sytuację mieszkaniową polskich gospodarstw domowych .....	254
<b>Marek Lubicz, Maciej Zięba, Konrad Pawelczyk, Adam Rzechonek, Jerzy Kołodziej:</b> Modele eksploracji danych niezbilansowanych – procedury klasyfikacji dla zadania analizy ryzyka operacyjnego.....	262
<b>Aleksandra Łuczak:</b> Zastosowanie rozmytej hierarchicznej analizy w tworzeniu strategii rozwoju jednostek administracyjnych .....	271
<b>Marcin Pelka:</b> Rozmyta klasyfikacja spektralna <i>c</i> -średnich dla danych symbolicznych interwałowych.....	282
<b>Małgorzata Machowska-Szewczyk:</b> Klasyfikacja obiektów reprezentowanych przez różnego rodzaju cechy symboliczne .....	290
<b>Ewa Chodakowska:</b> Indeks Malmquista w klasyfikacji podmiotów gospodarczych według zmian ich względnej produktywności działania .....	300
<b>Beata Bieszk-Stolorz, Iwona Markowicz:</b> Wykorzystanie modeli proporcjonalnego i nieproporcjonalnego hazardu Coxa do badania szansy podjęcia pracy w zależności od rodzaju bezrobocia .....	311
<b>Marcin Salamaga:</b> Weryfikacja teorii poziomu rozwoju gospodarczego J.H. Dunninga w ujęciu sektorowym w wybranych krajach Unii Europejskiej .....	321
<b>Justyna Wilk, Michał Bernard Pietrzak, Stanisław Matusik:</b> Sytuacja społeczno-gospodarcza jako determinanta migracji wewnętrznych w Polsce. ....	330
<b>Hanna Gruchociak:</b> Delimitacja lokalnych rynków pracy w Polsce na podstawie danych z badania przepływów ludności związanych z zatrudnieniem .....	343

<b>Radosław Pietrzyk:</b> Efektywność inwestycji polskich funduszy inwestycyjnych z tytułu doboru papierów wartościowych i umiejętności wykorzystania trendów rynkowych .....	351
<b>Sabina Denkowska:</b> Procedury testowań wielokrotnych .....	362

## Summaries

<b>Józef Pocięcha:</b> Financial ratios and classification models of bankruptcy prediction .....	22
<b>Eugeniusz Gatnar:</b> Analysis of FX reserve adequacy measures .....	32
<b>Marek Walesiak:</b> Automatic determination of the number of clusters using spectral clustering .....	43
<b>Joanicjusz Nazarko, Joanna Ejdys, Anna Kononiuk, Anna M. Olszewska:</b> Structural analysis as a method of data classification in foresight research .....	53
<b>Andrzej Bąk:</b> Linear ordering methods in Polish taxonomy – pllord package .....	62
<b>Aleksandra Łuczak, Feliks Wysocki:</b> The application of spatial median of Weber and the method TOPSIS in positional formulation for the construction of synthetic measure of standard of living .....	73
<b>Ewa Roszkowska:</b> Application of the fuzzy TOPSIS method to the estimation of negotiation offers.....	84
<b>Jacek Batóg:</b> Sensitivity analysis of ELECTRE III method for outliers and change of thresholds .....	92
<b>Jerzy Korzeniewski:</b> Modification of the HINoV method of selecting variables in cluster analysis .....	100
<b>Małgorzata Markowska, Danuta Strahl:</b> Implementation of reference limit system for the European regional space classification regarding smart growth pillar – creative regions .....	110
<b>Elżbieta Sobczak:</b> Smart workforce structures versus structural effects of employment changes in the European Union countries .....	119
<b>Elżbieta Gołata, Grażyna Dehnel:</b> Divergence in National Census 2011 and LFS estimates.....	130
<b>Iwona Foryś:</b> Event history analysis in the resale study on the local housing market .....	141
<b>Hanna Dudek, Joanna Landmesser:</b> Impact of the relative deprivation on subjective income satisfaction .....	150
<b>Grażyna Łaska:</b> Numerical syntaxonomy in classification, identification and analysis of changes of secondary communities .....	160
<b>Magdalena Osińska, Marcin Faldziński, Tomasz Zdanowicz:</b> Analysis of relations between fundamental processes and capital market in China.....	166
<b>Andrzej Bąk, Tomasz Bartłomowicz:</b> Microeconomic polynomial models and their application in the analysis of preferences using R program.....	179

