

Marcin DUDZIŃSKI<sup>1</sup>, Konrad FURMAŃCZYK<sup>2</sup>

<sup>1</sup>Katedra Ekonometrii i Statystyki SGGW  
Department of Econometrics and Statistics WAU

<sup>2</sup>Katedra Zastosowań Matematyki SGGW  
Department of Applied Mathematics WAU

## Wykorzystanie pakietu R w zagadnieniach dotyczących procedur jednoczesnego testowania wielu hipotez Application of the R-project package in the area of the multiple-comparison procedures

**Słowa kluczowe:** procedury jednoczesnego testowania wielu hipotez, miary błędów testowania wielu hipotez, pakiet R-project

**Key words:** multiple-comparison procedures, familywise error rate, false discovery rate, R-project package

### Wprowadzenie

#### Miary błędów występujących przy jednoczesnej weryfikacji wielu hipotez

Tematyka jednoczesnego testowania wielu hipotez stała się w ostatnim okresie przedmiotem intensywnych studiów i analiz. Zyskuje ona z dnia na dzień na popularności, przede wszystkim z racji swych licznych zastosowań, zwłaszcza w biotechnologii i medycynie.

Żałómy, że dokonujemy jednoczesnej weryfikacji  $m$  hipotez zerowych  $H_i$ , przy założeniach, że: 1) liczba testowanych hipotez jest duża, 2) dysponujemy

odpowiednimi testami weryfikacyjnymi, oddzielnie dla każdej z hipotez zerowych, o statystykach testowych  $T_i$ , dla  $i = 1, \dots, m$ .

Klasycznym podejściem do zagadnień dotyczących jednoczesnego testowania wielu hipotez jest konstrukcja procedury testowej, zapewniającej kontrolę miary błędu, którą to miarę stanowi prawdopodobieństwo fałszywego odrzucenia co najmniej jednej hipotezy zerowej, przy czym przez procedurę testową będziemy rozumieli regułę określającą, które spośród hipotez zerowych ( $H_i$ ) należałoby przyjąć, a które odrzucić. Ową miarę nazwano familywise error rate – w skrócie FWER.

Niech zatem  $V$  oznacza liczbę fałszywych odrzuceń. Wtedy:

$$\text{FWER} = \Pr(V \geq 1) \quad (1)$$

sama zaś procedura kontrolująca FWER przy ustalonym poziomie  $\alpha$  to procedura, która zapewnia spełnienie warunku:

$$\text{FWER} \leq \alpha \quad (2)$$

Procedura kontrolująca FWER przy założeniu, że wszystkie hipotezy zerowe są prawdziwe, zapewnia tzw. słabą kontrolę FWER, w odróżnieniu od mocnej kontroli tej miary, gdy potrafimy znaleźć ograniczenie na FWER dla wszystkich możliwych konfiguracji prawdziwych i fałszywych hipotez zerowych.

Ze względu na to, iż liczba fałszywie odrzucanych hipotez rośnie wraz ze wzrostem liczby wszystkich odrzucanych hipotez, powstał pomysł, by zamiast liczby fałszywych odrzuceń badać stosunek liczby fałszywych odrzuceń do liczby wszystkich odrzuceń. Niech  $V$  oznacza liczbę fałszywych odrzuceń, natomiast  $R$  – liczbę wszystkich odrzuceń. Benjamini i Hochberg (1995) zaproponowali rozważanie procedur kontrolujących miarę o nazwie *false discovery rate* – w skrócie FDR. Zdefiniowano ją jako:

$$\begin{aligned} \text{FDR} &:= E\left(\frac{V}{R \vee 1}\right) = \\ &= E\left(\frac{V}{R} \mid R > 0\right) \Pr(R > 0) \end{aligned} \quad (3)$$

gdzie  $R \vee 1 := \max(R, 1)$  natomiast  $E(\cdot)$  oznacza wartość oczekiwaną.

Kontrola FDR na poziomie  $\alpha$  polega na znalezieniu procedury, przy której zachodzi relacja:

$$\text{FDR} \leq \alpha \quad (4)$$

Należy pamiętać o tym, że w naszych rozważaniach zarówno prawdopodobieństwo  $\Pr(\cdot)$ , jak i wartość oczekiwana  $E(\cdot)$  są wyznaczane warunkowo – przy ustalonej konfiguracji hipotez zerowych.

## Procedury kontroli miar FWER i FDR

Założmy, że dane  $\mathbf{X}$ , rozumiane jako wynik eksperymentu i zapisane jako wektory zmiennych losowych wymiaru  $n$ , przy czym  $n$  jest rozmiarem próby, pochodzą z rozkładu  $P_\theta$  dla pewnego  $\theta \in \Theta$ , gdzie  $\Theta$  jest zbiorem parametrów. Hipotezę  $H_i$  możemy utożsamić z pewnym podzbiorem  $\Theta_i \subset \Theta$  i zapisać ją w formie następującej  $H_i: \theta \in \Theta_i$  dla  $i = 1, \dots, m$ .

Ze względu na to, iż szereg procedur jednoczesnego testowania wielu hipotez wykorzystuje pojęcia  $p$ -wartości ( $p$ -value) bądź dopasowanej  $p$ -wartości (adjusted  $p$ -value), podamy teraz definicje obu tych terminów.

Zwykłą  $p$ -wartość określamy następująco:  $p_i := \inf\{\alpha: \text{hipoteza } H_i \text{ jest odrzucana na poziomie istotności } \alpha\}$ . Innymi słowy,  $p_i$  oznacza najmniejszy poziom istotności, przy którym hipotezę zerową ( $H_i$ ) należy odrzucić.

Niech z kolei  $\mu$  oznacza dowolną miarę błędu procedury testowania wielu hipotez (np. FWER lub FDR). Wtedy dopasowaną  $p$ -wartość definiujemy następująco:  $\tilde{p}_i := \inf\{\alpha: \text{hipoteza } H_i \text{ jest odrzucana na poziomie błędu } \mu = \alpha\}$ . Dopasowana  $p$ -wartość jest zatem najmniejszą wartością rozważanej miary błędu testowania wielu hipotez, przy której hipotezę zerową ( $H_i$ ) należy odrzucić.