<b>Andrzej Dudek, Bartosz Kwaśniewski:</b> Parallel processing of clustering algorithms in CUDA technology .....	187
<b>Michał Trzęsiok:</b> Real estate market value estimation based on multivariate statistical analysis .....	196
<b>Joanna Trzęsiok:</b> On some simulative procedures for comparing nonparametric methods of regression.....	205
<b>Artur Mikulec:</b> Mojena and Wishart criterion in cluster analysis – the case of clusters with different covariance matrices .....	215
<b>Artur Zaborski:</b> Unfolding analysis by using gravity model .....	224
<b>Justyna Wilk:</b> Determination of problem and growth areas in Dolnośląskie Voivodship as regards human capital.....	235
<b>Karolina Bartos:</b> Risk analysis of bachelor students' university abandonment – the use of MLP networks .....	245
<b>Ewa Genge:</b> Clustering of industrial holiday participants with the use of latent class analysis.....	253
<b>Izabela Kurzawa:</b> Multinomial logit model as a tool to identify the factors affecting the housing situation of Polish households.....	261
<b>Marek Lubicz, Maciej Zięba, Konrad Pawelczyk, Adam Rzechonek, Jerzy Kołodziej:</b> Modelling class imbalance problems: comparing classification approaches for surgical risk analysis .....	270
<b>Aleksandra Łuczak:</b> The application of fuzzy hierarchical analysis to the evaluation of validity of strategic factors in administrative districts.....	281
<b>Marcin Pełka:</b> A spectral fuzzy c-means clustering algorithm for interval-valued symbolic data .....	289
<b>Małgorzata Machowska-Szewczyk:</b> Clustering algorithms for mixed-feature symbolic objects .....	299
<b>Ewa Chodakowska:</b> Malmquist index in enterprises classification on the basis of relative productivity changes .....	310
<b>Beata Bieszk-Stolorz, Iwona Markowicz:</b> Using proportional and non proportional Cox hazard models to research the chances for taking up a job according to the type of unemployment .....	320
<b>Marcin Salamaga:</b> Verification J.H. Dunning's theory of economic development by economic sectors in some EU countries .....	329
<b>Justyna Wilk, Michał Bernard Pietrzak, Stanisław Matusik:</b> Socio-economic situation as a determinant of internal migration in Poland .....	342
<b>Hanna Gruchociak:</b> Delimitation of local labor markets in Poland on the basis of the employment-related population flows research.....	350
<b>Radosław Pietrzyk:</b> Selectivity and timing in Polish mutual funds performance measurement .....	361
<b>Sabina Denkowska:</b> Multiple testing procedures.....	369

**Sabina Denkowska**

Uniwersytet Ekonomiczny w Krakowie

---

## PROCEDURY TESTOWAŃ WIELOKROTNYCH

---

**Streszczenie:** W artykule przedstawiono procedury testowań wielokrotnych, a szczególnie procedury łączne, które uwzględniają łączne rozkłady statystyk testowych, dzięki czemu są mniej konserwatywne od procedur brzegowych. W przykładzie empirycznym do porównań parami przeciętnych wynagrodzeń brutto wybranych grup zawodowych w modelu zrównoważonej jednoczynnikowej analizy wariancji zastosowano klasyczną procedurę Tukeya, wybrane procedury brzegowe oraz procedury łączne testowań wielokrotnych.

**Słowa kluczowe:** testowanie wielokrotne, FWER, repróbkiwanie.

### 1. Wstęp

W wielowymiarowej analizie danych często mamy do czynienia z testowaniem wielokrotnym. Typową sytuacją jest porównywanie parami wartości przeciętnych w celu wyodrębnienia jednorodnych podgrup wartości przeciętnych w sytuacji, gdy analiza wariancji spowoduje odrzucenie hipotezy zerowej o równości wartości przeciętnych we wszystkich grupach. W tym przypadku rozwiązaniem mogą okazać się klasyczne procedury *post-hoc* powszechnie dostępne w pakietach statystycznych. W przypadku modelu zrównoważonej analizy wariancji zalecane w literaturze tematu jest zastosowanie procedury Tukeya<sup>1</sup>. Testowanie wielokrotne to jednak nie tylko porównywanie parami wartości przeciętnych, gdy ANOVA spowoduje odrzucenie hipotezy zerowej, to również powszechne w statystycznych analizach wielowymiarowych testowanie istotności współczynników korelacji w macierzach korelacji czy testowanie istotności współczynników regresji w modelu regresji wielorakiej. Niestety, nadal często zdarza się, że liczne testowania są prowadzone każde na poziomie istotności  $\alpha$ , a wnioski są podsumowywane łącznie, a przecież wraz ze wzrostem liczby rozpatrywanych hipotez rośnie prawdopodobieństwo wykrycia pozornie istotnych statystycznie związków. Jeśli rozważymy teoretycznie testowanie  $m$  prawdziwych, niezależnych hipotez zerowych, każdą na poziomie istotności  $\alpha$ , to prawdopodobieństwo odrzucenia przynajmniej jednej prawdziwej hipotezy zerowej wynosi

---

<sup>1</sup> Procedura typu *single-step* uwzględniająca zależności pomiędzy statystykami testowymi, oparta na studentyzowanym rozstępie [Hochberg, Tamhane 1987; Bretz, Hothorn, Westfall 2011].

$1 - (1 - \alpha)^m$ . Już w przypadku 20 niezależnych, prawdziwych hipotez zerowych, testowanych każda na poziomie istotności 0,05, prawdopodobieństwo odrzucenia co najmniej jednej prawdziwej hipotezy wynosi 0,64, a wartość oczekiwana liczby błędnych odrzuceń wynosi 1. W praktyce niezmiernie rzadko mamy do czynienia z niezależnymi testowaniami, co znacznie utrudnia kontrolę efektu testowania wielokrotnego.

Celem artykułu jest przedstawienie zarówno zalet, jak i wad procedur testowań wielokrotnych umożliwiających kontrolę efektu testowania wielokrotnego, a szczególnie procedur łącznych, które uwzględniają łączny rozkład statystyk testowych, dzięki czemu są mniej konserwatywne od procedur brzegowych.

## 2. Miary błędu I rodzaju dla rodziny wnioskowań

W celu wprowadzenia najczęściej spotykanych miar błędu I rodzaju dla rodziny wnioskowań przyjmijmy pomocniczo następujące oznaczenia. Niech  $V$  oznacza zmienną losową określającą liczbę prawdziwych hipotez zerowych odrzuconych w procesie testowania  $m$  hipotez zerowych, a  $R$  – zmienną losową określającą liczbę odrzuconych hipotez zerowych.

### Wybrane miary błędu I rodzaju dla rodzin wnioskowań:

- **FWER** (Family-Wise Error Rate):  $FWER = P(V > 0)$ ,
- **gFWER** (generalized FWER):  $gFWER = P(V > k), k = 0, \dots, m$ ,
- **FDR** (False Discovery Rate): 
$$FDR = \begin{cases} E\left(\frac{V}{R}\right) & \text{dla } R > 0 \\ 0 & \text{dla } R = 0 \end{cases}$$
.

Miara FWER nawiązuje do tradycyjnego podejścia, gdy chcemy się ustrzec błędnych odrzuceń prawdziwych hipotez zerowych. W 1953 r. w monografii pt. *The Problem of Multiple Comparisons* Tukey porównywał różne miary kontroli błędu I rodzaju dla rodziny wnioskowań, podsumował je wnioskiem, iż „kontrola FWE powinna być standardem” [Hochberg, Tamhane 1987]. Niestety, wraz ze wzrostem liczby weryfikowanych hipotez maleje moc procedur kontrolujących FWER rozumiana jako zdolność procedur do wykrywania fałszywych hipotez zerowych. Lehmann i Romano [2005] zaproponowali kontrolę miary gFWER, będącej uogólnieniem miary FWER. Nie rozwiązuje to jednak problemu spadku mocy procedur w przypadku bardzo licznych zbiorów, złożonych z tysięcy, milionów wnioskowań, np. w genetyce. Indywidualne poziomy istotności są tak małe, iż rzadko dochodzi do odrzuceń hipotez zerowych. W przypadkach bardzo licznych zbiorów wnioskowań warto rozważyć kontrolę FDR [Benjamini, Hochberg 1995], czyli wartości oczekiwanej frakcji błędnych odrzuceń pośród wszystkich odrzuceń hipotez zerowych.