Przedstawimy teraz najważniejsze procedury kontroli miar. Będziemy zakładać, że weryfikujemy jednocześnie  $m$  hipotez zerowych ( $H_i$ ) dla  $i = 1, \dots, m$ .

**Procedura Bonferroniego kontroli FWER.** Procedura Bonferroniego jest procedurą, która przy ustalonym poziomie istotności  $\alpha$ , odrzuca hipotezę  $H_i$ , jeśli odpowiadająca tej hipotezie

$p$ -wartość  $p_i$  spełnia warunek  $p_i \leq \alpha/m$ , gdzie  $m$  – liczba weryfikowanych hipotez zerowych. Można pokazać, że przy pewnym założeniu na  $p$ -wartości  $p_i$ , dla  $i = 1, \dots, m$ , procedura Bonferroniego zapewnia mocną kontrolę miary FWER – wzór (1), na poziomie  $\alpha$  – wzór (2).

Warto dodać, iż w przypadku procedury Bonferroniego kontroli FWER dopasowane  $p$ -wartości są dane wzorem  $\tilde{p}_i = \min(mp_i, 1)$ .

#### Procedura Holma kontroli FWER.

Niech  $\alpha_i = \alpha/(m - i + 1)$  dla  $i = 1, \dots, m$ . Oznaczmy przez  $p_{(1)} \leq p_{(2)} \leq \dots \leq p_{(m)}$  uporządkowany ciąg  $p$ -wartości, a przez  $H_{(1)}, H_{(2)}, \dots, H_{(m)}$  – odpowiadające im hipotezy zerowe. Algorytm procedury Holma można przedstawić następująco.

Jeśli  $p_{(1)} > \alpha_1$ , to nie odrzucamy żadnej hipotezy zerowej.

W przeciwnym przypadku, gdy

$$p_{(1)} \leq \alpha_1, \dots, p_{(r)} \leq \alpha_r \quad (5)$$

odrzucamy hipotezy  $H_{(1)}, \dots, H_{(r)}$ , gdzie  $r$  jest największą liczbą, spełniającą warunek (5).

Dla procedury Holma dopasowane  $p$ -wartości są określone wzorem (Dudoit i in. 2003):

$$\tilde{p}_i = \max_{k=1, \dots, i} \{ \min[(m - k + 1)p_{(k)}, 1] \}$$

#### Procedura Hochberga kontroli

**FWER.** Niech podobnie jak w przypadku procedury Holma:  $\alpha_i = \alpha/(m - i + 1)$  dla  $i = 1, \dots, m$ ,  $p_{(1)} \leq p_{(2)} \leq \dots \leq p_{(m)}$  oznacza uporządkowany ciąg  $p$ -wartości, odpowiadających hipotezom zerowym  $H_{(1)}, H_{(2)}, \dots, H_{(m)}$ .

Algorytm procedury Hochberga można opisać w sposób następujący.

Jeśli  $p_{(m)} \leq \alpha_m$ , to nie przyjmujemy żadnej hipotezy zerowej.

W przeciwnym przypadku, gdy

$$p_{(m)} > \alpha_m, \dots, p_{(r+1)} > \alpha_{r+1} \quad (6)$$

przyjmujemy hipotezy  $H_{(r+1)}, \dots, H_{(m)}$ , gdzie  $r$  jest najmniejszą liczbą, spełniającą relację (6).

Warto nadmienić, iż w przypadku procedury Hochberga dopasowane  $p$ -wartości są określone wzorem (Dudoit i in. 2003):

$$\tilde{p}_i = \min_{k=i, \dots, m} \{ \min[(m - k + 1)p_{(k)}, 1] \}$$

#### Procedura Benjaminiego i Hochberga kontroli FDR.

Zostanie tu przedstawiona procedura kontroli FDR – wzór (3), na poziomie  $\alpha$  – wzór (4), opracowana przez Benjaminiego i Hochberga (1995). Zaproponowana przez Benjaminiego i Hochberga procedura kontroli FDR jest procedurą realizowaną w sposób identyczny jak procedura Hochberga, z tym jedynie wyjątkiem, że w przypadku procedury Benjaminiego i Hochberga ciąg  $\{\alpha_i\}$  jest dany wzorem  $\alpha_i = i\alpha/m$ , dla  $i = 1, \dots, m$ .

Dla procedury Benjaminiego i Hochberga dopasowane  $p$ -wartości są określone zależnością (Dudoit i in. 2003):

$$\tilde{p}_i = \min_{k=i, \dots, m} \left\{ \min \left( \frac{m}{k} p_{(k)}, 1 \right) \right\}$$

#### Cele pracy

W artykule przedstawimy zastosowania pakietu statystycznego R-project w zagadnieniach dotyczących procedur jednoczesnego testowania wielu hipotez oraz związanej z tym problematyki kontroli miar błędów popełnianych przy korzystaniu ze wspomnianych procedur.

W naszych rozważaniach ograniczymy się do trzech przykładów testów parametrycznych. Pierwszy z nich to test dla średniej z rozkładu normalnego, wykorzystujący jako statystykę testową odpowiednio przeskalowaną średnią arytmetyczną. Drugi test dotyczy weryfikacji hipotez o średniej z rozkładu normalnego przy użyciu statystyki *t*-Studenta. Ostatni test stanowi test *t*-Studenta o równości średnich w dwóch populacjach o normalnych rozkładach badanej cechy. Dla każdego z wymienionych rodzajów testów omówimy otrzymane przez nas rezultaty symulacji komputerowych, prezentujące liczby fałszywych odrzuceń weryfikowanych hipotez, uzyskane w wyniku zastosowania czterech wybranych przez nas procedur jednoczesnego testowania wielu hipotez. Stanowią je opisane powyżej procedury Bonferroniego, Holma, Hochberga oraz Benjaminiego i Hochberga.