Niezwykle wygodnym pojęciem w teorii testowań wielokrotnych są skorygowane prawdopodobieństwa testowe. Analogicznie do zwykłych prawdopodobieństw testowych (*p-value*) **skorygowanym prawdopodobieństwem testowym** (*adjusted*



$p$ -value)  $\tilde{p}_i$  dla dowolnej hipotezy  $H_{0,i}$  vs.  $H_{A,i}$  nazywamy najmniejszą wartość FWER, przy której dana hipoteza zerowa  $H_{0,i}$  zostałaby odrzucona, gdy rozpatrywana jest cała rodzina hipotez. Analogicznie są definiowane skorygowane prawdopodobieństwa testowe w przypadku innych miar błędu I rodzaju dla rodziny wnioskowań.

Skorygowane prawdopodobieństwa mają liczne zalety. Są łatwe do interpretacji, gdyż mając podane ich wartości, decyżję o ewentualnym odrzuceniu hipotezy podejmujemy, porównując odpowiadające jej skorygowane prawdopodobieństwo testowe z przyjętym łącznym poziomem istotności dla całej rodziny wnioskowań. Wskazują, jak mocne są podstawy do odrzucenia hipotezy zerowej w kontekście kontroli wybranej miary błędu I rodzaju dla całego zbioru wnioskowań. Można również łatwo porównywać różne procedury, porównując ich prawdopodobieństwa skorygowane (mniejsze wartości skorygowanych prawdopodobieństw testowych wskazują na mniej konserwatywną procedurę).

### 3. Procedury brzegowe testowań wielokrotnych kontrolujące FWER

W ostatnich latach znaczną popularność zyskały proste obliczeniowo, o szerokim zakresie zastosowań (mogą być stosowane w przypadku porównań wartości przeciętnych, badania istotności współczynników korelacji czy współczynników regresji) brzegowe procedury testowań wielokrotnych. Proces testowania w przypadku tych procedur polega na analizie zbioru prawdopodobieństw testowych otrzymanych z indywidualnych wnioskowań.

Najstarszą, a zarazem najprostszą procedurą brzegową testowań wielokrotnych jest procedura Bonferroniego. Procedura Bonferroniego jest procedurą uniwersalną, czyli można ją stosować w przypadku dowolnej rodziny wnioskowań, bez względu na typ zależności pomiędzy statystykami testowymi. Metoda Bonferroniego jest metodą bardzo konserwatywną, czyli metodą o małej mocy. Konserwatyzm ten jest tym poważniejszy, im silniejsze są zależności pomiędzy statystykami testowymi lub im liczniejsza jest rodzina wnioskowań. Mniej konserwatywna jest uniwersalna procedura Holma, która jest wieloetapową modyfikacją procedury Bonferroniego. Każda hipoteza odrzucona przez metodę Bonferroniego jest odrzucona również przez metodę Holma, natomiast hipotezy odrzucone przez metodę Holma mogą nie zostać odrzucone przez metodę Bonferroniego.

W przypadku, gdy rozpatrywane statystyki testowe tworzą wielowymiarowy rozkład normalny lub rozkład  $t$ -Studenta o niezależnych składowych [Shaffer 1995], a rozważane hipotezy alternatywne mają dwustronne zbiory krytyczne, do kontroli efektu testowania wielokrotnego można zastosować modyfikację procedury Bonferroniego oraz procedury Holma oparte na nierówności Šidáka [Hochberg, Tamhane 1987, s. 366; Westfall i in. 1999; Denkowska 2005]. Jak wykazali B. Holland,

M.D. Copenhaver [1987], procedury ŠidákaSS (*single-step*) oraz ŠidákaSD (*step-down*) zapewniają kontrolę FWE również w przypadku, gdy statystyki testowe mają dodatnią zależność orthantową [Denuit, Scaillet 2004; Shaffer 1995].

Pośród procedur brzegowych największą moc mają procedury wieloetapowe typu *step-up*, zapewniające kontrolę FWER w przypadku statystyk testowych niezależnych lub silnie dodatnio skorelowanych. Skomplikowana obliczeniowo procedura Hommela daje nieznacznie lepsze wyniki od procedury Hochberga.