### Narzędzia wykorzystywane do celów symulacji

W naszych badaniach zastosowaliśmy pakiet matematyczno-statystyczny R-project (w skrócie R). Jest on objęty licencją GPL, można pobrać go bezpłatnie ze strony pod adresem <http://www.r-project.org>. Do symulacji prób z wielowymiarowego rozkładu normalnego oraz numerycznej analizy weryfikowanych testów użyliśmy dodatkowo rozszerzeń pakietu R o nazwach Mass i Multitest.

Poniżej przedstawimy najczęściej wykorzystywane przez nas komendy pakietu R i jego wspomnianych rozszerzeń (osoby zainteresowane poznaniem wszystkich najważniejszych elementów

składni oraz poleceń środowiska R odsyłamy do opracowania Komsty (2004):

- a < – `as.vector(m)` – deklaracja wektora *m*-wymiarowego,
- b < – `matrix(, n, m)` – deklaracja macierzy o *n* wierszach i *m* kolumnach,
- c < – `rnorm(n, mi, sigma)` – generowanie *n*-elementowej próby (*n*-wymiarowego wektora) z rozkładu normalnego o średniej *mi* i odchyleniu standardowym *sigma*,
- d < – `rnorm(n)` – generowanie *n*-elementowej próby (*n*-wymiarowego wektora) ze standardowego rozkładu normalnego,
- e < – `mean(rnorm(n, mi, sigma))` – obliczanie średniej dla *n*-elementowej próby (*n*-wymiarowego wektora) z rozkładu normalnego o średniej *mi* i odchyleniu standardowym *sigma*,
- f < – `sd(rnorm(n, mi, sigma))` – obliczanie odchylenia standardowego dla *n*-elementowej próby (*n*-wymiarowego wektora) z rozkładu normalnego o średniej *mi* i odchyleniu standardowym *sigma*,
- g < – `pnorm(t)` – obliczanie wartości dystrybuanty standardowego rozkładu normalnego w punkcie *t*,
- h < – `pt(t, df=n)` – obliczanie wartości dystrybuanty rozkładu *t*-Studenta o *n* stopniach swobody w punkcie *t*,
- i < – `mvrnorm(n, mi, sigma)` – generowanie *n*-elementowej próby (*n*-wymiarowego wektora) z wielowymiarowego rozkładu normalnego o wektorze średnich *mi* oraz macierzy kowariancji *sigma* (komenda pochodzi z dodatku Mass),
- j < – `mt.rawp2adjp(p, proc=c("Bonferroni", "Holm", "Hochberg", "BH"))` – wyznaczanie, na podstawie wektora zwykłych *p*-wartości *p*, odpowiada-

jących weryfikowanym hipotezom zerowym, macierzy dopasowanych  $p$ -wartości dla czterech wybranych procedur: Bonferroniego, Holma, Hochberga, Benjaminiego i Hochberga (komenda pochodzi z dodatku Multitest),

$k < -mt.reject(adjp, alpha)$  – wyznaczanie, przy założonym poziomie błędu testowania wielu hipotez  $\alpha$ , liczby odrzuceń hipotez zerowych na podstawie macierzy (wektora) dopasowanych  $p$ -wartości  $adjp$ , otrzymanych przy użyciu polecenia  $mt.rawp2adjp$  (komenda pochodzi z dodatku Multitest).

- testujemy jednocześnie  $m$  hipotez zerowych  $H_i: \mu_i = 0$ , wobec hipotez alternatywnych, odpowiednio:  $H_i'$ :  $\mu_i \neq 0$  dla  $i = 1, \dots, m$ .

### Model 1

**Szczegółowy opis modelu 1.** Do weryfikacji hipotezy zerowej ( $H_i$ ) w modelu 1, na poziomie błędu  $\alpha$ , stosujemy statystykę testową o wzorze:

$$T_i = T_{i,n} = \sqrt{n} \left| \overline{X_{i,n}} \right|$$

$$\text{gdzie: } \overline{X_{i,n}} = \frac{1}{n} \sum_{j=1}^n X_{i,j}.$$

Hipotezę  $H_i$  odrzucamy, jeśli  $T_i \geq u_{1-\alpha/2}$ , gdzie  $u_{1-\alpha/2}$  jest kwantylem rzędu  $1 - \alpha/2$  standardowego rozkładu normalnego, tzn. wartością taką, że  $\Phi(u_{1-\alpha/2}) = 1 - \alpha/2$ , przy czym  $\Phi$  oznacza dystrybuantę standardowego rozkładu normalnego.

Odpowiednie  $p$ -wartości dla powyższych testów wyznaczamy z relacji:

$$\begin{aligned} p_i &= P(T_i \geq t_i | H_i) = \\ &= P\left(\sqrt{n} \left| \overline{X_{i,n}} \right| \geq t_i | H_i\right) = \\ &= P\left(\sqrt{n} \left| \overline{X_{i,n}} \right| \geq t_i \mid \overline{X_{i,n}} \sim N(0,1)\right) = \\ &= 2 - 2\Phi(t_i) \end{aligned}$$

**Algorytm do modelu 1.** Poniżej prezentujemy algorytm, który zastosujemy w późniejszych symulacjach dotyczących procedur jednoczesnej weryfikacji hipotez rozważanych powyżej.

1. Generujemy  $m_0$  prób  $n$ -elementowych  $(X_{i,1}, X_{i,2}, \dots, X_{i,n})$  dla  $i = 1, \dots, m_0$ , z rozkładu  $N(0, 1)$ , oraz  $m_1$  prób

### Rozważane modele (rodzaje testów)

W niniejszym rozdziale przedstawimy trzy przykłady testów parametrycznych, które wykorzystamy w dalszych częściach pracy do celów symulacji odnoszących się do zagadnień dotyczących procedur jednoczesnej weryfikacji wielu hipotez statystycznych.

Będziemy zakładać, że spośród  $m$  testowanych hipotez  $m_0$  hipotez stanowią hipotezy prawdziwe, a  $m_1 = m - m_0$  hipotez to hipotezy fałszywe, natomiast prawdopodobieństwo ( $P$ ) jest wyznaczone przy ustalonej konfiguracji prawdziwych i fałszywych hipotez zerowych.