Wymieniając najważniejsze procedury brzegowe, należy wspomnieć o procedurze Shaffer [1986] dla hipotez logicznie powiązanych. Przykładem hipotez logicznie powiązanych mogą być hipotezy o równości wartości przeciętnych parametrów dla co najmniej trzech populacji. Zauważmy, że w rzeczywistości niemożliwe jest, aby  $\mu_1 = \mu_2$  oraz  $\mu_2 = \mu_3$ , ale  $\mu_1 \neq \mu_3$ . J. Shaffer uznała więc, że w przypadku porównywania wartości przeciętnych trzech populacji nie ma potrzeby rozpatrywać sytuacji, gdy odrzucamy jedną hipotezę zerową, a przy dwóch stwierdzamy, że nie mamy podstaw do ich odrzucenia i zaproponowała modyfikację uniwersalnej procedury Holma, która dzięki uwzględnieniu logicznych relacji pomiędzy hipotezami ma większą moc, a kontrola FWE na poziomie  $\alpha$  jest nadal zagwarantowana.

Podsumowując zalety brzegowych procedur testowań wielokrotnych, należy podkreślić ich szeroki zakres zastosowań, prostotę obliczeniową oraz niewielkie wymagania odnośnie do założeń modelu statystycznego. Niektóre z procedur brzegowych zapewniają kontrolę wybranej miary błędu I rodzaju tylko w przypadku pewnych typów zależności pomiędzy statystykami testowymi<sup>2</sup>. Procedury te charakteryzują się zazwyczaj większą mocą w stosunku do procedur uniwersalnych, ale wymogi dotyczące zależności pomiędzy statystykami komplikują ich użycie i ograniczają zakres zastosowań. Natomiast wadą uniwersalnych procedur brzegowych jest fakt, iż nie uwzględniają łącznego rozkładu statystyk testowych, przez co mają mniejszą moc od procedur łącznych.

#### 4. Procedury łączne Westfalla i Younga kontrolujące FWER

Procedury zaproponowane przez Westfalla i Younga [1993] oparte na regule domknięcia [Domański, Pruska 2000, s. 201; Hochberg, Tamhane 1987] wykorzystują repróbkiowanie. Resampling umożliwia przeprowadzanie testowania wielokrotnego mimo braku normalności czy też braku znajomości struktury kowariancyjnej danych. Procedury Westfalla i Younga oparte są na maksimach statystyk testowych lub minimach prawdopodobieństw testowych. Wadą tych procedur jest wymóg „obrotowości podzbioru” (*subset pivotality*), który w przypadku procedur opartych na maksimach statystyk testowych oznacza, że rozkład maksimum statystyk testowych dla dowolnego podzbioru  $I$  zbioru wszystkich rozważanych wnioskowań  $\{1, \dots, m\}$ , musi być taki sam zarówno w przypadku prawdziwości wszystkich hipotez zero-

<sup>2</sup> Np. procedury ŠidákaSS, ŠidákaSD, Hommela, Hochberga zapewniają kontrolę FWER w przypadku typów zależności wymienionych powyżej w artykule.

wych  $H_i$  dla  $i \in I$ , jak i w przypadku prawdziwości globalnej hipotezy zerowej  $H_0^C$ , głoszącej, że wszystkie hipotezy zerowe  $H_i (i \in \{1, \dots, m\})$  są prawdziwe.

Warunek obrotowości podzbioru jest bardzo istotny, zwłaszcza gdy resampling wykorzystuje rozkład generujący dane przy założeniu prawdziwości wszystkich hipotez zerowych, pozwala to bowiem uprościć algorytm procedury opartej na regule domknięcia i zamiast testować  $2^m - 1$  przecięć hipotez zerowych, wystarczy przeprowadzić  $m$  testowań. Niestety w wielu sytuacjach badawczych warunek ten nie jest spełniony. Należy do nich np. testowanie istotności współczynników korelacji. Rozkład generujący dane przy założeniu prawdziwości hipotez zerowych może dawać łączny rozkład statystyk testowych inny od prawdziwego (rzeczywistego) rozkładu. Rozważmy badanie istotności trzech współczynników korelacji  $\rho_{12}, \rho_{13}, \rho_{23}$ . Aitken wykazał [Westfall, Young 1993], że gdy  $H_{0,12}$  oraz  $H_{0,13}$  są prawdziwe, a  $H_{0,23}$  jest fałszywa, to łączny rozkład statystyk testowych odpowiadających prawdziwym hipotezom zerowym jest w przybliżeniu normalny, zależny od współczynnika korelacji  $\rho_{23}$ , czyli warunek obrotowości podzbioru nie jest spełniony.