W przypadku dwóch pierwszych z prezentowanych modeli przyjmujemy, że:

- obserwujemy  $m$  prób  $n$ -elementowych:  $(X_{i,1}, X_{i,2}, \dots, X_{i,n})$  dla  $i = 1, \dots, m$ , o rozkładach odpowiednio:  $N(\mu_i, 1)$  dla  $i = 1, \dots, m$ ,

$n$ -elementowych  $(Y_{i,1}, Y_{i,2}, \dots, Y_{i,n})$  dla  $i = 1, \dots, m_1$ , z rozkładu  $N(b, 1)$ , gdzie  $b \neq 0$ .

2. Obliczamy  $m_0$  średnich próbkowych  $\overline{X}_{i,n}$  dla  $i = 1, \dots, m_0$ , odpowiadających  $m_0$  hipotezom prawdziwym oraz  $m_1 = m - m_0$  średnich próbkowych  $\overline{Y}_{i,n}$ , dla  $i = 1, \dots, m_1$ , odpowiadających  $m_1$  hipotezom fałszywym.
3. Wyznaczamy  $t_i(0)$  – realizacje statystyk testowych dla  $m_0$  hipotez prawdziwych, dane wzorem:  $t_i(0) = \sqrt{n}|\overline{X}_{i,n}|$  dla  $i = 1, \dots, m_0$ , oraz  $t_i(1)$  – realizacje statystyk testowych dla  $m_1 = m - m_0$  hipotez prawdziwych, dane wzorem:  $t_i(1) = \sqrt{n}|\overline{Y}_{i,n}|$  dla  $i = 1, \dots, m_1$ .
4. Obliczamy  $p$ -wartości  $p_0(i)$ ,  $p_1(i)$ , odpowiadające  $m_0$  hipotezom prawdziwym ( $m_1$  hipotezom fałszywym), według wzoru:
 
$$p_i(0) = 2 - 2\Phi(t_i(0))$$

$$p_i(1) = 2 - 2\Phi(t_i(1))$$
5. Zliczamy liczbę odrzuceń hipotez zerowych w obu grupach, korzystając z procedur: Bonferroniego, Holma, Hochberga, kontroli FWER na poziomie błędu  $\alpha = 0,05$  oraz procedury Benjaminiego i Hochberga, kontroli FDR na poziomie błędu  $\alpha = 0,05$ .

## Model 2

**Szczegółowy opis modelu 2.** Do weryfikacji hipotezy zerowej ( $H_i$ ) w modelu 2, na poziomie błędu  $\alpha$ , wykorzystujemy statystykę testową o wzorze  $T_i := T_{i,n} = \sqrt{n}|\overline{X}_{i,n}|/S_{i,n}(x)$

gdzie:  $\overline{X}_{i,n} = \frac{1}{n} \sum_{j=1}^n X_{i,j}$ ,

$$S_{i,n}(x) = \left\{ \frac{1}{n} \sum_{j=1}^n (X_{i,j} - \overline{X}_{i,n})^2 \right\}^{1/2}$$

Hipotezę  $H_i$  odrzucamy, jeśli  $T_i \geq t_{\alpha, n-1}$ , gdzie  $t_{\alpha, n-1}$  jest wartością krytyczną rzędu  $\alpha$  rozkładu  $t$ -Studenta o  $n - 1$  stopniach swobody, tzn. wartością taką, że  $F_{t_{n-1}}(t_{\alpha, n-1}) = 1 - \alpha/2$ , gdzie  $F_{t_{n-1}}$  oznacza dystrybuantę rozkładu  $t$ -Studenta o  $n - 1$  stopniach swobody. Odpowiednie  $p$ -wartości dla powyższych testów wyznaczamy z relacji:

$$p_i = P(T_i \geq t_i | H_i) =$$

$$= P\left(\sqrt{n}|\overline{X}_{i,n}|/S_{i,n}(x) \geq t_i | H_i\right) =$$

$$= \left(P\sqrt{n}|\overline{X}_{i,n}|/S_{i,n}(x) \geq t_i | \overline{X}_{i,n} \sim \right.$$

$$\left. \sim N(0,1)\right) = 2 - 2F_{t_{n-1}}(t_i)$$

**Algorytm do modelu 2.** Poniżej przedstawiamy algorytm, który wykorzystamy później w symulacjach dotyczących procedur jednoczesnego testowania hipotez rozważanych powyżej.

1. Identyczny jak krok 1 w algorytmie dotyczącym modelu 1.
2. Identyczny jak krok 2 w algorytmie dotyczącym modelu 1.
3. Wyznaczamy  $t_i(0)$  – realizacje statystyk testowych dla  $m_0$  hipotez prawdziwych, dane wzorem  $t_i(0) = \sqrt{n}|\overline{X}_{i,n}|/S_{i,n}(x)$  dla  $i = 1, \dots, m_0$ , oraz  $t_i(1)$  – realizacje statystyk testowych dla  $m_1 = m - m_0$

hipotez prawdziwych, dane wzorem: gdzie:

$$t_i(1) = \sqrt{n} \left| \overline{Y_{i,n}} \right| / S_{i,n}(y) \quad \text{dla } i = 1, \dots, m_1.$$

4. Obliczamy  $p$ -wartości  $p_i(0)$ ,  $p_i(1)$ , odpowiadające  $m_0$  hipotezom prawdziwym ( $m_1$  hipotezom fałszywym), według wzoru:

$$p_i(0) = 2 - 2F_{t_{n-1}}(t_i(0))$$

$$p_i(1) = 2 - 2F_{t_{n-1}}(t_i(1))$$

5. Identyczny jak krok 5 w algorytmie dotyczącym modelu 1.

$$\overline{X_{i,\cdot}} = \frac{1}{n} \sum_{j=1}^n X_{i,j},$$

$$\overline{Y_{i,\cdot}} = \frac{1}{n} \sum_{j=1}^n Y_{i,j},$$

$$[S_{i,\cdot}(x)]^2 = \frac{1}{n} \sum_{j=1}^n (X_{i,j} - \overline{X_{i,\cdot}})^2,$$

$$[S_{i,\cdot}(y)]^2 = \frac{1}{n} \sum_{j=1}^n (Y_{i,j} - \overline{Y_{i,\cdot}})^2.$$

### Model 3

**Szczegółowy opis modelu 3.** Badaemy dwie populacje. Oznaczmy je przez  $A$ ,  $B$ . W przypadku populacji  $A$  rozważamy  $n$  wektorów  $m$ -wymiarowych  $(X_{1,j}, X_{2,j}, \dots, X_{m,j})$  dla  $j = 1, \dots, n$ , o  $m$ -wymiarowym rozkładzie normalnym  $N_m(\mu_A, \Sigma)$ , natomiast w przypadku populacji  $B$  rozważamy  $n$  wektorów  $m$ -wymiarowych  $(Y_{1,j}, Y_{2,j}, \dots, Y_{m,j})$  dla  $j = 1, \dots, n$ , o  $m$ -wymiarowym rozkładzie normalnym  $N_m(\mu_B, \Sigma)$ , gdzie:  $\mu_A = (\mu_{1,A}, \mu_{2,A}, \dots, \mu_{m,A})$  oraz  $\mu_B = (\mu_{1,B}, \mu_{2,B}, \dots, \mu_{m,B})$  są  $m$ -wymiarowymi wektorami średnich, a  $\Sigma$  oznacza wspólną macierz kowariancji.

Testujemy jednocześnie  $m$  hipotez zerowych  $H_i: \mu_{i,A} = \mu_{i,B}$ , wobec  $m$  hipotez alternatywnych  $H_i: \mu_{i,A} \neq \mu_{i,B}$  dla  $i = 1, \dots, m$ . Do weryfikacji hipotezy zerowej ( $H_i$ ) w modelu 3, na poziomie błędów  $\alpha$ , wykorzystujemy statystykę testową  $t$ -Studenta, stosowaną przy weryfikacji hipotez o dwóch średnich, daną wzorem:

$$T_{i,n} = \frac{\overline{X_{i,\cdot}} - \overline{Y_{i,\cdot}}}{\sqrt{\frac{[S_{i,\cdot}(x)]^2}{n} + \frac{[S_{i,\cdot}(y)]^2}{n}}}$$

Hipotezę  $H_i$  odrzucamy, jeśli  $|T_i| \geq t_{\alpha, 2n-2}$ , gdzie  $t_{\alpha, 2n-2}$  jest wartością krytyczną rzędu  $\alpha$  rozkładu  $t$ -Studenta o  $2n - 2$  stopniach swobody, tzn. wartością taką, że  $F_{t_{2n-2}}(t_{\alpha, 2n-2}) = 1 - \alpha/2$ , gdzie  $F_{t_{2n-2}}$  oznacza dystrybuantę rozkładu  $t$ -Studenta o  $2n - 2$  stopniach swobody. Odpowiednie  $p$ -wartości dla powyższych testów wyznaczamy na podstawie relacji

$$p_i = P(|T_i| \geq |t_i| | H_i) = 2 - 2F_{t_{2n-2}}(|t_i|).$$

**Algorytm do modelu 3.** Poniżej proponujemy algorytm, z którego będziemy korzystać w naszych analizach poświęconych zagadnieniom jednoczesnej weryfikacji hipotez rozważanych powyżej.

1. Generujemy  $n$ -elementową próbę z  $m$ -wymiarowego rozkładu normalnego  $N_m(\mu_A, \Sigma)$ , gdzie:

$$\mu_A = \left( \underbrace{1, \dots, 1}_{m_0}, \underbrace{1, \dots, 1}_{m_1} \right), \quad \Sigma = (\sigma_{i,j})$$

dla  $i = 1, \dots, n$ ,  $j = 1, \dots, m$ ,

$$\sigma_{i,j} = \begin{cases} 1 & \text{dla } i = j \\ q & \text{dla } i \neq j \end{cases}$$

oraz  $n$ -elementową próbę z  $m$ -wymiarowego rozkładu normalnego

$N_m(\mu_B, \Sigma)$ , gdzie:

$$\mu_B = \left( \underbrace{1, \dots, 1}_{m_0}, \underbrace{2, \dots, 2}_{m_1} \right), \Sigma(\sigma_{i,j})$$

– określone jak poprzednio.

- Obliczamy wartości średnich  $\overline{X_{i,\cdot}}$ ,  $\overline{Y_{i,\cdot}}$  oraz wariancji  $[S_{i,\cdot}(x)]^2$ ,  $[S_{i,\cdot}(y)]^2$ .
- Wyznaczamy realizacje  $t_i(0)$  dla  $i = 1, \dots, m_0$  statystyk testowych ( $T_i$ ) dla  $m_0$  prawdziwych hipotez zerowych, oraz realizacje  $t_i(1)$  dla  $i = 1, \dots, m_1$ , statystyk testowych  $T_i$  dla  $m_1$  fałszywych hipotez zerowych.
- Wyznaczamy  $p$ -wartości  $p_i(0)$ ,  $p_i(1)$ , odpowiadające  $m_0$  hipotezom prawdziwym ( $m_1$  hipotezom fałszywym), według wzoru:
 
$$p_i(0) = 2 - 2_{t_{2n-2}}(t_i(0))$$

$$p_i(1) = 2 - 2_{t_{2n-2}}(t_i(1))$$
- Identyczny jak krok 5 w algorytmach dotyczących modeli 1, 2.

jest  $m_0 = 0,4 \cdot 3000$  hipotez prawdziwych,

- ustalamy następującą wartość parametru  $b$ :  $b = 1$ ,
- przyjmujemy następującą wartość błędu  $\alpha$ :  $\alpha = 0,05$ .