Procedury Westfalla i Younga uwzględniają łączny rozkład statystyk testowych i dzięki temu mają większą moc niż brzegowe procedury testowań wielokrotnych. Procedury te niekoniecznie muszą się opierać na resamplingu, gdy rozkład statystyk testowych jest znany i jest to np. wielowymiarowy rozkład normalny lub  $t$ -Studenta.

#### 4.1. Przykład empiryczny

W przykładzie do porównywania parami wartości przeciętnych w modelu zrównoważonej analizy wariancji zastosowano procedury łączne oparte na maksimach statystyk testowych, procedurę „free”<sup>3</sup> oraz Westfalla [1997] dla hipotez logicznie powiązanych<sup>4</sup>. Otrzymane wyniki porównano z wynikami otrzymanymi za pomocą klasycznej procedury Tukeya oraz za pomocą wybranych procedur brzegowych<sup>5</sup>. Przykład empiryczny dotyczył porównywania przeciętnych wynagrodzeń brutto w Polsce w 2010 r. w pięciu grupach zawodowych, takich jak<sup>6</sup>: rolnictwo, leśnictwo, łowiectwo i rybactwo ( $r$ ); przemysł ( $p$ ); przetwórstwo przemysłowe ( $pp$ ); edukacja ( $e$ ); opieka zdrowotna i pomoc społeczna ( $z$ ). Dane dotyczące przeciętnych miesięcznych wynagrodzeń brutto dla województw Polski zaczerpnięto z „Rocznika Statystycznego Rzeczypospolitej Polskiej 2011”<sup>7</sup>. W badaniu przyjęto poziom istotności 0,01.

<sup>3</sup> Procedura „free” jest procedurą step-down Westfalla i Younga [1993] wykorzystującą fakt, iż łączny rozkład statystyk testowych w modelu zrównoważonej analizy wariancji jest wielowymiarowym rozkładem normalnym lub  $t$ -Studenta. Dostępna poprzez metodę *summary* w pakiecie *multcomp* w R.

<sup>4</sup> Modyfikacja procedury maxT Westfalla i Younga [1993]. Dostępna poprzez metodę *summary* w pakiecie *multcomp* w R.

<sup>5</sup> Skorygowane prawdopodobieństwa testowe dla procedur brzegowych wyznaczono za pomocą funkcji *mt.rawp2adjp* dostępnej w pakiecie *multtest* w R.

<sup>6</sup> W nawiasach podano oznaczenia grup, które przyjęto w badaniu.

<sup>7</sup> [http://www.stat.gov.pl/cps/rde/xbr/gus/rs\\_rocznik\\_statystyczny\\_rp\\_2011.pdf](http://www.stat.gov.pl/cps/rde/xbr/gus/rs_rocznik_statystyczny_rp_2011.pdf).

**Tabela 1.** Surowe prawdopodobieństwa testowe oraz skorygowane prawdopodobieństwa testowe wyznaczone dla procedury klasycznej Tukeya, procedur brzegowych Holma, Šidáka SS, Šidáka SD oraz procedur łącznych „free”, Westfalla

Porównania	Surowe prawd. $p_i$	Skorygowane prawdopodobieństwa testowe $\tilde{p}_i$						
		Tukey	Holm	Shaffer	Šidák SS	Šidák SD	„free”	Westfall
$r - pp$	0,000000	<0,001	<0,001	<0,001	<0,001	<0,001	<0,001	<0,001
$pp - e$	0,000000	<0,001	<0,001	<0,001	<0,001	<0,001	<0,001	<0,001
$r - p$	0,000331	<b>0,00299</b>	<b>0,00265</b>	<b>0,0019</b>	<b>0,0033</b>	<b>0,0026</b>	<b>0,0024</b>	<b>0,0018</b>
$p - e$	0,000833	<b>0,00727</b>	<b>0,00583</b>	<b>0,0025</b>	<b>0,0083</b>	<b>0,0058</b>	<b>0,0053</b>	0,0024
$z - r$	0,001224	0,01039	<b>0,00734</b>	<b>0,0049</b>	0,0122	<b>0,0073</b>	<b>0,0067</b>	<b>0,0047</b>
$z - pp$	0,002567	0,02103	0,01283	0,0103	0,0254	0,0128	0,0117	<b>0,0096</b>
$z - e$	0,002890	0,02348	0,01283	0,0103	0,0285	0,0128	0,0117	<b>0,0096</b>
$pp - p$	0,008152	0,06062	0,02446	0,0163	0,0786	0,0243	0,0234	0,0162
$z - p$	0,688864	0,99439	1,00000	1,0000	0,9999	0,9032	0,9026	0,9026
$r - e$	0,779654	0,99861	1,00000	1,0000	0,9999	0,9032	0,9026	0,9026