Oto kod algorytmu do modelu 1:

```
n <- 30
m <- 3000
m0 <- 0.4*m
m1 <- m-m0
b <- 1
alpha <- 0.05
x0 <- as.vector(m0)
t0 <- as.vector(m0)
p0 <- as.vector(m0)
x1 <- as.vector(m1)
t1 <- as.vector(m1)
p1 <- as.vector(m1)
v0 <- as.vector(4)
v1 <- as.vector(4)
for (i in 1:m0){
  x0[i] <- mean(rnorm(n))
  for (j in 1:m1){
    x1[j] <- mean(rnorm(n,b,1))
  }
  t0 <- sqrt(n)*abs(x0)
  t1 <- sqrt(n)*abs(x1)
  p0 <- 2-2*pnorm(t0)
  p1 <- 2-2*pnorm(t1)
  g0 <- mt.rawp2adjp(p0, proc
=c("Bonferroni","Holm","Hochberg",
"B"))
  g1 <- mt.rawp2adjp(p1, proc
=c("Bonferroni","Holm","Hochberg",
"B"))
  h0 <- g0[1]$adjp
  h1 <- g1[1]$adjp
  for (k in 1:4){
    v0[k] <- mt.reject(h0[,k+1],alpha)[1]
    v1[k] <- mt.reject(h1[,k+1],alpha)[1]
  }
  print(v0)
  print(v1)
}
```

## Kody algorytmów w terminach języka R

### Kod algorytmu do modelu 1

Rozważając powyższy model, przyjęliśmy następujące założenia o wartościach parametrów:

- rozpatrujemy próbę o liczebności  $n = 30$ ,
- testujemy jednocześnie  $m = 3000$  hipotez zerowych, wśród których



## Kod algorytmu do modelu 2

Rozważając powyższy model, przyjęliśmy następujące założenia o wartościach parametrów:

- rozpatrujemy próbę o liczebności  $n = 60$ ,
- testujemy jednocześnie  $m = 3000$  hipotez zerowych, spośród których  $m_0 = 0,9 \cdot 3000$  to hipotezy prawdziwe,
- ustalamy następującą wartość parametru  $b$ :  $b = 0,5$ ,
- przyjmujemy następującą wartość błędu  $\alpha$ :  $\alpha = 0,05$ .

Oto kod algorytmu do modelu 2.

Parametry:  $n, m, m_0, m_1, b, \alpha$ , oraz wektory:  $x_0, x_1, t_0, t_1, p_0, p_1, v_0, v_1$ , deklarujemy według składni algorytmu do modelu 1. Dodatkowo deklarujemy nowe wektory:  $s_0$  – jako wektor  $m_0$ -wymiarowy, tzn.  $s_0 < -as.vector(m_0)$ , i  $s_1$  – jako wektor  $m_1$ -wymiarowy, tzn.  $s_1 < -as.vector(m_1)$ .

Wektory:  $x_0, s_0, x_1, s_1, t_0, t_1, p_0, p_1$ , określamy jak następuje:

```
for (i in 1:m0){
x0[i] <- mean(rnorm(n))
s0[i] <- sd(rnorm(n))}
for (j in 1:m1){
x1[j] <- mean(rnorm(n,b,1))
s1[j] <- sd(rnorm(n,b,1))}
t0 <- sqrt(n)*abs(x0)/s0
t1 <- sqrt(n)*abs(x1)/s1
p0 <- 2-2*pt(t0, df=n-1)
p1 <- 2-2*pt(t1, df=n-1)
```

Pozostałe linijki kodu, w szczególności te dotyczące definicji obiektów:  $g_0, g_1, h_0, h_1, v_0, v_1$ , są identyczne jak w algorytmie do modelu 1.

## Kod algorytmu do modelu 3

Rozważając powyższy model, przyjęliśmy następujące założenia o wartościach parametrów:

- rozpatrujemy próbę o liczebności  $n = 30$ ,
- testujemy jednocześnie  $m = 1000$  hipotez zerowych, wśród których  $m_0 = 0,2 \cdot 3000$  to hipotezy prawdziwe,
- ustalamy następującą wartość parametru  $q$ :  $q = 0,5$ ,
- przyjmujemy następującą wartość błędu  $\alpha$ :  $\alpha = 0,05$ ,

Oto kod algorytmu do modelu 3.

Parametry:  $n, m, m_0, m_1, q, \alpha$ , jak również:  $m$ -wymiarowe wektory:  $\mu, \mu_1$ ,  $m_0$ -wymiarowe wektory:  $x_{s0}, y_{s0}, s_{x0}, s_{y0}$ ,  $m_1$ -wymiarowe wektory:  $x_{s1}, y_{s1}, s_{x1}, s_{y1}$ , oraz 4-wymiarowe wektory:  $t_0, t_1, p_0, p_1, v_0, v_1$ , deklarujemy według składni algorytmu do modelu 1.

Kolejne instrukcje algorytmu przedstawiają się następująco:

```
sig <- matrix(m,m)
for (i in 1:m){mu[i]=1; if (i<=m0)
mul[i]=1 else mul[i]=2}
for (j in 1:m){
for (k in 1:m){if (j== k) sig[j,k]=1 else
sig[j,k]=q}}
a1 <- mvnrm(n,mu,sig)
b1 <- mvnrm(n,mul,sig)
for (r in 1:m0){
xs0[r] <- mean(a1[,r]);
ys0[r] <- mean(b1[,r]);
sx0[r] <- sd(a1[,r]);
sy0[r] <- sd(b1[,r])}
for (s in 1:m1){
xs1[s] <- mean(a1[,s+m0]);
ys1[s] <- mean(b1[,s+m0]);
```

```

sx1[s] <- sd(a1[,s+m0]);
sy1[s] <- sd(b1[,s+m0]);
t0 <- (xs0-ys0)/sqrt((sx0^2+sy0^2)/n)
t1 <- (xs1-ys1)/sqrt((sx1^2+sy1^2)/n)
t01 <- abs(t0)
t11 <- abs(t1)
n1 <- 2*n-2
p0 <- 2-2*pt(t01, df=n1)
p1 <- 2-2*pt(t11, df=n1)

```

Pozostałe linijki kodu, w szczególności te dotyczące definicji obiektów: g0, g1, h0, h1, v0, v1, są identyczne jak w algorytmie do modelu 1.

## Wyniki symulacyjne do opracowanych algorytmów

W rozdziale tym przedstawimy rezultaty symulacji dla algorytmów do modeli 1, 2, 3, otrzymane przy wykorzystaniu pakietu statystycznego R dla czterech wybranych przez nas procedur: Bonferroniego (B), Holma (H), Hochberga (Ho), Benjaminiego i Hochberga (B-H). Użyjemy następujących oznaczeń:  $n$  – liczność próby,  $m$  – liczba weryfikowanych hipotez,  $m_0$  – liczba hipotez prawdziwych,  $m_1$  – liczba hipotez fałszywych,  $V_0$  – liczba odrzuceń hipotez prawdziwych,  $V_1$  – liczba odrzuceń hipotez fałszywych,  $\alpha$  – miara błędu.

### Wyniki symulacyjne dla algorytmu dotyczącego modelu 1

**Przypadek 1.** Zakładamy, że:  $n = 30$ ,  $m = 3000$ ,  $b = 0,2$ ,  $\alpha = 0,05$ .

Tabela 1	B	H	Ho	B-H
$m_0 = 300, V_0$	0	0	0	0
$m_1 = 2700, V_1$	2	2	2	31
$m_0 = 900, V_0$	0	0	0	0

cd. tab. 1

$m_1 = 2100, V_1$	4	4	4	18
$m_0 = 1500, V_0$	0	0	0	0
$m_1 = 1500, V_1$	1	1	1	13
$m_0 = 2100, V_0$	0	0	0	0
$m_1 = 900, V_1$	1	1	1	4
$m_0 = 2700, V_0$	0	0	0	0
$m_1 = 300, V_1$	0	0	0	0

**Przypadek 2.** Zakładamy, że:  $n = 60$ ,  $m = 3000$ ,  $b = 0,2$ ,  $\alpha = 0,05$ .

Tabela 2	B	H	Ho	B-H
$m_0 = 300, V_0$	0	0	0	0
$m_1 = 2700, V_1$	12	12	12	300
$m_0 = 900, V_0$	0	0	0	0
$m_1 = 2100, V_1$	7	7	7	231
$m_0 = 1500, V_0$	0	0	0	0
$m_1 = 1500, V_1$	9	9	9	180
$m_0 = 2100, V_0$	0	0	0	0
$m_1 = 900, V_1$	4	4	4	87
$m_0 = 2700, V_0$	0	0	0	0
$m_1 = 300, V_1$	5	5	5	29

**Przypadek 3.** Zakładamy, że:  $n = 60$ ,  $m = 3000$ ,  $b = 0,5$ ,  $\alpha = 0,05$ .

Tabela 3	B	H	Ho	B-H
$m_0 = 300, V_0$	0	0	0	0
$m_1 = 2700, V_1$	947	1053	1053	2630
$m_0 = 900, V_0$	0	0	0	3
$m_1 = 2100, V_1$	776	880	880	2028
$m_0 = 1500, V_0$	0	0	0	0
$m_1 = 1500, V_1$	556	625	625	1456
$m_0 = 2100, V_0$	0	0	0	0
$m_1 = 900, V_1$	386	455	455	877
$m_0 = 2700, V_0$	1	1	1	1
$m_1 = 300, V_1$	150	175	175	290

### Wyniki symulacyjne dla algorytmu dotyczącego modelu 2

**Przypadek 4.** Zakładamy, że:  $n = 30$ ,  $m = 3000$ ,  $b = 0,2$ ,  $\alpha = 0,05$ .

Tabela 4	B	H	Ho	B-H
$m_0 = 300, V_0$	0	0	0	0
$m_1 = 2700, V_1$	0	0	0	0
$m_0 = 900, V_0$	0	0	0	0
$m_1 = 2100, V_1$	2	2	2	9
$m_0 = 1500, V_0$	0	0	0	0
$m_1 = 1500, V_1$	1	1	1	2
$m_0 = 2100, V_0$	0	0	0	0
$m_1 = 900, V_1$	0	0	0	0
$m_0 = 2700, V_0$	0	0	0	0
$m_1 = 300, V_1$	2	2	2	10

**Przypadek 5.** Zakładamy, że:  $n = 60$ ,  $m = 3000$ ,  $b = 0,2$ ,  $\alpha = 0,05$ .

Tabela 5	B	H	Ho	B-H
$m_0 = 300, V_0$	0	0	0	0
$m_1 = 2700, V_1$	4	4	4	176
$m_0 = 900, V_0$	0	0	0	0
$m_1 = 2100, V_1$	6	6	6	213
$m_0 = 1500, V_0$	0	0	0	0
$m_1 = 1500, V_1$	3	3	3	120
$m_0 = 2100, V_0$	0	0	0	0
$m_1 = 900, V_1$	3	3	3	74
$m_0 = 2700, V_0$	0	0	0	0
$m_1 = 300, V_1$	5	5	5	52

**Przypadek 6.** Zakładamy, że:  $n = 60$ ,  $m = 3000$ ,  $b = 0,5$ ,  $\alpha = 0,05$ .

Tabela 6	B	H	Ho	B-H
$m_0 = 300, V_0$	0	0	0	0
$m_1 = 2700, V_1$	617	681	681	2594
$m_0 = 900, V_0$	0	0	0	3
$m_1 = 2100, V_1$	536	595	595	2026

cd. tab. 6

$m_0 = 1500, V_0$	0	0	0	0
$m_1 = 1500, V_1$	426	489	489	1456
$m_0 = 2100, V_0$	0	0	0	0
$m_1 = 900, V_1$	290	330	330	866
$m_0 = 2700, V_0$	0	0	0	0
$m_1 = 300, V_1$	135	164	164	287