Źródło: obliczenia własne (pogrubiono prawdopodobieństwa skorygowane mniejsze od 0,01).

Po sprawdzeniu założeń modelu analizy wariancji zweryfikowano hipotezę mówiącą o równości przeciętnych wynagrodzeń brutto w tych grupach zawodowych. Hipoteza ta została odrzucona ( $p = 0,000000$ ). Następnie zastosowano zalecaną w literaturze tematu klasyczną procedurę *post-hoc* Tukeya. W przypadku procedur brzegowych kontrolujących FWER zaskakująco dobrze wypadła brzegowa procedura Holma, w której skorygowane prawdopodobieństwa testowe są wyznaczane tylko na podstawie surowych prawdopodobieństw brzegowych, bez wnikania w strukturę korelacyjną danych. Klasyczna metoda Tukeya spowodowała odrzucenie czterech hipotez zerowych, a procedura Holma aż pięciu, na poziomie FWE wynoszącym 0,01. Również pięć odrzuceń, ale znacznie mniejsze prawdopodobieństwa skorygowane, otrzymano w przypadku procedury Shaffer dla hipotez logicznie powiązanych. Natomiast modyfikacje procedur Bonferroniego oraz Holma oparte na nierówności Šidáka dają nieznaną poprawę mocy w stosunku do modyfikowanej procedury.

Najlepszą procedurą spośród procedur kontrolujących FWE, która wykryła aż siedem istotnych różnic dla par:  $\{r, pp\}$ ,  $\{pp, e\}$ ,  $\{r, p\}$ ,  $\{p, e\}$ ,  $\{z, r\}$ ,  $\{z, pp\}$ ,  $\{z, e\}$  okazała się procedura łączna Westfalla [1997], wykorzystująca logiczne powiązania pomiędzy hipotezami. W przypadku procedur nieuwzględniających logicznych relacji między hipotezami najlepiej wypadła procedura łączna „free”, która mimo, iż dała tyle samo odrzuceń co inne procedury brzegowe wieloetapowe (Holma, ŠidákaSD), to prawdopodobieństwa skorygowane były najmniejsze.

## 5. Procedura łączna testowań wielokrotnych Dudoit i van der Laana

Dudoit i van der Laan [2008] przedstawili procedurę łączną, która w odróżnieniu od propozycji Westfalla i Younga nie opiera się na rozkładzie<sup>8</sup> generującym dane, który spełnia hipotezę zupełną, że wszystkie hipotezy zerowe są prawdziwe, ale na **rozkładzie „zerowym” statystyk testowych**, czyli rozkładzie statystyk testowych przy założeniu prawdziwości hipotezy zupełnej. Przedstawili dwa rodzaje „zerowych” transformacji bootstrapowych statystyk testowych: przesunięcie i skalowanie oraz kwantylowe przekształcenie. Procedura Dudoit i van der Laana jest dostępna w pakiecie *multtest* w R pod nazwą MTP. Zapewnia ona kontrolę wybranej miary błędu I rodzaju dla rodziny wnioskowań: FWER, gFWER, FDR, TPPFP<sup>9</sup>. Można ją wykorzystywać do porównań parami wartości przeciętnych, do testowania istotności dla współczynników regresji, testowania istotności współczynników korelacji oraz w wielu innych sytuacjach badawczych.

Dudoit i van der Laan [2008] zadykowali procedurę MTP badaniom genetycznym, specyficznym ze względu na bardzo liczne rodziny wnioskowań, składające się z tysięcy hipotez zerowych. Wstępne eksperymenty symulacyjne pokazują, że wskazane są dodatkowe badania potwierdzające kontrolę miar błędów I rodzaju w przypadku mniej licznych zbiorów wnioskowań.