### Wyniki symulacyjne dla algorytmu dotyczącego modelu 3

**Przypadek 7.** Zakładamy, że:  $n = 30$ ,  $m = 1000$ ,  $q = 0,4$ ,  $\alpha = 0,05$ .

Tabela 7	B	H	Ho	B-H
$m_0 = 100, V_0$	0	0	0	0
$m_1 = 2900, V_1$	203	223	223	873
$m_0 = 300, V_0$	0	0	0	0
$m_1 = 2700, V_1$	230	280	280	694
$m_0 = 500, V_0$	0	0	0	0
$m_1 = 2500, V_1$	91	102	102	490
$m_0 = 700, V_0$	0	0	0	0
$m_1 = 2300, V_1$	255	300	300	300
$m_0 = 900, V_0$	0	0	0	0
$m_1 = 2100, V_1$	88	100	100	100

**Przypadek 8.** Zakładamy, że:  $n = 30$ ,  $m = 1000$ ,  $q = 0,6$ ,  $\alpha = 0,05$ .

Tabela 8	B	H	Ho	B-H
$m_0 = 100, V_0$	0	0	0	0
$m_1 = 2900, V_1$	128	138	138	898
$m_0 = 300, V_0$	0	0	0	3
$m_1 = 2700, V_1$	72	82	82	695
$m_0 = 500, V_0$	0	0	0	0
$m_1 = 2500, V_1$	306	429	500	500
$m_0 = 700, V_0$	0	0	0	0
$m_1 = 2300, V_1$	153	210	300	300
$m_0 = 900, V_0$	0	0	0	0
$m_1 = 2100, V_1$	95	100	100	100

**Przypadek 9.** Zakładamy, że:  $n = 30$ ,  $m = 1000$ ,  $q = 0,8$ ,  $\alpha = 0,05$ .

Tabela 9	B	H	Ho	B-H
$m_0 = 100, V_0$	0	0	0	0
$m_1 = 2900, V_1$	134	170	900	900
$m_0 = 300, V_0$	0	0	0	3
$m_1 = 2700, V_1$	517	700	700	700
$m_0 = 500, V_0$	0	0	0	0
$m_1 = 2500, V_1$	460	500	500	500
$m_0 = 700, V_0$	0	0	0	0
$m_1 = 2300, V_1$	275	300	300	300
$m_0 = 900, V_0$	0	0	0	0
$m_1 = 2100, V_1$	100	100	100	100

## Wnioski

W tej części pracy przedstawimy spostrzeżenia, uzyskane na podstawie dokonanych przez nas symulacji.

W przypadkach dotyczących modelu 1 obserwacje te są następujące. Otóż gdy  $b = 0,5$ , a wielkość próby  $n = 60$ , bardzo wysoką moc wykazały procedury Holma, Hochberga oraz Benjaminiego i Hochberga, natomiast procedura Bonferoniego okazała się procedurą najslabszą, przy czym moc procedury Benjaminiego i Hochberga zdecydowanie przewyższała moce pozostałych procedur. Tendencje te dostrzegamy również w sytuacji, gdy  $b = 0,2$ . Dodatkowo obserwujemy, iż wraz ze zmniejszeniem wartości parametru  $b$ , maleje moc rozpatrywanych procedur. Z kolei zmniejszając w modelu 1 rozmiar próby do  $n = 30$ , widzimy znaczną obniżkę mocy rozważanych procedur i słabszą przewagę mocy procedury Benjaminiego i Hochberga (w sensie liczby odrzuceń hipotez fałszywych). Podobne

tendencje do opisanych powyżej można dostrzec również w przypadku wyników symulacji dotyczących modelu 2 (przy czym poszczególne procedury wykazywały tu, ogólnie rzecz biorąc, słabszą moc). Z kolei w większości przypadków odnoszących się do modelu 3, zaobserwowaliśmy silną dominację procedury Benjaminiego i Hochberga nad pozostałymi, z wyjątkiem sytuacji, gdy  $n = 30$  i  $q = 0,8$ , kiedy to procedura Benjaminiego i Hochberga oraz procedura Hochberga zachowywały się niemal identycznie.

## Literatura

- BENJAMINI Y., HOCHBERG Y. 1995: Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. Ser. B* 57: 289–300.
- DUDOIT S., SHAFFER J.P., BOLDRICK J.C. 2003: Multiple hypothesis testing in microarray experiments. *Statistical Science* 18 (1): 71–103.
- HOLM S. 1979: A simple sequentially rejective multiple test procedure. *Scand. Journal. Statist.* 6: 65–70.
- KOMSTA Ł. 2004: Wprowadzenie do środowiska R. Opracowanie dostępne na stronie internetowej <http://www.komsta.net>.
- LEHMANN E.L. 1957: A theory of some multiple decision problems. I. *Ann. Statist.* 28: 1–25.
- LEHMANN E.L., ROMANO J.P. 2005: Generalizations of the familywise error rate. *Ann. Statist.* 33: 1138–1154.
- POLLARD K.S., van der LAAN M.J., BIRKNER M.D., DUDOIT S. 2005: Test statistics null distributions in multiple testing: simulation studies and applications to genomics. U. C. Berkeley division of biostatistics working paper series 184.
- STOREY J. 2003: The positive false discovery rate: A Bayesian interpretation and the  $q$ -value. *Ann. Statist.* 31: 2013–2035.

## Summary

**Application of the R-project package in the area of the multiple-comparison procedures.** In our paper, we give the simulation results concerning the multiple-comparison procedures, especially their error measures, such as FWER and FDR. We compare the powers of four popular multiple-comparison procedures, applied to three standard statistical models. All of our simulations were carried out by using the R-project package.

### Authors' addresses:

Marcin Dudziński  
Szkoła Główna Gospodarstwa Wiejskiego  
Katedra Ekonometrii i Statystyki  
ul. Nowoursynowska 166, 02-787 Warszawa  
Poland  
e-mail: mdudzinski@mors.sggw.waw.pl

Konrad Furmańczyk  
Szkoła Główna Gospodarstwa Wiejskiego  
Katedra Zastosowań Matematyki  
ul. Nowoursynowska 159, 02-776 Warszawa  
Poland  
e-mail: konfur@wp.pl