## 6. Podsumowanie

Bezsprzecznie kontrola efektu testowań wielokrotnych jest konieczna. Zaleca się przez merytoryczną oraz formalno-statystyczną analizę maksymalnie zredukować liczbę rozpatrywanych wnioskowań oraz dążyć do wyboru procedury o największej mocy. Nieskomplikowane, o szerokim zakresie zastosowań, uniwersalne procedury brzegowe nie uwzględniają łącznego rozkładu statystyk testowych, przez co charakteryzują się gorszą mocą od procedur łącznych. Z kolei zakres zastosowań procedur łącznych zaproponowanych przez Westfalla i Younga jest ograniczony ze względu na wymóg obrotowości podzbioru. Ciekawą alternatywą zatem wydaje się dedykowana badaniom genetycznym procedura łączna zaproponowana przez Dudoit oraz van der Laana. Szeroki zakres zastosowań, możliwość wyboru miary błędu I rodzaju dla zbioru wnioskowań i dostępność w pakiecie *multtest* to jej istotne zalety. Użyteczność tej procedury w przypadku mniej licznych zbiorów wnioskowań wymaga jednak dalszych badań.

---

<sup>8</sup> Rozkład generujący dane może dać w efekcie rozkład łączny statystyk testowych o innej strukturze zależnościowej niż ich prawdziwy rozkład (gdy niespełniony jest warunek obrotowości podzbioru).

<sup>9</sup> TPPFP (Tail Probability for Proportion of False Positives):  $TPPFP = P(\frac{V}{R} > q)$ , gdzie  $q \in (0,1)$ .

## Literatura

- Benjamini Y., Hochberg Y., *Controlling the false discovery rate: a practical and powerful approach to multiple testing*, "Journal of the Royal Statistical Society", Ser. B, 1995, vol. 57, no. 1, 289-300.
- Bretz F., Hothorn T., Westfall P., *Multiple Comparisons using R*, Chapman and Hall, Boca Raton, 2011.
- Denkowska S., *Zastosowanie procedur testowań wielokrotnych opartych na uporządkowanych prawdopodobieństwach testowych do wydzielenia jednorodnych podgrup wartości przeciętnych*, „Przeгляд Statystyczny” 2005, t. 52, nr 1, 115-131.
- Denuit M., Scaillet O., *Nonparametric tests for positive quadrant dependence*, "Journal of Financial Econometrics" 2004, 2, 422-450.
- Domański Cz., Pruska K., *Nieklasyczne metody statystyczne*, PWE, Warszawa 2000.
- Dudoit S., van der Laan M., *Multiple Testing Procedures with Applications to Genomics*, Springer Series in Statistics, 2008.
- Hochberg Y., Tamhane A.C., *Multiple Comparison Procedures*, John Wiley & Sons, NY 1987.
- Holland B., Copenhaver M. D., *An improved sequentially rejective bonferroni test procedure*, "Biometrics" 1987, 43, 417-423.
- Lehmann E.L., Romano J.P., *Generalizations of the familywise error rate*, "Annals of Statistics" 2005, vol. 33, no. 3, 1138-1154.
- Shaffer J.P., *Modified sequentially rejective multiple test procedures*, "Journal of the American Statistical Association" 1986, 81, 826-831.
- Shaffer J.P., *Multiple hypothesis testing*, "Annual Review of Psychology" 1995, no. 46, 561-84.
- Westfall P.H., *Multiple testing of general contrasts using logical constraints and correlations*, "Journal of the American Statistical Association" 1997, 92, 299-306.
- Westfall P.H., Young S.S., *Resampling Based Multiple Testing*, Wiley, New York 1993.
- Westfall P.H., Tobias R.D., Rom D., Wolfinger R.D., Hochberg Y., *Multiple Comparisons and Multiple Tests, Using the SAS System*, SAS Institute Inc., 1999.

## MULTIPLE TESTING PROCEDURES

**Summary:** The article presents multiple testing procedures, in particular joint procedures taking into account the joint distribution of the test statistics, which makes them less conservative than popular marginal multiple testing procedures. The empirical example uses the Tukey's procedure, selected marginal and joint multiple testing procedures for pairwise comparisons of average gross earnings in selected professions in the balanced one-way analysis of variance model.

**Keywords:** multiple testing, FWER, resampling.